

370



АКАДЕМИЯ НАУК СССР

# ФИЗИОЛОГИЯ РЕЧИ

## ВОСПРИЯТИЕ РЕЧИ ЧЕЛОВЕКОМ

А К А Д Е М И Я    Н А У К    С С С Р  
НАУЧНЫЙ СОВЕТ ПО КОМПЛЕКСНЫМ ПРОБЛЕМАМ  
ФИЗИОЛОГИИ ЧЕЛОВЕКА И ЖИВОТНЫХ

---

РУКОВОДСТВО  
ПО ФИЗИОЛОГИИ

ФИЗИОЛОГИЯ  
РЕЧИ

ВОСПРИЯТИЕ  
РЕЧИ  
ЧЕЛОВЕКОМ



---

ИЗДАТЕЛЬСТВО «НАУКА»  
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ  
ЛЕНИНГРАД • 1976

вно  
лю  
ечь,  
звек  
, что  
жи-  
пре-  
от  
ную

ели,  
ний  
и и  
пра-  
оль-  
зала  
све-  
зна-  
ния  
ихо-  
гать  
рия-  
ком

ться  
отся  
вно  
до-

## Авторы:

Л. А. ЧИСТОВИЧ, А. В. ВЕНЦОВ, М. П. ГРАНСТРЕМ, С. Я. ЖУКОВ,  
М. Г. ЖУКОВА, Э. Г. КАРНИЦКАЯ, В. А. КОЖЕВНИКОВ, Д. М. ЛИСЕНКО,  
В. В. ЛЮБЛИНСКАЯ, В. Н. МУШНИКОВ, Н. А. СЛЕПОКУРОВА,  
Н. А. ФЕДОРОВА, Р. Х. ХААВЕЛЬ, И. А. ЧИСТОВИЧ, В. С. ШУПЛЯКОВ.

УДК 612.821.8 : 612.789 : 534.75

Физиология речи. Восприятие речи человеком. (Чистович Л. А., Венцов А. В., Гранстрем М. П. и др.). В серии «Руководство по физиологии». Л., «Наука», 1976. 388 с.

Книга посвящена рассмотрению процессов обработки речевого сигнала слуховой системой и мозгом человека. В первом разделе описываются свойства речевого сигнала и обсуждаются проблемы понимания смысла устного сообщения. Во втором разделе излагаются данные и теоретические представления относительно сегментации речевого потока, фонемной классификации звуков речи и восприятия ритмико-мелодических признаков речевых последовательностей. В третьем разделе описываются результаты исследований и моделирования периферического слухового анализа и рассматриваются новейшие физиологические и психоакустические данные об обработке сигнала в слуховой системе. Ил. — 205, табл. — 6, лит. — 551 назв.

## Редакционная коллегия Руководства:

Н. П. Бехтерева (зам. отв. редактора), В. А. Кисляков, Г. П. Конради,  
П. Г. Костюк, К. А. Ланге (отв. секретарь), Б. Ф. Ломов, В. И. Медведев,  
А. Л. Поленов, Т. Н. Турпаев, А. М. Уголев, В. Н. Черниговский  
(отв. редактор), А. И. Шаповалов, М. М. Хананашвили.

## Редакционная коллегия тома:

А. В. Венцов, В. А. Кожевников, Д. М. Лисенко,  
Л. А. Чистович (отв. редактор).

## ПРЕДИСЛОВИЕ

На протяжении последних 10—15 лет постепенно сформировалось научное направление, ставящее своей целью разработку системы, способной воспринимать и понимать речь, причем делать это так, чтобы общающийся с этой системой человек не мог бы даже догадаться, что его собеседник — машина. Ясно, что такая система может быть лишь функциональной моделью живого прототипа, ее задачей является только воспроизведение преобразований информации, осуществляемых в живой системе; отнюдь не предполагается, что она должна повторять конкретную «конструкцию» живой системы.

Так как обязательное требование, предъявляемое к модели, состоит в том, что она должна работать, первый этап исследований был направлен на анализ самой проблемы понимания речи и поиски общей структуры модели, способной, в принципе, справиться с поставленными задачами. Основные сведения, использованные на этом этапе работы, касались свойств речевого сигнала и организации языка. Ограничения, накладываемые этими сведениями на возможную структуру модели, оказались столь значительными, что вначале предполагалась возможность решения проблемы вообще без обращения к данным физиологии и психологии. Лишь за последние годы стало ясным, что пренебрегать данными этих наук и экспериментальным исследованием восприятия и понимания речи человеком вряд ли возможно и, во всяком случае, нецелесообразно.

При разработке модели неизбежно приходится пользоваться рядом допущений. Некоторые из них очевидны, другие являются спорными. Именно в случае спорных допущений естественно обратиться к исследованию живой системы и попытаться эти допущения проверить.



При разработке модели возникает ряд вариантов решения одной и той же задачи. Если разрабатываемая система отличается большой сложностью, проверить, к чему приведет выбор того или иного варианта, очень трудно, а иногда и практически невозможно до тех пор, пока система не будет полностью реализована. В этом случае «дешевле» обратиться к исследованию живого прототипа и попытаться выяснить, какой из вариантов лучше согласуется с полученными экспериментальными фактами.

Легко заметить, что в данном случае наблюдается весьма непривычное для физиолога и психолога отношение между теорией и экспериментальными фактами. Физиолог и психолог привыкли считать, что сначала должно быть как можно больше фактов, потом уже можно строить на их основании теорию, т. е. попытаться дать некоторое обобщенное описание этих фактов. Здесь же как будто бы все поставлено с ног на голову — ведущим является изобретатель системы, эксперименты нужны лишь для ограничения его фантазии. На самом деле все, конечно, обстоит далеко не так. Изобретатель системы с самого начала был отнюдь не свободен в своих фантазиях. Он начинал свою работу с фактов, которые уже были обобщены с помощью более или менее формализованных теорий. Только эти факты относились не к физиологии и психологии, а к другим областям науки — акустике речи и лингвистике.

Для работ данного направления функциональная модель восприятия и понимания речи и теория этих процессов — просто одно и то же. Соответственно, целью экспериментального психологического и физиологического изучения является разработка теории или, что то же самое, уточнение структуры и определение параметров этой функциональной модели.

Для того чтобы очертить круг вопросов, рассматриваемых в настоящей книге, необходимо коротко остановиться на том, какова, по распространенным сейчас представлениям, общая структура полной модели восприятия и понимания речи.

Принимается, что полная модель состоит из трех последовательно соединенных моделей. Первая из них, она обычно называется моделью восприятия, осуществляет преобразование поступающего на ее вход акустического речевого сигнала в последовательность фонетических элементов. Модель включает блок (блоки) слухового анализа сигнала и блок фонетической интерпретации. Сведения о языке, содержащиеся в блоке фонетической интерпретации, еще очень ограничены и касаются фонетики языка.

Упрощенно говоря, модель умеет переводить воспринятый ею акустический речевой сигнал в артикуляторные инструкции-указания о том, как нужно произнести то, что модель «услышала». Эта модель не знает ни словарного состава языка, ни его грамматики и, тем более, не «понимает» смысла услышанного. Вторая модель преобразует последовательность фонетических элементов в описание смысла фразы. Она осуществляет морфологический анализ и синтаксический анализ, используя для этого словарь (словари) и грамматические правила. Короче говоря, это действующая модель анализирующей части данного языка. Описание смысла, получаемое на выходе модели, является описанием тех сведений о «действительности», которые содержались в проанализированной фразе. Это описание таково, что по нему уже нельзя установить, на каком языке была произнесена исходная фраза. Третья модель — ее разработка в настоящее время только начинается — занимается интерпретацией и оценкой полученных сведений о событиях, явлениях и т. д. Она решает, являются эти сведения истинными или ложными, важными или безразличными, что нужно предпринять в результате их получения и т. д. Иначе говоря, модель делает какую-то часть из того, что обозначается как интеллектуальная деятельность.

Уже по характеру задач, решаемых разными моделями, отчетливо видно, что их разработкой занимаются специалисты совершенно разного профиля, т. е. разные модели относятся к компетенции разных областей науки.

По этой причине сейчас приходится сделать допущение, хотя оно, возможно, и несколько рискованно, что эти модели являются чисто последовательными. Другими словами, принимается, что первая модель не получает никакой информации с выходов второй и третьей моделей, а вторая модель ничего не знает о том, что решает третья модель. В такой ситуации главный вопрос «стыковки» моделей заключается в согласовании выхода модели предыдущего уровня со входом модели следующего уровня. Конкретно речь идет о том, чтобы задаться описанием последовательности фонетических элементов (какая информация должна в ней содержаться и как она должна быть представлена) и задаться описанием смысла.

Ясно, что разработка функциональных моделей требует обязательного четкого определения того, что является сигналом на входе и что необходимо получить на выходе. При рассмотрении

вопроса о стыковке моделей, естественно, приходится исходить, с одной стороны, из того, какое входное описание необходимо для модели следующего уровня, и, с другой стороны, какое описание реально можно получить на выходе модели предыдущего уровня.

В настоящей книге рассматриваются экспериментальные данные и теоретические вопросы, касающиеся только первой из этих трех моделей, определяемой как модель восприятия.

Две первые главы являются вводными. В первой главе даются элементарные сведения об акустических свойствах речевого сигнала и приводится краткое и весьма схематизированное изложение основных идей, использовавшихся при разработке систем автоматического фонемного распознавания речи. Во второй главе обсуждается вопрос стыковки модели восприятия речи с моделью следующего уровня, осуществляющей морфологический и синтаксический анализ фразы.

Главы 3—6 посвящены вопросам фонетической интерпретации речевого сигнала, а главы 7—12 рассматривают проблемы предварительной слуховой обработки этого сигнала.

Интерес к слуховой обработке речевого сигнала и стремление разобраться в том, в какой мере теоретические представления относительно фонетической интерпретации речевого сигнала согласуются с представлениями и данными физиологии слуха и психоакустики, в значительной мере обусловлены научными традициями коллектива, к которому принадлежат авторы настоящей книги. Исходное ядро этого коллектива было образовано из учеников основателя современной советской физиологии сенсорных систем — Григория Викторовича Гершуни. Г. В. Гершуни внушил своим ученикам и последователям убеждение в том, что ни нейрофизиология, ни психоакустика, ни экспериментальная психология (или фонетика) не смогут привести к удовлетворительному пониманию принципов обработки информации мозгом, если они будут развиваться как внутренне замкнутые, самостоятельные в теоретическом плане дисциплины. Г. В. Гершуни одним из первых в мире понял также и то, что исследование восприятия речевых и других естественных звуковых сигналов поставит совершенно новые проблемы перед физиологией слуха и психоакустикой и потребует в конечном итоге существенного пересмотра теоретических представлений, сформировавшихся в этих областях.

Работая над настоящей книгой, авторы пытались решить вполне определенную задачу. Она состояла в том, чтобы выяснить,

какие ограничения на возможную структуру или, еще лучше, параметры тех или иных блоков функциональной модели восприятия речи накладываются экспериментальными данными, полученными при исследовании слуха и исследовании восприятия речевых и речеподобных сигналов.

При такой постановке задачи круг рассматриваемых экспериментальных фактов оказался достаточно ограниченным. В книге не обсуждаются данные, касающиеся обработки информации на более высоких уровнях полной модели восприятия и понимания речи, в ней нет также описания данных по очень популярной в настоящее время проблеме анатомической локализации «речевых функций».

## ПОЯСНЕНИЯ К ТРАНСКРИПЦИИ

В настоящей книге для обозначения звуков использовались знаки международной фонетической транскрипции. Следует, однако, отметить, что при обозначении синтетических звуков не преследовалась цель отразить их звучание наиболее точно. В этих случаях символика употреблялась для обозначения не фонетического качества отдельного звука, а для обозначения способа интерпретации слушателями целого множества сигналов. Таким образом, транскрипция была не столько фонетической, сколько фонематической.

Особых пояснений требуют следующие обозначения. Знак [i] был принят для обозначения русского гласного *ы*; для обозначения мягкости согласного использовался штрих, расположенный вверху справа от основного символа, например [d'].

При цитировании работ зарубежных авторов сохранялась символика оригинала.



## РЕЧЕВОЙ СИГНАЛ И ПРОБЛЕМЫ ЕГО ОПИСАНИЯ

### Глава I

#### СВОЙСТВА РЕЧЕВОГО СИГНАЛА И НЕКОТОРЫЕ ВОПРОСЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

Представление о современных подходах к описанию свойств речевого сигнала дает рис. 1.1, весьма обобщенно и условно показывающий основные типы его обработки в уже используемых или еще разрабатываемых технических системах.

Определим исходный речевой сигнал как функцию  $p(t)$ , представляющую изменения звукового давления, происходящие в течение того времени, когда говорящий произносит фразу. Функция  $p(t)$  может быть

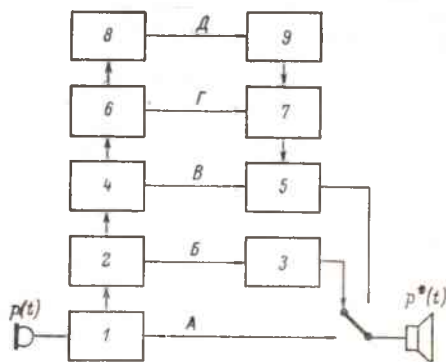


Рис. 1.1. Соотношение основных типов преобразований речевого сигнала в технических системах.

1 — линейное усиление; 2 — спектральное описание; 3 — синтез по спектральному описанию; 4 — параметрическое описание; 5 — функциональная модель речевого аппарата; 6 — фонетическое описание; 7 — синтез сигналов управления; 8 — описание смысла; 9 — синтез фраз. Остальные обозначения см. в тексте.

выражена электрическим напряжением или другой физической величиной, может быть зарегистрирована тем или иным способом. Ясно, что во всех случаях  $p(t)$ , определенная на отрезке фразы, полностью характеризует последнюю. При таком определении речь представляется множеством функций  $p(t)$ , соответствующих разным фразам, произнесенным множеством дикторов.

Будем считать, что существуют процедуры обработки, в результате которых функция  $p(t)$  может быть представлена некоторым набором величин. Эти величины являются характеристиками

сигнала по определенным признакам, вся их совокупность представит описание сигнала. Задачу обычных каналов речевой связи (уровень *A*, рис. 1.1) можно трактовать как передачу полных описаний  $p(t)$ . Известно, что любой колебательный процесс может быть представлен без искажений последовательностью дискретных величин, следующих с частотой  $2F$ , где  $F$  — максимальная частота спектральных составляющих сигнала. Последовательность этих величин при условии, что каждая из них точно воспроизводит значение соответствующей мгновенной амплитуды, составит так называемое полное описание сигнала.

Считается, что для получения полного описания речевого сигнала (обладающего качеством сигнала, переданного по телефону) необходимо производить не менее 8000 отсчетов в секунду и каждый из отсчетов должен иметь точность, соответствующую представлению амплитуды по крайней мере 128 уровнями. Объем описания составит 56 000 бит в секунду [121].

Для нас, однако, наибольший интерес имеют описания  $p(t)$ , результатом которых является получение нового представления речевого сообщения, заведомо более сжатого по сравнению с исходным.<sup>1</sup> Будем считать, что эти описания являются обратимыми, т. е. по ним можно синтезировать новые функции  $p^*(t)$ , сохраняющие определенные свойства исходного сигнала. Считаем также, что качество и особенности восстановленного сигнала можно оценить, предъявив  $p^*(t)$  слушателю. Соотношение процедур синтеза с процедурами описания в общих чертах может быть понято на основании рис. 1.1.

Наиболее распространенным типом сокращенного описания  $p(t)$  является спектральное описание (уровень *B*, рис. 1.1), выражающееся в спектре амплитуд (значения интенсивности частотных составляющих сигнала в зависимости от частоты) и спектре фаз (значения фаз составляющих сигнала в зависимости от частоты). Результаты спектрального анализа зависят от времени наблюдения сигнала. Получение формально точного спектра требует бесконечно долгого наблюдения.

Длительные наблюдения имеет смысл производить только в случае стационарных сигналов, речевой же сигнал по своей природе представляет колебательный процесс, у которого и форма, и периоды волн изменяются довольно быстро и на всем протяжении сигнала. В сигнале реально отсутствуют стационарные участки, и он как бы является непрерывной последовательностью переходных процессов.

Получить спектральное описание подобного процесса, удовлетворяющее требованиям практики (из них принципиальное —

<sup>1</sup> Количественные оценки степени сжатия описания, достигаемой с помощью различных преобразований сигнала, можно найти в [121, 137]. Оценки эти, однако, основываются на формальных расчетах, известных в теории информации, и вряд ли могут быть применены для таких уровней, как описание смысла сообщения.

восстановление  $p^*(t)$  без существенных задержек по сравнению со временем поступления  $p(t)$ , возможно только с помощью методов, обеспечивающих быстрое получение мгновенных спектров [142]. В таких методах время наблюдения сигнала достаточно мало и само спектральное изображение является функцией времени. В использующих этот метод приборах — «динамических спектрографах» — дело ограничивается получением «текущих» во времени изображений относительной энергии частотных составляющих сигнала. Фазовый спектр теряется.

Для речевых исследований особое значение имеет рассматриваемый в данной главе метод динамической спектрографии, носящий название «видимая речь». Примечательно, что принцип метода в определенной степени согласуется с теми преобразованиями, которые, как известно, происходят в периферических отделах слуховой системы (быстрый спектральный анализ с почти полной потерей фазовой информации; преобразование частотной шкалы в шкалу координат вдоль оси улитки внутреннего уха). Однако основания выбора конкретных характеристик приборов типа «видимая речь» являются довольно условными [330]. Исследования частотного анализа сигналов, осуществляемого слуховой системой (см. главу 7), могут послужить основой для их уточнения.

Еще более сжатое описание (уровень  $B$ , рис. 1.1) основывается на результатах изучения физических процессов, происходящих в речевом аппарате. При образовании акустического речевого сигнала имеют место явления двух основных типов: собственно создание звуковых колебаний и изменение спектра этих колебаний в воздушных полостях речевого тракта, действующих как акустические фильтры.

Существо процессов может быть представлено функциональной моделью речевого аппарата. Состояния модели, соответствующие созданию определенных речевых звуков, описываются небольшим числом параметров, характеризующих работу генераторов звуковых колебаний и передаточную функцию речевого тракта. Соответственно в пределах набора этих же величин составляется параметрическое описание звукового речевого сигнала. В первом приближении параметрическое описание является как бы сокращенным описанием по наиболее важным признакам картины, наблюдаемой на изображении спектрографов «видимая речь». Реально же для получения параметрического описания речевых сигналов применяются довольно сложные измерительные и вычислительные процедуры [35, 121, 137].

Уровень  $\Gamma$  рис. 1.1 соответствует фонетическому описанию, которое предполагает изображение речевого сигнала в виде дискретной последовательности символов, набор которых весьма ограничен. Для создания фонетического описания необходимо применение особых процедур, прежде всего процедуры фонемной классификации последовательных участков речевого сигнала и процедуры сегментации — разделения непрерывного речевого сиг-

нала на отрезки, в пределах которых должны приниматься решения о их фонемной принадлежности.

Следует отметить, что в фонетическом описании должны быть отражены также и так называемые просодические характеристики речи (см. главу 5), которые, в частности, указывают, выражает ли фраза вопрос, восклицание, выделяются ли в ней определенные слова и т. п. Решения о фонетическом описании сигнала не могут быть приняты мгновенно, производящая описание система должна работать над более или менее протяженным изображением речевого сигнала и соответственно должна обладать оперативной памятью.

В технике автоматического распознавания речи известны попытки создания систем, производящих фонетическое описание как на основании спектральных изображений речевого сигнала, так и на основании его параметрического описания. В фонетических и психологических работах, связанных с исследованием восприятия речи, обычно используются параметрические описания. Вопрос о том, каким описанием сигнала реально пользуется человек, осуществляющий фонетическую интерпретацию сигнала, является одним из главных предметов обсуждения в настоящей книге.

Существенной особенностью  $p^*(t)$ , синтезированной на основании фонетического описания  $p(t)$ , является то, что она в принципе не содержит сведений об индивидуальности диктора. В ней сохраняется информация только о том, что сказано и на каком языке.

Уровень описания смысла речевого сообщения (уровень  $D$ , рис. 1.1) обеспечивает еще более общее представление содержания  $p(t)$ . Описание это не должно быть связано со словарем и грамматикой языка, на котором было сделано исходное сообщение. Вопрос о необходимости таких описаний возник при разработке проблемы автоматического перевода. Некоторых аспектов подхода к описанию смысла касается глава 2.

## 1.1. ЭЛЕМЕНТЫ ТЕОРИИ РЕЧЕОБРАЗОВАНИЯ

Акустический речевой сигнал возникает в результате сложных координированных движений, происходящих в ряде органов, вся совокупность которых и называется речевым аппаратом (рис. 1.2, А). Легкие со всей дыхательной мускулатурой обеспечивают развитие давлений и возникновение воздушных потоков в речевом тракте. Последний (рис. 1.2, Б, В) представляется гортанью и рядом воздушных полостей, конфигурация которых существенно изменяется в процессе речеобразования. Ведущую роль играют движения небной занавески, языка, губ и нижней челюсти.

Механизмы возбуждения акустических колебаний связаны либо с работой гортани, либо с возникновением шумных или импульсных звуков при прохождении воздушного потока через сужения,

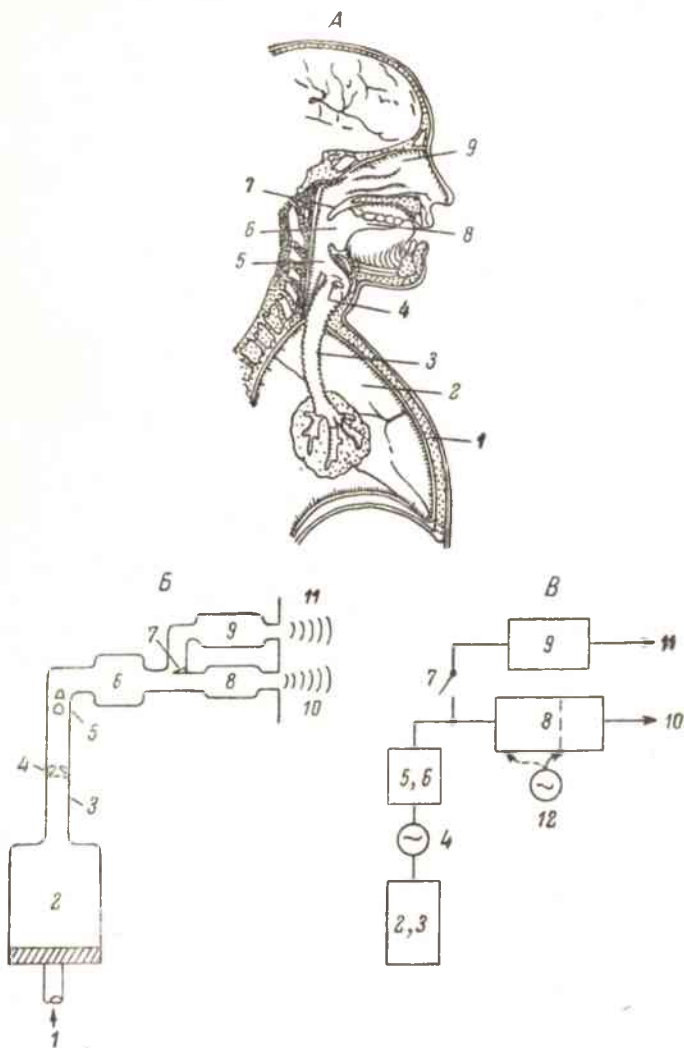


Рис. 1.2. Схема речеобразующего аппарата.

А — анатомическое изображение; Б — функциональные элементы; В — эквивалентная блок-схема. На А: 1 — грудная клетка, 2 — легкие, 3 — трахея, 4 — голосовые связки, 5 — гортанная трубка, 6 — полость глотки, 7 — небная занавеска, 8 — полость рта, 9 — полость носа. На Б: 1 — сила дыхательных мышц, 2 — объем легких, 3 — трахея, 4 — голосовые связки, 5 — гортанная трубка, 6 — полость глотки, 7 — небная занавеска, 8 — полость рта, 9 — полость носа, 10 — излучение из ротового отверстия, 11 — излучение из носовых отверстий. На В: 2, 3 — емкость легких и трахеи, 4 — голосовой источник колебаний, 5, 6 — емкость гортани и глотки, 7 — механизм небной занавески, 8 — емкость полости рта, 9 — емкость полостей носа, 10 — выходной сигнал ротового тракта, 11 — выходной сигнал носового тракта, 12 — шумовой источник.

образующиеся в определенных местах речевого тракта. Особенности этих источников акустической энергии будут описаны ниже.

Возбужденные акустические колебания подвергаются частотной фильтрации в воздушных полостях речевого тракта, действующих как акустические частотные фильтры. Конфигурация и объемы этих полостей в процессе речеобразования определенным образом изменяются. Соответственно этому изменяется и спектр исходных звуковых колебаний, создаваемых акустическими источниками.

Образование воздушных потоков, работа механизма гортани, все движения органов, образующих речевой тракт («артикуляторов»), происходят закономерно и координированно. Благодаря этой динамически слаженной деятельности и возникают сигналы связанной речи.

#### 1.1.1. КЛАССИФИКАЦИЯ ЗВУКОВ РЕЧИ

Перед изложением современных представлений акустической теории речеобразования коснемся подходов к классификации звуков речи, основывающихся на рассмотрении особенностей работы артикуляторного аппарата. общепринятая классификация звуков речи базируется на ряде упрощающих допущений, из которых наиболее существенными являются следующие:

1) каждый язык может обойтись весьма ограниченным набором действий органов, участвующих в речеобразовании (набор «артикуляторных жестов»);

2) каждый артикуляторный жест есть некоторое характерное для него состояние речевого аппарата (особенности работы источников звуковой энергии, конфигурация речевого тракта) и ведет к возникновению определенного звука речи;

3) артикуляторные жесты выполняются последовательно один за другим.

В результате этих допущений мы имеем дело с идеализированной речью, состоящей из предельно четко выраженных, характерно различающихся между собой звуковых элементов. При этом обеспечивается возможность разобраться в исходной структуре звукового материала, на которой основывается естественная речь того или иного языка. Кратко рассмотрим классификацию звуков на примере русского языка.

Все звуки речи делятся на два основных типа: гласные и согласные. При образовании гласных воздушный поток свободно проходит через весь речевой тракт; работает голосовой источник (для нормальной нешепотной речи); речевой тракт имеет определенную конфигурацию, благодаря чему обеспечивается специфическая форма спектра, типичная для данного звука. Гласные могут искусственно продолжительно «тянуться», и в речевом потоке при нормальном темпе ударные гласные обычно имеют участок, где их характеристики оказываются относительно стационарными.



При образовании другого типа звуков — согласных — воздушный поток встречает более или менее выраженное препятствие, расположенное в том или ином месте речевого тракта. Образование согласных характеризуется, как правило, быстрыми изменениями в действиях речеобразующего механизма, большинство согласных, по существу, определяется спецификой возникающих при этом переходных процессов. Классификация согласных рус-

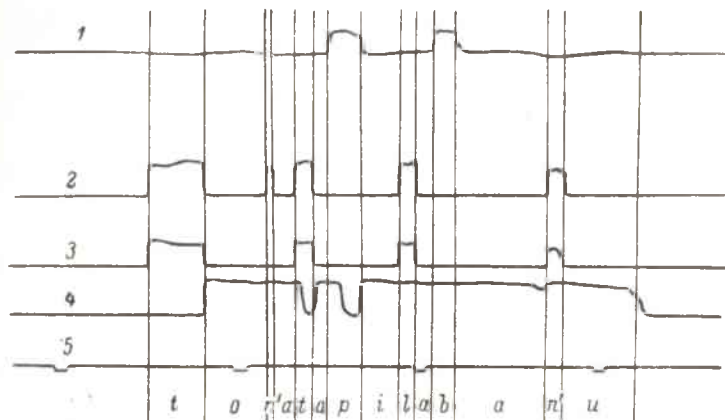


Рис. 1.3. Запись основных артикуляторных параметров при произнесении фразы *Тоня топила баню*.

Обозначения см. в тексте.

ского языка [72] базируется на следующих альтернативных признаках:

- 1) участие голосового источника: глухие, звонкие;
- 2) способ образования: щелевые, смычные, дрожащие;
- 3) место образования: губные, переднеязычные, среднеязычные, заднеязычные;
- 4) участие носовых полостей: назальные, чистые;
- 5) характер положения спинки языка: твердые, мягкие.

Если попытаться поочередно произносить разные звуки речи, то легко можно установить, что для каждого звука имеются свои отличительные особенности в работе речеобразующего механизма. Соответственные отличия имеют место и в акустических сигналах. Пониманию этих особенностей образования речевых звуков помогают специальные способы регистрации артикуляторных явлений. Так, на рис. 1.3 приведена запись показателей работы речеобразующего механизма при произнесении фразы *Тоня топила баню*, полученная с помощью техники комплексной регистрации артикуляторных параметров [159].

Рассмотрим рисунок. Кривая 1 показывает факты смыкания губ при произнесении [p] и [b]. Кривые 2 и 3 показывают касание языком нёба в точках, расположенных в передней части твердого

нёба по центральной линии (кривая 2) и несколько латеральнее этой линии (кривая 3). Наличие сигнала хотя бы на одной из этих кривых означает переднюю нёбно-язычную смычку, которая в рассматриваемом случае образовывалась при произнесении [t], [n'], [t], [l], [n]. Кривая 4 показывает интенсивность сигнала, получаемого от ларингофона («горлового микрофона») и позволяет судить о наличии или отсутствии работы голосового источника (фонации). Фонация началась на первом гласном и имела лишь небольшие перерывы в моменты смычных согласных [t] и [p]. Кривая 5 — отметка времени (1 с).

Записи, подобные приведенной, помогают характеризовать состояния артикуляторного аппарата и их изменения, имеющие место при произнесении того или иного речевого материала (звуков, слов, фраз). Однако четкая артикуляторная картина в случае слитной речи наблюдается только при тщательном и небыстром произнесении, т. е. при так называемом полном стиле. При нормальной устной речи, особенно при ускоренном ее темпе, дело существенно усложняется. Ранее указывалось, что соответствующие определенным звукам речи положения и движения всех органов, участвующих в речеобразовании, можно рассматривать как артикуляторные жесты. Ясно, что каждый из этих жестов явно отличается от других только при намеренно четком, еще лучше раздельном, его исполнении.

При непрерывном и особенно при ускоренном произнесении четкость их стирается. Типичной является такая ситуация, когда элементы движений, характеризующих образование звуков последовательности, начинают осуществляться заранее на участках, которые относятся к предшествующим звукам. К тому же каждый жест может оказаться зависящим от жестов, которые были осуществлены перед ним и которые должны быть осуществлены после него. Все это означает, что полностью предсказать данный жест исходя только из отдельно взятого соответствующего ему звука речи оказывается невозможным [372].

Анализ трудностей, которые возникают при описании речевого сигнала, основывающемся на данных артикуляции, выходит за рамки настоящей книги. Отметим это, перейдем к изложению основных данных акустической теории речеобразования, выделив описание источников акустической энергии и явлений частотной фильтрации в речевом тракте.

#### 1.1.2. ИСТОЧНИКИ АКУСТИЧЕСКОЙ ЭНЕРГИИ

Воздушный поток, создаваемый в речевом тракте благодаря действию дыхательной системы, может модулироваться тремя способами, показанными на рис. 1.4. В результате модуляции энергия воздушного потока в какой-то своей части преобразуется в акустические колебания.

Наиболее мощным источником акустической энергии является гортань, представляющая собой орган, приспособленный для создания звуковых колебаний. Находящиеся внутри полости гортани две эластичные складки — «голосовые связки» — образуют своеобразный и тонко управляемый клапанный механизм. При наличии определенных условий — достаточной величины подвязочного давления, сведения и соответствующего натяжения голосовых связок — последние приходят в колебательное движение.

Колебания эти происходят благодаря взаимодействию сил, направленных в разные стороны. Сила, создаваемая подвязоч-

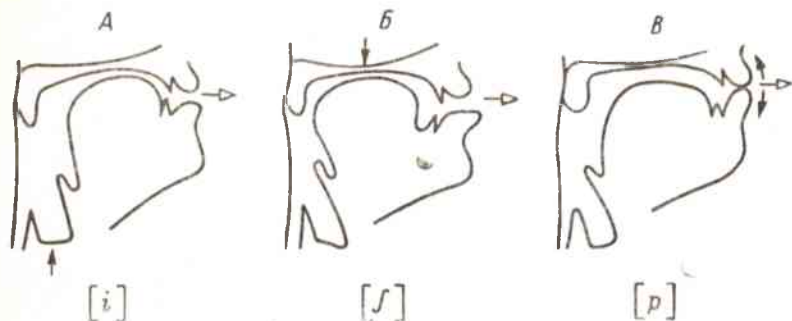


Рис. 1.4. Основные способы модуляции воздушного потока, обеспечивающие возникновение звуков речи. По [479].

*А* — периодическая модуляция колеблющимися голосовыми связками; *Б* — возникновение турбулентности в потоке, проходящем через место резкого сужения речевой трубки; *В* — возникновение взрывных звуков при быстром открывании полной смычки. Белые стрелки — звуковой сигнал, черные — место образования звука.

ным давлением, стремится раздвинуть голосовые связки; благодаря упругости соответственно натянутых мышцами связок возникает сила, пытающаяся свести их вместе. Определенное значение имеет также эффект Бернулли (возникновение силы, сближающей стенки канала, если в нем с большой скоростью протекает поток жидкости или газа). Площадь отверстия между голосовыми связками (голосовая щель) периодически изменяется. Как правило, в течение некоторой доли периода колебаний голосовая щель оказывается полностью сомкнутой, воздушный поток полностью прерывается. Типичной является картина, изображенная на рис. 1.5, *Б*.

Частотный спектр, соответствующий подобной последовательности воздушных толчков, представляет рис. 1.5, *В*. Расстояния по шкале частот между отдельными гармоническими компонентами линейчатого спектра определяются частотой, с которой следуют импульсы. Эта важнейшая характеристика голосового источника — основная частота голоса — характерным образом изменяется в потоке речи. Изменения эти имеют существенное значение для выражения повествовательной, вопросительной, восклиц-

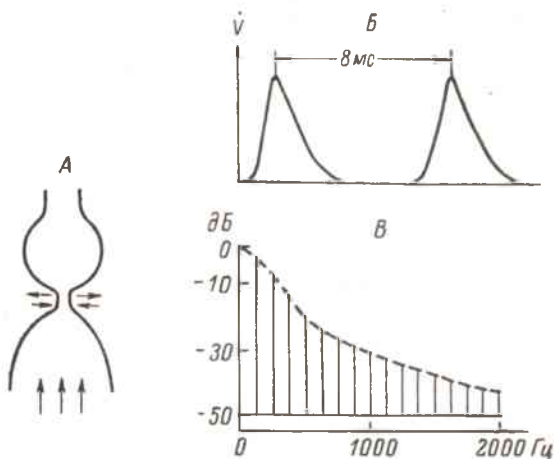


Рис. 1.5. Характеристики работы голосового источника.

— схематическое изображение фронтального сечения гортани в области голосовых связок, движения которых показаны горизонтальными стрелками; стрелки внизу показывают турбулентность воздуха. Б — типичная форма изменений скорости воздушного потока, возникающих благодаря действию колеблющихся голосовых связок; по оси абсцисс — время; по оси ординат — объемная скорость. В — спектр колебаний, оциллограмму которых изображает рис. 1.5, Б; по оси абсцисс — частота колебаний; по оси ординат — интенсивность.

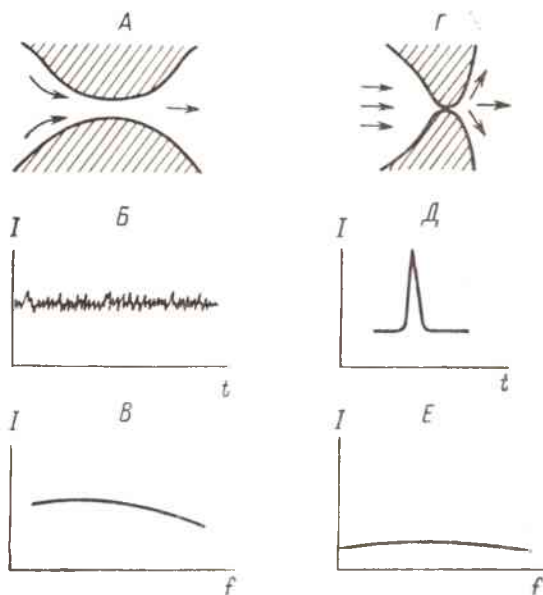


Рис. 1.6. Характеристики работы шумовых источников.

— протекание воздушной струи через сужение в речевой трубке; Б — шумовой сигнал, возникающий благодаря явлению турбулентности; В — широкополосный спектр шумового сигнала; Г — образование воздушного толчка при открывании сомкнутых артикуляторов, которыми создано дополнительное давление; Д — импульсный сигнал, возникающий в момент открывания смычки артикуляторов; Е — спектр импульсного сигнала. Стрелки на А, Г характеризуют движение воздушных потоков.

цательной интонации, для словесных и особенно логических ударений, для обозначения незаконченности или оконченности фразы и т. п. Для мужских голосов типичной областью изменений основной частоты голоса является 80—150 Гц, для женских — более высокая область 120—400 Гц.

В сложном механизме регулирования основной частоты голоса основную роль играют: задаваемая мышцами гортани степень натяжения голосовых связок и величина подсвязочного давления, создаваемого дыхательной системой. Интенсивность звуковых колебаний, создаваемых гортанью, зависит от максимальной величины объемной скорости воздуха, проходящего через голосовую щель в момент ее открытия, и в основном определяется величиной подсвязочного давления, создаваемого дыхательной системой. Форма воздушного толчка — наклон фронтов треугольника, изображенного на рис. 1.5, *Б*, отношение длительности фазы открытия голосовой щели к длительности всего периода — также изменяется в зависимости от подсвязочного давления и от степени натяжения связок. С этими изменениями связано изменение спектрального состава колебаний, возбуждаемых голосовым источником, что в конечном итоге также оказывает свое влияние на характер звукового сигнала речи.

Механизм возбуждения шумных звуков сводится к возникновению турбулентции в потоке воздуха, проходящего через речевой тракт. Возникновение шума связано как с величиной линейной скорости частиц воздуха, так и с геометрическими свойствами (величиной «эффективной площади») канала, в котором проходит воздушный поток. В речевом тракте условия, необходимые для возникновения шума, имеют место при образовании сужения в каком-либо сечении речевой трубки и создании достаточного напора воздуха за этим сужением.

Механизм создания таких шумных звуков, как [f], [s], [ʃ] и т. п., называемых фрикативными, соответствует возникновению шума при продувании воздуха через сопло (рис. 1.6, *А*). Возбуждаются беспорядочные звуковые колебания, спектр которых отличается значительной протяженностью в области высоких частот. Подобные согласные звуки можно производить продолжительное время, поддерживая их параметры стационарными.

Наоборот, краткостью отличаются шумные взрывные звуки ([p], [t], [k] и др.). При их артикуляции производится полное смыкание стенок в определенном месте речевого тракта, глубже места этой «смычки» создается повышенное давление, и смычка быстро размыкается (рис. 1.6, *Г*). Быстрый толчок воздуха создает короткий импульсный звук.

Голосовой и шумовые (фрикативный или взрывной) источники могут действовать и одновременно (звонкие согласные). При шепотном произнесении гласных звуков источником возбуждения является турбулентный воздушный поток у частично сомкнутых

голосовых связок. Динамика работы всех этих источников в слитной речи отличается высокой степенью координации с движениями всех органов речеобразующего аппарата.

### 1.1.3. ЧАСТОТНАЯ ФИЛЬТРАЦИЯ В РЕЧЕВОМ ТРАКТЕ

Звуковые колебания, возникающие в речевом тракте в результате действия рассмотренных выше акустических источников, представляют собой как бы исходный «сырой» материал, подвергающийся в процессе создания речевого сигнала дальнейшим преобразованиям. Ведущую роль здесь играют частотно-избирательные свойства воздушных полостей всего речевого тракта. Если звуковые колебания имеют место в некотором ограниченном объеме, то будут проявляться акустические резонансные

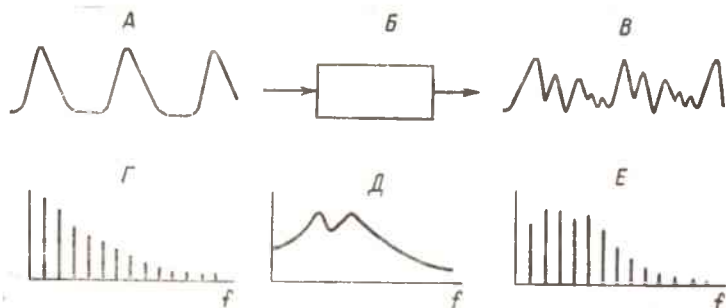


Рис. 1.7. Схематическое изображение преобразований сигнала в речевом тракте. По [134].

А — сигнал голосового источника; Б — фильтрующая система речевого тракта; В — выходной речевой сигнал; Г — спектр сигнала голосового источника; Д — вид частотной характеристики речевого тракта; Е — спектр речевого сигнала.

свойства этого объема, зависящие от его размеров и конфигурации. Рассматриваемый в этом аспекте речевой тракт представляет собой последовательно расположенные полости довольно сложной конфигурации, к тому же существенно изменяющиеся в динамике речевого потока. В акустическом отношении подобная система является сложным частотным фильтром с переменными параметрами.

Современная акустическая теория речеобразования [134] показывает, что весь процесс создания речевого сигнала может быть сведен к возбуждению акустических источников и управляемой фильтрации создаваемых ими сигналов. Схему происходящих явлений можно пояснить рис. 1.7. Спектральные составляющие исходных колебаний, проходя через фильтр, умножаются на величины соответствующих ординат частотной характеристики фильтра. Преобразованный таким образом спектр и характеризует излучаемый речевой сигнал.



Фильтрующее действие речевого тракта зависит от его геометрической формы и места расположения акустического источника. При работе голосового источника место его расположения не изменяется, ведущее значение оказывается за конфигурацией полостей речевого тракта. Характерный пример показывает рис. 1.8. При фонации звука [i] к нёбу поднимается передняя часть спинки языка, большим оказывается объем полостей, расположенных позади места этого сужения, малым — объем между местом сужения и

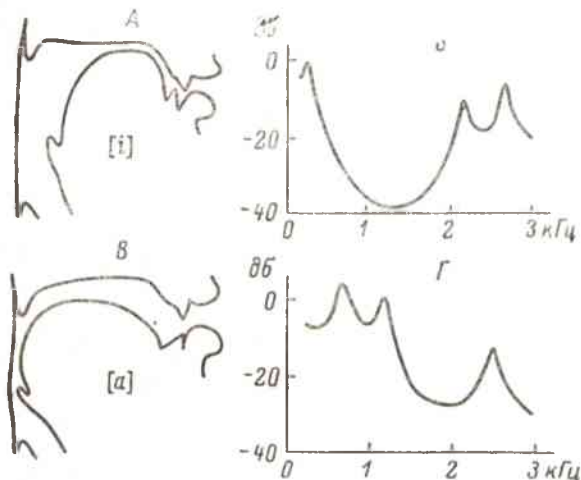


Рис. 1.8. Схематическое изображение конфигурации артикуляторов (А, В) и огибающие спектров (Б, Г) для гласных звуков [i] и [a]. По [479].

По оси абсцисс — частота; по оси ординат — относительная интенсивность.

ротовым отверстием; поднятая вверх занавеска мягкого нёба «отключает» носовые полости. В результате «настроенной» таким образом частотной фильтрации звуковые колебания голосового источника, спектр которых показан на рис. 1.5, В, характерным образом преобразуются. Огибающая спектра окончательного сигнала представляется рис. 1.8, Б. При образовании звука [a] объем полостей передней части речевого тракта оказывается значительно большим, так как спинка языка более удалена от нёба (рис. 1.8, В); спектр звукового сигнала оказывается существенно отличным (рис. 1.8, Г), хотя спектр голосового источника в обоих случаях одинаков.

Подобного же рода частотная фильтрация имеет место и при действии шумовых источников, но процесс усложняется изменением места возникновения звука. В первом приближении источник может быть привязан к месту смычки артикуляторов. Фильтрующие свойства речевого тракта определяются совокупностями объемов, расположенных как за артикуляторной смычкой, так и перед ней.

При опускании небной занавески роль акустических резонаторов начинают играть и носовые полости. Образование назальных гласных [m] и [n] происходит при закрытом ротовом отверстии. Звук излучается через носовые отверстия, однако полость рта выступает в качестве параллельно подключенного резонатора. Сообщение между ротовым и носовым трактами может возникать и при фонации гласных — при их «назализации». В данном случае основная энергия излучается через ротовое отверстие, но звук приобретает характерную окраску.

Резонансные пики, проявляющиеся в картине спектра и соответствующие набору собственных частот речевого тракта, называются «форманты». Максимальное число их зависит от общей длины речевого тракта. Считается, что у взрослого мужчины их может быть до 7. Существенное значение для различия звуков имеют первые 2—3 низкочастотные форманты.

Понятие «форманта» не является, однако, достаточно однозначным. Дело в том, что при работе голосового источника возбуждаются колебания, имеющие спектр, приближающийся к линейчатому. Частоты гармоник могут находиться в разных положениях относительно резонансных частот речевого тракта, и соответственно тому оказывается, что спектральные максимумы результирующего акустического сигнала не являются непосредственным отражением пиков фильтрующей функции речевого тракта (особенно при высоких частотах основного тона, как, например, в женских голосах). Вопрос о том, как может быть определена форманта точки зрения восприятия сигнала, будет обсуждаться в главе 10.

К настоящему времени проведено весьма значительное число исследований, целью которых были попытки найти и количественно выразить зависимости между важными для речеобразования акустическими свойствами и геометрическими размерами тракта. Основой здесь являются измерения, осуществляемые с помощью рентгенографической техники, речевой тракт обычно аппроксимируется моделью в виде последовательно соединенных отрезков труб. Достаточно точные расчеты [134, 137] могут быть получены лишь в том случае, если звуки рассматривать заведомо стационарными. Несмотря на это, получаемые количественные результаты представляют явный интерес, особенно в связи с разработкой методов искусственного синтеза речевых сигналов.

## 1.2. АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ

Общепризнанным способом анализа акустического речевого сигнала, с помощью которого получена подавляющая часть данных, составивших современные сведения о его структуре, является динамическая спектрография. Среди большого числа таких методов (см. обзоры [37, 121, 137]) он представляется наиболее важным для исследователей речи. По этой причине только он и будет рассмотрен в настоящем разделе.



Рис. 1.9. Динамическая спектрограмма фразы *Тоня пошла баню.*  
Обозначения см. в тексте.

Принцип метода и соответствующая аппаратура были описаны в 1946 г. [330], и вскоре же появилась солидная монография под названием «Видимая речь» [434], представившая богатый иллюстративный материал: множество динамических спектрограмм отдельных звуков, звуко сочетаний, слов, фраз и текстов (последние не только на английском, но и на многих других языках). С этих пор устройства, использующие этот метод, часто называются приборами «видимая речь».

В наиболее распространенных модификациях приборов анализируемый отрезок сигнала (длительностью обычно около 2.5 с), записанный на магнитный носитель, многократно воспроизводится и подается на вход полосового фильтра. На диаграмме, представляющей результаты анализа, вычерчивается линия, степень черноты которой отражает интенсивность выходного сигнала упомянутого фильтра. Запись производится на тепловой, электрочувствительной или фотографической бумаге. С каждым циклом частота настройки фильтра сдвигается, на диаграмме вычерчивается новая линия, несколько отодвинутая от предыдущей. В результате возникает трехмерное изображение с осями: время, частота, интенсивность. В приборах обычно используется принцип гетеродинного анализатора, абсолютная величина ширины полосы фильтра сохраняется постоянной по всей частотной шкале. Наиболее часто ширина полосы пропускания фильтра берется довольно большой — порядка 300 Гц.

На рис. 1.9 показана динамическая спектрограмма фразы *Тоня топил баню* (запись ее артикуляторных параметров изображает рис. 1.3). Видно, что на динамической спектрограмме на протяжении всей фразы, за исключением участков, отнесенных к [t], [t] и [p'], имеются области почернения (наличия звуковой энергии) в тех или иных зонах частотной шкалы. Наибольшей степенью почернения отличается область нижних частот до 1000 Гц, здесь располагаются максимально выраженные спектральные составляющие гласных. Можно видеть, что на участках гласных звуков наблюдается несколько областей почернения, положение которых по шкале частот изменяется на протяжении звука. Эти области соответствуют формантам гласного.

Более точная количественная характеристика амплитудно-частотного спектра, соответствующая некоторому моменту изучаемого сигнала, может быть получена с помощью регистрации на той же диаграмме «видимой речи» так называемого спектрального среза. На рис. 1.9 темная вертикальная линия отмечает момент, к которому относится измерение, а кривая в правой части диаграммы отражает спектр в логарифмическом масштабе.

При рассмотрении рис. 1.9 отмечается также характерная структура динамической спектрограммы в виде вертикальных полос (в областях гласных звуков и звонких согласных). Эта структура отражает работу голосовых связок. Частота следования вертикальных полос определяется частотой основного тона голоса.

На спектрограмме, например, четко отмечается уменьшение этой частоты к концу фразы.

На спектрограмме можно легко видеть также общее уменьшение энергии (по сравнению с гласными) и некоторое снижение частоты основного тона на участках, относимых к звонким согласным [n'], [l], [b], [n']; шум взрывных согласных [t], [t], [p'] представлен узкой вертикальной полоской, занимающей большую часть частотной шкалы, что и характерно для импульсного звука.

Огромное число исследований, выполнявшихся с помощью метода «видимой речи», было направлено на выявление характерных инвариантных свойств тех или иных речевых звуков, нахождение различительных признаков соответствующих звуковых групп и многих других фонетически значимых особенностей сложной картины речевого сигнала. Большая практика позволила сформулировать определенные правила чтения «видимой речи» [135, 254, 358], однако уже сравнительно давно стало ясно, что возможности метода в отношении расшифровки структуры речевого сигнала далеко не беспредельны и по сути исчерпались на этапах работ, проводившихся в пятидесятых и шестидесятых годах [344].

Были поставлены специальные исследования для выявления предельных возможностей «чтения» речи на основе динамических спектрограмм. Выяснилось [53, 327, 362], что полностью достоверного распознавания достичь не удастся, но окончательные результаты распознавания могут быть существенно улучшены, если спектрограммы относятся только к ограниченному набору фраз, содержание которых заранее известно.

### 1.3. СИНТЕЗ РЕЧЕВЫХ СИГНАЛОВ

Кроме прикладного использования (создание «говорящих машин»), методы искусственного синтеза речевых сигналов применяются и в исследовательских целях. Здесь выделяются две основные группы задач. Во-первых, синтезаторы разрабатываются как функциональные модели процессов, имеющих место при речеобразовании. В данном случае они используются для разработки и подтверждения правильности тех или иных положений теории речеобразования. Во-вторых, синтезаторы используются как устройства, способные генерировать речеподобные звуки с точно задаваемыми параметрами. Звуки эти применяются и при исследовании механизмов восприятия речи.

Широкое использование синтезированных звуков для исследования особенностей восприятия речи началось со времени работ, проводившихся в пятидесятых годах в Хаскинских лабораториях в США [75, 220]. Идея исследований сводилась к попыткам найти полезные признаки звуковых сигналов, на основании которых человеком производится различение определенных звуков речи. В опытах создавались звуки, физические характеристики которых точно задавались и изменялись согласно желанию эксперимента-



тора. Слушатели идентифицировали эти звуки с теми или иными элементами речи.

Было изготовлено специальное электро-оптическое устройство, создающее речеподобные звуки. Управляющие сигналы рисовались на лентах; вдоль ленты шла ось времени, поперек — ось частот [221]. С помощью этого в общем достаточно примитивного устройства и были начаты современные исследования связи физических особенностей речевого сигнала с особенностями его восприятия.

К настоящему времени созданы значительно более совершенные средства синтеза речевых сигналов. Наибольшее приближение

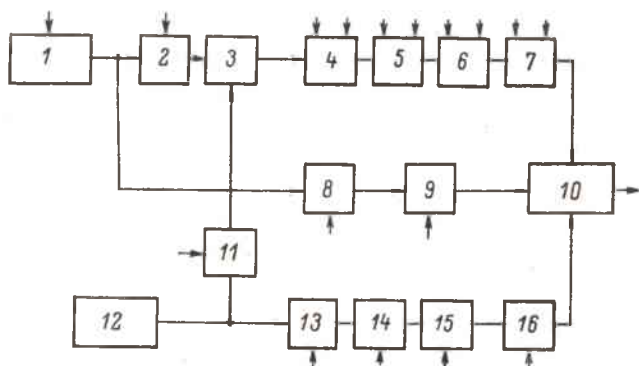


Рис. 1.10. Блок-схема синтезатора речевых сигналов OVE-III. По [357].

Обозначения см. в тексте.

к естественной речи дают системы, в той или иной форме моделирующие основные явления, которые имеют место при речеобразовании. Не имея возможности хоть в какой-то мере охватить все разнообразие систем [45, 121, 138, 261], рассмотрим в качестве примера принципы устройства хорошо известного современного аналогового синтезатора (OVE-III), разработанного в Лаборатории передачи речи Стокгольмского технологического института [35].

В синтезаторе (рис. 1.10) источниками энергии создаваемых колебаний служат два генератора. Генератор последовательности импульсов треугольной формы (1) представляет голосовой источник. Под влиянием управляющего сигнала его частота изменяется, что отражает изменение основной частоты голоса. Вторым генератор (12) создает шумовые колебания, воспроизводящие действие шумовых источников.

Колебания от генераторов попадают в систему, в которой основными являются элементы с управляемым усилением (2, 9, 11, 16), управляемые частотные фильтры (4—8, 13—15) и смесители сигналов (3, 10). Фильтры, у которых под влиянием электрических



управляющих сигналов изменяется как положение резонансной частоты, так и ширина полосы пропускания, являются важными элементами формантных синтезаторов. Эти фильтры моделируют резонансные свойства полостей речевого тракта.

Для образования гласных звуков служат четыре таких последовательно включенных фильтра (4, 5, 6, 7), обеспечивающих появление в выходном сигнале четырех формант. Управляемыми параметрами являются: уровень интенсивности сигнала от голосового источника, резонансные частоты и ширина полос всех формантных фильтров.

Эффект подключения носовых полостей при образовании назальных звуков достигается благодаря прохождению колебаний голосового источника через параллельную ветвь, имеющую лишь один формантный фильтр (8) и управляемый усилитель (9). Третья ветвь служит для образования шумных фрикативных согласных. Она состоит из двух формантных фильтров (14, 15) и одного «антиформантного» фильтра (13), производящего подавление частот в определенной области, положение которой также изменяется отдельным управляющим сигналом. Необходимость введения специальных фильтров для образования шумных звуков связана с тем, что место положения шумового источника перемещается по трубке речевого тракта и шумные звуки образуются при разведенных голосовых связках. Влияние этих факторов делает акустические условия отличными от тех, которые имеются при образовании гласных.

Важный для английского и ряда иных языков эффект возникновения аспирационного шума (результат турбуленции, возникающей в районе голосовой щели при неколеблющихся голосовых связках) производится с помощью регулируемого усилителя (11), подмешивающего определенную долю шумовых колебаний к сигналу голосового источника.

Выходные сигналы всех трех ветвей — гласных, назальных и шумных звуков — складываются на электронном сумматоре (10) и образуют выходной речеподобный сигнал.

Прибор типа OVE-III является так называемым синтезатором последовательного типа; фильтры в нем соединены один за другим, воспроизводя таким образом частотную фильтрацию в последовательно расположенных резонансных полостях речевого тракта. Известны разработки синтезаторов, в которых трубка речевого тракта моделируется еще более детально в виде длинной цепочки резонансных элементов, воспроизводящих изменения ее поперечных сечений [285, 487].

Однако речевые сигналы достаточно успешно можно воспроизвести и с помощью так называемых синтезаторов параллельного типа. Управляемые фильтры в них параллельно подключаются к генераторам, а их выходные сигналы суммируются [299].

При проведении исследований восприятия речи система параллельного синтеза имеет определенные преимущества. Так, частоты

и интенсивности каждой форманты управляются независимо, в то время как в последовательном синтезаторе они оказываются взаимосвязанными в силу того, что управляемые фильтры подключаются друг за другом. В параллельном синтезаторе легче добиться большего динамического диапазона, проще могут быть получены речеподобные сигналы, у которых все физические характеристики точно задаются по желанию экспериментатора.

В значительной части исследований восприятия речи могут быть использованы достаточно краткие речеподобные сигналы относительно простой структуры. В этом случае управление синтезатором оказывается относительно простым и для этой цели могут быть использованы удобные и точные системы генераторов различных функций, работающие на основе аналоговых электронных схем.

#### 1.4. АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ РЕЧИ

Проблема машинного распознавания речевых сигналов имеет особое значение для исследователей восприятия речи по следующей элементарной причине. Признаком достаточной изученности процесса восприятия речи является возможность построения функциональной модели этого процесса. Такая модель должна прежде всего быть способной распознавать речь, т. е. она по определению должна быть системой автоматического распознавания речи. Хотя обратное положение — система автоматического распознавания речи должна моделировать процесс восприятия речи человеком — формально не является обязательным, за предлагавшимися системами автоматического распознавания, как правило, стояли идеи их авторов об определенных принципах организации процесса восприятия.

Соответственно эволюция работ по автоматическому распознаванию речи достаточно хорошо отражает эволюцию представлений относительно процесса восприятия речи человеком. Ниже мы коротко расскажем только об основных идеях — принципиальных подходах, — не останавливаясь ни на конкретных математических методах, ни на технической реализации систем. Эти вопросы подробно освещены в ряде обзорных работ [1, 44, 46, 88, 148].

##### 1.4.1. АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ РЕЧИ И ПРОБЛЕМА РАСПОЗНАВАНИЯ ОБРАЗОВ

Первая попытка создать техническое устройство, способное распознавать звуки речи, была предпринята Мясниковым в начале сороковых годов [111, 112]. Однако серьезный интерес к проблеме автоматического распознавания речи со стороны математиков и инженеров возник примерно 15 лет тому назад. Задача автоматического распознавания речи рассматривалась в то время как частный случай общей проблемы распознавания образов.

Основные усилия были направлены на анализ и формализацию проблемы, определение необходимой системы понятий, поиск адекватного математического аппарата.

Читателю, желающему познакомиться с теорией распознавания образов, следует обратиться к специальной литературе [50, 68, 69, 124]. В настоящей книге мы ограничимся очень кратким и схематическим изложением основных идей подхода.

Распознавание образа в живой и неживой системе рассматривается как результат обучения. В процессе обучения системе предъявляются изображения сигналов и сообщаемые «учителем» названия классов, к которым принадлежат эти изображения (сигналы). Изображения сигнала соответствуют указанию значений сигнала по некоторому произвольно выбранному набору признаков  $(x_1, x_2, \dots, x_n)$ . Соответственно каждое единичное изображение может быть представлено точкой в  $n$ -мерном пространстве измеряемых признаков сигнала. Обучающаяся система использует гипотезу о том, что изображения, относящиеся к одному и тому же классу, образуют компактные множества в пространстве признаков сигнала, т. е. каждому классу соответствует определенная область в этом  $n$ -мерном пространстве. Задача системы состоит в том, чтобы найти наилучшие границы между областями. При предположении, что граница между двумя областями соответствует гиперплоскости, определение границы сводится к подбору коэффициентов  $a_0, a_1, \dots, a_n$  в уравнении, описывающем эту гиперплоскость:  $a_0 + \sum_{i=1}^n a_i x_i = 0$ . Это уравнение называется линей-

ной решающей функцией. Если значение  $a_0 + \sum_{i=1}^n a_i x_i$  больше нуля, то изображение относится к одному классу, если оно меньше нуля, то изображение относится к другому классу. Значения коэффициентов  $a_0, a_1, \dots, a_n$  подбираются так, чтобы минимизировать вероятность ошибочной классификации.

В случае автоматического распознавания речевых сигналов такой общий подход был использован как для фонемного распознавания звуков речи, так и для распознавания изолированных слов-команд. В случае изолированных слов каждое слово рассматривалось как самостоятельный образ. Изображение слова включало признаки звукового сигнала, соответствующего слову в целом.

Хотя задача автоматического распознавания изолированных слов-команд из ограниченного набора (от 5—10 до 50—200 слов) решается довольно успешно [28, 33, 80, 133] и находит практическое применение [119], в научном отношении она не представляет большого интереса. От нее не может быть перехода к распознаванию естественной связной речи.

Не существует сомнений, что распознавание связной речи должно основываться на фонемном или близком к фонемному распознавании. При фонемном распознавании изолированных зву-

ков речи звуки речи рассматривались как сигналы, а фонемы — как названия классов. Решающие функции, найденные в процессе обучения, соответствовали границам между фонемными областями в пространстве выбранных исследователями признаков сигнала. Как выяснилось, автоматическое фонемное распознавание изолированно произнесенных звуков речи (преимущественно исследовались гласные) может не уступать по надежности фонемному распознаванию их человеком [88]. Однако система, обученная на изолированных звуках речи, оказывается практически неработоспособной в условиях, когда ей предъявляется для распознавания связная речь.

Мы остановимся лишь на нескольких из огромного числа вопросов, с которыми столкнулись исследователи, пытавшиеся разработать системы для автоматического фонемного распознавания связной речи.

Как уже говорилось, система автоматического распознавания образов формируется под влиянием «учителя», который знает возможные классы сигналов и сообщает относительно каждого сигнала, к какому классу он относится. В роли такого учителя системы автоматического распознавания речи обычно выступает лингвист.

Предполагалось, что, слушая речевой поток, лингвист может точно указать моменты времени, когда начинается и кончается данная фонема, причем у него не возникает сомнений в том, какая именно фонема произносится. Реально оказалось, что оба эти предположения не оправдываются (см. главы 2 и 3). Последовательно прослушивая отрезки речевого потока, соответствующие примерно слогам, фонетисты могут более или менее однозначно записать сигнал последовательностью фонетических символов, соответствующих фонемным вариантам или аллофонам (число их значительно превышает число фонем). Для фонемной записи звука речи нужно в ряде случаев услышать или слово в целом, или даже несколько соседних слов.

Уже на первых этапах работы по автоматическому распознаванию звуков речи стало ясно, что нужны какие-то гипотезы, направляющие выбор полезных признаков сигнала. Математические методы позволяют сравнить эффективность выбранных систем признаков или уменьшить размерность описания, однако сам изначальный выбор признаков оказывается вне пределов теории распознавания. Вопрос о выборе подходящих признаков становится особенно важным, когда дело идет об описании принципиально нестационарных сигналов, как это имеет место в случае естественного речевого потока. С вопросом выбора признаков тесно связан вопрос о сегментации речевого потока. Если возможно предварительное разделение потока на элементарные отрезки, то в качестве полезных признаков могут выступать свойства отрезков.

Таким образом, первый этап работ по автоматическому распознаванию речи позволил проанализировать задачу и выявить

несколько узловых вопросов, без решения которых практически невозможно продолжать работу. К ним относится определение системы полезных признаков сигнала, определение набора распознаваемых образов (фонемы или варианты фонем), выяснение отношений между фонемами (вариантами фонем) и отрезками звукового потока.

Можно видеть, что все перечисленные вопросы находятся за пределами общей теории распознавания образов, их решение требует выяснения тех частных принципов кодирования и декодирования информации, которые характерны именно для речевой коммуникации.

Очевидно, что для получения нужных сведений необходимо обратиться к исследованию и моделированию деятельности или источника речевого сигнала (системы речеобразования), или приемника речевого сигнала (системы восприятия).

#### 1.4.2. «МОТОРНАЯ ТЕОРИЯ» И «АНАЛИЗ С ПОМОЩЬЮ СИНТЕЗА»

До последнего времени существовала надежда на то, что ключ к решению проблемы автоматического распознавания речи лежит в изучении речеобразования. Подход этот легко объясним, если учесть, что наиболее общее определение сигнала состоит в том, что он является процессом, переносящим сведения о состоянии порождающей его системы [140].

Значительные усилия исследователей были направлены на разработку методов, позволяющих с достаточной точностью восстанавливать по текущим значениям речевого сигнала текущие значения параметров передаточной функции речевого тракта и характеристики источника возбуждения [181, 182, 190]. За последние годы на этом пути были достигнуты весьма существенные успехи [375].

Казалось, что остается сделать еще один шаг и восстановить по картине изменения состояния тракта во времени те сигналы управления — моторные команды, которые вызвали эти изменения состояния.

На перспективность такого подхода указывала, казалось бы, и весьма популярная примерно 10 лет назад моторная теория восприятия речи [350] (см. обзоры [40, 339]). Моторная теория постулировала, что, воспринимая речевой сигнал, слушатель определяет моторные команды, которые необходимо применить для создания аналогичного сигнала, и запоминает далее уже не акустическое изображение сигнала, а последовательность этих моторных команд.<sup>2</sup>

---

<sup>2</sup> Следует заметить, что экспериментальные факты, на которых основывалась моторная теория, были в равной мере совместимы с представлением,



Привлекательность моторной теории для специалистов, занимающихся проблемой автоматического распознавания речи, заключалась в том, что моторные команды можно было в первом приближении отождествить со ступенчатыми импульсами управления, характеризующимися всего двумя параметрами — временем появления ступеньки и ее величиной.

Рассматривая движение управляемого артикуляторного органа как отклик инерционного звена на ступенчатое возмущение и считая, что параметры этого звена постоянны, можно было надеяться составить модель, описывающую преобразование управляющих команд во временную картину изменения конфигурации речевого тракта и соответствующую ей временную картину изменений параметров тракта как акустической системы.

Если такая модель создана, то появляется возможность применить метод анализа с помощью синтеза, т. е. записать восстановленную по речевому сигналу картину изменений параметров речевого тракта во времени и подобрать такие значения управляющих команд в модели, при которых отклик модели будет наилучшим образом соответствовать реально наблюдавшейся картине. Естественно, что реализация такого метода возможна лишь при использовании ЭВМ. Метод требует предварительного запоминания изображения сигнала, параметров тракта — как функции времени, — он включает поиск одновременно по многим переменным наилучшей аппроксимирующей функции.

Важно, что при таком подходе фонемы существуют как дискретная последовательность лишь на уровне сигналов управления. Если управляющие ступенчатые импульсы быстро следуют друг за другом, то конфигурация тракта может вообще не достигать стационарных «целевых» состояний. В случае метода анализа с помощью синтеза это не является препятствием — для аппроксимации используются переходные участки процесса.

Рассмотрение фонемы как группы моторных команд, адресованных разным артикуляторным органам, открывало также, как казалось, принципиальную возможность параллельного поиска команд, относящихся к гласным и согласным. Факт переслаивания артикуляции гласных и согласных, известный как эффект коартикуляции, является одним из дополнительных источников трудностей при автоматическом распознавании речевого потока.

Легко заметить, что перспективность указанного очень логичного, но технически сложного подхода критически зависит от того, насколько стандартны моторные команды и что стабильнее — моторные команды или акустический эффект, достигаемый в результате реализации этих команд.

---

что человек текущим образом распознает дифференциальные признаки фонем (их вариантов) и запоминает не акустический, а фонетический образ сообщения. Именно такая интерпретация этих данных принята в настоящее время.



Последнее десятилетие характеризовалось чрезвычайно интенсивными исследованиями характеристик артикуляторных движений и связанной с этими движениями биоэлектрической активности мышц артикуляторов. Оказалось, что предположение о стандартности моторных команд не подтверждается [360, 372]. Как и в случае других сложных движений, обнаруживается высокая вариативность движений, зависимость их от многих конкретных условий исполнения. В качестве примера, хорошо известного каждому из обыденной жизни, можно привести тот факт, что человек достаточно легко говорит лежа, с сигаретой в зубах и т. д. Конкретные характеристики движений в этих разных условиях исполнения оказываются различными.

Исследователи речеобразования приходят сейчас к тому, что для описания и моделирования процесса управления речеобразованием необходимо использовать такое понятие, как цель движения. Очевидно, что целью в данном случае может быть лишь создание определенного акустического эффекта, доступного и «понятно» слушателю.

Если это так, то ключ к решению проблемы автоматического распознавания речи следует искать в изучении не источника, но приемника речевой информации, т. е. в изучении процесса восприятия речи.

#### 1.4.3. ИЕРАРХИЧЕСКИЕ СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

Наиболее характерной особенностью современного этапа работ по автоматическому распознаванию речи является стремление разработать многоуровневые иерархические системы, способные использовать разного рода сведения о языке и ситуации [18, 255, 441].

Утверждение, что только системы, знающие язык и понимающие смысл сообщения, окажутся в конечном итоге работоспособными и практически полезными, содержалось в письме-заметке Пирса [422]. Считается, что это письмо оказало очень большое влияние на организацию работ по автоматическому распознаванию речи не только в США, но и во всем мире. Оно привело к определенному сокращению разработок технических систем автоматического распознавания речи, но способствовало более серьезному отношению к теоретическим аспектам речевых исследований [256, 259, 324, 436].

Вопрос о том, почему система должна иметь иерархическую многоуровневую структуру, будет достаточно подробно рассмотрен в главе 2. Сейчас мы только укажем на то, что первым (самым нижним) блоком системы, в структуре которого уже должно отражаться знание языка, является блок фонетической интерпретации.

Следующие, более высокие, блоки должны моделировать морфологический и синтаксический анализ сообщения, обеспечивая в конечном итоге получение описания смысла сообщения (см. гл. 2).

Вопрос о том, как и в какой форме может быть заложено в систему знание языка, пока отнюдь не решен.

При обсуждении проектов систем указывается, что блок фонетической интерпретации должен использовать сведения о характерных для данного языка закономерностях сочетаний фонем, а также о фонетических и фонологических взаимовлияниях [407]. Для использования этих сведений, очевидно, необходимо, чтобы отрезок речевого потока, рассматриваемый при фонетической интерпретации, соответствовал интервалу, на котором наблюдаются эти взаимовлияния.

Общепринятый подход состоит в том, что если информации недостаточно, то выходом фонетического интерпретатора является не единственное решение (например, одна цепочка фонем), но несколько возможных кандидатов — несколько совместимых с сигналом цепочек фонем. Выбор между этими кандидатами производится на следующем уровне обработки, результатом которой может быть также указание нескольких кандидатов, но теперь уже не цепочек фонем, но последовательностей слов.

Использование языковой избыточности на уровне морфологического и синтаксического анализа может быть весьма разнообразным. Простейший вариант, применяемый в ряде реализованных систем автоматического распознавания [11, 34, 38, 55], заключается в том, что сообщение считается состоящим только из известных системе слов (набор их ограничен) и что в памяти системы указаны разрешенные сочетания слов.

Элементарный и поверхностный способ учета ситуации состоял бы в том, чтобы словам в словаре системы были приписаны вероятности, значения которых динамически зависели бы от ситуации. Однако разработка этого вопроса требует решения ряда сложнейших проблем, относимых к области моделирования интеллекта (форма описания смысла сообщения, организация сведений об окружающем мире в долговременной памяти системы и т. д.). Работа в этом направлении пока только начинается.

При обсуждении возможности использования языковых закономерностей выделяются закономерности, имеющие вид правил, и статистические характеристики. Очевидно, что привлечение при распознавании статистических сведений о языке (например, такое-то слово априори более вероятно, чем другие слова) чревато опасностями искажения сообщения. Поэтому разработка методов, реализующих догадку о сообщении, ни в коей мере не снимает проблему максимально полного извлечения информации, содержащейся в самом речевом сигнале. Соответственно все вопросы, связанные с определением набора полезных признаков речевого сигнала, набора фонетических элементов, правил фонетической интерпретации и т. д., в полной мере остаются актуальными и для иерархических распознающих систем. Последние обзорные статьи по автоматическому распознаванию речи [147, 260, 283, 407] свидетельствуют о том, что они все еще не решены.

## Глава 2

### ОБ ОПИСАНИИ РЕЧЕВОГО СИГНАЛА ПОСЛЕ ЕГО ФОНЕТИЧЕСКОЙ ИНТЕРПРЕТАЦИИ

Полная модель восприятия речи должна обеспечивать преобразование речевого сигнала в описание смысла сообщения. Анализ возможной структуры полной модели показывает, что она должна состоять из ряда последовательных блоков обработки информации.

Исследование отдельного блока такой сложной системы необходимо предполагает использование определенных допущений относительно описания сигнала на входе блока и описания на выходе. Так как описание на выходе  $i$ -того блока является входным описанием для  $(i+1)$ -го блока, формулировка требований к этому описанию начинается с того, что исследователь  $(i+1)$ -го блока сообщает, какое описание было бы удобным для его целей. Исследователь  $i$ -того блока принимает это в качестве исходной гипотезы и пытается ее проверить. Если она не подтверждается, происходят дальнейшие уточнения, ведущие к изменению представлений о предполагаемой структуре как  $i$ -того, так и  $(i+1)$ -го блока.

Одно из фундаментальных положений науки о языке заключается в том, что переход к смыслу сообщения возможен лишь после того, как речевой сигнал преобразован в последовательность дискретных элементов. Более частным является представление о том, что эта последовательность может быть изображена цепочкой символов-фонем и что число фонем в каждом языке очень мало (так, по мнению московской фонологической школы, в русском языке существует 39 фонем). В соответствии с этим основным положением модель восприятия речи должна включать блок фонетической интерпретации, обеспечивающий переход от речевого сигнала к последовательности элементов.

Проблема фонетической интерпретации первоначально формулировалась как проблема распознавания фонем, набор которых принят лингвистами для описания данного языка. При этом пред-

полагалось, что вся информация, содержащаяся в описании сигнала, полученном в результате фонетической интерпретации, исчерпывается цепочкой фонем. В настоящее время ни у кого не вызывает сомнений, что такая формулировка проблемы была ошибочной. Таким образом, снова появляется необходимость определить требования к описанию сигнала на выходе блока фонетической интерпретации, исходя при этом из того, какое описание необходимо для обработки информации на следующих, более высоких уровнях и какое описание в принципе можно получить из речевого сигнала.

## 2.1. ОПИСАНИЕ СМЫСЛА И ПРЕДШЕСТВУЮЩИЕ ЕМУ ОПИСАНИЯ СООБЩЕНИЯ

Настоящий раздел главы предназначается для читателя, не имеющего специального лингвистического образования. В нем делается попытка популярно объяснить, что для понимания смысла сообщения необходим синтаксический анализ предложения, который в свою очередь предполагает предварительный морфологический анализ, а последний требует фонемного или близкого к фонемному описания речевого сигнала. Для достижения этой цели мы сочли возможным избрать один весьма частный подход к описанию смысла и использовать ряд упрощающих предположений (подробно о проблеме «текст  $\leftrightarrow$  смысл» см. в работах [36, 61, 103, 104]).

Попытаемся прежде всего представить себе, что может подразумеваться под смыслом услышанного предложения. Для этого примем, что человек является информационной системой и что любое услышанное сообщение потенциально является либо информацией, которая должна быть записана в долговременную память системы, либо запросом, требованием выдать информацию, хранящуюся в системе. Реальная судьба сообщения будет, естественно, зависеть от того, что система уже знает, есть ли в ней запрашиваемая информация, является ли сообщаемая информация новой, не приходит ли она в противоречие с уже имеющимися сведениями.

Для того чтобы решить, что делать с полученным сообщением, и быть способной или приобщить его к уже известным данным, или найти запрашиваемую им информацию, система должна располагать удобным для обозрения описанием сообщения, согласованным с описанием данных в долговременной памяти.

Существует мнение [89], что единицей, используемой для описания (упорядочения) сведений в долговременной памяти человека, является событие. Событие имеет примерно один-два десятка признаков, имеющих смысл вопросов о том, какое действие совершается, кто его совершает, на что оно направлено, что используется в качестве инструмента, где происходит событие, когда оно происходит (происходило) и т. д. Характеристиками события по

признакам, т. е. ответами на эти вопросы, являются или понятия, или другие события.

Не останавливаясь на обсуждении привлекательных сторон такого подхода (например, возможность увеличения упорядоченности сведений с накоплением опыта), обратим внимание на то, что признаки события приблизительно соответствуют членам предложения, тем вопросам, на которые отвечают члены предложения (сказуемое, подлежащее, обстоятельство места, обстоятельство времени и т. д.).

Примем для простоты, что простое предложение является сообщением только об одном событии, причем в качестве характеристик события выступают только понятия (а не другие события). Будем также, заведомо упрощенно, считать, что каждому признаку события соответствует специальный член предложения. Тогда описание события, о котором сообщается в простом предложении, можно представить себе как частично заполненную анкету со стандартным числом вопросов (каждый вопрос — возможный признак события и одновременно возможный в языке член предложения). Это описание можно условно принять в качестве смысла предложения, заметив, однако, что в анкете должна быть по крайней мере еще одна графа, указывающая, является ли предложение запросом информации (вопросом) или сообщением информации.

Мы априорно приняли, что ответом на вопрос анкеты должно быть понятие. Несомненно, что понятие соответствует не одному частному объекту, но множеству объектов, оно обозначается в большинстве случаев не одним словом, но группой слов-синонимов.

Для наглядности примем, что понятие — это номер, указывающий множество слов разных языков, обозначающих одно и то же. Тогда становится вполне понятным, что по анкете уже нельзя восстановить язык, на котором было произнесено предложение. Таким образом, описание смысла предложения с помощью анкеты удовлетворяет требованию свободы смысла сообщения от конкретного языка, на котором оно сделано.

Теперь мы можем определить задачу понимания смысла предложения как задачу заполнения указанной анкеты.

Легко заметить, что для этого нужно решить два вопроса: найти номера граф анкеты, на которые отвечает предложение, и записать в каждую графу ответ. Для рассматриваемого идеализированного случая, где каждой графе соответствует специальный член предложения, задача сводится к тому, чтобы определить, какие члены предложения есть в данном предложении, идентифицировать каждый член предложения (найти подлежащее, сказуемое и т. д.) и записать соответствующее ему конкретное понятие.

Информация о номере графы анкеты передается с помощью разнообразных грамматических средств, набор которых различен в разных языках. Существовало, что для описания этих средств необходимо пользоваться понятием слова. Система, осуществляющая



щая синтаксический анализ предложения (упрощенно говоря, производящая поиск подлежащего, сказуемого и т. д.), должна трактовать предложение как состоящее из слов. Тогда она может воспользоваться такими характеристиками, как порядок слов в предложении, грамматический класс слова (глагол, существительное и т. д.), грамматическое значение слова (падеж, число и т. д.), наличие/отсутствие грамматической связи между словами.

Слово как единица необходимо и для описания конкретных понятий. Заметим, что классификация конкретных объектов может осуществляться по нескольким различным признакам. Соответственно конкретное понятие может быть описано «аналитически» — путем указания двух или более классов, к которым оно одновременно принадлежит. Такой метод широко используется в языке. Например, слово *козленок* определяет конкретное понятие через два класса (*{козленок, козел, коза}* и *{козленок, слоненок, львенок и т. д.}*).

Для перехода от предложения к его смыслу необходимо представить предложение как последовательность единиц, где единицы — это организованные группы информации о слове, его характеристики. Характеристики слова разделяются на лексические (номер конкретного понятия или классов понятий при аналитическом описании) и грамматические (грамматический класс слова, падеж, число, залог и т. д.).

Из этого следует, что модель восприятия речи обязана включать промежуточный блок, который обеспечивал бы представление сообщения в виде последовательности слов, описанных в терминах их лексических и грамматических характеристик. Обработка информации, выполняемая этим блоком, обозначается как морфологический анализ. Все представления о правилах морфологического анализа основываются на допущении, что на входе этого блока слово представлено как последовательность дискретных элементов-фонем.

Понятие «фонема» является весьма конструктивным, так как оно не только позволяет связать слово с соответствующими ему артикуляторными жестами и речевым сигналом, но и связывает слуховое изображение речевого сигнала с движениями. Не используя понятий, близких по смыслу к фонемам, нельзя описать таких явлений, как имитация (речевое воспроизведение) услышанного бессмысленного звукосочетания или его буквенная запись.

Как только слово определено как последовательность фонем, появляются понятия начала и конца слова. «Входными» признаками слова, которые можно использовать при определении его грамматических и лексических характеристик, становятся последовательности фонем, расположенные на определенном месте в слове. Такие последовательности называются грамматическими и лексическими морфемами. В частном случае морфема может состоять всего из одной фонемы.



Тот факт, что человек действительно производит морфологический анализ, доказывается возможностью частичного понимания смысла фраз, составленных из незнакомых человеку слов.

Приведем в качестве примера фразу, придуманную известным русским фонетистом Л. В. Щербой: *Глокая куздра штеко будланула бокра и курдячит бокренка*.

Носители русского языка не только однозначно определяют синтаксическую структуру этой фразы, они согласны в том, что участниками описываемого события являются животные, что *бокренка* — это детеныш *бокра*; представления о характере упоминаемых действий также оказываются довольно близкими у разных людей [2]. Легко заметить, что такой результат был бы невозможным, если бы человек пользовался нерасчлененным на фонемы суммарным образом слова и производил бы не анализ слова, а его идентификацию с эталоном (число эталонов равно числу возможных слов).

Приняв, что человек производит морфологический анализ слов и пользуется при этом фонемным или близким к фонемному «входным» описанием речевого сигнала, мы при этом оставляем открытыми еще множество вопросов. Они касаются как процедур морфологического анализа, так и самого «входного» описания (исчерпывается ли оно последовательностью фонем, что такое фонема, какой набор фонем использует человек).

Легко, например, показать, что 39 фонем русского языка нам отчетливо не хватает. Так как в определение морфемы входит понятие ее места в слове, мы обязаны допустить, что в описании речевого сигнала на выходе блока фонетической интерпретации или используется еще дополнительная «фонема», соответствующая знаку пробела между словами, или есть какая-то группа особых конечных (начальных) фонем, или существуют какие-то особые знаки, позволяющие производить членение на слова.

Имея в виду не абстрактное описание языка, а модель восприятия речи, мы должны также потребовать, чтобы фонема реально существовала, т. е. чтобы в речевом сигнале была информация, на основе которой фонетический интерпретатор сможет выработать фонемный символ.

Необходимо подчеркнуть, что вопрос об описании речевого сигнала, используемом при морфологическом анализе (преобразовании фонетической последовательности в слова — их грамматические и лексические характеристики), тесно связан с вопросом о процедурах этого морфологического анализа. В следующих разделах главы мы будем рассматривать эти вопросы совместно.

## 2.2. ПОТЕРИ ИНФОРМАЦИИ В УСТНОЙ РЕЧИ

Существующее представление о том, что в устной речи какая-то часть информации теряется, основано, естественно, на допущении, что где-то изначально эта информация существовала.

Для удобства изложения мы примем, что на каком-то уровне описания предложения при речеобразовании слова отделены друг от друга и каждое слово представлено своей фонематической транскрипцией. При описании потерь фонемной информации используется понятие позиционных и комбинаторных вариантов фонем [72]. В частности, очень значительные потери связаны с позиционными вариантами фонем, т. е. с положением фонемы в безударном (заударном, предударном) слоге [21].

Пять (по представлениям московской фонологической школы) русских гласных фонем реализуется лишь в положении под ударением. В безударном положении число отличающихся друг от друга вариантов гласных резко сокращается. Особенно сильной редукции подвергаются гласные в заударных слогах. Так как в русском языке значительная часть грамматической информации передается с помощью окончаний слов — флексий, очень важен вопрос о фонетических характеристиках заударных флексий. Бондарко и Вербицкая [14] показали, что гласные элементы заударных флексий подвергаются столь же сильной редукции, что и заударные гласные, не несущие грамматической нагрузки.

Из этого следует, что полного описания слова в терминах его лексических и грамматических характеристик принципиально нельзя получить на основании анализа речевого сигнала, соответствующего одному этому слову.

Обратимся теперь к вопросу о том, есть ли в устной речи какие-то сигналы, которые играли бы роль пробелов между словами, т. е. указывали бы, где кончается одно слово и начинается другое слово. Такие предполагаемые сигналы называются в русской фонетической литературе «пограничными сигналами».

В работе [20] были исследованы динамические спектрограммы и осциллограммы 300 фраз, составленных таким образом, что одно и то же сочетание фонем гласный + согласный (ГС) или согласный + гласный (СГ) то принадлежало одному слову, то относилось к двум разным соседним словам. Например, *Ракета кажется старой* и *Ракет окажется мало*.

Результаты показали, что при произнесении одной и той же фразы (где С и Г относятся к разным словам) наблюдаются два крайних случая. В одном случае между согласным и гласным имеется отчетливая пауза (длительностью 55—80 мс), в другом случае нет ни паузы, ни каких-либо иных акустических признаков границы между словами. Из всех случаев, когда наблюдалась пауза, 95% приходилось на фразы, где первое слово кончалось согласным (С—пауза—Г) и только 5% — на фразы, где первое слово заканчивалось гласным (Г—пауза—С). Во второй части работы исследовалась способность слушателей произвести членение на слова по вырезанной части фразы. Оказалось, что при наличии паузы граница между словами определялась практически безошибочно. При отсутствии паузы ответы достоверно не отличались от случайных.

Вывод о том, что специальные сигналы (в частности, пауза), указывающие место границы между словами, могут наблюдаться в устной речи, но что это отнюдь не обязательно, был сделан также рядом исследователей, работавших с другими языками ([<sup>132</sup>: 412], см. также главу 5).

Еще один вполне очевидный источник потерь информации в случае устной речи составляют разнообразные акустические помехи и искажения, наличие которых в естественной ситуации речевого общения можно считать скорее правилом, чем исключением.

Приведенные данные исключают предположение, что морфологический анализ устной речи (преобразование фонетической последовательности в слова, описанные по их лексическим и грамматическим характеристикам) производится поочередно слово за словом. В случае устной речи последовательность фонетических элементов, соответствующих слову, может быть не выделена, а если она и выделена, то ее фонемный состав может быть или не полностью определенным, или искаженным.

Это означает, что вместо одного слова мы в лучшем случае имеем дело с некоторым набором гипотез о возможных словах. Соответственно необходимо предположить, что уже при выборе между этими гипотезами используются грамматические и семантические ограничения окружающего контекста.

Вопрос о том, как и в каких пределах используются эти ограничения, критически зависит от объема оперативной памяти — длины фонетической последовательности, запоминаемой до перехода к словам.

### 2.3. ОБЪЕМ ОПЕРАТИВНОЙ ПАМЯТИ НА ФОНЕТИЧЕСКИЕ ЭЛЕМЕНТЫ

Исследованию оперативной памяти у человека посвящено колоссальное количество работ. Увлекательный обзор достигнутых результатов содержится в книге Линдсея и Нормана [<sup>89</sup>].

Пожалуй, наиболее интересный вывод из исследований заключается в том, что объем оперативной памяти равен некоторому постоянному числу ( $7 \pm 2$ ) независимых «единиц», предъявляемых для запоминания, и не зависит от количества информации, содержащейся в этих «единицах». Что такое «единица» — формально не определено, однако интуитивно очевидно, что это нечто такое, что человек может описывать признаками, характеризующими это нечто как самостоятельное целое.

Показано, что в качестве единиц, в частности, могут выступать слоги. Это весьма любопытно, так как слог обычно рассматривается как единица речи, но не языка. В описании языка понятие слог не фигурирует — в качестве единиц, более мелких, чем слово, используются морфемы и фонемы.

В психологических экспериментах по исследованию памяти слоги предъявляются изолированно, с паузами между ними. По этой причине полученные данные позволяют лишь предполагать, что человек использует слоги как единицы также и при записи в память фонетической последовательности. Для проверки этого предположения следовало провести такой эксперимент, где стимулами служили бы отрезки нормального речевого потока, но человек не мог бы запоминать их как слова или события, о которых идет речь в сообщении.

В приводимых ниже экспериментах [159] принималось допущение, что в случае сильно искаженных фраз человеку требуется

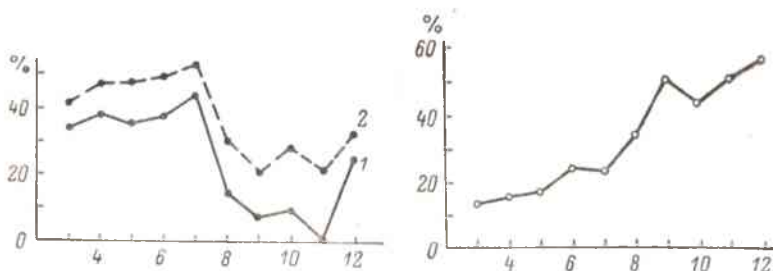


Рис. 2.1. Фразовая (1) и словесная (2) разборчивость при разном числе слогов в сообщении. По [159].

По оси абсцисс — количество слогов в сообщении; по оси ординат — разборчивость (относительное количество правильно воспринятых слов или фраз).

Рис. 2.2. Вероятность пропуска слова при восприятии сообщения заданной длины. По [159].

По оси абсцисс — количество слогов в сообщении; по оси ординат — относительное количество пропущенных слов.

значительное время для перехода от запомненной последовательности фонетических элементов к решению о смысле сообщения. Если это так, а объем оперативной памяти на фонетические элементы ограничен, то предъявление длинных фраз должно приводить не к улучшению разборчивости, как это обычно имеет место (см. обзор [485]), а наоборот, к ее ухудшению за счет пропуска части фразы. Повышение разборчивости с увеличением длины фразы должно наблюдаться до тех пор, пока длина фразы не превысит объема оперативной памяти на фонетическую последовательность.

Всего было записано 10 таблиц, по 20 фраз в каждой. Фразы различной длины (от трех до двенадцати слогов) следовали друг за другом в случайном порядке. Каждая фраза содержала от двух до пяти фонетических слов, составляла одну синтагму и при чтении произносилась слитно, без пауз между словами. Интервал между фразами составлял 20 с. Фразы произносил диктор, для которого русский язык был родным.

Искажения создавались путем пропуска речевого сигнала через полосовой фильтр (906—1141 Гц). Фразовая разборчивость, определенная по всему материалу суммарно, составила 30%. Зависимость фразовой (1) и словесной (2) разборчивости от числа слогов во фразе показана на рис. 2.1. Можно видеть, что разборчивость несколько возрастает в пределах до семи слогов, затем резко падает. Это падение обусловлено преимущественно пропуском слов, о чем свидетельствует рис. 2.2.

Способность человека имитировать речевой сигнал без понимания его смысла дает еще одну возможность исследования объема оперативной памяти на фонетическую последовательность. Хурмэ провел эксперименты [304], в которых исследовалась имитация финских фраз разной длины десятью русскими испытуемыми, которые совсем не знали финского языка. Записанные на пленку фразы-ответы затем прослушивались и оценивались группой слушателей-финнов. Зависимость числа правильно воспроизведенных русскими финских фраз от количества слогов во фразе приведена на рис. 2.3. Видно, что фразы длиной в 5 слогов воспроизводятся достаточно хорошо; воспроизведение фраз, состоящих из 9 слогов, становится для испытуемых практически невозможным.

Анализ зависимости вероятности правильного воспроизведения слога от его положения (порядкового номера) во фразе показал, что для фраз, состоящих из 5 и 6 слогов, отчетливой зависимости заметить не удастся. Начиная с фраз длиной в 7 слогов проявляется хорошо известный эффект места элемента в последовательности — первые и последние слоги во фразе правильно воспроизводятся чаще, чем слоги, находящиеся в середине фразы.

Заметим, что описание слога как единицы речи не исчерпывается фонемной информацией. Характеристикой слога является также отсутствие или наличие ударения.

Если объем оперативной памяти на фонетическую последовательность определяется действительно числом слогов, а не количеством информации, необходимой для описания этой последовательности, логично ожидать, что те же ограничения проявятся в условиях, когда испытуемым нужно будет запомнить всего только ритмический рисунок последовательности, т. е. порядок следования ударных и безударных слогов.

В экспериментах Лисенко и Федоровой [92] исследовалось речевое воспроизведение ритмического рисунка последовательности тональных посылок в виде последовательности слогов [tatata. . .]. Длительность «безударной» посылки ( $T$ ) равнялась длительности паузы и составляла 88 или 176 мс (в разных вариантах опыта). Длительность «ударной» посылки равнялась  $2T$ . Ритмические рисунки синтезированных последовательностей схематически повторяли ритмические рисунки выбранных естественных русских фраз длиной от 5 до 13 слогов.



Последовательности были записаны на магнитную пленку с большими паузами между ними. Во время паузы испытуемый должен был воспроизвести услышанную последовательность.

Полученные результаты приведены на рис. 2.4. Можно видеть, что более быстрые последовательности, где длительность «слогов» близка к естественной речи, воспроизводятся несколько лучше.

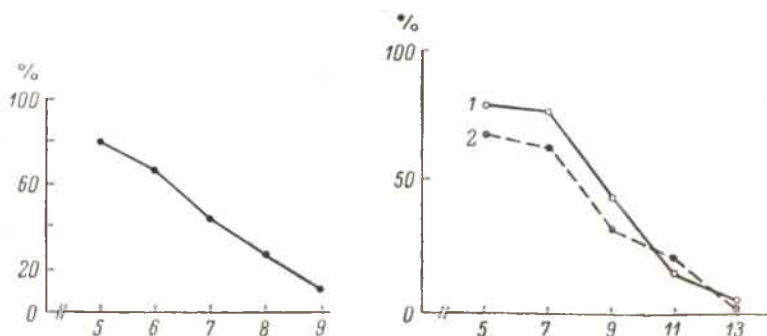


Рис. 2.3. Качество воспроизведения русских испытуемыми финских фраз. По [304].

По оси абсцисс — количество слогов в сообщении; по оси ординат — процент правильно воспроизведенных фраз.

Рис. 2.4. Точность речевого воспроизведения ритмического рисунка tonальной последовательности. По [92].

По оси абсцисс — количество посылок тона в стимуле; по оси ординат — процент правильно воспроизведенных стимулов. 1 — стимулы с «короткими» посылками; 2 — стимулы с «длинными» посылками.

Существенно, что для обоих вариантов эксперимента характерно отчетливое ухудшение воспроизведения ритмического рисунка последовательности при увеличении числа элементов последовательности. Когда число элементов больше семи, правильное воспроизведение наблюдается менее чем в 50% случаев.

Таким образом, все приведенные выше данные достаточно хорошо согласуются между собой и позволяют считать, что объем оперативной памяти на последовательность фонетических элементов действительно составляет примерно семь слогов.

## 2.4. ЗАДАЧА ПЕРЕХОДА ОТ ПОСЛЕДОВАТЕЛЬНОСТИ ФОНЕТИЧЕСКИХ ЭЛЕМЕНТОВ К СЛОВАМ ЯЗЫКА

Семь слогов соответствуют не более чем двум-трем знаменательным русским словам. Таким образом, приходится считать, что модель «видит» как последовательность фонетических элементов только часть анализируемого предложения. Эта «фонетически видимая» часть соответствует настоящему и ближайшему прошлому речевого сигнала. Дальнее прошлое сигнала



(им будет, например, начало длинного предложения в момент окончания предложения) должно, очевидно, запоминаться или в форме слов как единиц языка (описанных в лексических и грамматических терминах), или в неязыковой форме.

Очевидно, что модель морфологического анализатора, преобразуя последовательность фонетических элементов в слова языка, может воспользоваться тем обстоятельством, что соседние слова в пределах одной синтагмы в естественных предложениях в большинстве случаев связаны по форме и по смыслу.

Прежде чем обсуждать возможные способы использования этой грамматической и семантической связанности, подчеркнем тот факт, что соседние слова в предложении могут быть синтаксически независимыми и семантически несовместимыми. Например, слова *куртке* и *громко* во фразе *Человек в кожаной куртке громко закричал* являются независимыми, они принадлежат разным синтаксическим группам.

Если программа основывается в своей работе на гипотезе о связанности соседних слов, она, естественно, должна получать информацию о том, когда эту гипотезу нельзя использовать. Иначе говоря, необходимы специальные сигналы, указывающие границы между синтаксическими группами; информация о границах групп не только также должна записываться в оперативную память, она должна быть как-то привязана к фонетическим элементам.

Обратимся теперь к случаю, когда на гипотезу о связанности соседних слов можно положиться. Основная трудность в этом случае состоит в том, что на уровне последовательности фонетических элементов слова не отделены друг от друга (могут быть не отделены), а проверка грамматической согласованности может быть осуществлена только после того, как слова выделены.

Это обусловлено тем, что русские грамматические морфемы по своему фонемному составу, а следовательно и по своим фонетическим и акустическим характеристикам, в принципе не отличаются от других возможных сочетаний фонем внутри слова и на стыках слов. Их особенностью является то, что они находятся либо в конце слова (флексии), либо перед словом (предлоги), либо в начале слова (приставки).

Кажется, что единственный разумный способ поведения в такой ситуации состоит в том, чтобы осуществлять пробные членения последовательности и проверять, получится ли при данном членении сочетание из грамматически согласованных осмысленных слов.

Применение такого способа требует какой-то стратегии пробных членений, обеспечивающей минимизацию времени поиска и ошибок членения. Иначе говоря, у человека должны быть какие-то правила применения пробных членений, использующие как свойства самой входной последовательности фонетических элементов, так и общие сведения о языке (речи).

## 2.5. ЧЛЕНЕНИЕ НА «СЛОВА» БЕССМЫСЛЕННЫХ РЕЧЕВЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ И ПСЕВДОПРЕДЛОЖЕНИЙ

Указания на то, что носители русского языка используют какие-то стандартные правила членения фонетической последовательности, были впервые получены в работе Лисенко [90]. Эксперименты преследовали другую цель — автор хотел выяснить, совпадают ли услышанные «слова» с теми, которые были произнесены диктором.

Слушателям (16 человек) была предъявлена магнитофонная запись бессмысленных «двухсловных» сочетаний (табл. 2.1), произнесенных диктором с разными вариантами членения на «слова». Из таблицы следует, что испытуемые не воспринимали

Т а б л и ц а 2.1

Членение человеком на «слова» бессмысленных слогосочетаний

| Передаваемые сообщения<br>и их ритмический рисунок                                                              |   | Количество принятых сообщений<br>со следующими ритмическими рисунками: |    |   |   |    |   |    |
|-----------------------------------------------------------------------------------------------------------------|---|------------------------------------------------------------------------|----|---|---|----|---|----|
|                                                                                                                 |   | ↗                                                                      | ↘  | ↗ | ↘ | ↗  | ↘ | ↗  |
|                                                                                                                 |   |                                                                        |    |   |   |    |   |    |
|                                                                                                                 |   | —                                                                      | —  | — | — | —  | — | —  |
|                                                                                                                 |   | ↗                                                                      | ↘  | ↗ | ↘ | ↗  | ↘ | ↗  |
| му <sup>о</sup> ма <sup>о</sup> ка <sup>о</sup> би <sup>о</sup> ни <sup>о</sup> та <sup>о</sup>                 | { | ↗                                                                      | —  | — | — | —  | — | —  |
|                                                                                                                 | { | ↗                                                                      | —  | — | — | —  | — | —  |
|                                                                                                                 | { | ↗                                                                      | —  | — | — | —  | — | —  |
| на <sup>о</sup> ви <sup>о</sup> ла <sup>о</sup> ки <sup>о</sup> ша <sup>о</sup> ма <sup>о</sup>                 | { | —                                                                      | 15 | — | — | —  | — | —  |
|                                                                                                                 | { | —                                                                      | 15 | — | — | —  | — | —  |
|                                                                                                                 | { | —                                                                      | 15 | — | — | —  | — | —  |
| ли <sup>о</sup> ба <sup>о</sup> та <sup>о</sup> пу <sup>о</sup> ка <sup>о</sup> то <sup>о</sup>                 | { | 2                                                                      | 11 | — | — | —  | — | —  |
|                                                                                                                 | { | 4                                                                      | 8  | 1 | — | —  | — | —  |
|                                                                                                                 | { | 3                                                                      | 11 | — | — | —  | — | —  |
| на <sup>о</sup> ви <sup>о</sup> та <sup>о</sup> ки <sup>о</sup> на <sup>о</sup> ма <sup>о</sup>                 | { | 1                                                                      | 14 | — | — | —  | — | —  |
|                                                                                                                 | { | 1                                                                      | 13 | — | — | —  | — | —  |
|                                                                                                                 | { | 2                                                                      | 11 | — | — | —  | — | —  |
| га <sup>о</sup> ти <sup>о</sup> ла <sup>о</sup> та <sup>о</sup> ба <sup>о</sup> па <sup>о</sup>                 | { | —                                                                      | —  | — | — | 15 | — | —  |
|                                                                                                                 | { | —                                                                      | —  | — | — | 15 | — | —  |
|                                                                                                                 | { | —                                                                      | —  | — | — | 14 | 1 | —  |
| ла <sup>о</sup> то <sup>о</sup> би <sup>о</sup> ни <sup>о</sup> то <sup>о</sup> на <sup>о</sup>                 | { | —                                                                      | —  | — | 1 | 14 | — | —  |
|                                                                                                                 | { | —                                                                      | —  | — | — | 15 | — | —  |
|                                                                                                                 | { | —                                                                      | —  | — | — | 15 | — | —  |
| па <sup>о</sup> та <sup>о</sup> ва <sup>о</sup> ту <sup>о</sup> ба <sup>о</sup> па <sup>о</sup> ка <sup>о</sup> | { | —                                                                      | —  | — | — | —  | — | 30 |
|                                                                                                                 | { | —                                                                      | —  | — | — | —  | — | 32 |
|                                                                                                                 | { | —                                                                      | —  | — | — | —  | — | —  |

Примечание. <sup>о</sup> обозначает ударный слог; — обозначает неударный слог; | обозначает границу между словами.

намерений диктора в отношении членения, что хорошо согласуется с данными Бондарко и др. [20], приведенными в разделе 2.2.

Неожиданным результатом эксперимента оказалось поразительное единообразие членений у всех слушателей, хотя, казалось бы, отсутствие грамматической и семантической информации должно было привести к случайному поведению. Некоторые представления о том, на какой информации основывается членение, были предложены Клаас [79].

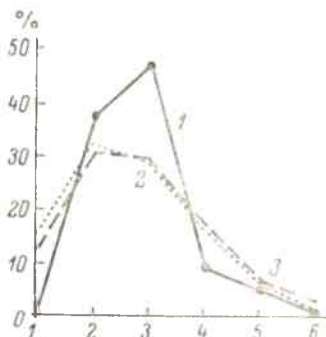
Исходной задачей автора было исследование влияния грамматического контекста псевдопредложения на решение о границе между первым и вторым словами в предложении.

Были взяты пары фраз типа:

1. а) *Рѡхѡм бѣха не перетѡльпѣшь*  
 б) *Рѡх амбѣха затѡльпѣл*
2. а) *Тифѣльнику ганзѣтъ несантъчно*  
 б) *Тифѣльник уганзѣл всантък и т.д.*

Рис. 2.5. Сравнение результатов членения бессмысленных предложений на псевдослова (1) с распределением русских слов по количеству слогов (2, 3). По [79].

По оси абсцисс — количество слогов в слове;  
 по оси ординат — частота встречаемости слов  
 соответствующей длины.



Фразы были в случайном порядке записаны на магнитную пленку и предъявлены слушателям.

Гипотезу о влиянии контекста проверить не удалось, так как примерно в 40% случаев слушатели членили предложения иначе, чем предполагалось автором (письменным текстом) и, соответственно, диктором. При этом слушатели вели себя весьма единообразно, а записанные ими предложения были грамматически правильными (об этом можно было судить по орфографии). Пытаясь найти причину отклонений, автор проанализировала все содержащиеся в записях испытуемых псевдослова с точки зрения количества слогов и положения ударения. Оказалось, что полученные распределения достаточно хорошо согласуются с данными по статистике русского языка [71, 164].

На рис. 2.5 приведены распределения русских слов по числу входящих в них слогов (кривые 2 и 3) и распределения записанных псевдослов по числу слогов (кривая 1). Можно видеть, что кривая 1 является как бы утрированным вариантом кривых 2 и 3.

Заметим, что процедура пробных членений на слова наверняка требует предварительного обнаружения ударения (ударного гласного или слога) как опорного сигнала. Если два таких опорных сигнала (два ударения) уже есть, можно установить пробную границу так, чтобы число заударных слогов первого слова и число

предударных слогов второго слова наилучшим образом соответствовало статистике русской речи. При этом, естественно, подразумевается, что процедура умеет в каком-то смысле считать и что «считаемыми» единицами являются именно слоги (или слогообразующие гласные), но не фонемы вообще.

## 2.6. ФОНЕТИЧЕСКИЙ ОБРАЗ СЛОВА

Можно заметить, что до сих пор мы все время пользовались допущением, что существует некоторый алгоритм, который, получив на вход последовательность фонетических элементов, трактует ее как слово, т. е. пытается описать ее в терминах грамматических и семантических характеристик. Если описание окажется непротиворечивым (например, не возникнет такая ситуация, что слово определено как глагол в родительном падеже), то оно условно принимается как возможное и сравнивается с описаниями соседних слов.

Вне зависимости от того, как конкретно организован такой алгоритм, он, несомненно, весьма часто будет сталкиваться с такими случаями, когда по входной последовательности нельзя полностью определить слова. Например, входная последовательность не совпадает полностью ни с одним словом русского языка, но частично совпадает сразу с несколькими русскими словами.

Если соседние слова также не снимают неопределенности, очевидный выход в такой ситуации состоит в том, чтобы предпочесть фонетически наиболее похожее слово. Понятно, что величина сходства будет определяться тем, какие признаки используются для описания фонетического образа слова.

В теоретической фонологии принимается, что слово является последовательностью фонем, где каждая фонема в свою очередь описана набором значений по дифференциальным признакам (более подробно об этом см. главу 4). В таком случае максимально близкими словами считаются слова, отличающиеся по одному дифференциальному признаку, например *дочка* и *бочка*, *кот* и *пот*.

Легко заметить, что в этих примерах подразумевается, что сходство определяется путем поэлементного сравнения двух слов: первая фонема одного слова сравнивается с первой фонемой другого слова, вторая фонема со второй и т. д. Для поэлементного сравнения логически необходимо, чтобы элементы были пронумерованы. При допущении, что фонетический образ слова есть пронумерованная последовательность элементов, сразу же возникает вопрос, что является точкой отсчета при нумерации и что именно (и как) пронумеровано.

Основными источниками информации для ответа на эти вопросы являются ошибки приема слов при искажениях речи, замены бессмысленных слов на осмысленные при восприятии псевдопредложений, результаты шкалирования сходства звучания слов. Данные, полученные всеми этими методами, являются достаточно

однозначными и подтверждают то, что давно эмпирически известно относительно «ослышек». Необходимым условием близости слов является совпадение их ритмических рисунков (одинаковое число слогов и одинаковое положение ударений).

Замена сочетания из нескольких согласных на один согласный или, наоборот, замена одного согласного на сочетание согласных производится очень легко [79, 159]. Оценки субъективной близости слов в таких парах, как *мали́на*—*кали́на* и *мали́на*—*масли́на*, примерно совпадают.

Можно видеть, что эти результаты несовместимы с представлением о пофонемной нумерации. Удобный способ нумерации заключается в том, что за точку отсчета (элемент с номером, равным нулю) принимается ударный гласный. Относительно него нумеруются остальные гласные. Согласному (или сочетанию согласных) приписываются два номера, соответствующие тем гласным, между которыми этот согласный находится; внутри сочетания согласных применяется собственная нумерация, т. е. сочетание согласных рассматривается как подпоследовательность.

## 2.7. ТРЕБОВАНИЯ К ОПИСАНИЮ ФОНЕТИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Из приведенных ранее данных с несомненностью следует, что записанная в оперативной памяти последовательность фонетических элементов должна содержать в явном (выделенном) виде как фонемную, так и просодическую информацию.

Относительно фонемной информации можно утверждать:

1) выгодно сохранить позиционные варианты гласных, отражающие положение гласного в слове относительно ударения, т. е. записываемая информация должна ближе соответствовать фонетической, чем фонематической транскрипции русских гласных;

2) очевидно, что следует избегать необоснованных решений; с этой точки зрения было бы удобно описывать фонему не символом, но ее значениями по дифференциальным признакам; отсутствие решения по признаку было бы эквивалентно указанию группы возможных фонем;

3) гласные и согласные должны быть в памяти разделены;

4) запись согласных должна обеспечивать возможность отнесения согласного как к предыдущему, так и к последующему слогу, т. е. предпочтительно, чтобы при преобразовании речевого сигнала в фонетическую последовательность вопрос об открытом или закрытом характере слога остался нерешенным.

Относительно просодической информации можно только утверждать, что она должна быть представлена в максимально готовом для использования виде. Нужны сигналы, указывающие границы между грамматически и семантически несогласованными словами, и очень важно, чтобы решение об ударенности слога (гласного) принималось достаточно быстро, так как до обнаружения ударенного гласного нельзя начинать анализ слова.



## ФОНЕТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕЧЕВОГО СИГНАЛА

### Глава 3

#### ПРИРОДА ФОНЕТИЧЕСКОГО ОБРАЗА

В основе науки о языке лежит допущение о том, что речевой сигнал представлен в сознании слушателя последовательностью дискретных фонетических элементов. Заметим, что справедливость этого допущения самой лингвистикой не проверяется. Важно, что, не приходя в противоречие с эмпирически известными фактами, это допущение позволяет дальше строить конструктивную теорию как самого языка, так и процесса речевого общения.

Нужно подчеркнуть, что эти дискретные фонетические элементы, по предположению, появляются лишь в «сознании», т. е. в изображении сообщения, записанном в оперативную память слушателя. В самом речевом сигнале они присутствуют в неявном виде, для их получения нужна определенная обработка сигнала.

Для исследователя, изучающего восприятие речи человеком или разрабатывающего систему автоматического распознавания речи, указанное допущение означает, что модель (система) должна включать блок, превращающий звуковой речевой сигнал в последовательность дискретных фонетических элементов. Действия, выполняемые этим блоком, обычно обозначаются как фонетическая интерпретация сигнала.

Предметом настоящей главы является рассмотрение фундаментального вопроса теории фонетической интерпретации — вопроса об отношениях между фонетическим элементом и отрезком (сегментом) речевого сигнала.

#### 3.1. ПОНЯТИЕ ФОНЕТИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Для того чтобы отчетливо представить себе, что подразумевается под понятием фонетической последовательности, примем, что оперативная память представляет собой ряд ячеек; в каждую ячейку записывается только один фонетический элемент,  $(k+1)$ -ый элемент записывается в ближайшую ячейку справа относительно той, куда записан  $k$ -тый элемент. Записывае-



мой в ячейку информации являются характеристики данного фонетического элемента. Некоторые из этих характеристик можно представить как числа, указывающие значения данного элемента по определенным признакам; другие характеристики соответствуют названиям классов, к которым принадлежит данный элемент.

Отрезок фонетической последовательности, содержащийся в настоящем моменте времени в оперативной памяти, можно в таком случае рассматривать как функцию, описывающую связь между содержанием ячейки и ее порядковым номером. Подчеркнем, что в отличие от осциллограммы, динамической спектрограммы и ряда других изображений речевого сигнала как физического процесса независимой переменной (аргументом) для фонетической после-

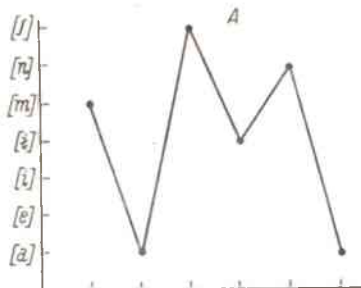
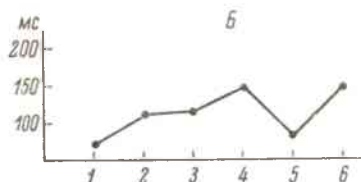


Рис. 3.1. Представление связи порядкового номера звука речи в слове *машина* с его фонемной характеристикой (А) и характеристикой по длительности (Б).

По оси абсцисс — порядковый номер звука в слове; по оси ординат — фонемные символы (А) и длительность звука (Б).



довательности является не время, а порядковый номер элемента. Соответственно значения аргумента в данном случае принципиально дискретны. Значения функции — характеристики элемента по признакам — могут, по крайней мере в некоторых случаях, трактоваться как непрерывные.

Легко заметить, что преобразование речевого сигнала в фонетическую последовательность требует решения двух задач. Одна из них состоит в обнаружении моментов смены элементов и переходе к следующей по порядку ячейке памяти. Вторая задача заключается в установлении значений характеристик, которые должны быть приписаны элементу — записаны в ячейку.

При описании фонетической последовательности используются понятия «фонемный состав» и «ритмико-мелодический рисунок». Естественно считать, что эти понятия относятся к одному и тому же явлению, т. е. что в обоих случаях подразумеваются те же самые фонетические элементы с теми же самыми порядковыми номерами; речь идет о разных характеристиках (группах характеристик) этих фонетических элементов.

Например, если мы условно примем, что фонетический элемент соответствует звуку речи, то фонемный состав фонетической по-

последовательности, соответствующей слову *машина*, можно представить в виде графика *A* на рис. 3.1. Ритмический рисунок этой же последовательности (если мы примем для простоты, что он определяется абсолютными значениями длительностей элементов) можно представить графиком *B* на рис. 3.1.

Ось ординат на рис. 3.1, *A* соответствует порядковому номеру фонемы в списке фонем. Подчеркнем, что этот номер условен, — он зависит от того, по какому принципу составлен алфавит — упорядоченный в ряд список фонем. Значения по оси ординат, которые может принимать элемент, в данном случае принципиально дискретны. Символ фонемы — условное название класса, к которому относится звук речи.

Ось ординат на рис. 3.1, *B* соответствует длительности. Значения по этому признаку можно считать непрерывными. (Переход к дискретному описанию — классификация звуков на ударные и неударные — требует дополнительной обработки последовательности).

Легко заметить, что для превращения фонетической последовательности в слово, описанное в терминах его лексических и грамматических характеристик, необходимы и график *A*, и график *B*. Имея в своем распоряжении только график *A*, мы не сможем выбрать между словами *машина* и *Машина* (принадлежащая Маше).

Для совместного использования разных характеристик при определении слова необходимо в данном случае, чтобы эти характеристики были как-то «привязаны» друг к другу. Предположение о том, что они записаны в одной и той же ячейке памяти, является просто наглядным способом представления идеи о связанности данных.

Под фонетическим элементом понимается, с одной стороны, некоторая единица организации (связанности) данных, получаемых в результате фонетической интерпретации. С другой стороны, под фонетическим элементом понимается отрезок речевого сигнала, обработка которого обеспечивает получение этой единицы.

Рассмотрим, какие преобразования предполагаются необходимыми для получения ритмического рисунка и фонемного состава фонетической последовательности и какие фонетические элементы при этом подразумеваются.

### 3.2. РИТМИЧЕСКИЙ РИСУНОК ФОНЕТИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Ритмический рисунок фонетической последовательности чрезвычайно интересен тем, что в качестве основной характеристики элемента выступает длительность. Длительность является признаком отрезка физического процесса; измерять длительность можно лишь после того, как речевой сигнал сегментирован, определены моменты начала и конца отрезка.

Фонетическим элементом, принимаемым при описании ритмического рисунка, обычно служит слог. Это основывается на том, что фонетические последовательности, соответствующие двум разным словам или двум разным строкам стиха, воспринимаются как ритмически совпадающие, если число слогов в этих последовательностях одинаково и порядковые номера ударенных слогов совпадают [144]. Число звуков речи (фонем) может при этом быть различным, однако это касается только согласных, так как общепринято, что каждый слог обязательно включает гласный, причем только один гласный (случаи, когда роль гласного выполняет слогаобразующий сонант, в русском языке сравнительно редки, и мы пока их рассматривать не будем).

Если элементом является слог, то самое простое предположение состоит в том, что длительность, присваиваемая элементу, — это длительность слогового отрезка в речевом сигнале.

Один из способов экспериментальной проверки этого предположения заключается в том, чтобы получить субъективные оценки длительностей слогов во фразе и выяснить, коррелируют ли они с длительностями слоговых отрезков. Результаты эксперимента [143] показали, что субъективная длительность слога коррелирует с длительностью не слогового отрезка, но отрезка гласного.

Другой возможный способ проверки состоит в таком изменении длительностей гласного и согласного в слоге, чтобы суммарная длительность слога сохранялась постоянной. Если сделанное предположение верно, то ритмический рисунок не должен при этом меняться и должен наблюдаться эффект «компенсации» длительностей гласного и согласного. Эксперименты показали, что эффекта «компенсации» нет [303] и что ударенность слога определяется длительностью отрезка гласного [31].

Эти данные не опровергают предположения, что именно слог является фонетическим элементом. Однако они говорят о том, что обработка речевого сигнала должна обеспечивать выделение отрезков гласных.

Обратим внимание на два существенных обстоятельства. Во-первых, ритмический рисунок фонетической последовательности в первом приближении не зависит от фонемного состава этой последовательности и, соответственно, от конкретных спектральных характеристик речевого сигнала на отрезках гласных и согласных. Во-вторых, ритмический рисунок может быть передан с помощью сигнала, состоящего из тональных посылок (см. главу 2), или сигнала, образуемого при «произнесении» фразы с закрытым ртом [159]. В обоих этих случаях вместо согласных в сигнале имеется пауза.

Приходится считать, что отрезки, на которые сегментируется речевой сигнал, должны соответствовать не согласным и гласным, но гласным и негласным, где негласный может быть как согласным

(сочетанием согласных), так и паузой. Формально это то же самое, что обнаружение (выделение) гласных в речевом потоке.

Понятно, что выделение отрезков гласных в речевом потоке автоматически обеспечивает возможность выделения слоговых отрезков. Слог определяется как интервал между началами соседних гласных в речевом потоке (закрытый слог) или как интервал между концами соседних гласных (открытый слог).

### **3.3. ФОНЕМНЫЙ СОСТАВ ФОНЕТИЧЕСКОЙ ПОСЛЕДОВАТЕЛЬНОСТИ**

Распространенная точка зрения, на которой основывается большинство работ по автоматическому распознаванию речи, заключается в том, что элементами речевого сигнала являются звуки речи и что фонемы распознаются последовательно одна за другой. Исследователи восприятия речи за последнее время все больше склоняются к тому, что фонемы могут распознаваться параллельно и что в качестве фонетического элемента выступает слог [19, 195, 220]. Далее мы начнем с изложения этой наиболее привычной точки зрения и попытаемся пояснить разницу между последовательным и параллельным распознаванием фонем.

#### **3.3.1. ФОНЕМНАЯ КЛАССИФИКАЦИЯ ТЕКУЩИХ ЗНАЧЕНИЙ РЕЧЕВОГО СИГНАЛА**

Существо традиционных представлений относительно связи между речевым сигналом и фонемой, используемых в теориях восприятия речи и автоматическом распознавании, сводится к следующим основным допущениям: 1) фонема является характеристикой звука речи — указанием класса, которому этот звук принадлежит; 2) каждому звуку речи соответствует определенное состояние речевого аппарата и соответственно определенная форма спектра звука (и тип источника возбуждения); 3) звуки речи произносятся последовательно один за другим.

На первом этапе развития работ по автоматическому распознаванию речи было сделано дополнительное допущение, что длительность квазистационарных участков, в течение которых сохраняется спектр, типичный для данного звука речи, существенно больше длительности участков, соответствующих переходам от одного звука речи к другому.

Если все эти допущения верны, то для фонемного распознавания речи нужно только научиться классифицировать спектры, т. е. найти оптимальные решающие правила, обеспечивающие переход от значения спектра к фонеме — названию класса спектров. Реализовав эти правила в соответствующем классификаторе и подавая на его вход отклик на речевой сигнал спектрального

анализатора (гребенки фильтров), мы автоматически получим и фонемное распознавание, и сегментацию речевого потока.

Классификатор будет сохранять на выходе один и тот же фонемный символ, пока спектр сигнала будет находиться внутри обозначаемого этим символом класса. Как только значения сигнала выйдут за пределы данного класса, символ нужно записать в ячейку памяти и перейти к следующей по порядку ячейке.

Быстро выяснилось, что системы, построенные по этому принципу, не являются работоспособными. Основная причина неудачи, по распространенному мнению, кроется в том, что относительная длительность переходных участков реально оказалась не меньше, а больше длительности стационарных участков.

Если длительность переходных участков велика, то число «фонем», выработанных классификатором при предъявлении слова, будет неизбежно больше числа фонем, которые обнаруживает в этом слове человек. Так как такой эффект будет наблюдаться в случае любой системы, работающей по принципу фонемной классификации текущих значений сигнала, приходится сделать вывод, что фонемная классификация, осуществляемая человеком, основана на других принципах.

### 3.3.2. ПОСЛЕДОВАТЕЛЬНОЕ ИЛИ ПАРАЛЛЕЛЬНОЕ РАСПОЗНАВАНИЕ ФОНЕМ

В литературе неоднократно указывалось на то, что фонемы «переслаиваются», что один и тот же участок сигнала содержит информацию более чем об одной фонеме и найти такую точку в сигнале, которая соответствовала бы концу одной и началу другой фонемы, в принципе невозможно.

Основанием для этой точки зрения послужили в основном результаты экспериментов по фонемной идентификации синтетических слогов [350, 351] и выделенных отрезков естественного речевого сигнала [59]. Весьма показателен в этом смысле пример с так называемыми переходами второй форманты (см. главы 4 и 12). Контур частоты второй форманты на отрезке, соответствующем артикуляции гласного, закономерно зависит от фонемного качества как этого гласного, так и соседних согласных. Участки, на которых частота  $F_2$  изменяется (они находятся вблизи начала и конца отрезка гласного), названы участками переходов  $F_2$ . На основании характера переходов (направления изменения частоты  $F_2$ ) человек различает согласные фонемы, и по этому признаку переход должен быть отнесен к участку согласного. Однако исключение этих участков приводит к ухудшению разборчивости гласного. Если гласный короткий, то при исключении из него переходов от гласного может вообще ничего не остаться.

Такого рода факты привели к довольно распространенному сейчас представлению, что согласные и гласные фонемы распознаются не последовательно, но параллельно. Вопрос о том, что



формально означает параллельное распознавание и при каких условиях оно допустимо, пока, насколько нам известно, в общем виде не рассматривался. Мы ограничимся тем, что приведем частный пример, на котором видна возможная принципиальная разница между последовательным и параллельным распознаванием.

Допустим, мы знаем, что некоторый наблюдатель обозначает отрезок функции  $x(t)$ , изображенный на рис. 3.2, последовательностью двух символов  $AB$ . Необходимо составить некоторые предварительные гипотезы относительно принципов, которыми он при этом пользуется.

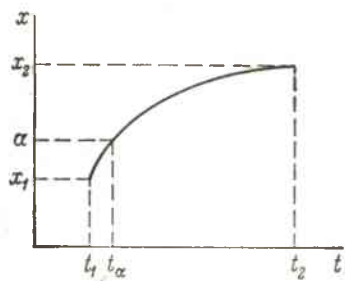


Рис. 3.2.

Одна гипотеза может состоять в том, что где-то между  $x_1$  и  $x_2$  расположена граница между двумя классами значений  $x$ , обозначим ее  $\alpha$ . Пока  $x(t)$  меньше  $\alpha$ , выбирается символ  $A$ , при  $x(t) > \alpha$  выбирается  $B$ . В результате классификации отрезок функции разбивается на два участка, граница между ними соответствует тому значению  $t$ , при котором  $x(t) = \alpha$ . Очевидно, что это чистый случай последовательного распознавания.

Предположим теперь, что  $A$  и  $B$  обозначают разные признаки отрезка. Например,  $A$  обозначает, что  $x$  возрастает ( $\frac{dx}{dt} > 0$ ), а  $B$  обозначает, что среднее по отрезку значение  $x$  находится в определенных пределах  $\left(a_1 < \frac{1}{t} \int_{t_1}^{t_2} x(t) dt < a_2\right)$ . В этом случае оба

символа относятся к отрезку в целом, и естественно говорить о параллельном распознавании. В приведенном примере параллельного распознавания порядок записи символов ( $AB$  или  $BA$ ) не определяется сигналом, он должен быть предусмотрен специальными правилами.

Чистый случай параллельного распознавания необходимо предполагает предварительное указание отрезка, на котором производится анализ — определяются признаки функции. В приведенном примере в признак, обозначаемый символом  $B$ , в явном виде входит понятие отрезка — интегрирование производится в пределах от начала ( $t_1$ ) до конца отрезка ( $t_2$ ).

Очевидно, что вопрос о том, в какой мере (или в какой комбинации) последовательное и параллельное распознавание фоном используется человеком, выгодно решать уже после того, как будут созданы хотя бы предварительные представления относительно сегментации звукового потока.

Заметим, что фонетист или психолог, описывающий связь между динамической спектрограммой речевого сигнала и его



фонемным описанием, свободно пользуется понятиями отрезка гласного и слога.

Использование этих понятий (гласного и слога) в модели фонетической интерпретации означает необходимость включения дополнительного блока обработки информации, заключенного между первичным спектральным анализатором и фонемным классификатором.

Для того чтобы пойти на такое усложнение модели, нужно быть уверенным в том, что понятия отрезка гласного и слога не изобретены учеными для удобства описания наблюдаемых ими явлений, но реально используются человеком в процессе фонетической интерпретации. Экспериментальные данные, приводимые в следующих разделах главы, как кажется, не оставляют в этом сомнений.

### 3.4. ЭКСПЕРИМЕНТАЛЬНЫЕ ДОКАЗАТЕЛЬСТВА ВЫДЕЛЕНИЯ ОТРЕЗКОВ ГЛАСНЫХ ЗВУКОВ В РЕЧЕВОМ СИГНАЛЕ

#### 3.4.1. ВЫДЕЛЕНИЕ НАЧАЛА И КОНЦА ГЛАСНОГО В РЕЧЕВОМ СИГНАЛЕ (ОПЫТЫ ПО СИНХРОНИЗАЦИИ)

Данные о том, что человек не только выделяет моменты начала и конца гласного в воспринимаемом стимуле, но и может использовать соответствующие этим моментам сигналы при управлении речеобразованием, были получены в опытах по синхронизации [93, 161].

Экспериментальный прием состоял в том, что испытуемым предъявлялось периодически повторяющееся (приблизительно 1 раз в секунду) звуко сочетание, записанное на кольцо из магнитной ленты.

Испытуемый должен был произносить то же или другое звуко сочетание «синхронно» с внешним стимулом. Звуко сочетания, использованные в качестве стимулов, и соответствующие им реакции приведены ниже.

|         |   |        |        |        |
|---------|---|--------|--------|--------|
| Стимул  | { | [ata]  | [ata]  | [ata]  |
|         |   | [ada]  | [ada]  | [ada]  |
|         |   | [asa]  | [asa]  | [asa]  |
|         |   | [atma] | [aba]  | [amsa] |
|         |   | [amta] | [adma] | [asma] |
| Реакция |   |        | [amda] | [atsa] |
|         |   |        | [azda] | [afa]  |
|         | { | [ata]  | [ada]  | [asa]  |

Ответы испытуемого (приблизительно 50 в каждом отдельном опыте) регистрировались на экране двухлучевого осциллоскопа в виде линейчатого раstra (рис. 3.3), где по горизонтальной оси — время относительно начала стимула (или относительно предшествующего ему синхронизирующего импульса, запускающего луч осциллоскопа), по вертикальной оси сверху вниз — порядковый номер ответа ( $j$ ), а интенсивность засветки характеризует огибающую реакции. В верхней части кадра регистрировалась

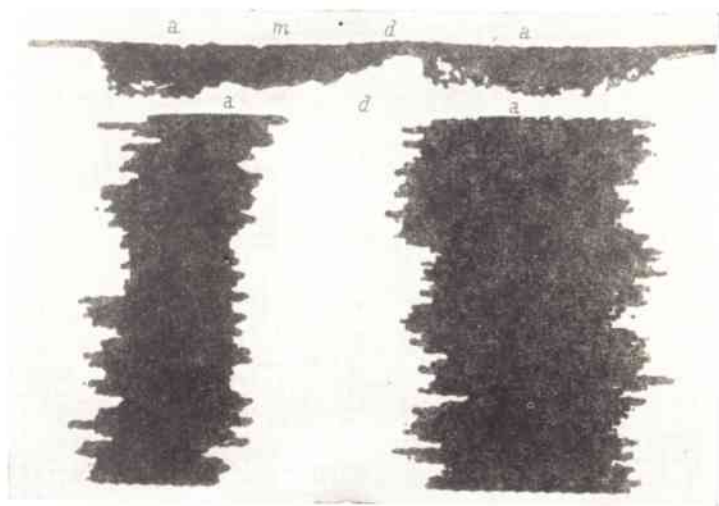


Рис. 3.3. Фотография изображения, полученного на электронном осциллоскопе в одном из опытов по синхронизации произнесения звуко сочетания [ada] (ответы испытуемого) с речевым стимулом [amda]. По [93].

(с помощью второго луча) осциллограмма предъявляемого стимула.

На рис. 3.4 приведены результаты измерений временной картины стимулов (значения  $t_2$ ,  $t_3$ ,  $t_4$ ) и реакций-ответов испытуемых (средние по 5 испытуемым значения  $t_1^*$ ,  $t_2^*$ ,  $t_3^*$ ,  $t_4^*$ ). Как видно из рисунка, временная картина произносимого звуко сочетания приблизительно воспроизводит временную картину стимула. Полученные результаты показывают, что в среднем момент окончания первого гласного в реакции ( $t_2^*$ ) достаточно хорошо соответствует моменту окончания первого гласного в стимуле ( $t_2$ ); это же можно сказать и о моменте начала второго гласного стимула и реакции ( $t_3$  и  $t_3^*$ ). Такое соответствие достигалось испытуемыми за счет управления длительностью согласного при речеобразовании: в тех случаях, когда в стимуле было сочетание из двух согласных, интервал согласного в реакции растягивался (см. на рис. 3.4

стимулы [atma], [adma], [asma] и т. д.; точечной кривой обозначена граница между согласными в стимуле).

Интересная особенность наблюдалась в тех случаях, когда вторым согласным в стимуле был сонант (стимулы [adma], [atma], [asma]). Было обнаружено, что длительность согласного в реак-

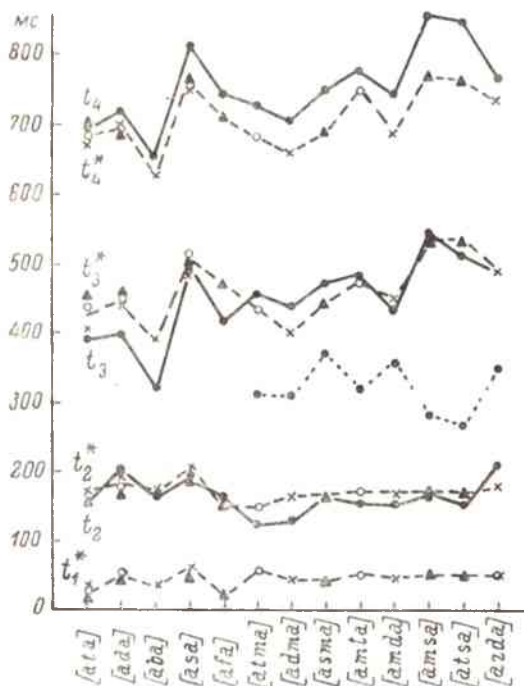


Рис. 3.4. Временные отношения между стимулами и реакциями. По [93].

По оси абсцисс — стимулы; по оси ординат — время (за нуль принято начало стимула). Сплошными линиями обозначены временные положения границ между гласными и согласными в стимулах; прерывистыми — в реакциях. Точечной кривой обозначено положение границы между двумя согласными в соответствующих стимулах. Разные реакции обозначены разными значками: [asa] — треугольниками, [ada] — крестиками, [ata] — кружками.

циях, создаваемых в ответ на эти стимулы, оказывается в среднем уменьшенной по сравнению с длительностью согласных в стимуле. Как выяснилось, это было связано с тем, что в половине случаев испытуемые принимали начало сонанта за начало второго гласного.

В экспериментах по синхронизации был обнаружен еще один чрезвычайно интересный эффект, заключающийся в том, что флюктуации во времени более поздних явлений (границ сегментов) в реакции были меньше, чем флюктуации более ранних явлений (рис. 3.5).

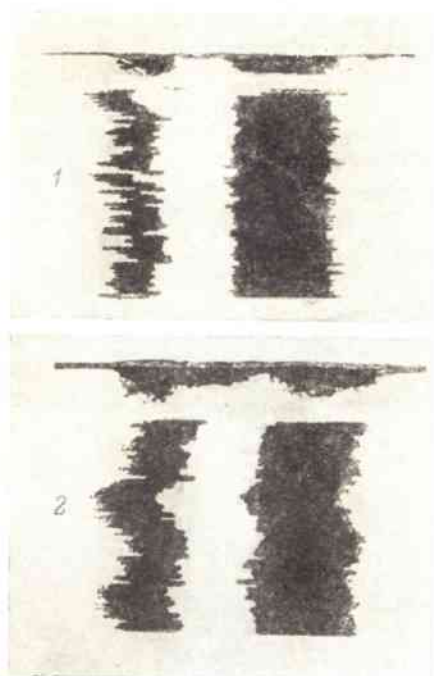


Рис. 3.5. Фотографии опытов по синхронизации произнесения звуко сочетания [ada] со стимулом [aba] (1) и со стимулом [amda] (2). По [161].

Рис. 3.6. Результаты классификации опытов по признаку зависимости между величиной флюктуаций сегментной границы и порядковым номером границы внутри звуко сочетания. По [161].  
Обозначения см. в тексте.

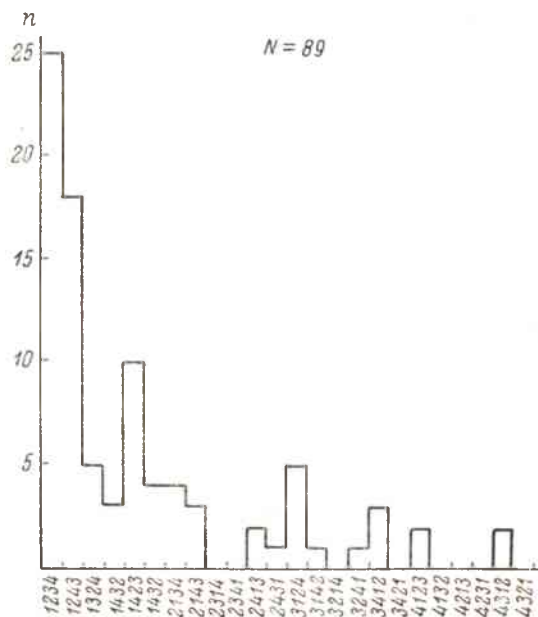


Рис. 3.6 позволяет судить о выраженности этого эффекта по всему полученному материалу. На рисунке представлены результаты классификации данных 89 опытов (89 фотографий) по размаху флюктуаций в опыте каждой из границ сегментов в реакции. Ряд чисел 1 2 3 4, например, означает, что самый большой размах флюктуаций соответствует началу первого гласного, затем в порядке убывания размаха следует конец первого гласного, начало второго гласного и конец второго гласного. Можно видеть, что именно этот случай (класс 1 2 3 4) наблюдался наиболее часто. Очевидно, что если при управлении речеобразованием в качестве точек отсчета длительностей последовательных сегментов используются только некоторые явления в самой реакции, то размах флюктуаций для каждой следующей границы должен быть таким же (или большим), как размах флюктуаций начала первого гласного.

Для объяснения полученного эффекта (убывание флюктуаций с возрастанием номера сегментной границы) приходится принять, что испытуемый выделяет во внешнем речевом стимуле какие-то явления, служащие индикатором  $i$ -той сегментной границы, и использует их в качестве пусковых сигналов при выработке  $(i+1)$ -ой команды.

Полученные данные позволяют считать, что требование синхронизации произносимого звуко сочетания с внешним речевым сигналом интерпретируется испытуемым как требование добиться того, чтобы гласные в реакции приходились по времени на гласные стимула, а согласный в реакции приходился по времени на согласный или на сочетание согласных стимула. Очевидно, что такая интерпретация возможна лишь в том случае, если испытуемый располагает понятием интервалов гласных и интервалов согласных, т. е. членит воспринимаемый сигнал на соответствующие отрезки и может управлять длительностью этих отрезков при речеобразовании.

#### 3.4.2. ГЛАСНЫЙ ЗВУК И ФОРМАНТНЫЕ ПЕРЕХОДЫ

В правилах, предлагаемых фонетистами относительно членения осциллограмм и спектрограмм речевого сигнала на отрезки гласных и согласных [13: 416], формантный переход относится к отрезку гласного. Если исходить из переносимой им фонемной информации, его нужно было бы отнести к согласному [346].

Попытка выяснить, входит ли участок перехода в воспринимаемый гласный звук, была предпринята в эксперименте Жукова и Чистович. В качестве стимулов были использованы два синтетических звуко сочетания ( $A$  и  $B$  на рис. 3.7), образованных из трех отрезков: стационарного гласного  $[i]$  ( $I$ ); паузы ( $II$ ); отрезка ( $III$ ), состоящего из переходного участка, на котором частота  $F_2$  меня-



лась от некоторого начального значения до значения, характерного для гласного [u], и стационарного участка.

Единственное физическое отличие стимулов *А* и *Б* (рис. 3.7) состояло в разной длительности переходного участка 3-го отрезка; она составляла 72 мс для стимула *А* и 143 мс для стимула *Б*. Стимул *А* воспринимался как звуко сочетание [idu], стимул *Б* — как звуко сочетание [id'u].

Испытуемые (5 человек) управляли длительностью интервала от начала 3-го отрезка до конца стимула — меняли момент окончания стимула. Задача испытуемых состояла в нахождении границы по ударению — такой длительности второго гласного, когда ударение в стимуле начинает перемещаться с одного гласного на другой.

Если человек определяет второй гласный как интервал от момента усиления звука (размыкание смычки) в стимуле до окончания стимула, т. е. относит фор-

мантный переход к гласному, тогда устанавливаемые испытуемым длительности должны быть одинаковыми для стимулов *А* и *Б*.

Если же второй гласный определяется человеком как интервал от момента пересечения второй формантой области, характерной для гласного [u], до окончания стимула, тогда длительность, устанавливаемая для стимула *Б*, должна превышать таковую для стимула *А* примерно на 71 мс (разность длительностей формантных переходов).

Рис. 3.7. Формантная структура стимулов *А* и *Б*.

Обозначения см. в тексте.

мантный переход к гласному, тогда устанавливаемые испытуемым длительности должны быть одинаковыми для стимулов *А* и *Б*.

Если же второй гласный определяется человеком как интервал от момента пересечения второй формантой области, характерной для гласного [u], до окончания стимула, тогда длительность, устанавливаемая для стимула *Б*, должна превышать таковую для стимула *А* примерно на 71 мс (разность длительностей формантных переходов).

Результаты экспериментов для разных испытуемых приведены ниже ( $M$  — среднее арифметическое из 10 установленных значений длительности интервала от начала 3-го отрезка до конца стимула в мс;  $s$  — стандартное отклонение в мс):

|               | Стимул А |       | Стимул Б |       | $M_B - M_A$ |
|---------------|----------|-------|----------|-------|-------------|
|               | $M_A$    | $s_A$ | $M_B$    | $s_B$ |             |
| Л. Ч. . . . . | 181      | 22.6  | 189      | 16.0  | 8           |
| С. Ж. . . . . | 175      | 15.6  | 178      | 19.1  | 3           |
| В. Т. . . . . | 207      | 18.2  | 211      | 29.4  | 4           |
| Н. С. . . . . | 237      | 13.8  | 243      | 19.7  | 6           |
| Т. М. . . . . | 147      | 17.1  | 141      | 14.9  | -6          |

Как видно, различия ( $M_E - M_A$ ) очень малы и для всех испытуемых не являются статистически достоверными.

Таким образом, полученные данные позволяют считать, что участок формантного перехода  $F_2$  в применявшихся стимулах действительно включается при восприятии в отрезок гласного и испытуемый принимает за начало гласного момент начала звука (конец паузы).

Нужно подчеркнуть, что сделанный вывод не следует распространять на все случаи формантных переходов, особенно на переходы первой форманты. Есть основание полагать, что связанное с повышением  $F_1$  увеличение энергии в средней области спектра или сам факт повышения  $F_1$  являются для человека сигналом начала отрезка гласного.

#### 3.4.3. ДЛИТЕЛЬНОСТЬ ГЛАСНОГО ЗВУКА КАК ФОНЕМАТИЧЕСКИЙ ПРИЗНАК

В большей части языков признаки, ответственные за различение гласных фонем, заключены в спектральной (формантной) картине звука, а длительности гласных звуков несут информацию прежде всего о просодических и эмоциональных характеристиках высказывания. Однако существуют языки, в которых длительности гласных имеют фонологическое значение; система гласных фонем в таких языках включает долгие и краткие гласные. При этом в ряде языков (например, финском, чешском, норвежском) члены каждой пары «долгий — краткий» имеют одинаковые спектральные характеристики, и длительность гласных звуков является единственным различительным признаком этих фонем. В ряде других языков (немецком, шведском, голландском) члены каждой такой пары различаются не только длительностью, но и своими спектральными характеристиками.

Если человек действительно использует длительность отрезка гласного для идентификации фонем, это может служить доказательством того, что сегментация (выделение отрезка гласного) логически предшествует фонемной классификации, т. е. признаком фонемного класса является не текущее значение спектра, но характеристика отрезка звукового потока.

Доказательства использования длительности в качестве фонемного признака были получены двумя экспериментальными методами: методом идентификации и методом установки «характерной» для долгих и кратких фонем длительности в синтетических стимулах.

На рис. 3.8 приведены данные [410], полученные в опытах по идентификации синтезированных голландских слов, которые представляли собой минимальные пары, различающиеся только длительностью гласного. Длительность гласного изменялась ступенями по 5 и по 10 мс; его спектральные характеристики сохранялись неизменными. Испытуемые должны были записать, какое

из голландских слов: [as] или [a:s], [at] или [a:t] (рис. 3.8, А), [at] или [a:t], [atəxə] или [a:təxə] (рис. 3.8, Б) — они слышат.

Убывание количества ответов «короткий» при увеличении длительности гласного является экспериментальным подтверждением

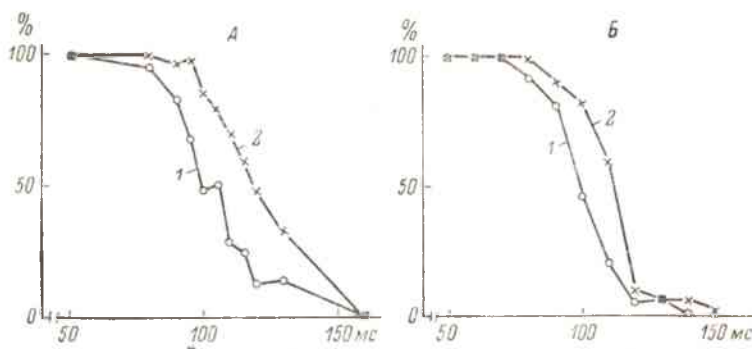


Рис. 3.8. Функция идентификации синтетического гласного по признаку долготы-короткости. По [410].

По оси абсцисс — длительность гласного; по оси ординат — относительное количество ответов «короткий». На А: 1 — для стимулов [at]; 2 — для стимулов [as]. На Б: 1 — для стимулов [atəxə]; 2 — для стимулов [a:təxə].

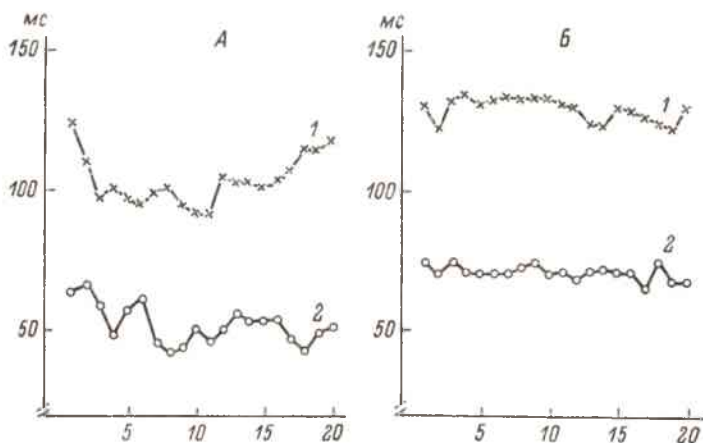


Рис. 3.9. Двадцать последовательных установок длительности ударного гласного в синтезированном слове двумя испытуемыми (А и Б). По [409].

По оси абсцисс — номер установки; по оси ординат — длительность гласного. 1 — установка длительности [a:] в слове [rəra:rəp]; 2 — установка длительности [a] в слове [rərarəp].

того, что длительность гласного может быть использована человеком в качестве признака при фонемной идентификации.

Аналогичные результаты были получены и на носителях японского языка [267]. Стимулы представляли собой 17 синтезированных

(осмысленных) слов, в которых длительность гласного [o] изменялась так, чтобы покрыть весь диапазон длительностей от слова [oi] до слова [o:i] (от 80 до 240 мс ступенями по 10 мс). В тестах по идентификации эти синтетические стимулы предъявлялись как изолированно, так и в контексте предложения.

В работе [409] приведены данные по установке «оптимальной» длительности долгого и короткого гласного в бессмысленных звуко-сочетаниях. Испытуемых (носителей голландского языка) просили установить длительность ударного гласного [a:] в слове [рэра:рэр] и [a] в слове [рэррэр], руководствуясь только тем критерием, чтобы все слово целиком звучало как можно естественнее.

Результаты приведены на рис. 3.9. Как видно из рисунка, устанавливаемые длительности для кратких гласных оказываются значительно меньше, чем для долгих (примерно в 2 раза), что согласуется с результатами, полученными при измерениях в естественной речи.

Известно, что в речи длительности гласных могут изменяться в достаточно широких пределах. Это является результатом действия целого ряда фонетических факторов, в том числе контекстуальных (характера окружающих согласных, позиции по отношению к ударению, количества слогов в слове, темпа речи и т. д.). Приведенные выше данные (рис. 3.8) указывают на то, что контекстуальные влияния в какой-то мере учитываются при принятии решения о фонематической длительности гласного. В естественной речи гласные имеют большую длительность перед щелевыми, чем перед смычными согласными. Как видно из рис. 3.8, А, фонемная граница между [a:] и [a] сдвигается влево, если следующее за этим гласным [s] (кривая 2) заменяется на [t] (кривая 1).

Из рис. 3.8, Б видно, что увеличение числа слогов в слове сдвигает фонемную границу между [a:] и [a] в направлении более коротких длительностей, что также согласуется с фактами из речеворазования.

Вопрос же о том, какова процедура учета этих контекстуальных влияний при восприятии длительности гласного, пока остается нерешенным.

#### 3.4.4. ДЛИТЕЛЬНОСТЬ ГЛАСНОГО КАК ПРИЗНАК ДЛЯ ИДЕНТИФИКАЦИИ ГЛУХОСТИ-ЗВОНКОСТИ ПОСЛЕДУЮЩЕГО СОГЛАСНОГО

Изменения длительности гласного, связанные с характером последующего согласного — его звонкостью или глухостью, являются известным фонетическим фактом.

Наиболее ярко этот эффект выражен в английском языке, в котором длительность гласного, предшествующего глухому согласному, составляет приблизительно от 1/2 до 2/3 длительности того же гласного в идентичных условиях, но перед звонким [439].

Такие большие различия между длительностями гласных, предшествующих глухим и звонким согласным, привлекают внимание исследователей речи по той причине, что в ряде случаев

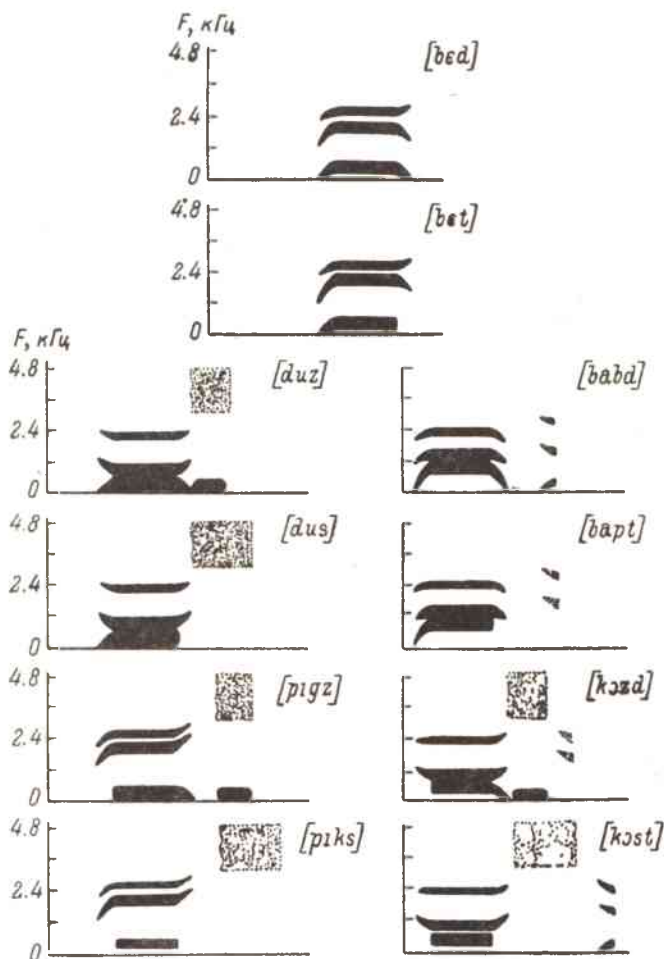


Рис. 3.10. Схематическое изображение формантной структуры стимулов, использованных в эксперименте. По [440].

Обозначения см. в тексте.

длительность предшествующего гласного является единственным признаком для идентификации глухости или звонкости последующего согласного. Так, например, в парах слов типа английских *bit*—*bid*, которые теоретически различаются только глухостью-звонкостью последнего согласного, самый надежный признак звонкости этого согласного — систематически большая длитель-

ность предшествующего гласного. Акустические признаки собственно согласного, несущие информацию о звонкости (наличие

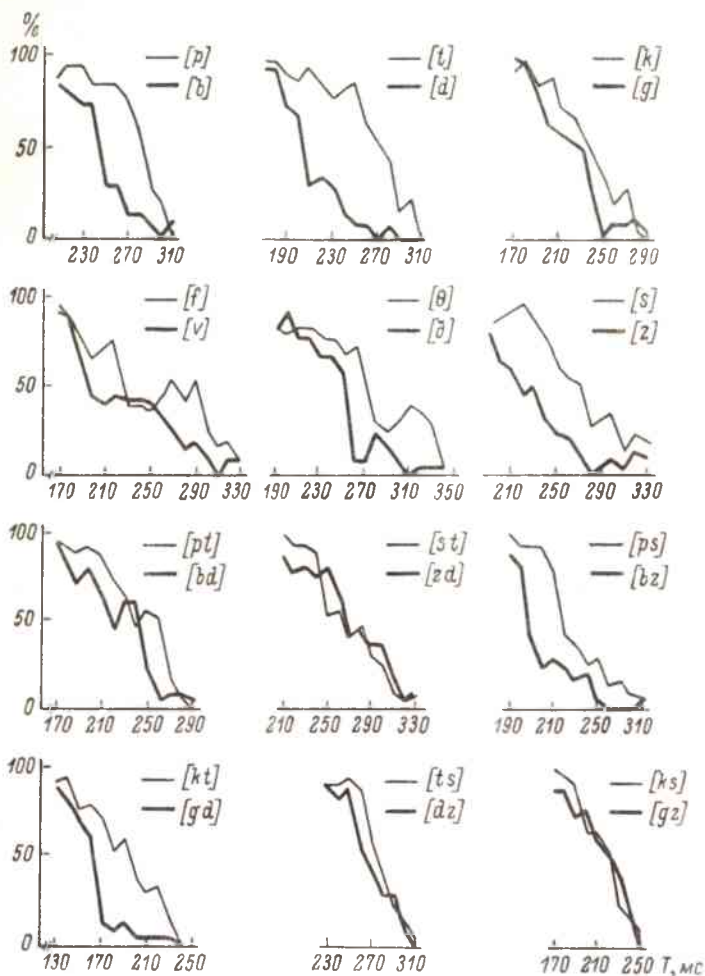


Рис. 3.11. Зависимость восприятия признака глухости-звонкости конечного согласного или сочетания согласных от длительности предшествующего гласного (для соответствующих стимулов, синтезированных с «глухими» и «звонкими» конечными согласными). По [440].

По оси абсцисс — длительность гласного; по оси ординат — относительное количество ответов «глухой».

голоса во время смычки и взрыва), в конечном положении могут быть выражены очень слабо или даже полностью отсутствовать.

Возникает естественный вопрос, используется ли информация о длительности гласного для принятия решения о глухости-звон-



кости соседнего согласного при восприятии речи. Положительный ответ, с одной стороны, означал бы, что человек способен выделять отрезок речевого потока, соответствующий гласному звуку, и, с другой стороны, что при принятии фонемного решения должно происходить объединение фонемной информации, заключенной в соседних сегментах (в ГС слоге).

На рис. 3.10 приведено схематическое изображение синтетических стимулов, использованных в работе [440]. Конечная часть стимулов синтезировалась таким образом, чтобы воспроизводить типичные акустические явления в естественных сочетаниях, оканчивающихся на глухой или на звонкий согласный (сочетание согласных). Переменным параметром при синтезе была длительность гласного, изменяющаяся в диапазоне от 150 до 350 мс. Записанные на пленку стимулы предъявлялись испытуемым, которые должны были ответить, звонким или глухим был конечный согласный.

Полученные результаты приведены на рис. 3.11, где сравниваются функции идентификации конечного согласного (сочетания согласных) для соответствующих стимулов, синтезированных с «глухими» и «звонкими» конечными согласными. Как видно из этого рисунка, влияние признаков конца стимула оказалось значительно более слабым, чем влияние длительности предшествующего гласного.

Таким образом, оказывается, что длительность предшествующего гласного при восприятии может служить для человека признаком глухости-звонкости последующего конечного согласного, причем имеет значительный «вес».

### 3.5. РАЗНЫЕ ИНТЕРПРЕТАЦИИ ПОНЯТИЯ «ФОНЕТИЧЕСКИЙ ЭЛЕМЕНТ»

Приведенные в главе данные заставляют считать, что человек действительно каким-то образом выделяет в речевом сигнале отрезки гласных звуков. Можно с достаточной вероятностью полагать, что гласная фонема является характеристикой выделенного отрезка, т. е. распознавание гласных фонем основывается на измерении ряда признаков именно этого отрезка. Следовательно, для гласных выдерживаются требования, предъявляемые ранее к фонетическому элементу: элемент есть отрезок речевого потока, фонема является обозначением класса, в который по ряду своих физических свойств попадает данный отрезок. Характеристиками элемента, кроме фонемного символа, могут быть длительность, интенсивность (громкость) и, вероятно, какие-то признаки основной частоты голоса.

В отношении согласных дело обстоит сложнее. Удобно предположить, что для записи согласного (группы согласных) используется отдельная ячейка памяти, и в этом смысле согласные как бы «приписываются» отрезку сигнала между концом предыдущего

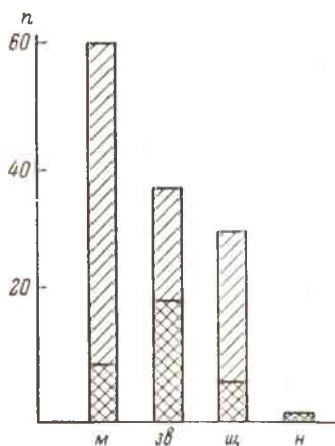
и началом следующего гласного. Однако акустические признаки, на основе которых производится идентификация согласных, находятся не только на этом отрезке, но и на соседних отрезках гласных.

Если определить фонетический элемент как отрезок сигнала, характеристикой которого является фонема, то для согласной фонемы фонетическим элементом оказывается слог. При этом выясняется еще одна трудность: полезная информация о согласном в сочетании ГСГ содержится на отрезках обоих гласных.

Тот факт, что человек «собирает» информацию о согласном, заключенную на переходе от гласного к смычке (щели) согласного и на переходе от согласного к гласному, отчетливо был продемонстрирован в экспериментах по текущей имитации звукосочетаний типа ГСГ [159]. Оказалось, что, начав имитировать согласный

Рис. 3.12. Количество ошибок в согласных при имитации, заключающихся в неправильном воспроизведении места образования согласного (ж), звонкости или глухости (зв), щелевого или смычного характера согласного (щ), назальности или неназальности (н). По [159].

Остальные обозначения см. в тексте.



раньше, чем диктор успевал полностью произнести его, испытуемый в ряде случаев начинал имитацию с неправильных движений, которые могли быть им исправлены после того как он слышал конец согласного. Этот факт отражен в диаграмме на рис. 3.12, где всей высотой столбика показано количество ошибок (по разным признакам), наблюдавшихся в начальный момент имитации согласных; двойной штриховкой показано количество ошибок, сохранившихся до конца произнесения согласного. Из диаграммы следует, что в большинстве случаев по мере накопления информации происходит исправление артикуляции.

Следовательно, допустив, что фонетические элементы (слоги) представляют собой соседние неперекрывающиеся отрезки речевого потока, мы неизбежно потеряем часть информации о согласной фонеме. Это произойдет вне зависимости от того, примем ли мы открытые или закрытые слоги, т. е. будем ли мы считать границей слога начало или конец отрезка гласного.

Казалось бы заманчивым предположить, что блок фонетической интерпретации использует в качестве единиц анализа речевого потока как интервалы между началами соседних отрезков

гласных, так и интервалы между их концами, т. е. последовательно рассматривает то закрытый, то открытый слог. Такой метод перекрывающихся единиц анализа позволил бы обеспечить полное извлечение информации. Экспериментальные данные в пользу того, что такая процедура применяется человеком при решении вопроса о том, является ли рассматриваемый отрезок слогом, приводятся в главе 6.

Итак, вопрос о природе фонетического элемента оказался достаточно сложным. Вероятно, имеет смысл отдельно рассматривать, с одной стороны, систему промежуточных понятий относительно речевого потока, используемых блоком фонетической интерпретации при преобразовании речевого сигнала в последовательность выходных данных, и, с другой стороны, принципы организации (объединения в группы) самих выходных данных.

В отношении речевого потока, очевидно, нужно допустить такие понятия, как отрезок гласного и интервалы между началами и между концами соседних отрезков гласных. Если эти понятия введены, то понятие фонетического элемента может стать излишним.

В отношении выходной информации приходится предполагать, что должны быть выделены группы сведений об отрезках гласных. Каждая такая группа должна включать, кроме указания гласной фонемы, ряд просодических характеристик. Вопрос о том, можно ли все просодические характеристики включить в такую группу и что, кроме последовательности фонем, должна включать группа сведений относительно согласных, пока остается открытым. Обсуждение этих вопросов содержится в главе 5. Можно видеть, что и для описания организации выходной информации нужны более дифференцированные понятия, чем понятие фонетического элемента. Из этого, естественно, не следует, что термином «фонетический элемент» нельзя пользоваться в описательных целях.

Заметим, что необходимость уточнения понятий никак не влияет на существо исходных лингвистических представлений о том, что речевой сигнал описывается в «сознании» слушающего последовательностью дискретных элементов.

Главный для настоящей главы вывод о выделении в речевом потоке отрезков гласных влечет за собой по крайней мере два важных следствия. Во-первых, появляется возможность ожидать, что полезными признаками фонем служат признаки отрезков; это расширяет область поисков этих признаков при исследовании восприятия речи и при разработке устройств для автоматического распознавания речи. Во-вторых, появляется самостоятельная и достаточно сложная задача изучения способов обработки сигнала, обеспечивающих выделение отрезков гласных.

## Глава 4

# ЛИНГВИСТИЧЕСКИЕ И ПСИХОЛОГИЧЕСКИЕ ФОНЕМЫ

### 4.1. ЛИНГВИСТИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ О ФОНЕМАХ

Речевое сообщение рассматривается в лингвистике как линейная последовательность фонем. Под фонемой понимается кратчайшая смыслоразличительная (фонологическая) единица. Безусловное требование, выдвигаемое при определении фонемного состава языка, состоит в том, что цепочки фонем, соответствующие двум различающимся по смыслу речевым сообщениям, должны быть разными (оставим пока в стороне сравнительно редкие случаи омонимии сигналов.) Фонематически значимыми при этом признаются те и только те различия в звуковом содержании сообщений, которые используются в данном языке для различения смысла. Такой подход обеспечивает наиболее экономное описание единиц фонологического уровня языковой структуры. Однако при решении вопроса о соотносительности фонетических и фонологических единиц разные лингвистические школы пользуются разными критериями.

Сторонники московской фонологической школы, например, ставят требование, чтобы корневые морфемы разных словоформ одного и того же слова совпадали, т. е. были выражены одной и той же цепочкой фонем [118]. Поэтому первые гласные в словах *водá* [vɔdá], *водý* [vódi], *водонóс* [vɔdnɔs] рассматриваются в качестве вариантов одной фонемы — [o] (фонема всегда определяется по сильной позиции, для гласных таковой является позиция под ударением). Таким образом, согласно концепции московской фонологической школы, фонема является абстрактной фонологической единицей, которая не может быть отождествлена ни с какой конкретной звуковой единицей. Хотя использование этой концепции облегчает задачу формального описания языка, ясно, что для решения о том, что первым гласным в слове [vɔdá] является фонема [o], распознающее устройство (человек или машина) должно сначала понять смысл слова.

Основной принцип установления фонемной принадлежности сообщения, используемый представителями ленинградской фонологической школы, заключается в следующем: если две словоформы одного и того же слова отличаются друг от друга по звучанию, то кратчайшие фонологические единицы, несущие это различие, должны быть признаны разными фонемами [72, 170]. Поэтому в словах *водá* [vɔdá] и *вóды* [vódi] мы имеем разные корневые гласные фонемы. В одну фонему, следовательно, сводятся сходные в физиолого-акустическом отношении звуки.

Таким образом, ленинградские фонетисты рассматривают фонему прежде всего как определенный звукотип. Такой подход обеспечивает единство классификации фонем по их физиолого-акустическим свойствам. Сложность, однако, состоит в том, что количество звукотипов, существующих в языке, по-видимому, значительно превышает число фонем, выделяемых на основе принципа смыслоразличения. Вопрос о том, сколько разных звукотипов используют носители русского языка, становится для ленинградской фонологической школы предметом исследования.

Из сказанного следует, что как фонемный состав всего языка, так и фонемный состав отдельного слова в лингвистической литературе определяется по-разному, в зависимости от тех допущений и требований, которые предъявляются к описанию языка.

Выдающееся значение для фонологии имела теория дифференциальных признаков, разработанная Трубецким [132]. В огромном количестве смыслоразличительных противопоставлений, на основе которых определяется фонемный состав языка, он сумел увидеть ограниченное количество разных типов противопоставлений по какому-либо определенному признаку. Число таких признаков в каждом языке ограничено и сравнительно невелико. Понятие фонемы является, по Трубецкому, производным от понятия признак: она определяется как совокупность фонологически существенных, т. е. дифференциальных, признаков. Например, дифференциальными признаками фонемы [g] в русском языке будут:

1) место образования — заднеязычность, поскольку по этому признаку [g] противопоставляется губной и переднеязычной фонемам [b], [d];

2) звонкость — по этому признаку [g] противопоставляется фонеме [k];

3) твердость<sup>1</sup> — по этому признаку [g] противопоставляется фонеме [g'].

Способ образования, взрывность, не является дифференциальным признаком, так как в русском языке противопоставление [g] взрывное—[g] щелевое фонологически не существенно, т. е. не влечет за собой противопоставления по смыслу. Это же относится и к признаку назальности в паре [ŋ]—[g].

---

<sup>1</sup> Сторонники московской фонологической школы не считают [g'] самостоятельной фонемой в русском языке.



## 4.2. ПСИХОЛОГИЧЕСКИЕ ПРИЗНАКИ ФОНЕМ

Концепция дифференциальных признаков Трубецкого, получившая дальнейшее развитие в ряде работ [215, 305], оказалась весьма плодотворной не только для теоретической лингвистики, но и для теории восприятия речи.

Согласно этой концепции, множество воспринимаемых фонем не является совокупностью отдельных символов, но представляет собой строгую упорядоченную систему, в которой каждая фонема задается перечислением значений по признакам, число которых существенно меньше количества возможных фонем.

Идея о наличии в сознании носителей языка системы признаков, по которым воспринимаются и запоминаются фонемы, подтверждается, в частности, существованием у человека представления о субъективном сходстве фонем. Информация о таком сходстве, получаемая в разных экспериментальных условиях (шкалирование, забывание, восприятие в условиях помех), как правило, хорошо объясняется с позиции дифференциальных признаков: чем меньшее число таких признаков разделяет фонемы, тем более субъективно близкими, похожими они являются.

Какова природа этих признаков? Предполагается, что в системе, распознающей речь, имеется фонемный классификатор — устройство, выдающее фонемное решение в ответ на входной акустический сигнал. Очевидно, что дифференциальные признаки могут быть отражением либо акустических признаков на входе фонемного классификатора, либо каких-то абстрактных психологических признаков на его выходе. Попытка поиска прямых соответствий между дифференциальными признаками и непосредственно физическими свойствами речевого сигнала [305] не дала положительных результатов. Это заставляет думать, что верным является второе предположение.

Трудность, однако, состоит в том, что до недавнего времени было неизвестно, с какого рода и каким количеством признаков имеет дело человек при восприятии: в упоминавшихся исследованиях системы таких признаков вводились априорно и заимствовались, как правило, из литературы по теоретической фонологии. Лишь в последнее время появились работы, в которых рассматриваются перцептивные признаки фонем, полученные апостериорным путем, в результате анализа реакций испытуемых в экспериментах по восприятию. Возможность появления такого рода работ связана с разработкой специальных вычислительных процедур, позволяющих находить минимальную достаточную размерность пространственной конфигурации речевых стимулов на основе информации о субъективной близости этих стимулов, полученной экспериментально [209, 466]. Количество выявленных ортогональных осей этой пространственной конфигурации соответствует числу независимых перцептивных признаков, по которым данное множество входных фонем упорядочено в сознании испы-



туемых. Интерпретация этих найденных в результате вычислительного анализа осей-признаков зависит от экспериментатора.

В работе [468] изложены результаты применения подобной вычислительной процедуры к данным большого количества экспериментов по нахождению субъективной близости между согласными. Эти эксперименты проводились в широком диапазоне условий, включающих разные языки, разные наборы исходных стимулов, разные методы их предъявления (акустическое, визуальное, тактильное), разных испытуемых (нормально- и слабослышащих) и т. д. Оказалось, что во всех без исключения экспериментах выявляются одни и те же перцептивные признаки согласных, обозначаемые автором статьи [468] как назальность, озвонченность, сибилантность, непрерывность, место образования и sonorность.

Тот факт, что как при слуховом, так и неслуховом предъявлении согласных стимулов были обнаружены одни и те же признаки, подтверждает уже высказанное предположение о том, что эти признаки не являются акустическими и что по ним описываются не сигналы на входе фонемного классификатора, но решения, принятые классификатором.

#### 4.3. ФОНЕМА КАК ЗАПОМИНАЕМОЕ СОКРАЩЕННОЕ ОПИСАНИЕ СИГНАЛА

Переход от описания сигнала в терминах акустических признаков к фонемному описанию (описанию в терминах дифференциальных признаков) не только обеспечивает возможность дальнейшего смыслового (лексического и синтаксического) анализа сообщения, он также резко сокращает количество информации, необходимой для записи сигнала в памяти. Гипотеза о том, что при запоминании речевого сигнала человек использует не его подробное акустическое изображение, но принятое о сигнале фонемное решение, проверялась в многочисленных экспериментах.

##### 4.3.1. СОГЛАСНЫЕ

Все данные, касающиеся согласных, были получены методом сравнения функций идентификации с функциями различения. В экспериментах такого рода применяются синтезированные стимулы, образующие ряд по какому-то акустическому признаку стимула.

Рассмотрим в качестве примера опыт со стимулами, различающимися по так называемому времени начала голоса. В различных языках выделяются три группы смычных согласных: полностью звонкий согласный, глухой неаспирированный согласный и глухой аспирированный согласный. В случае полностью звонкого согласного (например, русские [b], [d], [g]) голосовая щель смы-

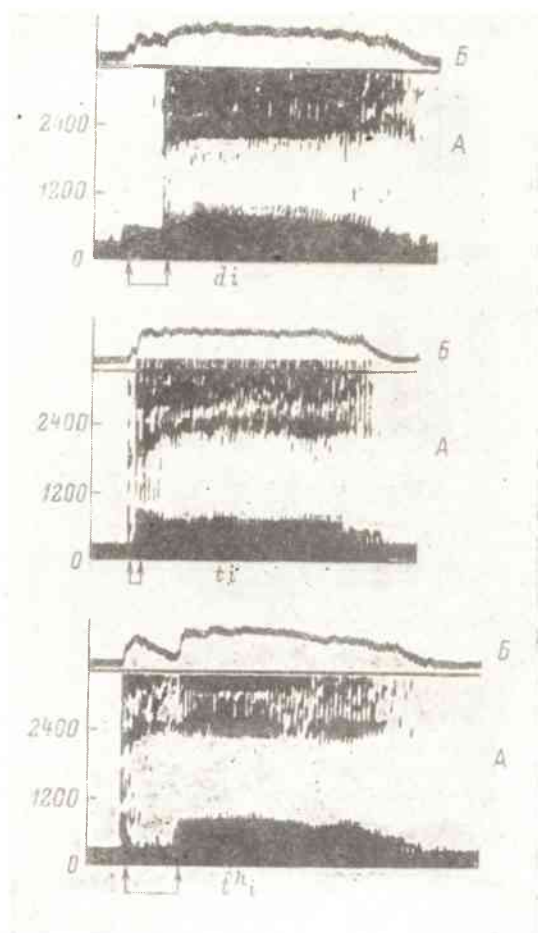


Рис. 4.1. Определение параметра «время начала голоса». По [176].

На А: — динамические спектрограммы слогов  $[di]$ ,  $[ti]$  и  $[t^hi]$ ; по оси абсцисс — время, по оси ординат — частота. На Б — энергетические огибающие тех же слогов. По оси абсцисс — время, по оси ординат — интенсивность. Стрелками отмечен интервал времени между началом колебаний голосовых связок и размыканием смычки в  $[di]$  и интервал времени между появлением шума взрыва и аспирации и началом колебаний голосовых связок в  $[ti]$  и  $[t^hi]$ .

кается значительно раньше размыкания смычки, в случае глухого аспирированного согласного (например, английские  $[p^h]$ ,  $[t^h]$ ,  $[k^h]$ ) голосовая щель остается открытой некоторое время после размыкания смычки.

В качестве акустического коррелята этого признака было предложено [367, 368] принять параметр, названный «время начала голоса». Рис. 4.1, где приведены спектрограммы слога с полностью

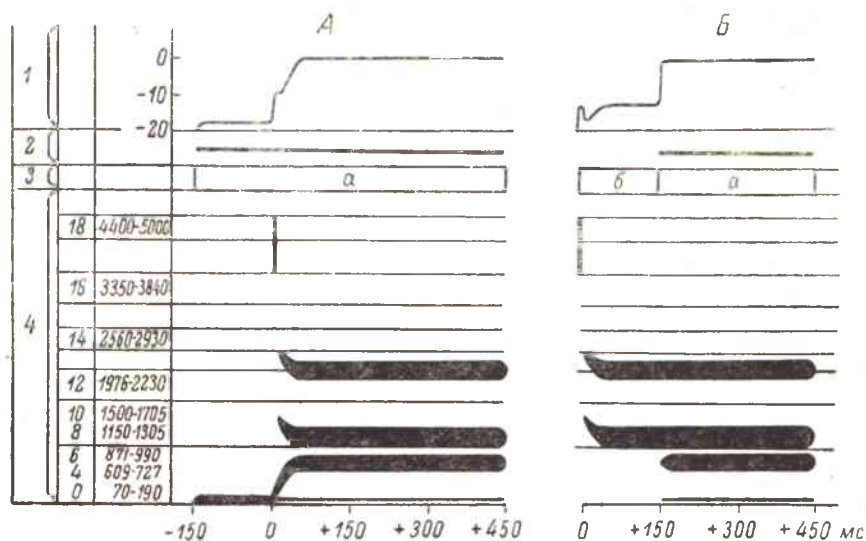


Рис. 4.2. Схематическое изображение синтетических слогов, начинающихся со звонкого согласного (А) и глухого аспирированного согласного (Б). По [176].

По оси абсцисс — «время начала голоса», начало отсчета соответствует моменту размыкания смычки согласного; по вертикали — характеристики каналов управления синтезатором: 1 — уровень интенсивности сигнала (в дБ); 2 — частота основного тона (сохраняется постоянной и равной 114 Гц); 3 — характер возбуждения (а — голосовой, б — шумовой); 4 — спектр сигнала, заданного в 18 частотных полосах (номера и границы полос в Гц указаны в соответствующих столбцах) так, что энергия в полосе имеет максимальное значение в зачерненных областях.

звонким согласным  $[di]$ , глухим неаспирированным согласным  $[ti]$  и глухим аспирированным согласным  $[t^h i]$ , иллюстрирует способ измерения этого параметра на динамических спектрограммах. За точку отсчета принимается момент размыкания смычки согласного, характеризующийся появлением энергии в высокочастотной части спектра (шум взрыва, появление энергии в области второй и более высоких формант). Начало голоса, если оно предшествует размыканию смычки, определяется по появлению энергии в области основной частоты голоса. В случае начала голоса после размыкания смычки имеют место не только появление энергии на основной частоте и проявление периодического характера

всего сигнала, но и резкое увеличение амплитуды первой форманты и общей энергии сигнала.

Слоги, приведенные на рис. 4.1, характеризуются по параметру «время начала голоса» значениями —85, 15 и 110 мс. Для исследования восприятия был синтезирован набор стимулов, изменяющихся в широком диапазоне значений этого параметра. Характеристики двух крайних стимулов показаны на рис. 4.2. В данном случае стимулы соответствуют слогу с гласным [a] и переднеязычным взрывным согласным [d] и [t<sup>h</sup>]. Аналогичным образом были синтезированы слоги с губным согласным.

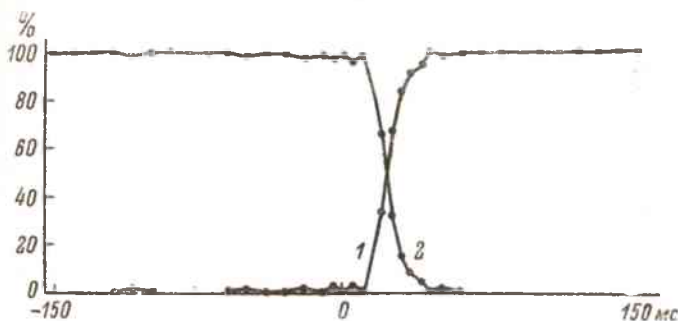


Рис. 4.3. Функции идентификации синтетических слогов, упорядоченных по параметру «время начала голоса». По [369].

По оси абсцисс — «время начала голоса» (начало отсчета — то же, что и на рис. 4.2); по оси ординат — относительное число ответов [b] (кривая 2) и [p<sup>h</sup>] (кривая 1).

Для получения функции идентификации стимулы, следующие друг за другом в случайном порядке, записывались на магнитофонную ленту и предъявлялись испытуемым. Испытуемые должны были записывать в ответ на каждый стимул букву, обозначающую ту фонему, на которую этот стимул больше всего похож. Набор букв был ограничен экспериментатором (метод форсированного выбора).

На рис. 4.3 приведены результаты идентификации губных согласных группой слушателей-американцев [369]. Можно видеть, что использованный ряд стимулов разделился на две области, одна из которых связана с фонемой [b], вторая — с [p<sup>h</sup>]. Значение параметра «время начала голоса», при котором ответы [b] и [p<sup>h</sup>] оказались равновероятными, называется фонемной границей между [b] и [p<sup>h</sup>] по указанному признаку.

После получения функции идентификации можно проверить предположение, что при запоминании и сравнении стимулов испытуемый действительно пользуется фонемами как субъективными описаниями этих стимулов.

Для этого используется метод АВХ-сравнения. Испытуемым предъявляются триады стимулов типа  $x_i$ ,  $x_{i+k}$ ,  $x_i$  и  $x_i$ ,  $x_{i+k}$ ,  $x_{i+k}$

( $i$  — порядковый номер стимула в ряду по значениям акустического параметра,  $k=1, 2, 3, 4$ ). Задачей испытуемых является определение того, с каким звуком триады — первым или вторым — совпадает последний звук. Построив такие триады для всех значений  $i$ , можно получить функцию различения — вероятность правильного ответа испытуемого в зависимости от положения стимулов по шкале акустического параметра. Зная функцию идентификации, легко вычислить теоретические значения функции различения при предположении, что вся информация о стимуле, которую испытуемый может запомнить, сводится к знанию выбранного в ответ на стимул фонемного символа. Не останавли-

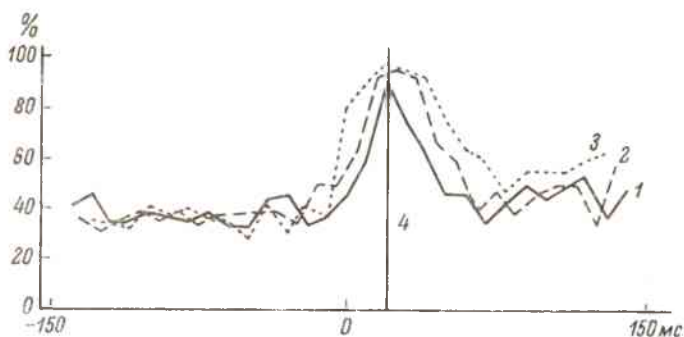


Рис. 4.4. Функции различения синтетических слогов, упорядоченных по параметру «время начала голоса». По [177].

По оси абсцисс — «время начала голоса»; по оси ординат — относительное число случаев, когда данный стимул отличался от остальных. Кривые 1—3 соответствуют стимулам, различающимся на две, три и четыре градации параметра «время начала голоса». Прямая 4 соответствует положению фонемной границы между [b] и [p<sup>h</sup>] на рис. 4.3.

ливаясь на процедуре вычислений, укажем только, что в этом случае значения функции различения должны быть близки к вероятности случайного отгадывания на интервалах, лежащих внутри фонемных категорий, и должны достигать максимальных значений вблизи фонемных границ.

На рис. 4.4 приведены функции различения описанных стимулов, полученные на группе слушателей-американцев [369]. Сравнение рис. 4.3 и 4.4 показывает, что максимум на функции различения совпадает с положением границы между [b] и [p<sup>h</sup>] на функции идентификации. Различение же стимулов внутри фонемных категорий оказывается очень плохим, близким к уровню случайного отгадывания.

Иная картина была получена для носителей языка таи. В этом языке имеются три губных согласных, различающихся по параметру «время начала голоса»: звонкий [b], глухой неаспирированный [p] и глухой аспирированный [p<sup>h</sup>]. Эксперименты, описанные в [177, 178, 369], показали, что на функции различения, получен-

ной на испытуемых-тайландцах, имеются два максимума, соответствующие границам между [b], [p] и [p], [p<sup>h</sup>], определенным по функции идентификации.

Данные, аналогичные только что описанным, были получены также в большом количестве экспериментов с синтезированными взрывными согласными, различающимися по направлению переходов  $F_2$  и  $F_3$  [242. 353. 378. 492]. При идентификации таких стимулов обычно удается получить три фонемные категории, соответствующие согласным с разным местом образования: [b], [d] и [g] или [p], [t] и [k]. Функции различения, полученные на таких стимулах, всегда имели максимумы в области фонемных границ.

#### 4.3.2. ГЛАСНЫЕ

Значительно более сложным оказался вопрос о субъективных речевых образах гласных. Сравнение функций различения с функциями идентификации не позволило сделать определенных выводов, так как было показано, что человек способен различать большое число градаций стимулов (синтезированных гласных), относимых им к одной и той же фонемной категории [265. 480. 489]. Этот результат можно было интерпретировать двояким образом. Одна возможная интерпретация состояла в том, что множество фонетических образов гласных образует континуум, и в качестве субъективных признаков гласных используются координаты гласных в фонетическом пространстве (например, конфигурация речевого тракта). Другая интерпретация заключалась в том, что, кроме фонемного решения о гласном, запоминается еще какая-то дополнительная информация (например, данные о спектре гласного).

Доказательство того, что, запоминая гласный, человек использует фонему в качестве описания этого гласного, было получено в экспериментах по прямому численному шкалированию субъективного расстояния между гласными-стимулами, приведенными на рис. 4.5,А [47]. Фонемная идентификация этих стимулов группой русских слушателей показала, что стимулы 1—3 воспринимаются как [i], стимулы 4—9 — как [e], стимулы 10—12 — как [a].

На рис. 4.5,Б приведена зависимость субъективного расстояния от расстояния между стимулами на двухформантной плоскости (положение стимулов в двухформантной плоскости показано на рис. 4.5, А). Можно видеть, что зависимость имеет отчетливый ступенчатый характер: субъективное расстояние резко возрастает при переходе от одной фонемной категории к другой, оно мало меняется в пределах фонемной категории.

Это позволяет утверждать, что испытуемые пользуются фонемами в качестве описания стимулов и что расстояния, которые они указывают, это расстояния между фонемами. Однако такой



ступенчатый характер зависимости ярко выражен лишь в случае, если стимулы в оцениваемой паре принадлежат к разным фонемным категориям.

Другая картина наблюдается в том случае, когда оба стимула в паре относятся к одной и той же фонемной категории. На рис. 4.6 приведены зависимости, полученные для пар [e]—[e]. Можно видеть, что субъективное расстояние монотонно возрастает с увели-

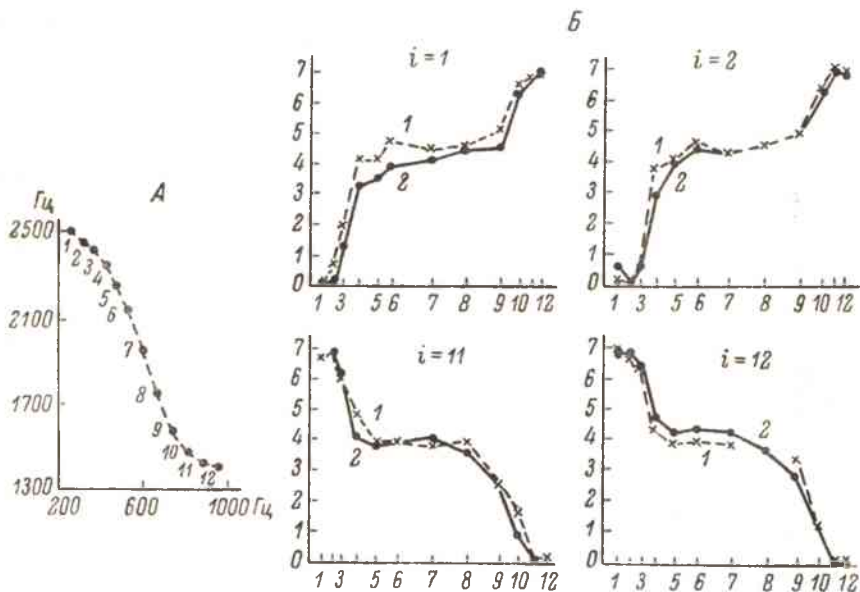


Рис. 4.5. Результаты экспериментов по психологическому шкалированию. По [47].

На А: положение в плоскости первой и второй формант ( $F_1, F_2$ ) синтетических гласных, использованных в экспериментах; по оси абсцисс — частота  $F_1$ , по оси ординат — частота  $F_2$ ; цифрами около точек обозначены номера стимулов. На Б: субъективные расстояния между парами синтетических гласных  $x_i, x_j$ ; по оси абсцисс — номер стимула  $x_j$ , по оси ординат — субъективное расстояние в баллах между  $x_j$  и стимулом, номер которого ( $i$ ) указан на каждом из четырех графиков; кривые 1 и 2 соответствуют разному порядку следования стимулов в парах ( $x_i, x_j$  или  $x_j, x_i$ ).

чением физической разницы между стимулами. Этого не могло бы быть, если бы информация о гласном, полученная при его восприятии, ограничивалась фонемным символом. Необходимо допустить, что человек способен на какое-то время запоминать не только фонему, выбранную на основании услышанного стимула, но и более подробное описание сигнала; назовем это условно «тембральной информацией».

В работе [266] исследовались идентификация и различение синтетических [i] и [e] гласных в условиях, когда после каждого стимула следовал дополнительный мешающий звук. Полученные

результаты оказались весьма близкими к тем, которые ранее наблюдались для согласных. Из рис. 4.7 видно, что на функции различения имеется отчетливый пик, соответствующий границе между

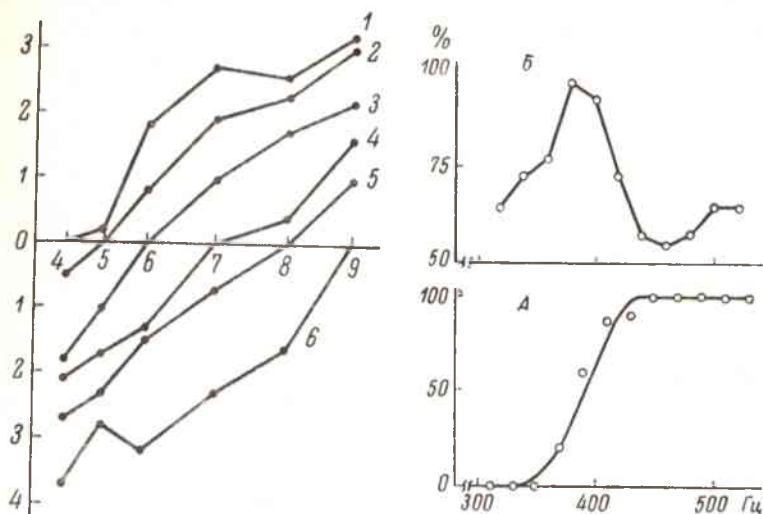


Рис. 4.6. Субъективные расстояния между парами гласных  $x_i x_j$ , принадлежащими к одной и той же фонемной категории (стимулы 4—9 на рис. 4.5, А). По [158].

По оси абсцисс — номера стимулов  $x_j$  в соответствии с рис. 4.5, А; по оси ординат — субъективное расстояние в баллах. Параметром кривых 1—6 является номер стимула  $x_i$  (соответственно  $i=4, 5, 6, 7, 8, 9$ ). Вверх от горизонтальной прямой отложено субъективное расстояние для пар стимулов с  $j > i$ , вниз — для пар с  $i > j$ .

Рис. 4.7. Сравнение функций идентификации (А) и различения (Б) синтетических гласных. По [266].

По оси абсцисс — частота  $F_1$  синтетических гласных; по оси ординат — относительное число ответов [е] (А) и относительное число случаев, в которых данный стимул отличался от остальных (Б).

фонемами [i] и [e]. Различение стимулов, лежащих внутри фонемной категории, оказывается плохим.

Итак, полученные данные позволяют сделать вывод, что множество субъективных образов гласных также является дискретным.

#### 4.4. НАБОР ПСИХОЛОГИЧЕСКИХ ФОНЕМ

Сокращение информации, обеспечиваемое переходом к фонемному описанию, зависит, естественно, от числа идентифицируемых при восприятии фонем.

В работах [16, 32] была предпринята попытка выяснить, сколько классов гласных идентифицирует в русской речи носитель русского языка. Испытуемым предъявлялись изолированные рус-

ские гласные, вырезанные из записей естественной речи. От испытуемых требовалось обозначать услышанные гласные русскими или латинскими буквами с добавлением в случае необходимости диакритических знаков. Ответы испытуемых показали, что число идентифицируемых ими классов гласных звуков гораздо больше числа лингвистических русских гласных: совершенно надежно можно утверждать, что гласные между твердыми и гласные между мягкими согласными относятся испытуемыми к разным звуковым типам.

Сходные результаты были получены в работе [489]. Оказалось, что испытуемые-американцы способны идентифицировать огубленные гласные переднего ряда, хотя они и отсутствуют в английском языке как самостоятельные лингвистические фонемы.

Выделяемые при восприятии классы звуков были названы психологическими фонемами. Естественно думать, что психологические фонемы соответствуют неперекрывающимся областям в пространстве акустических признаков сигнала. Легко заметить, что фонемы, определенные согласно критериям московской фонологической школы, не отвечают этому требованию. Первые гласные в словах *водá* и *валѹн* акустически не различаются, вместе с тем предполагается, что они представляют разные фонемы.

Вышеприведенные данные позволяют без всякого риска утверждать, что число психологических фонем превышает число лингвистических фонем любого языка, однако неясно, как велико это превышение.

Следует заметить, что существование каких-то минимальных единиц восприятия, аналогичных по функции психологическим фонемам (аллофонов, звукотипов), допускается во многих моделях восприятия речи [215. 491. 527]. Однако авторы этих моделей считают, что на более высоких уровнях восприятия эти единицы каким-то образом трансформируются в лингвистические фонемы. Следует подчеркнуть, что в пользу существования такой трансформации нельзя привести ни одного веского довода.

#### 4.5. ФОНЕМНАЯ КЛАССИФИКАЦИЯ

Для описания связи между речевым сигналом и фонемным решением необходимо указать признаки сигнала, на основе которых осуществляется классификация, и решающие правила. Единственный способ указания признаков заключается в описании алгоритмов (процедур) обработки сигналов, обеспечивающих выделение и измерение этих признаков. Пока признаки остаются неизвестными, — а ситуация сейчас именно такова, — исследователи вынуждены использовать следующую стратегию постепенного приближения к истине. Сначала на основании каких-то априорных сведений, например полученных при исследовании речеобразования, выбирается некоторый условный набор признаков. Далее синтезируется множество стимулов, различа-

ющихся по этим признакам, и эти стимулы предъявляются испытуемым для фонемной классификации. Результатом эксперимента является установление фонемных границ в пространстве априорно выбранных условных признаков. Аналитическое описание фонемных границ дает решающие функции, определенные, естественно, также в терминах этих условных признаков.

Эти фонемные границы (решающие функции) являются промежуточным результатом, позволяющим дальше проверять и уточнять сами гипотезы о природе признаков, а также исследовать некоторые общие характеристики фонемного классификатора (постоянство правил или наличие динамической подстройки и т. д.).

Необходимо подчеркнуть, что все фонемные границы, имеющиеся сейчас в литературе по восприятию речи, являются условными в том смысле, как было определено выше. Непосредственная их полезность для автоматического распознавания речи пока весьма проблематична, так как трудно надеяться на то, что исследователю так повезло, что он выбрал в качестве условных признаков именно те признаки, которые выделяет в сигнале слуховая система человека.

#### 4.5.1. ФОНЕМНЫЕ ГРАНИЦЫ СТАЦИОНАРНЫХ ГЛАСНЫХ В ПРОСТРАНСТВЕ ЧАСТОТ ФОРМАНТ

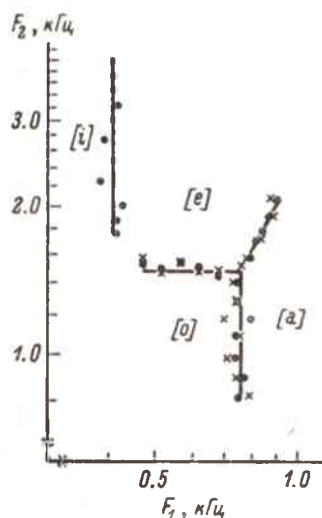
Попытки определить фонемные области стационарных гласных в пространстве частот формант стимулов предпринимались рядом исследователей на материале разных языков [208, 306, 337, 364, 458, 505].

На рис. 4.8 приведены данные, полученные для русских испытуемых. Граница между гласными [i] и [e] была определена методом активного поиска в экспериментах с двухформантными синтетическими гласными [107]. Частота второй форманты в пределах каждого данного измерения была фиксированной (принимала одно из значений, указанных по оси  $F_2$ ); испытуемые, управляя частотой  $F_1$ , искали на слух то ее значение, которое соответствовало границе между [i] и [e].

Три другие границы на рис. 4.8 были найдены в эксперименте по идентификации четырехформантных синтезированных гласных методом форсированного выбора [129]. Испытуемым предъявлялись два набора стимулов, различавшихся лишь фиксированными значениями частот  $F_3$  и  $F_4$  (2660 и 3500 Гц для первого и 3540 и 4000 Гц для второго набора). Частоты  $F_1$  и  $F_2$  этих стимулов изменялись в широком диапазоне (470—920 Гц и 800—2070 Гц соответственно) логарифмически равными шагами ( $\Delta F = 0.1 F$ ). Точки и крестики на рис. 4.8 представляют разные наборы стимулов и соответствуют условию равновероятности выбора соседних фонем. Видно, что границы в плоскости  $F_1 F_2$  практически не зависят от значений частот третьей и четвертой форманты.

Рис. 4.8 показывает, что три фонемных границы из четырех аппроксимируются вертикальными или горизонтальными прямыми линиями. Это означает, что решения о фонемах могут приниматься на основе измерения единственного параметра — частоты  $F_1$  или  $F_2$ .

Этот факт можно рассматривать как подтверждение правильности распространенной гипотезы о том, что слуховая система



выделяет в гласных какие-то признаки, или совпадающие с частотами формант, или близкие к ним (подробное обсуждение этого вопроса см. в главе 10).

Более сложной оказалась граница между областями гласных [a] и [e]. В данном случае возможной аппроксимацией этой границы является уравнение  $\lg F_1 - 0.5 \lg F_2 = 1.3$ , т. е. указанная граница является функцией частот обеих формант. Дан-

Рис. 4.8. Границы между синтетическими гласными в плоскости  $F_1 F_2$ . По [313].

По оси абсцисс — частота первой форманты; по оси ординат — частота второй форманты. Остальные обозначения см. в тексте.

ные, полученные на испытуемых чехах, шведах и поляках [208, 306, 337], подтверждают этот факт и, кроме того, указывают на то, что аналитическое выражение границы между [a] и [e] может иметь более простой вид:  $\lg F_1 - \lg F_2 = \text{const}$ . Это означает, что в качестве решающего правила может использоваться постоянство разности значений  $F_1$  и  $F_2$  по логарифмической шкале.

Очевидно, что исследование различения гласных [a] и [e] является наиболее перспективным для понимания принципов обработки спектральной информации.

#### 4.5.2. ОБЛАСТЬ НЕОПРЕДЕЛЕННОСТИ И КОНТЕКСТУАЛЬНЫЕ ВЛИЯНИЯ ПРИ ВОСПРИЯТИИ ГЛАСНЫХ

Фонемная граница соответствует значениям параметров стимула, при которых две соседние фонемы выбираются с равной вероятностью. Интервал значений параметров стимула, заключенный между 0.25 и 0.75 ответов (данной фонемой) на функции идентификации, называется областью неопределенности.

Экспериментально показано, что время (задержка) двигательной реакции, связанное с принятием фонемного решения (па-



жатие на кнопку, запись буквы и т. д.), закономерно возрастает внутри области неопределенности и достигает максимума на фонемной границе [86, 213, 492]. Этот факт указывает на то, что для описания перехода от акустических (слуховых) параметров стимула к фонемам недостаточно «знать» фонемные границы (решающие правила) в пространстве указанных параметров: необходимо еще допустить существование некоторых промежуточных переменных (сигналов), отражающих величину «сходства» с фонемой (см. обсуждение в [152, 158]).

Можно думать, что в экспериментах по идентификации изолированных гласных фонемная интерпретация «неопределенного» стимула зависит в основном от значений функций сходства с соседними фонемами. Если же неопределенный гласный предъявляется в некотором контексте, то логично ожидать, что фонемное решение о нем будет определяться преимущественно другими, «определенными» фонемами.

В работе [115] было показано, что при пересадке синтезированного фонетически неопределенного гласного в естественные слоги СГС, предъявлявшиеся изолированно или в составе слов и фраз, оценка фонемного качества гласного зависит от окружающих согласных: например, стимул с параметрами  $F_1=320$  Гц и  $F_2=1420$  Гц в окружении твердых согласных чаще интерпретируется как [i], в окружении мягких согласных — как [u].

Эффекты подобного рода получили в литературе название контекстуальных влияний. Хотя основной интерес представляют контекстуальные влияния в связной речи — внутри слова и предложения, — анализ механизмов контекстуальных влияний в этой ситуации оказывается очень сложным, так как один и тот же контекстуальный эффект может достигаться под действием разных факторов. Априорно можно предполагать, что к числу таких факторов относится, например, разрешение неопределенности на уровнях анализа смысла, сравнение запомненной информации (значений сходства с фонемами) о соседних гласных в последовательности, изменения состояния самого фонемного классификатора под действием предшествующего стимула, т. е. эффекты последствия, и т. д.

По этой причине исследования сейчас ограничиваются в основном значительно более простой экспериментальной ситуацией, когда влияющий стимул предшествует тестирующему и отделен от него достаточно большой паузой. Ответ о тестирующем стимуле дается сразу же после его предъявления.

В ряде работ [86, 210, 265, 363] было продемонстрировано два эффекта последствия — эффект контраста и эффект ассимиляции (уподобления). Эти эффекты удобно описывать, используя понятия среднего и текущего значения фонемной границы. Если, например, влияющий стимул находится справа от среднего значения фонемной границы, то эффект контраста проявится в том, что текущее значение границы после предъявления влияющего стимула уста-



новится правее среднего ее значения. Эффект ассимиляции в этом случае будет, наоборот, заключаться в том, что текущее значение границы установится левее среднего ее значения.

Данные работы [210] говорят о том, что эффект контраста и эффект ассимиляции связаны с разными уровнями обработки сигнала. В эксперименте [210] испытуемым предъявлялась для идентификации ([u] или [o]) случайная последовательность всевозможных пар стимулов, параметры которых указаны в подписи к рис. 4.9. Полученные данные позволяют судить о положении границы между фонемами [u] и [o] в зависимости от решения о фонемной принадлежности влияющего стимула (первого стимула каждой пары) и положения его на шкале параметров. На рис. 4.9 показаны границы, вычисленные при условии, что влияющий стимул был опознан как [o] (прямая 1) и как [u] (прямая 2).

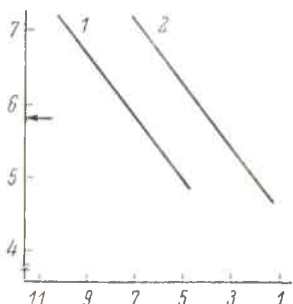


Рис. 4.9. Положение фонемной границы между гласными [u] и [o], вычисленной по данным для второго стимула в паре, при идентификации обоих стимулов пары. По [210].

По горизонтали отложен номер первого стимула в паре, по вертикали — номер второго стимула в паре (приведены только стимулы в пограничной между [u] и [o] области). Прямая 1 аппроксимирует положение границы в случае, когда первый стимул в парах принимался за [o], прямая 2 — когда первый стимул принимался за [u]. Стрелка указывает среднее положение границы. Ниже приведены номера и частоты формант (в Гц) стимулов (стимул 1 соответствует [u], стимул 11 — [o]):

|                | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   |
|----------------|------|------|------|------|------|------|------|------|------|------|------|
| F <sub>1</sub> | 250  | 267  | 284  | 301  | 318  | 335  | 352  | 369  | 386  | 403  | 420  |
| F <sub>2</sub> | 750  | 780  | 770  | 780  | 790  | 800  | 810  | 820  | 830  | 840  | 850  |
| F <sub>3</sub> | 2400 | 2420 | 2440 | 2460 | 2480 | 2500 | 2520 | 2540 | 2560 | 2580 | 2600 |

Обе прямые свидетельствуют о наличии эффекта контраста: повышение формантных частот влияющего стимула приводит к соответствующему сдвигу границы между [u] и [o]. Тот факт, что положение фонемной границы изменяется под влиянием физически разных стимулов, опознаваемых как одна и та же фонема, говорит о том, что эффект контраста определяется акустическими параметрами гласных стимулов.

Разница между прямыми 1 и 2 в области влияния стимулов 5, 6 и 7 на рис. 4.9 демонстрирует эффект ассимиляции. Действительно, если влияющий стимул 7 был опознан как [u], то фонемная граница устанавливалась таким образом, что неопределенные стимулы 6 и 7 из тестирующего множества также воспринимались как [u]. Если же влияющий стимул 7 принимался за [o], то тестирующие стимулы 6 и 7 также воспринимались как [o]. Это свидетельствует о том, что в отличие от эффекта контраста эффект ассимиляции определяется фонемным решением о влияющем стимуле.

Существенно отметить, что эффект последействия типа контраста наблюдался также в такой ситуации, когда влияющими стимулами были гласные [u] и [o], а тестируемыми стимулами — гласные, воспринимаемые как [i] и [e] [91]. Было обнаружено, что фонемная граница по  $F_1$  между гласными [i] и [e], определенная методом активного поиска, существенно ниже, когда тестирующий (переменный) стимул следует за [u], чем когда он следует за [o]. Это подтверждает, что эффект контраста определяется акустическими параметрами стимулов и действует на уровне обнаружения отдельных признаков (в данном случае — частот формант).

#### 4.5.3. ДЕТЕКТОРЫ ПРИЗНАКОВ СОГЛАСНЫХ

Наиболее популярными экспериментальными моделями, используемыми при исследовании восприятия согласных, являются наборы стимулов СГ, различающихся по «времени начала голоса» (рис. 4.2), и наборы стимулов СГ, различающихся по характеру быстрых переходов второй и третьей форманты.

Первый набор используется для определения фонемной границы между глухими и звонкими начальными взрывными согласными, второй набор — для определения фонемных границ между взрывными согласными, различающимися по месту образования.

Популярность указанных стимулов определяется тем, что фонематически полезная информация заключена в этом случае или в последовательности акустических явлений, или в характере быстрых изменений спектра во времени. Для извлечения этой информации необходимы явно другие методы обработки сигнала, чем те, которые предполагаются для случая стационарных гласных [483]. Можно считать, что эти стимулы, с одной стороны, и стационарные гласные, с другой, представляют собой как бы две крайние ситуации, с которыми встречается человек при классификации речевых сигналов. Исследование фонемной идентификации такого типа стимулов в обычных экспериментальных условиях (разные стимулы следуют с паузами друг за другом) не обнаруживало эффектов последействия типа контраста [265].

Однако за последние два года выяснилось, что такие эффекты можно получить, если использовать длительное (многократное) предъявление влияющего стимула [179, 185, 222, 231, 243]. В качестве примера можно привести описание методики, примененной Купером [222], и полученные им результаты. В эксперименте Купера использовался ряд синтезированных слогов типа СГ с гласным [æ], различавшихся по исходным значениям и направлению переходов  $F_2$  и  $F_3$  на начальном участке гласного длительностью в 40 мс. При идентификации этих стимулов группой испытуемых-американцев было получено три фонемных категории, соответствующих слогам с билабиальным [bæ], альвеолярным [dæ] и велляр-

ым [gæ] согласными. На рис. 4.10, А показано расположение указанных областей согласных у одного из испытуемых.

После этого проводился эксперимент по идентификации в присутствии влияющих стимулов. Как и в первом опыте, испытуемые слушали случайную последовательность исходных стимулов, каждый из которых надо было идентифицировать с одним из трех

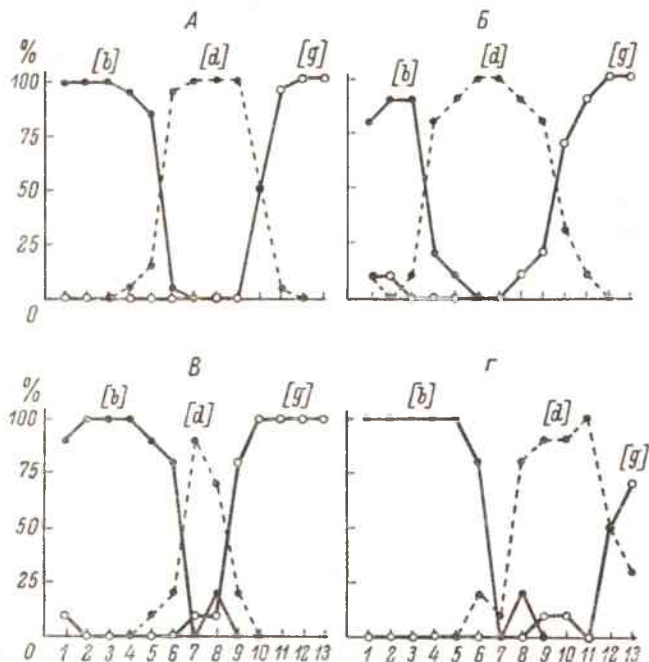


Рис. 4.10. Функции идентификации синтетических слогов, различающихся по направлению переходов  $F_2$  и  $F_3$ , полученные в условиях влияния предшествующих стимулов. По [222].

По оси абсцисс — номер стимула; по оси ординат — относительное число ответов [b], [d] и [g]. На А: функции идентификации слогов [bæ], [dæ] и [gæ] в отсутствии контекста. На Б, В, Г: функции идентификации тех же слогов при использовании в качестве предшествующего стимула соответственно слогов [bæ], [dæ] и [gæ].

указанных согласных. Но в этом случае предъявлению серии из четырех тестирующих стимулов предшествовало многократное (90 раз в течение минуты) предъявление влияющего стимула. В качестве последнего использовались (в разных сериях прослушиваний) слоги [bæ], [dæ] и [gæ], взятые из исходного набора стимулов и идентифицированные в качестве таковых в 100% случаев. Рис. 4.10, Б, В, Г показывает функции идентификации, полученные у одного из испытуемых под влиянием вышеуказанных предшествующих стимулов. Видно, что предъявление влияющих стимулов привело к сдвигу фонемных границ по принципу контраста. Эти данные были подтверждены в экспериментах

по различению стимулов в присутствии влияющих стимулов. Максимумы на функциях различения, полученных в этих условиях, хорошо совпадают с положением границ, найденных при идентификации стимулов в тех же условиях.

Обнаруженный эффект последствия получил название «селективной адаптации». Он сразу же привлек к себе большое внимание исследователей. Варьируя характеристики влияющих и тестирующих стимулов, можно попытаться выяснить, с какими уровнями обработки сигнала связан этот эффект.

Предполагается, что эффект селективной адаптации действует на уровне обнаружителей (детекторов) признаков согласных и вызывается утомлением детекторов, происходящим из-за многократного предъявления одного и того же стимула. Это, очевидно, означает, что выходом классификатора согласных являются не фонемы, а признаки, т. е. имеются, например, детекторы звонкости-глухости, работающие независимо от места или способа образования согласных, или детекторы места образования, работающие независимо от того, глухие или звонкие согласные подаются на его вход.

Доказательством такого представления является тот факт, что в экспериментах [<sup>222</sup>] сдвиг границы между областями [b] и [d] происходил не только под действием влияющего стимула [bæ], но и под действием стимула [p<sup>h</sup>æ], согласный которого совпадает с [bæ] по месту образования. Точно так же в работе [<sup>243</sup>] граница между согласными [d] и [t<sup>h</sup>], различающимися по признаку звонкости-глухости, смещалась при использовании в качестве влияющих стимулов слогов с согласными [b] и [p<sup>h</sup>].

При предположении детекторов психологических (дифференциальных) признаков встает вопрос о том, имеются ли универсальные детекторы, работающие во всей частотной области, или существует множество детекторов каких-то акустических признаков. Первое допущение означает, что если, например, рассматривать признак места образования, то детектор этого признака сработает одинаково при предъявлении слогов [di] и [du], согласные в которых различаются акустически. Имеются указания [<sup>185</sup>], что это допущение неверно.

Вопрос о том, какие акустические признаки могут измеряться при восприятии согласных, пока не решен. Например, имеются данные, что информация о месте образования согласных содержится в формантных переходах (см. главу 12). Это может означать, что в локальных частотных областях существуют измерители (детекторы), реагирующие на направление изменения частоты. Решение о признаке — месте образования согласного принимается в результате суммирования данных таких детекторов. С другой стороны, можно представить, что в качестве самостоятельных параметров выделяются частоты формант и что имеются схемы, измеряющие направление изменения этих переменных. В экс-

периментах [185] была засвидетельствована адаптация к отдельной частоте форманты, однако этих данных недостаточно для выбора между указанными возможностями.

В приведенном кратком обзоре мы не останавливались на большом числе фактов, касающихся того, какие вообще особенности речевого сигнала могут быть существенными для фонемной классификации. Соответствующие более подробные данные можно найти, например, у Стивенса и Хауза [485].

Опыт работы по синтезу по правилам [218] говорит о том, что при восприятии речи используется очень большое число признаков сигнала и что труднее найти такие изменения в сигнале, которые прошли бы незамеченными, чем наоборот. Это, очевидно, означает, что чисто эмпирическая работа по поиску признаков малоперспективна. Необходимо иметь хотя бы предварительные гипотезы о том, какого типа обработку речевого сигнала осуществляет слуховая система и какие признаки сигнала имеет смысл выбирать в качестве условных при исследовании фонемных границ. Такую информацию можно в какой-то мере извлечь из данных физиологии слуха и психоакустики.

## Глава 5

### ВОСПРИЯТИЕ ПРОСОДИЧЕСКОЙ ИНФОРМАЦИИ

По определению Трубецкого [132], «звуковые признаки, образующие в разных языках смыслоразличительные оппозиции, можно подразделить на три класса: вокалические, консонантные и просодические» (стр. 102).

Отсюда вытекает, что для понимания смысла речевого сообщения необходима фонемная и просодическая информация и что в последнюю входит все воспринимаемое в речевом сигнале, помимо последовательности фонем. Совершенно очевидно, что при таком определении просодической информации ее конкретное содержание в значительной степени зависит от содержания фонемной информации. Так, если фонемная информация соответствует последовательности фонемных вариантов (см. главу 4), то сведения о словесной границе, извлекаемые из факта появления пограничных аллофонов [218. 270. 342. 343], следует отнести к фонемной, а не к просодической информации. Напротив, если фонемная информация представлена фонемами московской фонологической школы, то сведения о словесных границах придется отнести в разряд просодической информации.

Приступая к исследованию восприятия просодической информации, целесообразно разделить эту сложную проблему на несколько менее крупных, связанных с решением следующих вопросов.

1. Какие (какого типа) сведения, необходимые для понимания смысла устного сообщения, извлекаются из акустического сигнала при восприятии речи?

2. В каких параметрах сигнала эти сведения содержатся? При этом надо иметь в виду, что параметры — это результат обработки акустического речевого сигнала слухом, т. е. специфическим измерительным прибором.

3. Что представляют собой процедуры перехода от параметров к промежуточным смыслоразличительным решениям (просодическим признакам)?



Имеющиеся в литературе данные позволяют представить, что :анализа просодической информации, содержащейся в речевом сигнале, могут быть получены следующие важные для понимания смысла сообщения сведения.

1. Прежде всего сведения, необходимые для установления количества слов в высказывании, их выделения и опознавания, т. е. для морфологического анализа. Эти сведения могут быть получены из анализа ритмической структуры высказывания и распределения пограничных сигналов в нем. Поскольку ритм речи есть упорядоченное чередование выделенных и невыделенных элементов (слов), часть задачи решается выяснением места (номера) выделенного слога (ударного) в последовательности слогов, описывающей данный речевой сигнал.

2. Сведения, существенные для синтаксического анализа высказывания, т. е. для установления отношений между словами в нем.

3. Данные об особенностях конкретного процесса речевого общения: является ли высказывание запросом или сообщением информации, какой именно информации, как относится говорящий к сообщаемой им или ему информации и т. д. Так, предложение *Мама мыла Маню?*, реализованное в виде высказывания, может иметь, в зависимости от ситуации, несколько разных смыслов: *Кто же мыл Маню? Кого мыла мама? Мыла ли уже мама Маню?*

Очень важно отметить, что п а р а м е т р ы речевого сигнала, используемые для передачи просодической информации, и п р о с о д и ч е с к и е п р и з н а к и речи, по всей вероятности, не одно и то же. Здесь нам кажется уместной аналогия с дифференциальными признаками фонов. Так, частота форманты, будучи параметром речевого сигнала, может нести информацию и о фоне, и об особенностях диктора, а возможно, и просодическую информацию. Но дифференциальный признак фонемы — это результат сравнения измеренной в сигнале частоты форманты с некоторым пороговым значением, т. е. уже не только сам параметр, но и определенное решающее правило.

Соответственно этому и просодические признаки появляются в результате применения некоторых решающих правил к параметрам речевого сигнала. Естественно, что при таком определении одни и те же параметры сигнала могут быть источником не только просодической информации, но и сведений об эмоциональном и даже физическом состоянии диктора [39], которые, строго говоря, к просодической информации не относятся, так как не являются смыслообразительными. Для этого достаточно использовать разные решающие правила применительно к одним и тем же параметрам.

Ясно также, что такие физические параметры речевого акустического сигнала, как частота основного тона, длительность и интенсивность звуков и речевой темп — общепризнанные источники

просодической информации — не могут быть отнесены к просодическим признакам речи, так как после обработки в слуховой системе из них может быть извлечена не только просодическая информация.

Таким образом, из акустического речевого сигнала человек способен, по всей вероятности, извлекать не только необходимую ему для понимания смысла речевого сообщения информацию, но и признаки, характеризующие диктора и его состояние. Работы по автоматическому распознаванию речи не только доказывают возможность автоматического выделения ударных слогов [141, 341], границ между словами и синтаксическими группами [340, 529], но показывают также, что одновременное использование просодической и фонемной информации (фонемных вариантов) дает возможность существенно повысить качество грамматического анализа речевого сообщения [362, 409].

Однако попытки исследовать просодические признаки с использованием естественного речевого материала приводят подчас к обескураживающим результатам: восприятие того или иного признака оказывается подверженным влиянию лексического состава экспериментального материала, контекста и даже степени лингвистической подготовки слушателя. Оказалось, например, что при изучении восприятия слушателями изменений темпа последние обнаруживали изменения при их полном физическом отсутствии и могли не заметить изменений темпа на 100—120% [148]. Аналогичные явления наблюдались и при исследовании восприятия паузы в речевом сигнале. В пределах одного высказывания пауза длительностью 200 мс могла не восприниматься совсем или восприниматься в 100% случаев. Случалось также, что в одном высказывании слушатели не воспринимают паузу в 100 мс и прекрасно воспринимают паузу длительностью 40 мс [51]. При экспериментальном увеличении длительности паузы на стыке слов в естественном сигнале до 500 мс она воспринималась слушателями всего в 50% случаев [196].

В опытах Либермана [355] была следана попытка выяснить, каким физическим признакам в речевом сигнале и в какой мере соответствует широко применяемая в английском языке система транскрипции интонации Трэгера и Смита [507]. Испытуемым (опытным лингвистам) предлагалось транскрибировать по этой системе естественный речевой материал, а затем синтетические «фразы», состоявшие из протяжного гласного [a], но сохранявшие ритмико-мелодическую структуру исходных естественных фраз. Выяснилось, что при исключении лексической информации даже опытные лингвисты в состоянии обнаружить только две степени ударения вместо четырех в естественном сигнале. Исключение лексической информации приводило также к более точному восприятию мелодических изменений, из чего автор делает вывод, что при транскрибировании естественной речи лингвисты ориентируются не столько на конкретные физические признаки,

сколько на свои знания того, какими транскрипционными символами должна быть описана данная комбинация слов.

Из сказанного можно сделать вывод, что переходя от просодических признаков к их лингвистической интерпретации в процессе восприятия естественного речевого сигнала, слушатели теряют способность давать осознанную оценку этих признаков как таковых. По-видимому, исследование просодических признаков целесообразно проводить либо психоакустическими методами с применением сигналов, лишенных лексического содержания, либо с применением таких естественных сигналов, в которых изменение одного признака меняет смысл всего высказывания.

Как уже говорилось в главе 2, система лингвистического анализа имеет дело с описанием речевого потока в виде последовательности дискретных элементов (фонемных вариантов), внесенных в ячейки оперативной памяти. Естественно предположить, что просодическая информация заносится в те же ячейки оперативной памяти и, следовательно, тоже должна быть представлена последовательностью дискретных величин. Хотя о структуре оперативной памяти в настоящее время еще ничего не известно, есть основания считать, что ячейка памяти соответствует слогу в речевом потоке. Сам же слог при восприятии речи человеком может члениться на «гласный» и «негласный» элементы (см. главу 3), что делает вполне вероятным допущение об аналогичном разбиении ячейки оперативной памяти, отведенной слогу (подробнее об этом см. в главе 6).

Из сказанного следует, что при записи просодической информации в оперативную память некоторые дискретные величины могли бы быть «приписаны» всему слогу или «гласному» и «негласному» его элементам. Возможно также, что часть информации «приписывается» слогу, а часть — его составляющим. Отсюда — дополнительный вопрос, на который должна дать ответ теория восприятия речи: в какой форме и вместе с какими элементами речевого потока производится запись просодической информации в оперативную память?

В этой главе мы рассмотрим вначале, что известно о признаках речевого сигнала, несущих просодическую информацию, а затем — некоторые вопросы записи этих признаков в память.

#### **5.1. ИНТЕНСИВНОСТЬ И ДЛИТЕЛЬНОСТЬ КАК САМОСТОЯТЕЛЬНЫЕ ИСТОЧНИКИ ПРОСОДИЧЕСКОЙ ИНФОРМАЦИИ. ДАННЫЕ ИССЛЕДОВАНИЙ ВОСПРИЯТИЯ СЛОВЕСНОГО УДАРЕНИЯ**

Сегодня ни у кого из исследователей речи не вызывает сомнения, что интенсивность и длительность элементов речевого потока могут нести просодическую информацию и, в частности, сведения о месте словесного ударения. Не существует,

однако, единого мнения в вопросе о том, являются ли эти признаки самостоятельными или при восприятии речи они образуют некоторый сложный, но один признак. Так, предпринимались попытки объединить их в параметр типа «текущей энергии», чтобы избежать измерения двух самостоятельных параметров. Как было установлено при анализе акустического речевого сигнала, акустическая энергия гласного достаточно хорошо коррелирует со степенью его ударенности в английском [354] и украинском [22] языках. Использование энергии гласного для принятия решения о месте словесного ударения предполагает наличие устройства, интегрирующего интенсивность речевого сигнала за время, соответствующее длительности гласного, и сбрасывающего до нуля вычисленное значение энергии после снятия отсчета, чтобы сделать возможным вычисление энергии следующего гласного.

В главе 3 было показано, что при восприятии речи слуховая система человека должна осуществлять сегментацию непрерывного речевого сигнала на последовательность отрезков «гласных» и «негласных». Этим создается возможность интегрирования интенсивности сигнала на участке гласного. Однако данных о существовании в слуховой системе измерителя энергии со сбросом пока нет и, кроме того, психоакустические исследования показывают, что при восприятии словесного ударения человек не пользуется полной энергией гласного [136].

Некий эквивалент «текущей энергии» могла бы дать обычная схема выделения общей энергетической огибающей речевого сигнала при соответствующем подборе параметров ее элементов: детектора и фильтра нижних частот. Выходной сигнал такого устройства имел бы множество максимумов, соответствующих слоговым ядрам (гласным), и задача поиска места словесного ударения решалась бы путем выделения и сравнения этих максимумов без сегментации речевого потока. Очевидно, с точки зрения автоматического распознавания такой подход весьма привлекателен, но с точки зрения психоакустики он кажется весьма неправдоподобным.

Из психоакустических измерений следует, что аналогом «текущей энергии» в слухе может быть громкость звука. Поскольку все модели слухового измерителя громкости имеют интегрирующее звено (как правило, в виде фильтра нижних частот), при определенном соотношении постоянной времени такого фильтра и длительности звука громкость последнего может быть не только функцией интенсивности, но и длительности. Не исключено поэтому, что при восприятии словесного ударения различия в длительности звуков речи при равной их интенсивности могли бы восприниматься не сами по себе, а через различия в громкости этих звуков.

Экспериментально показано, что при средних уровнях интенсивности фильтр нижних частот модели слухового измерителя громкости должен иметь постоянную времени в 30—50 мс [442].

Если минимальную длительность гласного в речи принять равной 50 мс, то при такой постоянной времени увеличение длительности гласного до 100 мс при неизменной интенсивности даст наибольшее приращение его громкости, эквивалентное увеличению интенсивности на 2—3 дБ. А изменение длительности гласного, скажем, от 100 до 200 мс уже не даст практически никакого увеличения его громкости.

Исследования, посвященные восприятию словесного удара в двусложных словах, показывают, что при равной длительности гласных увеличение интенсивности одного из них на 3—5 дБ еще не приводит к восприятию его как ударного [307, 314, 398]. Из этого можно сделать вывод, что изменение длительности звуков речи вряд ли может восприниматься через изменение их громкости.

С другой стороны, известна способность человека оценивать звуки по длительности, причем субъективная длительность звука в широком диапазоне изменения интенсивности не зависит от последней [541]. Кроме того, относительный дифференциальный порог по длительности, полученный в опытах, где оценивалась именно длительность стимула, в интересующем нас диапазоне длительностей (50 мс и более) не превышает 10—15% [173, 223, 267] и достаточно хорошо совпадает с величинами, полученными в экспериментах, где критерием служила степень ударенности (см. раздел 5.2). Еще одним аргументом в пользу возможности независимого измерения слухом интенсивности и длительности элементов речевого потока может быть тот факт, что в ряде языков длительность является единственным фонематическим признаком некоторых гласных (см. главу 3).

Таким образом, есть все основания предположить, что в процессе восприятия речи интенсивность и длительность являются самостоятельно измеряемыми параметрами, но при принятии решения о том или ином просодическом признаке они используются одновременно, причем вполне возможно, что влияние («вес») каждого из них не одинаково.

Экспериментально-фонетические исследования признаков словесного удара привели к выводу, что в русском языке интенсивность менее существенна, чем длительность [60, 74, 125], что «относительная интенсивность гласного зависит не столько от его ударности-неударности, сколько от его места в слове» [15].

Попытки выяснить роль интенсивности и длительности в восприятии словесного удара приводят подчас к противоречивым результатам; это либо является следствием использования разных методических подходов, либо отражает различие между языками. Так, в голландском языке ни увеличение интенсивности, ни увеличение длительности гласного в синтетическом сигнале не вызывает эффекта восприятия его ударным [314].<sup>1</sup> В английском языке по

<sup>1</sup> Необходимо отметить, что в голландском языке длительность гласного является фонематическим признаком. В связи с этим вопрос о возможности использования этого признака для передачи просодической информации



данным одних авторов интенсивность не играет роли в создании ударения [<sup>386</sup>], другие полагают, что и длительность, и интенсивность создают эффект выделенности, но только «вес» каждого из этих признаков неодинаков [<sup>263</sup>]. К аналогичным выводам приходят и другие исследователи, изучавшие роль этих параметров в восприятии словесного ударения в русском [<sup>7, 136</sup>], польском [<sup>307</sup>] и английском [<sup>398</sup>] языках.

На рис. 5.1. и 5.2 представлены частоты восприятия ударения на первом слоге синтетической двусложной последовательности носителями разных языков (английского, польского, русского) в функции разности длительностей и интенсивностей первого и второго гласных. Эти результаты получены в экспериментах, где англичанам и полякам предъявлялись синтетические слова [sasa], [sɔsɔ] и [sisi], а испытуемым-русским — последовательность гласных, разделенных короткой паузой (такой сигнал воспринимается как слово [aga]). Носителям английского и польского языков разрешались три вида ответов на каждый стимул: «первый ударный», «второй ударный» и «не знаю, не уверен, ударные оба». Носителям русского языка неопределенные ответы были запрещены.

Приведенные на рис. 5.1 данные о восприятии словесного ударения англичанами и поляками получены в экспериментах [<sup>307, 308</sup>], где носителям разных языков предъявлялись практически одинаковые синтетические сигналы. Если бы признаком словесного ударения была громкость, следовало бы ожидать одинаковых результатов для носителей разных языков, так как в этом случае восприятие ударения определялось бы свойствами измерителя громкости в слуховой системе человека. Совершенно очевидно, однако, что распределения на рис. 5.1, А и 5.1, Б существенно различаются. Такой результат возможен при условии, что длительность и интенсивность измеряются как независимые признаки и по-разному используются носителями разных языков.

По виду приведенных на рис. 5.1 и 5.2 распределений можно сделать заключение, что в исследованных языках (английском, польском и русском) и интенсивность, и длительность являются признаками словесного ударения. Ясно также, что влияние каждого из этих признаков на восприятие словесного ударения (его «вес») неодинаково в разных языках.

Семейство гистограмм, приведенных, например, на рис. 5.2, можно описать одним выражением, учитывающим зависимость вероятности ответа «первый слог ударный» сразу от двух параметров сигнала: разности интенсивностей и длительностей гласных, — применив следующую процедуру.

---

требует специального изучения. Видимо, говорить о «весе» конкретного признака при передаче просодической информации имеет смысл лишь применительно к тем языкам, в которых данный признак не является фонематическим.



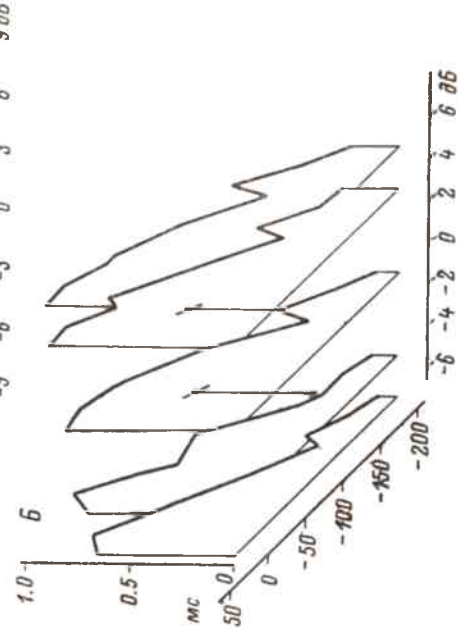
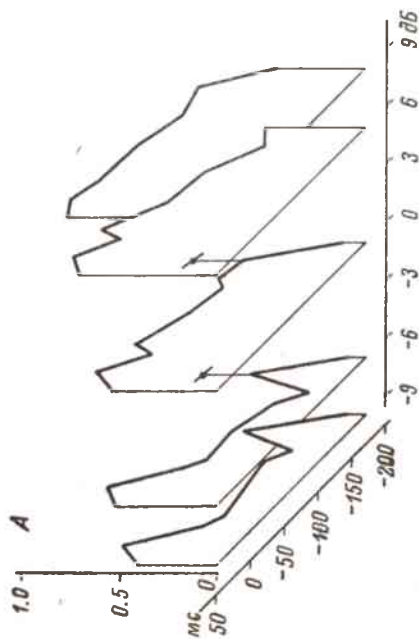


Рис. 5.2. Зависимость частоты ответов «первый слог ударный» от соотношения интенсивностей и длительностей гласных при восприятии испытуемыми-русскими ударения в синтетической двусложной последовательности (данные [136]).

Обозначения те же, что на рис. 5.1.

Рис. 5.1. Зависимость восприятия испытуемыми-англичанами (А) и испытуемыми-поляками (Б) ударения на первом слоге синтетической двусложной последовательности от соотношения интенсивностей и длительностей входящих в нее гласных.

По оси абсцисс — разность интенсивностей первого и второго гласных; по оси ординат — разность длительностей этих же гласных; по вертикали — частота ответов «первый слог ударный». Неопрделенные ответы исключены из общего числа.

Допустим, что при заданной разности интенсивностей гласных зависимость вероятности ответа «первый слог ударный» от разности длительностей может быть аппроксимирована интегральной кривой нормального распределения с параметрами  $\Delta \bar{T}$  и  $\sigma$ . Вычислив значения этих параметров для каждой из приведенных на рис. 5.2

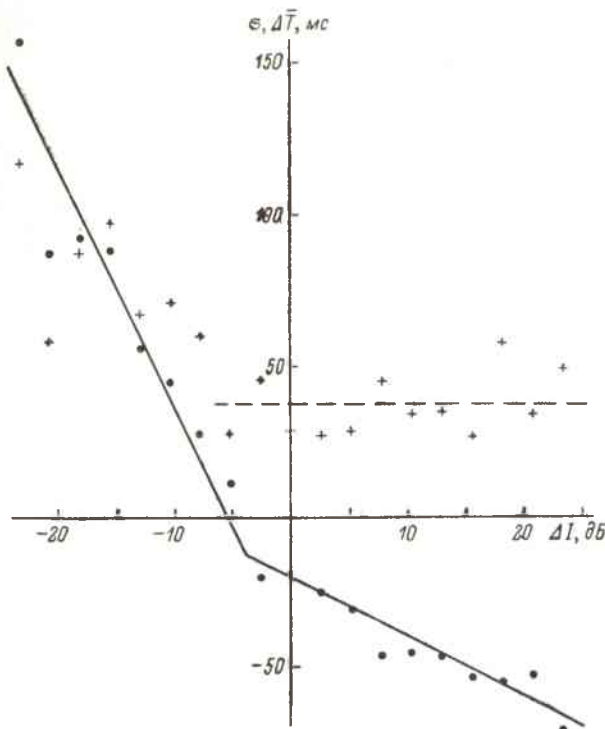


Рис. 5.3. Параметры интегральных кривых нормального распределения, аппроксимирующих гистограммы ответов «первый слог ударный» (рис. 5.2), при разном соотношении интенсивностей гласных в двусложном синтетическом стимуле.

По оси абсцисс — разность интенсивностей первого и второго гласных; по оси ординат — разность длительностей этих гласных, при которой оба слога воспринимаются ударными с равной вероятностью, и величина среднеквадратического отклонения. Крестики — среднеквадратическое отклонение, точки — пограничная величина разности длительностей.

гистограмм, получим зависимости  $\Delta \bar{T}$  и  $\sigma$  от разности интенсивностей  $\Delta I$  (рис. 5.3). На этом рисунке отчетливо выделяются две области: область сравнительно постоянных  $\sigma$  и область, в которой  $\sigma$  увеличивается с увеличением разности интенсивностей. Последнее, вероятнее всего, является следствием неестественности части использованных в эксперименте сигналов с точки зрения носителя русского языка: в естественной речи первый глас-

ный в двусложном слове не может быть существенно меньше второго по интенсивности. Судя по данным рис. 5.3, этот эффект начинает сказываться при  $\Delta I = -4$  дБ.

Рассмотрим далее только ту область, в которой  $\sigma$  постоянна. Здесь усредненное среднеквадратическое отклонение ( $\sigma$ ) равно 37.2 мс, а зависимость  $\Delta \bar{T}$  от разности интенсивностей (в дБ) может быть представлена выражением  $\Delta \bar{T} = a + b \cdot \Delta I$ , где  $a = -20$  мс и  $b = -1.95$  мс/дБ. Вероятность восприятия удара на первом слоге при заданных разностях интенсивностей и длительностей гласных теперь может быть получена из таблиц интегральных вероятностей по известной величине нормированного отклонения:  $z = (\Delta T - \Delta \bar{T})/\sigma = (\Delta T - a - b\Delta I)/\sigma = (\Delta T - a)/\sigma - b \cdot \Delta I/\sigma$ , или  $z = (\Delta T - a)/\sigma - \Delta I/\sigma_I$ , где  $\sigma_I = \sigma/b$ .

Это выражение дало бы возможность оценить «вес» каждого из параметров, если бы удалось найти способ сравнивать столь разные параметры, как длительность (в мс) и интенсивность (в дБ).

Одна из возможностей заключается в установлении того количества дифференциальных порогов, на которое надо изменить каждый из рассматриваемых признаков, чтобы получить одно и то же приращение нормированного отклонения. В результате аппроксимации экспериментальных данных имеем  $\sigma = 37.2$  мс и  $\sigma_I = 19.1$  дБ. Исходная длительность гласного в этих экспериментах [136] составляла 100 мс, что при относительном дифференциальном пороге 10% дает величину абсолютного дифференциального порога 10 мс. При  $\sigma = 37.2$  мс изменение  $z$  на единицу будет достигнуто при изменении длительности гласного на 3.72 величины дифференциального порога.

Для интенсивности абсолютный дифференциальный порог равен 0.3—0.7 дБ [153], и, следовательно, изменение  $z$  на единицу достигается при изменении интенсивности гласного на 27—64 величин дифференциального порога.

При этих условиях следует признать, что длительность гласного имеет больший «вес» при определении места словесного удара.

## 5.2. ДЛИТЕЛЬНОСТИ РАЗНЫХ СЕГМЕНТОВ СЛОГА — РАЗНЫЕ ПРОСОДИЧЕСКИЕ ПРИЗНАКИ

Итак, необходимо принять, что и длительность, и интенсивность гласного являются измеряемыми слухом параметрами речевого сигнала и могут использоваться в разных языках в качестве просодических признаков, в частности, для принятия решения о месте словесного удара. Но поскольку во всех описанных выше экспериментах изменялась и контролировалась только длительность гласного (изменение его длительности, естественно, приводило и к изменению длительности всего слога), по их результатам еще нельзя сказать с уверенностью, важна ли

для восприятия ударения длительность только гласного или всего слога. Возможна и более общая постановка вопроса: для передачи какой просодической информации используются длительность слога и длительности его составляющих («гласного» и «негласного»).

Что касается интенсивности, то без большой ошибки можно считать интенсивность слога равной интенсивности входящего в него гласного, так как она, как правило, существенно больше интенсивности согласного (согласных).

#### 5.2.1. ВЛИЯНИЕ ДЛИТЕЛЬНОСТИ ГЛАСНОГО И СОГЛАСНОГО В СЛОГЕ НА ВОСПРИЯТИЕ СЛОВЕСНОГО УДАРЕНИЯ

Исследованиями процесса речеобразования установлено, что в русской речи словесное ударение характеризуется большей длительностью и гласного, и согласного выделенного слога [60, 74, 125]. В английском языке различий по длительности между согласными в ударных и безударных слогах не обнаружено [263]. При изучении субъективной оценки испытуемыми длительности слогов в естественной русской речи оказалось, что субъективная длительность слога наилучшим образом коррелирует с длительностью гласного в этом слоге [433].

Детальное исследование влияния длительности согласного и гласного на восприятие ударения было проведено с использованием синтетического звуко сочетания [sasasa], в котором одновременно изменялись длительности и гласного, и согласного во втором слоге при постоянных частоте основного тона и интенсивностях согласных и гласных [31]. Были выбраны следующие длительности (в миллисекундах) гласных и согласных в слогах этой последовательности.

|                     | I слог | II слог | III слог |
|---------------------|--------|---------|----------|
| Согласный . . . . . | 200    | 80—360  | 100      |
| Гласный . . . . .   | 160    | 90—380  | 90       |

Если предположить, что на восприятие ударения одинаково влияет и длительность гласного, и длительность согласного в слоге, т. е., другими словами, восприятие ударения определяется суммарной длительностью всего слога, то вероятность получить ответ «второй слог ударный» должна оставаться постоянной при постоянной суммарной длительности слога независимо от характера ее распределения между составляющими слог гласным и согласным. На рис. 5.4, А приведено гипотетическое распределение ответов «второй слог ударный» (прямая *ab* описывает закон изменения длительностей гласного и согласного в слоге при постоянной длительности самого слога). На рис. 5.4, Б приведены экспериментально полученные частоты ответов «второй слог ударный». Видно, что экспериментальные зависимости имеют мало общего с гипотетическими: при увеличении длительности глас-

ного и неизменной суммарной длительности слога (этот случай описывается прямой  $ab$  в плоскости  $T_c - T_r$  на рис. 5.4,  $B$ ) частота ответов «второй слог ударный» возрастает от нуля до единицы.

Как следует из рис. 5.4,  $B$ , при увеличении свыше 200 мс длительность согласного начинает оказывать влияние на частоту ответов

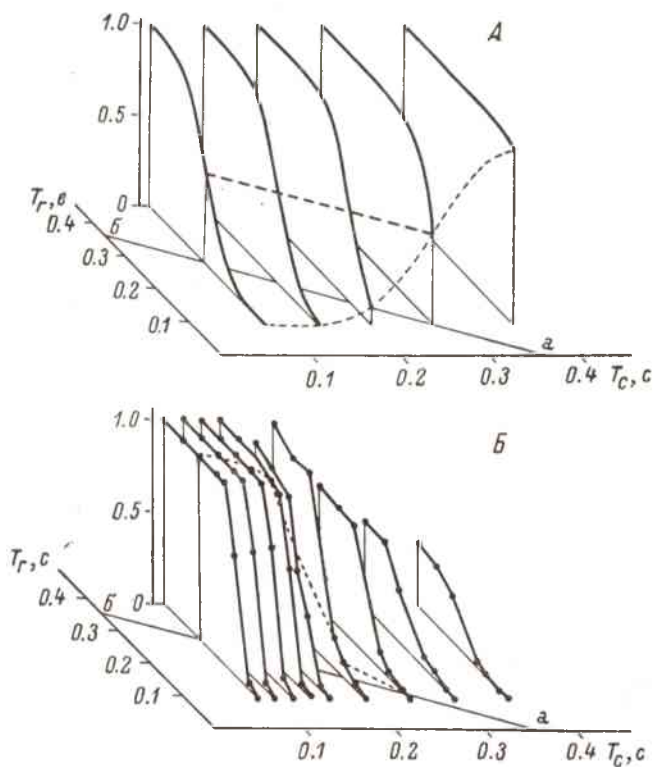


Рис. 5.4. Распределение гипотетических вероятностей ( $A$ ) и полученных экспериментально частот ( $B$ ) ответов «второй слог ударный» при восприятии ударения в синтетической последовательности [sasasa].

По оси абсцисс — длительность второго согласного; по оси ординат — длительность второго гласного; по вертикали — гипотетическая вероятность ( $A$ ) и экспериментальная частота ( $B$ ) ответов «второй слог ударный». Остальные обозначения см. в тексте.

«второй слог ударный». Причины этого явления будут частично рассмотрены в разделе 5.2.3. Здесь важно отметить, что увеличение длительности согласного понижает частоту ответов «второй слог ударный», хотя при этом возрастает длительность слога и следовало бы ожидать увеличения вероятности восприятия его ударным, если бы восприятие словесного ударения определялось длительностью всего слога.

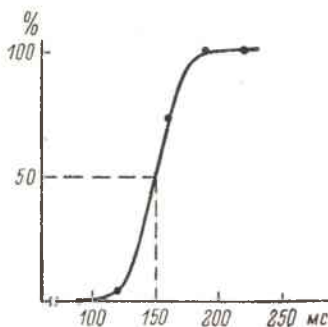
Таким образом, мы можем сделать вывод, что для восприятия словесного ударения существенна только длительность гласного в слоге.

Экспериментально полученные гистограммы ответов «второй слог ударный» (рис. 5.4, Б) дают возможность оценить величину дифференциального порога по длительности, когда сигналы (гласные в слогах) сравниваются по критерию «ударный—безударный». В первом приближении эта величина может служить мерой точности измерения длительности при восприятии человеком просодических признаков речи.

Независимость восприятия ударения от длительности согласного

Рис. 5.5. Частота ответов «второй слог ударный» при разных длительностях второго гласного в синтетической последовательности [sasasa].

По оси абсцисс — длительность второго гласного; по оси ординат — частота ответов. Усредненные данные для длительностей второго согласного 80, 100 и 120 мс.



(при малых длительностях последнего) дает возможность объединить экспериментальные результаты для длительностей согласного до 120 мс включительно и построить суммарную гистограмму ответов «второй слог ударный» (рис. 5.5). Аппроксимация этой гистограммы интегральной кривой нормального распределения вероятностей дает нам длительность второго гласного в последовательности [sasasa], при которой первые два слога воспринимаются равноударными,  $T_r = 150$  мс и среднеквадратическое отклонение  $\sigma = 17$  мс. Зная эти величины, мы можем определить величину дифференциального порога (ДП) и относительного дифференциального порога (ОДП):

$$\text{ДП} = 0.675\sigma = 11.5 \text{ мс}, \text{ ОДП} = \text{ДП}/T_r = 0.077 (7.7\%).$$

Полученная таким образом величина относительного дифференциального порога совпадает с имеющимися в литературе данными об относительных порогах по длительности, полученных в результате субъективного сравнения длительностей [173, 223, 267].

Таким образом, изложенные выше экспериментальные результаты позволяют сделать вывод, что длительность гласного в слоге является признаком словесного ударения, и при этом точность принятия решения о месте словесного ударения определяется в основном точностью субъективного измерения длительностей.



### 5.2.2. РЕЧЕВОЙ ТЕМП

Выше было показано, что для выделения элементов слова (гласных (согласных) или слогов) может быть использована длительность этих элементов. Оказывается, что аналогичный способ выделения используется и на более высоком уровне: для выделения слов и словосочетаний во фразе тоже используется длительность [76, 78, 132, 159, 205, 244]. Такое выделение называется фразовым (логическим) ударением, и при этом в той или иной степени изменяются длительности всех составляющих слово звуков. Можно предположить, что наиболее простым способом реализации такого изменения длительности при речеобразовании и, соответственно, наиболее простым признаком, используемым при восприятии, было бы изменение темпа произнесения слов.

Известно, что темп речи может быть использован и как один из признаков при опознавании индивидуальности диктора, а в ряде случаев его изменения могут служить показателем эмоционального и даже физического состояния диктора. Вероятно, узнавая на слух конкретного диктора по характерному для него темпу речи, мы пользуемся некоторой средней величиной, полученной в результате анализа достаточно большого отрезка речевого сигнала.

Напротив, для обнаружения логической выделенности слова во фразе необходимо установить, в каком темпе было произнесено именно это слово. Использование для этой цели средних значений темпа на некотором интервале предполагало бы членение речевого потока на слова до начала измерения. Но членение на слова возможно только на этапе лингвистического анализа (см. главу 2), а к этому моменту все измерения в речевом сигнале должны быть закончены. Видимо, преодолеть это противоречие можно при использовании параметра типа «мгновенного» темпа. Измеряя величину «мгновенного» темпа и записывая ее в оперативную память одновременно с фонемной информацией, мы получили бы возможность сравнивать величины «мгновенных» темпов по обе стороны словесной границы, обнаруживаемой при лингвистическом анализе, и тем самым — оценивать степень выделенности слов во фразе. Кроме того, не исключено, что в некоторых случаях резкое изменение величины «мгновенного» темпа само могло бы служить пограничным сигналом.

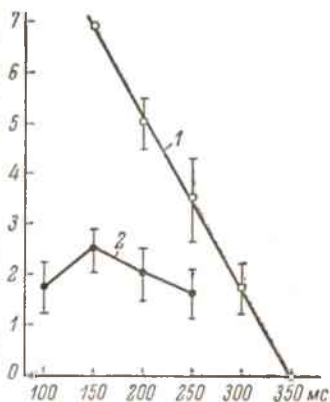
Попытаемся установить, что могло бы быть мерой «мгновенного» темпа в речевом сигнале. Обычно принято характеризовать темп речи на заданном отрезке средним числом некоторых элементов речевого потока, произнесенных в единицу времени. В зависимости от конкретных задач применяемые элементы могут быть различны: от фразы до фонемы [533]. Однако большинство исследователей пользуется для этой цели слогом. Анализ различных элементов речевого потока с точки зрения получения минимального коэффициента вариации их длительности при поддержании постоянного темпа и возможности измерения с их помощью

темпа сравнительно коротких отрезков речевого сигнала показал, что этим условиям наиболее полно удовлетворяет слог [148].

Исследователь, измеряющий темп речи по осциллограмме или по динамической спектрограмме, может в равной степени пользоваться для определения темпа и понятием частоты (число слогов в единицу времени), и понятием периода — длительности слога. Однако предположение, что при восприятии темпа человек использует процедуру подсчета числа слогов, произнесенных за эталонный интервал времени, представляется в высшей степени невероятным. К такому же выводу приводят и некоторые результаты психофизических экспериментов с периодическими сигналами [478]. Вместе с тем известно, что человек способен с достаточно высокой точностью измерять интервалы времени (длительности звуков, интервал между двумя щелчками или посылками тона или

Рис. 5.6. Зависимость субъективной величины темпа от длительности периода следования посылок гласного в стимуле (1) и от длительности самого гласного (2).

По оси абсцисс — длительность гласного или периода их следования; по оси ординат — субъективное «расстояние» между стандартным и тестовым стимулами (в среднеквадратических отклонениях).



шума) [150, 173, 223, 267]. Поэтому более естественным кажется предположение, что субъективная оценка темпа слушателями базируется на измерении периода, т. е. длительности слога.

Интересны в этом отношении результаты выполненного Венцовым эксперимента по психологическому шкалированию темпа последовательностей, образованных из четырех посылок синтетического гласного. Испытуемым предъявлялась таблица записанных на магнитную ленту стимулов, в которой либо изменялся период следования посылок гласного (от 150 до 350 мс) при постоянной длительности посылки гласного (100 мс), либо изменялась длительность посылки гласного (от 100 до 250 мс) при постоянном периоде следования этих посылок — 300 мс. Испытуемые должны были оценить темп каждого стимула в баллах от 1 до 7. Результаты эксперимента, приведенные на рис. 5.6, показывают, что субъективные оценки темпа таких синтетических последовательностей являются линейной функцией периода следования гласных в них, т. е. длительности интервала «гласный+пауза», и не зависят от длительности образующих стимул гласных при постоянном периоде их следования.

Так как интервал «гласный+пауза» соответствует в речевом сигнале слогу, а также принимая во внимание сказанное выше

о слоге как о том минимальном элементе речевого потока, которым определяется темп, мы можем принять, что мерой «мгновенного» темпа является период следования гласных, который в большинстве случаев есть не что иное, как длительность слога. Существенно отметить, что при таком определении «мгновенного» темпа мы в то же время не теряем возможности пользоваться и понятием среднего темпа, определяемого как средняя длительность слога на заданном отрезке речевого сигнала.

Если «мгновенный» темп определяется длительностью слога и в качестве меры точности его субъективной оценки принята величина относительного дифференциального порога, то следует ожидать, что точность измерения темпа будет равна точности субъективной оценки человеком длительности (относительный дифференциальный порог 5—15%). Хотя интуитивно очевидно, что слушатель в состоянии воспринять темп даже изолированно произнесенного двусложного слова, например *иду*, имеющего всего

Таблица 5.1

Относительный (в процентах) дифференциальный порог по периодичности. Данные разных авторов

| Период,<br>мс | Миллер<br>и Тейлор [385] | Поллак<br>[428] | Маубрей<br>и др. [309] | Крылов<br>[84] | Мичон<br>[384] |
|---------------|--------------------------|-----------------|------------------------|----------------|----------------|
| 3.12          | 13.1                     |                 |                        |                |                |
| 4.16          | 23.5                     |                 |                        |                |                |
| 5.4           | —                        | 2.46            |                        |                |                |
| 6.25          | 23.7                     |                 |                        |                |                |
| 8.3           | 21.4                     | 3.71            |                        |                |                |
| 12.5          | 19.4                     |                 |                        |                |                |
| 21.7          | —                        | 3.76            |                        |                |                |
| 25.0          | 15.6                     |                 |                        |                |                |
| 31.2          | —                        | 3.18            |                        |                |                |
| 50            | 13.8                     | —               | 0.8                    |                |                |
| 67            | —                        | —               | —                      | —              | 2.3            |
| 100           | —                        | 1.85            | 0.8                    | —              | 0.9            |
| 143           | —                        | —               | —                      | —              | 0.84           |
| 156           | —                        | 2.70            |                        |                |                |
| 200           | —                        | —               | 0.8                    | 4.9            | 0.86           |
| 250           | —                        | 3.10            | —                      | —              | 0.96           |
| 333           | —                        | —               | —                      | —              | 1.82           |
| 400           | —                        | 3.42            | 1.2                    |                |                |
| 500           | —                        | —               | —                      | 5.9            | 1.85           |
| 667           | —                        | —               | —                      | —              | 1.96           |
| 1000          | —                        | —               | 2.0                    | 8.0            | 1.85           |
| 1430          | —                        | —               | —                      | —              | 2.98           |
| 2000          | —                        | —               | —                      | 8.8            | 3.70           |
| 2500          | —                        | 8.05            | —                      |                |                |
| 2700          | —                        | —               | —                      | —              | 6.3            |

Примечание. Величину дифференциального порога, полученную Поллаком для периода 2500 мс, правильно было бы отнести к дифференциальному порогу по длительности, так как по условиям его эксперимента при этой величине периода испытуемым предъявлялся всего один интервал и о периоде в этом случае не могло быть и речи.

один период, обычно принято считать, что темп характеризует сравнительно длинную последовательность повторяющихся интервалов. Соответственно этому при оценке точности восприятия темпа в перечевых сигналах использовались последовательности неограниченной длины из периодически повторяющихся интервалов: определялся дифференциальный порог по периодичности. Результаты этих работ приведены в табл. 5.1 и на рис. 5.7. Видно, что в диапазоне характерных для речи темпов (максимальный темп — 12 слогов/с, период следования гласных 83 мс [73, 467, 533]; минимальный — 120 слогов/мин, период — 500 мс [4]) относительный дифференциальный порог составляет 1–6%. Можно предпо-

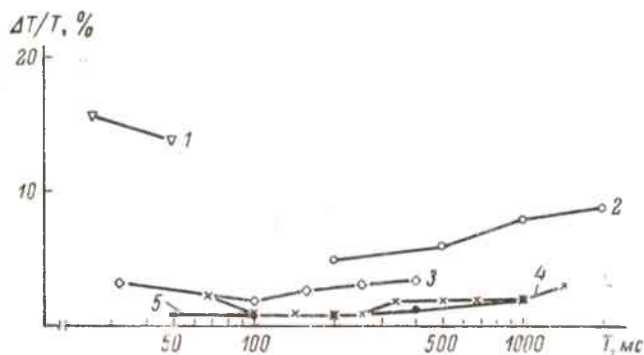


Рис. 5.7. Зависимость относительного дифференциального порога по периодичности от длительности периода.

По оси абсцисс — длительность периода; по оси ординат — относительный дифференциальный порог. Данные разных авторов: 1 — [383]; 2 — [84]; 3 — [428]; 4 — [284]; 5 — [309].

ложить, что в основе этих различий между дифференциальными порогами по длительности и периодичности лежит известный в психоакустике эффект накопления, заключающийся в повышении точности субъективных оценок, в частности высоты, с увеличением числа доступных системе отсчетов данного параметра. Так, восприятие высоты оказывается невозможным при числе периодов меньшем трех [428].

Если такое предположение правильно, точность субъективной оценки слушателем темпа последовательности гласных должна возрастать по мере увеличения в ней числа интервалов между гласными (количества гласных в последовательности). Именно такой результат (рис. 5.8) был получен в эксперименте, где от испытуемого требовалось уравнивать темп стандартной и тестовой последовательностей, изменяя период следования гласных в последней. Количество последовательных интервалов между гласными в стандартном и тестовом стимулах было одинаковым и изменялось от опыта к опыту от одного до трех. Измерения точности установки были проведены для 2 значений периода следования

гласных 200 и 300 мс при длительности самого гласного 100 мс. Каждый испытуемый проводил по 15 установок для каждого из периодов при данном числе периодов в стимуле. В качестве меры точности установки использовалась абсолютная величина разности длительностей стандартного и установленного испытуемым периодов следования гласных, выраженная в процентах от длительности стандартного периода.

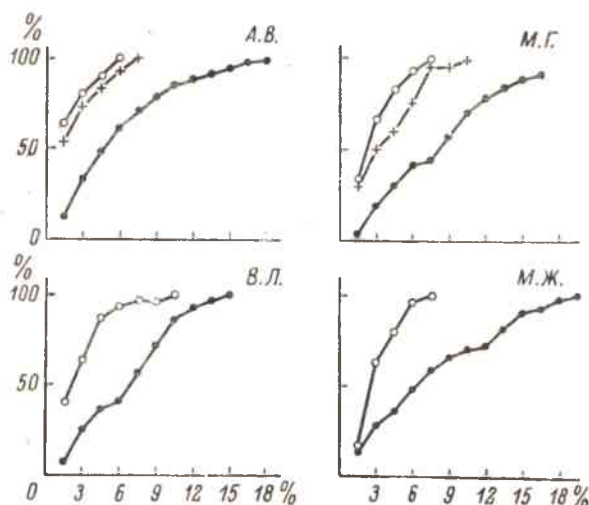


Рис. 5.8. Точность восприятия темпа при разном числе измеряемых периодов в сигнале.

По оси абсцисс — относительная ошибка уравнивания темпов; по оси ординат — относительное число случаев, когда ошибка не превышает данную. Точки — при одном периоде, крестики — при двух периодах, кружки — при трех. Данные четырех испытуемых.

Надо сказать, что полученная в таком эксперименте точность уравнивания темпов может быть использована в качестве меры точности восприятия темпа в речи с большими оговорками, так как неизвестен критерий, используемый испытуемым при принятии решения о равенстве темпов. Однако в первом приближении эти результаты все же дают представление о характере возрастания точности с увеличением числа доступных измерению периодов, особенно если данные разных испытуемых анализировать порознь.

Причины повышения точности субъективной оценки темпа при увеличении числа измеряемых интервалов еще не выяснены. Можно предположить, что это явление так или иначе связано с памятью, но на основании имеющихся данных нельзя сказать, производится ли при этом усреднение измеряемых величин или многомерное описание последовательности с одновременным использованием сведений обо всех измеренных интервалах. Тем не менее необходимо иметь в виду, что, говоря о точности субъек-



тивной оценки периодичности или темпа, необходимо указывать, при каком числе последовательных периодов в сигнале получена эта точность.

Высокая точность восприятия периодичности (низкие дифференциальные пороги) в психоакустических экспериментах с речевыми сигналами могла быть следствием использования в этих экспериментах стимулов с максимальными стабильными периодами следования посылок. Какова будет точность восприятия темпа последовательностей с нестационарным периодом следования гласных, свойственных естественной речи, сейчас сказать невозможно, так как исследование такого рода сигналов только начинается [94, 430]. Очевидно, что применительно к восприятию речевого темпа данные о дифференциальных порогах по периодичности могут быть использованы лишь в качестве нижнего предела возможных ошибок измерения темпа.

В заключение необходимо сказать, что возможность использования «мгновенного» темпа — длительности слога — в качестве одного из просодических признаков кажется вполне правдоподобной, хотя прямых доказательств этого пока нет. Надо отметить также, что в определении «мгновенного» темпа как периода следования гласных содержится некоторая неопределенность: так как при сегментации речевого потока выделяются и начало, и конец гласного (см. главу 3), этот период может измеряться от начала до начала последовательных гласных, что соответствует закрытому слогу, или от конца до конца (открытый слог). Какой из периодов используется при восприятии темпа естественной речи, неизвестно. Не исключено, однако, что в зависимости от конкретных условий человек может пользоваться любым из этих периодов.

### 5.2.3. ДЛИТЕЛЬНОСТЬ СОГЛАСНОГО КАК ПРИЗНАК СЛОВЕСНОЙ ГРАНИЦЫ

В предыдущих разделах мы установили, что длительность слога может быть использована в качестве меры темпа речи, а длительность гласного в слове определяет место словесного ударения. Естественно, возникает вопрос, не используется ли и длительность согласного в качестве просодического признака и, если используется, то какого рода информацию она несет. Разумеется, под длительностью согласного в данном случае мы понимаем длительность интервала от конца одного гласного до начала следующего, вне зависимости от того, каким числом согласных фонем будет описан этот интервал при его записи в оперативную память.

В работе Пикетта и Декера [421] впервые было показано, что длительность согласного может служить признаком словесной границы. При увеличении длительности смычки глухого взрывного согласного [p] в слове *topic* (в эксперименте изменялась длительность паузы между концом предшествующего гласного и взрывом



согласного путем вклеивания отрезков чистой магнитной ленты разной длины) испытуемые начинают воспринимать этот согласный как двойной (рис. 5.9). Так как в английском языке нет долгих и кратких согласных, это приводило к членению исходного слова *topic* на два слова — *top pick*, при условии, конечно, что такое членение не противоречило смыслу той фразы, в которой это слово было произнесено.

Можно представить себе, что при восприятии таких долгих согласных в оперативную память записывается либо два согласных, либо один согласный с указанием его длительности. Для выбора одной из этих возможностей пока нет достаточного количества

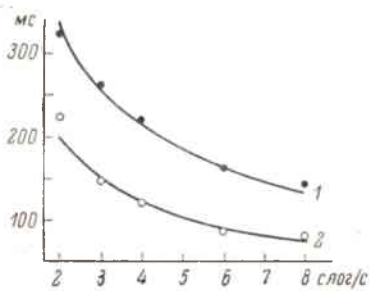
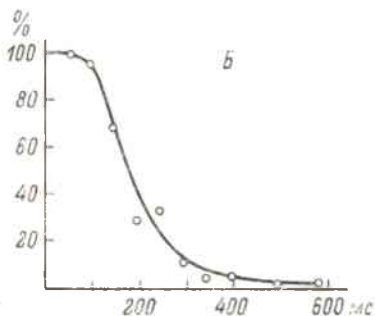
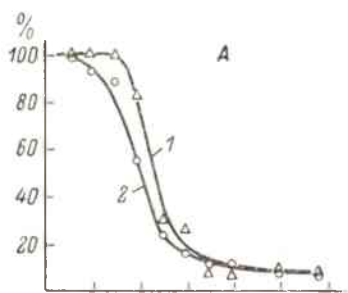


Рис. 5.9. Зависимость частоты ответов «одинарный согласный» от длительности согласного. По [421].

На А — при изменении длительности взрывного согласного [p] во фразе *He was the topic of the year*, произнесенной разными дикторами (кривые 1 и 2). На Б — то же для фразы *He was the topic for discussion*. По оси абсцисс — длительность фазы смычки взрывного согласного; по оси ординат — частота опознавания согласного как одинарного.

Рис. 5.10. Длительность фазы смычки взрывного согласного как функция темпа речи. По [421].

По оси абсцисс — темп речи; по оси ординат — длительность смычки. 1 — длительность смычки, при которой в 60% случаев испытуемые оценивали согласный как одинарный; 2 — длительность одинарного согласного в естественной речи.

данных, но ясно, что записи в память в обоих случаях должно предшествовать нормирование измеренной длительности согласного по темпу речи (рис. 5.10), хотя совершенно неясно, каким темпом надо пользоваться при нормировании («мгновенным» или средним), а если средним, то на каком интервале его надо измерять.

Появление словесной границы при увеличении длительности согласного наблюдалось и при восприятии носителями русского

языка синтетической звуковой последовательности [sasasa] [31]: увеличение длительности второго согласного приводит к тому, что в этой последовательности испытуемые слышат два ударных гласных одновременно.<sup>2</sup> По правилам русского языка это означает, что данная последовательность воспринимается состоящей из двух слов (рис. 5.11). Весьма примечательно то обстоятельство, что при постоянной длительности слога (этому условию соответствуют длительности гласного и согласного, представленные прямой  $ab$  в плоскости  $T_r - T_c$ ) именно длительность согласного является тем параметром, изменение которого приводит к восприятию в данной последо-

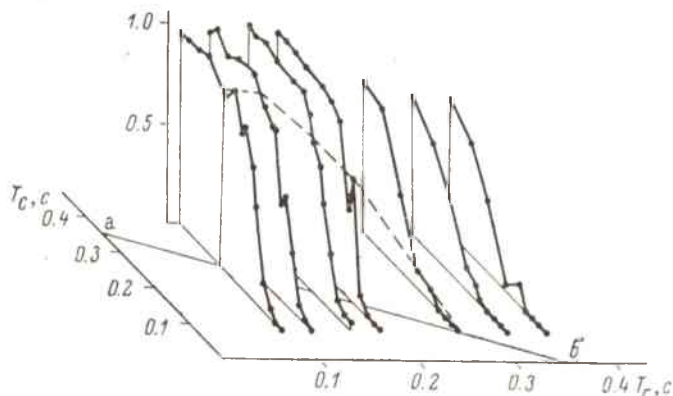


Рис. 5.11. Членение синтетической последовательности на два «слова» при разных длительностях согласного и гласного во втором слоге.

По оси абсцисс — длительность второго гласного; по оси ординат — длительность второго согласного; по вертикали — относительное число случаев, в которых последовательность воспринималась состоящей из двух «слов». Прямая  $ab$  описывается уравнением  $T_r + T_c = \text{const}$ .

вательности двух слов. Однако при больших длительностях гласного (больше 200 мс) увеличение длительности согласного уже не приводит к стопроцентному членению последовательности на два слова из-за появления ответов испытуемых «второй слог ударный». Причина этого заключается, по всей вероятности, в том, что для восприятия двух слов в такой последовательности параметры входящих в нее гласных (в данном случае их длительности) должны допускать возможность восприятия двух ударных гласных при соответствующей длительности согласного. Видимо, это условие перестает соблюдаться в полной мере, когда длительность второго гласного существенно превосходит длительности остальных гласных и он-то и воспринимается как главноударный. При этом, естественно, появляются ответы «второй слог ударный» вместо ответов «два слова».

<sup>2</sup> Параметры гласных и согласных в последовательности приведены в разделе 5.2.1.

Интересно отметить, что при существенном различии между русским и английским языком (в английском нет долгих и кратких согласных, а в русском они есть [<sup>113</sup>]) зависимости вероятности появления словесной границы от длительности согласного оказываются принципиально одинаковыми (ср. рис. 5.9 и рис. 5.12). Зависимости эти отличает, при всем их сходстве с обычной психометрической функцией, чрезмерно затянутый «хвост» при больших длительностях согласного. Это означает, что стопроцентное членение может не достигаться даже при очень больших длительностях согласного. Удовлетворительного объяснения этому явлению в настоящее время не существует.

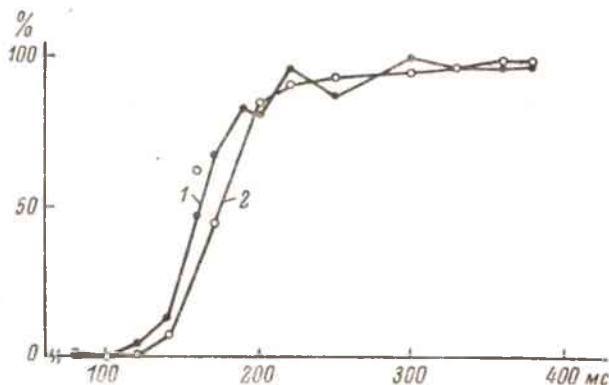


Рис. 5.12. Зависимость членения психодной последовательности [sasasa] на два «слова» от длительности согласного.

По оси абсцисс — длительность второго согласного; по оси ординат — относительное число членений. 1 — при длительности второго гласного 90 мс; 2 — при длительности второго гласного 190 мс.

В заключение необходимо сказать, что длительность согласного не единственный признак словесной границы: в ряде случаев признаком границы служит появление так называемых пограничных аллофонов, а также образование паузы между двумя следующими друг за другом звуками.

Аллофонные изменения на границах заключаются в том, что звуки, между которыми проходит граница, как правило, имеют более четкую формантную структуру, не оказывают друг на друга влияния типа коартикуляции и имеют большую длительность [218, 270, 342, 343]. Результаты эти получены путем анализа акустического сигнала и параметров звуков в тех случаях, когда слушатели четко устанавливали наличие в этом месте границы. Правил синтеза этих пограничных аллофонов в настоящее время не существует, и вопрос о том, насколько надо изменить длительность звука, чтобы он превратился в пограничный аллофон, совершенно не исследован.

Пауза чаще всего появляется в случаях, когда граница проходит между двумя фонетически различными звуками. При на-

личии паузы граница опознается слушателями в 95% случаев [20], и в некоторых случаях может оказаться, что она является единственным признаком границы [90]. Длительность такой паузы составляет 55—90 мс (средняя длительность звука речи) [20, 90]. Если пауза разделяет согласный и гласный, то гласный лишается формантных переходов, определяемых обычно качеством предшествующего согласного [20]. Относительно обратного влияния через паузу сведения противоречивые: согласный, предшествующий паузе, оказывается огубленным, когда за паузой следует губной гласный [20]; гласный же перед паузой не назализуется, если пауза разделяет гласный и носовой согласный [270]. Таким образом, оказывается, что звуки, разделенные паузой, могут не оказывать друг на друга влияний типа коартикуляции. Это явление сравнительно легко объяснить, если допустить, что пауза представляет собой некоторую остановку в непрерывной артикуляторной программе. С этой точки зрения паузу можно было бы определить как нулевой звук речи и в таком виде записывать ее в оперативную память.

И все же, хотя в арсенале и слушающего речь, и говорящего имеется достаточно средств для выражения словесной границы, бывают случаи, когда ни аудиторское прослушивание, ни аппаратный анализ омонимичных фраз не обнаруживают никаких пограничных сигналов, несмотря на старание диктора произнести их только с одним из возможных вариантов смысловой интерпретации [90, 342, 347].

### 5.3. ЧАСТОТА ОСНОВНОГО ТОНА И ПРОСОДИЧЕСКИЕ ПРИЗНАКИ РЕЧИ

Анализ литературных данных показывает, что частота основного тона является едва ли не самым универсальным параметром: любая просодическая информация может быть передана частотой основного тона (или ее изменениями). При этом оказывается, что частота основного тона еще и очень сильный параметр: изменение частоты — более весомый признак словесного ударения, чем длительность гласного [264].

Так как частота основного тона может использоваться для передачи любого вида просодической информации, возникает необходимость не только установить, какие именно параметры контура частоты основного тона высказывания используются в качестве просодических признаков, но и насколько специализированы эти признаки. Другими словами, существуют ли специальные признаки словесного ударения, фразовой интонации, логической выделенности и т. д. Если признаки специализированы, то необходимо установить правила их выделения из речевого сигнала, особенно в тех случаях, когда несколько признаков реализуется одновременно. Если признаки окажутся неспециализированными, то придется устанавливать те правила, по которым из

одного и того же признака может извлекаться разная информация в зависимости от условий его появления в сигнале.

Имеющиеся в литературе экспериментальные данные слишком немногочисленны, чтобы на их основании можно было сделать вывод о степени специализации признаков. К тому же оказывается, что данные, полученные на носителях разных языков, зачастую невозможно сравнивать из-за различий в системах просодических признаков. Так, общий характер контуров частоты основного тона ( $F_0$ ) оказывается принципиально разным в языках, различающихся наличием — отсутствием смысложнакобующего словесного ударения [512]. Установлено также, что контуры  $F_0$ , соответствующие вопросительной интонации, совершенно различны в разных языках и сравнение их практически невозможно [5].

#### 5.3.1. МЕЛОДИЧЕСКИЙ КОНТУР ФРАЗЫ. МЕЛОДИЧЕСКИЕ ИЗМЕНЕНИЯ ВО ФРАЗЕ И СИНТАКСИЧЕСКИЕ ОТНОШЕНИЯ В ЯЗЫКЕ

В этом разделе мы остановимся только на самом общем анализе контуров  $F_0$  высказываний и попытаемся установить, какого рода признаки таких контуров могли бы быть использованы в качестве просодических. К сожалению, имеющийся экспериментальный материал еще недостаточен для количественного анализа, поэтому мы рассмотрим только качественные характеристики контуров. Поскольку в этом разделе будут рассмотрены только те признаки, которые могут нести информацию для лингвистического анализа высказывания, анализу подлежат контуры  $F_0$  фраз, произнесенных нейтрально с точки зрения логическиситуативной.

Сравнение контуров частоты основного тона большого числа фраз русского языка с морфо-синтаксической структурой «подлежащее — сказуемое — дополнение» [82] позволило сделать вывод, что все они создаются по одной и той же схеме (рис. 5.13). Причем было убедительно показано, что наблюдаемые при произнесении этих фраз изменения частоты основного тона отражают именно фразовую интонацию.

Похожие результаты были получены при качественном анализе контуров частоты основного тона во фразах, произнесенных дикторами-американцами (рис. 5.14): в своей основе эти контуры могут быть описаны последовательностью элементов типа «hat pattern». Этот термин используется для обозначения элементов мелодического контура, имеющих форму трапеции, и был впервые введен для описания контуров  $F_0$  во фразах голландского языка [294].

Принято считать, что любое высказывание может быть расчленено на конечное число последовательных элементов, характеризующихся тем, что их дальнейшее членение приводит либо к нарушению единства смысла, либо к нарушению целостности инто-

наций. Такой единый с точки зрения смысла и интонации элемент называют синтагмой [72, 170].

Совершенно очевидно, что при таком определении интонационной единицы ее отражением в контуре частоты основного

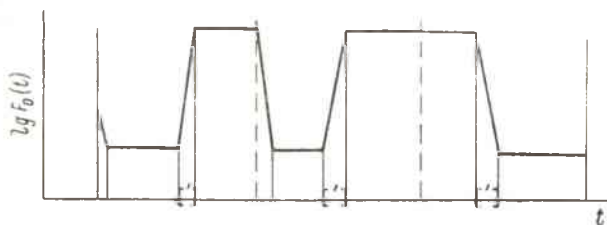


Рис. 5.13. Схематическое изображение интонационного контура высказывания, соответствующего предложению типа «подлежащее—сказуемое—дополнение». По [83].

По оси абсцисс — время; по оси ординат — логарифм частоты основного тона. Сплошные вертикальные линии — границы предложений, прерывистые — границы слов, квадратные скобки — ударные слоги.

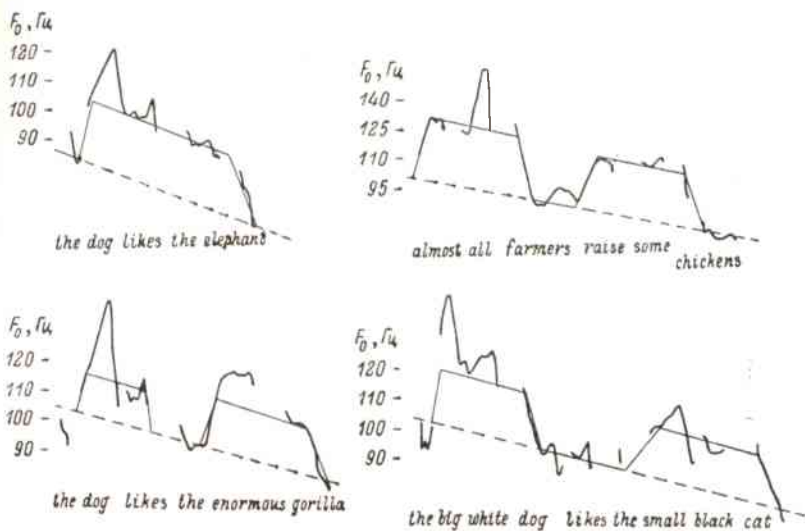


Рис. 5.14. Контуры изменений частоты основного тона и схематическое представление интонационных единиц для предложений с синтаксической структурой «группа подлежащего+группа сказуемого». По [373].

тона не может быть уровень частоты (ни для русского, ни для американского варианта английского языка). Принятие такого признака в качестве характеристики интонационной единицы привело бы к тому, что фраза на рис. 5.13 была бы расчленена



на пять интонационных единиц, соответствующих пяти разным уровням частоты, но при этом в одну интонационную единицу попали бы совершенно лишние смыслы части фразы: предупредная часть слова, заударная одного и предупредная другого и т. д. То же самое можно сказать и обо всех контурах на рис. 5.14.

Исследование на материале русского языка [83] показало, что каждое из двух характерных изменений в контуре частоты основного тона (рис. 5.13) — повышение с последующим понижением частоты — задается единой программой реализации фразовой интонации и, следовательно, может быть принято как показатель интонационной единицы. Из этого можно заключить, что количество интонационных единиц (синтагм) может быть определено



Рис. 5.15. Схематическое изображение экспериментально полученного мелодического контура группы «определение—подлежащее». По [83].

Обозначения те же, что на рис. 5.13.

по числу таких импульсных изменений в контуре частоты основного тона высказывания. Применительно к мелодическим контурам, полученным на материале английского языка, количество интонационных единиц следовало бы определять по числу реализованных в контуре «hat pattern» (рис. 5.14).

При этих условиях возникает необходимость количественной оценки указанных изменений частоты основного тона. Дело в том, что похожие импульсные изменения могут быть реализованы в контуре вне связи с фразовой интонацией. Например, во фразе с морфо-синтаксической структурой «определение—подлежащее—сказуемое—дополнение» изменения частоты в ударном слове определения связаны только с реализацией словесного ударения (рис. 5.15). Чтобы такое изменение в контуре не было принято как проявление дополнительной интонационной единицы, оно должно отличаться от изменений, свойственных последней. По имеющимся данным [83], выделение словесного ударения достигается изменением частоты основного тона на 20—50 Гц у разных испытуемых, но у тех же испытуемых при интонационном выделении единицы оно составляет 90—115 Гц. Следует, однако, иметь в виду, что еще совсем неясно, измеряет слуховая система величину приращения частоты или знак и величину скорости ее изменения. Поэтому необходимы дальнейшие кропотливые исследования с целью выяснения правил, по которым человек различает разного рода изменения в мелодическом контуре.

Итак, создается впечатление, что в результате анализа контура частоты основного тона может быть получена информация о количестве интонационных единиц (синтагм), реализованных

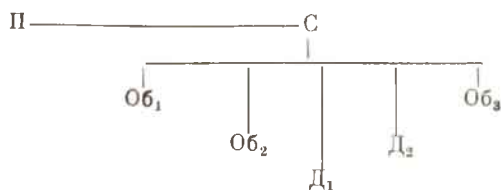
в данном высказывании. Однако мало знать количество интонационных единиц, необходимо получить еще и сигналы их начал и концов. Так как любое высказывание должно члениться на такого рода единицы без остатка, границы интонационной единицы не могут быть определены как границы соответствующего изменения частоты — импульса или «hat pattern», — ибо при таком определении часть высказывания неизбежно останется за пределами интонационных единиц (рис. 5.13, 5.14). Очевидно, что при таком членении определение (рис. 5.15) не попадет в одну интонационную единицу с подлежащим, что явно противоречит смыслу.

При внимательном рассмотрении обобщенного интонационного контура русских фраз (рис. 5.13) нетрудно заметить, что некое понижение частоты (неконечный отрицательный перепад частоты) приходится на начало слова. Это дает возможность высказать предположение, что интонационная единица — это часть мелодического контура и соответствующая ей часть фразы, заключенная между двумя отрицательными перепадами частоты, началом фразы и отрицательным перепадом или между отрицательным перепадом и концом фразы при обязательном наличии положительного перепада на указанном интервале. Легко убедиться, что при таком определении интонационной единицы высказывание с мелодическим контуром, приведенном на рис. 5.13, членится на соответствующие единицы без остатка. В пределах одной интонационной единицы окажутся теперь и определение с подлежащим (рис. 5.15).

Интересно, что в эту схему достаточно хорошо укладывается так называемая интонация незавершенности. Принято считать, что незавершенность характеризуется повышением частоты основного тона, хотя количественных данных о величинах изменений частоты при этом и о точном месте реализации этих изменений в слове найти не удалось. Тогда, с точки зрения предложенного выше определения интонационной единицы, как незавершенность должно восприниматься не просто само повышение частоты, а то обстоятельство, что за этим повышением нет завершающего интонационную единицу понижения частоты.

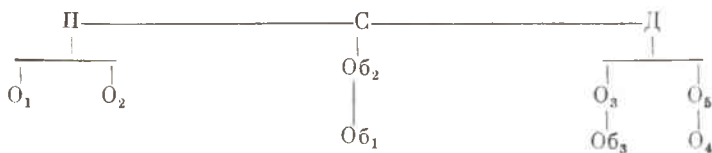
Экспериментальных данных пока еще слишком мало, чтобы можно было считать предложенное выше правило интонационного членения фразы универсальным. Если применить это правило к контурам частоты основного тона английских фраз (рис. 5.14), то окажется, что сам контур будет расчленен на единицы без остатка, но при этом в одной из них окажется предупредительная часть слова, а в другой — его ударная часть, так как оба изменения частоты, образующие «hat pattern» (и повышение, и понижение), по утверждению автора, приходится на ударные слоги. Не исключено, что и в русском языке это правило будет приложимо к ограниченному набору фраз, произносимых изолированно, так как именно в таких условиях были получены мелодические контуры, из которых оно выведено.

А



$(Об_1) + (Об_3) + (П) + (С) + (Д_1) + (Д_2 + Об_3) - 6 \text{ синтагм}$   
 $(Об_1) + (Об_3) + (П) + (С + Д_1) + (Д_2 + Об_3) - 5 \text{ синтагм}$   
 $(Об_1) + (Об_3) + (П) + (С) + (Д_1 + Д_2 + Об_3) - 5 \text{ синтагм}$

Б



$(О_1 + О_2 + П) + (Об_1 + Об_2) + (С + Об_3 + О_3 + Д) + (О_4 + О_5) - 4 \text{ синтагмы}$   
 $(О_1 + О_2 + П) + (Об_1 + Об_2) + (С) + (Об_3 + О_3 + Д) + (О_4 + О_5) - 5 \text{ синтагм}$   
 $(О_1 + О_2 + П) + (Об_1 + Об_2) + (С) + (Об_3 + О_3) + (Д) + (О_4 + О_5) - 6 \text{ синтагм}$

Рис. 5.16. Морфосинтаксическая структура и синтагматическое членение двух типов (А и Б) высказываний.

П — подлежащее, С — сказуемое, Д — дополнение, О — определение, Об — обстоятельство.

В заключение рассмотрим вопрос о характере связи интонационного членения фразы с ее синтаксической структурой. Одни авторы считают, что разграничительная функция мелодии обусловлена синтаксической структурой предложения, а выделяемые группы являются его синтаксическими составляющими [356, 507, 525]. Другие отрицают подобное влияние синтаксической структуры на мелодику и предполагают, что разграничительная функция интонации подчинена принципу, в соответствии с которым выделяемые ею группы едины прежде всего с точки зрения коммуникации и не обязательно должны совпадать с синтаксическими составляющими предложения [312].

Была сделана попытка выяснить, в какой степени членение на синтагмы зависит от синтаксических связей между элементами высказывания [82]. Рассматривались два наиболее интересных с точки зрения синтаксических связей типа высказываний. Один тип — это высказывания с максимально самостоятельными синтаксическими членами, каждый из которых зависит только от сказуемого. На рис. 5.16, А приведена морфо-синтаксическая структура одного из предложений этого типа: *Утром на месте мамы послала письмо Мане почтой*. На этом же рисунке приведены варианты объединения элементов высказывания в синтагмы, наблюдавшиеся в эксперименте (отнесенные дикторами в одну синтагму элементы взяты в скобки). Высказывания второго типа содержали синтаксические группы, элементы которых связаны

между собой теснее, чем с элементами соседних синтаксических групп. Морфо-синтаксическая структура одного из предложений этого типа (*Моя милая мама довольно медленно намыливала очень мелкую миску маленькой Мани*) и варианты членения соответствующего высказывания на синтагмы приведены на рис. 5.16, Б.

Из рис. 5.16 видно, что синтагматическое членение высказывания в значительной степени определяется его синтаксической структурой, хотя количество и состав синтагм в одном и том же высказывании может быть различным как при произнесении разными дикторами, так и при повторных произнесениях одного диктора.

Пробное прослушивание записей одной из фраз французского языка [511] в произнесении четырех дикторов обнаружило, что слушатели в состоянии уловить степень синтаксической связи входящих в нее слов и их оценки хорошо совпадают с данными анализа контуров частоты основного тона.

Носители русского языка на основании информации об изменениях частоты основного тона при произнесении двух предложений без паузы между ними в состоянии были определить, сколько было произнесено фраз и где находилась граница между ними [336].

Наконец, опыты, проведенные на испытуемых-голландцах [219], кажется, убеждают в существовании совершенно определенной связи между параметрами мелодического контура и синтаксической структурой фразы. Испытуемым предъявляли синтетическую последовательность слогов, напоминающую по звучанию производимую с закрытым ртом речь, с контуром изменения частоты основного тона, аналогичным приведенному на рис. 5.13. При постоянном числе слогов в каждой последовательности в ходе эксперимента изменяли номер слога, на который приходилось некое понижение частоты. Задача испытуемых заключалась в придумывании предложений, идентичных заданной последовательности по своему ритмико-мелодическому рисунку. Анализ придуманных предложений показал, что в 72% случаев слог, содержащий некое понижение частоты, соответствовал первому слогу очередной синтаксической составляющей предложения.

Таким образом, хотя результаты упомянутых выше исследований нельзя считать окончательными и проблема требует дальнейшего изучения, наличия связи мелодических изменений с синтаксической структурой высказывания отрицать невозможно.

### 5.3.2. ЧАСТОТА ОСНОВНОГО ТОНА И ПРИЗНАКИ РИТМИЧЕСКОЙ СТРУКТУРЫ ВЫСКАЗЫВАНИЯ. ВОСПРИЯТИЕ СЛОВЕСНОГО УДАРЕНИЯ

Исследуя влияние изменений частоты основного тона на восприятие словесного ударения, следует иметь в виду, что изменения частоты, связанные с реализацией фразовой интонации, тоже осуществляются на гласных ударного слога и, по

| Month          | Temperature (T4) in °C |
|----------------|------------------------|
| June 1968      | 125                    |
| July 1968      | 155                    |
| August 1968    | 125                    |
| September 1968 | 110                    |
| October 1968   | 125                    |
| November 1968  | 155                    |
| December 1968  | 125                    |
| January 1969   | 150                    |
| February 1969  | 125                    |
| March 1969     | 125                    |
| April 1969     | 155                    |
| May 1969       | 105                    |

По оси ординат — частота основного тона. В нижней части рисунка приведено относительное число случаев (в %), в которых каждый из гласных последовательности воспринимался ударным.

Возможны, однако, случаи, когда на одном из слов высказывания программой реализации фразовой интонации изменения

120



частоты основного тона не задаются (рис. 5.15). Тогда выделение ударного слога могло бы задаваться не фактом изменения частоты, а более высоким ее уровнем на ударном слоге. Указывая на это обстоятельство, Фрай [264] при исследовании роли изменений частоты основного тона в восприятии словесного ударения в английском языке стремился свести к минимуму возможное влияние фразовой интонации. Для изучения восприятия ударения в минимальных парах *subject—sùbject* он применял стационарные частоты на всех слогах, но на разных уровнях. При этом предполагалось, что в качестве ударного будет восприниматься слог с более высокой частотой основного тона. Именно этот результат он и получил в своих экспериментах, причем уже при разности частот в 5 Гц восприятие ударения оказывалось почти стопроцентным. Это хорошо согласуется с тем фактом, что дифференциальный порог по высоте достаточно мал — 0.5 Гц [323].

Приведенные выше данные могли бы означать, что при отсутствии изменений частоты основного тона, связанных с реализацией фразовой интонации, производится измерение абсолютных величин частоты и при определении места словесного ударения применяется правило «слог (гласный) с более высокой частотой основного тона — ударный».

Однако в опытах по восприятию испытуемыми-англичанами ударения в синтетических двусложных «словах» [sasa], [səsə], [sisi] [398] этот вывод не подтвердился. При стационарных, но различных по величине частотах основного тона гласных в этих «словах» восприятие ударения определяется не столько величиной и знаком разности этих частот, сколько тем, на каком из гласных частота изменяется в ходе эксперимента: при одной и той же разности частот первый гласный воспринимается ударным, когда его частота изменяется от стимула к стимулу, и наоборот (рис. 5.18, А).<sup>4</sup>

<sup>4</sup> Надо отметить, что в этом и подобных ему экспериментах, описанных ниже, вряд ли можно говорить о том, что испытуемые воспринимали словесное ударение как его понимают лингвисты и как это было в опытах Фрая [264]. В этих опытах каждая комбинация частот первого и второго гласных представлялась испытуемым по одному разу в каждом из трех «слов». Разрешались ответы «ударный первый гласный», «ударный второй» и «не знаю, не уверен, ударные оба». В естественной английской речи не может быть изолированного слова, ни один слог которого не имеет ударения или все слоги равноударны. Следовательно, способность испытуемых оценивать в ряде случаев оба гласных в синтетических «словах» как равноударные говорит о том, что они оценивали не столько ударение (в его лингвистическом смысле), сколько степень выделенности того или иного гласного. Естественно, выделенность могла бы восприниматься как при положительных, так и при отрицательных перепадах частоты основного тона, тогда как для восприятия ударения необходим, по Фраю [264], положительный перепад (заметим, однако, что в опытах Фрая отрицательные перепады не использовались). В этом случае результаты опытов с синтетическими двусложными словами дают возможность предположить, что испытуемые действительно пользовались оценками абсолютных величин частоты основного тона на каждом из гласных «слова», но при этом они не просто сравнивали два гласных в одном «слове», а принимали во внимание то обстоятельство, что частота основного тона 120 Гц



Набор стимулов, использованный в опытах с испытуемыми-англичанами, был предъявлен с той же инструкцией слушателям-полякам. В этих экспериментах [307] был получен совершенно аналогичный результат (рис. 5.18, Б).

Немного отчетливее проявляется правило «гласный с более высокой частотой основного тона — ударный» в результатах эксперимента (рис. 5.19) по восприятию ударения в синтетических двусложных словах носителями немецкого языка [288].

Использование в синтетических двусложных «словах» монотонно изме-

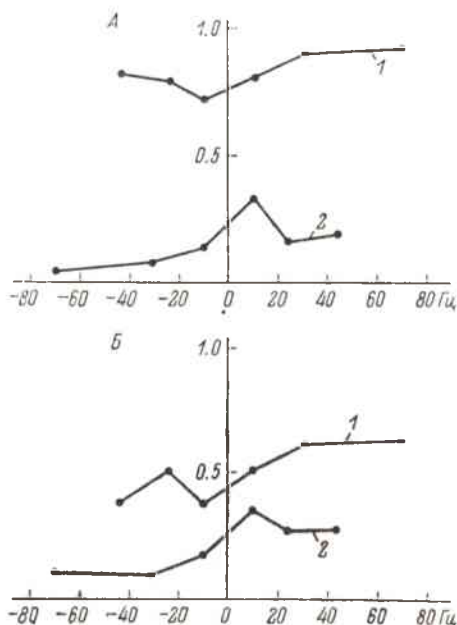


Рис. 5.18. Восприятие словесного ударения испытуемыми-англичанами (А) и испытуемыми-поляками (Б) при стационарных частотах основного тона гласных в синтетических двусложных «словах» [sasa], [sɔsɔ], [sisi].

По оси абсцисс — разность частот основного тона первого и второго гласных в слове; по оси ординат — относительное количество ответов «первый гласный ударный» при исключенных из общего числа неопределенных ответов. 1 — при частоте основного тона второго гласного 120 Гц; 2 — при частоте основного тона первого гласного 120 Гц.

Рис. 5.19. Зависимость восприятия испытуемыми-немцами ударения в синтетических двусложных «словах» [sasa], [sɔsɔ], [sisi] от соотношения стационарных частот основного тона входящих в них гласных.

По оси абсцисс — разность частот основного тона первого и второго гласных; по оси ординат — относительное количество ответов «не знаю, не уверен» (1) и «первый гласный ударный» (2). Кривая 2 получена при исключенных из общего числа неопределенных ответов.

няющихся на протяжении гласного частот основного тона приводит к тому, что испытуемые-англичане [398], испытуемые-поляки [307] и испытуемые-русские [123] довольно последова-

в пределах предложенной им для прослушивания таблицы стимулов является стандартом. Поэтому выделенными всегда воспринимались гласные, частота основного тона которых отличалась от стандарта.

Рис. 5.20. Восприятие словесного ударения испытуемыми-англичанами (А) и испытуемыми-поляками (Б) при монотонно изменяющейся частоте основного тона одного из гласных в синтетических двусложных «словах» [sasa], [soso], [sisi].

По оси абсцисс — величина изменения частоты на гласном (положительные величины соответствуют повышению частоты, отрицательные — понижению; частота основного тона стационарного гласного — 120 Гц); по оси ординат — относительное количество ответов «первый гласный ударный» при исключенных из общего числа неопределенных ответов.

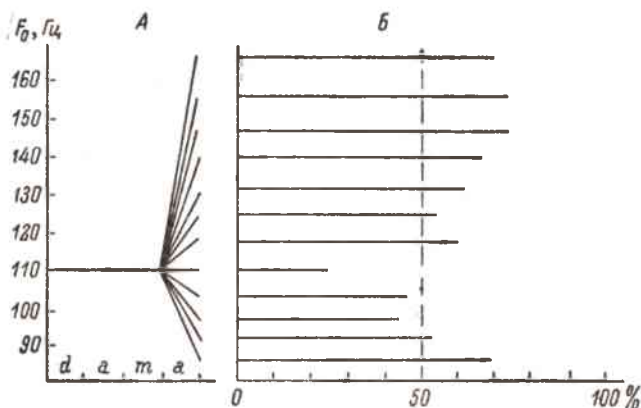
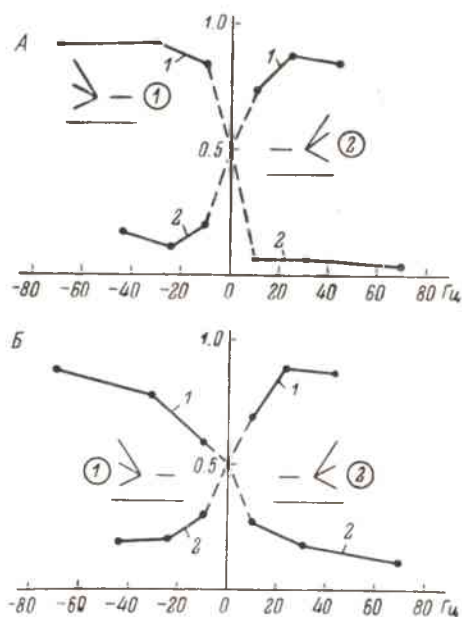


Рис. 5.21. Восприятие словесного ударения испытуемыми-русскими при монотонном изменении частоты основного тона на одном из гласных синтетического двусложного слова. По [123].

На А — контуры изменения частоты основного тона; по оси ординат — частота основного тона. На Б — соответствующие каждому из контуров относительные количества ответов «второй слог ударный».

тельно следуют правилу «слог с изменяющейся частотой основного тона — ударный» (рис. 5.20, 5.21). Данные, имеющиеся для голландского языка [315], также свидетельствуют, что ударение лучше воспринимается при изменяющейся частоте основного тона. Однако, как видно из приведенных на рисунках графиков, и в этом случае стопроцентное восприятие ударения не достигается.

Таким образом, имеющиеся в литературе данные позволяют допустить, что в речевом сигнале не существует специфических признаков словесного ударения, извлекаемых из мелодического контура. Решение об ударении в каждом конкретном случае может приниматься на основании разных признаков, только для этого необходимы специальные правила принятия решений и критерии оценки признаков. То обстоятельство, что во всех описанных выше экспериментах при изменяющейся частоте основного тона не удавалось добиться безошибочного восприятия испытуемыми ударения, вероятно, является следствием несоответствия использованных стимулов тем правилам, которыми испытуемые пользуются при восприятии естественной речи. Сами же исследователи отмечают, что ударение в изолированном двусложном слове лучше всего воспринимается русскими слушателями, когда заданные изменения частоты основного тона соответствуют естественной фразовой интонации [123].

### 5.3.3. ИСПОЛЬЗОВАНИЕ МЕЛОДИЧЕСКИХ ИЗМЕНЕНИЙ ДЛЯ ВЫДЕЛЕНИЯ СЛОВ ВО ФРАЗЕ. ЛОГИЧЕСКОЕ УДАРЕНИЕ

Логическое ударение используется для выделения в высказывании одного из слов (или словосочетания), несущего наиболее важную, с точки зрения говорящего, смысловую информацию. В зависимости от коммуникативной ситуации оно может быть утвердительным, вопросительным (вопрос без вопросительного слова), побудительным (просьба, приказ и т. д.), восклицательным. Вполне вероятно, что перечисленные типы не исчерпывают всего набора используемых в общении логических выделений и что в пределах каждого из типов возможны различные оттенки.

В русском языке произнесение логически выделенного слова сопровождается увеличением внутрилегочного давления, возрастанием частоты основного тона [3] и длительности логически выделенного слова [159]. В английском языке выделение слова приводит к увеличению длительности входящих в него слогов [76, 78] и повышению частоты основного тона [200]. То же самое наблюдается и в шведском языке [205, 244, 370].

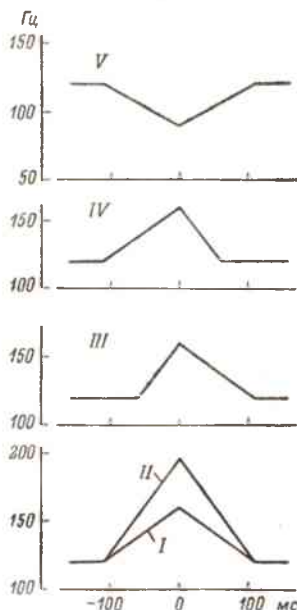
Логично допустить существование аналогии в способах реализации логического выделения (слова во фразе) и словесного ударения (слога в слове). Тогда применительно к контуру частоты основного тона выделение одного из слов должно приводить

к нейтрализации изменений частоты, связанных с нормальной фразовой интонацией, и к выделению на этом фоне одного из слов соответствующим изменением частоты.

Исследование зависимости восприятия такого рода выделенности от параметров мелодического контура было проведено Хаавелем [289, 290]. Для этой цели синтезировалась фраза *Мама мыла Маню* (ее временная структура приведена в нижней части рис. 5.23), в которой на стационарный основной тон накладывались изменения частоты нескольких типов (рис. 5.22). Создавался набор стимулов-фраз, в каждом из которых указанное изменение приходилось на разное место во фразе. Затем таблица стимулов предъявлялась группе auditors, задача которых заключалась в шкалировании степени выделенности слова (без указания типа выделенности) с записью результата так, чтобы сумма оценок на стимул-фразу составляла 3 балла (например, 0,0,3; 1,1,1; 2,0,1 и т. п.). Результаты эксперимента приведены на рис. 5.23 в виде зависимости средней оценки степени выделенности от положения

Рис. 5.22. Типы (I—V) контуров частоты основного тона, использованных в эксперименте по восприятию выделенности. По [289].

По оси абсцисс — длительность элементов контура, отсчитанная от положения экстремальной точки; по оси ординат — частота основного тона.



экстремальной точки каждого из контуров относительно фразы. Оказалось, что перемещение изменяющейся части мелодического контура вдоль фразы приводит к закономерному переносу оценок выделенности с одного слова на другое. При этом контуры разных типов оказываются эквивалентными в смысле создания эффекта выделенности, если начало увеличения частоты приходится на одно и то же место во фразе (рис. 5.24).

Судя по полученным результатам (рис. 5.23), для восприятия выделенности слова достаточно, чтобы соответствующее изменение частоты основного тона произошло в момент его произнесения, причем не требуется соблюдения каких-либо специальных условий согласования этих изменений со слоговой структурой слова. Однако при анализе частного случая вопросительного логического ударения (во фразе из одного слова) [70] было установлено, что соответствующее изменение частоты основного тона приходится на ударный гласный (рис. 5.25). К тому же аналогичная «при-

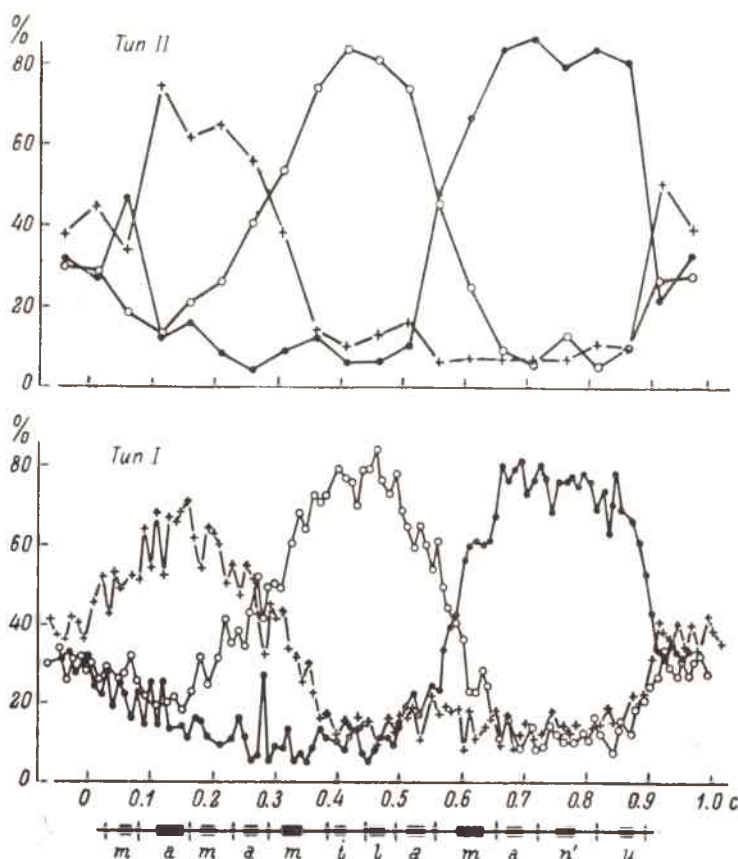


Рис. 5.23. Зависимость вероятности восприятия логического ударения на одном из слов фразы от положения в ней экстремума контура частоты основного тона. По [289].

По оси абсцисс — время; по оси ординат — относительная оценка выделенности. Крестики — ударение на первом слове, кружки — на втором, точки — на третьем. В нижней части рисунка указано положение фонемных границ (вертикальные линии) и стационарных участков фонем (черные прямоугольники).

вязка» изменений частоты к ударному слогу наблюдается и при реализации фразовой интонации.

Возможно, причина отмеченного противоречия заключается в особенностях использованного при исследовании восприятия выделенности [289] речевого материала (фраза из двусложных слов с ударением на первом слоге) и формы мелодического контура (участок изменения частоты слишком растянут и может охватывать слово целиком). Очевидно, в дальнейшем важно проверить, останется ли столь же несущественным согласование изменений частоты основного тона со слоговой структурой высказывания при использовании в нем более длинных слов с различным

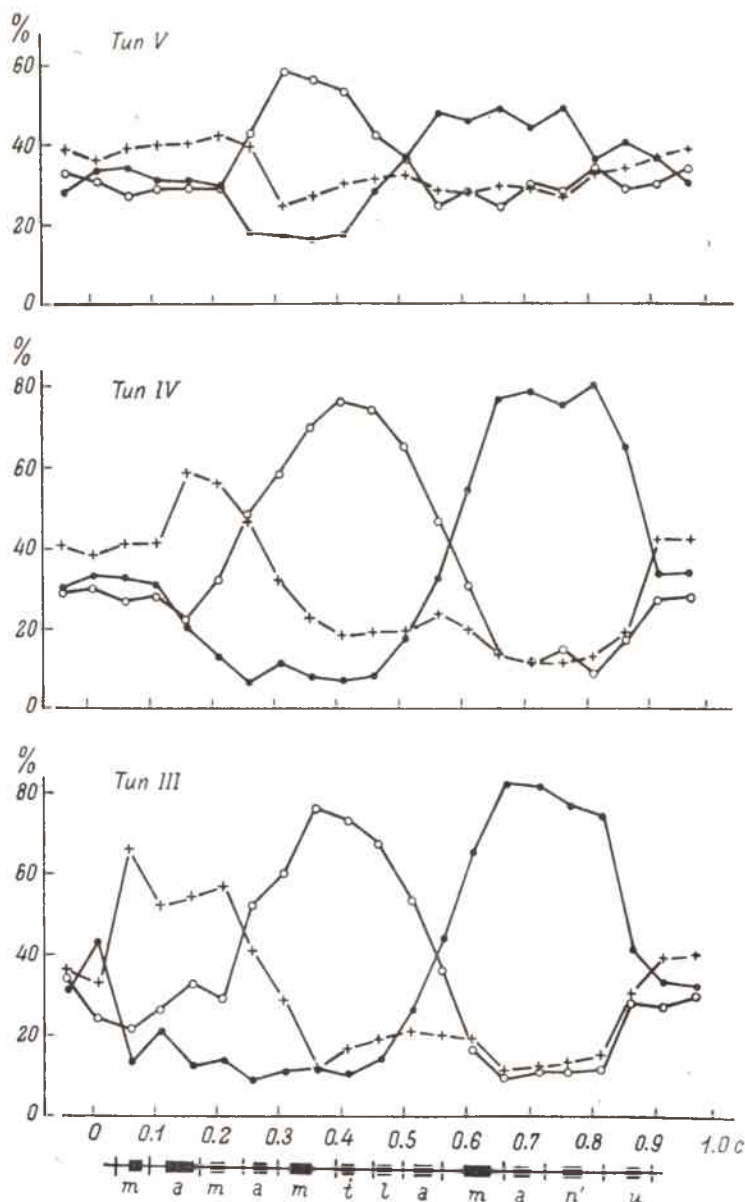


Рис. 5.23 (продолжение).

положением ударных гласных в слове и при разных формах мелодического контура. Интересно также сравнить полученные результаты с данными анализа мелодических контуров в естественной речи.



Это тем более интересно, что позволило бы получить сведения для решения вопроса об использовании при восприятии речи «анализа через синтез». Ведь если при произнесении выделенного слова во фразе изменения частоты основного тона действительно «синхронизированы» с произнесением ударного гласного, а для восприятия выделенности такой «синхронизации» не требуется, то это означало бы невозможность процедуры «анализа через синтез».

Затем Хаавелем была сделана попытка найти параметры мелодического контура, наиболее информативные в качестве признака выделенности слова при неизменном положении экстремальной точки контура во фразе *Мама мыла Маню*. Использовались

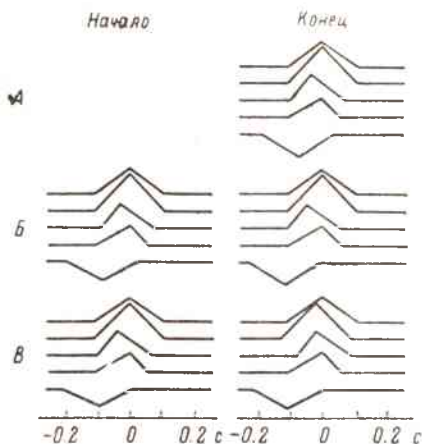


Рис. 5.24. Равноценные, в смысле вероятности восприятия логического ударения, положения контуров частоты основного тона разных типов. По [290].

По оси абсцисс — сдвиг во времени экстремальных точек контуров II—V типов относительно экстремальной точки контура I типа. Столбец «начало» соответствует увеличению вероятности восприятия ударения, столбец «конец» — уменьшению вероятности. А — ударение на первом слове, Б — ударение на втором слове, В — на третьем.

контуры частоты основного тона, аналогичные приведенным на рис. 5.22. По результатам предыдущего эксперимента были выбраны такие положения экстремальной точки контура во фразе, при которых оценка выделенности одного из слов была максимальной и не зависела от небольших смещений этой точки контура вдоль фразы. Это соответствовало моментам времени 200, 480 и 750 мс от начала фразы.

В качестве основных параметров контуров были выбраны: максимальная частота ( $F$ ), время ( $T_{\uparrow}$ ) и скорость ( $V_{\uparrow}$ ) повышения частоты, время ( $T_{\downarrow}$ ) и скорость ( $V_{\downarrow}$ ) понижения частоты. На основании этих параметров было составлено два полнофакторных комплекта стимулов. В первом использовались контуры с параметрами:  $F=155$  и  $175$  Гц,  $T_{\uparrow}=110$  и  $170$  мс,  $T_{\downarrow}=110$  и  $170$  мс. Во втором:  $F=150$  и  $175$  Гц,  $V_{\uparrow}=320$  и  $500$  Гц/с,  $V_{\downarrow}=320$  и  $500$  Гц/с. Общее количество стимулов с разными контурами основного тона на одно слово составляло 16. Оценка степени выделенности каждого из слов фразы производилась описанным выше способом. В качестве меры информативности параметра были использованы коэффициенты корреляции каждого из них с оценками выделенности слов. При обработке результатов были проанализированы

дополнительные параметры  $Q_T = T_n/T_v$  и  $Q_V = V_n/V_v$ ,  $S_v = \Delta F \times T_v/2$ ,  $S_n = \Delta F \times T_n/2$ ,  $S = S_v + S_n$ , где  $\Delta F = (F - 120)$  Гц. Результаты приведены в табл. 5.2.

Судя по данным таблицы, самым информативным параметром является площадь повышенной части контура, что может свидетельствовать об использовании человеком не максимальной частоты контура, а некоторой средней частоты. Оказывается также,

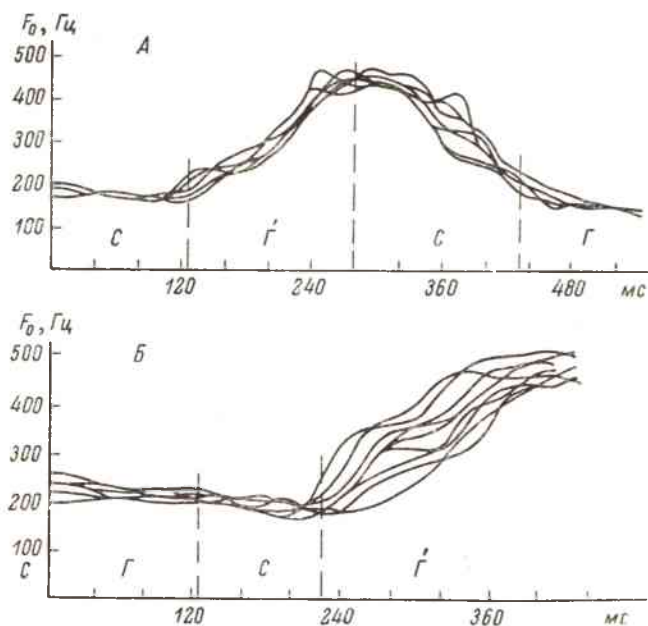


Рис. 5.25. Семейства кривых частоты основного тона, полученные при многократном произнесении вопросительных фраз одним диктором. По [70].

По оси абсцисс — время; по оси ординат — частота основного тона. На А — мелодический контур для структуры СГСГ (слово *баба*). На Б — то же для структуры СГСГ (слово *дала*). Вертикальными штриховыми линиями отмечены границы звуков.

что для восприятия выделенности большее значение имеет время, в течение которого частота изменяется с постоянной скоростью, чем сама скорость. Мало сказывается на восприятии выделенности асимметрия контура (параметры  $Q_T$  и  $Q_V$ ).

При исследовании зависимости восприятия выделенности слова во фразе [289] от положения в ней экстремальной точки мелодического контура (любого из пяти типов на рис. 5.22) оказалось, что при некоторых положениях этой точки во фразе одно из слов воспринимается выделенным, но не воспринимается вопросительным (рис. 5.26). Как видно из этого рисунка, для восприятия вопросительной выделенности контур изменения частоты основного тона должен занимать совершенно определенное положение

относительно фонемной последовательности со сравнительно малым допуском на изменение этого положения (зона, в которой

Таблица 5.2

Коэффициенты корреляции между параметрами контура основного тона и оценками выделенности для каждого слова во фразе

| Параметр   | I слово | II слово | III слово |
|------------|---------|----------|-----------|
| $\Delta F$ | 0.78    | 0.76     | 0.73      |
| $T^a$      | 0.71    | 0.69     | 0.60      |
| $T^b$      | 0.52    | 0.69     | 0.35      |
| $V^a$      | -0.13   | -0.14    | -0.10     |
| $V^b$      | -0.01   | -0.16    | -0.07     |
| $Q_T$      | -0.17   | -0.06    | -0.02     |
| $Q_V$      | 0.15    | -0.07    | -0.19     |
| $S^a$      | 0.92    | 0.87     | 0.83      |
| $S^b$      | 0.81    | 0.85     | 0.67      |
| $S^c$      | 0.92    | 0.92     | 0.80      |

вопросительность воспринимается с вероятностью, большей 50%. не превышает 160 мс). Ранее подобный эффект наблюдался при исследовании восприятия интонации в односложной синтетической фразе [6]. Смещение контура частоты основного тона, описываемого Гауссовой кривой, приводило к последовательной смене восприятия интонаций: утвердительной, восклицательной, вопросительной. Так как такие перемещения контура относительно фонемной последовательности, образующей фразу, неизбежно приводят к изменению набора признаков, описывающих контур на участке каждой из фонем, можно предположить, что «при-

вязка» определенных изменений в контуре к конкретным фонемам более важна для установления типа выделенности, чем выделенности как таковой.

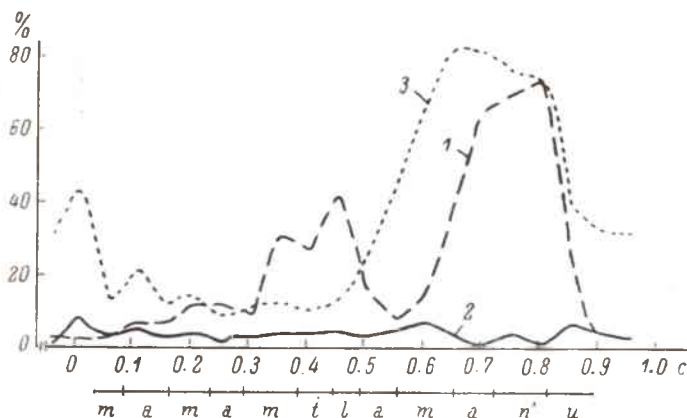


Рис. 5.26. Восприятие разных типов логической выделенности в зависимости от положения во фразе изменяющейся части контура частоты основного тона. По [289].

По оси абсцисс — время; по оси ординат — относительная оценка выделенности. 1 — вопросительная выделенность для контура типа III (см. рис. 5.22); 2 — то же для контура типа V; 3 — выделенность без указания ее конкретного типа при контуре частоты основного тона типа III. В нижней части рисунка приведено членение исследуемой фразы на звуки.

В связи с этим возникает вопрос о том, каковы же те признаки, которые определяют восприятие того или иного типа выделенности.

В русском языке вопросительная интонация характеризуется значительным повышением частоты основного тона на ударном гласном слова, к которому относится вопрос (рис. 5.25). При воспроизведении в синтетическом речевом сигнале такие изменения частоты довольно уверенно оцениваются слушателями как интонация вопроса [6, 122]. Оказывается, однако, что не всякое повышение частоты может восприниматься как интонация вопроса. Фраза *Мама мыла Маню* с мелодическим контуром типа V (рис. 5.22), расположенным в разных местах фразы [289], никогда не воспринимается как вопросительная (рис. 5.26). Особенность контура типа V заключается в том, что участку повышения частоты в нем предшествует участок понижения частоты. Из этого можно заключить, что для восприятия определенного типа логической выделенности существенным может оказаться не просто появление некоторого признака в мелодическом контуре, а закономерное чередование этих признаков. Характерно, что чередование в пределах одного гласного повышения и понижения частоты основного тона дает наилучшее восприятие восклицания [6].

Особый интерес представляет проблема синхронизации изменений в мелодическом контуре с элементами речевого потока. На рис. 5.25, А приведен контур частоты основного тона в вопросительной фразе СГСГ. Аналогичный вид имеет контур частоты основного тона фразы СГСГ, но произнесенной с утвердительной интонацией: подъем частоты происходит на безударном гласном, а понижение — на ударном [123]. Возникает вопрос, заключается ли разница в параметрах самих контуров (скоростях изменения частоты, разнице минимальной и максимальной частот и т. п.) или важно положение контура относительно ударного гласного (происходит ли повышение частоты на ударном или безударном гласном).

#### 5.4. ОБ ИЗМЕРЕНИИ ПРОСОДИЧЕСКИХ ПАРАМЕТРОВ В НЕПРЕРЫВНОМ РЕЧЕВОМ ПОТОКЕ И ПРАВИЛАХ ПРИНЯТИЯ РЕШЕНИЙ О НИХ

В начале настоящей главы мы говорили о том, что в оперативную память просодическая информация записывается одновременно с фонемной, т. е. в одни и те же ячейки памяти. Другими словами, это означает, что при записи в оперативную память просодическая информация должна быть представлена в дискретном виде.

Возникает вопрос, как должна быть организована система выделения и измерения просодических признаков, чтобы из физических параметров речевого сигнала, часть которых непрерывна по своей сущности, можно было получить последовательность

дискретных признаков для записи в оперативную память одновременно с фонемной информацией.

К такому результату может, очевидно, привести одна из двух измерительных процедур. В первой сегментация речевого сигнала предшествует выделению просодических параметров, и последние получаются в результате обработки сигнала на соответствующих участках речевого потока. Вторая предполагает, что выделение марок сегментации и непрерывное преобразование исходного речевого сигнала в описание по признакам производится в независимых каналах, а в момент появления сигнала окончания некоторого отрезка речевого потока производится считывание «показаний» измерителей и запись результатов в соответствующую ячейку оперативной памяти. Следует иметь в виду, однако, что обе процедуры должны обеспечивать возможность измерения не только локальных изменений в сигнале и мгновенных значений просодических параметров, но и таких «средних» характеристик, как темп и громкость речи данного диктора, средняя высота голоса за время произнесения некоторых слов или словосочетаний во фразе. Последняя может использоваться, например, для выделения авторских слов в прямой речи [132]. Попытаемся теперь рассмотреть, насколько успешно эти процедуры обеспечивают выделение и использование просодической информации.

Длительность любого элемента в речевом потоке может быть измерена только при условии, что предварительно были выделены метки начала и конца этого элемента. Это значит, что при использовании длительности звуков речи, а также и темпа, измеряемого как период следования гласных, для извлечения просодической информации сегментация в принципе должна предшествовать измерению.

Уже с громкостью дело обстоит сложнее. Суть в том, что и измерение громкости, и сегментация базируются, очевидно, на одних и тех же исходных сигналах — огибающей плотности импульсации в частотных каналах. Различие заключается в том, что для сегментации необходимо дифференцирование огибающей (или полосовая фильтрация в сравнительно высокочастотной области), а для измерения громкости — фильтр низких частот или полосовая фильтрация, но в значительно более низкочастотной области. Таким образом, системы сегментации и измерения громкости, работая с одним и тем же входным сигналом, оказываются в значительной степени независимыми. Это позволяет предположить, что для измерения громкости обе процедуры могут быть использованы с одинаковой вероятностью: на фильтр нижних частот может подаваться сегментированный сигнал и марки сегментации могут использоваться для считывания результатов измерения. Однако когда мы говорим о двух процедурах измерения громкости и пытаемся анализировать возможные различия в конечных результатах, мы имеем в виду то обстоятельство, что любая модель измерения громкости в слуховой системе человека



имеет сглаживающее звено (фильтр низких частот). Так как интенсивность гласных и согласных в речевом потоке различна, то в зависимости от величины постоянной времени такого фильтра громкость звуков может оказаться разной при использовании разных измерительных процедур: в одном случае на фильтр будет подаваться непрерывный сигнал, в другом — он будет работать в переходном режиме.

Изменения высоты речевого сигнала, определяющие мелодический компонент интонации, создаются изменением частоты основного тона, управление которым в процессе речеобразования осуществляется непрерывно [3, 29]. Результирующий контур частоты, как правило, представляет собой достаточно сложную кривую, для описания которой применяют ответ сложного апериодического фильтра на ступенчатое возмущение [268, 414]. Если бы при восприятии просодической информации человек пользовался процедурой «анализа через синтез», а любые мелодические изменения можно было описать реакцией одного и того же фильтра, то достаточно было бы указать, на каком элементе в речевом потоке произошло ступенчатое изменение управляющего сигнала и какова была величина этого изменения. Однако анализ экспериментально полученных контуров частоты основного тона показывает, что параметры фильтров должны зависеть от темпа речи и могут изменяться от произнесения к произнесению, а точность аппроксимации, используемая в качестве критерия правильности априорических построений, практически не зависит от формы аппроксимирующей функции и может быть достаточно высокой даже при использовании линейно изменяющихся функций [30]. Оказалось также, что при задании на контуре частоты основного тона ряда «особых» точек восприятие мелодии слушателем почти не зависит от вида траектории изменений частоты между этими точками [291]. Меловероятно в этих условиях, чтобы человек пользовался методом «анализ через синтез» при восприятии мелодических изменений в речи. Скорее он пользуется результатом измерения некоторых параметров в контуре изменяющейся частоты основного тона. Что это за параметры и как они могут измеряться в речевом сигнале, будет показано в главе 12. Здесь мы попытаемся рассмотреть только вопрос о том, на каких отрезках в речевом потоке производится их измерение.

В предыдущих разделах главы говорилось о том, что существенные с точки зрения восприятия просодической информации изменения частоты основного тона происходят в основном на гласных (как правило, ударных). Возможен вывод, что соответствующие высотные изменения оцениваются слуховой системой человека именно на отрезках гласных. Так как слуховая система выделяет марки начал и концов гласных (см. главу 3), такой результат, казалось бы, проще всего получить при условии, что сегментация предшествует измерению. Изменения же частоты основного тона на отрезке (в данном случае на гласном) могут



описываться в слуховой системе двумя отсчетами: в начале и конце отрезка (см. главу 12). Совершенно очевидно, что такое описание обеспечивает независимость восприятия данного контура от траектории изменения частоты между двумя точками отсчета.

Однако надо иметь в виду, что способность описывать изменения частоты начальным и конечным ее значениями на отрезке была показана с использованием сигналов, в которых начало и конец изменения совпадали с началом и концом самого сигнала. С другой стороны, если сегментация мелодического контура предшествует измерениям частоты основного тона, а описание контура производится в терминах начальных и конечных значений частоты на отрезках, то может быть затруднено или просто окажется невозможным доказанное экспериментально использование разности частот на стыке согласного и гласного для различения [b] и [m] <sup>[151]</sup>, а изменений на начальном участке гласного для различения [p] и [b] <sup>[292]</sup>. В первом случае препятствием может стать неточность сегментации, во втором — отсутствие членения гласного на более мелкие элементы. Таким образом, описание контура частоты основного тона только начальным и конечным отсчетом частоты на отрезке гласного может оказаться недостаточным для восприятия интонации и совершенно недостаточным с точки зрения получения дополнительных признаков для фонемной идентификации.

Поэтому надо предположить либо существование двух независимых систем измерения частоты основного тона (одной для выделения просодических параметров, другой — фонематических), либо совершенно иную структуру измерителя, в которой выделение и измерение признаков мелодического контура производится непрерывно и независимо от сегментации потока, а считывание и запись в оперативную память результатов измерения — в моменты появления сигналов сегментации. В пользу такой структуры измерителя говорит способность слуховой системы выделять в контуре изменения частоты основного тона некоторые особые точки: начало изменения частоты, знак (направление) изменения, экстремумы (изменение направления) и т. д. <sup>[291]</sup>. В результате контур оказывается как бы расчлененным особыми точками на последовательность отрезков и может быть описан набором признаков, каждый из которых характеризует один из отрезков. При появлении марок сегментации речевого потока каждому элементу, выделяемому ими, приписывается тот набор параметров мелодического контура (в порядке их появления), который был выделен измерителем за время между появлением марки начала и конца данного элемента речевого потока.

Обратимся теперь к вопросу о том, в каком виде могла бы быть записана в оперативную память просодическая информация: в виде ли окончательных решений или в виде набора просодических признаков, обеспечивающих принятие того или иного решения. В главе 2 речь шла о том, что для успешного лингвистиче-

ского анализа высказывания необходимо использовать просодическую информацию уже при синтаксическом и морфологическом анализе. На практике это означает, что она должна быть записана в оперативную память одновременно с последовательностью фонемных вариантов и, по возможности, в виде окончательных решений, т. е. в виде указаний места словесного и логического ударений, границ слов и синтагм и т. п. Однако после перехода от признаков к решениям система последующего анализа теряет возможность в дальнейшем обращаться к признакам и начинает работать только с решениями. При ошибке в системе измерения или недостаточно четко выраженных признаках возникает опасность принятия ошибочного решения, а снова обратиться к признакам для его коррекции уже невозможно. Поэтому желательно в процессе анализа как можно дольше работать с признаками и не переходить к решениям. Это чревато увеличением числа анализируемых вариантов, перебором большого количества гипотез. Необходимо поэтому стремиться к некоторой «золотой середине».

Очень может быть, что тот или иной ответ на вопрос о способе и форме записи просодической информации в оперативную память в немалой степени зависит от характера процедур принятия решений и от структуры оперативной памяти.

Относительно структуры оперативной памяти можно предположить два варианта ее построения: пассивный, при котором оперативная память является только хранилищем информации, и активный, допускающий некоторые логические операции с записанной информацией и перезапись результатов. При первом варианте мы вынуждены заносить в память готовые решения, при втором — записывать признаки, а решения принимать уже на основе записанных в память признаков. Ясно, что активная структура оперативной памяти наилучшим образом отвечала бы требованию сохранять как можно дольше в процессе восприятия речи признаки и не переходить к решениям.

В первом приближении правила принятия решений можно отнести к двум уровням: уровню выделения просодических признаков из параметров речевого сигнала, измеренных слухом, и уровню лингвистической интерпретации этих признаков.

Правила первого уровня должны, по-видимому, работать с параметрами речевого сигнала на отрезке, не превышающем длительности двух—трех слогов, и обеспечивать не только выделение признака из сложного параметра, но и объединение данных о нескольких параметрах. Можно предположить, что характер решающих правил будет изменяться в зависимости от величины конкретного параметра. Так, из данных анализа длительностей ударных и безударных гласных в русском языке [17] следует, что независимо от темпа речи существуют неперекрывающиеся диапазоны длительностей, характерные только для ударных или только для безударных гласных. Это означает, что при величинах параметров, больших или меньших некоторых критических значений,

возможны правила, работающие с абсолютными значениями параметров. Но как только значение параметров попадает в область, где возможны взаимоисключающие решения, необходимо использовать правила принятия решений, в которых учитывались бы конкретные условия произнесения. Здесь надо иметь в виду две возможности: сравнение с эталонами, т. е. с некоторыми усредненными параметрами сигнала, и сравнение с ближайшими соседями в речевом потоке. Сравнение с эталонами требует предварительного нормирования измеренных для данного отрезка параметров по средним величинам громкости, высоты и темпа, и в этом недостаток этой процедуры. Сравнение с ближайшими соседями нормирования не требует (в пределах семи слогов, записываемых в оперативную память, среднюю громкость, высоту и средний темп можно считать постоянными) и сравнительно просто реализуется при активной структуре оперативной памяти.

В обоих случаях имеет место сравнение двух величин и решение представляет собой ответ на вопрос, равны ли сравнительные величины, а если нет, то какая из них больше (меньше). Система сравнения не может работать без некоторой зоны нечувствительности, в пределах которой сигналы принимаются равными. В общем случае зона нечувствительности должна быть принята несимметричной, т. е. ее ширина не одинакова по обе стороны от точки равенства. Показано, что, используя предположение о существовании зоны нечувствительности и о нормальном законе распределения ошибок, удастся удовлетворительно интерпретировать результаты психоакустических экспериментов [474].

В обычном психоакустическом эксперименте, когда от испытуемого требуется оценивать сигналы по принципу «короче—длиннее», «тише—громче», «ниже—выше» (если речь идет о восприятии частоты основного тона), на окончательный результат могут влиять зашумленность исходного акустического сигнала, «шумы» в измерителе, флуктуации величины зоны нечувствительности, эффект забывания параметров первого по порядку сигнала и т. д. В результате физически равные сигналы не всегда оказываются субъективно одинаковыми.

Ранее было показано, что в естественной речи просодическая информация передается одновременно несколькими параметрами. Следовательно, решающие правила должны «уметь» объединять эти параметры, учитывая при этом степень важности каждого из них, его «вес». Известно, что человек способен оценивать не только сам факт различия двух сигналов по данному признаку, но и степень своей уверенности в этом [241, 522]. Эта оценка могла бы быть использована в качестве меры «веса» данного признака. В главе 4 уже говорилось, что такая система дополнительных оценок должна существовать в устройствах фоновой идентификации.

Правила второго уровня базируются на правилах и статистике языка, носителем которого является слушатель, понимании им

конкретной ситуации общения и используют в качестве входных сигналов признаки, полученные на первом уровне. В этой связи становится понятной отмеченная в начале главы неспособность испытуемых оценивать просодические признаки как таковые при восприятии осмысленного речевого сообщения: при переходе к решениям лингвистического и коммуникативного уровня происходит «забывание» исходных просодических признаков.

На этом этапе существенное влияние может оказывать отступление от естественности, привычности набора признаков, извлеченных из речевого сигнала. Задание части признаков в стимуле вне их естественного диапазона может привести к иному восприятию остальных. Так, при исключении свойственных интонации завершенности изменений интенсивности в конце фразы один и тот же контур частоты основного тона опознается как завершающий значительно хуже (на 20 и более процентов).<sup>[122]</sup>

Следовательно, использование при исследовании процесса восприятия просодической информации ответов испытуемых, основанных на применении решающих правил высокого уровня, чревато опасностью получить ошибочное представление о структуре просодических признаков, записываемых в оперативную память одновременно с фоновой информацией.

## Глава 6

### ВРЕМЕННАЯ ОРГАНИЗАЦИЯ ПРОЦЕССА ФОНЕТИЧЕСКОЙ ИНТЕРПРЕТАЦИИ

Данные, приведенные в предыдущих главах, позволяют в первом приближении определить, какое описание речевого сигнала должно содержаться в оперативной памяти после того, как фонетическая интерпретация окончена, а морфологический анализ еще не произведен. Это описание можно представить таблицей (6.1). Вертикальная организация таблицы обеспечивает возможность записи как фонемной, так и просодической информации и отражает привязку фонемной информации к двум классам отрезков сигнала: гласным (Г) и негласным (Г). Дополнительное разделение столбца Г дает возможность сохранить информацию о последовательности согласных (С) в сочетаниях согласных — «кластерах» — в условиях, когда отдельные согласные распознаны только частично — установлены их значения лишь по некоторым из психологических (дифференциальных) признаков. Подчеркнем, что признаки гласных и согласных в таблице — это признаки «на выходе» фонемного классификатора, а не свойства акустического сигнала, используемые для классификации (см. главу 4). Сведения, записанные в табл. 6.1, мы будем обозначать как фонетическое описание сигнала.

Процесс, обеспечивающий преобразование слухового изображения воспринимаемого речевого сигнала в фонетическое описание элементов, обозначается как фонетическая интерпретация или фонетический анализ. Модель, реализующая этот процесс, называется «фонетическим интерпретатором», «фонетическим процессором», «фонетическим анализатором».

Алгоритм работы фонетического интерпретатора пока остается неизвестным, о чем, в частности, свидетельствует отсутствие серьезных успехов в автоматическом распознавании речи. Вместе с тем некоторые представления о возможных принципах его действия сформировались за последние 2—3 года благодаря довольно интенсивным психологическим исследованиям.

Таблица 6.1

Описание речевого сигнала на «входе» блока морфологического анализа

| Признаки<br>согласных     | Г |   |   |   | Г | Г |   |   |   | Г |
|---------------------------|---|---|---|---|---|---|---|---|---|---|
|                           | с | с | с | с |   | с | с | с | с |   |
|                           |   |   |   |   |   |   |   |   |   |   |
| Признаки<br>гласных       | + |   |   |   |   | + |   |   |   |   |
| Просодические<br>признаки |   |   |   |   |   |   |   |   |   |   |

Приводимое ниже изложение существа этих представлений основывается преимущественно на статьях Массаро [376] и Уоррена [521]. Мы вынуждены, однако, признаться, что, так как эти авторы используют каждый свою собственную и довольно неопределенную терминологию, возможно, мы упустили некоторые нюансы и наше изложение грешит известным упрощением.

Принимается, что модель работает в реальном времени со сравнительно небольшой задержкой интерпретации относительно сигнала. Модель обладает «сенсорной» памятью, т. е. использует в качестве входного сигнала изображение некоторого отрезка речевого стимула. Это изображение можно представить себе, например, как картинку, соответствующую или отрезку динамической спектрограммы, или отрезку записи параметров сигнала (частоты формант, основная частота и т. д.) при их параллельной регистрации. Модель одновременно «видит» всю картину в целом и преобразует ее в набор фонетических сведений, записываемых в определенную группу соседних столбцов табл. 6.1. Такая группа может, например, состоять из столбцов ГГ или ГГ. Преобразование осуществляется или путем идентификации картинок или путем их анализа. В случае идентификации предполагается, что имеется набор типовых картинок. Каждому номеру в этом наборе



сопоставлены, с одной стороны, правила распознавания картинки, отнесения ее к этому номеру, а с другой стороны — группа фонетических сведений, играющих роль развернутого названия картинки. В случае анализа предполагается набор правил, связывающих определенные свойства картинки с определенными фонетическими сведениями. При этом оказывается возможным частичное распознавание картинок — часть сведений получена, а часть осталась неизвестной.

Наиболее существенное допущение, используемое при обоих указанных вариантах (идентификация и анализ), состоит в том, что модель работает последовательными циклами. Закончив обработку одной картинки и записав ее название, программа возвращается в начальное состояние и начинает обрабатывать следующую картинку. Таким образом, после каждого цикла производится заполнение нескольких (а не одного!) столбцов табл. 6.1, новый цикл соответствует переходу к новой группе столбцов.

Легко заметить, что обработка, организованная таким образом, должна использовать два ряда правил. Кроме правил, описывающих связь между признаками картинки и ее названием (фонетическими сведениями), алгоритм должен еще включать правила, обеспечивающие окончание цикла и переход к следующему циклу. Информация, используемая этими правилами, также должна извлекаться из речевого сигнала. Иначе говоря, отрезок речевого потока, соответствующий одному циклу работы модели, должен обладать некоторыми формальными свойствами, т. е. быть в каком-то смысле элементом потока.

Имея в виду описанный принцип обработки, сведения, записанные в табл. 6.1, можно рассматривать как последовательность названий картинок, где каждое название в свою очередь может включать последовательность фонем. Число названий, записанных в таблице, равно числу полных циклов работы модели и числу элементов в обработанном сигнале. Внутри одного названия связь между числом фонем, входящих в название, и числом мелких кусочков, на которые можно было бы разделить одну картинку, может быть сложной. Признаки одного кусочка картинки могут содержать информацию относительно нескольких фонем.

Признаком одной фонемы могут быть определенные отношения между соседними кусочками картинки. Другими словами, попытки механически разложить элемент потока на такую последовательность более мелких составных частей, чтобы число этих частей равнялось числу фонем, входящих в название элемента, могут быть в принципе безнадёжными.

Приведенные выше модельные представления дают возможность сгруппировать имеющиеся экспериментальные данные вокруг двух основных вопросов: 1) элемент потока и возможные методы его выделения при восприятии и 2) описание отрезка речевого сигнала на входе фонетического интерпретатора — картинка или последовательность акустических событий.

## 6.1 СЛОГ КАК ЭЛЕМЕНТ ПОТОКА

Вероятность того, что именно слог должен соответствовать одному полному циклу работы фонетического интерпретатора, представляется весьма высокой из чисто априорных соображений. Два соседних слога в потоке могут иметь совпадающие акустические характеристики, и для фонетической интерпретации этих слогов необходимо повторно применить одни и те же правила. Слог является полным элементом с точки зрения описания ритмического рисунка слова, каждому слогу могут быть приписаны такие характеристики, как наличие / отсутствие ударения и длительность слога (см. главу 5).

Повторяемости элементов внутри слога практически нет. Так, если описание слога в табл. 6.1 включает подряд два одинаковых согласных, то это отнюдь не означает, что для получения этого описания нужно два раза применить правила, соответствующие обнаружению одного такого согласного. Напротив того, как результаты сравнительного исследования акустической картины сочетаний типа [ata]—[atta], [asa]—[assa] и т. д., так и результаты исследования восприятия таких сочетаний при их синтезе говорят о том, что двойной и единичный согласный должны рассматриваться как два представителя одной и той же группы. Для различения обоих представителей группы [t] и [tt] от других возможных групп [s] и [ss], [p] и [pp] и т. д. нужно применять одни и те же правила, для различения внутри группы, т. е. [t] от [tt], нужно применить некоторое специальное правило, например оценить длительность паузы [267]. Для идентификации первого и второго [t] в сочетании [atsta] необходимо использовать разные правила.

Данные о том, что решение о согласных фонемах основывается на акустической информации, распределенной на слоговом отрезке, приводились в главе 3.

Эксперименты, направленные на проверку гипотезы о том, что слог соответствует одному полному циклу работы фонетического интерпретатора, исходили из следующих допущений. Если модель обрабатывает слоги последовательно один за другим, то должны существовать ограничения по частоте следования слогов. При частоте следования большей, чем предельная, распознавание слогов должно ухудшиться или стать невозможным. Эти ограничения могут быть обусловлены как тем, что сам процесс интерпретации требует времени [376], так и тем, что длительность может быть одним из формальных признаков элемента потока, используемых при его выделении (элементом не может быть отрезок короче некоторой величины).

Предположим, что процесс интерпретации действительно требует некоторого времени  $t_c$  от пачала поступления изображения слога на вход интерпретатора до выработки на выходе соответствующего названия. Тогда при поступлении второго слога с ин-

тервалом  $T < t_c$  от начала первого интерпретатор должен будет или прекратить обработку первого слога и перейти к следующему, или пропустить второй слог. В обоих случаях восприятие последовательности будет нарушено, т. е. описание стимула в табл. 6.1 будет искажено по сравнению со случаем, когда  $T \gg t_c$ .

Существенно, что определяющим должен быть именно интервал между слогами, ограничения по слоговому интервалу должны проявляться при разных распределениях длительностей гласного и согласного внутри слога.

### 6.1.1. РАСПОЗНАВАНИЕ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЛАСНЫХ. ОБРАТНАЯ МАСКИРОВКА ГЛАСНЫХ

В первой работе, посвященной исследованию восприятия последовательностей гласных [157], применялись трехчленные последовательности, образованные путем перестановок гласных [u], [a], [e]. Гласные стимулы представляли собой отрезки, вырезанные из естественных, протяжно произнесенных гласных. Стимулы, составляющие последовательность, имели одинаковую длительность (от 100 до 500 мс) и непосредственно примыкали друг к другу. Испытуемые, прослушав последовательность, записывали ее буквами. Специального обучения не проводилось.

Результаты показали, что наилучшее распознавание (90% принятой информации) соответствовало длительности гласного, равной 200 мс. Для длительности, равной 100 мс, принятая информация снижалась до 60%. Анализ ошибок показал, что вероятность правильного приема является наименьшей для среднего элемента последовательности.

Экспериментальный метод, при котором все элементы последовательности находятся в одинаковых условиях, был предложен Уорреном [518]. Метод заключается в том, что последовательность стимулов записывается на магнитную пленку, которая склеивается в кольцо. Испытуемый слушает этот «бесконечный» стимул (он может делать это достаточно долго) и должен записать последовательность элементов. Ответ считается правильным, если последовательность записанных букв соответствует последовательности элементов в стимуле.

Данные по восприятию последовательности гласных [i:au] в «бесконечном» стимуле были получены Томасом и др. [504]. Стимулами являлись отрезки заданной длины, вырезанные из протяжно произнесенных естественных гласных; между стимулами не было пауз. Испытуемые (25 человек) были осведомлены относительно того, из каких именно гласных образована последовательность; перед основным экспериментом им предлагалось прослушать последовательность, составленную из элементов длительностью по 500 мс. В основном эксперименте испытуемым последовательно предъявлялись все более трудные стимулы — по-

следовательности с постепенно уменьшающейся длительностью элементов.

Полученные результаты приведены на рис. 6.1. Можно видеть, что резкое ухудшение распознавания последовательности происходит на интервале от 125 до 100 мс. Авторы отмечают, что и при наименьших длительностях (75 и 100 мс) испытуемые продолжают слышать гласные «раздельно».

Легко заметить, что результаты приведенных выше двух работ еще не позволяют решить, что является фактором, ограничиваю-

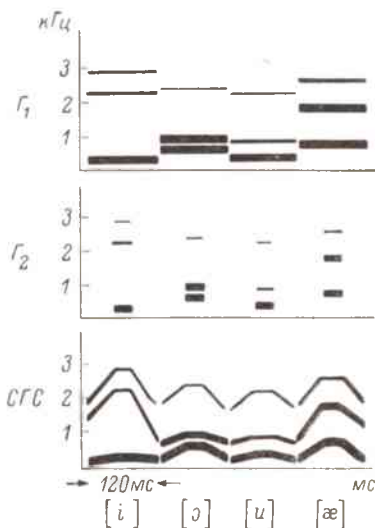
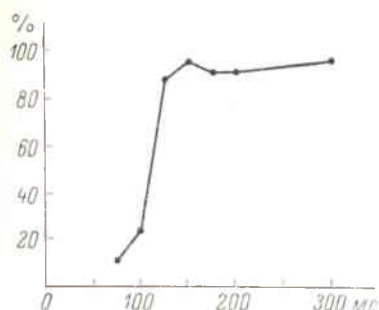


Рис. 6.1. Распознавание последовательности гласных в «бесконечном» стимуле. По [504].

По оси абсцисс — длительность гласного; по оси ординат — процент испытуемых, правильно определивших порядок следования гласных.

Рис. 6.2. Схематические динамические спектрограммы синтетических стимулов, применявшихся в работе [236].

По оси абсцисс — время; по оси ординат — формантные частоты. Остальные обозначения см. в тексте.

щим распознавание быстрых последовательностей. Им может быть как период следования гласных (интервал от начала одного гласного до начала следующего), так и длительность гласного.

Для получения ответа на этот вопрос Томас и др. [503] повторили описанные выше эксперименты с той разницей, что длительность посылок гласных [i], [ɛ], [a], [u] была теперь фиксированной (50 мс), а величина периода следования гласных менялась от 75 до 300 мс за счет изменения длительности паузы. Оказалось, что в этих условиях распознавание последовательностей не только не ухудшилось, но, наоборот, несколько улучшилось. При длительности периода, равной или большей 100 мс, распознавание было практически совершенным, при сокращении длительности периода до 87 мс правильные ответы уменьшались до 65%, а при

сокращения периода до 75 мс ответы уже становились случайными (16,7% правильных ответов).

Убедительным доказательством того, что определяющим является именно период следования гласных, является результат эксперимента, выполненного в Хаскинских лабораториях [236]. Авторы использовали в качестве стимулов длительные (20 с) отрезки «бесконечных» последовательностей из элементов, следующих друг за другом с периодом в 130 мс (рис. 6.2). Элементами были или синтетические стационарные гласные [i], [æ], [ɔ], [u] длительностью 120 мс ( $G_1$  на рис. 6.2), или 30-миллисекундные отрезки тех же гласных ( $G_2$  на рис. 6.2), или синтетические сигналы, воспринимаемые при изолированном предъявлении как сочетания согласный—гласный—согласный [bib], [bæb], [bɔb], [bub] (СГС на рис. 6.2). В этих сочетаниях длительность стационарного участка стимула составляла 30 мс, остальное время (90 мс) приходилось на формантные переходы. Для каждого из этих типов стимулов были образованы шесть разных последовательностей, различающихся порядком следования элементов. Результаты распознавания приведены в табл. 6.2.

Т а б л и ц а 6.2

Зависимость правильного распознавания последовательностей от типа стимулов и от порядка следования гласных (в %)

| Последовательность | $G_1$ | $G_2$ | СГС |
|--------------------|-------|-------|-----|
| [i], [æ], [ɔ], [u] | 88    | 50    | 88  |
| [i], [æ], [u], [ɔ] | 44    | 38    | 78  |
| [i], [ɔ], [æ], [u] | 50    | 13    | 50  |
| [i], [ɔ], [u], [æ] | 39    | 50    | 67  |
| [i], [u], [æ], [ɔ] | 63    | 75    | 88  |
| [i], [u], [ɔ], [æ] | 50    | 63    | 50  |
| Среднее . . . . .  | 52    | 47    | 71  |

Можно видеть, что последовательности распознаются существенно лучше в случае сигналов СГС. Практически одинаковое распознавание последовательностей для сигналов  $G_1$  и  $G_2$  говорит о том, что роль длительности отрезка гласного весьма невелика.

Худшее распознавание последовательностей из отрезков стационарных гласных по сравнению с сигналами СГС связано с тем, что когда два соседних гласных имеют существенно отличающиеся значения  $F_1$ , они перестают восприниматься как принадлежащие одному и тому же сообщению. Правильность этого объяснения была проверена дополнительным экспериментом, в котором гласные были соединены формантными переходами. Распознавание последовательностей для таких сигналов оказалось значительно лучшим, чем для  $G_1$ , и совпало с таковыми для СГС.

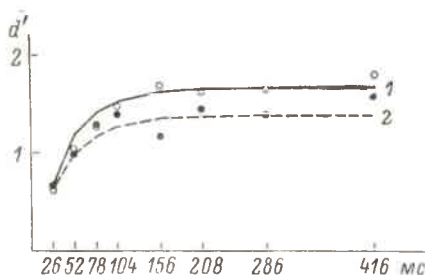


Вторым методом определения минимального интервала между гласными, необходимого для обработки первого гласного, является метод так называемой обратной маскировки.

Испытуемым предъявляются двучленные последовательности; при этом они должны идентифицировать первый гласный, игнорируя второй гласный, служащий маскером. Результаты, получаемые этим методом, сильно различаются в разных работах [237, 377, 423], значительно зависят от индивидуальных характеристик испытуемого и от числа выборов — возможных фонем [237]. С точки зрения чистоты эксперимента этот метод значительно уступает методу распознавания «бесконечных» последовательностей.

Рис. 6.3. Распознавание гласного в условиях обратной маскировки. По [377].

По оси абсцисс — длительность интервала от начала гласного до начала маскера; по оси ординат — обнаружимость  $d'$ . 1 — конец гласного совпадает с началом маскера, 2 — длительность гласного равна 26 мс.



В описываемом ниже эксперименте Массаро [377] гласные стимулы были получены на основе естественных, произнесенных женщиной гласных [i], [ɪ], [a], [u]. Из гласного, записанного в память вычислительной машины, выбирался стандартный стационарный отрезок, длительностью 26 мс (пять периодов основного тона 192 Гц). Использовались два варианта тестирования: с паузой и без паузы. В варианте с паузой длительность первого гласного равнялась 26 мс, после паузы переменной длительности включался маскер. Интервал от начала гласного до начала маскера обозначим  $T$ . В варианте без паузы гласный выключался в момент начала маскера. Маскером служил «бессмысленный» гласный, длительностью 208 мс. Он был составлен из стандартных отрезков четырех тестирующих гласных; каждый стандартный отрезок повторялся в маскере два раза.

Использовалась методика форсированного выбора (четыре возможных ответа); первый опытный день был посвящен тренировке в распознавании тестирующих гласных разной длительности. При обработке данных в качестве меры разборчивости была использована «обнаружимость» —  $d'$ , принятая в теории обнаружения.

На рис. 6.3 приведены в качестве примера данные для гласного [ɪ]. Можно видеть, что при тестировании с паузой (точки) обнаружимость растет при увеличении интервала  $T$  до 104 мс. Близкий вид зависимости наблюдается и при тестировании без паузы (кружки). Так как значения  $d'$  выше при тестировании без



паузы, очевидно, что длительность гласного, равная 26 мс, является еще слишком малой для надежного распознавания. Известно, что человек может научиться распознавать очень короткие изолированные гласные [483], однако для этого необходима специальная тренировка, в некоторых случаях очень длительная [10]. Итак, на основании экспериментов по распознаванию последовательностей гласных и данных по обратной маскировке гласных можно считать, что период следования гласных в звуковом потоке не должен быть меньше примерно 100—150 мс. В противном случае фонемное распознавание гласных оказывается нарушенным.

### 6.1.2. МИНИМАЛЬНАЯ ДЛИТЕЛЬНОСТЬ СЛОГА, НЕОБХОДИМАЯ ДЛЯ ЕГО ОБНАРУЖЕНИЯ

На основании приведенных выше данных можно сделать вывод, что для нормальной работы фонетического интерпретатора частота следования слогов в речевом потоке не должна превышать 10 в 1 с. Это примерно соответствует максимальной частоте повторения движений при артикуляции [рарара...], [tatata...] [311, 345, 467].

Однако известно, что человек способен воспринимать ускоренную (компрессированную по времени) речь при частоте следования фонем, достигающей 30 в 1 с [172, 329, 415].

Исходя из того, что понимание смысла сообщения отнюдь не требует полного и правильного описания всех фонем в оперативной памяти на входе анализатора смысла, заманчиво предположить, что нарушения работы фонетического интерпретатора при частоте следования слогов 10 в 1 с, будут выражаться не в том, что слоги будут пропущены, а в том, что о слогах будет записана лишь неполная информация. Иначе говоря, можно ожидать, что при этой частоте будут еще работать правила, обеспечивающие последовательное заполнение табл. 6.1, переход от одной группы столбцов к следующей. Нарушения будут касаться лишь содержания записанной информации.

Описываемые ниже эксперименты были направлены на определение минимальной длительности слогов, при которой слог еще уществует.

Ориентировочные данные были получены в экспериментах амплитудно-модулированным [а]-образным гласным [63]. При прямоугольной модуляции с достаточной глубиной такой стимул воспринимается как последовательность слогов, если частота модуляции низка. Испытуемым предлагалось, изменяя частоту модуляции, найти значения частот, соответствующие отчетливой мере характера восприятия.

На рис. 6.4 приведены распределения значений для трех границ: граница между нормальной и быстрой последовательностью слогов (1), граница между быстрой последовательностью слогов звуком [г] (2), верхняя граница восприятия [г] (3).

Можно видеть, что граница нормальной последовательности слогов устанавливается, как правило, на частоте модуляции, не превышающей 10 Гц. Однако и при больших частотах модуляции

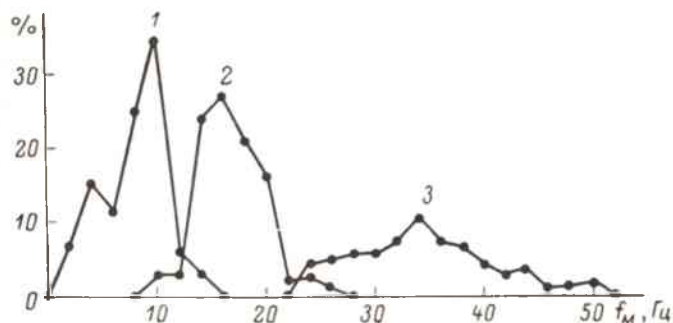


Рис. 6.4. Распределение частот модуляции, соответствующих изменению качества звучания стимула. По [63].

По оси абсцисс — частота модуляции; по оси ординат — доля установок данной границы на интервале в 2 Гц. 1 — граница между нормальной и быстрой последовательностью слогов, 2 — граница между быстрой последовательностью слогов и [г], 3 — граница между [г] и хриплым звуком.

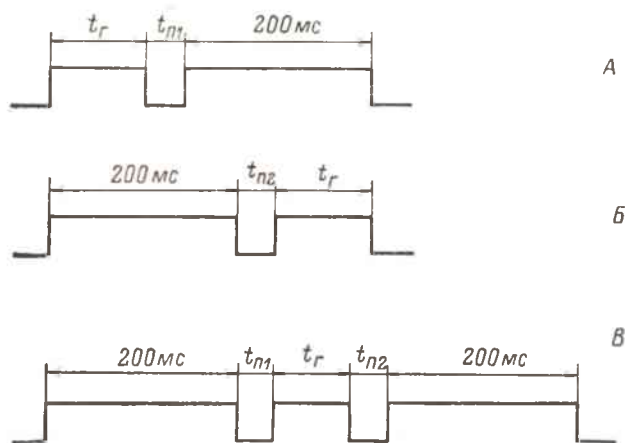


Рис. 6.5. Схематическое изображение стимулов, применявшихся в работе [62].  
Обозначения см. в тексте.

стимул продолжает восприниматься как слоговая последовательность. Среднее значение частоты, при которой слоги исчезают и сигнал превращается в непрерывный [г]-образный звук (среднее для 2), составляет 16.2 Гц, что соответствует периоду модуляции, равному 65 мс.

Описываемые ниже эксперименты [52] ставили своей целью более детальный анализ условий, необходимых для обнаружения слога.

Применявшиеся стимулы представляли собой последовательности из двух или трех отрезков синтетического гласного (рис. 6.5). Параметрами стимула, управляемыми экспериментатором, служили длительности пауз ( $t_{n1}$ ,  $t_{n2}$ ). Задачей испытуемого являлось нахождение минимальной длительности отрезка гласного ( $t_r$ ),

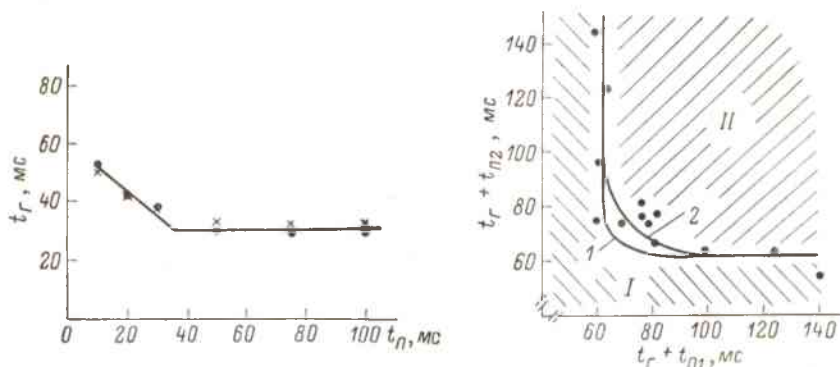


Рис. 6.6. Условия обнаружения слога в двухсложной последовательности. Зависимость между длительностью паузы ( $t_n$ ) и длительностью отрезка гласного ( $t_r$ ).

Обозначения см. в тексте.

Рис. 6.7. Условия обнаружения второго слога в трехсложной последовательности.

Область отсутствия восприятия второго гласного (слога) — I и область обнаружения второго гласного (слога) — II в пространстве двух параметров:  $t_r + t_{n1}$  (интервал от конца первого до конца второго отрезка гласного) и  $t_r + t_{n2}$  (интервал от начала второго до начала третьего отрезка гласного), см. рис. 6.5. В. Точки — экспериментальные данные, кривые 1 и 2 — результаты расчета.

при которой стимулы A и B воспринимались как двусложные «слова», а стимул B — как трехсложное «слово».

На рис. 6.6 приведены данные для стимулов A (точки) и B (крестики). Из рисунка следует, что при больших значениях паузы минимальная длительность отрезка гласного постоянна. При малых значениях паузы примерно постоянным является суммарный интервал из отрезка гласного и паузы.

На основании рис. 6.6 можно сделать вывод, что для восприятия гласного (слога) необходимо совместное выполнение трех условий: длительность отрезка гласного должна быть не меньше некоторого порога ( $t_r \geq C_1$ ), длительность интервала от начала этого гласного до начала следующего гласного должна быть не меньше некоторого порога ( $t_r + t_{n2} \geq C_2$ ), длительность интервала от конца предыдущего гласного до конца данного гласного должна быть не меньше некоторого порога ( $t_r + t_{n1} \geq C_3$ ).

Приведенное выше условие обнаружения слога можно сокращенно записать как  $(t_r \geq C_1) \wedge (t_r + t_{n2} \geq C_2) \wedge (t_r + t_{n1} \geq C_3)$ .

Для выяснения совместимости этого вывода с данными, полученными в экспериментах со стимулами *B* (рис. 6.5), была принята следующая элементарная модель.

Критерии  $C_1$ ,  $C_2$ ,  $C_3$  считались нормально распределенными случайными величинами со следующими средними значениями, определенными из данных экспериментов со стимулами *A* и *B*:  $C_1=32$ ,  $C_2=C_3=62$  мс. Принималось, что испытуемый устанавливает такую длительность отрезка гласного ( $t_r$ ), при которой вероятность обнаружить слог равна 0.5.

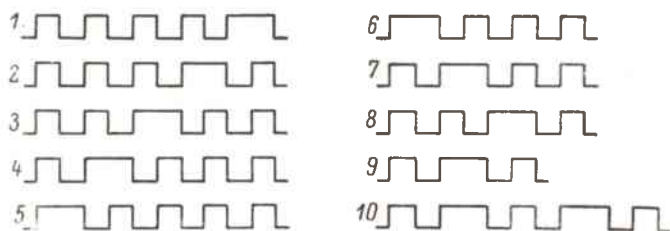


Рис. 6.8. Схематическое изображение стимулов, применявшихся в работе [92].

Обозначения см. в тексте.

Кривые на рис. 6.7 показывают вычисленные комбинации значений  $(t_r + t_{n1})$  и  $(t_r + t_{n2})$ , соответствующие вероятности обнаружения слога, равной 0.5. Они определены для случая, когда условие  $(t_r \geq C_1)$  всегда автоматически выполняется, если выполняется одно из двух остальных условий. При вычислениях было принято, что флуктуации  $C_2$  и  $C_3$  являются независимыми,  $\sigma_{C_2} = \sigma_{C_3}$  и составляет 10 мс (кривая 1) или 20 мс (кривая 2). Точками на рис. 6.7 показаны экспериментальные данные, полученные для стимулов *B* при следующих значениях пауз:  $t_{n1}=15$  мс;  $t_{n2}=10, 15, 20, 30, 50, 75, 100$  мс;  $t_{n2}=15$  мс,  $t_{n1}=10, 20, 30, 50, 75, 100$  мс.

Кривую на рис. 6.7 можно рассматривать как границу области обнаружения гласного в пространстве таких параметров, как интервал между концами соседних отрезков  $(t_r + t_{n1})$  и интервал между началами соседних отрезков  $(t_r + t_{n2})$ . Область II соответствует обнаружению, область I — отсутствию обнаружения.

Для понимания существа выявленных временных ограничений очень важно учитывать, что пропуск (отсутствие обнаружения) гласного отнюдь не означает пропуска отрезка стимула. Человек слышит в этих условиях кластер — сочетание согласных. Например, стимул *B* воспринимается как [atra] (если  $t_{n1} > 50$  мс, а  $t_{n2} \leq 30$  мс) или как [arta] (если  $t_{n1} \leq 30$  мс, а  $t_{n2} > 50$  мс).

Таким образом, полученные данные описывают условия, при которых отрезок стимула интерпретируется или как самостоя-

тельный элемент — гласный (слог), или как составная часть другого элемента.

Заметим, что с точки зрения записи информации в табл. 6.1 обнаружение слога (гласного) означает переход к некоторой следующей группе столбцов таблицы, пропуск слога означает продолжение записи в ту же группу столбцов таблицы, с которой до этого работала модель.

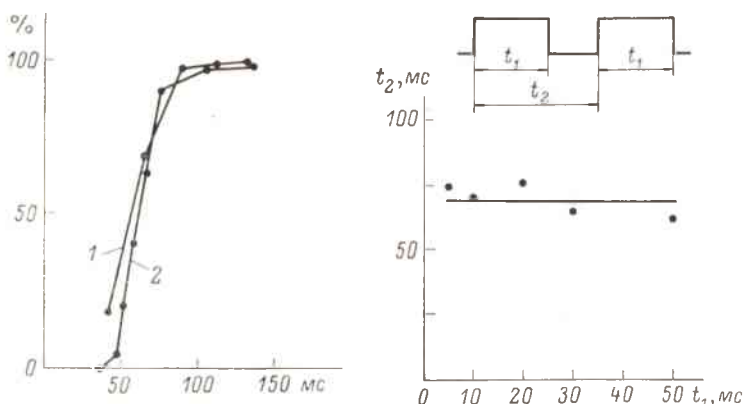


Рис. 6.9. Временные ограничения восприятия сигнала как последовательности элементов. По [92].

По оси абсцисс — длительность интервала «посылка плюс пауза» или период модуляции; по оси ординат — процент правильных ответов или случаев, когда установленное граничное значение периода модуляции меньше значения, указанного по оси абсцисс. 1 — процент правильного воспроизведения ритмического рисунка последовательности как функция интервала «посылка плюс пауза»; 2 — накопленная гистограмма значений установок периода амплитудной модуляции, соответствующего границе между [г] и последовательностью слогов.

Рис. 6.10. Условия, соответствующие порогу восприятия «двух звуков». По [49].

По оси абсцисс — длительность послышки; по оси ординат — длительность интервала «посылка плюс пауза». Вверху — схематическое изображение стимула. Остальные обозначения см. в тексте.

Согласно представлениям, изложенным в предыдущих главах книги, информация о длительностях, определяющая ритмическую структуру последовательности, приписана гласным. Это означает, что пропуск гласного (слога) должен приводить к искажению ритмической структуры воспринимаемой последовательности.

В работе [92] исследовалось воспроизведение ритмической структуры последовательностей тональных посылок (с частотой 1 кГц) при разной частоте следования посылок. Применявшиеся последовательности схематически показаны на рис. 6.8. Длительность короткой послышки ( $T$ ) равнялась длительности паузы. Длительность длинной послышки равнялась или  $2T$ , или  $3T$  (разные варианты эксперимента). Последовательности, записанные на магнитофонную ленту, следовали друг за другом с интервалом в 4 с.

За время этого интервала испытуемый должен был воспроизвести ритмический рисунок последовательности как [tatata. . .] в удобном для него темпе. Разные таблицы (100 последовательностей) отличались друг от друга значениями  $T$ .

На рис. 6.9 кривой 1 показан процент правильно воспроизведенных последовательностей как функция от длительности интервала, занятого короткой посылкой плюс паузой ( $2T$ ). Кривой 2 показана накопленная гистограмма значений границы между [г] и последовательностью слогов, определенных в работе [63]. При построении этой гистограммы использовались данные, приведенные на рис. 6.4, частота амплитудной модуляции пересчитывалась в период модуляции ( $1/f_m$ ). Можно видеть, что обе кривые практически совпадают.

Анализ искажений ритмического рисунка при его воспроизведении (он был сделан для тех таблиц стимулов, где количество ошибочных воспроизведений превышало 30%) показал, что случаи, когда ответная последовательность содержит больше элементов, чем переданная, очень редки. Они составляют всего 7% от всех ошибок. В 76% случаев ответная последовательность содержит меньше элементов, чем переданная. Примерно 45% всех ошибочных ответных последовательностей соответствует таким последовательностям, которые были бы получены из переданных, если бы в переданной последовательности была пропущена одна короткая посылка.

Таким образом, можно утверждать, что основной причиной искажения ритмического рисунка является пропуск одного или нескольких элементов.

Дополнительные эксперименты показали, что определяющим фактором является длительность «слога», т. е. интервала, состоящего из посылки и паузы. Перераспределение времени между посылкой и паузой внутри этого интервала мало влияет на результаты эксперимента.

Совпадение результатов, полученных в трех описанных выше работах, позволяет думать, что правила обнаружения элемента в последовательности не являются специфическими для речи, т. е. скорее заложены генетически, чем приобретены в процессе обучения.

Доводом в пользу этого предположения являются результаты, полученные Гольдбурт [49]. Автор предъявляла испытуемым две следующие друг за другом одинаковые по частоте тональные посылки и меняла как длительность посылки ( $t_1$  на рис. 6.10), так и величину паузы между ними. Испытуемые должны были сообщать, слышат ли они два сигнала или один сложный звук.

Полученные результаты приведены на рис. 6.10 (график построен нами по данным [49], с. 154; вычислены средние значения порога по всем испытуемым). Из рис. 6.10 следует, что порог восприятия двух сигналов определяется интервалом  $t_2$ , равным сумме длительности посылки и паузы. Как распределен этот



интервал между посылкой и паузой, безразлично. Значение  $t_2$  находится в районе 60—70 мс, т. е. практически совпадает с приведенными выше оценками минимальной длительности слога, необходимой для его обнаружения.

Приведенные в настоящем разделе данные показывают, что выделение элемента основывается на временных правилах такого типа, что два «одинаковых» события (начала двух соседних посылок или концы двух соседних посылок) должны появляться не раньше, чем через определенный интервал времени. Для реализации таких правил необходимо располагать специальными сигналами, обозначающими начало и конец посылки — марками начала и конца.

Если марка начала первого отрезка предшествует марке начала второго отрезка меньше, чем на примерно 65 мс, то первый отрезок не принимается за гласный звук. Таким образом, решение о гласном должно быть задержано не менее чем на 65 мс от начала отрезка. Естественно предположить, что определение того, к какой именно гласной фонеме относится отрезок, должно начаться уже после того, как выяснено, что отрезок действительно является гласным звуком. С этой точки зрения кажется довольно логичным, что обнаружение гласных (слов) осуществляется при более высокой частоте следования слогов, чем это допустимо для фонемного распознавания гласных (рис. 6.11).

### 6.1.3. ПРОБЛЕМА КОНЦА ГЛАСНОГО

Система, производящая сегментацию речевого потока на отрезки гласных и негласных и пользующаяся понятием допустимых интервалов между началами и концами гласных, должна, естественно, пользоваться каким-то формальным определением гласного звука. Логично предположить, что гласный звук определяется как отрезок сигнала между маркой начала и маркой конца. При этом определении мы сталкиваемся с той трудностью, что гласный звук может быть длинным.

Очевидно, что если сенсорная память ограничена — модель «видит» только небольшой участок прошлых значений сигнала — фонемная интерпретация длинного отрезка должна осуществляться до того, как этот отрезок закончится. В противном случае может оказаться так, что к моменту начала интерпретации акустическая информация о начале отрезка будет потеряна. Тот факт, что человек может осуществлять фонемное распознавание гласного звука, не дожидаясь конца этого звука, доказывается экспериментами по быстрой имитации гласных. Средние значения скрытых периодов имитации обычно не превышают 200 мс [<sup>156</sup> <sup>213</sup>]. Характер зависимости между формантными характеристиками гласных-стимулов и гласных-реакций [<sup>213</sup>] говорит о том, что классификация (на «психологические» фонемы) действительно имеет место. Надежность классификации достаточно велика — разборчивость (про-

Рис. 6.11. Временные ограничения восприятия сигнала как последовательности элементов (1 и 2) и фонемного распознавания последовательности (3 и 4).

3 и 4 — зависимости правильного распознавания последовательности гласных в «бесконечном» стимуле от периода следования гласных по [503] и [504] соответственно. Остальные обозначения те же, что на рис. 6.9.

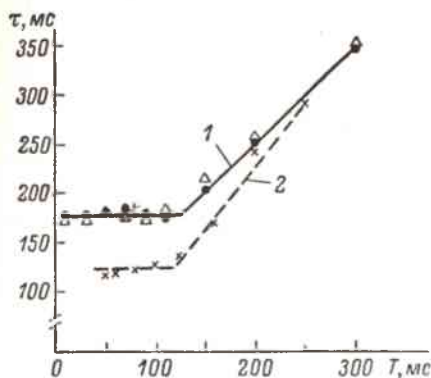
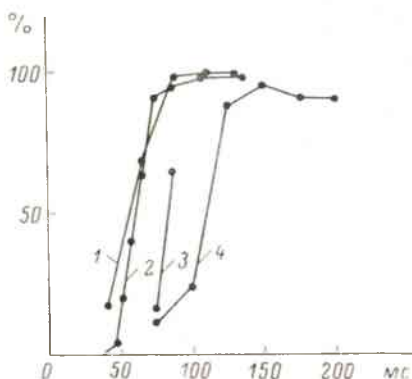
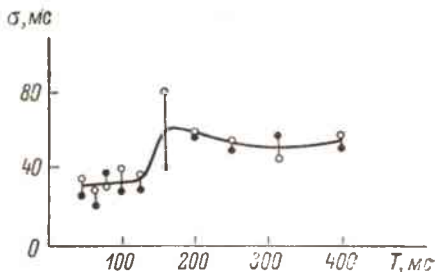


Рис. 6.12. Критический интервал от начала звука по данным [240] и [154].

Обозначения см. в тексте.

Рис. 6.13. Зависимость стандартного отклонения ( $\sigma$ ) длительностей гласного-реакции, создаваемого при текущей имитации, от длительности ( $T$ ) гласного-стимула. По [154].

Точки и кружки — разные гласные.



цент правильного узнавания) гласных-стимулов, определенная при прослушивании вызванных этими стимулами гласных-реакций, составляет 91—98% [156]. Спектр гласного-реакции, создаваемого при быстрой имитации, как правило, не меняется на протяжении звука [213]. Это говорит о том, что к моменту начала фонации, т. е. примерно через 200 мс от начала стимула (длительность стимула составляла 300 мс), речевой тракт имитирующего испытуемого уже принял соответствующую данной фонеме конфигурацию.

Распознавание стимула, основанное на наблюдении отрезка стимула, но не трубящее информации об окончании стимула, могло бы обеспечиваться «опросом», задержанным относительно марки начала на фиксированный интервал времени.

Данные о существовании некоторого критического интервала от начала стимула были получены Эфроном [240]. Эксперимент состоял в том, что на одно ухо испытуемого подавалась посылка белого шума с разными значениями длительности, на второе ухо — тон 2 кГц, длительностью 500 мс, запаздывающий по времени относительно шума. Время запаздывания начала тона изменялось, испытуемый должен был давать один из двух ответов: «стимулы переслаиваются», «между стимулами имеется пауза». Пороговым считалось то значение запаздывания начала тона относительно начала шума, которое соответствовало смене этих ответов.

Полученные значения порогового запаздывания как функция длительности шумовой посылки приведены на рис. 6.12, кривая 1 (график построен по данным Эфрона [240], приведенным в табличной форме; точки и треугольники — данные разных испытуемых).

Можно видеть, что, пока длительность шума не достигнет 120—130 мс, значение запаздывания не меняется, т. е. не зависит от длительности шума. При больших длительностях шума величина запаздывания равна длительности шума плюс примерно 50 мс, т. е. начало тона привязано к концу шума.

Принципиально аналогичные результаты были получены раньше [154] при исследовании вопроса о том, с каким отрезком стимула работает человек, преобразуя сигнал в последовательность артикуляторных движений при быстрой имитации. Испытуемые должны были как можно быстрее повторять синтетические стационарные гласные, длительности которых варьировали от стимула к стимулу в пределах от 50 до 400 мс.

Ясно, что стимулы, кончающиеся позже того как испытуемый уже начал их воспроизводить, должны трактоваться как последовательность событий «начало звука» и «конец звука», а их воспроизведение — как последовательность реакции «начало фонации» и «конец фонации».

Была сделана попытка найти, какой длительностью должен обладать стимул, чтобы трактоваться как последовательность из двух событий. На рис. 6.12 (кривая 2) приведена средняя длительность гласного-реакции (по оси ординат) как функция длитель-

ности гласного-стимула.<sup>5</sup> Можно видеть, что пока длительность стимула не превышает 100—125 мс, испытуемые создают реакции практически одинаковой длительности.

При больших длительностях стимула длительность реакции резко возрастает и начинает зависеть от длительности стимула. Свидетельством скачкообразного перехода от одного способа реагирования к другому является зависимость стандартного отклонения длительности реакции от длительности стимула (рис. 6.13).

Обе кривые на рис. 6.12 свидетельствуют о существовании некоторой критической длительности звука, причем значения ее оказываются одинаковыми в обеих работах.

Подчеркнем, что на основании этих работ можно говорить лишь о существовании такого феномена, как критическая длительность. Эксперименты слишком сложны, чтобы можно было понять, что скрывается за этой критической длительностью и имеет ли она какое-либо значение при восприятии речи. Тем не менее любопытно проверить, не были ли обнаружены при исследовании восприятия речи временные константы, равные критической длительности. Оказывается, что примерно такое значение имеет граница по длительности между долгими и краткими гласными фонемами (см. главу 3). Кроме того, пороговая длительность слога, необходимая для его обнаружения (см. раздел 6.1.2), равна половине критической длительности. Хотя очень вероятно, что это является случайным совпадением, заманчиво предположить, что марка начала посылки вызывает генерацию импульсов опроса, следующих с фиксированной низкой частотой.

Вернемся теперь к проблеме конца гласного в речевом потоке и возможным экспериментальным методам ее исследования.

Попытки разработать технические схемы для сегментации речевого потока на гласные и негласные наталкиваются на то, что по своим физическим характеристикам гласные звуки довольно близки к сонантам. Следовательно, существует опасность, что явление, принятое за марку начала гласного, в действительности окажется началом сонанта. Выяснить, была ли допущена ошибка, можно, дождавшись следующей по времени марки. В случае гласного она должна оказаться маркой конца, в случае сонанта — маркой начала (начало следующего за сонантом гласного). Однако, если сенсорная память ограничена, ждать слишком долго нельзя, т. е. после некоторой задержки решение должно быть принято, даже если вторая марка не поступила.

Попытка выяснить, как будет вести себя человек в ситуации, когда неизвестно, с чего (гласного или сонанта) начинается стимул, была предпринята в работе [85] и в экспериментах Жукова.

На рис. 6.14 приведены два типа огибающих синтетического гласного [а], применявшиеся в экспериментах Жукова. Конечный сегмент стимула в обоих типах стимулов (*A* и *B*) был одинаковым, его длительность ( $t_2$ ) составляла 200 мс. Как стимул *A*, так и стимул *B* начинался с сегмента, уровень интенсивности которого

был на 12 дБ ниже уровня интенсивности конечного сегмента. Разница между стимулами заключалась в том, что стимул *А* состоял из трех сегментов (уровень интенсивности среднего сегмента был на 6 дБ ниже начального уровня), а стимул *Б* — из двух сегментов.

По предположению, стимулу *А* должна соответствовать последовательность марок границ, характерная для сочетания ГСГ — за первой маркой начала следует марка конца. Стимулу *Б* должна соответствовать последовательность марок, характерная для

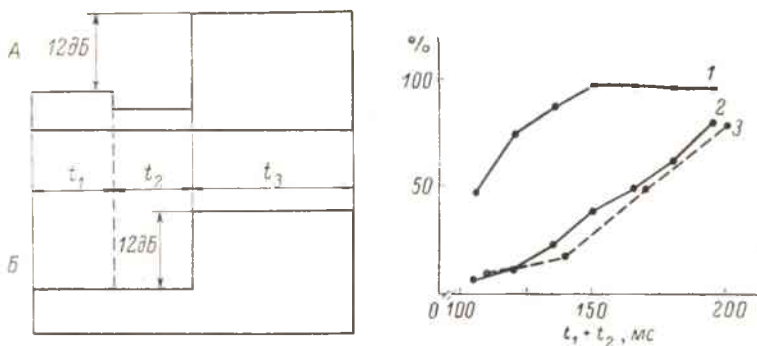


Рис. 6.14. Схематическое изображение трехсегментного (*А*) и двухсегментного (*Б*) стимулов.

Обозначения см. в тексте.

Рис. 6.15. Обнаружение первого гласного в трехсегментных и двухсегментных стимулах.

По оси абсцисс — длительность интервала от начала стимула до момента возрастания его интенсивности; по оси ординат — относительное количество случаев, когда стимул воспринимался как последовательность ГСГ. 1 — трехсегментный стимул (*А* на рис. 6.14); 2 и 3 — двухсегментный стимул (*Б* на рис. 6.14). Остальные обозначения см. в тексте.

слога СГ — за первой маркой начала следует еще одна марка начала.

Длительность среднего сегмента в стимуле *А* ( $t_2$ ) (и соответствующей части первого сегмента в стимуле *Б*) равнялась 75 мс. Переменным параметром была длительность  $t_1$  (рис. 6.14).

В таблице, записанной на пленку, стимулы обоих типов следовали в случайном порядке. Испытуемые должны были определять, является ли стимул одним гласным, слогом СГ или сочетанием ГСГ.

Полученные данные (% ответов ГСГ) приведены на рис. 6.15 (сплошные кривые). Штриховой кривой приведены данные работы [85], где применялись только стимулы типа *Б*.

Данные для трехсегментного стимула *А* практически совпали с данными, полученными ранее (см. раздел 6.1.2) для последовательностей посылок гласного, разделенных паузами. 50% ответов ГСГ соответствует длительности первого сегмента ( $t_1$ ), равной

35 мс ( $t_1 + t_2 = 110$  мс). Таким образом, если у отрезка гласного выделяется конец, обнаружение гласного может осуществляться достаточно быстро.

Для восприятия двухсегментного стимула  $B$  как сочетания ГСГ требуется большая длительность первого сегмента. 50% ответов ГСГ приходится на значение длительности, равное 170 мс. Судя по тому, что и при 200 мс ГСГ воспринимается еще не в 100% случаев, максимальная задержка фонемной интерпретации первого отрезка может, очевидно, превышать 200 мс. Если верны приведенные выше соображения о связи максимальной задержки с длиной отрезка сигнала, хранимого в «сенсорной» памяти, то длина этого отрезка должна быть не меньше 200 мс.

Вся сумма данных, приведенных в разделах 6.1.1—6.1.3, подтверждает представление о том, что слоги обрабатываются последовательно и что понятие слога используется при членении потока на элементы. Для формирования в модели понятия слога необходимы такие промежуточные понятия, как марки начала и конца отрезка. Вопрос о том, как и где образуются эти марки, тесно связан с проблемой описания сигнала в сенсорной памяти.

## 6.2. ОПИСАНИЕ СИГНАЛА НА ВХОДЕ ФОНЕТИЧЕСКОГО ИНТЕРПРЕТАТОРА

До сих пор мы говорили о том, что отрезок сигнала, обрабатываемый фонетическим интерпретатором, можно представить себе как картинку, изображающую или динамическую спектрограмму этого отрезка, или кривые параметров сигнала на этом отрезке. Сейчас необходимо более точно определить, что можно понимать под картинкой и какие допущения скрываются за предположением о том, что интерпретатор «видит» картинку.

Картинкой естественно считать последовательность из  $n$  равноотставленных (на  $\Delta\tau$ ) временных отсчетов значений или спектра сигнала, или его параметров (частот формант, основного тона, огибающей и т. д.).

Говоря, что фонетический интерпретатор «видит» картинку стимула на временном интервале  $\tau_k = n\Delta\tau$ , мы подразумеваем, что он одновременно имеет на входе информацию о всех  $n$  отсчетах и что упорядоченность значений изображения стимула по оси  $\tau$  воспроизводит порядок появления этих значений стимула во времени. Это означает, что мы допустили, что до интерпретатора имеется память. Элементы этой памяти образуют упорядоченную шкалу, на которой отображается время. Память, отвечающая этим требованиям, может быть реализована или в виде набора линий задержки, где каждая линия имеет  $n$  отводов с интервалами  $\Delta\tau$ , или в виде упорядоченного множества запоминающих элементов, в которые последовательно, по мере поступления сигнала, записываются значения его временных отсчетов.



В случае первого предположения (линии задержки) мы имеем дело с «временным окном», мимо которого скользит изображение. При этом возникает проблема выбора моментов времени, когда это изображение будет передаваться на вход интерпретатора. В случае второго предположения, если  $\tau_k$  не превышает длительности слога, возникает проблема управления записью и стиранием информации.

При представлении сигнала картинкой появляется неприятный эффект: одному и тому же слогу, многократно произнесенному одним и тем же диктором, будут соответствовать разные изображения. Это обусловлено случайными колебаниями темпа речи и тем, что движения разных артикуляторных органов не являются идеально синхронизированными. В результате возникает очень сложная проблема временной нормализации изображения. Нормализация должна логически предшествовать распознаванию.

За последние годы все более отчетливо проявляется стремление найти такую форму изображения речевого сигнала, при которой последовательность наблюдаемых в сигнале явлений была бы сохранена, но точные временные отношения между этими явлениями были бы несущественными.

Кажется, что такая возможность появляется, если допустить, что на вход интерпретатора поступает не спектральное или параметрическое изображение речевого сигнала, но поток акустических событий, обнаруженных в этом сигнале слуховой системой. В качестве примера возможных событий можно назвать появление или, наоборот, исчезновение энергии в определенной области спектра, сдвиг спектрального максимума в определенном направлении, кратковременный импульс или, наоборот, паузу в сигнале.

Если фонетический интерпретатор имеет дело с потоком уже выделенных событий, то для него время как одна из координатных осей изображения может не существовать. Вместо времени появляются такие качественные понятия, как «одновременно», «следует за», «предшествует», «отставлено на  $t > C$ » и т. д.

Допустив, что слуховая система действительно обнаруживает такие (или какие-то другие) события, мы все равно столкнемся с проблемой запоминания. Для того чтобы избирательно реагировать на последовательность событий  $\langle A, B \rangle$  при неодновременном появлении  $A$  и  $B$  во времени, очевидно, необходимо запоминать  $A$  на интервале времени, в течение которого  $B$  может появиться. Однако такое запоминание просто реализовать, оно отнюдь не требует записи временных отсчетов процесса.

Итак, кажется возможным назвать две гипотезы относительно природы описания сигнала на входе фонетического интерпретатора. Первую из них мы будем обозначать как гипотезу картинок, вторую — как гипотезу потока событий. Они принципиально отличаются друг от друга уже тем, что предполагают разное разделение задач по обработке речевого сигнала между слуховой системой и фонетическим интерпретатором. Согласно первой гипотезе,

роль слуховой системы в обработке речевого сигнала сводится к частотному анализу и, возможно, выделению текущих значений параметров. При этом слуховая система должна обладать памятью (например, линиями задержки), допускающей подробное описание отрезка стимула длительностью порядка слога. Задачей фонетического интерпретатора (в процессе обучения) является формирование полезных признаков и выработка решающих правил, необходимых для распознавания сложных изображений.

Согласно второй гипотезе, слуховая система не только производит частотный анализ сигнала, но и частично решает задачу распознавания, обнаруживая в сигнале ряд существенных событий. Задачей фонетического интерпретатора (в процессе обучения) является формирование схем, реагирующих на определенные последовательности (комбинации) из поступающих на него событий.

Так как вторая гипотеза предполагает весьма специфическую обработку сигнала слуховой системой, проверка ее относится к области физиологии слуха.

Данные по восприятию согласных и быстрых неречевых последовательностей, которые будут приведены ниже, показывают, что допущение о том, что слуховая система выделяет определенные события в речевом потоке, является удобным в том смысле, что позволяет проще и экономичнее описать экспериментальные результаты. Кроме того, они содержат некоторые указания на то, какие события следовало бы искать, если бы вторая гипотеза оказалась верной.

### 6 2.1. ОБНАРУЖЕНИЕ [г] РУССКИМИ СЛУШАТЕЛЯМИ

Кратковременное (не более чем примерно на 40 мс) прерывание гласного приводит к тому, что носители таких языков, как болгарский и русский, слышат [г] и стимул воспринимается ими как последовательность ГСГ [67, 216].

Если длительность исходного гласного стимула фиксировать и смещать положение паузы относительно краев стимула, то человек воспринимает перемещение ударения с одного гласного на другой. Если сделать паузу совсем короткой, то [г] исчезает, исчезает и эффект перемещения ударения при смещении паузы. Испытуемый начинает слышать один непрерывный гласный, на фоне которого действует щелчок.

На этом изменении восприятия при укорочении паузы была основана методика определения минимальной [г]-паузы в синтетических гласных [52, 62]. Экспериментатор определял пороговую паузу методом границ, постепенно ее уменьшая или увеличивая шагом в 1 или 0.5 мс. Испытуемый имел возможность смещать положение паузы относительно краев стимула. Если при этом происходило перемещение ударения, испытуемый говорил, что пауза обнаруживается.

Определенные этим методом значения пороговой [r]-паузы у двух испытуемых оказались следующими: для [u] — 8.5, 8.0 мс; для [i] — 8.2, 7.9 мс; для [a] — 5.7, 5.6 мс.

Так как [r]-пауза является очень короткой, ее можно использовать в эксперименте в качестве примера точно локализованного во времени события и попытаться выяснить, что для человека является характеристикой события. Логично подозревать, что характеристикой события может быть значение спектра (значение частот формант) в момент его возникновения.

В русском языке есть такая «удобная» особенность, что согласные фонемы различаются по твердости /мягкости, т. е. имеются

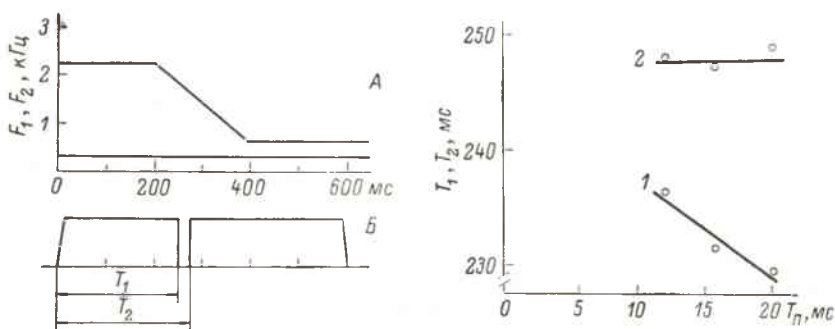


Рис. 6.16. Схематическое изображение стимулов, применявшихся в работе [62].

А — формантный контур, Б — огибающая стимула.

Рис. 6.17. Положение паузы в стимуле, показанном на рис. 6.16, соответствующее границе между [iru] и [ir'u]. По [62].

По оси абсцисс — длительность паузы; по оси ординат — положение начала паузы  $T_1$  (1) и положение конца паузы  $T_2$  (2).

фонема [r] (твердая) и фонема [r'] (мягкая). Установлено также, что основным признаком, используемым для различения твердых и мягких согласных, является частота второй форманты [167, 169, 230].

Это дало возможность использовать такую методику, когда испытуемому предъявлялся стимул с формантным контуром, соответствующим [iu], и паузой в огибающей (рис. 6.16), а испытуемый должен был установить положение паузы, соответствующее фонемной границе между [r] и [r'] ([iru]/[ir'u]), при разных длительностях паузы. Предварительные опыты показали, что пороговое значение [r]-паузы в таком нестационарном стимуле практически совпадает с таковым для стационарных [u]- и [i]-стимулов.

В основном эксперименте (на группе из пяти испытуемых) использовались паузы, длительностью в 12, 16 и 20 мс. Каждый испытуемый осуществлял по 10 установок положения паузы, соответствующего границе между [iru] и [ir'u].

На рис. 6.17 приведены средние значения установленного положения паузы. Можно видеть, что при всех трех значениях паузы испытуемый устанавливал ее так, что конец паузы находился на одном и том же расстоянии от краев стимула, т. е. попадал на одно и то же значение  $F_2$ .

Прямое доказательство того, что в качестве характеристики [г]-паузы выступает значение  $F_2$  в момент ее осуществления, было получено в экспериментах с дихотической стимуляцией [64]. На одно ухо подавался стимул с изменяющейся во времени  $F_2$ , который не прерывался. На другое ухо одновременно подавался второй стимул, в котором вторая форманта вообще отсутствовала, а частоты  $F_1$  и  $F_3$  не изменялись во времени. Этот второй стимул прерывался паузой (12—15 мс), положение которой устанавливалось испытуемым. Оказалось, что испытуемые легко определяют границу между [г] и [г'], причем значение  $F_2$ , приходящееся на конец паузы, находится в районе 1600 Гц, что хорошо согласуется со значениями границы по твердости / мягкости, определенной для стационарных изолированных [s] и [ʃ] [167].

#### 6.2.2. ВОСПРИЯТИЕ НАЧАЛЬНЫХ ЗВОНКИХ И ГЛУХИХ СМЫЧНЫХ СОГЛАСНЫХ РУССКИМИ СЛУШАТЕЛЯМИ

При обсуждении данных, касающихся восприятия начальных звонких и глухих согласных, целесообразно выделить два вопроса. Один из них касается условий обнаружения согласного в слого, т. е. определения тех минимальных требований, которым должен удовлетворять стимул для того, чтобы он был воспринят как слог СГ, а не Г. Второй вопрос касается различия слогов с начальным глухим или звонким согласным.

Наиболее элементарный стимул, воспринимаемый носителями русского языка как слог, начинающийся со звонкого смычного согласного, представляет собой последовательность из двух стационарных сегментов. Второй сегмент соответствует гласному, первый может быть чистым тоном, равным или более низким по частоте, чем первая гармоника гласного ( $F_0$ ). Критическим параметром стимула является длительность интервала от начала первого сегмента до начала гласного. Порог (50% идентификации стимула, как слога, начинающегося со звонкого согласного) соответствует 18—20 мс при  $F_0=300$  Гц [68].

Такую же величину имеет (при  $F_0=300$  Гц) пороговая длительность сегмента [m] в элементарном двухсегментном синтетическом слого [ma]. При  $F_0=200$  Гц пороговая длительность увеличивается примерно до 23 мс [65].

Наиболее элементарный стимул, воспринимаемый русскими слушателями как слог, начинающийся с глухого согласного ([p], [t], [k]), является последовательностью из сегмента гласного

и предшествующей ему посылки шума. Пороговый интервал от начала шума до начала гласного, насколько нам известно, не определялся, однако при интервале в 10—15 мс восприятие [p] и [t] является уже вполне отчетливым [81].

Существенно, что для восприятия согласного сегмент гласного столь же необходим, как и первый (тональный или шумовой) сегмент. Другая существенная особенность состоит в том, что сегменты не обязательно должны примыкать друг к другу — наличие короткой паузы не мешает восприятию. Изменение длительности первого тонального сегмента в звонких согласных, если эта длительность превосходит порог, возможно в очень широких пределах.

Таким образом, используя даже такие элементарные двухсегментные стимулы, возможно создать множество разных, различающихся между собой картинок, которые, тем не менее, будут восприниматься как слог с тем же самым начальным согласным.

Связь между стимулом и воспринимаемым согласным легко описать, если воспользоваться такими понятиями, как события («появление голоса», «появление шума», «появление гласного») и отношения порядка следования («следует за»). Во всех слогах с начальным глухим согласным выполняется условие: «появление гласного» «следует за» «появлением шума». Во всех слогах с начальным звонким согласным выполняется условие: «появление гласного» «следует за» «появлением голоса».

Однако легко заметить, что приведенные описания глухих и звонких согласных не являются взаимоисключающими, т. е. не являются полными (начальный согласный не может быть одновременно и звонким, и глухим).

Дополнительные условия могут быть определены из результатов экспериментов по восприятию синтетических стимулов, различающихся по параметру «время начала голоса» (см. главу 4). На рис. 4.2 приведены схематические изображения двух крайних в ряду из набора стимулов, синтезированных в Хаскинских лабораториях. Напомним, что «время начала голоса» определяется как интервал между началом шума взрыва и появлением голоса. Отрицательные значения параметра соответствуют опережению голоса.

Указанные стимулы были применены в работе Авакяна, исследовавшего положение фонемных границ между глухими и звонкими согласными у русских слушателей. Значения границ оказались следующими: —18.6 мс ([b]—[p]); —6.3 мс ([d]—[t]); —10.9 мс ([g]—[k]). В разделе 6.2.5 будут приведены данные, показывающие, что задержка обнаружения начала сигнала сильно зависит от его физических свойств. В частности, задержка обнаружения начала тона резко возрастает на низких частотах. Так как в стимулах, синтезированных в Хаскинских лабораториях, частота основного тона составляла 100 Гц, естественно ожидать, что задержка обнаружения начала голоса должна была превышать задержку обнаружения шума взрыва.



Приведенные выше значения границы отрицательны и малы по абсолютной величине. Это заставляет думать, что наиболее вероятное объяснение состоит в том, что граница соответствует просто условию одновременности обнаружения появления голоса и появления шума. Иначе говоря, русский слушатель выбирает тот согласный, характерное для которого событие обнаруживается первым. Для того чтобы сформировать логическую схему, которая воспроизводила бы такое поведение, нужно еще ввести такое понятие, как «первое событие в последовательности».

### 6.2.3. РАЗЛИЧИЕ НАЧАЛЬНЫХ ЗВОНКИХ И ГЛУХИХ СМЫЧНЫХ СОГЛАСНЫХ НОСИТЕЛЯМИ АНГЛИЙСКОГО ЯЗЫКА

Представление о том, что для описания распознавания начальных звонких и глухих английских согласных необходимо учитывать три типа акустических событий и порядок их появления, было выдвинуто Стивенсом и Клаттом [488].

Хотя английские [b], [d], [g] могут начинаться с появления голоса, это не является их обязательным признаком. Значение границы между глухими и звонкими согласными по параметру «время начала голоса» всегда больше +20 мс и сильно зависит как от места образования согласного, так и от конкретных характеристик синтетических стимулов, применявшихся в исследовании [535].

Типичные динамические спектрограммы естественных [b], [d], [g] характеризуются начальным шумом взрыва, затем появлением голоса и отчетливо выраженным формантным переходом. Общим для [b], [d], [g] является повышение  $F_1$ .

Начальные глухие согласные, особенно в положении перед ударным гласным, являются отчетливо аспирированными ([p<sup>h</sup>], [t<sup>h</sup>], [k<sup>h</sup>]). Включение голоса значительно задержано относительно начала шума взрыва и происходит в тот момент времени, когда артикуляторный тракт уже принял характерное для данного гласного положение. Соответственно начального перехода  $F_1$  в этом случае нет.

Исходя из этих различий в акустической картине, Стивенс и Клатт предложили выделять три события: а) скачкообразное появление (и быстрое изменение) энергии в средне- или высокочастотной области спектра — взрыв, б) появление энергии в низкочастотной области спектра, в) формантный переход, т. е. перемещение спектрального максимума.

В отношении первых двух событий используется правило, что глухой согласный может быть воспринят только в том случае, если второе событие «следует за» первым событием, т. е. первое событие должно быть обнаружено раньше, чем будет обнаружено



второе событие. Третье событие (формантный переход) является положительным признаком звонкого согласного.

При экспериментальной проверке этой гипотезы использовалось дополнительное предположение, что для обнаружения формантного перехода необходимо некоторое время его существования. Иначе говоря, быстрые переходы должны обнаруживаться хуже, чем медленные. Были синтезированы стимулы, различающиеся длительностью формантных переходов и длительностью интервала от начала шума до начала голоса. Полученные данные подтвердили исходную гипотезу, они показали, что оба параметра являются существенными. Наименьшие значения границы между [d] и [t<sup>h</sup>] по признаку «время начала голоса» были получены для стимулов с наиболее коротким формантным переходом.

Очень интересным результатом работы является установление различия в «весах» проверявшихся признаков у разных испытуемых. Это является одним из доводов в пользу того, что стимул действительно представлен как последовательность выделенных в нем событий, а не как последовательность временных отсчетов его спектра или параметров.

#### 6.2.4. РАСПОЗНАВАНИЕ ПОРЯДКА СЛЕДОВАНИЯ ФОНЕМ В СОЧЕТАНИЯХ СОГЛАСНЫХ

Сочетания согласных — «кластеры», — отличающиеся друг от друга порядком следования одних и тех же фонем, чрезвычайно интересны по той причине, что гипотеза картинок и гипотеза потока событий предсказывают разную трудность (или легкость) их распознавания.

Например, динамические спектрограммы сочетаний [ats] и [ast] очень сильно отличаются друг от друга, и, пользуясь соответствующими картинками, спутать [ast] и [ats] неизмеримо труднее, чем спутать [ast] и [ask]. Если распознавание основывается на последовательности событий, то ситуация меняется. Различение [ast] и [ats] может оказаться более трудной задачей, так как в обоих сочетаниях имеют место практически одни и те же события; разница заключается в порядке их следования. Для различения порядка следования необходимо располагать специальной логической схемой. Различение [ast] и [ask] может основываться на разнице в самих событиях.

В экспериментах Бонда [194] стимулами служили 15 пар коротких английских слов. Слова в паре отличались друг от друга только порядком следования глухого смычного согласного и [s] ([ps]—[sp], [ts]—[st], [ks]—[sk]). Из этих 30 слов, следующих друг за другом в случайном порядке, были составлены таблицы, которые были прочтены диктором — носителем английского языка. Испытуемые прослушивали пленку и записывали услышанные слова. При анализе материала все ошибки разделялись на три

группы: перестановки (например, [ts] превратилось в [st]), перестановки вместе с заменами (например, переход [ts] в [sk]), замены (например, переход [ts] в [ks]).

Полученные данные приведены на рис. 6.18. Можно видеть, что наиболее частой ошибкой является перестановка.

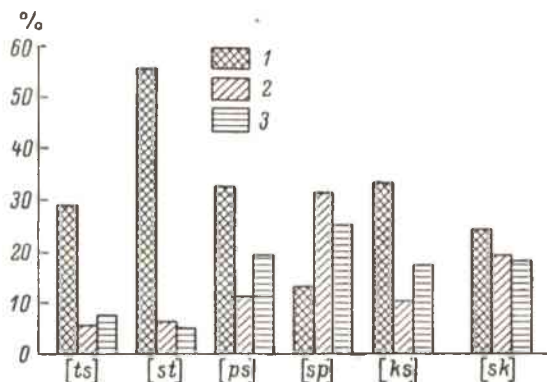


Рис. 6.18. Ошибки при распознавании сочетаний согласных. По [184].

Процент ошибок трех типов: 1 — перестановки, 2 — перестановки вместе с заменами согласных, 3 — замены.

Этот результат практически невозможно объяснить с позиций гипотезы картинок. Вместе с тем он хорошо согласуется с приводимыми ниже данными о том, что распознавание порядка следования элементов в быстрых неречевых последовательностях является трудной задачей для человека и требует специального обучения.

#### 6.2.5. ВРЕМЕННОЙ ПОРОГ ОБНАРУЖЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ

Данные о минимальном интервале между событиями *A* и *B*, при котором человек может установить отношение «*B* следует за *A*», были получены в работе Бюрка и др. [202] и работе Хирша [297]. Бюрк и др. применяли сложные стимулы, образованные путем сложения двух чистых тонов с частотами  $f_1$  и  $f_2$ ; один из тонов следовал за вторым с регулируемой задержкой по времени. Испытуемые сравнивали стимул, начинающийся тоном с частотой  $f_1$ , со стимулом, начинающимся тоном с частотой  $f_2$ . Определялось пороговое значение задержки, при котором испытуемые различали стимулы. Заметим, что такие стимулы легко ассоциируются со слогами СГ, различающимися по характеру начального согласного.

Оказалось, что если частоты тонов не слишком близки (различаются больше, чем примерно на 10%), порог практически не зависит от расстояния между тонами по частоте. Вместе с тем наблю-

дается отчетливая зависимость величины пороговой задержки от абсолютного положения стимула по оси частот. Такая зависимость, определенная для стимула с  $f_2/f_1=1.1$ , приведена на рис. 6.19. Можно видеть, что порог минимален в области частот 500—4000 Гц. В области низких частот порог закономерно возрастает с понижением частоты: на частоте 100 Гц он примерно в 3.5 раза превышает порог на 1000 Гц. Авторы нашли также, что порог отчетливо зависит от постоянной времени включения то-

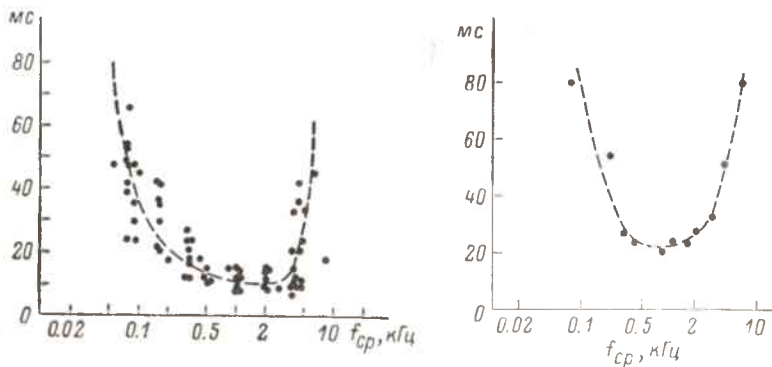


Рис. 6.19. Величина порогового сдвига по времени между началом двух тонов как функция частоты тонов. По [202].

По оси абсцисс — средняя частота тонов в паре,  $f_{cp}=(f_1+f_2)/2$ ; по оси ординат — пороговый сдвиг.

Рис. 6.20. Зависимость порогового сдвига от средней частоты при постоянной времени включения тонов, равной 12 мс. По [202].

Обозначения те же, что на рис. 6.19.

нов — он минимален при резком включении и возрастает с увеличением постоянной времени.

На рис. 6.20 приведена зависимость порога от частоты, определенная при постоянной времени включения, равной 12 мс ( $f_2/f_1=1.1$ ). Можно видеть, что на частоте 1000 Гц порог возрастает примерно вдвое по сравнению с тем случаем, когда применялось резкое включение (рис. 6.19).

В экспериментах Хирша [297] применялись комплексы из двух тонов, частоты которых находились в диапазоне 250—4800 Гц. Постоянная времени нарастания составляла 20 мс. Полученные значения порога находились в районе 20 мс. Хирш применял также комбинации из щелчка и шума, щелчка и тона и двух различающихся по спектру щелчков. Приводимый в статье [297] график свидетельствует о том, что для щелчков значение пороговой задержки отчетливо меньше 20 мс.

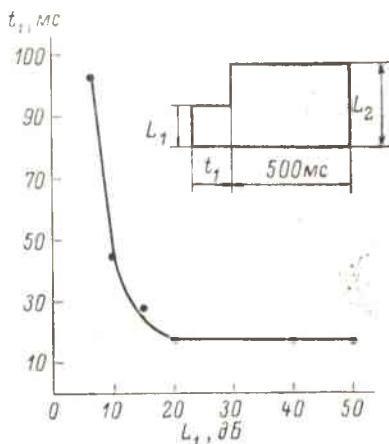
Пороговая задержка находится на интервале 10—20 мс и в том случае, если она определена как та минимальная длительность первого, более слабого сегмента в двухсегментном тональном

стимуле (см. рис. 6.21), при которой испытуемый обнаруживает присутствие этого первого сегмента. Согласно представлениям, изложенным в разделе 6.1.3, стимул такого типа должен приводить к возникновению двух следующих друг за другом марок начала. В таком случае можно ожидать, что пороговая задержка отражает или разрешающую способность «схемы», вырабатывающей марки начала, или характеристики «схемы», отличающей «последовательные» события от «одновременных».

Данные, приведенные на рис. 6.21, получены для тона 1000 Гц; применялось экспоненциальное сглаживание фрон-

Рис. 6.21. Пороговый интервал между началом тона и моментом увеличения его интенсивности как функция уровня ощущения стимула.

По оси абсцисс — начальный уровень интенсивности ( $L_1$ ) в дБ от порога ощущения; по оси ординат — пороговая длительность первого сегмента. Вверху — схематическое изображение стимула. Приращение уровня интенсивности ( $L_2 - L_1$ ) равно 20 дБ.



тов с постоянной времени 5 мс. Точки соответствуют средним для пяти испытуемых значениям порога. Результаты показывают, что когда уровень интенсивности первого сегмента превышает 20 дБ над порогом ощущения, пороговая задержка не зависит от уровня интенсивности.

Из приведенных в настоящем разделе данных можно сделать вывод, что в среднем диапазоне частот и при отчетливо надпороговых уровнях интенсивности звука пороговый интервал зависит преимущественно от формы фронтов стимула.

Если слуховая система обнаруживает (выделяет) такие события, как начало звука, очевидно, что форма (скорость изменения) огибающей сигнала должна иметь значение.

#### 6.2.6. РАСПОЗНАВАНИЕ ПОРЯДКА СЛЕДОВАНИЯ ЗВУКОВЫХ ПОСЫЛОК В БЫСТРЫХ НЕРЕЧЕВЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ

Частота следования звуков в речевых последовательностях, при которой человек может распознать (идентифицировать) каждый из элементов, оказывается или значительно более низкой, чем частота следования слогов, необходимая для фонемного распознавания гласных, или примерно такой же.

Существенно большая длительность периода следования, чем это необходимо в случае гласных, наблюдается в том случае, когда

звуки, входящие в последовательность, резко отличаются друг от друга и воспринимаются как независимые сигналы [157, 520, 521].

Если элементами последовательности являются тоны с не очень сильно различающимися частотами, то человек воспринимает последовательность как мелодию, и временные ограничения оказываются примерно такими же, как в случае гласных [199, 238].

Естественно, что если бы мы захотели сделать действующую модель фонетического интерпретатора, работающую в реальных акустических условиях, мы столкнулись бы с необходимостью задать правила, по которым интерпретатор будет или принимать новый отрезок сигнала за новый слог в том же самом сообщении и записывать информацию о нем в табл. 6.1, или отвергать его как посторонний сигнал (речевой сигнал от другого диктора, неречевая помеха).

С этой точки зрения распознавание последовательности неречевых резко различных звуков можно скорее сравнивать с распознаванием последовательности дикторов по последовательности гласных, взятых из речи этих дикторов. Такого эксперимента, насколько нам известно, пока никто не произвел.

Наиболее интересный факт, касающийся неречевых последовательностей, состоит в том, что если элементы следуют быстро друг за другом и каждый элемент в отдельности не распознается, последовательность воспринимается как некоторый качественно новый сигнал. Он может напоминать естественные звуки, такие как стрекотание сверчка, чириканье и т. д. Это дает возможность использовать быстрые неречевые последовательности для исследования кардинального вопроса о том, что же такое для человека один звук, какими параметрами он описывается и чему именно учится человек, обучаясь распознавать разные звуки.

За последние годы появилось несколько работ, показывающих, что человек может научиться различать и абсолютно опознавать быстрые последовательности звуковых посылок, отличающиеся друг от друга только порядком следования посылок [233, 348, 519], однако для этого необходима тренировка, которая может занимать несколько дней.

Опишем подробно эксперимент Уоррена [519], направленный на исследование способности обучаться распознаванию быстрых последовательностей. Испытуемых учили абсолютно опознавать две четырехэлементные последовательности, отличающиеся порядком следования двух средних элементов:  $N$  — посылки шума с полосой 500—4000 Гц и  $SW$  — посылки из прямоугольных импульсов с частотой 1 кГц, пропущенных через фильтр 500—4000 Гц. В качестве первого и последнего элемента обеих последовательностей использовались посылки тона с частотой 2.5 кГц —  $T$ .

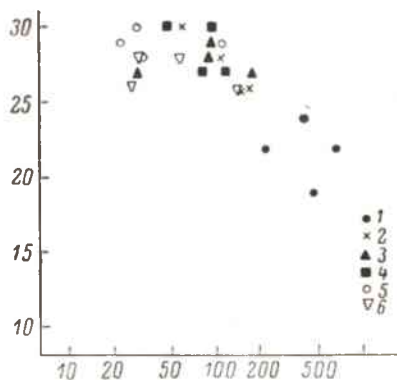
Во время периода обучения перед испытуемым находились две кнопки, на которых были написаны названия последовательностей:  $T, N, SW, T$  и  $T, SW, N, T$ . Нажимая соответствующую кнопку, испытуемый вызывал ту последовательность, которую он хотел прослушать, и тренировался столько времени, сколько ему



казалось необходимым. Когда испытуемый считал, что он подготовился, он сообщал об этом экспериментатору, который предъявлял ему тестирующий набор сигналов, состоящий из 30 последовательностей, следующих в случайном порядке. После каждого из сигналов тестирующего набора испытуемый сообщал название последовательности и получал от экспериментатора сообщение о том, был ли он прав или ошибся. Если испытуемый правильно опознал 25 из 30 предъявленных сигналов, эксперимент считался законченным. В противном случае испытуемому предлагалось тренироваться дальше, и затем снова проводилось тестирование.

Рис. 6.22. Характеристики распознавания четырехэлементных последовательностей. По [519].

По оси абсцисс — число предъявлений стимулов в период тренировки; по оси ординат — число правильных ответов при тестировании. Разные значки соответствуют разным длительностям элементов в последовательности: 1 — 5 мс, 2 — 10 мс, 3 — 10 мс и пауза в 5 мс, 4 — 20 мс, 5 — 50 мс, 6 — 100 мс.



Всего допускалось пять повторных тестирований. Если и при пятом тестировании испытуемый не достигал критерия (25 правильных ответов из 30), то эксперимент все равно заканчивался.

В результате Уоррен [519] характеризовал распознавание двумя параметрами: числом правильных ответов при последнем тестировании  $N$  ( $N \leq 30$ ) и числом предъявлений сигналов ( $n$ ) в период тренировки.

В экспериментах приняло участие четверо испытуемых; применялось 6 вариантов последовательностей: в пяти из них звуковые посылки непосредственно примыкали друг к другу, и варианты различались между собой значением длительности посылки (все посылки внутри последовательности были одинаковой длительности). В одном варианте между посылками имелись паузы в 5 мс.

Полученные Уорреном данные приведены нами в графической форме (рис. 6.22). Можно видеть, что, когда длительность посылки составляет 5 мс, испытуемые практически не научаются распознаванию последовательностей, хотя период тренировки превышает 200—600 предъявлений. При длительностях посылок 10—100 мс испытуемые научаются распознаванию.

Из графика видно также, что для более медленных последовательностей (длительность посылки 50 и 100 мс)  $n < 60$  в 6 из 8 случаев. Для быстрых последовательностей (длительности 10 и 20 мс)  $n > 60$  в 9 из 12 случаев. К сожалению, из этих данных еще нельзя сделать вывода, что обучение распознаванию более медленных



последовательностей происходит быстрее. Так как в экспериментах испытуемые сначала учились распознавать быстрые последовательности, а потом все более медленные, возможно, что наблюдаемый эффект отражает перенесение опыта.

В кратком сообщении о работе по обучению распознаванию последовательностей тональных посылок («Высокий, Средний, Низкий, Высокий» и «Высокий, Низкий, Средний, Высокий» тоны) отмечается, что порог, соответствующий 10—20 мс, достигается на восьмой день тренировки [233].

Заметим, что в экспериментах Уоррена применялись стимулы со спектром в средней частотной области и с резким включением. Полученный порог различения последовательности (10 мс) близок к тому, что было найдено Бюрком и др. (рис. 6.19).

Естественно думать, что обнаружение акустических событий, соответствующих  $N$  и  $SW$  в экспериментах Уоррена, обеспечивается слуховой обработкой сигнала, которая является стандартной и не зависит от того, обучен человек или нет. В таком случае обучение может заключаться в формировании схемы, реализующей правила: если  $SW$  «следует за»  $N$ , то  $A$ ; если  $N$  «следует за»  $SW$ , то  $B$ .

### 6.3. РЕАЛЬНОСТЬ АКУСТИЧЕСКИХ СОБЫТИЙ

Данные, приведенные в разделе 6.1, указывают на то, что модель фонетической интерпретации включает специальную программу (механизм) сегментации речевого потока и что сигналами, используемыми этой программой, являются марки границ, т. е. события, соответствующие началу и окончанию гласноподобного отрезка. Данные, приведенные в разделе 6.2, показывают, что использование таких понятий, как акустические события и отношения порядка следования, удобны для описания связи между сигналом и его фонетической интерпретацией. Следовательно, можно утверждать, что гипотеза потока событий не столь уж фантастична и достойна дальнейшего рассмотрения.

Однако, приняв эту гипотезу, мы сразу же сталкиваемся с необходимостью определить, что такое событие. Формально определить события можно лишь описав алгоритм (механизм) их обнаружения. Так как вся психоакустика основывалась до сих пор на допущении, что «ощущение является непрерывной функцией времени», ни прямых данных о слуховом обнаружении событий, ни разработанных экспериментальных подходов к этой проблеме пока, естественно, нет. Обсуждение некоторых конкретных вопросов, возникающих при попытках определить возможные классы (типы) и параметры событий, содержится дальше.

Таким образом, сейчас акустические события реальны в том смысле, что без их допущения трудно обойтись при построении модели фонетической интерпретации, но они же нереальны в том смысле, что их пока нельзя ни описать, ни даже перечислить.

## СЛУХОВАЯ ОБРАБОТКА РЕЧЕВОГО СИГНАЛА

### Глава 7

#### СПЕКТРАЛЬНЫЙ СЛУХОВОЙ АНАЛИЗ И ЕГО МОДЕЛИРОВАНИЕ

В настоящей главе рассматриваются экспериментальные данные относительно свойств периферической слуховой системы как спектрального анализатора и вопросы, связанные с моделированием этого анализатора.

То, что эти сведения нужны для специалистов, исследующих восприятие речи или занимающихся его моделированием, не требует доказательств. В общем виде известно, что полезная информация, обеспечивающая восприятие речевого сигнала, заключена в его динамическом спектре. Однако, какова именно эта информация и как конкретно представляется сам динамический спектр, зависит от свойств анализатора. Применив для анализа речи прибор, в чем-то существенно отличающийся от слухового спектрального анализатора, мы наверняка потеряем какую-то часть входной информации, а какую-то часть, наоборот, представим с излишними подробностями.

Необходимость использования для исследования восприятия речи анализаторов, по своим характеристикам близких к слуховому, в последнее время признается все большим числом специалистов [130, 211, 319, 543]. При этом в большинстве случаев, в принципе, нужна лишь функциональная модель, формально воспроизводящая слуховой спектральный анализ.

Если бы слуховой спектральный анализатор представлял собой линейную систему, описание такой функциональной модели можно было бы свести к указанию амплитудно-частотных и фазово-частотных характеристик всех слуховых фильтров; при этом характеристикой каждого фильтра была бы сквозная характеристика ряда последовательно соединенных частотно-зависимых элементов реальной системы.

В настоящее время, однако, имеются достаточно убедительные свидетельства о том, что система слухового спектрального анализа является нелинейной и ее частотно-избирательные свойства изменяются с изменением интенсивности входных сигналов. При опи-

сании свойств такой системы представляется более целесообразным рассмотреть отдельно характеристики составляющих ее последовательно включенных элементов — систем наружного, среднего и внутреннего уха — тем самым, по возможности, локализовав нелинейность.

В дальнейшем мы будем пользоваться допущением, что выходными сигналами «фильтров» слухового спектрального анализатора являются сигналы, существующие еще в аналоговой форме и логически предшествующие возникновению нервных импульсов. Вопрос о характеристиках преобразования этих аналоговых сигналов в нервную импульсацию будет рассмотрен в следующей главе.

### 7.1. СИСТЕМА НАРУЖНОГО УХА

К системе наружного уха принято относить (рис. 7.1) ушную раковину и наружный слуховой проход, который заканчивается барабанной перепонкой.

Свойства системы наружного уха при стационарном гармоническом входном воздействии могут быть описаны с помощью ампли-

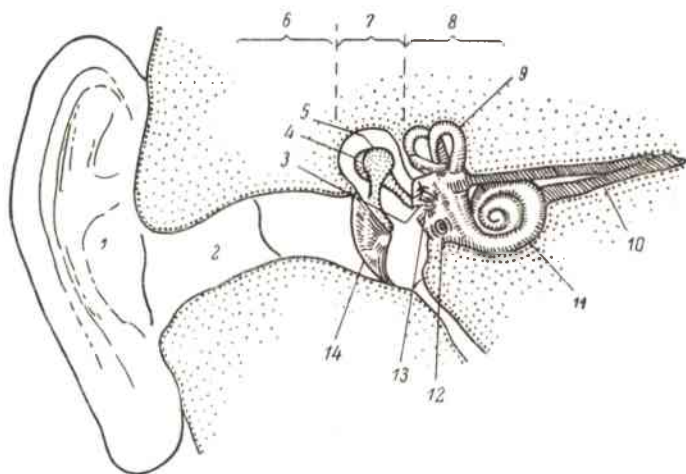


Рис. 7.1. Схематическое изображение уха человека. По [137].

1 — ушная раковина; 2 — наружный слуховой проход; 3, 4, 5 — слуховые косточки, соответственно стремечко, наковаленка и молоточек; 6 — наружное ухо; 7 — среднее ухо; 8 — внутреннее ухо; 9 — вестибулярный аппарат; 10 — слуховой нерв; 11 — улитка; 12 — круглое окно; 13 — овальное окно; 14 — барабанная перепонка.

тудно-частотной характеристики,  $P_0(\omega)/P_{вх}(\omega)$ , связывающей амплитуду звукового давления у барабанной перепонки,  $P_0(\omega)$ , с амплитудой звукового давления,  $P_{вх}(\omega)$ , в свободном акустическом поле в точке, соответствующей центру головы испытуемого. Эту характеристику, выраженную в логарифмическом масштабе,

в свою очередь можно представить в виде суммы двух характеристик:

$$20 \lg \frac{P_0(\omega)}{P_{\text{нх}}(\omega)} = 20 \lg \frac{P_0(\omega)}{P_{\text{нх}}(\omega)} + 20 \lg \frac{P_0(\omega)}{P_0(\omega)}, \quad (1)$$

где  $P_0(\omega)$  — амплитуда звукового давления в начале наружного слухового прохода.

Зависимость амплитуды звукового давления в начале наружного слухового прохода от частоты при постоянной амплитуде звукового давления  $P_{\text{нх}}(\omega)$  неоднократно измерялась [235, 523] и приведена на рисунке 7.2 (кривая 1). Эта характеристика учиты-

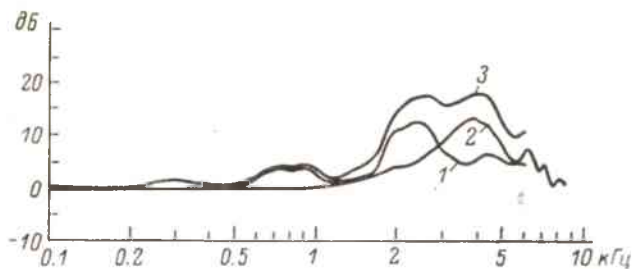


Рис. 7.2. Амплитудно-частотные характеристики системы наружного уха. Но [528].

По оси абсцисс — частота; по оси ординат — относительная амплитуда. 1 — амплитудно-частотная характеристика ушной раковины, определяемая как  $P_0(\omega)/P_{\text{нх}}(\omega)$ , 2 — наружного слухового прохода,  $P_0(\omega)/P_0(\omega)$ , 3 — всей системы наружного уха,  $P_0(\omega)/P_{\text{нх}}(\omega)$ .

вает также дифракцию акустических волн вокруг головы. Необходимо отметить, что форма этой характеристики на высоких частотах несколько изменяется с изменением ориентации головы испытуемого относительно источника звука. Частотная характеристика  $P_0(\omega)/P_0(\omega)$  также измерялась и представлена на рис. 7.2 (кривая 2). Кривая 3 показывает суммарную характеристику.

Система наружного уха линейна. Поэтому в большинстве встречающихся на практике случаев и, в частности, при моделировании процессов, происходящих в органе слуха при восприятии речевых сигналов, оказывается достаточным представить систему наружного уха в виде некоторого линейного четырехполюсника, имеющего ту же частотную характеристику, что и система наружного уха.

Ввиду того что частотная характеристика наружного уха обладает сравнительно малой неравномерностью, влиянием наружного уха на вид результирующей характеристики органа слуха нередко вообще пренебрегают [137].

## 7.2. СИСТЕМА СРЕДНЕГО УХА

К системе среднего уха (рис. 7.1, см. также [390, 392]) принято относить барабанную перепонку и три слуховые косточки: молоточек, наковаленку и стремечко, расположенные в костной полости, заполненной воздухом. Молоточек одним своим концом прикреплен к барабанной перепонке и через наковаленку и стремечко передает колебания барабанной перепонки мембране овального окна внутреннего уха; при этом образуется система рычагов с соотношением плеч, равным для уха человека  $1.3 : 1$ . Поскольку при этом еще и эффективная поверхность барабанной перепонки примерно в 14 раз больше площади основания стремечка, это приводит, вследствие трансформации давления, к значительно лучшим условиям передачи энергии из воздушной среды в каналы улитки, заполненные жидкостью, чем если бы колебания передавались непосредственно по воздуху [287, 390]. Удаление слуховых косточек, однако, приводит к значительно большим потерям слуха, чем можно было бы ожидать исходя лишь из нарушения трансформации давления. Объясняется это тем, что колебания давления воздуха в этом случае подаются синфазно на оба (круглое и овальное) окна улитки и в результате вызывают значительно меньшие амплитуды колебания заключенной в улитке жидкости.

Измерения, выполненные на препаратах человеческого уха [187, 228], показывают, что в диапазоне частот от 25 до 2000 Гц амплитуда колебаний стремечка линейно возрастает с повышением уровня звукового давления входных сигналов до 100 дБ над порогом. Однако в полости среднего уха имеются две мышцы, одна из которых (*m. tensor tympani*) подсоединена к молоточку, а другая (*m. stapedius*) — к стремечку. При подаче на ухо сильных звуков у человека возникает сокращение *m. stapedius*, в результате чего характеристики системы изменяются (см. обзор: [392]). Этот эффект, по-видимому, и является основным источником нелинейности в среднем ухе человека.

Пороговый уровень входного тонального сигнала, при котором обнаруживается рефлекторное сокращение мышц среднего уха у человека [381, 417], составляет около 80 дБ над абсолютным порогом слуха при частотах сигналов от 250 до 4000 Гц. Скрытый период рефлекса в ответ на импульсный сигнал сильно зависит от интенсивности входного сигнала и у человека лежит в пределах от 150 мс при пороговых уровнях до 25—35 мс при больших интенсивностях [392]. С учетом этих данных приходится считать, что система среднего уха человека линейна при уровнях звукового давления входных сигналов не выше 80 дБ над порогом.

Измерения частотной зависимости амплитуды колебания стремечка, а также других косточек среднего уха относительно амплитуды звукового давления у барабанной перепонки [287, 309, 387, 445], несмотря на различия, характерные для разных животных, согласуются в том, что ухо в диапазоне низких и средних частот



(примерно до 3000 Гц) может рассматриваться как минимально-фазовая система, амплитудно-частотные характеристики которой практически горизонтальны от низких частот до 800—2000 Гц, а затем начинают спадать (рис. 7.3) с крутизной около 12 дБ на октаву [287, 387].

Такой спад частотной характеристики среднего уха прекращается на частотах около 3 кГц, где начинает сказываться влияние полости среднего уха и неравномерность частотной характери-

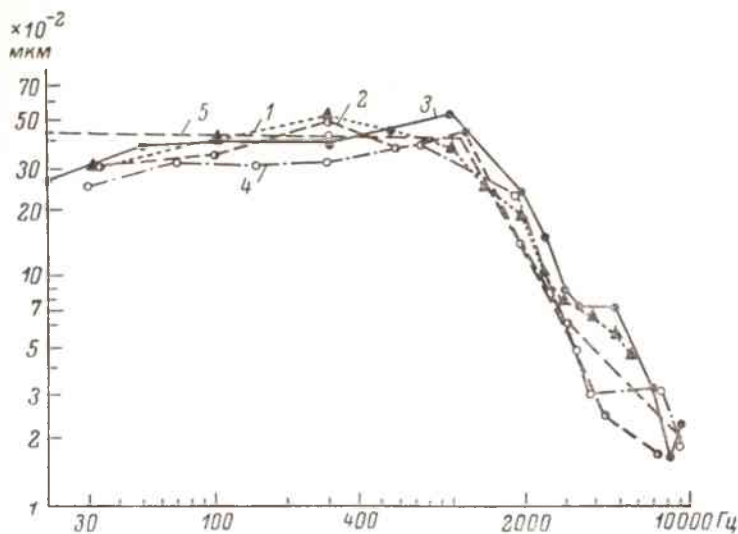


Рис. 7.3. Зависимость от частоты амплитуды колебания стремечка при постоянном звуковом давлении сигнала (120 дБ над порогом) у барабанной перепонки уха кошки. По [287].

По оси абсцисс — частота; по оси ординат — размах. 1, 2, 3, 4 — результаты измерений на четырех кошках; 5 — аппроксимирующая характеристика.

стики самой барабанной перепонки [506]. В этой области частот (от 3 кГц и выше) отмечается ряд максимумов [392] и общий наклон характеристики уменьшается примерно до  $-6$  дБ на октаву. На высших частотах воспринимаемого диапазона характеристика среднего уха, по-видимому, имеет значительно больший наклон.

Амплитудно-частотные характеристики колебания молоточка, наковаленки и стремечка, измеренные при постоянном уровне звукового давления у барабанной перепонки, на низких и средних частотах в значительной мере совпадают, и, следовательно, система косточек колеблется как единое целое [387]. На более высоких частотах (выше 3 кГц) начинает сказываться различие в массе и податливости отдельных косточек; кроме того, сама барабанная перепонка перестает колебаться как жесткий поршень и на ней



образуются стоячие волны [506]. Анализ свойств системы в этом случае значительно затрудняется и по существу до настоящего времени никем не выполнялся.

### 7.3. СИСТЕМА УЛИТКИ ВНУТРЕННЕГО УХА

Внутреннее ухо человека представляет собой сложный лабиринт соединенных друг с другом каналов (рис. 7.4), расположенных внутри височной кости и заполненных жидкостью. Часть костного лабиринта вместе с разделяющими мембранами и

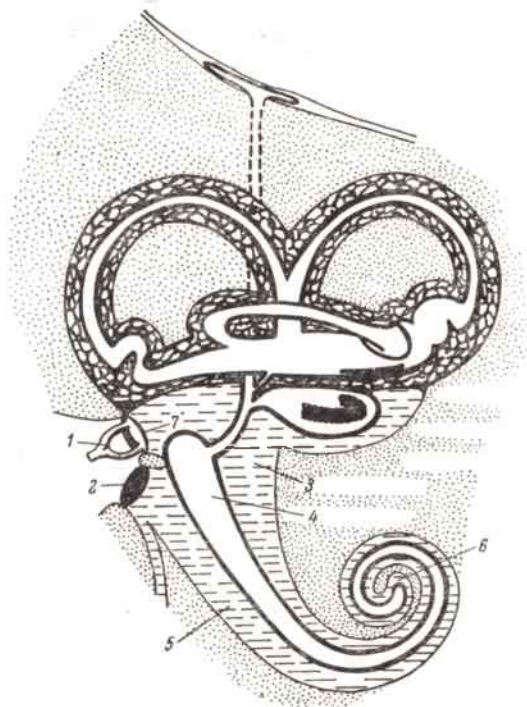


Рис. 7.4. Схематическое изображение строения внутреннего уха млекопитающих. По [203].

1 — стремечко, 2 — круглое окно, 3 — вестибулярный канал, 4 — средний, или кохлеарный, канал, 5 — тимпанальный канал, 6 — геликотрема, 7 — овальное окно.

рецепторными клетками образует воспринимающий отдел вестибулярной системы; один из каналов, имеющий спиралевидную форму, улитка, относится к слуховой системе.

Костный канал улитки человека имеет длину около 35 мм и с одной стороны (базальной) заканчивается двумя отверстиями (круглым и овальным окнами), выходящими в полость среднего уха и отделенными от нее эластичными мембранами. По мере уда-

ления от базальной части к вершине, или апикальному концу, площадь поперечного сечения канала улитки уменьшается примерно в 6 раз [8, 548].

Расположенные вдоль улитки две мембраны — Рейсперова и базилярная — делят ее на три части, или канала: тимпанальный, вестибулярный и средний (рис. 7.5). При этом тимпанальный и вестибулярный каналы соединяются у вершины улитки небольшо-

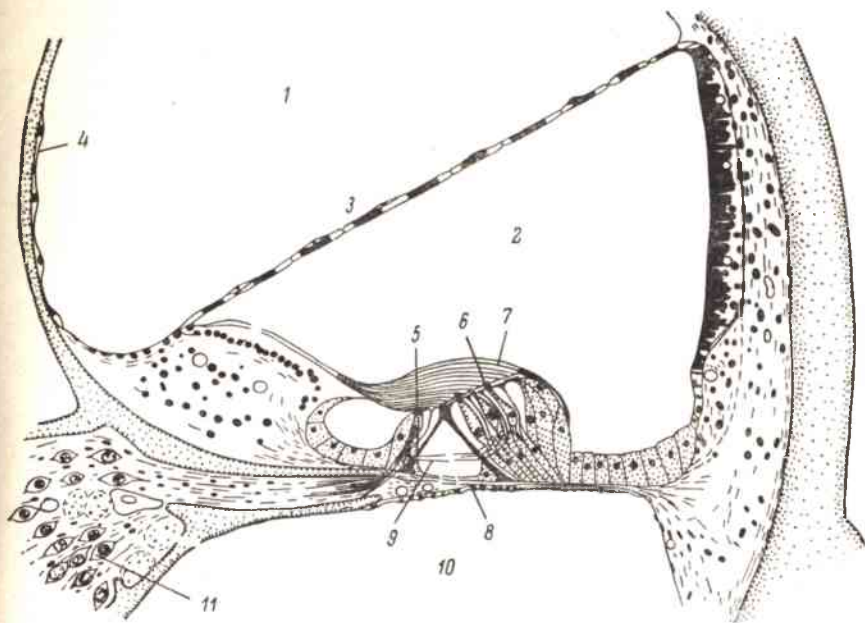


Рис. 7.5. Поперечный разрез улитки. По [54].

1 — вестибулярный канал, 2 — средний канал, 3 — Рейсперова мембрана, 4 — костная стенка улиткового канала, 5 — внутренние волосковые клетки, 6 — наружные волосковые клетки, 7 — текториальная мембрана, 8 — базилярная мембрана, 9 — нервные волокна, 10 — тимпанальный канал, 11 — клетки спирального ганглия.

шим отверстием (геликотремой) и заполнены одинаковой жидкостью (перилимфой) со свойствами, близкими к свойствам плазмы крови.

Средний канал образует замкнутую полость, заполненную более вязкой жидкостью (эндолимфой) отличного от перилимфы химического состава. В этом канале находится рецепторный аппарат улитки — кортиев орган, представляющий у человека систему из одного ряда внутренних и трех рядов внешних волосковых клеток, образуемых двумя группами различных рецепторов [476].

Чувствительные клетки равномерно распределены вдоль улитки и расположены между базилярной и покровной, или текториальной, мембранами. Каждая внутренняя волосковая клетка (всего их около 3500) соединяется посредством синаптических

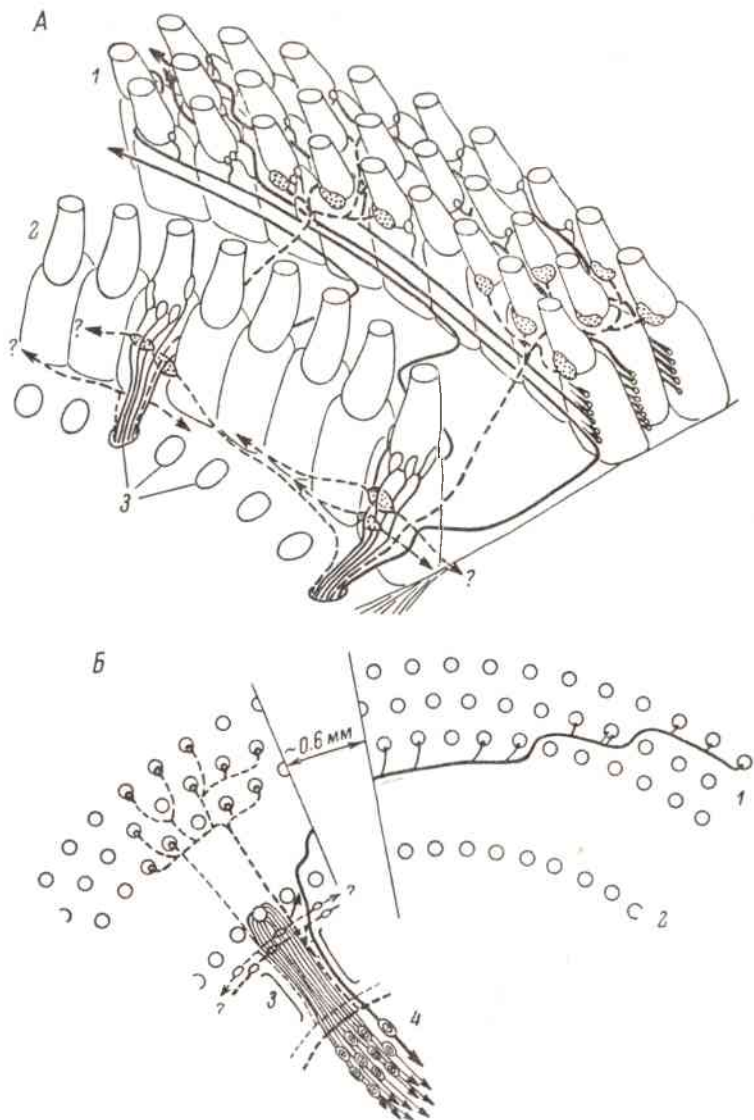


Рис. 7.6. Схема иннервации Кортиева органа (А) и схема связей нервных волокон с волосковыми клетками (Б). По [478].

1 — наружные волосковые клетки; 2 — внутренние волосковые клетки; 3 — отверстия, через которые проходят нервные волокна; 4 — спиральный ганглий; сплошные линии — афферентные волокна; штриховые линии — эфферентные волокна; ? — предполагаемые эфферентные связи.

контактов примерно с десятью радиальными волокнами биполярных клеток слухового нерва (рис. 7.6). Информация от наружных волосковых клеток (около 12 000) передается спиральным волокнам нервных клеток, каждое из которых имеет синаптические соединения с большим числом клеток.

Механизм преобразования колебаний стремечка среднего уха в импульсацию волокон слухового нерва можно схематически представить следующим образом: смещения стремечка и связанной с ним мембраны овального окна сопровождаются изменениями

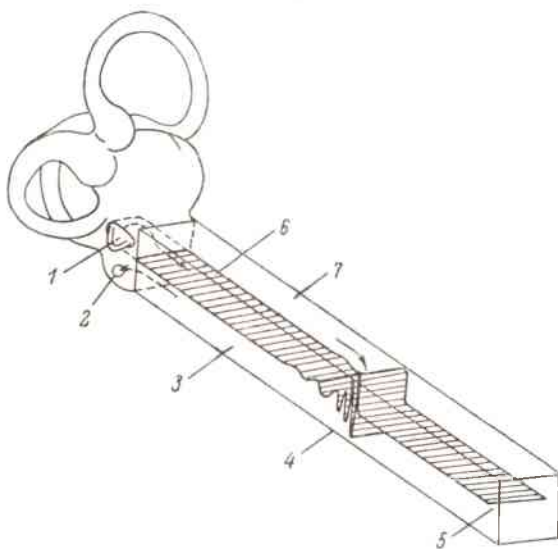


Рис. 7.7. Схематическое изображение «бегущей волны» в улитке. По [537].

1 — стремечко, 2 — круглое окно, 3 — тимпанный канал, 4 — улитка, 5 — геликотрема, 6 — базилярная мембрана, 7 — вестибулярный канал.

давления жидкости (перилимфы) в базальной части вестибулярного канала; под действием этого приводится в движение Рейснера мембрана, вязкая жидкость (эндолимфа) среднего канала улитки и базилярная мембрана (рис. 7.4 и 7.5), что в свою очередь сопровождается колебаниями давления жидкости в тимпанальном канале улитки и смещениями мембраны круглого окна. Разностное давление, возникающее в базальной части улитки между вестибулярным и тимпанальными каналами, распространяется по направлению к геликотреме, образуя «бегущую волну» давления (рис. 7.7).

Вследствие большой податливости мембран, ограничивающих средний канал улитки, скорость распространения бегущей волны давления значительно меньше скорости распространения звука в жидкости со свойствами перилимфы. Так как податливость среднего канала изменяется вдоль улитки, скорость волны зависит

еще и от расстояния от базальной части улитки, сильно уменьшаясь по мере продвижения к геликотреме. По этой же причине амплитуда колебания среднего канала улитки достигает максимума в определенной точке, положение которой зависит от частоты входного сигнала, а затем, при дальнейшем продвижении, быстро затухает. Это значит, что если улитку внутреннего уха представить в развернутом виде (рис. 7.7) и расстояние каждой точки базилярной мембраны от базального конца улитки обозначить с помощью некоторой координаты  $x$ , то различным частотам входных акустических сигналов будут соответствовать различные положения, или координаты, максимума колебания на базилярной мембране, т. е. улитка является преобразователем «частота—координата».

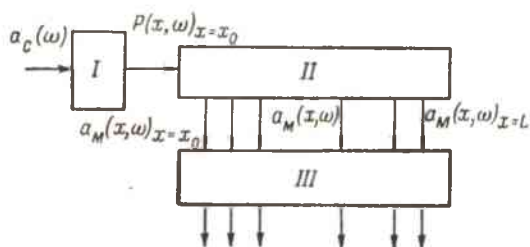


Рис. 7.8. Схематическое изображение последовательности преобразований сигналов в улитке.

$I$  — преобразователь амплитуды колебания стремечка,  $a_0(\omega)$ , в амплитуду разностного давления,  $P(x, \omega)_{x=x_0}$ , в базальной части улитки;  $II$  — преобразователь  $P(x, \omega)_{x=x_0}$  в амплитуды колебания точек вдоль базилярной мембраны улитки,  $a_m(x, \omega)$ ;  $III$  — преобразователь амплитуды колебаний точек базилярной мембраны в сигналы, ведущие к возбуждению волокон слухового нерва.

Колебания базилярной мембраны сопровождаются смещением расположенных на ней волосковых клеток относительно текториальной мембраны и возбуждением их: возбуждение волосковых клеток передается окончаниям первичных слуховых нейронов — клеток спирального ганглия.

Такая последовательность операций над входными сигналами в системе внутреннего уха в настоящее время общепризнана. Однако сложность процессов, происходящих при этом, и трудность получения экспериментальных данных о физических параметрах и функциях отдельных звеньев этой системы могут служить объяснением тому, что, несмотря на более чем столетний период интенсивных исследований, единая непротиворечивая теория функционирования систем внутреннего уха пока отсутствует. Только в самые последние годы, благодаря появлению принципиально новых экспериментальных возможностей, стали быстро накапливаться необходимые данные и появилась надежда, что подобная теория наконец будет разработана.

Представляется целесообразным представить систему улитки в виде трех последовательно соединенных звеньев (рис. 7.8):



преобразователя амплитуды колебания стремечка,  $a_c(\omega)$ , в амплитуду разностного давления в базальной части улитки,  $P(x, \omega)_{x=x_0}$ ; преобразователя амплитуды разностного давления,  $P(x, \omega)_{x=x_0}$ , в амплитуду колебания базилярной мембраны улитки,  $a_m(x, \omega)$ , при различных значениях координаты  $x$  и, наконец, третьего звена, в котором механические колебания базилярной мембраны улитки преобразуются в сигналы, ведущие к возбуждению импульсов в волокнах слухового нерва. О механизмах последнего преобразования практически ничего не известно, и они рассматриваться здесь не будут.

### 7.3.1. АМПЛИТУДНО-ЧАСТОТНАЯ ХАРАКТЕРИСТИКА БАЗАЛЬНОЙ ЧАСТИ СИСТЕМЫ УЛИТКИ

Частотная зависимость амплитуды разностного давления в базальной части улитки,  $P(x, \omega)_{x=x_0}$ , при постоянной амплитуде звукового давления у барабанной перепонки,  $P_6(\omega)$ , определялась непосредственно, с помощью миниатюрных датчиков

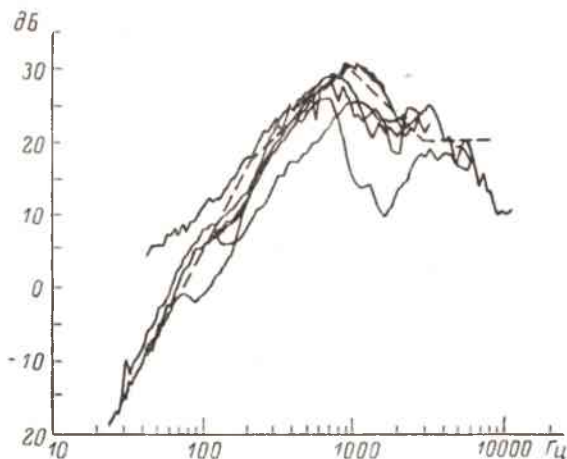


Рис. 7.9. Амплитудно-частотная характеристика  $P(x, \omega)_{x=x_0}/P_6(\omega)$  улитки кошки. По [405].

По оси абсцисс — частота; по оси ординат — амплитуда разностного звукового давления. Данные пяти измерений. Штриховой линией обозначена аппроксимирующая характеристика

давления (диаметром 0.3 мм), помещаемых в вестибулярный и тимпальный каналы улитки кошки [405]. Измерения были выполнены в диапазоне частот 0.02—6 кГц, при уровне звукового давления у барабанной перепонки 90 дБ над порогом.

Измеренные кривые (рис. 7.9) имеют сложную форму. Если в качестве первого приближения описать их с помощью отрезков прямых, изображенных на рис. 7.9 штриховыми линиями, то полу-



чается, что от низких частот и до 300—400 Гц характеристики могут быть аппроксимированы прямой с положительным наклоном около 12 дБ на октаву, затем в диапазоне до 1 кГц наклон характеристики уменьшается до 6 дБ на октаву; в диапазоне частот 1—3 кГц характеристики падают с наклоном около —6 дБ на октаву, а на более высоких частотах (до 6 кГц) характеристики горизонтальны.

Так как амплитудно-частотная характеристика среднего уха кошек,  $a_c(\omega)/P_6(\omega)$ , известна (рис. 7.3), искомая амплитудно-частотная характеристика системы «стремечко—базальная часть

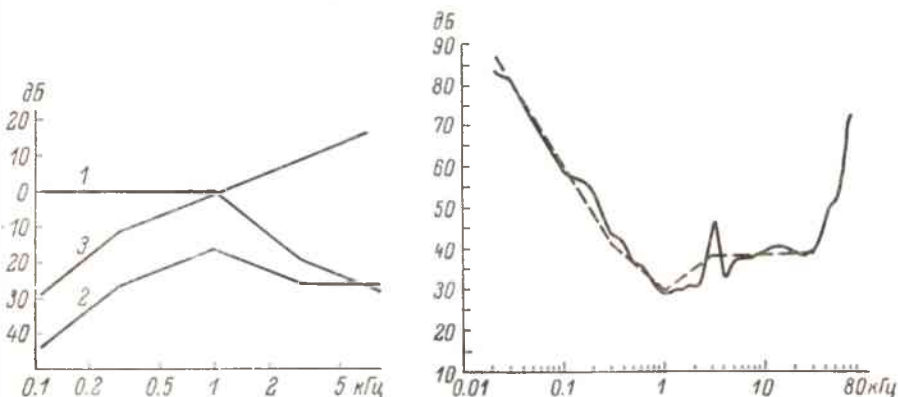


Рис. 7.10. Идеализированные амплитудно-частотные характеристики различных систем механической части слухового анализатора.

По оси абсцисс — частота; по оси ординат — относительная амплитуда. 1 — амплитудно-частотная характеристика среднего уха, определяемая как  $a_c(\omega)/P_6(\omega)$ ; 2 — среднего уха и базальной части улитки,  $P(x, \omega)_{x=x_0}/P_6(\omega)$ ; 3 — базальной части улитки,  $P(x, \omega)_{x=x_0}/a_c(\omega)$ . Характеристики для наглядности сдвинуты по вертикали.

Рис. 7.11. Усредненная зависимость (по данным трех независимых измерений) уровня звукового давления у барабанной перепонки, необходимого для возникновения микрофонного потенциала с амплитудой 1 мкВ в базальной части улитки кошки, от частоты. По [523].

По оси абсцисс — частота; по оси ординат — звуковое давление относительно  $2 \cdot 10^{-5}$  Н/м<sup>2</sup>. Аппроксимирующая характеристика обозначена штриховой линией.

улитки», которую обозначим  $P(x, \omega)_{x=x_0}/a_c(\omega)$ , может быть получена как разность характеристик, выраженных в логарифмическом масштабе:

$$20 \lg \frac{P(x, \omega)_{x=x_0}}{a_c(\omega)} = 20 \lg \frac{P(x, \omega)_{x=x_0}}{P_6(\omega)} - 20 \lg \frac{a_c(\omega)}{P_6(\omega)}, \quad (2)$$

где  $a_c(\omega)$  — амплитуда колебания стремечка.

Сами характеристики  $a_c(\omega)/P_6(\omega)$  и  $P(x, \omega)_{x=x_0}/P_6(\omega)$ , представленные в виде отрезков прямых, а также результат их вычитания, характеристика  $P(x, \omega)_{x=x_0}/a_c(\omega)$ , изображены на

рис. 7.10. Видно, что характеристика системы «стремечко—базальная часть улитки» на частотах до 300 Гц имеет наклон 12 дБ на октаву, а на более высоких частотах — +6 дБ на октаву.

Существенно, что данные, полученные при непосредственном измерении характеристики  $P(x, \omega)_{x=x_0}/P_0(\omega)$ , полностью подтверждаются и экспериментами по определению частотных характеристик системы с помощью регистрации микрофонных потенциалов<sup>1</sup> дифференциальными электродами, помещаемыми в вестибулярный и тимпальный каналы базальной части улитки кошки [224, 523, 524, 526].

Аппроксимирующая характеристика, построенная на основе этих данных (рис. 7.11), практически совпадает<sup>2</sup> с аналогичной характеристикой, приведенной на рис. 7.9.

Экспериментами с измерением микрофонных потенциалов, кроме того, было показано, что система на уровне базальной части улитки линейна [224, 523] и что величина наклона характеристики в области низких частот различна у различных животных [224] и составляет у кошек и шиншил 12 дБ на октаву, а у морских свинок и мартышек — +6 дБ на октаву.

Таким образом, на основе приведенных выше данных может быть сделан вывод, что характеристика звена, преобразующего амплитуду колебания стремечка в амплитуду разностного звукового давления в базальной части улитки, у кошки и, по-видимому, у человека [224] имеет наклон 12 дБ на октаву от самых низких частот и до 300—400 Гц, т. е. величина разностного давления у овального окна оказывается пропорциональной ускорению движения стремечка, что не согласуется с распространенной в настоящее время гипотезой о механизме передачи энергии колебания стремечка базальной части улитки [258, 548], согласно которой амплитуда давления пропорциональна скорости движения стремечка.

Кроме того, эти данные свидетельствуют о том, что на частотах выше 300 Гц и по крайней мере до 6 кГц частотная характеристика  $P(x, \omega)_{x=x_0}/a_0(\omega)$  почти полностью компенсирует спад характеристики среднего уха на частотах выше 1 кГц.<sup>3</sup>

По-видимому, при моделировании процессов преобразования

<sup>1</sup> Допустимость использования этих данных для оценки частотной зависимости амплитуды разностного давления в базальной части улитки обосновывается тем, что на низких и средних частотах амплитуда микрофонного потенциала пропорциональна амплитуде смещения базилярной мембраны [189], а она, в свою очередь, на этих частотах пропорциональна амплитуде разностного давления [545].

<sup>2</sup> Поскольку при получении характеристики, представленной на рис. 7.11, применялся пороговый метод, т. е. по существу измерялась характеристика  $P_0(\omega)/a_{\text{мп}}(\omega)$ , где  $a_{\text{мп}}(\omega)$  — амплитуда микрофонного потенциала, амплитудно-частотная характеристика  $a_{\text{мп}}(\omega)/P_0(\omega)$  получается после поворота кривой, изображенной на рис. 7.11, на 180° относительно горизонтальной оси.

<sup>3</sup> Имеются данные, основанные на измерении микрофонных потенциалов и также полученные на кошках [435], согласно которым спад характеристики среднего уха компенсируется полностью до частоты порядка 10—20 кГц.

сигналов слуховым спектральным анализатором воспроизведение отдельно характеристики среднего уха и характеристики звена, преобразующего амплитуду колебания стремечка в амплитуду разностного давления в базальной части улитки, может оказаться не всегда целесообразным. В ряде случаев проще воспроизвести сразу суммарную характеристику.

### 7.3.2. ХАРАКТЕРИСТИКИ ГИДРОМЕХАНИЧЕСКОЙ ЧАСТИ СИСТЕМЫ УЛИТКИ

Как было показано выше, разностное давление в базальной части улитки,  $P(x, \omega)_{x=x_0}$ , преобразуется в бегущую вдоль улитки волну давления,  $P(x, \omega)$ , которая, в свою очередь, вызывает колебания структур среднего канала улитки и базилярной мембраны.

Таким образом, амплитуда колебания базилярной мембраны улитки,  $a_m(x, \omega)$ , является функцией двух переменных — частоты  $\omega$  входного сигнала и координаты  $x$  вдоль улитки и, следовательно, для полного описания свойств системы необходимо располагать семейством характеристик  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$ , каждая из которых в зависимости от того, какая переменная ( $x$  или  $\omega$ ) фиксируется, характеризует амплитудно-частотные или амплитудно-координатные свойства системы.

Фактические данные о частотно-избирательных свойствах базилярной мембраны улитки были получены с помощью разнообразных методов: визуального наблюдения колебаний мембраны при синхронном с частотой сигнала освещении, измерения скорости колебания точек базилярной мембраны с помощью эффекта Мессбауэра (путем регистрации изменения потока гамма-лучей от помещаемого на базилярную мембрану радиоактивного источника), визуальной регистрации величины области «размытости» на базилярной мембране при лазерном освещении, регистрации амплитуды колебания базилярной мембраны с помощью микромостовых датчиков.

Основные результаты, полученные с помощью этих методов, приведены в табл. 7.1. Видно, что все они свидетельствуют о резкой несимметрии амплитудно-частотных характеристик базилярной мембраны. При этом абсолютные значения крутизны характеристик, полученные в ранних работах, отличаются от более поздних измерений весьма значительно. Наряду с этим имеются и принципиальные качественные различия.

Измерения, выполненные на препаратах улитки человека [189], показали, что улитка ведет себя как линейная система и, следовательно, форма амплитудно-частотных и фазо-частотных характеристик ее не зависит от уровня интенсивности стимула. Вместе с тем в экспериментах на живых марышкиках [443, 446, 447] подобная зависимость была обнаружена (рис. 7.12), причем характер изменения таков, что частотная избирательность системы уменьшается

Таблица 7.4

Параметры амплитудно-частотных характеристик точек базальной мембраны, полученные различными методами

| Авторы                              | Объект                    | Метод                                                 | Место измерения                                    | Наклон характеристики, дБ/окт                            |                        | Примечание                                                                                  |
|-------------------------------------|---------------------------|-------------------------------------------------------|----------------------------------------------------|----------------------------------------------------------|------------------------|---------------------------------------------------------------------------------------------|
|                                     |                           |                                                       |                                                    | низко-частотная ветвь                                    | высоко-частотная ветвь |                                                                                             |
| Бекеш [188, 189]                    | Препараты улитки человека | Визуальное наблюдение при стробоскопическом освещении | Аппальный конец улитки, диапазон частот 25—4000 Гц | 6                                                        | От -20 до -30          | Измерения относительно колебаний стремечка, уровень 150 дБ над порогом                      |
| Джонстон, Бойль [308]               | Морские свинки            | Эффект Мессбауэра                                     | 1.4 мм от стремечка (резонанс 18—21 кГц)           | 10 на частотах ниже 3 кГц и 13 на более высоких частотах | -60                    | Измерения относительно колебаний стремечка, уровни входных сигналов более 90 дБ над порогом |
| Джонстон и др. [309]                | То же                     | То же                                                 | 2 мм от стремечка (резонанс в области 18 кГц)      | 12                                                       | -95                    | То же                                                                                       |
| Джонстон, Ятес [310], Вильсон [530] | » »                       | » »                                                   | То же                                              | 6 на частотах до 800 Гц и 12 на более высоких частотах   | -340                   | » »                                                                                         |

Таблица 7.1 (продолжение)

| Авторы                                                                  | Объект         | Метод                                        | Место измерения                                           | Наклон характеристики, дБ/окт                                                                              |                                                                     | Примечание                                                                                  |
|-------------------------------------------------------------------------|----------------|----------------------------------------------|-----------------------------------------------------------|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
|                                                                         |                |                                              |                                                           | низкочастотная ветвь                                                                                       | высокочастотная ветвь                                               |                                                                                             |
| Роде, Дейслер [445], Роде [443], Роде, Роблес [446], Роблес, Роде [447] | Мартышки       | Эффект Мессбауэра                            | На координатах с резонансом на частотах 6—9 кГц           | 6—10 на низких частотах и до 20—30 у максимума (наклон уменьшается при увеличении уровня входных сигналов) | —                                                                   | Измерения относительно колебаний молоточка, уровни входных сигналов более 70 дБ над порогом |
| Роде [444]                                                              | »              | То же                                        | То же                                                     | От 24 сразу после смерти, до 6 через несколько суток                                                       | От —50 до —150                                                      | То же                                                                                       |
| Колбфелль [332, 333, 335]                                               | Морские свинки | Визуальные наблюдения при лазерном освещении | На координатах с резонансом на 18—30 кГц                  | 9 — у живой, 8 — 4 день, 6 — 2 дня, 4.5 — 3 дня после смерти                                               | —130 — у живой, —99 — 1 день, —57 — 2 дня, —46 — 3 дня после смерти | Измерения при входных сигналах более 100 дБ                                                 |
| Колбфелль [334]                                                         | То же          | То же                                        | То же                                                     | 7—9                                                                                                        | От —100 до —130                                                     | То же                                                                                       |
| Вильсон, Джонстон [532]                                                 | »              | Микрометрический датчик                      | 1.6—4.4 мм от стрелочка (резонансные частоты выше 18 кГц) | 9—12                                                                                                       | —100                                                                | Измерения относительно колебаний наковальни, уровни входных сигналов 90—120 дБ над порогом  |



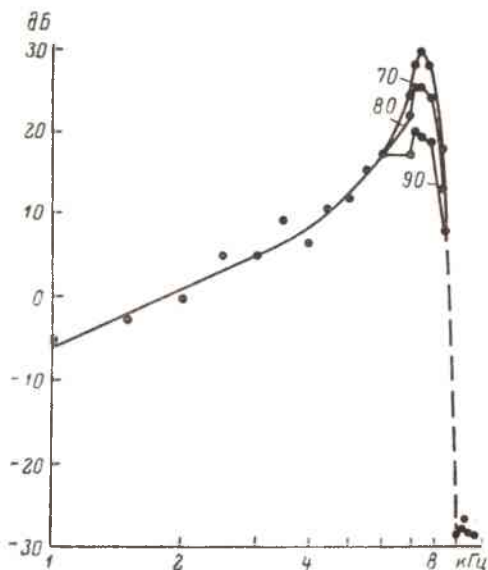
с увеличением уровня. Существенно, что нелинейный характер зависимости амплитуды колебания улитки от амплитуды входного сигнала был обнаружен только на частотах вблизи резонанса. На других частотах система не отличается от линейной.

Изменения амплитудно-частотной характеристики вблизи резонанса с увеличением уровня сигнала проявляются не только в уплощении кривой, но и в сдвиге вниз (в сторону более низких частот) самой резонансной частоты [443].

Следует отметить, что на другом живом объекте, использованном в экспериментах (морские свинки), нелинейные свойства базилярной мембраны не были обнаружены, несмотря на то, что попытки выявить нелинейность неоднократно предпринимались [531, 532]. Причины этого несо-

Рис. 7.12. Зависимость амплитудно-частотной характеристики точки базилярной мембраны улитки мартышки от уровня звукового давления входного сигнала у барабанной перепонки. По [443].

По оси абсцисс — частота; по оси ординат — отношение амплитуды колебания базилярной мембраны к амплитуде колебания стремечка (в дБ). Цифры у кривых — уровни звукового давления у барабанной перепонки (в дБ).



ответствия пока не ясны и в настоящее время активно обсуждаются [310, 443, 444, 446].

На обоих исследованных объектах (и мартышках, и морских свинках) было обнаружено, что амплитудно-частотные свойства системы начинают изменяться сразу же после смерти животного; по имеющимся данным [333, 335], этот процесс продолжается в течение нескольких дней и выражается в уменьшении наклона обеих (низкочастотной и высокочастотной) ветвей характеристики. На третий день после смерти наклоны характеристик составляют 4,5 дБ на октаву и —46 дБ на октаву соответственно, т. е. (табл. 7.1) оказываются очень близкими к результатам экспериментов, проводившихся на препаратах улитки человека [189].

По данным [444], полученным на мартышках, уже через несколько минут после смерти животного улитка начинает вести себя как линейная система. Постепенные изменения формы амплитудно-частотной характеристики вблизи резонанса, наблюдаемые после смерти, аналогичны тем, которые наблюдаются при увеличении



интенсивности сигнала (см. выше): кривая уплощается, резонансная частота сдвигается в сторону низких частот (рис. 7.13).

Таким образом, сейчас уже не вызывает сомнений, что живая улитка обладает большей избирательностью, чем препараты улитки.

Естественно, прямых измерений на улитке живого человека сделать нельзя, однако, как будет показано в последнем разделе настоящей главы, психоакустические данные, полученные на человеке, говорят о нелинейности частотно-избирательных свойств слухового анализатора и, следовательно, не противоречат измерениям, полученным на животных.

Из амплитудно-частотных характеристик базиллярной мембраны, приведенных выше (рис. 7.12, табл. 7.1), в принципе возможно определить вид характеристики  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$ , если известна амплитудно-частотная характеристика преобразователя колебаний стремечка в разностное давление в базальной части улитки. Для этого можно воспользоваться выражением  $\lg[a_m(x, \omega)/P(x, \omega)_{x=x_0}] = \lg[a_m(x, \omega)/a_c(\omega)] - \lg[P(x, \omega)_{x=x_0}/a_c(\omega)]$ .

В настоящее время, как это видно из табл. 7.1, характеристики  $a_m(x, \omega)/a_c(\omega)$  получены лишь на морских свинках и мартишках. Характеристики же  $P(x, \omega)_{x=x_0}/a_c(\omega)$ , как было показано выше, непосредственно измерялись только на кошках, поэтому они не могут быть использованы при определении характера частотной зависимости  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$ .

Сведения о форме зависимости  $P(x, \omega)_{x=x_0}/a_c(\omega)$  на низких частотах для морских свинок могут быть получены, если воспользоваться данными о частотных характеристиках микрофонных потенциалов в базальной части улитки [224] (см. также раздел 7.3). Вычитание этой характеристики из экспериментально полученных для морских свинок характеристик  $a_m(x, \omega)/a_c(\omega)$  показывает, что на низких частотах (по крайней мере, до 300—500 Гц) характеристика  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$  должна мало отличаться от горизонтальной. Вид этой характеристики в области частот резонанса кривой  $a_m(x, \omega)/a_c(\omega)$  остается неизвестным, поскольку прямые измерения  $P(x, \omega)_{x=x_0}/a_c(\omega)$  в этой области частот пока еще не выполнялись, а использование данных по измерению микрофонных потенциалов затруднительно, так как амплитуда микрофонного потенциала на этих частотах заведомо не пропорциональна амплитуде разностного давления.

Наряду с амплитудно-частотными характеристиками большой интерес представляют амплитудно-координатные характеристики базиллярной мембраны улитки. Объясняется это прежде всего тем, что реальная разрешающая способность системы по частоте определяется именно характером распределения амплитуды колебания мембраны по координате, т. е. формой амплитудно-координатных характеристик.

Кроме того, исследование формы этих характеристик может дать дополнительные сведения о колебательных свойствах улитки,

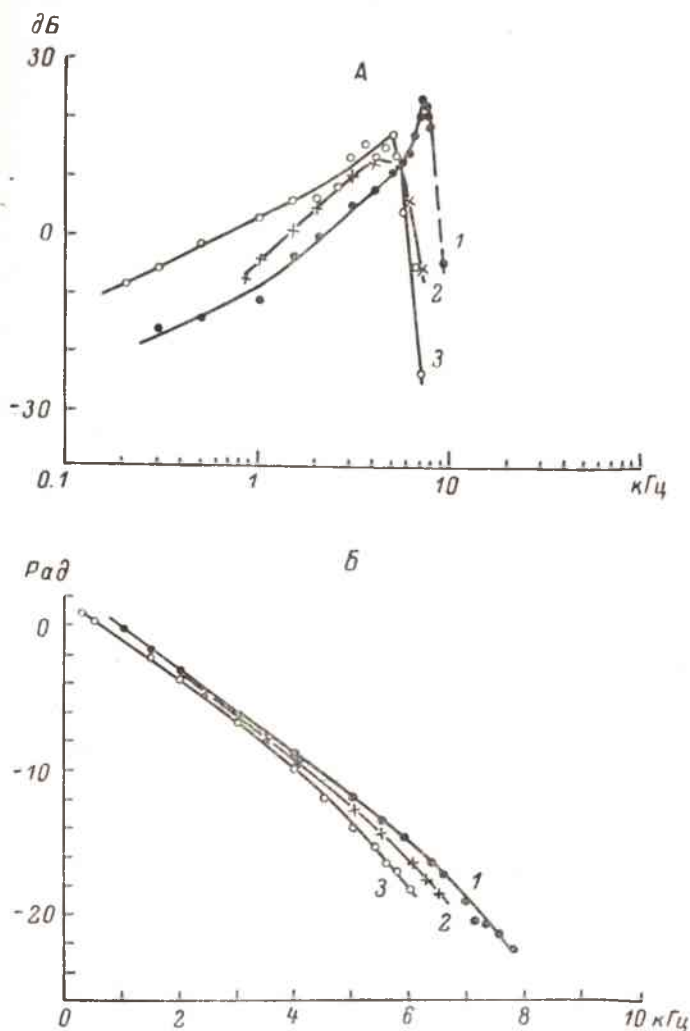


Рис. 7.13. Изменение амплитудно-частотных (А) и фазо-частотных (Б) характеристик базилярной мембраны мартышки после ее смерти. По [443].

По оси абсцисс — частота; по оси ординат: на А — отношение амплитуды колебаний базилярной мембраны к амплитуде колебаний молоточка; на Б — сдвиг по фазе колебания базилярной мембраны по отношению к колебанию молоточка среднего уха (в радианах).

1 — получено на живых мартышках, 2 и 3 — через 1 и 31 час после смерти.

так как на форме этих характеристик, в отличие от амплитудно-частотных, не должны сказываться амплитудно-частотные свойства звеньев системы, расположенных до улитки (наружного и среднего уха), а также преобразователя амплитуды колебания стремечка в амплитуду разностного давления в базальной части улитки.

К сожалению, к настоящему времени амплитудно-координатные характеристики непосредственно измерялись всего в двух известных нам работах [189, 335] на препаратах улитки и при уровнях звукового давления входных сигналов существенно больше 100 дБ над абсолютным порогом. Было показано, что амплитудно-координатные характеристики имеют максимум, положение которого смещается к геликотреме с понижением частоты входного сигнала.

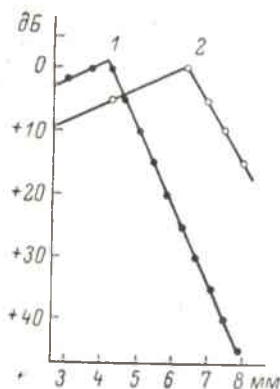


Рис. 7.14. Амплитудно-координатные характеристики улитки морской свинки. По [335].

По оси абсцисс — расстояние от базального конца улитки; по оси ординат — уровень звукового давления, необходимый для обнаружения колебания на данной координате. 1 — при частоте входного сигнала 9 кГц, 2 — при частоте входного сигнала 5 кГц. 0 дБ соответствует уровню звукового давления 100 дБ над абсолютным порогом.

Характеристики несимметричны относительно максимума и могут быть представлены в виде двух прямых (рис. 7.14) с пологой базальной частью характеристики и крутой апикальной.

Амплитудно-координатные характеристики, также как и амплитудно-частотные, сильно уплощаются после смерти животного [335]. В частности, эксперименты, выполненные на морских свинках, показывают, что если спустя 3 дня после смерти базальный наклон характеристики составляет 10 дБ на октаву координатных частот<sup>4</sup> и апикальный наклон около —56 дБ на октаву, то наклоны амплитудно-координатных характеристик, измеренные через 7 дней после смерти, оказываются равными соответственно 5 дБ на октаву и —12 дБ на октаву. Данные экспериментов, произведенных на препаратах улитки человека [189], находятся в соответствии с этими результатами.

Представляется интересным сопоставить характеристики и свойства, непосредственно наблюдаемые на улитках различных животных, с характеристиками, которые могут быть получены на современных моделях улитки, и тем самым оценить степень соответствия этих моделей экспериментальным данным.

<sup>4</sup> Поскольку между частотой входного сигнала и координатой максимума колебаний базиллярной мембраны имеется однозначная зависимость, в ряде случаев оказывается более целесообразным вместо шкалы координат пользоваться шкалой частот, которые в этом случае называются «координатными».

### 7.3.3. МОДЕЛИРОВАНИЕ УЛИТКИ КАК СПЕКТРАЛЬНОГО АНАЛИЗАТОРА

Гидродинамическая теория улитки устанавливает связь между характером распределения вдоль улитки амплитуд и фаз разностного давления,  $P(x, \omega)$ , возникающего при гармоническом входном сигнале и действующего на средний канал улитки, и физическими параметрами самой улитки, большая часть которых является функцией координаты  $x$ .

Эти параметры могут быть или определены экспериментально (что возможно лишь для некоторых параметров), или подобраны так, чтобы наблюдалось соответствие между характеристиками модели, базирующейся на теории, и экспериментально измеренными характеристиками, полученными на улитках различных животных.

Анализ математических моделей улитки (см. обзоры [459, 547]) показывает, что при составлении дифференциальных уравнений, описывающих колебательные свойства базилярной мембраны, обычно принимаются следующие основные допущения:

- 1) система линейна;
- 2) весь средний канал, заключенный между базилярной и Рейснеровой мембранами вместе с жидкостью и расположенными в ней структурами, заменяется одной мембраной с эквивалентными массой, сопротивлением потерь и податливостью;
- 3) элементарные участки на эквивалентной мембране никак не связаны, т. е. скорость каждой частицы мембраны зависит лишь от величины давления в данной точке и не зависит от скорости движения соседних точек;
- 4) вдоль улитки распространяются только поперечные волны, причем предполагается, что длина волны существенно больше поперечных размеров канала улитки;
- 5) закрученность улитки не оказывает никакого влияния на условие распространения волн;
- 6) величины поперечных сечений вестибулярного и тимпанального каналов имеют одинаковую зависимость от координаты;
- 7) вязкость перилимфы незначительна, и ею можно пренебречь.

Оказывается возможным свести все полученные дифференциальные уравнения к одному уравнению [459], которое для стационарного гармонического входного воздействия записывается в виде:

$$\frac{d}{dx} \left[ S(x) \frac{dP(x, \omega)}{dx} \right] + \left[ \left( \frac{\omega}{u} \right)^2 \cdot S(x) - j \frac{2\rho\omega b(x)}{W(x, \omega)} \right] \cdot P(x, \omega) = 0, \quad (3)$$

где  $\omega$  — угловая частота входного сигнала;  $\rho$  — плотность перилимфы;  $P(x, \omega)$  — разностное давление, действующее на мембрану;  $S(x)$  — суммарное поперечное сечение вестибулярного и тимпанального каналов;  $b(x)$  — ширина эквивалентной мембраны;  $u$  — скорость звука в перилимфе;  $W(x, \omega)$  — комплексное со-

противление элемента эквивалентной мембраны, которое, в свою очередь, представляется как сумма трех компонент:

$$W(x, \omega) = R(x) + j\omega M(x) + \frac{1}{j\omega C(x)}, \quad (4)$$

где  $R(x)$  характеризует погонное сопротивление трения, определяющее потери при колебании мембраны;  $C(x)$  — упругие свойства эквивалентной мембраны, ее податливость, а  $M(x)$  — инерционные свойства мембраны, определяемые массой.

В настоящее время из пяти входящих в уравнение коэффициентов ( $R(x)$ ,  $M(x)$ ,  $C(x)$ ,  $S(x)$  и  $b(x)$ ) лишь один — податливость базилярной мембраны,  $C(x)$ , — непосредственно измерялся [189]. Остальные могут быть оценены лишь приблизительно, поскольку они представляют некоторые эквивалентные величины, непосредственно не измеряемые или весьма трудно измеряемые. Например, параметр  $M(x)$ , характеризующий массу колеблющейся эквивалентной мембраны, не только должен включать массу базилярной и Рейснеровой мембран вместе с жидкостью и кортиевым органом, заключенными между ними, но должен учитывать лишь ту часть этой массы, которая синфазно колеблется с мембранами, а также часть синфазно колеблющейся с мембранами перилимфы вестибулярного и тимпанального каналов [262].

Приблизительные оценки параметров показывают [548], что податливость мембраны,  $C(x)$ , в наибольшей степени зависит от координаты, при этом экспоненциальная зависимость податливости от координаты

$$C(x) = C_0 e^{\beta(x-x_0)}, \quad (5)$$

где  $C_0$ ,  $x_0$  и  $\beta$  — постоянные, хорошо описывает экспериментальные данные.

Характер зависимости от координаты остальных параметров улитки в силу вышесказанного обычно задается произвольно таким образом, чтобы получаемые дифференциальные уравнения были по возможности решаемы в аналитическом виде. Так, например, в теории улитки Петерсона и Богерта [418] допускалось, что трением при движении мембраны можно пренебречь —  $R(x) = 0$ , а в теории Звислоцкого [546] пренебрегалось инерционной составляющей —  $M(x) = 0$ .

Как правило, дифференциальные уравнения, описывающие свойства более сложных моделей улитки, не имеют аналитического решения.

Отсутствие конкретных данных о параметрах и большая сложность самой системы улитки приводят к необходимости создавать такие модели, в которых можно было бы легко изменять различные параметры и получаемые результаты сравнивать с экспериментально измеряемыми на живых объектах. Поэтому особенное значение и распространение получают модели, в которых на основании принципов динамических аналогий система уравнений вопло-



щается в эквивалентную ей электрическую схему. Модели такого типа могут быть выполнены либо в виде физических устройств (см. обзор [459]), либо в виде соответствующих программ для цифровых вычислительных машин [325] или гибридных систем [272].

При построении электрических аналогов обычно исходят из подобия дифференциального уравнения, описывающего распределение давления  $P(x, \omega)$  вдоль улитки и дифференциального уравнения, описывающего распределение напряжения  $U(x, \omega)$  вдоль непрерывной неоднородной линии и имеющего вид

$$\frac{d}{dx} \left[ \frac{1}{Z(x, \omega)} \cdot \frac{dU(x, \omega)}{dx} \right] - Y(x, \omega) \cdot U(x, \omega) = 0, \quad (6)$$

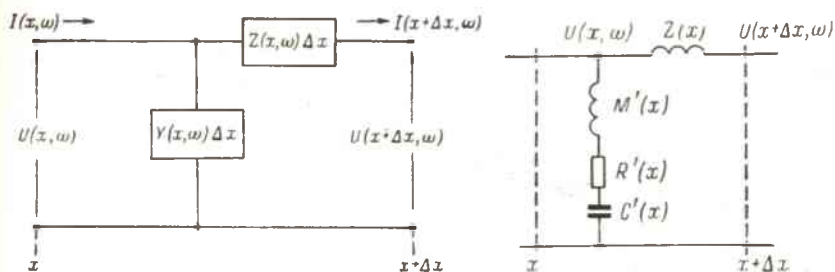


Рис. 7.15. Блок-схема элементарного звена  $\Delta x$  неоднородной линии — аналога улитки.

Обозначения см. в тексте.

Рис. 7.16. Электрическая схема элементарного звена аналога улитки.

Обозначения см. в тексте.

где  $Z(x, \omega)$  и  $Y(x, \omega)$  — комплексные погонные продольный импеданс и поперечная проводимость элементарного звена  $\Delta x$  линии, представленного на рис. 7.15.

Сопоставляя уравнения (3) и (6), возможно определить элементы  $Z(x, \omega)$  и  $Y(x, \omega)$  электрического аналога через параметры улитки.

Наиболее простая связь между элементами получается, если в уравнении (3) пренебречь вторым членом, характеризующим сжимаемость перилимфы, а  $S(x)$  считать независимым от координаты. В этом случае значения продольного импеданса и поперечной проводимости элементарного звена аналога улитки оказываются равными соответственно

$$Z(x, \omega) = j\omega \frac{2\rho}{S} \quad (7)$$

и

$$Y(x, \omega) = \frac{b(x)}{W(x, \omega)} = \frac{b(x)}{R(x) + j\omega M(x) + \frac{1}{j\omega C(x)}}. \quad (8)$$



Теперь электрическая схема элементарного звена аналога улитки может быть представлена так, как изображено на рис. 7.16.

Системы из последовательно включенных звеньев такого типа, при различных предположениях относительно величины и характера зависимости параметров от координаты, были исследованы рядом авторов [12, 87, 186, 193, 325, 413, 418, 490, 517]. При достаточно большом числе звеньев такие системы вполне удовлетворительно воспроизводят свойства непрерывной линии и могут рассматриваться как специализированные решающие устройства, преобразующие в реальном масштабе времени сигнал любой формы на входе в соответствующие сигналы на выходах вдоль цепочки звеньев.

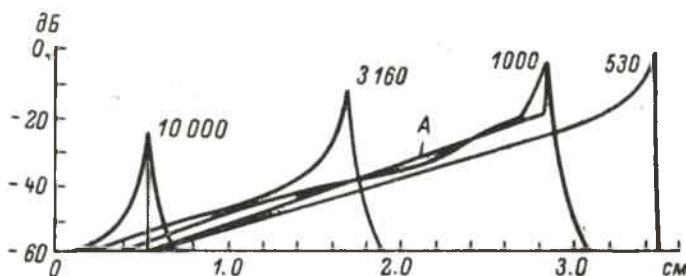


Рис. 7.17. Семейство амплитудно-координатных характеристик электрического аналога улитки. По [193].

По оси абсцисс — «координата» вдоль улитки (эквивалент номера звена модели); по оси ординат — относительная амплитуда. А — прямая с наклоном 12 дБ на октаву «координатных» частот. Цифры у кривых — частоты входных сигналов (в Гц).

Переходя к рассмотрению характеристик моделей улитки, отметим, что к настоящему времени имеется значительное количество моделей [130, 186, 232, 272, 418, 461, 517, 537, 548], отличающихся в основном видом зависимости параметров от координаты.

Поскольку обычно принимается экспоненциальный закон изменения податливости мембраны,  $C(x)$ , от координаты и именно этот параметр изменяется в наибольшем по сравнению с другими параметрами диапазоне значений, влияние характера зависимости других параметров от координаты на форму характеристик моделей оказывается не очень сильным, и всеми авторами отмечается хорошее качественное совпадение характеристик моделей с характеристиками, измеренными на улитке. В частности, на моделях воспроизводится «бегущая волна» с уменьшающейся по мере продвижения к апикальному концу улитки фазовой скоростью распространения, отмечается образование максимумов колебаний на определенной координате, зависящей от частоты входного сигнала (рис. 7.17).

Можно показать, что наклон амплитудно-частотных характеристик в области низких частот и наклоны амплитудно-координатных характеристик в области малых значений координат, полу-

чаемые на моделях, должны практически совпадать с измеряемыми на улитках.

Известно, что аналогом амплитуды колебания базиллярной мембраны в модели является заряд на конденсаторе поперечной цепи:

$$q_M(x, \omega) = \frac{U(x, \omega) \cdot Y(x, \omega)}{j\omega}, \quad (9)$$

где  $Y(x, \omega)$  — проводимость поперечной цепи эквивалентного звена модели, выражаемая как

$$Y(x, \omega) = \frac{1}{R(x) + j\omega M(x) + \frac{1}{j\omega C_0 e^{\beta(x-x_0)}}}, \quad (10)$$

а  $U(x, \omega)$  — амплитуда напряжения, падающего на поперечном звене модели (аналог разностного давления  $P(x, \omega)$  в улитке). При этих условиях амплитудно-частотные (при фиксированных  $x$ ) и амплитудно-координатные (при фиксированных  $\omega$ ) характеристики модели могут быть представлены в виде

$$\frac{q_M(x, \omega)}{U(x, \omega)_{x=x_0}} = \frac{U(x, \omega) \cdot Y(x, \omega)}{U(x, \omega)_{x=x_0} \cdot j\omega}, \quad (11)$$

где  $U(x, \omega)_{x=x_0}$  — амплитуда напряжения на входе модели (аналог разностного давления в базальной части улитки,  $P(x, \omega)_{x=x_0}$ ).

Подставив выражение (10) в (11) и введя обозначения

$$\omega_r(x) = \frac{1}{\sqrt{M(x) \cdot C_0 e^{\beta(x-x_0)}}} \text{ и } \tilde{\omega}(x) = R(x) \sqrt{\frac{C_0 e^{\beta(x-x_0)}}{M(x)}}, \quad (12)$$

после логарифмирования получим:

$$20 \lg \frac{q_M(x, \omega)}{U(x, \omega)_{x=x_0}} = 20 \lg \frac{U(x, \omega)}{U(x, \omega)_{x=x_0}} + 20\beta(x-x_0) \lg e + 20 \lg C_0 + \\ + 20 \lg \frac{1}{1 + j\tilde{\omega}(x) \frac{\omega}{\omega_r(x)} - \left[ \frac{\omega}{\omega_r(x)} \right]^2}. \quad (13)$$

Таким образом, каждая амплитудно-частотная (или амплитудно-координатная) характеристика моделей улитки может быть представлена в виде суммы четырех характеристик: амплитудно-частотной (или амплитудно-координатной) характеристики  $20 \lg \frac{U(x, \omega)}{U(x, \omega)_{x=x_0}}$ , прямой  $20\beta(x-x_0) \lg e$ , постоянной  $20 \lg C_0$  и резонансной кривой  $20 \lg \frac{1}{1 + j\tilde{\omega}(x) \frac{\omega}{\omega_r(x)} - \left[ \frac{\omega}{\omega_r(x)} \right]^2}$  с частотой резонанса  $\omega_r(x)$ .

Эксперименты на моделях, а также теоретический анализ уравнения (6) [461, 537] показывают, что характеристики

$20 \lg \frac{U(x, \omega)}{U(x, \omega)_{x=x_0}}$  при тех параметрах моделей, которые представляются достаточно правдоподобными, являются монотонно падающими функциями с почти горизонтальной частью в области малых частот (координат) и круто спадающей частью (с наклоном, достигающим десятков или даже сотен дБ на октаву) в области больших частот (координат) (рис. 7.18).

Из выражения (13) видно, что на форму амплитудно-частотных характеристик модели оказывают влияние только первая и четвертая составляющие. Поскольку и та и другая практически горизонтальны на частотах, значительно более низких, чем частота резонанса, результирующая ха-

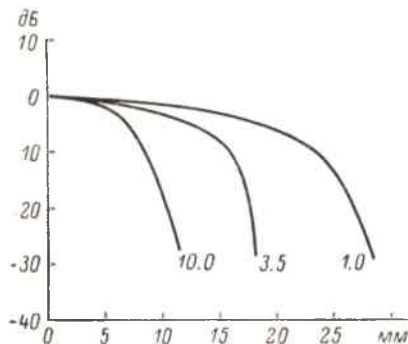


Рис. 7.18. Амплитудно-координатные характеристики  $U(x, \omega)/U(x, \omega)_{x=x_0}$  модели улитки.

По оси абсцисс — координата (эквивалент номера звена модели); по оси ординат — отношение амплитуды напряжения на поперечных цепочках звеньев модели (в дБ) к амплитуде напряжения на входе модели. Цифры у кривых — частоты входных сигналов (в кГц).

рактеристика в этой области также должна мало отличаться от горизонтальной.

Как было нами показано в предыдущем разделе, такую же форму имеет и амплитудно-частотная характеристика улитки  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$ , получаемая на основе экспериментальных данных.

Из выражения (13) также следует, что наклон амплитудно-координатных характеристик моделей в области малых координат определяется в основном второй компонентой и составляет около  $20\beta \lg e$ . Чтобы выразить этот наклон характеристики в децибелах на октаву координатных частот, воспользуемся выражением (12), согласно которому для отношения двух резонансных частот модели при  $M(x) = \text{const}$  можно записать:

$$\frac{\omega_{r2}(x)}{\omega_{r1}(x)} = e^{\frac{\beta \Delta x}{2}} = 2^F, \quad (14)$$

где  $F$  — число октав между частотами  $\omega_{r2}$  и  $\omega_{r1}$ . Отсюда следует, что максимумы на амплитудно-координатной характеристике, соответствующие частотам входного сигнала, отличающимся на одну октаву ( $F=1$ ), будут располагаться на расстоянии  $\Delta x$ , равном  $\Delta x = 2 \ln(2/\beta)$ . На этом интервале амплитуда сигнала изменится на  $40 \lg e \cdot \ln 2$  дБ, или крутизна амплитудно-координатных характеристик должна быть 12 дБ на октаву координатных частот.

Этот вывод о форме амплитудно-координатных характеристик моделей улитки в области малых координат подтверждается экспериментальными кривыми, полученными на конкретных моделях (рис. 7.17).

Как было показано в предыдущем разделе, базальная часть амплитудно-координатных характеристик улитки имеет наклон порядка 10 дБ на октаву координатных частот, т. е. и здесь наблюдается хорошее соответствие характеристик, получаемых на моделях и на улитке.

Можно ожидать, что путем соответствующего выбора параметров вполне достижимо еще лучшее совпадение этих и ряда других характеристик, получаемых на моделях и на живых объектах. Вместе с тем становится все более очевидным, что некоторые свойства, наблюдаемые в экспериментах на улитках, принципиально невозможно воспроизвести существующими моделями и, следовательно, требуется дальнейшее их усложнение. В частности, как это видно из выражения (13), амплитудно-частотные и амплитудно-координатные характеристики моделей в области максимума должны иметь форму, близкую к резонансной. Экспериментально полученные на улитках характеристики имеют другую форму. Кроме того, во всех моделях, предназначавшихся для воспроизведения характеристик улитки на низких частотах [186, 193, 517], наблюдаются отражения волн от конца модели, приводящие к сильному изменению амплитудно-частотных, амплитудно-координатных и других характеристик. В то же время эксперименты на улитках [189] таких отражений не обнаруживают.

Рассмотренные выше модели линейны и, следовательно, не воспроизводят нелинейные свойства улитки. Разработка же нелинейных моделей [168, 310, 322] находится в самой начальной стадии. Объясняется это прежде всего тем, что к настоящему времени остается неясным не только точное место нелинейности в слуховом спектральном анализаторе, но и ее вид [275, 277, 472, 538].

#### 7.4. ХАРАКТЕРИСТИКИ «ФИЛЬТРА УЛИТКИ» ПРИ ИЗМЕРЕНИИ НЕПРЯМЫМИ МЕТОДАМИ

Как указывалось выше, прямые измерения колебаний базилярной мембраны осуществимы пока лишь в условиях достаточно больших интенсивностей звукового сигнала. Если улитка является нелинейной системой, то полученные в этих условиях характеристики могут существенно отличаться от тех, которые имеют место при слабых звуках.

Вся информация о характеристиках «фильтров улитки» в условиях слабых сигналов получена косвенными методами. Трудность интерпретации этих характеристик заключается в том, что они, в принципе, могут отражать свойства не только преобразования амплитуды звукового давления входного сигнала в амплитуду

колебания базилярной мембраны, но и дополнительных преобразований}.

Важнейший непрямой метод основан на регистрации импульсной активности в одиночных волокнах слухового нерва. Как уже говорилось, колебания базилярной мембраны приводят к возбуждению рецепторов — волосковых клеток (рис. 7.6), рассматриваемых обычно как механо-электрические датчики. К рецепторным клеткам подходят афферентные волокна от нервных клеток (первый слуховой нейрон), расположенных в спиральном ганглии. Нейрон имеет два волокна, из которых одно подходит к рецепторной клетке (клеткам), другое (аксон) направляется в мозг в составе слухового нерва. Сигнал, генерируемый волосковой клеткой, и вызванный им сигнал, образуемый в начальном, близком к волосковой клетке, участке афферентного волокна, без сильных искажений воспроизводит однополупериодно выпрямленные колебания базилярной мембраны [450, 452]. Аналоговый сигнал, возникший в афферентном волокне, распространяется по нему со скоростью, зависящей от диаметра волокна, при этом постепенно затухая по амплитуде. В определенной точке волокна (положение ее точно неизвестно) происходит образование нервного импульса, который передается дальше по аксону в центральную нервную систему.

Окончательной модели, описывающей преобразование колебаний базилярной мембраны в последовательность нервных импульсов в волокнах слухового нерва, пока нет.

Исследование связей между нейронами и волосковыми клетками (рис. 7.6) показывает, что примерно 95% нейронов (всего их у кошки около 50 000, а у человека около 30 000) связано с внутренними волосковыми клетками, причем каждый из этих нейронов связан только с одной волосковой клеткой. Одна внутренняя волосковая клетка соединена с 10—12 нейронами [475, 476]. Таким образом, можно думать о том, что подавляющее число нейронов получают информацию от участков базилярной мембраны, ограниченных лишь площадями волосковых клеток.

Остальные 5% нейронов связаны с наружными волосковыми клетками, причем в данном случае одно афферентное волокно контактирует примерно с десятью клетками, расположенными вдоль мембраны (рис. 7.6). Есть основания считать, что преобразование аналогового сигнала в нервный импульс происходит на участке волокна, расположенного центрнее, ближе к телу нейрона, чем участки контакта волокна с волосковыми клетками [476]. Таким образом, возникает возможность суммирования в периферическом участке волокна аналоговых сигналов, поступающих от нескольких соседних волосковых клеток, связанных соответственно с расстроенными друг относительно друга «фильмами улитки». Очевидно, что результаты такого суммирования могут быть достаточно разнообразными в зависимости от фазовых отношений между сигналами в волосковых клетках, связанных



с одним волокном, скорости распространения аналоговых сигналов по волокну и т. д. [114].

Следует также заметить, что хотя гистологические данные говорят о том, что внутренние и наружные волосковые клетки связаны с разными нейронами, некоторые электрофизиологические факты, напротив, указывают на то, что один и тот же нейрон получает информацию и от тех и от других клеток [551].

Таким образом, сейчас нельзя исключить возможность того, что между колебаниями базилярной мембраны и возникновением импульса в одиночном волокне слухового нерва заключены некоторые преобразования, которые обеспечивают дополнительное

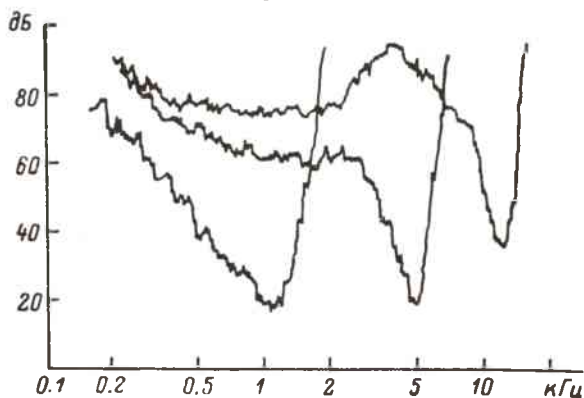


Рис. 7.19. Частотно-пороговые характеристики трех волокон слухового нерва кошки, полученные Кьянгом и Моксоном. По [543].

По оси абсцисс — частота; по оси ординат — уровень звукового давления у барабанной перепонки над абсолютным порогом.

увеличение частотной избирательности периферической слуховой системы как спектрального анализатора.

Ниже мы сделаем предположение, что таких дополнительных преобразований нет, и будем условно считать, что: 1) каждый нейрон (посылающий одно волокно в слуховой нерв) связан с определенной «точкой» мембраны; 2) условием возбуждения нейрона является отклонение мембраны на некоторую пороговую величину в одну определенную сторону.

При этих допущениях зависимость величины порогового звукового давления, необходимого для возникновения импульса, от частоты воздействующего тонального сигнала, так называемая частотно-пороговая кривая нейрона, может рассматриваться как сквозная амплитудно-частотная характеристика «фильтра улитки», измеренная пороговым методом. Частота тона, соответствующая наименьшему пороговому уровню звукового давления входного сигнала — так называемая характеристическая частота нейрона, — отражает при этих допущениях резонансную частоту «фильтра улитки».

Частотно-пороговые характеристики слухового нерва обычно получаются при уровнях сигнала на характеристической частоте, на 10—20 дБ превышающих уровень абсолютной чувствительности. Кривые, полученные этим методом, приведены на рис. 7.19.

Другой не прямой метод определения амплитудно-частотных характеристик «фильтра улитки» основывается на психоакустическом эксперименте [149, 469, 516]. Он заключается в том, что на различных частотах определяются значения амплитуды маскирующего тонального сигнала, необходимые для того, чтобы замаскировать восприятие фиксированного по частоте и уровню интенсивности тестирующего тона.

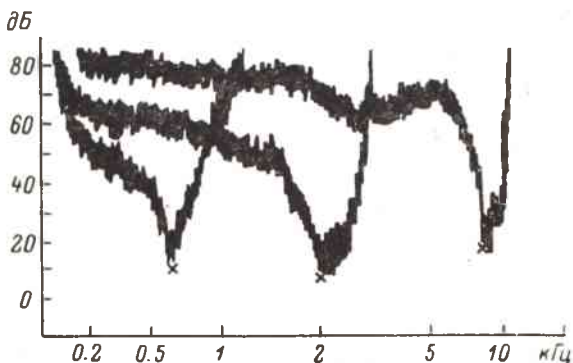


Рис. 7.20. Частотные характеристики эффекта маскировки. По [543].

По оси абсцисс — частота; по оси ординат — уровень звукового давления маскирующего тона над абсолютным порогом. Уровни маскируемых сигналов обозначены крестиками.

Допустив, что для обнаружения сигнала необходимо постоянное (не зависящее от частоты маскера) отношение амплитуды (энергии) сигнала к амплитуде (энергии) маскера на выходе «фильтра улитки», настроенного на сигнал, мы получаем в этом эксперименте пороговым методом амплитудно-частотную характеристику «фильтра улитки».

Частотные характеристики эффекта маскировки при уровнях тестирующего тона, всего на 5 дБ превышающих уровень абсолютного порога, приведены на рис. 7.20. Эти кривые и кривые, представленные на рис. 7.19, были получены примерно при одинаковых интенсивностях сигналов, т. е. в сравнимых условиях.

Сопоставление характеристик, полученных в нейрофизиологических и психоакустических экспериментах, указывает на их большое сходство, которое выражается прежде всего в том, что и те и другие имеют четко выраженную область частот, где система обладает повышенной чувствительностью, причем чувствительность системы в области характеристической частоты оказывается на 40—50 дБ больше чувствительности в области более низких частот, где характеристика становится относительно равномерной.

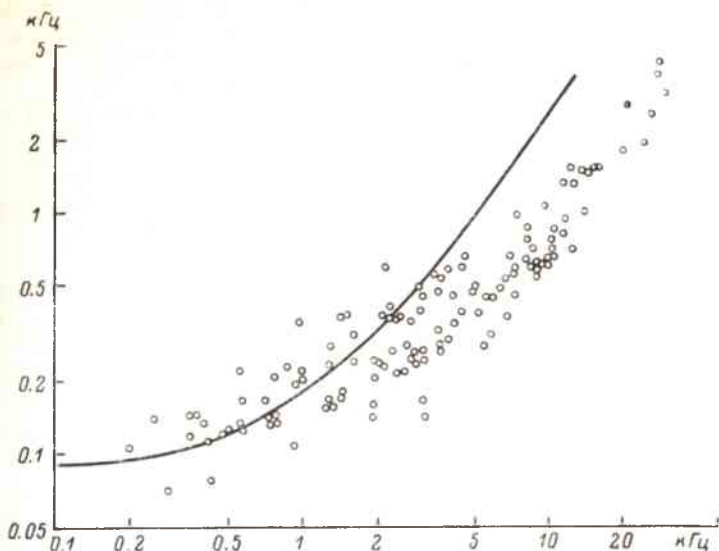


Рис. 7.21. Ширина полосы частотно-пороговых кривых волокон слухового нерва с различными характеристическими частотами, по данным измерений на кошках. По [252].

По оси абсцисс — частота; по оси ординат — ширина полосы. Кривая — зависимость ширины «критической полосы» от частоты, полученная в психоакустических экспериментах на человеке.

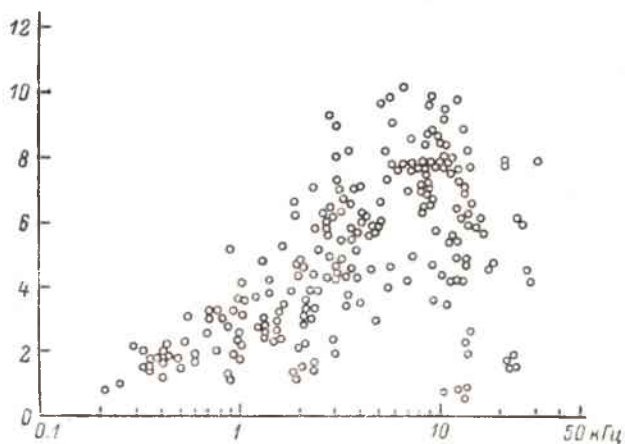


Рис. 7.22. «Добротность» частотно-пороговых характеристик нервных волокон с различной характеристической частотой. По [252].

По оси абсцисс — характеристическая частота волокна; по оси ординат — добротность частотно-пороговых характеристик, определяемая как отношение характеристической частоты к ширине полосы.

С учетом принятых выше допущений это значит, что амплитудно-частотные характеристики «фильтра улитки» имеют максимумы, на 40—50 дБ превышающие уровень характеристики на низких частотах.

Заметим, что зависимость ширины максимума амплитудно-частотной характеристики «фильтра улитки» от частоты, определяемая из кривых маскировки, хорошо совпадает с такой же зависимостью для ширины «критических полос» слухового спектрального анализатора [139, 540], измеряемых также с помощью большого числа других психоакустических методов [145].

Как можно видеть из рис. 7.21, зависимость ширины «критических полос» спектрального слухового анализатора от частоты, измеряемая на человеке, и зависимость от частоты ширины области максимальной чувствительности волокон слухового нерва кошки,<sup>5</sup> определенной по частотно-пороговым характеристикам, практически совпадают на низких и средних частотах и несколько различаются на частотах выше 4 кГц, что может быть объяснено различием объектов измерения. При этом максимальные значения «добротности» характеристик «фильтров улитки», определяемой как отношение характеристической частоты волокна к ширине полосы частотно-пороговой кривой, составляет не более десяти (рис. 7.22). С учетом того, что крутизна склонов амплитудно-частотных характеристик по обе стороны от максимума может достигать в сторону низких частот 100—150 дБ на октаву, а в сторону высоких частот несколько сотен дБ на октаву [246, 252], это означает, что амплитудно-частотные характеристики «фильтра улитки» по форме приближаются к характеристикам полосовых фильтров.

Другое очень важное свойство амплитудно-частотных характеристик «фильтра улитки», наблюдаемое и в нейрофизиологических, и в психоакустических экспериментах, заключается в том, что характеристики сильно изменяются с изменением интенсивности сигналов. Это наблюдается прежде всего в экспериментах по измерению частотных характеристик эффекта маскировки при различных уровнях тестирующего (маскируемого) сигнала (рис. 7.23). Видно, что с увеличением уровня сигнала характеристики уплощаются, причем наибольшие изменения наблюдаются на частотах вблизи частоты маскируемого тона.

На уровне волокон слухового нерва нелинейность характеристик была показана в экспериментах с измерением частотно-пороговых характеристик при добавлении к тестирующему сигналу узкополосного низкочастотного шумового сигнала [319]. При этом с увеличением уровня шумового сигнала форма характеристики волокна с высокой характеристической частотой сильно изменяется (рис. 7.24).

---

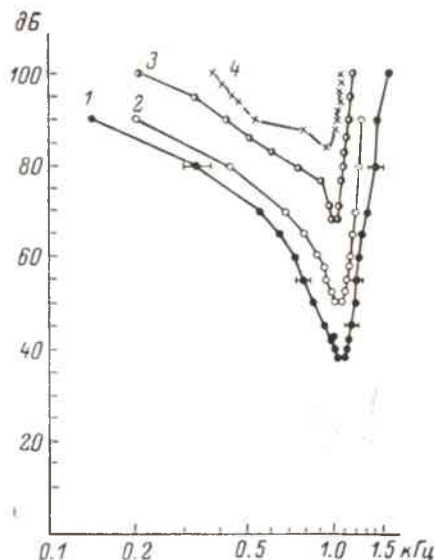
<sup>5</sup> Ширина частотно-пороговых характеристик обычно определяется на уровне 10 дБ относительно уровня на характеристической частоте.

На основании этих данных можно сделать вывод, что наиболее заметные изменения амплитудно-частотных характеристик происходят в области максимума, причем они в основном сводятся к уменьшению амплитуды максимума на 40—50 дБ и увеличению в несколько раз полосы пропускания. Это подтверждается также тем, что величины «добротности» частотно-пороговых кривых волокон слухового нерва кошки [232] на характеристических частотах в области 8—16 кГц изменяются в зависимости от пороговой чувствительности волокна на характеристической частоте (рис. 7.25).

Чрезвычайно большое сходство характеристик, получаемых на нейрофизиологическом и психоакустическом уровнях, позволяет при-

Рис. 7.23. Зависимость формы частотных характеристик эффекта маскировки от уровня звукового давления маскируемого тона. По [515].

По оси абсцисс — частоты маскирующих тонов; по оси ординат — уровни их звукового давления над абсолютным порогом. Длительность маскируемого сигнала 50 мс. Цифрами 1, 2, 3 и 4 обозначены кривые, полученные при уровнях звукового давления маскируемого сигнала, соответственно равных 10, 20, 40 и 50 дБ над абсолютным порогом.



нять, что основные параметры избирательных элементов слухового спектрального анализатора оказываются сформированными уже на уровне волокон слухового нерва [212, 248, 319, 543].

Представляет большой интерес сопоставить характеристики «фильтра улитки», определяемые косвенными методами, в частности на уровне волокон слухового нерва, с амплитудно-частотными характеристиками базилярной мембраны улитки, измеренными экспериментально. Поскольку амплитудно-частотные характеристики «фильтров улитки», определяемые косвенными методами, практически не изменяются с изменением интенсивности сигналов в области низких (на 1÷2 октавы ниже частоты максимума) частот, такое сравнение представляется вполне допустимым.

Учитывая, что амплитудно-частотные характеристики колебания базилярной мембраны у кошек пока еще не измерялись и что на низких и средних частотах характеристика  $P(x, \omega)_{x=x_0}/P_0(\omega)$  должна (вследствие горизонтального характера кривой  $a_m(x, \omega)/P(x, \omega)_{x=x_0}$  на этих частотах) воспроизводить форму амплитудно-частотной характеристики  $a_m(x, \omega)/P_0(\omega)$  (см. раз-



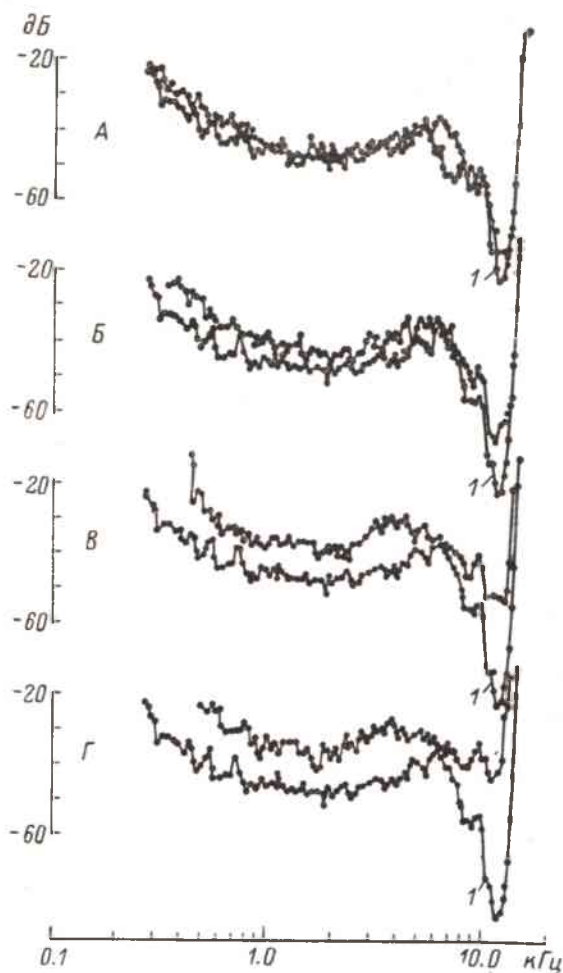


Рис. 7.24. Влияние уровня низкочастотного полосового шума (полоса 500 Гц, средняя частота 1 кГц) на форму частотно-пороговой характеристики нервного волокна с высокой характеристической частотой. По [319].

По оси абсцисс — частота; по оси ординат — уровень звукового давления сигнала относительно максимального значения. На А, Б, В и Г: частотно-пороговые кривые до подачи шумового сигнала (1) и при добавлении шумового сигнала с интенсивностями соответственно —45, —40, —35 и —30 дБ.

дел 7.3.1), частотно-пороговые характеристики волокон слухового нерва, полученные на кошках, целесообразно сравнивать с амплитудно-частотной характеристикой разностного давления в базальной части улитки, также полученной на кошках пороговым методом [405] (см. раздел 7.3.2). Можно видеть (рис. 7.26), что совпадение данных является практически идеальным.

Сопоставление сквозных амплитудно-частотных характеристик базилярной мембраны, измеряемых при постоянных уровнях звукового давления у барабанной перепонки (рис. 7.27), с амплитудно-

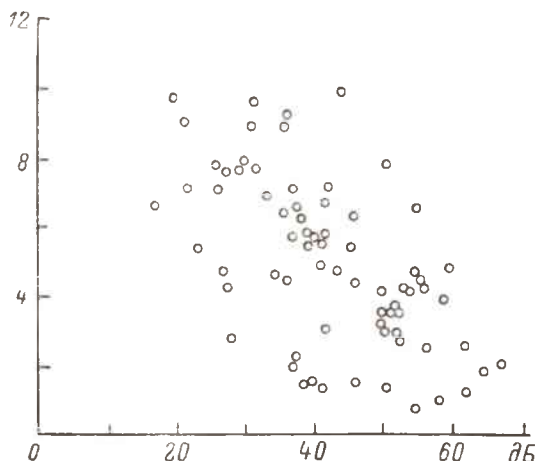


Рис. 7.25. «Добротность» частотно-пороговых кривых волокон слухового нерва кошки с характеристическими частотами в области 8—16 кГц, в зависимости от пороговой чувствительности волокна на характеристической частоте. По [282].

По оси абсцисс — пороговый уровень звукового давления, необходимый для возникновения импульсации волокна на характеристической частоте; по оси ординат — добротность частотно-пороговых кривых.

частотными характеристиками «фильтра улитки», получаемыми с помощью косвенных методов, показывает, что и в области максимума характеристики достаточно хорошо совпадают, если для сравнения использовать частотно-пороговые кривые волокон слухового нерва, полученные в условиях маскировки сильным шумом (рис. 7.24), или частотные характеристики эффекта маскировки (рис. 7.23), снятые при больших уровнях интенсивности тестирующего сигнала.

Значит ли это, что амплитудно-частотные характеристики базилярной мембраны при низких уровнях сигналов имеют такую же форму, как и частотно-пороговые характеристики при тех же условиях, пока неясно. Действительно, если характер нелинейного процесса, сопровождаемого изменением формы амплитудно-частотных характеристик базилярной мембраны и наблюдаемого

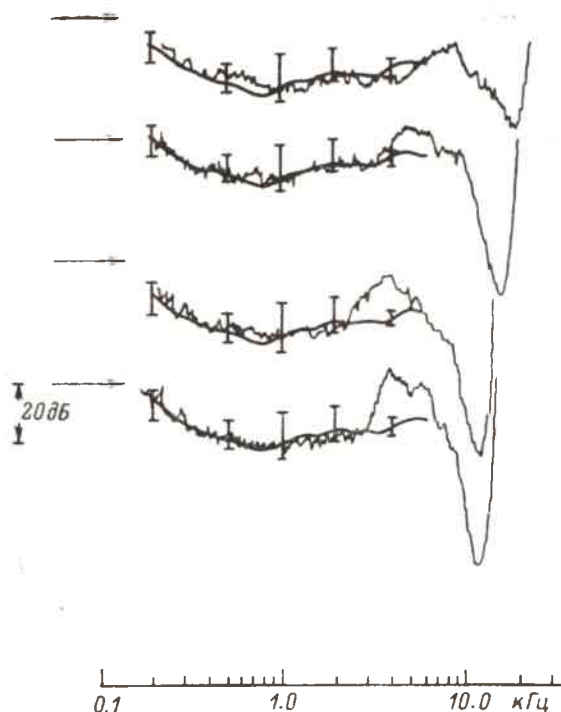


Рис. 7.26. Частотно-пороговые характеристики четырех волокон слухового нерва кошки (тонкие линии), полученные Кьянгом и Моксоном, и зависимость от частоты амплитуды звукового давления у барабанной перепонки, необходимого для создания постоянной амплитуды разностного давления в базальной части улитки (толстые линии). По [405].

По оси абсцисс — частота; по вертикали — уровень звукового давления у барабанной перепонки. Стрелками отмечен уровень звукового давления у барабанной перепонки 90 дБ над порогом. Кривые давления совмещены с пороговыми характеристиками на частоте 1 кГц.

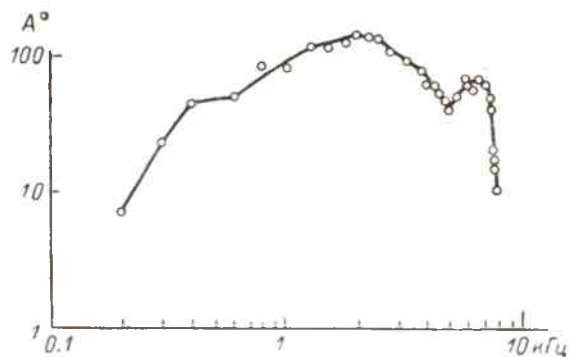


Рис. 7.27. Амплитудно-частотная характеристика базиллярной мембраны мартышки при постоянном уровне звукового давления у барабанной перепонки, равном 80 дБ над порогом. По [443].

По оси абсцисс — частота; по оси ординат — амплитуда (в ангстремах).

при уровнях входных сигналов 70—90 дБ над порогом (рис. 7.12), сохраняется и на более низких уровнях сигналов, этого может оказаться вполне достаточным, чтобы получить характеристики, подобные тем, которые наблюдаются на уровне слухового нерва.

Основные возражения против такого объяснения сводятся, во-первых, к тому, что для получения характеристик, которые наблюдаются на нейронном уровне, необходимы большие добротности избирательных элементов мембраны, что трудно допустить, если учесть, что колебания среднего канала улитки осуществляются в жидкости с достаточно большой вязкостью. Кроме того, при измерении частотно-пороговых характеристик волокон слухового нерва, даже на одном и том же объекте, отмечается очень сильный разброс параметров характеристик (рис. 7.22, 7.25), что опять-таки говорит о существовании между уровнем механических колебаний улитки и уровнем генерации нервных импульсов преобразователя с достаточно нестабильными свойствами. В этой связи интересно также отметить, что частотно-пороговые характеристики волокон слухового нерва значительно изменяются (и эти изменения аналогичны тем, которые происходят под воздействием сильных звуков) по мере развития отравления ототоксическим ядом или при гипоксии, т. е. в условиях, когда нарушается состояние рецепторных клеток [248].

С другой стороны, целый ряд фактов, наблюдаемых в психоакустических экспериментах, и прежде всего связанных с восприятием так называемого «кубического разностного тона» [277], пока невозможно объяснить, не предположив, что уже на уровне механических колебаний улитки должна быть нелинейность, особенно сильно выраженная при малых уровнях сигналов [275, 538].

Таким образом, сейчас вопрос сводится к тому, где — в механической или рецепторной части «фильтра улитки» — локализуется эта нелинейность или как она распределяется между этими двумя частями.

Можно утверждать, что в настоящее время это является наиболее актуальным вопросом теории слухового спектрального анализа, и очень большое число работ посвящено поиску, экспериментальному исследованию и теоретическому анализу эффектов, которых следовало бы ожидать от модели с нелинейными свойствами [225, 252, 275, 472].

## Глава 8

### ПРЕДСТАВЛЕНИЕ СИГНАЛА В «ЧАСТОТНОМ КАНАЛЕ» СЛУХОВОГО НЕРВА

Как уже упоминалось в главе 7, примерно 95% первичных слуховых нейронов связано с внутренними волосковыми клетками, причем каждый нейрон связан только с одной волосковой клеткой. Это означает, что стимулом для данного нейрона является не звуковой сигнал в целом, а тот колебательный процесс, который вызван сигналом в точке базилярной мембраны, где находится соответствующая волосковая клетка. Этот колебательный процесс — выходной сигнал «фильтра улитки» — мы будем дальше обозначать как  $x(t)$ .

При рассмотрении вопроса о том, как отражается  $x(t)$  в нервной импульсации, поступающей по слуховому нерву, целесообразно принять, что слуховой нерв представляет собой набор частотных каналов. В частотный канал входит группа волокон тех нейронов, которые связаны с одним и тем же или близкими фильтрами улитки, т. е. имеют совпадающие или близкие характеристические частоты. Сигналом, передаваемым по частотному каналу, является в таком случае плотность (мгновенная частота) импульсации. Обозначим ее  $g(t)$ . Плотность импульсации может рассматриваться как непрерывная функция времени.

Представление о том, что именно плотность импульсации в группе волокон, а не какие-то характеристики последовательности импульсов в единичном волокне, является сигналом, используемым центральными отделами слуховой системы, кажется физиологически достаточно обоснованным. Центральный нейрон получает, как правило, импульсы от многих волокон; возникающий в нейроне возбуждающий постсинаптический потенциал отражает, с определенным сглаживанием по времени, мгновенные значения суммарной приходящей импульсации (см. главу 9).

Введение понятий «частотный канал» и «плотность импульсации» в принципе позволяет строить такую функциональную модель периферической слуховой системы, которая не включает нейронов и рецепторных клеток. На выходе фильтров спектрального анали-



затора достаточно поставить преобразователи  $x(t)$  в  $g(t)$ , формально воспроизводящие отношения между этими переменными.

Заметим, что  $g(t)$  непосредственно недоступна прямому наблюдению. Излагаемые ниже представления о ее свойствах основаны на результатах исследования импульсации в одиночных волокнах слухового нерва и на психоакустических данных.

### 8.1. ВОСПРОИЗВЕДЕНИЕ ФОРМЫ КОЛЕБАНИЙ

Вывод о том, что  $g(t)$  достаточно точно воспроизводит форму одной полуволны возбуждающего сигнала  $x(t)$ , следует из ряда экспериментов, в которых исследовалась вероятность возникновения импульса в одиночном волокне слухового нерва на разных отрезках периода возбуждающего периодического звукового сигнала [201, 450, 452].

На рис. 8.1 приведены результаты работы [450]. В данном случае период (частота равна 907 Гц) разбивался на интервалы по 60 мкс. Стимул (сумма синусоидальных сигналов с частотами 907 и 1814 Гц и разными значениями сдвига фаз) действовал непрерывно в течение 10 с. За это время подсчитывалось число случаев, когда импульс попадал в первый, второй и т. д. отрезок периода, и строилась соответствующая гистограмма. Такие гистограммы, показывающие изменение вероятности импульсации на протяжении периода стимула, получили название цикло-гистограмм. Из рис. 8.1 видно, что огибающая цикло-гистограммы достаточно хорошо воспроизводит форму положительной полуволны возбуждающего колебания. На протяжении отрицательной полуволны импульсация близка к нулю. Существенно, что воспроизведение формы волны сохраняется вплоть до весьма высоких частот, по крайней мере до 4—5 кГц. На высоких частотах начинает сказываться эффект случайных колебаний скорости проведения импульса по нервному волокну.

Из полученных данных можно сделать заключение, что функциональная модель преобразования  $x(t)$  в  $g(t)$  должна включать однополупериодное выпрямление и, очевидно, не должна сильно сглаживать выпрямленные колебания.

Интересный факт был обнаружен при изменении уровня интенсивности воздействующего сигнала [451]. Оказалось, что хотя средняя частота импульсации сравнительно быстро достигает насыщения и уже не увеличивается при дальнейшем усилении сигнала, существенных искажений формы волны на цикло-гистограмме не наблюдается — не видно уплощения пиков, характерного для технического канала с амплитудными ограничениями.

### 8.2. ЭФФЕКТ ПЕРИФЕРИЧЕСКОЙ АДАПТАЦИИ

Распространенным методом электрофизиологического исследования является получение постстимульных гистограмм. Посылка стимула предъявляется многократно; отрезок времени,

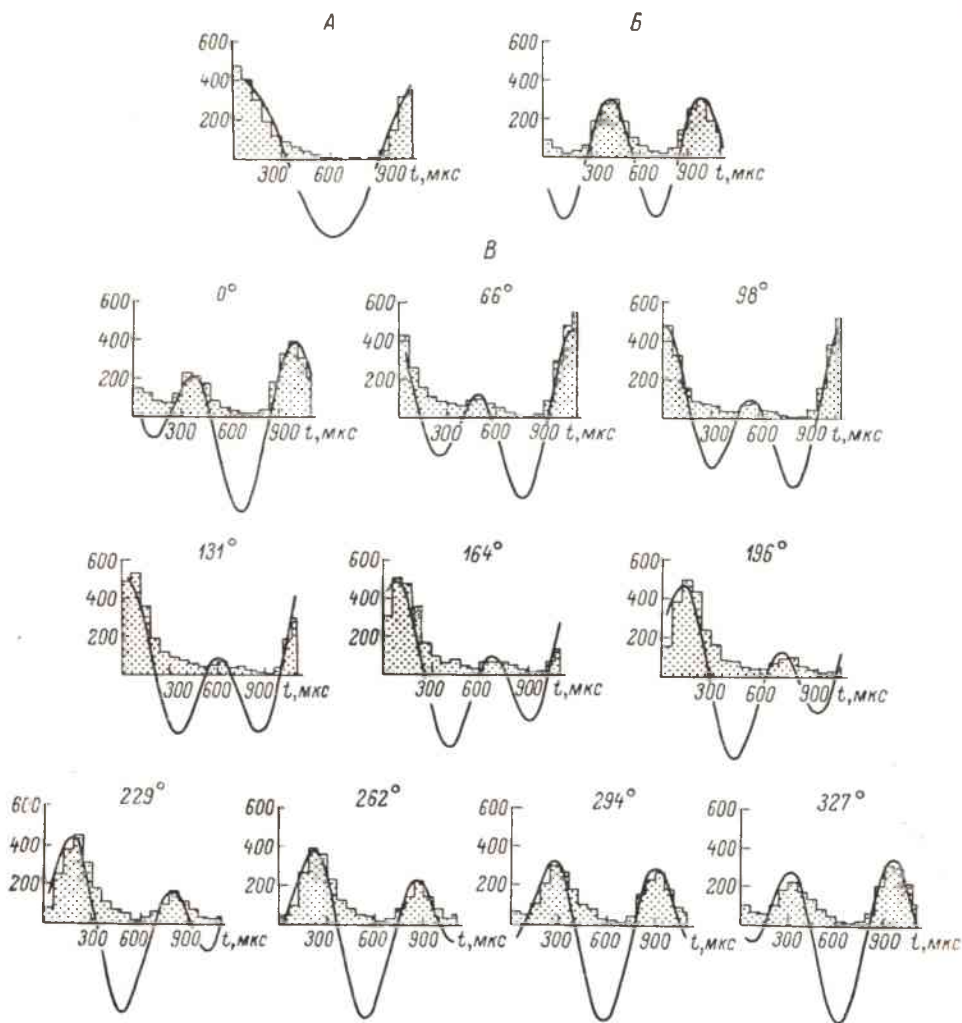


Рис. 8.1. Цикло-гистограммы ответов одиночного волокна слухового нерва при действии тонов с частотами 907 Гц (А) и 1814 Гц (Б) или при действии суммы этих тонов (Б). По [460].

По оси абсцисс — время относительно начала каждого периода колебания с частотой 907 Гц; по оси ординат — число импульсов в каждом из шестидесятимикросекундных интервалов, на которые разбит один период стимула (время накопления — 10 с). Цифры над графиками — значения фазового сдвига тона с частотой 1814 Гц. Кривые — один (А) или два (Б) периода аппроксимирующих синусоид с частотами, равными частоте стимула, или один период сложного колебания (остальные графики), соответствующего сумме этих двух синусоид.

привязанный к началу стимула, разбивается на интервалы; определяется количество импульсов, попавших в каждый из интервалов. Постстимульная гистограмма позволяет судить об изменении плотности импульсации на протяжении стимула.

Постстимульные гистограммы, полученные для волокон слухового нерва при действии тональной посылки с прямоугольной огибающей, показывают, что плотность импульсации увеличивается скачком в момент включения сигнала, а затем постепенно убывает до некоторого установившегося значения; после выключения сигнала плотность импульсации мгновенно падает до нуля,

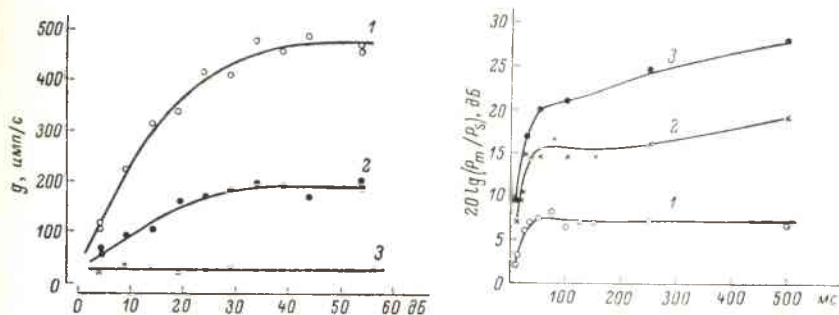


Рис. 8.2. Амплитудные характеристики одиночного волокна слухового нерва. По [471].

По оси абсцисс — интенсивность входного сигнала; по оси ординат — плотность импульсации. 1 — амплитудная характеристика в момент включения сигнала, 2 — амплитудная характеристика в установившемся режиме, 3 — спонтанная активность.

Рис. 8.3. Зависимость величины маскировки от длительности маскира. По [160].

По оси абсцисс — длительность маскира; по оси ординат — величина маскировки. Параметр кривых — уровень интенсивности маскира в тестируемом ухе относительно порога обнаружения маскира при длительности 500 мс. 1 — 20, 2 — 40, 3 — 60 дБ.

и в течение некоторого времени наблюдается подавление спонтанной активности. Уменьшение плотности импульсации на протяжении стимула (и временное угнетение спонтанной активности после окончания стимуляции) определяется как проявление адаптации. В эффекте адаптации можно различить по крайней мере два компонента — быстрый и медленный. Процессы, связанные с быстрым компонентом, устанавливаются в течение первых 100—200 мс после включения сигнала; медленное изменение плотности импульсации продолжается в течение десятков минут [321].

Эффект адаптации наблюдается во всем диапазоне интенсивностей стимула [471]. На рис. 8.2 приведены данные [471] для одиночного волокна слухового нерва. Стимулом служила посылка с частотой, равной характеристической частоте нейрона. Кривая 1 показывает среднюю частоту импульсации в течение первых 10 мс после включения сигнала; кривая 2 — среднюю частоту импульсации через 180 мс от начала тона.

Зависимость между средней частотой импульсации в волокне и уровнем интенсивности стимула называется амплитудной характеристикой нейрона. Из рис. 8.2 можно видеть, что амплитудная характеристика различна в момент начала стимула и в установившемся режиме. В установившемся режиме компрессия сигнала оказывается существенно большей.

Очевидно, что эффект адаптации должен, в принципе, обнаруживаться и в психоакустическом эксперименте. Так, естественно ожидать эффекта остаточной маскировки — повышения порога обнаружения тестирующего стимула, предъявляемого вскоре после окончания маскира. Величина маскировки определяется как  $20 \lg (P_m/P_s)$ , где  $P_m$  — пороговое звуковое давление тестирующего стимула после маскира,  $P_s$  — пороговое звуковое давление тестирующего стимула в тишине. Зависимость маскировки от величины паузы между маскиром и сигналом должна отражать процесс исчезновения адаптации. Эта зависимость исследовалась многими авторами [120, 160, 245, 549]. Результаты показывают, что нормальная чувствительность восстанавливается за 150—300 мс, что хорошо согласуется с данными, полученными при электрофизиологическом исследовании периферической адаптации, но при малых величинах паузы между маскиром и сигналом величина маскировки оказывается значительно большей, чем следует из электрофизиологических данных.

Имеются основания считать, что при малых паузах эффект остаточной маскировки определяется не только периферическими, но и центральными процессами. Это следует, в частности, из экспериментов [284], в которых маскир и сигнал предъявлялись бинаурально и применялись два варианта подачи сигнала: 1) сигналы в обоих ушах совпадают по фазе; 2) сигналы сдвинуты по фазе на  $180^\circ$ .

Во втором случае маскир и сигнал воспринимаются как разнесенные в пространстве. Оказалось, что фазовый эффект очень мал, если пауза между маскиром и сигналом превышает 30 мс, а для пауз, меньших 10 мс, сдвиг фазы на  $180^\circ$  приводит к резкому уменьшению маскировки. Отсюда следует, что для измерения периферических компонентов остаточной маскировки нужно создать такие условия, при которых маскир и сигнал будут восприниматься как разнесенные в пространстве звуки, и использовать данные, полученные для пауз, больших 30—40 мс. В экспериментах [160] тестирующий стимул подавался на одно ухо, а маскир — одновременно на оба, причем интенсивность маскира в тестируемом ухе была на 20 дБ ниже, чем в противоположном. При таких сигналах испытуемый воспринимал маскир и стимул как два независимых звука, действующих на разные уши. На рис. 8.3 приведены результаты исследования зависимости величины маскировки от длительности маскира [160]. Частота маскира и тестирующего стимула была равна 3 кГц, длительность стимула — 3 мс, пауза между концом маскира и началом сти-

мула — 40 мс. Можно видеть, что маскировка резко возрастает при увеличении длительности маскера до 50—70 мс, а затем либо остается постоянной, либо медленно увеличивается (для больших интенсивностей маскера). Отсюда можно заключить, что при больших интенсивностях сказывается медленный компонент адаптации; при небольших интенсивностях и длительностях маскера им можно пренебречь. Таким образом, при исследовании быстрого компонента периферической адаптации следует использовать результаты только тех экспериментов по остаточной маскировке, в которых маскер имел достаточно малую длительность и интенсивность.

### 8.3. МОДЕЛИ ПРЕОБРАЗОВАНИЯ СИГНАЛА В ПЕРИФЕРИЧЕСКОЙ СЛУХОВОЙ СИСТЕМЕ

Примем, что входной сигнал модели  $x(t)$  идентичен предъявляемому в эксперименте тональному стимулу, а выходной сигнал модели  $g(t)$  — плотность импульсации в частотном канале — идентичен вероятности возникновения импульса в одиночном нейроне, измеряемой с помощью постстимульных гистограмм. Будем также считать, что порогу обнаружения сигнала в эксперименте по остаточной маскировке соответствует некоторое постоянное значение амплитуды  $g(t)$ . Тогда можно считать, что в случае входного сигнала с постоянной частотой работает только один канал, и мы получаем возможность использовать для построения модели как нейрофизиологические, так и психоакустические данные.

Заметим, что пренебречь неидентичностью нейронов, составляющих канал, можно только при небольших интенсивностях входного сигнала. Дело в том, что плотность импульсации отдельного нейрона зависит от интенсивности входного сигнала в довольно узком диапазоне его интенсивностей (около 40 дБ). В то же время известно, например, что сигналы различаются по громкости при изменении интенсивности входного сигнала в гораздо более широких пределах. Следовательно, амплитудные характеристики отдельного нейрона должны существенно отличаться от амплитудных характеристик частотного канала за счет неидентичности составляющих канал нейронов.

В данном разделе будет рассмотрено пять моделей преобразования сигнала в периферической слуховой системе. Для краткости будем называть модель, описанную в работе [462], моделью I, в работе [471] — II, в работе [162] — III, в работе [211] — IV и в работе [338] — моделью V. Модели I и II построены в соответствии с нейрофизиологическими данными, в трех остальных моделях использованы данные по остаточной маскировке. Модели имеют следующие общие свойства:

1) все пять моделей описывают преобразование сигнала в одном частотном канале;



2) величина входного сигнала моделей предполагается пропорциональной звуковому давлению стимула;

3) в моделях II, III, IV предусмотрено однополупериодное выпрямление; в модели I используется «несовершенный» выпрямитель, пропускающий и отрицательную полуволну, но значительно ослабленную по уровню;

4) во всех моделях отсутствует сглаживание сигнала;

5) все модели описывают только быстрый компонент периферической адаптации.

При сравнении работы моделей с экспериментом в качестве входного сигнала будет, как правило, рассматриваться огибающая стимула. Это возможно потому, что адаптация — достаточно медленный процесс по сравнению с применявшимися в экспериментах несущими частотами.

Все предложенные модели можно разделить на две группы: в одной, к которой относятся модели I, III и V, используется переменный коэффициент передачи, в другой, модели II и IV, — переменный порог.

#### 8.3.1. МОДЕЛИ С ПЕРЕМЕННЫМ КОЭФФИЦИЕНТОМ ПЕРЕДАЧИ

Все три модели с переменным коэффициентом передачи качественно воспроизводят эффект периферической адаптации. На рис. 8.4 показаны реакции модели I на посылку тона для трех интенсивностей входного сигнала. Из рисунка видно, что входной сигнал убывает во времени до установившегося значения, зависящего от интенсивности входного сигнала; после окончания входного сигнала спонтанная активность на некоторое время подавляется.

Наиболее привлекательным свойством моделей с управляемым коэффициентом передачи является возможность получить узкий динамический диапазон выходного сигнала в установившемся режиме, не вводя сильно компрессирующей безынерционной нелинейности.

На рис. 8.5 показана амплитудная характеристика модели I в установившемся режиме. Можно видеть, что она напоминает амплитудные характеристики отдельных волокон слухового нерва (рис. 8.2).

Коротко опишем конкретные модели. На рис. 8.6 приведены блок-схемы двух из трех рассмотренных в работе [162] вариантов модели III. На рис. 8.6 блок 1 соответствует безынерционной нелинейности, блок 2 — инерционному звену 1-го порядка; 3 — блок переменного коэффициента передачи.  $x$  — входной сигнал модели,  $y$  — выходной сигнал блока 1,  $r$  — выходной сигнал блока 2,  $g$  — выходной сигнал модели.



Вариант А описывается уравнениями

$$\begin{aligned}\tau \dot{r} + r &= y, \\ g &= yk, \\ k &= \frac{1}{\alpha + \beta r},\end{aligned}\tag{1}$$

а вариант В —

$$\begin{aligned}\tau \dot{r} + r &= g, \\ g &= yk, \\ k &= \alpha - \beta r.\end{aligned}\tag{2}$$

$\tau$ ,  $\alpha$  и  $\beta$  — константы, разные для вариантов А и В.

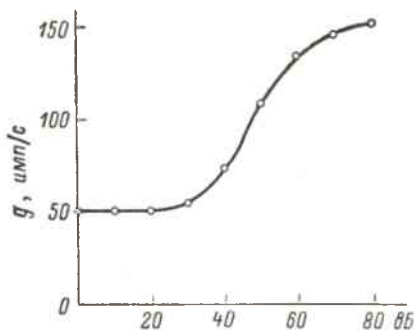
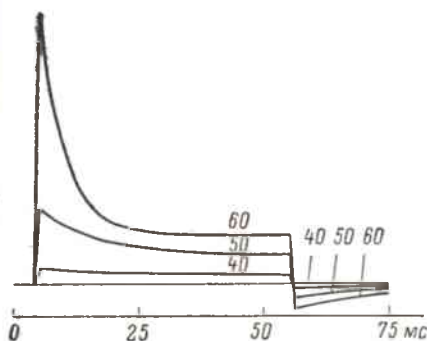


Рис. 8.4. Огибающие реакции модели I на посылку тона. По [462].

По оси абсцисс — время. Параметр кривых — интенсивность входного сигнала (в дБ от условно принятого порога).

Рис. 8.5. Амплитудная характеристика модели I в установившемся режиме. По [462].

По оси абсцисс — интенсивность входного сигнала (в дБ) от условно принятого порога по оси ординат — выходная плотность импульсации.

Нелинейность блока I описывается уравнениями

$$\begin{aligned}y &= \left(1 + 0.5 \lg \frac{x}{x_0}\right)^{1.6} \quad \text{при } x \geq x_0, \\ y &= x \quad \text{при } 0 \leq x < x_0, \\ y &= 0 \quad \text{при } x < 0,\end{aligned}\tag{3}$$

однако для качественного сравнения моделей конкретный вид нелинейности не имеет значения. Существенно только, что для достаточно большого диапазона входных сигналов зависимость  $y(x)$  мало отличается от линейной.

Амплитудная характеристика обоих вариантов модели III для момента включения определяется только безынерционной нелинейностью блока I. В установившемся режиме для варианта А  $g = y/(\alpha + \beta y)$ , т. е. имеет место сильная компрессия сигнала с насы-

щением (при  $\beta y \gg \alpha$   $g \approx 1/\beta$ ). Амплитудная характеристика варианта *Б* в установившемся режиме при надлежащем подборе постоянных совпадает с амплитудной характеристикой варианта *А*.

Для выбора между вариантами *А* и *Б* был определен для обоих вариантов аналитический вид зависимостей  $P_m/P_s$  (см. раздел 8.2) от длительности маскира и интервала между маскером и тестирующим сигналом. Оказалось, что экспериментальные данные хорошо согласуются с вариантом *А* и не согласуются с вариантом *Б*.

Модель I основана на «физиологических» представлениях о преобразовании движения базилярной мембраны в плотность импульсации в волокне слухового нерва. Преобразование осуществляет

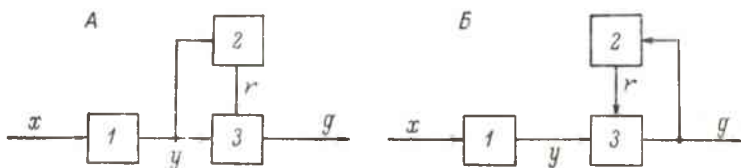


Рис. 8.6. Блок-схема вариантов *А* и *Б* модели III. По [142].

Обозначения см. в тексте.

некоторая гипотетическая «волосковая клетка» со следующими свойствами:

- 1) в волосковой клетке генерируются с постоянной частотой  $\gamma$  некоторые электрохимические «кванты»;
- 2) кванты исчезают, вызывая нервный импульс, с частотой, пропорциональной числу квантов  $n$  и зависящей от входного сигнала функции «проницаемости»,

$$y = y_0 \left[ \frac{1}{2} x + \left( \frac{1}{4} x^2 + 1 \right)^{\frac{1}{2}} \right], \quad (4)$$

где  $x$  — входной сигнал ( $x=1$  соответствует уровню звукового давления примерно 30 дБ),  $y_0$  — константа, связанная со спонтанной активностью нервного волокна;

- 3) кроме того, кванты исчезают, не вызывая импульса, с частотой  $\delta n$ , где  $\delta$  — константа.

Таким образом, выходная плотность импульсации в модели I

$$g = n y_0 \quad (5)$$

а  $n$  определяется дифференциальным уравнением

$$\dot{n} = \gamma - n y - n \delta. \quad (6)$$

Амплитудная характеристика модели I в установившемся режиме без учета входной нелинейности имеет такой же вид, как амплитудная характеристика модели III:

$$g = \gamma \frac{y}{\delta + y}.$$

Амплитудная характеристика модели I для момента включения определяется только входной безынерционной нелинейностью, т. е. для достаточно больших интенсивностей входного сигнала амплитудная характеристика линейна.

Если сравнить уравнения (5) и (6) с уравнениями (2), легко видеть, что модель I совпадает с вариантом B модели III; количество квантов  $n$  в модели I играет роль переменного коэффициента передачи. Модели отличаются только видом входной нелинейности. Это значит, что модель I, так же как вариант B модели III, несовместима с экспериментальными данными по остаточной маскировке.

К группе моделей с переменным коэффициентом передачи относится также модель V, блок-схема которой показана на рис. 8.7. Здесь блок 1 — блок переменного коэффициента передачи, блок 2 — инерционное звено I порядка,

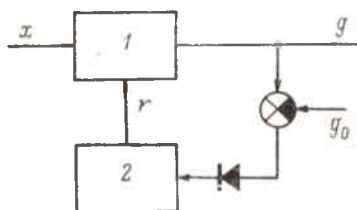


Рис. 8.7. Блок-схема модели V.  
По [338].

Обозначения см. в тексте.

$x$  — входной сигнал,  $r$  — выходной [сигнал блока 2,  $g$  — выходной сигнал модели.

Модель описывается уравнениями:

$$\begin{aligned} \dot{g} &= kx, \\ \tau \dot{r} + r &= g - g_0 \quad \text{при } g - g_0 > 0, \\ \tau \dot{r} + r &= 0 \quad \text{при } g - g_0 \leq 0, \\ k &= e^{-ar}, \end{aligned} \quad (7)$$

где  $\tau$  и  $a$  — константы.

Для построения модели были использованы данные по остаточной маскировке, но, к сожалению, только те из них, в которых маскер имел большую интенсивность (не менее 80 дБ) и большую длительность (не менее 300 мс) [120, 477, 549]. Как уже говорилось выше, такие данные малопригодны для моделирования быстрого компонента адаптации.

Модель V, как и остальные модели с переменным коэффициентом передачи, качественно воспроизводит эффект адаптации и позволяет получить узкий динамический диапазон выходного сигнала в установившемся режиме.

Однако все эти модели не согласуются с новыми электрофизиологическими данными, полученными Смитом [471]. Очевидно, что в любой модели с управляемым коэффициентом передачи реакция на приращение входного сигнала должна уменьшаться при увеличении промежутка времени  $T$  между началом сигнала и подачей приращения, так как коэффициент передачи со временем умень-

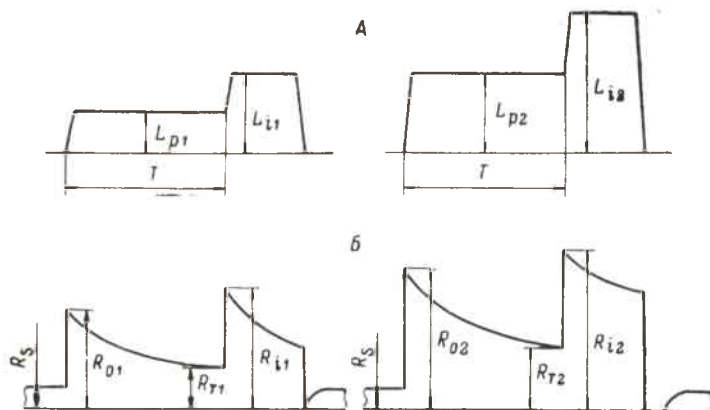


Рис. 8.8. Огибающие стимулов (А) и постстимульных гистограмм (Б) при исследовании ответов на приращение. По [471].

Обозначения см. в тексте.

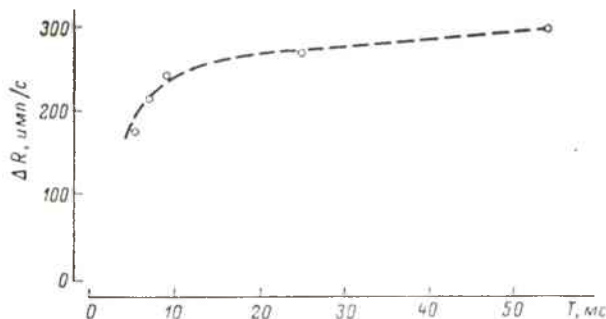


Рис. 8.9. Зависимость ответа ( $\Delta R$ ) на приращение от величины задержки ( $T$ ) при  $L_p=15$  дБ,  $\Delta L=21$  дБ. По [471].

По оси абсцисс — величина задержки; по оси ординат — величина ответа на приращение.

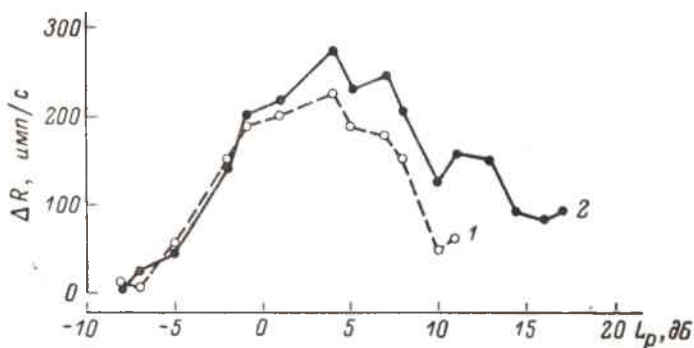


Рис. 8.10. Зависимости ответов ( $\Delta R$ ) на приращения от интенсивности пьедестала ( $L_p$ ) при  $\Delta L=6$  дБ. По [471].

По оси абсцисс — интенсивность пьедестала; по оси ординат — величина ответа на приращение. Параметр кривых — задержка приращения относительно начала стимула: 1 —  $T=0$ , 2 —  $T > 120$  мс.

шается. Смит показал, что для отдельного волокна слухового нерва это не так. В экспериментах Смита стимул представлял собой посылку тона с частотой, равной характеристической частоте нейрона. Огибающие стимулов и выходных сигналов (постстимульных гистограмм) показаны на рис. 8.8. Здесь  $L_p$  — интенсивность «пьедестала» в децибелах относительно порога,  $L_i$  — интенсивность сигнала после подачи приращения. Величина приращения определялась как  $\Delta L = L_i - L_p$ . Времена нарастания и убывания сигнала составляли 2.5 мс, шаг гистограммы — 10 мс. Ответы на приращения, поданные через время  $T$  от начала стимула, определялись как  $\Delta R = R_i - R_T$ . Так как выдерживалось соотношение  $L_{p2} = L_{p1}$ , можно было для той же величины приращения определить ответ в момент  $T=0$  как  $\Delta R = R_{02} - R_{01}$ .

На рис. 8.9 показана зависимость ответа на приращение  $\Delta R$  от величины задержки  $T$  для  $L_p = 15$  дБ,  $\Delta L = 21$  дБ. На рис. 8.10 приведены зависимости ответов на приращения от интенсивности пьедестала.

Как видно из рис. 8.9 и 8.10, реакция на приращение, поданное через время  $T$  от начала сигнала, либо равна реакции на приращение, поданное в момент включения сигнала (для малых  $L_p$  и  $\Delta L$ ), либо увеличивается с увеличением  $T$ . Кроме того, Смит обнаружил, что отношение величин ответов на пьедестал в момент включения сигнала и в установившемся режиме  $(R_0 - R_s)/(R_T - R_s)$  для одних волокон не зависит от интенсивности пьедестала и равно в среднем 2.5, для других — несколько уменьшается при больших интенсивностях пьедестала. В описанных выше моделях при увеличении входного сигнала это отношение должно увеличиваться.

### 8.3.2. МОДЕЛИ С ПЕРЕМЕННЫМ ПОРОГОМ

К моделям с переменным порогом относятся модели II и IV. Хотя модель IV основана на данных по остаточной маскировке, а модель II — чисто нейрофизиологическая, конструкции их весьма схожи.

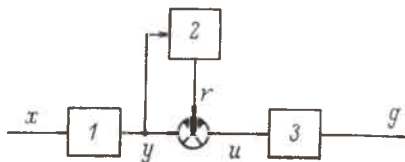


Рис. 8.11. Блок-схема моделей II и IV. По [211].

Обозначения см. в тексте.

На рис. 8.11 показана блок-схема, общая для этих моделей. Блоки 1 и 3 — безынерционные нелинейности, блок 2 — инерционное звено 1-го порядка;  $x$  — входной сигнал модели,  $y$  — выходной сигнал блока 1,  $r$  — выходной сигнал блока 2,  $g$  —



выходной сигнал модели. Линейная часть обеих моделей описывается уравнениями

$$\begin{aligned}\tau \dot{r} + r &= \alpha y, \\ u &= y - r,\end{aligned}\tag{8}$$

где  $\tau$  — постоянная времени инерционного звена,  $\alpha$  — константа. Постоянная времени  $\tau$  и в том, и в другом случае получалась равной 40 мс.

В модели II входная и выходная нелинейности не описаны аналитически, но предполагается, что обе они компрессирующие и выходной сигнал каждой из них равен нулю, если входной сигнал отрицателен.

В модели IV входная нелинейность такая же, как в модели III (см. уравнения (3)), а выходная нелинейность представляет собой просто постоянный порог:

$$\begin{aligned}g &= u - \beta \quad \text{при } u \geq \beta, \\ g &= 0 \quad \text{при } u < \beta.\end{aligned}\tag{9}$$

Если на вход линейной части модели подать ступеньку с амплитудой  $y_0$ , на выходе получим  $u = y_0(1 - \alpha) + y_0 e^{-t/\tau}$ .

По данным Смита, это выражение хорошо описывает огибающую постстимульных гистограмм, полученных для небольших интенсивностей входного сигнала при  $t > 25$  мс. На начальном участке ( $t < 25$  мс) в экспериментах проявляется действие еще одной экспоненциальной составляющей, с постоянной времени, примерно в 10 раз меньшей. Этот самый быстрый компонент адаптации подробно исследован не был, и в модели он не учитывается.

Рассмотрим, как модели II и IV согласуются с данными Смита по ответам на приращения интенсивности входного сигнала. При достаточно малых интенсивностях пьедестала  $L_p$  и приращения  $\Delta L$  обе модели можно считать линейными, поэтому реакция на приращение  $\Delta R$  в них не зависит от задержки  $T$  (рис. 8.9). При больших интенсивностях в модели II реакция на приращение будет увеличиваться с ростом  $T$  за счет выходной нелинейности. В модели IV реакция на приращение не зависит от  $T$  для любых интенсивностей входного сигнала. Аналогичным образом отношение величины ответа на пьедестал в момент включения входного сигнала к величине ответа в установившемся режиме в модели IV постоянно во всем диапазоне интенсивностей входного сигнала, в модели II это отношение уменьшается при больших интенсивностях входного сигнала за счет выходной нелинейности.

Что касается данных по остаточной маскировке, модель IV описывает их точно так же, как модель III вариант А.

Таким образом, для входных сигналов небольшой интенсивности модели II и IV эквивалентны и хорошо согласуются как с нейрофизиологическими, так и с психоакустическими данными. Для больших интенсивностей входных сигналов нейрофизиологические данные по отдельным нейронам лучше описываются мо-

делью II, т. е. нейрофизиологические данные указывают на необходимость компрессирующей выходной нелинейности.

На основании всего изложенного в настоящей главе можно сделать следующие выводы:

1) модели с переменным порогом воспроизводят преобразование сигнала в периферической слуховой системе лучше, чем модели с переменным коэффициентом передачи; однако модели с переменным порогом не могут обеспечить сильной компрессии сигналов, не искажая при этом формы сигнала;

2) пока остается неясным, должна ли функциональная модель частотного канала содержать одну или две компрессирующие нелинейности и каков их вид; поскольку динамический диапазон канала гораздо шире динамического диапазона отдельного нейрона, выяснить эти вопросы, вероятно, возможно только с помощью психоакустических экспериментов.

## Глава 9

### ПРИНЦИПЫ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА ЦЕНТРАЛЬНЫМИ ОТДЕЛАМИ СЛУХОВОЙ СИСТЕМЫ

Сигналом, поступающим в центральную слуховую систему, является пространственно-временной рисунок плотности импульсации в слуховом нерве. Мы будем его называть периферическим изображением стимула.

Примем, что функциональная модель периферической слуховой системы представляет собой упорядоченный по шкале частот (назовем ее шкалой  $z$ ) набор частотных каналов; каждый канал включает фильтр звуковых частот (см. главу 7) и преобразователь сигнала на выходе фильтра в плотность импульсации  $g(t)$  (см. главу 8). Соответственно периферическое изображение стимула обозначим как  $g(z, t)$ . Обработку сигнала центральными отделами слуховой системы естественно рассматривать как ряд преобразований  $g(z, t)$ .

Существующие электрофизиологические данные отнюдь не дают возможности точно описать эти преобразования, однако они позволяют составить качественные представления о том, на выделение какого типа признаков  $g(z, t)$  направлены эти преобразования.

Для формирования этих качественных представлений полезно начать с того, чтобы рассмотреть наиболее примитивную (включающую наименьшее число параметров) модель нейрона и выяснить, какие классы признаков  $g(z, t)$  могли бы выделяться такой моделью. Эти классы признаков можно будет дальше рассматривать как возможные гипотезы, правдоподобность которых должна быть оценена на основании экспериментальных данных электрофизиологии слуха.

#### 9.1. КАЧЕСТВЕННАЯ КЛАССИФИКАЦИЯ ВОЗМОЖНЫХ ПРИЗНАКОВ СИГНАЛА

Идеализированный центральный нейрон рассматривается как состоящий из двух элементов [114]. Первый элемент обеспечивает преобразование поступающего потока импульса-

ции в аналоговую переменную  $P(t)$ , характеризующую степень возбуждения нейрона. Вторым элементом обеспечивает преобразование  $P(t)$  в последовательность импульсов на выходе нейрона.

Физиологические явления, описываемые моделью первого элемента, обозначаются как пространственная и временная суммация процессов торможения и возбуждения и как взаимодействие этих процессов внутри клетки.

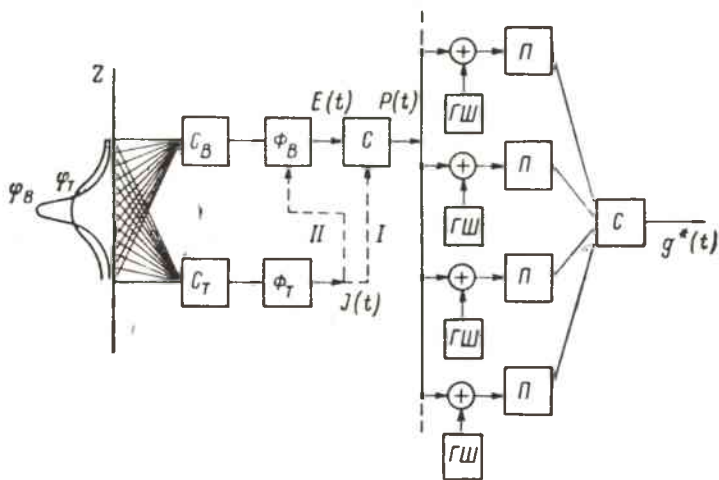


Рис. 9.1. Функциональная схема нейронного канала.

Обозначения см. в тексте.

Для описания пространственной суммации необходимо указать источники, от которых поступает импульсация на возбуждающие и тормозные входы, и приписать каждому источнику свой весовой коэффициент.

Допустим, что источниками импульсации для рассматриваемого нейрона является ряд примыкающих друг к другу каналов слухового нерва. Для указания этих каналов можно воспользоваться одномерной шкалой  $z$  и описать весовые коэффициенты возбуждающих и тормозных связей двумя функциями,  $\varphi_v(z)$  и  $\varphi_t(z)$  соответственно (рис. 9.1).

Для простоты принимается, что импульсация, поступающая на все возбуждающие входы, объединяется на одном сумматоре —  $C_v$ , а импульсация, поступающая на все тормозные входы, — на другом сумматоре —  $C_t$  (рис. 9.1). Распространенное представление относительно временной суммации состоит в том, что ее можно моделировать фильтрами нижних частот, расположенными на выходах сумматоров возбуждения и торможения —  $\Phi_v$  и  $\Phi_t$  (рис. 9.1). Часто принимается, что граничная частота фильтра торможения ниже таковой для фильтра возбуждения [43, 58].

Простейшее предположение относительно взаимодействия торможения и возбуждения состоит в том, что они просто алгебраически суммируются:  $P(t) = E(t) - kJ(t)$ , где  $E(t)$  — сигнал на выходе фильтра возбуждения,  $J(t)$  — сигнал на выходе фильтра торможения (вариант I на рис. 9.1). Используются и более сложные предположения. Например, принимается [68], что  $J(t)$  управляет параметрами фильтра возбуждения — возрастание  $J(t)$  приводит к уменьшению коэффициента передачи фильтра и расширению полосы частот (вариант II на рис. 9.1). Следует подчеркнуть, что основной интерес представляет, как правило, не модель единичного нейрона, но модель нейронного канала, описывающая поведение группы идентичных или близких по характеристикам нейронов. В этом случае логично принять, что входными сигналами, поступающими на каждый возбуждающий и тормозной вход элемента, являются не последовательности импульсов, но аналоговые переменные, отражающие плотность импульсации в соответствующих источниках.

Для описания плотности импульсации на выходе модели нейронного канала Родионов использовал следующий подход (рис. 9.1). Рассматривалась совокупность одинаковых независимых между собой пороговых элементов — II (рис. 9.1), на вход которых подается один и тот же сигнал  $P(t)$ , т. е. все элементы имеют на входе один и тот же механизм пространственной и временной суммации. Элементы обладают рефрактерностью: после «срабатывания» элемента его порог резко повышается, а затем с определенной скоростью снижается до стационарного уровня. Кроме того, на входе каждого элемента действует свой источник аддитивного шума — III (рис. 9.1). Выходом системы является величина, пропорциональная количеству сработавших элементов — назовем ее  $g^*(t)$ .

Для такой системы характерны, в частности, нелинейная зависимость установившейся реакции от уровня стационарного сигнала, колебательный переходный процесс, мощная кратковременная реакция на положительный скачок во входном сигнале и некоторая частотная избирательность, определяющаяся в основном параметрами рефрактерности и зависящая от уровня сигнала.

Можно надеяться, что принцип взаимодействия возбуждения и торможения внутри клетки достаточно универсален, т. е. функциональная зависимость  $P(t)$  от  $E(t)$  и  $J(t)$  может быть описана с помощью одного и того же алгоритма для всех нейронов слуховой системы. Одинаковой для разных нейронных каналов может оказаться также и зависимость между  $P(t)$  и плотностью импульсации на выходе канала  $g^*(t)$ .

В таком случае главное, что определяет специфику нейронного канала, это пространственные весовые функции возбуждающих и тормозных связей ( $\varphi_v$  и  $\varphi_r$ ) и функции передачи фильтров возбуждения и торможения.

Рассмотрим, нельзя ли составить каких-либо качественных предположений о возможных классах признаков  $g(z, t)$ , выделяемых нейронными каналами, основываясь на этих параметрах модели. При этом естественно начать с крайних случаев.

Допустим, что каждый нейронный канал высшего уровня получает на вход информацию только из ограниченного по ширине участка шкалы  $z$ . В этом случае признак, выделяемый каналом, можно определить как локальный по спектру. Если, наоборот, используется информация со всей шкалы  $z$ , можно говорить о признаках звука (спектра) в целом. Допустим теперь, что тормозных входов вообще нет —  $\varphi_t(z) = 0$ . Если при этом  $\varphi_v(z)$  равна нулю на всем интервале  $z$ , кроме одного частотного канала периферической слуховой системы, а фильтр возбуждения является фильтром нижних частот с высокой граничной частотой, то никакого преобразования информации практически не произойдет, передаваемым признаком останутся мгновенные значения  $g_i(t)$ . При тех же условиях в отношении  $\varphi_t(z)$  и  $\varphi_v(z)$ , но низкой граничной частоте фильтра возбуждения признаком станут усредненные по времени значения  $g_i(t)$ , т. е. нечто близкое к огибающей сигнала в  $i$ -той частотной полосе. Расширение  $\varphi_v(z)$  приведет к усреднению изображения по ряду частотных каналов.

Для всех моделей без тормозных входов и с  $\varphi_v(z)$ , характеризующейся одним максимумом, мгновенное значение плотности импульсации на выходе одного нейронного канала будет отражать усредненные по  $\Delta z$  и  $\Delta t$  значения  $g(z, t)$  в определенной точке  $z$  в данный момент времени.

Так как усредненное по времени значение  $g_i(t)$  отражает текущее значение энергии сигнала в  $i$ -той частотной полосе, мы приходим к тому, что упорядоченный по шкале  $z$  ряд центральных нейронных каналов без тормозных входов должен давать изображение, соответствующее мгновенному спектру стимула, полученному с тем или иным усреднением по шкале частот и по времени.

Если нейрон имеет и возбуждающие, и тормозные входы, признаки, которые он может выделять в  $g(z, t)$ , чрезвычайно многообразны. Поэтому рассматривать вопрос о признаках, не накладывая определенных ограничений на  $\varphi_v(z)$  и  $\varphi_t(z)$ , вообще не имеет смысла. Одно из важнейших ограничений состоит в том, что  $\varphi_v(z)$  и  $\varphi_t(z)$  заданы (имеют отличные от нуля значения) на достаточно близких или перекрывающихся интервалах значений  $z$ . В этом случае можно говорить о том, что нейронный канал выделяет локальную особенность или неравномерность на  $g(z)$  и соответственно на спектре стимула. Характер этой особенности будет, естественно, зависеть от вида  $\varphi_v(z)$  и  $\varphi_t(z)$  и их взаимного расположения по шкале  $z$ . Если обе эти весовые функции имеют один максимум, находящийся в одной и той же точке  $z$ , причем ширина  $\varphi_t(z)$  является большей, то мы имеем дело с моделью



«латерального торможения». Выделяемая ею неравномерность соответствует максимуму или краю на  $g(z)$ .

Легко заметить, что варьируя параметры фильтров возбуждения и торможения, можно получить самые разнообразные характеристики результирующего фильтра, соответствующего нейронному каналу. В частности, результирующий фильтр может оказаться полосовым, и признаком, выделяемым каналом, будет изменение текущего значения энергии (огнивающей) сигнала в соответствующей частотной полосе. Неравномерности  $g(z, t)$  во времени мы будем дальше называть событиями, в отличие от неравномерностей по  $z$ , названных особенностями.

Если нижние уровни центральной слуховой системы обеспечивают выделение особенностей и событий, можно предположить, что более высокие уровни работают уже как обнаружители определенных комбинаций особенностей и определенных последовательностей событий, т. е. осуществляют некоторую предварительную классификацию сигналов.

Итак, можно сформулировать следующие предположения относительно природы выделяемых признаков:

1) признаки являются локальными по спектру или относятся к спектру в целом;

2) признаки соответствуют текущим (мгновенным или усредненным) значениям  $g(z, t)$  или отражают только неравномерности  $g(z, t)$  по  $z$  и по  $t$ .

## 9.2. ТОНОТОПИЧЕСКАЯ ОРГАНИЗАЦИЯ СЛУХОВОЙ СИСТЕМЫ. ДАННЫЕ О ЛОКАЛЬНОМ ПО СПЕКТРУ ХАРАКТЕРЕ ПРИЗНАКОВ

Под центральной слуховой системой обычно понимается классический восходящий слуховой путь. Он схематически рассматривается как состоящий из нескольких последовательных уровней. Нейроны данного уровня получают информацию преимущественно от нейронов предыдущего уровня и посылают информацию преимущественно к нейронам ближайшего следующего уровня.

Волокна первичных нейронов — нейронов первого порядка — образуют слуховой нерв. Соответственно начальным (нижним) уровнем центральной слуховой системы являются скопления нейронов второго порядка — кохлеарные ядра ( $B1$  и  $B2$  на схеме рис. 9.2). Нейроны кохлеарных ядер получают импульсацию от слухового нерва своей стороны. Основная часть волокон, отходящих от нейронов кохлеарных ядер, переходит на противоположную сторону мозга.

В ядрах верхней оливы (рис. 9.2), получающих импульсацию от кохлеарных ядер обеих сторон,  $B1-B5$ , расположены преимущественно нейроны третьего порядка, в нижних (задних) буграх четверохолмия,  $G$ , — нейроны четвертого порядка, во внутреннем

коленчатом теле, *Д*, — нейроны пятого порядка, в первичной слуховой коре, *Е*, — нейроны шестого порядка.

Следует иметь в виду, что представление о том, что каждый данный уровень получает информацию только с выхода предыдущего уровня, является известным упрощением реальной картины. Имеются как связи через один и более уровней, так и связи внутри ядер одного и того же уровня. Кроме того, имеется система

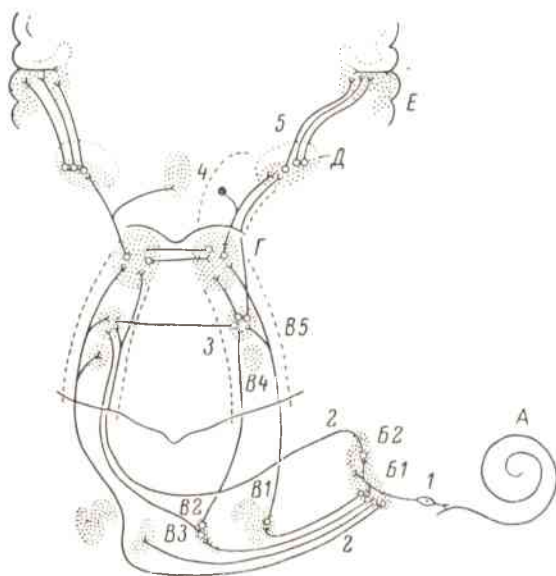


Рис. 9.2. Схема восходящего слухового пути. По [269].

*А* — улитка; *Б1* — вентральное кохлеарное ядро; *Б2* — дорсальное кохлеарное ядро; *Б1—Б5* — ядра, относимые к оливарному комплексу; *Г* — задний бугор четверохолмия; *Д* — внутреннее коленчатое тело; *Е* — первичная слуховая кора. 1—5 указывают порядковые номера нейронов, аксоны которых образуют основную часть данного участка слухового пути.

нисходящих слуховых путей (достаточно подробное и вместе с тем доступное неспециалисту описание слуховых путей дано Ратниковой [117]).

На всех уровнях слуховой системы обнаружена более или менее строгая тонотопическая организация. Под тонотопической организацией понимается наличие монотонной зависимости между пространственными координатами нейрона (местоположением нейрона в ядре) и частотой тона, к которому этот нейрон наиболее чувствителен.

Отчетливая тонотопическая организация была обнаружена в кохлеарных ядрах [249, 453], в ядре верхней оливы [508], в задних буграх четверохолмия [454], во внутреннем коленчатом теле [534] (цит. по [278]), в первичной слуховой коре [383].

С точки зрения интересующего нас вопроса о природе признаков, выделяемых на разных уровнях слуховой системы, важен не столько сам факт наличия тонотопической организации, сколько тот промежуточный результат, который позволяет ее обнаружить.

В эксперименте по определению тонотопической организации для каждого нейрона или снимается частотно-пороговая кривая, или определяется при постоянной интенсивности тестирующего тона полоса частот, внутри которой нейрон отвечает на тональный стимул. В качестве оптимальной частоты, приписываемой нейрону, используется или частота, соответствующая минимуму на частотно-пороговой кривой, или центральная частота найденной полосы частот.

Необходимым условием, без выполнения которого определение тонотопической организации было бы невозможным, является наличие у нейрона выраженной частотной избирательности: частотно-пороговая кривая должна иметь один выраженный минимум, полоса частот, на которую реагирует нейрон, должна быть существенно уже всего слухового диапазона.

Рассмотрение литературы показывает, что электрофизиологи слуха считают выполнение этого условия правилом, т. е. наиболее часто встречающимся случаем. Описываются как исключение те случаи, когда оно не выполняется. Хотя количественных оценок, характеризующих частоту отклонений от правила для разных уровней слуховой системы, найти не удалось, создается впечатление, что «неправильные» формы частотно-пороговых кривых наблюдаются преимущественно на уровне коленчатого тела и слуховой коры. Ограничимся рассмотрением данных о нейронах слуховой коры, т. е. наиболее сложного случая.

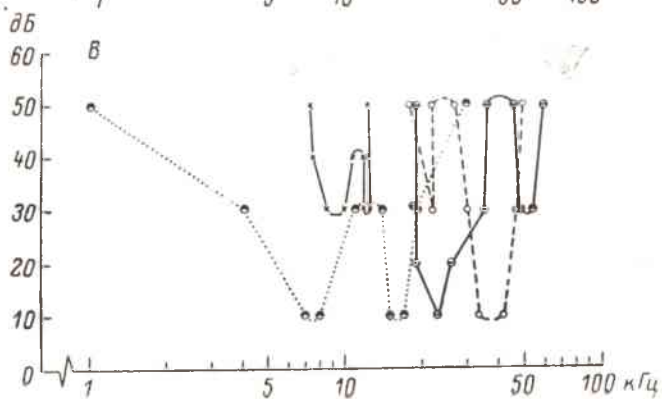
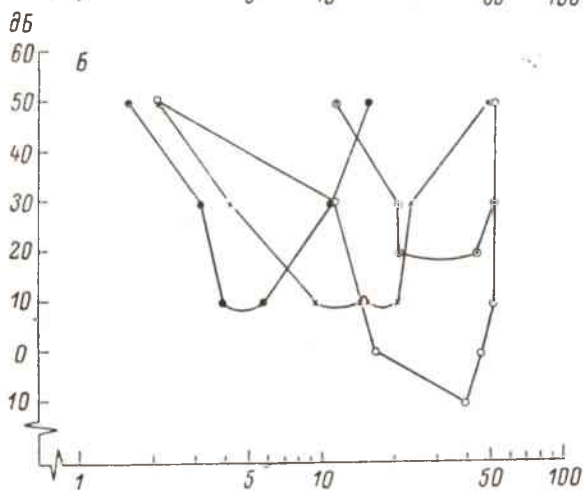
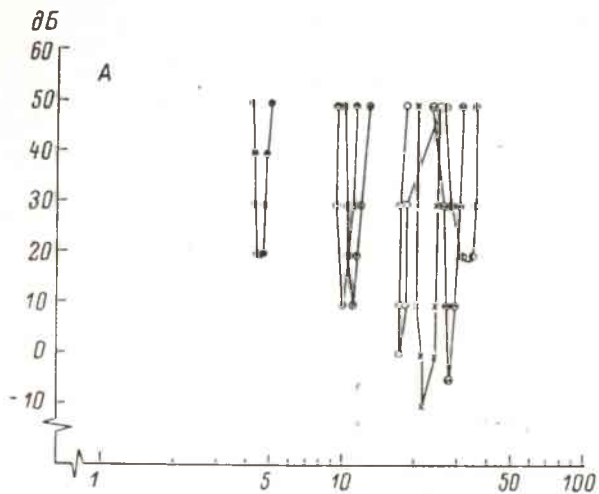
Необходимо прежде всего указать, что не у всех нейронов слуховой коры удается наблюдать реакцию на звуковой сигнал. Из нейронов, отвечающих на звук, только примерно 80—85% отвечают на тональные стимулы [247].

Три типа форм частотно-пороговых кривых, наблюдавшихся при тестировании 70 нейронов первичной слуховой коры тональными послылками длительностью 100 мс [174], приведены на рис. 9.3. Можно видеть, что часть нейронов (рис. 9.3, А) характеризуется очень узкими и близкими к полосовым частотно-пороговыми кривыми. Другие нейроны характеризуются или малой частотной избирательностью (рис. 9.3, Б), или наличием нескольких (обычно двух) минимумов (рис. 9.3, В).

---

Рис. 9.3. Типичные формы (А—В) частотно-пороговых кривых нейронов первичной слуховой коры. По [174].

По оси абсцисс — частота тональной послылки; по оси ординат — уровень звукового давления.



На рис. 9.4 приведена гистограмма распределения 236 нейронов по ширине полосы частот, на которые они отвечают [174]. Тестирование производилось при стандартном уровне звукового давления тона (50 дБ), полоса определялась как отношение  $F_2/F_1$ , где  $F_1$  — самая низкая, а  $F_2$  — самая высокая частота тестирующего тона, еще вызывающая ответ нейрона.

Из рис. 9.4 видно, что в большинстве случаев ширина полосы меньше октавы. Более 50% нейронов (нейроны типа А на рис. 9.3) характеризуются полосой, меньшей половины октавы.

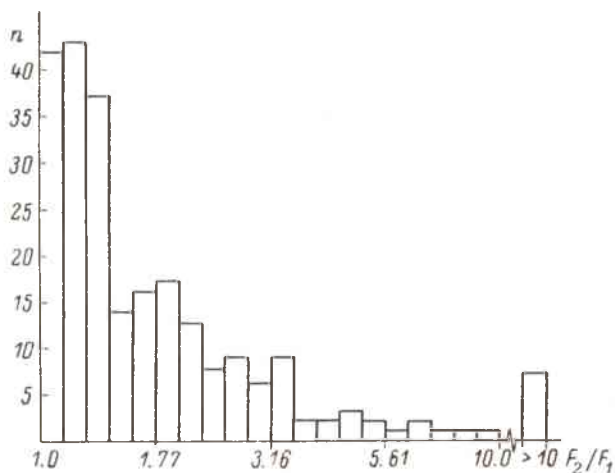


Рис. 9.4. Гистограмма распределения нейронов первичной слуховой коры по ширине полосы частот, вызывающих ответ нейрона. По [174].

По оси абсцисс — ширина полосы ( $F_2/F_1$ ); по оси ординат — число нейронов с данной шириной полосы.

Область частот, вызывающих положительный ответ (увеличение частоты разрядов) нейрона, называется зоной возбуждения. В случае модели латерального торможения зона возбуждения соответствует той области  $z$ , где значения  $\varphi_v(z)$  больше значений  $\varphi_r(z)$ , т. е. она может быть уже области возбуждающих связей (области задания  $\varphi_v(z)$ ). Для модели латерального торможения по бокам зоны возбуждения находятся две зоны торможения, соответствующие тем областям  $z$ , где значения  $\varphi_v(z)$  меньше значений  $\varphi_r(z)$ .

Для исследования зон торможения обычно применяется следующий метод. Тональная посылка оптимальной частоты предъявляется одновременно с тоном «боковой» частоты. Определяется зависимость между частотой и уровнем интенсивности тона боковой частоты, соответствующая некоторому критерию подавления реакции нейрона на тон оптимальной частоты.

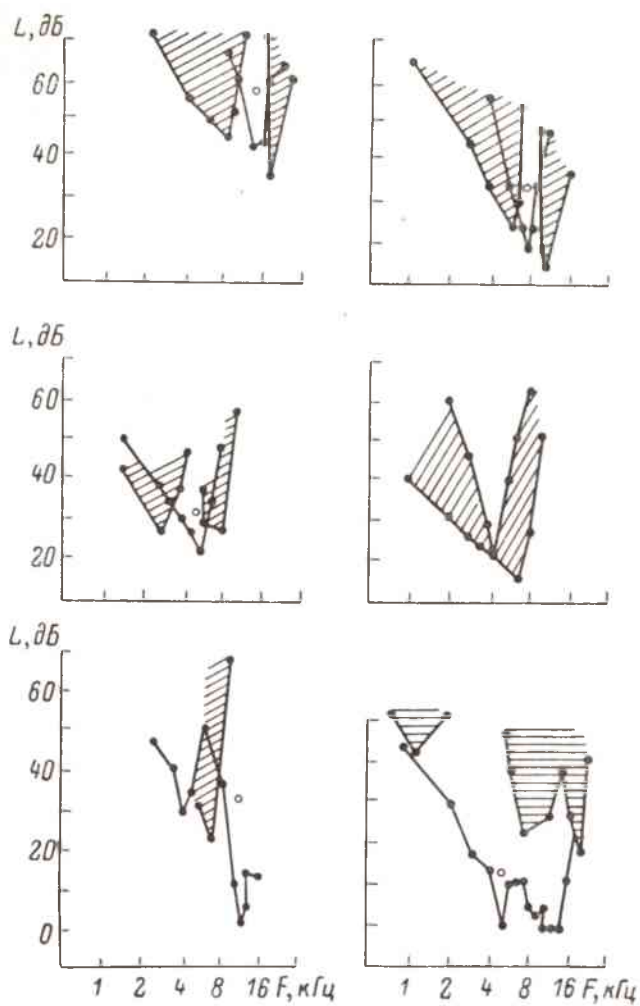


Рис. 9.5. Зоны возбуждения — частотно-пороговые кривые и зоны торможения (заштрихованные области) нейронов задних бугров четверохолмия. По [25].

По оси абсцисс — частота тона; по оси ординат — уровень интенсивности. Крестиком показано значение частоты и интенсивности тестирующего возбуждающего тона, применявшегося при исследовании зон торможения.



На рис. 9.5 приведены примеры зон возбуждения (частотно-пороговые кривые) и тормозных зон для нескольких нейронов задних бугров четверохолмия [25]. Тормозные зоны наблюдались более чем у 90% исследованных нейронов. Данные для четырех из шести нейронов (рис. 9.5) совместимы с предположением, что весовая функция  $\varphi_r(z)$  имеет один максимум, несколько смещенный относительно максимума на  $\varphi_v(z)$ . В работе приводятся указания на то, что, как правило, смещения не очень велики (не более чем на октаву) и что высокочастотный склон аппроксимирующей кривой значительно более крутой, чем низкочастотный. Отчетливо показано, что, хотя область тормозных входов достаточно широка, она несомненно ограничена по шкале  $z$ .

Таким образом, наиболее вероятным кажется предположение, что нейронные каналы всех отделов классического слухового пути обрабатывают информацию на ограниченном участке шкалы  $z$ , т. е. выделяют локальные по спектру признаки акустического процесса, а не признаки звука (спектра) в целом.

### 9.3. ПРИРОДА ПРИЗНАКОВ: ТЕКУЩИЕ ЗНАЧЕНИЯ АМПЛИТУДЫ (ЭНЕРГИИ) В ПОЛОСЕ ИЛИ ОСОБЕННОСТИ И СОБЫТИЯ

За последние годы получены данные [249, 250], показывающие, что уже в кохлеарных ядрах выделяются два типа нейронных каналов, существенно отличающихся по своей структуре и параметрам. Анатомически они соответствуют дорсальному и вентральному кохлеарным ядрам. По своей морфологической организации эти ядра сильно различаются [116, 117].

Постстимульная гистограмма нейронов вентральных ядер (рис. 9.6, В) близка к таковой для волокна слухового нерва, нейрон возбуждается узкой полосой частот стимула (рис. 9.6, А), торможения спонтанной импульсации тонами «боковых» частот не наблюдается (рис. 9.6, Б). Имеются данные [455], что импульсация синхронизирована с фазой тонального колебания вплоть до 4—5 кГц.

Приходится думать, что в вентральных ядрах вообще не производится существенных преобразований информации и их роль, по крайней мере частично, сводится к неискаженной передаче формы волны к ядрам верхней оливы, где сравниваются сигналы, поступающие из двух ушей.

Принципиально иная картина характерна для дорсального ядра. Нейроны этого ядра обладают собственной спонтанной активностью, она сохраняется при разрушении улитки [331]. При действии стимула наблюдается или увеличение импульсации, или ее подавление. На рис. 9.7 приведен пример нейрона, у которого возбуждающий эффект ограничен очень узкой частотной областью, боковые частоты вызывают торможение спонтанной активности.

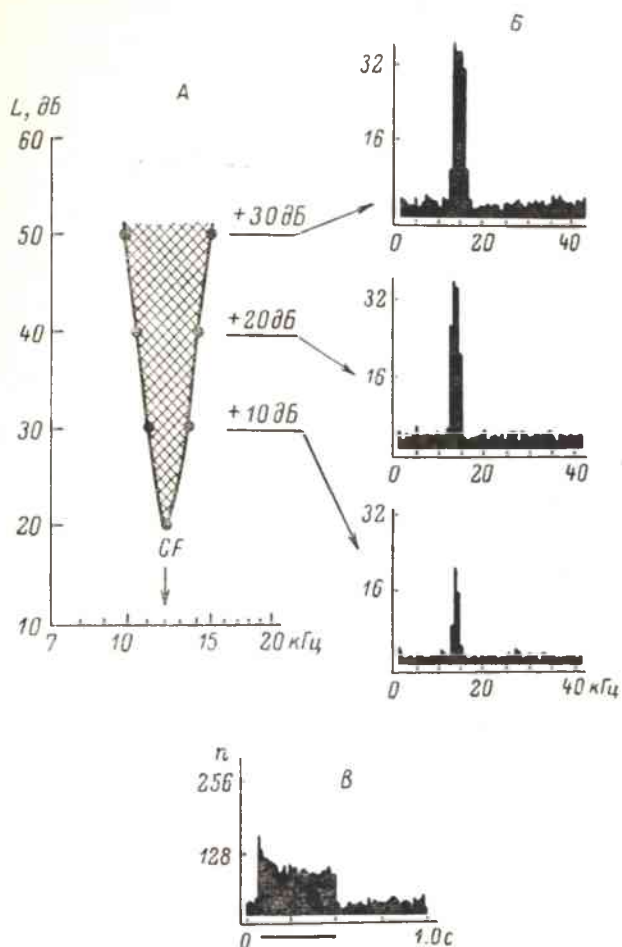


Рис. 9.6. Характеристики нейрона вентрального кохлеарного ядра. По [249].

А — частотно-пороговая кривая. По оси абсцисс — частота тональной послышки; по оси ординат — уровень интенсивности тона.  $CF$  — характеристическая частота нейрона. Б — зависимость числа импульсов (по оси ординат), возникших во время тональной послышки, от частоты тестирующего тона (по оси абсцисс). Уровни интенсивности тестирующего тона указаны над стрелками. В — постстимульная гистограмма, полученная при действии тональной послышки с частотой, равной характеристической частоте нейрона ( $CF$  на А) при уровне интенсивности, равном 30 дБ над порогом ответа нейрона. Период анализа 1 с;  $n$  — число импульсов за  $1/64$  периода анализа, полученное при 27 повторных предъявлениях тональной послышки. Чертой внизу показан интервал действия послышки.

На рис. 9.8 показан случай, когда во время действия тона (в узкой полосе частот) наблюдается торможение, после прекращения тона происходит резкое усиление импульсации.

Тормозной эффект является правилом для нейронов дорсальных ядер; область частот, вызывающих торможение, обычно значительно шире области возбуждающих частот [249]. Источником тормозящих сигналов является вентральное ядро [250]. На пост-стимульных гистограммах ответов нейронов дорсального ядра наблюдается выраженный переходный процесс в момент включения (или также и выключения) стимула [118, 249, 320, 420].

Основная информация о характеристиках нейронов кохлеарных ядер как каналов преобразования и передачи сигнала получена Мёллером. К сожалению, Мёллер не производил гистологического контроля положения электродов, однако, судя по наблюдавшимся им эффектам, данные его относятся к дорсальному ядру.

Как говорилось выше, для описания временных свойств отклика нейронного канала его можно рассматривать как фильтр. Если система является линейной (или ее допустимо считать линейной в каком-то диапазоне), то ее исчерпывающим описанием будет передаточная функция, т. е. ее амплитудно-частотная и фазово-частотная характеристики.

Один из методов, использованных Мёллером для определения функции передачи нейронов, состоял в следующем. Предъявлялись амплитудно-модулированные стимулы (тоны или шумы) с постоянной глубиной и переменной частотой ( $f_m$ ) модуляции. Снимались цикло-гистограммы ответов нейронов, показывающие изменение плотности импульсации на протяжении одного периода модуляции. Для аппроксимации огибающей плотности импульсации на цикло-гистограмме использовался один период синусоиды с частотой, равной частоте модуляции. Определялась относительная амплитуда этого аппроксимирующего колебания (глубина модуляции плотности импульсации) и его сдвиг по фазе относительно огибающей стимула ( $\varphi$ ). Коэффициент передачи ( $K$ ) нейрона на данной частоте определялся как отношение (в дБ) глубины модуляции плотности импульсации к глубине модуляции стимула.

Функции передачи группы нейронов, полученные при уровне стимула, равном 20 дБ над порогом нейрона, приведены на рис. 9.9. Можно видеть, что для всех нейронов существует некоторая область частот, где коэффициент передачи больше единицы. Значение частоты, соответствующее максимуму коэффициента передачи, колеблется в пределах от 50 до 200 Гц от нейрона к нейрону [389]. В области более высоких частот коэффициент передачи резко уменьшается с частотой, фазовый сдвиг (рис. 9.9, Б) растет. Подавление низких частот модуляции, как оказалось [394], резко возрастает с увеличением интенсивности стимула.

Таким образом, при малых интенсивностях стимула нейронный канал данного уровня слуховой системы действует как фильтр

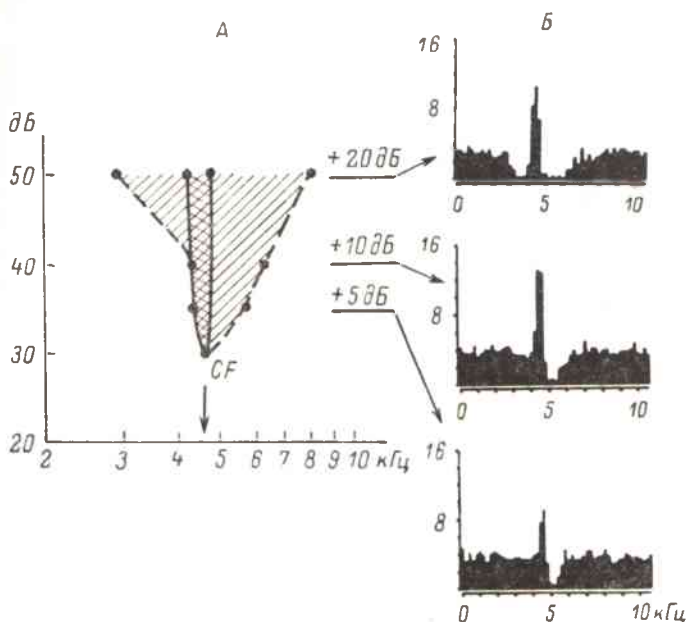


Рис. 9.7. Характеристики нейрона дорсального кохлеарного ядра. По [249].

*А* — частотно-пороговая кривая (двойная штриховка по диагонали); по оси абсцисс — частота тональной посылки, по оси ординат — уровень интенсивности. *CF* — характеристическая частота нейрона. *Б* — зависимость числа импульсов (по оси ординат), возникших во время действия тональной посылки, от частоты (по оси абсцисс) тестирующего тона. Уровни интенсивности тестирующего тона указаны над стрелками.

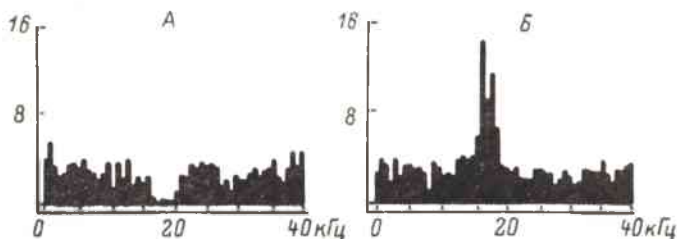


Рис. 9.8. Импульсация в нейроне дорсального кохлеарного ядра при действии тона (*А*) и после выключения тона (*Б*). По [249].

По оси абсцисс — частота тона; по оси ординат — число импульсов за 100 мс от начала тона (*А*) и от конца тона (*Б*).

нижних частот, при больших уровнях он приближается к полосовому фильтру.

Существенно, что в области частот около максимума коэффициент передачи сохраняется примерно постоянным в широком диапазоне уровней интенсивности стимула [394]. При этом среднее значение частоты импульсации может уже не зависеть от уровня интенсивности стимула, т. е., не передавая информации об абсолютных значениях амплитуды сигнала, нейрон передает информацию об ее изменениях.

Следует, естественно, иметь в виду, что характеристики «фильтра», приведенные на рис. 9.9, являются сквозными характеристиками системы, включающей не только нейрон кохлеарного ядра, но и все предшествующие ему более периферические элементы. Подавление низких частот модуляции может быть частично отнесено за счет механизма периферической адаптации, рассмотренного в главе 8. Подавление высоких частот модуляции, т. е. низкочастотная фильтрация  $g(t)$ , обусловлено, вероятно, временной суммацией в самих кохлеарных ядрах. Заметим, что граничная частота «фильтра» еще достаточно высока —  $100 \div 200$  Гц. Это означает, что при действии речевого сигнала во временной картине плотности импульсации на выходе кохлеарных ядер,  $g^*(t)$ , должна сохраняться информация об основной частоте (периоде) голоса.

Зависимости, изображенные на рис. 9.9, были получены при использовании несущей частоты, равной характеристической частоте нейрона. Применение в качестве несущей боковой частоты (посылка тона боковой частоты вызывает торможение импульсации) позволяет получить аналогичные характеристики. Отличие заключается только в сдвиге фазы на  $180^\circ$  — максимум плотности импульсации на цикло-гистограмме при несущей, равной характеристической частоте, соответствует минимуму при несущей, равной боковой частоте.

Если нейрон обладает спонтанной активностью, а схема его связей с первичными нейронами соответствует модели латерального торможения (см. раздел 9.1), наиболее эффективным для этого нейрона стимулом может быть сигнал, частота которого изменяется во времени, т. е. частотно-модулированный звук.

В цикле работ, посвященных исследованию реакций нейронов кохлеарных ядер на частотно-модулированные стимулы [388, 389, 393], использовалась модуляция разной формы, в качестве несущей служили как тоны, так и полосовые шумы. Исходным анализируемым материалом являлись цикло-гистограммы ответов одиночных нейронов, показывающие изменение плотности импульсации на протяжении периода модуляции. Пример двух таких цикло-гистограмм, соответствующих двум разным частотам трапецевидной модуляции, приведен на рис. 9.10. Из рисунка видно, что нейрон отвечает как на участке повышения, так и на участке понижения частоты. Хотя некоторая асимметрия ответов и имеет

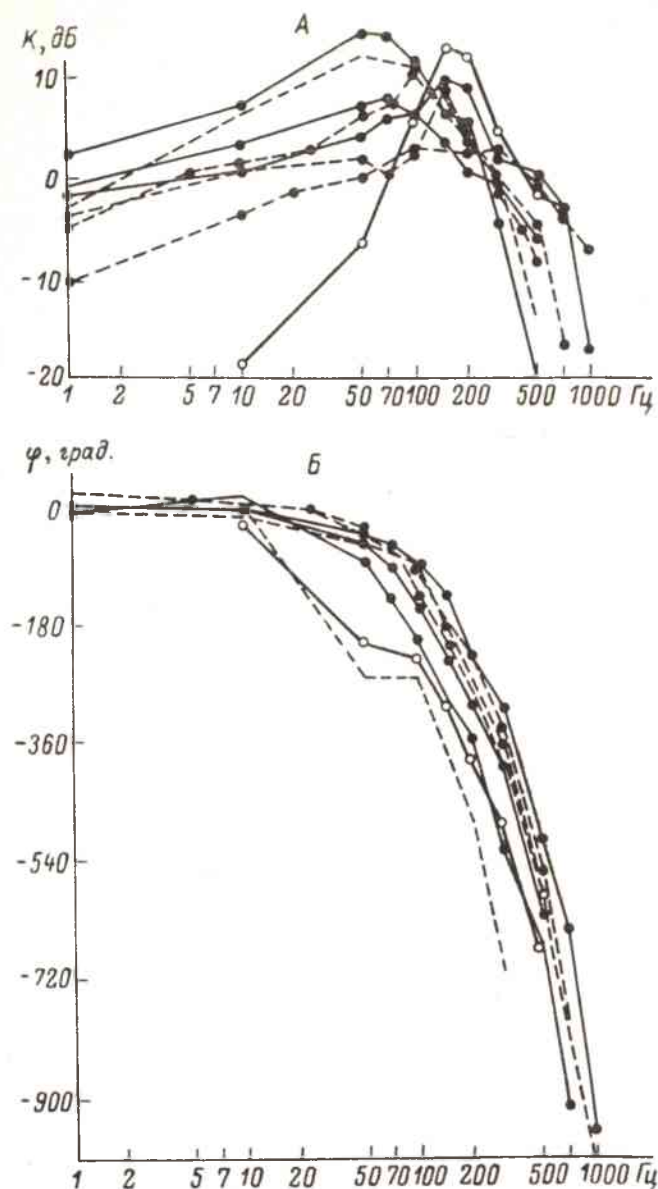


Рис. 9.9. Функции передачи огибающей нейронами кохлеарных ядер. По [391].

А — амплитудно-частотная характеристика; Б — фазово-частотная характеристика.  
Остальные обозначения см. в тексте.



место, она выражена не очень значительно и по-разному для разных нейронов.

Поразительно, что высота пика плотности импульсации (высота пика на гистограмме) возрастает с увеличением скорости изменения частоты стимула, причем это наблюдается вплоть до очень больших значений скорости (от 0.3 до 25 МГц/с у разных нейронов).

Если считать, что условием ответа нейрона является только попадание частоты стимула внутрь некоторой полосы  $W$  (с центральной частотой, соответствующей характеристической частоте нейрона), то полученные

данные пришлось бы истолковывать как сужение полосы  $W$ , т. е. как увеличение частотной избирательности нейрона с увеличением скорости изменения частоты стимула.

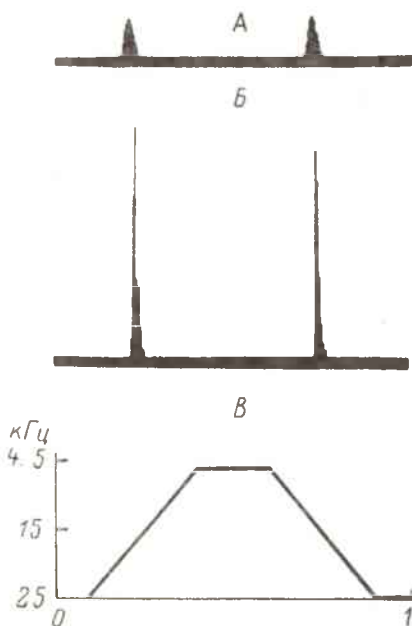


Рис. 9.10. Цикло-гистограммы ответов нейрона кохлеарного ядра с характеристической частотой 15 кГц при действии частотно-модулированного тона. По [389].

А — при скорости изменения частоты тона 7.3 кГц/с (частота модуляции — 0.1 Гц). Б — при скорости изменения частоты тона 470 кГц/с (частота модуляции — 6.4 Гц). В — форма изменения частоты тона на протяжении одного периода модуляции. По оси абсцисс — «время» (в долях периода модуляции) внутри одного периода модуляции; по оси ординат — частота тона.

Данные Мёллера качественно совместимы с моделью латерального торможения. В такой модели ЧМ-стимул будет вызывать  $P(t)$ , состоящий из трех фаз: отрицательной, положительной и снова отрицательной. При преобразовании  $P(t)$  в  $g^*(t)$  положительное колебание должно быть дополнительно подчеркнуто. Количественное моделирование явления связано с огромными трудностями, так как эффект должен зависеть и от параметров периферического слухового анализатора, т. е. модель должна включать периферическую слуховую систему.

Описать признаки сигнала, выделенные нейронными каналами дорсального ядра, естественно, нельзя до тех пор, пока модель не будет разработана. Можно только утверждать, что в изображении сигнала на выходе системы,  $g^*(z^*, t)$ , должны быть резко подчеркнуты участки, соответствующие быстрым изменениям огибающей и перемещениям спектральных максимумов по оси частот.

Если фильтр огибающей не пропускает нижних частот модуляции, то отклик на тональную посылку будет состоять из одного всплеска импульсации, соответствующего началу стимула.

Нейроны, отвечающие одним (или двумя-тремя) импульсом в момент включения стимула, наблюдаются на всех более высоких уровнях слуховой системы. Они получили название фазических нейронов, в отличие от тонических, отвечающих на всем протяжении действия стимула. Ясно, что фазические нейроны могут передавать информацию только об изменениях стимула — они не передают постоянной составляющей.

Для фазических нейронов заднего бугра четверохолмия было экспериментально показано, что они вообще не отвечают при амплитудной модуляции сигнала, если частота модуляции ниже 10—12 Гц [24, 27, 42]. Для нейронов этого же уровня обнаружено, что они обладают порогом по скорости изменения огибающей [495]. Автор применял стимулы с линейно нарастающей амплитудой на начальном участке и определял порог как функцию времени нарастания ( $T$ ) и конечной амплитуды ( $P$ ). Оказалось, что для  $T > 4\div 8$  мс приблизительно выполняется условие  $P/T = \text{const}$ ; отклонения от этой зависимости таковы, каких следовало бы ожидать, если бы между амплитудой стимула и ее эквивалентном в слуховой системе наблюдалась нелинейная зависимость.

В отношении тонических нейронов ситуация более сложна. Видя, что нейрон тонически отвечает на тональную посылку, можно в равной мере предполагать, или что он передает мгновенные или усредненные значения  $g_i(t)$ , или что он передает информацию о наличии (степени выраженности) спектрального максимума в полосе частот, соответствующей  $i$ -тому каналу периферической слуховой системы.

Выраженный эффект бокового торможения в нейронах заднего бугра четверохолмия (см. раздел 9.2) указывает на то, что второе предположение более вероятно, т. е. что тонические нейроны этого уровня передают информацию об особенностях спектра стимула, а не о значениях амплитуды (энергии) сигнала в частотных полосах. Если это действительно так, то информация о громкости стимула должна передаваться в управляющие и запоминающие отделы мозга по какому-то другому, а не классическому слуховому пути.

Итак, кажется достаточно правдоподобным, что изображение стимула на высших уровнях слуховой системы соответствует особенностям и событиям.

Имеющиеся данные указывают на то, что по крайней мере для четырех классов событий можно ожидать самостоятельных каналов, обеспечивающих обнаружение этих событий и передачу сведений об их выраженности. К ним относятся, во-первых, изменения огибающей разного знака (положительные и отрицательные приращения) и, во-вторых, сдвиги спектрального максимума в разном направлении по оси  $z$ .

В слуховой коре надежно установлены нейроны, избирательно реагирующие на выключение стимула, т. е. на отрицательное приращение огибающей. Они обозначаются как *off*-нейроны [175, 279], реакция нейрона (постстимульная гистограмма) характеризуется резким четко локализованным во времени пиком. *On*-нейроны, реагирующие на включение, широко представлены уже на уровне четверохолмия.

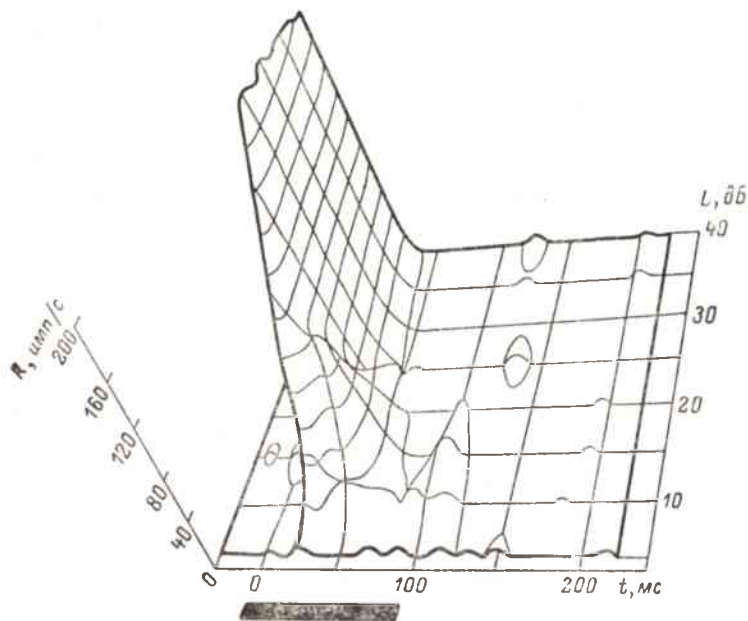


Рис. 9.11. Зависимость мгновенной частоты импульсации  $R$  (имп/с) от времени относительно начала стимула ( $t$ ) и от уровня интенсивности тона ( $L$ ). По [175].

Жирной чертой внизу отмечен стимул.

Избирательность нейронов к знаку частотной модуляции, т. е. направлению сдвига спектрального максимума, была обнаружена сначала для нейронов слуховой коры [192, 251]. В дальнейшем было показано, что она отчетливо выражена уже на уровне четверохолмия [26, 406, 494]. По данным Вартанян, 66% из 231 тестированных нейронов отвечают преимущественно или исключительно при изменении частоты в одном направлении (42% при повышении, 24% при понижении).

Существенно отметить, что если на уровне четверохолмия еще отчетливо представлены нейроны с тоническим разрядом, то на уровне слуховой коры ответ нейрона на протяжении всей посылки наблюдается преимущественно лишь при близкой к порогу интенсивности стимула [175]. Увеличение интенсивности стимула

(соответственно величины начального скачка) приводит к тому, что нейрон отвечает по типу *on*-нейрона, т. е. отклик становится резко локализованным по времени (рис. 9.11). Создается впечатление, что признаками становятся только события. Понятно, что если кодом является номер нейрона, указывающий его связь со шкалой  $z$  и класс выделяемого признака, то частота тонального стимула будет однозначно определена номерами нейронов, давших *on*- и *off*-ответы.

Важнейший вопрос касается запоминания событий, так как *on*- и *off*-ответы, естественно, образуются во времени. К этому вопросу мы обратимся в следующем разделе.

На основании данных, приведенных в настоящем разделе, можно с достаточной уверенностью утверждать, что принципом переработки изображения сигнала, используемым слуховой системой, является выделение неравномерностей в спектре стимула по шкале частот и по времени.

#### 9.4. ЦЕНТРАЛЬНЫЕ СЛУХОВЫЕ НЕЙРОНЫ И «ДЕТЕКТОРЫ ОБРАЗОВ»

Одно из направлений современной электрофизиологии состоит в поисках таких центральных нейронов, которые реагировали бы только на определенные классы биологически важных сигналов. Некоторые электрофизиологи считают, что для звуков речи такие «детекторы» уже найдены, причем находятся на уровнях ниже коры. Так, Кейдель [317] называет нейроны, избирательно чувствительные к знаку смещения спектрального максимума «детекторами переходов», а нейроны, реагирующие на частотную модуляцию вне зависимости от знака, — «детекторами согласных».

Если мы имеем в своем распоряжении систему, которая обнаруживает несколько типов событий в каждой из  $N$  частотных областей, то переход от описания сигнала на выходе такой системы к образу звука (указанию класса звуков, которому он принадлежит) требует решения двух весьма сложных проблем. Одна из них может быть названа проблемой объединения информации по спектру — всей области значений  $z$ , вторая — проблемой сбора информации по времени.

Допустив, что каждому классу сигналов соответствует определенный нейрон (нейронный канал), возбуждение которого означает установление принадлежности сигнала к этому классу, естественно предположить [247, 317], что объединение информации производится с помощью нейронов, выполняющих роль логического элемента «И» (логическое умножение). Рассмотрим правдоподобность такого предположения применительно к объединению информации по спектру.

Допустим, что слуховая кора представляет собой совокупность элементов, каждый из которых отвечает тогда и только тогда,

когда в нескольких определенных точках шкалы  $z$  обнаружены максимумы. Так как число естественных сигналов с больше чем одним спектральным максимумом превышает число сигналов с одним максимумом на спектре, элементы, отвечающие на чистый тон, т. е. сигнал с одним спектральным максимумом, должны быть редкостью. Это совершенно не согласуется с экспериментальными данными — подавляющее большинство нейронов отвечает на чистые тоны.

Аналогичные рассуждения показывают неправдоподобность предположения о том, что нейроны слуховой коры являются «детекторами» определенных последовательностей событий.

Элемент, избирательно реагирующий на последовательность событий АВ, не может ответить раньше, чем произойдет событие В. Ответ нейрона, реагирующего на последовательность АВ (наличие события А при отсутствии за ним события В), обязательно должен быть задержан относительно начала стимула. По данным Серкова и Сторожука [126], подавляющее большинство (109 из 135) нейронов, реагирующих на щелчок, отвечают на интервале от 7 до 25 мс от начала стимула. Такие же значения скрытого периода наблюдаются при действии шумовых посылок, причем величина скрытого периода не зависит от длительности посылки [23].

Таким образом, представление о том, что на уровне первичной слуховой коры происходит объединение информации по спектру и значительному интервалу времени, вероятно, нужно отвергнуть. Более вероятное предположение состоит в том, что на верхних уровнях слухового пути производится не только выделение событий в частотных полосах, но и формирование сигналов об этих событиях, приведение их к такому виду, при котором они могут быть использованы дешифратором.

Обратим внимание на то, что понятие последовательности событий обязательно включает время. Для различения последовательностей АВ и АВ̄ необходимо указать интервал времени от начала А, в течение которого нужно ждать В. Если В не появилось в течение этого интервала, то принимается решение АВ̄. Если такой интервал задан, то появляется возможность отличать двойные одинаковые события АА от единичных событий, следующих друг за другом: АА̄ АĀ. Наглядный пример такой классификации осуществляет человек в эксперименте, когда ему предъявляются два щелчка, разделенных разным интервалом времени  $t$ . При  $t < 10$  мс он воспринимает один хриплый щелчок, при  $10 \text{ мс} < t < 60 \div 100$  мс он слышит двойной щелчок, при больших интервалах воспринимаются независимые следующие друг за другом щелчки. Для воспроизведения такой классификации технической системой нужно, кроме прямого канала, воспроизводящего оба импульса, применить два канала с разной длительностью запрета (рефрактерности) после первого щелчка и, например, поставить на выходе сумматор. По величине сигнала на



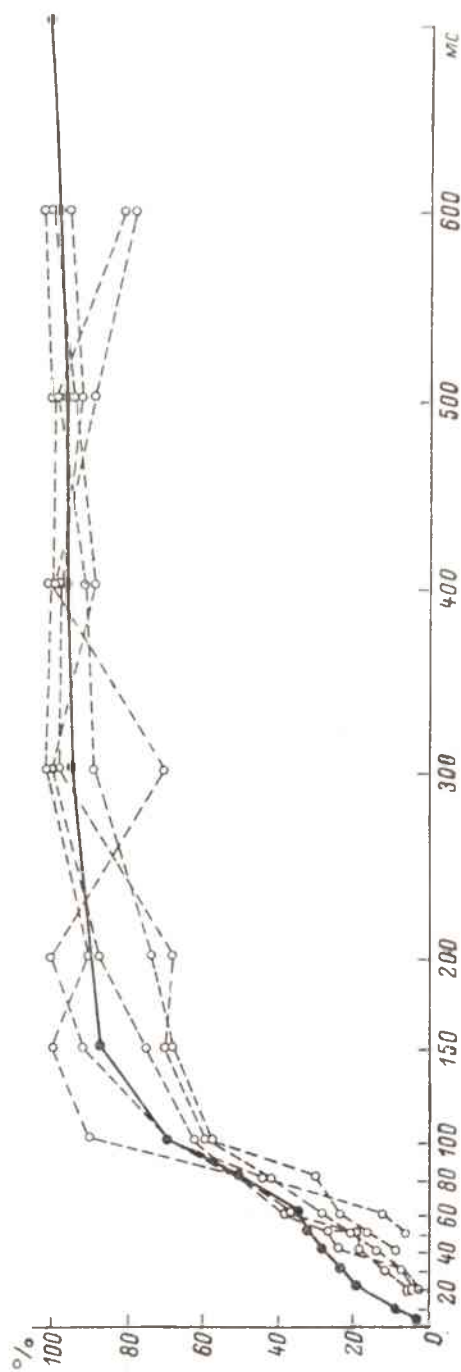


Рис. 9.12. Реакция нейронов слуховой коры на второй в паре щелчок. По [127].

По оси абсцисс — длительность интервала между щелчками (Т); по оси ординат — процент нейронов, отвечающих на второй щелчок при значениях Т, меньших указанных по оси абсцисс (сплошная кривая), и относительная величина амплитуды вызванного потенциала при действии второго в паре щелчка (штриховая кривая). Разные кривые — разная глубина погружения электрода.



сумматоре в момент второго щелчка мы сможем легко классифицировать эти три случая.

Таким образом, использование временных запретов после события является необходимым условием классификации последовательности событий.

В работе Серкова и Яновского [127] для нейронов первичной слуховой области был определен минимальный интервал времени между двумя щелчками, при котором нейрон отвечает на оба щелчка. В результате была получена зависимость между процентом нейронов, отвечающих на второй щелчок, и интервалом времени между щелчками. Она приведена сплошной кривой на

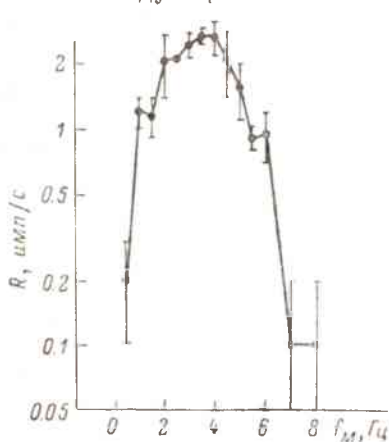


рис. 9.12. Так как ранее было показано [128], что начальная пачка импульсов на единичный щелчок хорошо совпадает по времени с начальным положительным отклонением так называемого суммарного коркового первичного ответа, были произведены также изме-

Рис. 9.13. Воспроизведение частоты модуляции корковым нейроном. По [247].

По оси абсцисс — частота синусоидальной частотной модуляции тона; по оси ординат — частота импульсации нейрона.

рения относительной амплитуды этого отклонения на второй щелчок в зависимости от интервала времени между щелчками. Отведения осуществлялись при разной глубине погружения электрода. Полученные зависимости приведены на рис. 9.12. Можно видеть, что сплошная кривая на этом рисунке свидетельствует о наличии двух распределений нейронов по рефрактерности (длительности запрета), суммарный ответ, очевидно, соответствует второму из них. Существенно заметить, что зависимость для суммарного ответа хорошо согласуется с зависимостью распознавания последовательности от интервала между элементами, приведенной в главе 6 (рис. 6.9).

На рис. 9.13 приведены данные по воспроизведению корковым нейроном частотно-модулированного тона при разных частотах синусоидальной модуляции [247]. Можно видеть, что ответ резко подавлен при частоте в 7 Гц и выше. Учитывая, что данный нейрон реагирует как на повышение, так и на понижение частоты, следует заключить, что он перестает воспроизводить обнаруживаемое им событие, если оно повторяется с частотой порядка 14 Гц, т. е. что рефрактерный период должен составлять примерно 70 мс. Это хорошо согласуется с данными рис. 9.12 и с упоминавшимися данными по восприятию последовательностей.

Причина этой рефрактерности может быть как в том, что сам нейрон получает тормозной сигнал, становится на некоторое время гиперполяризованным (этот эффект наблюдался при внутриклеточном отведении [<sup>126</sup>]), так и в том, что рефрактерность была на предыдущем уровне и корковый нейрон просто не получает второго сигнала. Иначе говоря, запрет может формироваться на разных уровнях обработки.

Имеющиеся данные говорят о том, что есть запреты, формируемые в специальных тормозных нейронах коры [<sup>128</sup>], но есть также и запреты, формируемые на уровне коленчатого тела [<sup>180</sup>].

Если есть два разных канала, обнаруживающих события А и В, и в каждом из них есть свой временной запрет, то следующий канал, использующий оба этих канала в качестве входов, может пропускать быстрые АВ- и ВА-последовательности, не пропуская АА- и ВВ-последовательности. Именно такая ситуация характерна для корковых нейронов с частотно-пороговыми характеристиками, имеющими несколько минимумов, т. е. отвечающих на несколько частотных полос [<sup>175</sup>]. Было обнаружено, что при действии двух тональных посылок ответ на вторую посылку подавлен в течение 100 и более миллисекунд от начала первой, если обе посылки принадлежат одной и той же частотной полосе  $W_1$  или  $W_2$ . Подавления ответа на вторую посылку нет, если одна из посылок принадлежит полосе  $W_1$ , а другая — полосе  $W_2$ .

Таким образом, можно думать, что если на начальных этапах обработки сигнала в центральной слуховой системе основной задачей является выделение и обнаружение пространственно-временных неравномерностей, высшие уровни представляют эти выделенные особенности как временные события и подготавливают возможность для опознавания последовательностей событий. Ведущую роль здесь начинают, очевидно, играть временные запреты.

#### 9.5. ЗНАЧЕНИЕ ДАННЫХ ЭЛЕКТРОФИЗИОЛОГИИ СЛУХА ДЛЯ ТЕОРИИ ВОСПРИЯТИЯ РЕЧИ

Приведенные выше данные говорят о том, что обработка речевого сигнала, осуществляемая слуховой системой, резко отличается по своим принципам от той обработки, которая применяется сейчас в технических системах автоматического распознавания речи. Соответственно трудно ожидать, что решающие правила, описанные в терминах тех признаков, которые сейчас используются в технике, будут простыми или достаточно универсальными.

Понятно, что задача нахождения правил распознавания, которыми пользуется человек, в принципе не решается до тех пор, пока признаки сигнала не определены.

К сожалению, данные электрофизиологии слуха позволяют только определить тот подход, по которому может пойти поиск моделей обработки и определение их параметров. Пока мы можем утверждать, что признаками являются спектрально-временные неравномерности, но совершенно не имеем возможности формально описать эти неравномерности или даже перечислить число их возможных классов. Очевидно, что единственный возможный выход состоит в том, чтобы попытаться извлечь необходимые сведения из другого источника, а именно из психоакустических данных по восприятию неравномерностей в сложном стимуле. Этому вопросу посвящены следующие главы настоящей книги.

## Глава 10

### ОБРАБОТКА «СЛУХОВОГО СПЕКТРА» СТИМУЛА

Под слуховым спектром стимула естественно понимать пространственный рисунок плотности импульсации в слуховом нерве —  $g(z)$ , где  $z$  обозначает характеристическую частоту нейрона (нейронного канала) и, следовательно, координату точки улитки, с которой связан нейрон. Из данных нейрофизиологии (см. главу 9) следует, что  $g(z)$  подвергается существенным преобразованиям в центральных отделах слуховой системы. Предполагается, что эти преобразования обеспечивают получение «удобного» сокращенного описания спектра стимула.

Под «удобным» понимается описание, по которому естественные сигналы, принадлежащие к разным классам, образуют компактные легко разделяемые множества. Кроме того, принимается, что признаки этого описания должны быть устойчивыми к разного рода естественным искажениям звука и приспособлены к «техническим» возможностям центральной нервной системы (удобство запоминания изображения, устойчивость его к собственным шумам системы и т. д.).

Логично думать, что при фонетической интерпретации речевых сигналов используется только часть из всего набора признаков, выделяемых центральной слуховой системой. Данные, приводимые в настоящей главе, касаются в основном еще более частного случая, а именно признаков  $g(z)$ , используемых при распознавании стационарных гласных.

#### 10.1. ГИПОТЕЗЫ О ПРИРОДЕ ПРИЗНАКОВ ГЛАСНЫХ

Обсуждая гипотезы о природе признаков гласных, исследователи часто основываются на известных технических методах, применяемых при автоматическом распознавании гласных. Если исходить из нейрофизиологии слуха, то более содержательным является разделение гипотез по источнику результирующей

информации — номерам (пространственным координатам) возбужденных сигналом нейронных каналов или величинам возбуждения на выходе нейронных каналов.

Первый класс гипотез мы будем называть гипотезами пространственного кодирования, второй класс — гипотезами кодирования величиной возбуждения.

#### 10.1.1. КОДИРОВАНИЕ ВЕЛИЧИНОЙ ВОЗБУЖДЕНИЯ

Принимается, что имеется некоторый небольшой набор суммирующих элементов, различающихся по виду весовых функций, т. е. по значениям коэффициентов возбуждающих и тормозных связей каждого элемента с нейронными каналами периферической слуховой системы (или какого-то другого уровня, где  $g(z)$  еще не подверглась существенным изменениям). В простейшем случае выходной сигнал  $j$ -того сумматора будет равен

$$u_j = \sum_{i=1}^n k_{ij} g_i,$$

где  $g_i$  — значение плотности импульсации в  $i$ -том канале периферической слуховой системы,  $n$  — число каналов периферической слуховой системы,  $k_{ij}$  — возбуждающий или тормозной коэффициент связи  $i$ -того канала с  $j$ -тым суммирующим элементом.

Описанием спектра гласного на выходе такой модели будет набор значений  $u_1, \dots, u_l$ , где  $l$  — число сумматоров. Такие представления лежат в основе многих работ по автоматическому распознаванию речи.

Мы коротко остановимся на группе работ, в которых этот подход был использован для описания восприятия гласных (обзор этих работ см. [425]). В работе [431] было показано, что субъективная разница по тембру между двумя стационарными сигналами довольно хорошо коррелирует с разностью их спектров, определенной как

$$D_{ij} = \sqrt{\sum_{n=1}^m (L_{in} - L_{jn})^2},$$

где  $L_{in}$  — уровень звукового давления одного сигнала в треть-октавной полосе под номером  $n$ ,  $L_{jn}$  — то же для второго сигнала,  $m$  — общее число полос.

Это послужило основанием для использования в качестве начального описания спектров естественных гласных (12 голландских гласных по 50 реализаций каждой) значений уровня звукового давления в 18 третьоктавных полосах. Оказалось, что исходное 18-размерное пространство может быть без заметных потерь сокращено до четырех- или трехмерного. Признаки (оси) этого нового пространства являются линейными комбинациями начальных признаков [328].

На основании оценок субъективной разницы между гласно-подобными стимулами (11 гласных) и на основании матрицы опшибок при идентификации гласных были сконструированы трехмерные и четырехмерные субъективные пространства гласных. Оказалось, что эти трехмерные и четырехмерные конфигурации очень хорошо совмещаются с соответствующими конфигурациями, полученными при анализе [328, 431].

Трудность интерпретации этих результатов состоит в том, что конфигурация гласных в пространстве двух главных признаков, найденных в работе [328], практически идеально совмещается с конфигурацией гласных в пространстве частот первой и второй формант [432]. Авторы делают вывод об эквивалентности использованного ими полосного метода анализа с формантным анализом гласных. Этот вывод означает, что примененный авторами подход фактически не дает возможности определить, какими признаками спектра пользуется человек при восприятии гласных.

Данные, приводимые ниже в настоящей главе, позволяют надежно исключить только крайний вариант гипотезы кодирования величиной возбуждения. Он состоит в том, что число суммирующих элементов мало и каждый элемент имеет отличные от нуля коэффициенты связи с частотными каналами на всем интервале шкалы  $z$ .

#### 10.1.2. ПРОСТРАНСТВЕННОЕ КОДИРОВАНИЕ

Основной смысл гипотезы пространственного кодирования состоит в том, что обработка  $g(z)$  обеспечивает выделение (обнаружение) некоторых локальных особенностей этой функции. Характеристикой выделенной особенности является ее положение на шкале  $z$ . Значения этих характеристик используются при фоновой интерпретации гласных.

Эта точка зрения соответствует распространенному представлению о том, что, воспринимая гласный, человек выделяет форманты и основывается при классификации на значениях частот формант. Вместе с тем гипотеза пространственного кодирования и формантная гипотеза не обязательно совпадают, они лишь частично перекрываются.

Гипотеза пространственного кодирования заранее не определяет, какие именно особенности выделяются на  $g(z)$ . Этих особенностей может быть, в принципе, как больше, так и меньше числа формант в гласном звуке, особенность может или совпадать с формантой, или быть ее приближенной оценкой.

С другой стороны, с позиции формантной гипотезы безразлично, как кодируется частота форманты в нервной системе. Она может быть представлена как номером возбужденного нейронного канала, так, например, и частотой (периодом) импульсации. Важно лишь, чтобы соблюдалась монотонная зависимость между частотой форманты и соответствующей выходной переменной.



Для рассмотрения вопроса о том, какие особенности  $g(z)$  могли бы служить признаками при распознавании гласных и как можно было бы выделять эти особенности, необходимо прежде всего выяснить, насколько подробно отражен в  $g(z)$  спектр гласного.

Так как слуховой анализатор обладает ограниченной разрешающей способностью, естественно, что слуховой спектр гласного должен быть «размыт» по сравнению с идеальным спектром. Если бы свойства улитки как анализатора были полностью известны и функциональная модель периферической слуховой системы была создана,  $g(z)$  можно было бы просто промерить. К сожалению, окончательной модели улитки пока нет (см. главу 7). По этой причине все еще необходимо пользоваться психоакустическими методами измерения и расчета слухового спектра; принципы их будут описаны в следующем разделе.

## 10.2. ОПРЕДЕЛЕНИЕ СЛУХОВОГО СПЕКТРА СТИМУЛА

Для психоакустики  $g(z)$  является изображением сигнала на выходе функциональной модели периферической слуховой системы. Это изображение можно сконструировать, используя результаты психоакустических измерений и приняв ряд априорных допущений.

Модель периферической слуховой системы считается состоящей из набора частотных каналов; каждый канал включает фильтр звуковых частот и преобразователь. Проблема нахождения  $g(z)$ , соответствующей исследуемому звуку, разделяется на две задачи. Первая из них состоит в том, чтобы экспериментально определить (или вычислить) для каждого частотного канала модели эквивалентный уровень интенсивности чистого тона той частоты, на которую настроен данный канал. Под эквивалентным понимается такой уровень интенсивности тона, при котором величину импульсации в данном канале можно считать такой же, как при действии исследуемого звука. Совокупность найденных эквивалентных уровней мы будем называть эквивалентным спектром. Вторая задача заключается в переходе от эквивалентного спектра к  $g(z)$ . Для этого нужно знать амплитудную характеристику преобразователя.

В случае стационарных сигналов под амплитудной характеристикой преобразователя можно понимать зависимость между энергией сигнала на входе преобразователя за достаточно большой интервал времени и суммарной величиной импульсации за это же время.

Прямых методов измерения требуемой характеристики не существует. Подходы к ее определению основаны на рассмотрении полных моделей, осуществляющих различение или оценку сигналов. Подбирается такая амплитудная характеристика преобразо-

вателя, при которой достигается согласие результатов психоакустического эксперимента и эксперимента на модели.

Как будет показано в главе 11, модель, обнаруживающая неравномерности на огибающей стимула, должна включать преобразователи с близкой к логарифмической характеристикой. Вместе с тем характеристика преобразователей в модели, вычисляющей громкость звука, описывается степенной функцией [145].

Правильный выбор амплитудной характеристики важен не столько для вычислений  $g(z)$ , сколько для моделирования процесса выделения признаков  $g(z)$ . В настоящее время работа в этом направлении только начинается. Были опробованы как логарифмическая характеристика [131], так и степенная «функция громкости» [313]. Какая из них лучше, пока сказать нельзя.

#### 10.2.1. ЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ ИЗМЕРЕНИЯ ЭКВИВАЛЕНТНОГО СПЕКТРА

Наиболее известный метод основан на измерении маскировки стимулом тональных сигналов, предъявляемых на фоне этого стимула.

Принимается, что за обнаружение тона ответствен только тот частотный канал, который настроен на данный тон. Прибавление тона к стимулу вызывает приращение «текущей» энергии, попадающей в этот канал. Приращение обнаруживается (человек слышит тон), если оно в  $k$  раз превышает значение «текущей энергии», попадающей в данный канал в отсутствие тона. Таким образом, зная энергию тона, соответствующую порогу его обнаружения на фоне стимула, и зная  $k(z)$  [145], легко вычислить ту энергию, которая была в канале, когда тон отсутствовал — действовал один маскирующий стимул.

Произведя измерения порогов обнаружения тонов во всем частотном диапазоне, мы получаем эквивалентный спектр стимула, т. е. уровень энергии стимула в канале, как функцию  $z$  — частоты, на которую настроен канал.

Принципиальные ограничения практической применимости этого метода связаны с тем, что мозг человека обнаруживает не только медленные изменения огибающей  $g(t)$  в частотном канале периферической слуховой системы, отражающие изменения «текущей энергии», но и достаточно быстрые изменения (см. главу 11). Присоединение тестирующего тона к такому, например, стимулу, как гласный звук, приводит к возникновению биений между тоном и гармоническими составляющими гласного. Эти биения, естественно, отражаются во флюктуациях огибающей  $g(t)$  в частотных каналах. Так как обнаружение этих флюктуаций огибающей  $g(t)$  зависит от их частоты (частоты биений), полученная зависимость маскировки от частоты тестирующего тона оказывается чрезвычайно сложной.

По этой причине этот метод реально не применяется для экспериментального измерения эквивалентного спектра любого произвольного сигнала. Он был применен для определения кривых маскировки, вызываемых полосовыми шумами с шириной полосы, равной «критической» (см. главу 7). По полученным кривым маскировки были построены «образцы возбуждения» (см. ниже), используемые для вычисления эквивалентного спектра произвольного стимула по значениям его энергетического спектра [544].

Прямой экспериментальный метод измерения эквивалентного спектра, применимый для любых стационарных стимулов, был недавно предложен Хоутгастом [301, 302]. При этом методе испытуемому предъявляется последовательность, составленная из чередующихся посылок исследуемого стимула и тестирующего тона. Метод основан на допущении, что если огибающая плотности импульсации в частотном канале, соответствующем частоте тона, не изменяется на переходе от стимула к тону, то человек слышит непрерывный тон, на фоне которого воспринимаются посылки стимула. Если плотность импульсации является большей на протяжении тестирующего тона, то человек слышит пульсации тона.

В эксперименте определяется порог пульсаций, т. е. такой максимальный уровень интенсивности тестирующего тона, при котором, согласно принятому допущению, плотность импульсации во время тона еще практически не отличается от таковой во время стимула.

На рис. 10.1 приведен эквивалентный спектр (порог пульсаций) стимула, представляющего собой 10 равных по интенсивности гармоник частоты 250 Гц. Можно видеть, что если нижние гармоники представлены самостоятельными максимумами на эквивалентном спектре, то выделение верхних гармоник на кривой уже затруднительно. Из рис. 10.1 видно также, что порог пульсаций на частотах гармоник во многих случаях ниже уровня интенсивности гармоник, т. е. тон (гармоника), действующий в составе сложного звука, оказывается как будто бы менее эффективным, чем тот же тон, действующий изолированно (тестирующий тон).

Основным результатом работ Хоутгаста [300, 302] является обнаружение этого эффекта подавления и установление ряда его свойств. Автор показал, что по своим характеристикам эффект весьма напоминает так называемое двухтоновое торможение, наблюдавшееся впервые в ответах одиночных волокон слухового нерва [456, 457]. Природа этого эффекта до сих пор остается невыясненной. Для нас существенны два обстоятельства, отчетливо показанные Хоутгастом. Во-первых, эффект обеспечивает увеличение контрастности изображения, подчеркивает разницу между минимумами и максимумами на спектре. Во-вторых, наблюдаемые закономерности таковы, как если бы подавление осуществлялось за счет уменьшения коэффициента передачи в соответствующем частотном канале анализатора. Если это так, то методом одновременной маскировки этого эффекта принципиально нельзя обнаружить. Нечувствительность метода одновременной маски-

ровки к эффекту подавления, выявляемому при измерении порога пульсаций, была экспериментально показана Хоутгастом [302].

Из приведенных данных следует, что процедура вычисления эквивалентного спектра должна в принципе учитывать как «размазывание» идеального спектра звука, связанное с ограниченной разрешающей способностью фильтров улитки, так и обострение

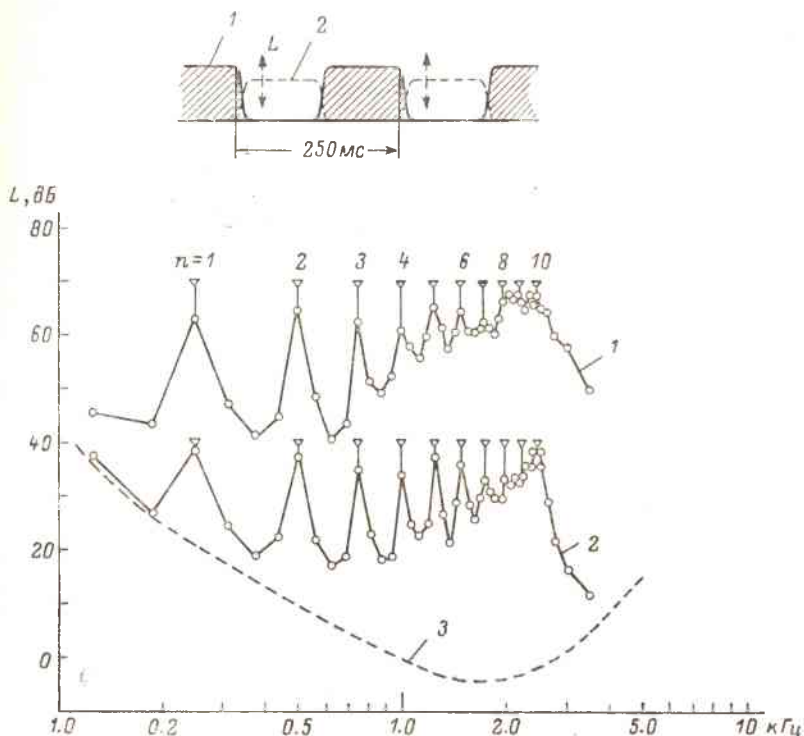


Рис. 10.1. Порог пульсаций для сложного стимула (10 гармоник частоты  $f_0=250$  Гц), измеренный при двух вариантах уровня интенсивности стимула (1 и 2). По [302].

По оси абсцисс — частота; по оси ординат — уровень звукового давления. Вверху дано схематическое изображение временной последовательности, состоящей из исследуемого стимула (1) и тестирующего тона (2); 3 — порог слышимости, треугольники — амплитуды гармоник в спектре стимула.

изображения, связанное с эффектом подавления. Однако, так как эти два эффекта действуют в противоположном направлении, достаточного приближения к действительности, возможно, удастся добиться, учитывая только фильтры (отражение их в характеристиках одновременной маскировки), но приписав им несколько большую добротность.

Иначе говоря, кажется возможным взять за основу процедуру вычисления эквивалентного спектра, предложенную Цвике-

ром [145 544], но изменить ее параметры так, чтобы получаемый результат максимально приближался к эквивалентному спектру, экспериментально получаемому методом порога пульсаций.

Основным параметром в процедуре расчета эквивалентного спектра является «образец возбуждения», повторяющий с неболь-

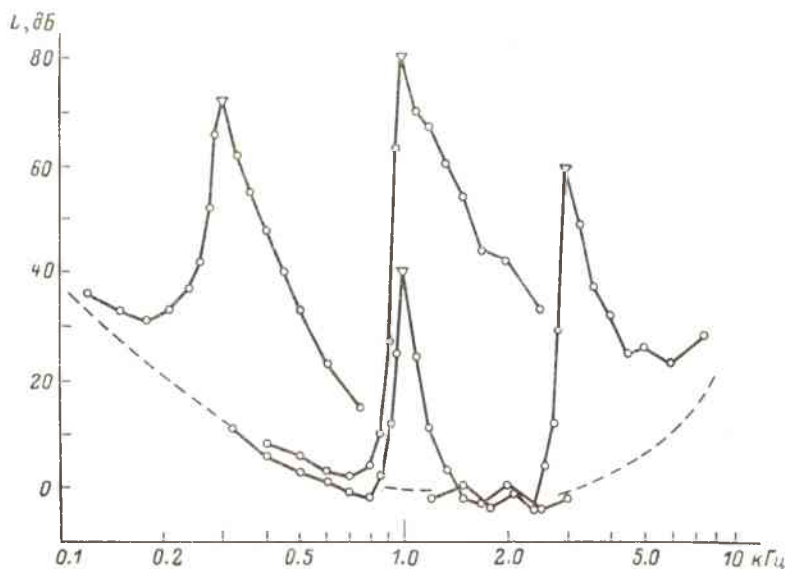


Рис. 10.2. Порог пульсаций для чистых тонов. По [302].

По оси абсцисс — частота; по оси ординат — уровень звукового давления. Треугольниками показаны характеристики (частота и уровень звукового давления) исследуемых тонов (маскеров). Штриховая кривая — порог слышимости.

шими поправками кривую маскировки. Цвикер аппроксимирует его трапецией с меньшим (верхним) основанием, равным ширине критической полосы. Данные Хоутгаста (рис. 10.2) свидетельствуют о том, что лучшей аппроксимацией должен быть треугольник с зависящим от уровня сигнала высокочастотным склоном.

### 10.2.2. ПРОЦЕДУРА ВЫЧИСЛЕНИЯ СЛУХОВОГО СПЕКТРА

Описываемая ниже процедура вычисления слухового спектра [77] является модификацией метода, предложенного Цвикером [145]. В согласии с Цвикером, принимается, что шкала  $z$  совпадает со шкалой высоты [540], называемой иначе шкалой Барк;<sup>1</sup> вид зависимости  $z(f)$  приведен на рис. 10.3. Кривые

<sup>1</sup> Один Барк соответствует расстоянию между двумя частотами, равному «критической» полосе (см. главу 7).



маскировки, получаемые при разных частотах маскера, практически совпадают, если они изображены на шкале высот [374]. Следовательно, приняв шкалу высот, можно резко упростить процедуру вычислений — использовать независимый от частоты «образец возбуждения» (см. ниже).

Вслед за Цвикером принимается также, что сигналом на выходе частотного канала является плотность громкости. Мы будем ее условно обозначать тем же символом,  $g$ , что и плотность импульсации. Использование плотности громкости дает возможность вычисления общей громкости стимула путем интегрирования  $g(z)$  по всему интервалу  $z$ .

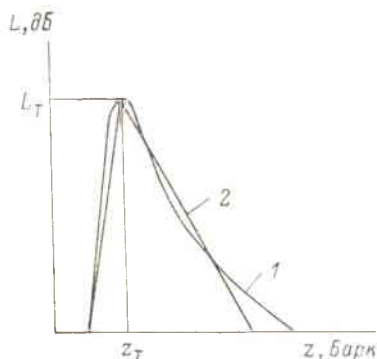
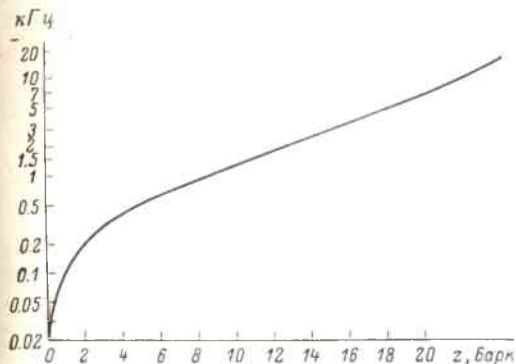


Рис. 10.3. Связь между частотой (по оси ординат) и высотой (по оси абсцисс). По [145].

Рис. 10.4. «Образец возбуждения» (1) и его аппроксимация (2).

По оси абсцисс — высота; по оси ординат — уровень интенсивности относительно абсолютного порога.

Проверка показала [77], что описываемая процедура вполне удовлетворительно воспроизводит известные психоакустические зависимости между громкостью звука и его спектром и является в этом смысле более мощной, чем исходный метод Цвикера.

Вычисление слухового спектра состоит из двух этапов. Первый этап интерпретируется как расчет энергий колебаний, вызванных стимулом, на выходе гребенки из 240 фильтров. Резонансные частоты фильтров образуют ряд по шкале  $z$  с шагом в 0.1 Барк.

Отклик гребенки фильтров на чистый тон с высотой  $z_T$  и уровнем  $L_T$  — «образец возбуждения» — аппроксимируется треугольником (рис. 10.4). Такой треугольник можно задать формулами

$$z < z_T, L(z) = L_T + 27(z - z_T),$$

$$z > z_T, L(z) = L_T - K(z - z_T).$$



Наклон  $K$  высокочастотной стороны зависит от уровня входного сигнала:

$$L_r < 60 \text{ дБ}, K = -0.27L_r + 25;$$

$$L_r > 60 \text{ дБ}, K = \frac{0.14L_r + 24.6}{0.078L_r - 1}.$$

При вычислении отклика на сложный стимул спектр стимула разбивается на полосы шириной 0.1 Барк; каждая полоса заменяется условной составляющей — чистым тоном с высотой  $z_r$ , равной высоте центра полосы, и уровнем  $L_r$ , определяемым энергией в полосе. Для каждого из этих тонов вычисляются по приведенной выше формуле уровни энергии отклика во всех точках шкалы  $z$ .

Общая энергия отклика в точке  $z$ , обозначим ее  $\alpha$ , определяется как сумма энергий откликов на условные составляющие в этой точке:

$$\alpha = \sum_{i=1}^m 10^{L_i(z)},$$

где  $m$  — количество условных составляющих.

Заметим, что  $\alpha(z)$  соответствует эквивалентному спектру стимула и должна, по идее, совпадать с эквивалентным спектром, измеренным в эксперименте.

Следующий этап расчета обеспечивает переход от  $\alpha(z)$ , или вычисленной описанным выше способом, или определенной в эксперименте, к  $g(z)$ .

Согласно Цвикеру, используется формула

$$g(z) = 0.08 \{ [\alpha(z)]^{0.25} - [\alpha_n(z)]^{0.25} \}, \quad (1)$$

где  $\alpha_n(z) = 10^{0.1L_n(z)}$ ,  $L_n(z)$  — порог слышимости (в дБ относительно абсолютного порога). Плотность громкости ( $g$ ) выражается в сон/Барк, где сон — принятая единица громкости [145].

Как говорилось выше, единственный существующий в настоящее время прямой метод измерения эквивалентного спектра, применимый для любых произвольных стимулов, основан на определении порога пульсаций. Поэтому для оценки качества описываемой процедуры существенно сравнить  $g(z)$ , полученную в результате вычислений  $\alpha(z)$  по энергетическому спектру стимула, с  $g(z)$ , полученной из экспериментально определенного (методом пульсаций) эквивалентного спектра.

На рис. 10.5 приведены данные для стимула из 10 гармоник, применявшегося в экспериментах Хоутгаста (рис. 10.1, Б, кривая 1). Кривая 1 показывает слуховой спектр (зависимость плотности громкости от частоты), полученный при использовании  $\alpha(z)$ , соответствующей кривой порога пульсаций. Кривая 2 — слуховой спектр, полученный при использовании  $\alpha(z)$ , вычисленной по линейчатому спектру стимула. Основной интерес при сравнении кривых 2 и 1 представляет изрезанность кривой (разница между

минимумами и максимумами) и изменение изрезанности с увеличением  $z$ . Эти характеристики отражают разрешающую способность слухового анализатора и зависимость разрешающей способности от частоты. Можно видеть, что по этим признакам обе кри-

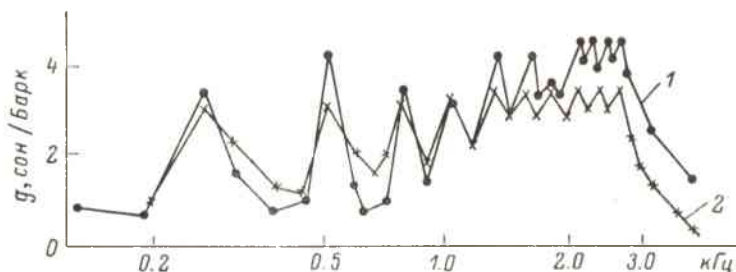


Рис. 10.5. Слуховой спектр сложного стимула, приведенного на рис. 10.1, Б.

По оси абсцисс — частота; по оси ординат — плотность громкости. 1 — результаты вычислений по значениям порога пульсаций  $[\alpha_{02}]$ ; 2 — результаты расчета по линейчатому спектру стимула.

вые достаточно похожи. Так как изрезанность кривой 1 в области нижних частот несколько больше, чем изрезанность кривой 2, следует считать, что, используя описанную выше процедуру вычисления, мы получаем слегка сглаженное по сравнению с истинным изображение слухового спектра.

### 10.2.3. ПРИМЕРЫ СЛУХОВЫХ СПЕКТРОВ ГЛАСНЫХ

На рис. 10.6 приведены пороги пульсаций для двух синтетических гласных  $[\alpha_{02}]$ . Так как порог определялся только на частотах гармонических составляющих стимула, полученные кривые следует рассматривать не как полный эквивалентный спектр стимула, но как его огибающую.

Рис. 10.7 показывает значения плотности громкости ( $g$ ) на частотах (значениях  $z$ ) гармоник синтетического  $[\alpha]$ , приведенного на рис. 10.6. Кривая 2 получена при использовании значений  $\alpha(z)$ , соответствующих порогу пульсаций. Кривая 1 получена при использовании значений  $\alpha(z)$ , вычисленных непосредственно по гармоническому спектру стимула. Можно видеть, что совпадение кривых является вполне удовлетворительным. Это означает, что вычислительная процедура обеспечивает достаточно близкую к реальной картину  $\alpha(z)$ .

На рис. 10.8 показан слуховой спектр естественного гласного  $[i]$ . Для вычислений использован гармонический спектр этого гласного, приведенный Фантом  $[\alpha_{53}]$ .

На рис. 10.9 показан слуховой спектр синтетического гласного  $[\alpha]$ , применявшегося в экспериментах Хирато и др.  $[\alpha_{96}]$ .

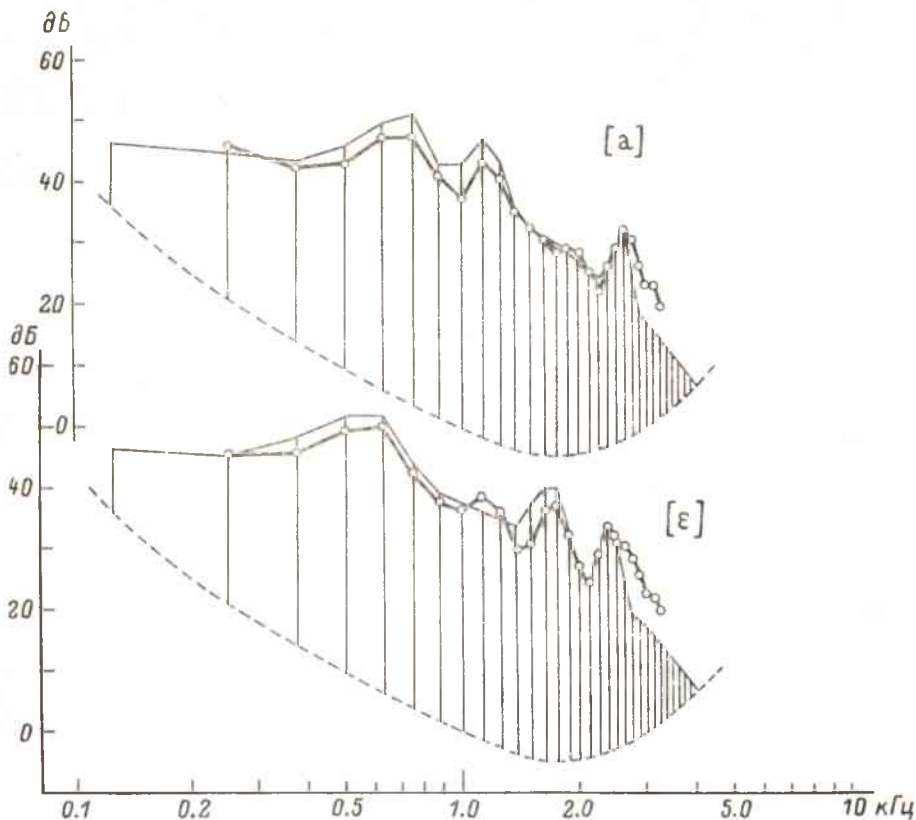


Рис. 10.6. Пороги пульсаций для синтетических гласных [а] и [ε]. По [302].

По оси абсцисс — частота; по оси ординат — уровень звукового давления. Вертикальные линии, соединенные кривой, — амплитуды гармонических составляющих в спектре гласного. Кружки, соединенные жирной кривой, — значения порога пульсаций на частотах гармоник стимула. Штриховая кривая — порог слышимости.

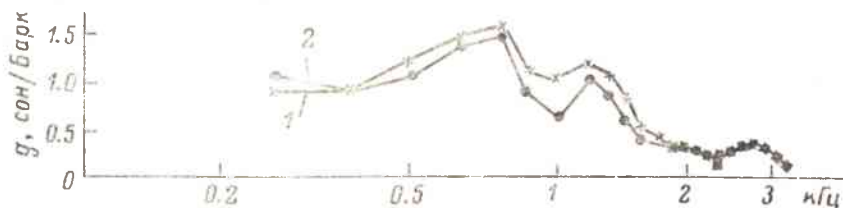


Рис. 10.7. Обходящее слухового спектра синтетического гласного [а], линейчатый спектр которого приведен на рис. 10.6.

По оси абсцисс — частота; по оси ординат — плотность громкости. 1 — результаты расчета по линейчатому спектру стимула; 2 — результаты вычислений по значениям порога пульсаций [302], приведенного на рис. 10.6.

Общий вывод, основанный на этих примерах, а также на результатах вычислений слуховых спектров большого числа естественных русских гласных, состоит в том, что в слуховом спектре обнаруживается, как правило, больше чем два формантных мак-

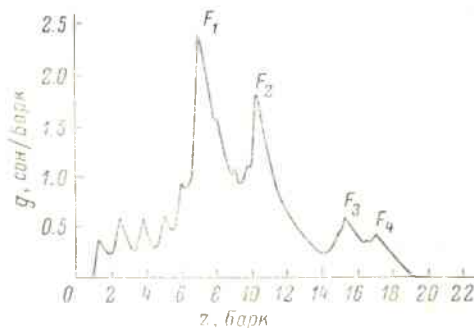
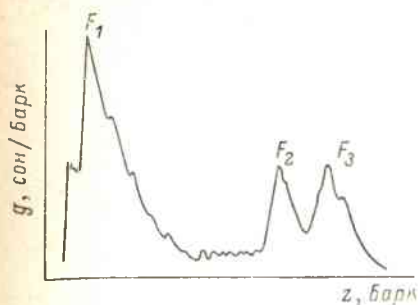


Рис. 10.8. Слуховой спектр естественного гласного [i], рассчитанный по линейчатому спектру, приведенному Фантом [253].

По оси абсцисс — высота; по оси ординат — плотность громкости.  $F_1$ — $F_3$  — формантные максимумы.

Рис. 10.9. Слуховой спектр синтетического гласного [a], рассчитанный по линейчатому спектру стимула, из работы Хирато и др. [296].

Обозначения те же, что на рис. 10.8.

симула (рис. 10.8 и 10.9), кроме того, имеются максимумы, соответствующие первым низкочастотным гармоникам (рис. 10.9).

Если верно распространенное допущение о том, что фонемное распознавание гласных основывается на весьма сглаженном изображении их спектров [191, 206, 425], то такое сглаживание следует, очевидно, искать на уровнях дополнительной обработки  $g(z)$ , обеспечивающих выделение признаков.

### 10.3. ПРИРОДА ПОЛЕЗНЫХ ПРИЗНАКОВ СПЕКТРА

Экспериментальные данные по фонемной классификации стационарных речеподобных стимулов позволяют утверждать, что по крайней мере часть из полезных признаков имеет локальный характер, т. е. касается особенностей спектра в ограниченной частотной области. Кроме того, данные свидетельствуют о том, что край и максимум на спектре в известном смысле эквивалентны.

#### 10.3.1. ЛОКАЛЬНЫЙ ХАРАКТЕР ПРИЗНАКОВ

Вывод о локальном характере признаков напрашивается уже из того приведенного в главе 4 факта, что для ряда пар гласных фонемная граница в пространстве формант определяется только частотой одной из формант и не зависит от час-

тот других формант. В случае русских слушателей такая ситуация была обнаружена для пар [i]—[e], [o]—[a], [o]—[e].

Другим доводом в пользу локального характера признаков является нечувствительность фонемной границы по частоте форманты к значительным искажениям общей формы спектра, заключающимся в искусственном усилении или, наоборот, подавлении энергии в области первой форманты.

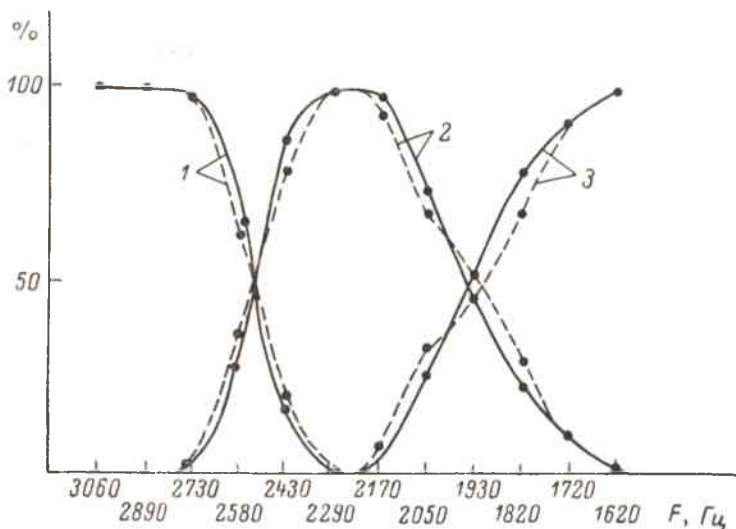


Рис. 10.10. Результаты идентификации синтетических гласных с ослабленной (сплошные кривые) или усиленной (штриховые кривые) первой формантой. По [365].

По оси абсцисс —  $F = \sqrt{F_2 F_3}$ , где  $F_2$  — частота второй форманты,  $F_3$  — частота третьей форманты стимула; по оси ординат — процент идентификации стимула с соответствующими гласными [i] — 1, [y] — 2 и [u] — 3.

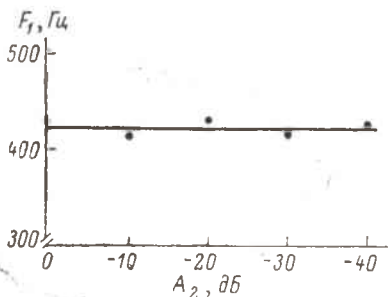
В работе Линдквиста и Паули [365] исследовалось восприятие синтетических гласных [i], [y], [u]. Все стимулы имели одинаковую частоту первой форманты, равную 266 Гц. Вторая и третья форманты изменялись таким образом, что среднее геометрическое их частот  $\sqrt{F_2 F_3}$  ( $F_3/F_2 = \text{const}$ ) принимало значения в диапазоне от 1620 до 3060 Гц. Стимулы были сначала синтезированы с нормальными для речевого тракта амплитудными отношениями между формантами. Затем с помощью полосовых фильтров вводились амплитудные искажения — область первой форманты в одном варианте усиливалась на 12.5 дБ по сравнению с нормой, во втором варианте подавлялась на 12.5 дБ.

Функции идентификации, полученные при этих двух вариантах искажений, приведены на рис. 10.10. Можно видеть, что функции идентификации, а следовательно и фонемные границы, в обоих случаях совпадают.

В работе Мушников и Чистович [107] определялась граница по  $F_1$  между [i] и [e] при разных значениях уровня интенсивности второй форманты двухформантного гласного. Полученные данные показали (рис. 10.11), что изменения уровня интенсивности в диапазоне 40 дБ не влияют на положение границы.

Рис. 10.11. Положение границы по частоте первой форманты между гласными [i] и [e] при разных значениях интенсивности второй форманты ( $F_2 = 2250$  Гц). По [107].

По оси абсцисс — уровень интенсивности второй форманты; по оси ординат — частота первой форманты.



Вывод о том, что существенным является именно частотное положение спектрального максимума, а форма спектра сигнала даже вблизи этого максимума не имеет большого значения, следует также из данных по фонемной классификации стационарных шум-

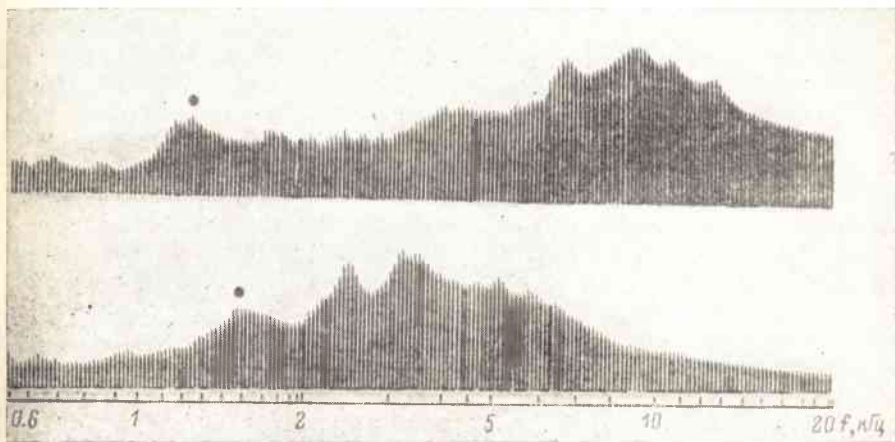


Рис. 10.12. Спектры естественных изолированно произнесенных согласных [s] (вверху) и [ʃ] (внизу). По [165].

ных согласных [165, 167]. На рис. 10.12 приведены спектры естественных изолированных согласных [s] и [ʃ]. Точкой отмечен спектральный максимум, частотное положение которого, как показано в [165], определяет высоту звука и одновременно используется носителями русского языка для различения твердых и мягких согласных.



В контрольных экспериментах [167] использовались полусинтетические стимулы: основная часть спектра стимула (выше 2500 Гц) создавалась за счет естественных согласных, нижняя часть спектра создавалась с помощью шума, пропущенного через резонансный контур с перестраиваемой частотой. Определялась частота контура (спектрального максимума), соответствующая границе между твердыми и мягкими согласными. Оказалось, что значения границы практически идеально совпали для [s] и [ʃ], хотя, как видно из рис. 10.12, спектры этих звуков в основной их части сильно различаются.

Для того чтобы приведенные данные были совместимы с гипотезой кодирования величиной возбуждения (см. раздел 10.1.1), необходимо допустить, что коэффициенты связи суммирующего элемента с частотными каналами периферической слуховой системы не равны нулю лишь на весьма ограниченных участках шкалы  $z$ . Лишь в этом случае отклик сумматора будет определяться только локальными свойствами спектра. Очевидно, что при этом допущении гипотеза теряет всю свою привлекательность.

### 10.3.2. НЕОДНОРОДНОСТЬ НА СПЕКТРЕ КАК ФОНЕТИЧЕСКИЙ ПРИЗНАК

В работе [108] была сделана попытка определить, какими особенностями должен характеризоваться спектр синтетического гласного, чтобы человек мог обнаружить присутствие в этом гласном второй форманты.

В экспериментах применялись двухформантные синтетические гласные с  $F_1=600$  Гц,  $F_0=300$  Гц и переменными значениями  $F_2$ . Когда вторая форманта в стимуле полностью подавлялась, сигнал воспринимался как [o] или [u]. Определялось минимальное значение амплитуды второй форманты, при котором звук изменял фонемное качество. В результате эксперимента был получен набор спектров, соответствующих порогу обнаружения второй форманты. Три из них приведены на рис. 10.13. Можно видеть, что в спектре, соответствующем стимулу с  $F_2=1500$  Гц, формально нет второго максимума (амплитуда гармоники с частотой 1500 Гц равна таковой для ближайшей более низкой гармоники). В данном случае приходится говорить не о максимуме, а о неоднородности на спектре.

В ряде психоакустических экспериментов показано, что высота полосового или широкополосного шума со срезом в области низких (высоких) частот определяется частотой края (или краев) на спектре этого шума [257, 470].

В одном из вариантов упоминавшихся выше экспериментов [166] по восприятию твердости/мягкости полусинтетических согласных [s] и [ʃ] низкочастотная часть спектра стимула создавалась за счет шума, пропущенного через фильтр высоких частот; исследуемой переменной была частота среза. Оказалось, что край на спектре

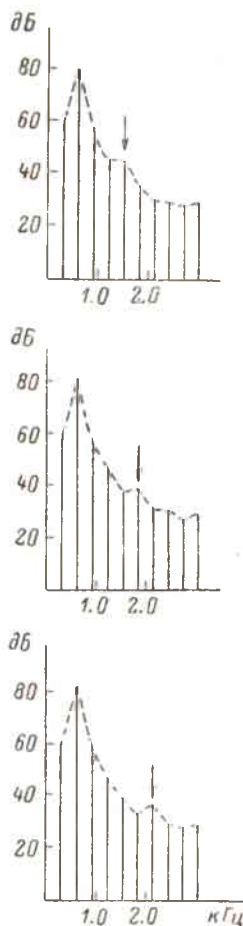
играет ту же роль, что и спектральный максимум. В зависимости от значения частоты среза согласный воспринимался как твердый или мягкий.

Таким образом, приведенные в настоящем разделе экспериментальные данные позволяют сформулировать два требования к модели, осуществляющей выделение признаков на слуховом спектре стимула: 1) модель должна выделять локальные особенности функции  $g(z)$ , 2) модель должна схожим образом реагировать на максимумы и на неоднородности на  $g(z)$ . Этим требованиям удовлетворяет модель латерального торможения. Использование латерального торможения для выделения неоднородностей на  $g(z)$  представляется достаточно правдоподобным с точки зрения физиологии слуха.

Подчеркнем еще раз, что допущение гипотезы латерального торможения еще отнюдь не означает, что тем самым автоматически принимается формантная гипотеза. В зависимости от параметров модели латерального торможения она будет выделять или форманты, или какие-то другие особенности на слуховом спектре гласного.

Рис. 10.13. Линейчатые спектры синтетических гласных, соответствующие порогу обнаружения второй форманты. По [108].

По оси абсцисс — частота; по оси ординат — уровень интенсивности. Стрелкой показана гармоника с частотой  $F_2$ .



#### 10.4. ВЫДЕЛЕНИЕ НЕОДНОРОДНОСТЕЙ НА СЛУХОВОМ СПЕКТРЕ МОДЕЛЬЮ ЛАТЕРАЛЬНОГО ТОРМОЖЕНИЯ

Хотя идея использования латерального торможения в моделях слуховой обработки сигнала является весьма популярной (см. обзор [114]), работы, направленные на определение значений параметров латерального торможения, совместимых с психоакустическими данными, находятся в самой начальной стадии.

Если вход модели определен как слуховой спектр стимула (см. раздел 10.2) и сделан ряд общепринятых допущений относительно структуры модели (см. главу 9), то требующими определе-

ния параметрами остаются весовые функции возбуждающих и тормозных связей и значения порога. При этом следует еще подчеркнуть, что вопрос о способе взаимодействия суммарных возбуждающих и тормозных сигналов пока остается открытым (см. главу 9). Неясно также, какую амплитудную характеристику преобразователя энергии в плотность импульсации (степенную или логарифмическую) следует предпочесть. В модели Карницкой [313] принималось первое допущение, Темов [31] рассматривал вариант, соответствующий второму допущению.

В обеих работах использовался следующий подход. Выбирался небольшой набор спектров стимулов, для которых было установлено, что человек обнаруживает в стимулах те или другие неоднородности. Принималось, что обнаружение неоднородности соответствует представлению этой неоднородности в отклике модели положительным выбросом  $g^*(z)$ , отделенным от соседних выбросов некоторым интервалом, где  $g^*(z)$  ниже порога. Требовалось найти параметры весовых функций (вид функций принимался априорно), при которых это условие, а также некоторые дополнительные условия выполняются. Существенный результат, полученный в обеих работах, состоял в том, что интервал  $\Delta z$ , на котором задаются весовые функции, должен быть достаточно узким, не должен превышать 2—3 Барк.

Остановимся в качестве иллюстрации на одной из попыток подбора параметров, предпринятой Карницкой. За основу при подборе параметров были взяты данные эксперимента Хирато и др. [296] по восприятию синтетического гласного [a]. Эксперимент состоял в том, что испытуемые сравнивали стандартный гласный (значения амплитуд его 32 гармоник показаны точками на рис. 10.14) с тем же гласным, одна из гармоник которого была уменьшена по амплитуде. Определялось минимальное уменьшение уровня интенсивности гармоники, при котором человек различает эти два стимула. Такие измерения поочередно проводились для всех гармоник. Полученные данные приведены в нижней части рис. 10.14. Крестиками отмечены те гармоники, исключение которых из спектра звука испытуемые не обнаруживали.

Можно видеть, что в области выше 1500 Гц человек чувствует уменьшение интенсивности только тех двух гармоник, которые соответствуют формантным максимумам (третьей и четвертой формантам). Рис. 10.9 показывает, что на слуховом спектре гласного эти верхние формантные максимумы отчетливо представлены.

При подборе параметров модели латерального торможения была принята следующая интерпретация данных Хирато и др. Считалось, что на отклике модели все четыре формантные пика должны быть представлены отделенными друг от друга выбросами. Модель «чувствует» уменьшение интенсивности гармоники, если оно приводит или к сдвигу положения выброса по шкале  $z$ , или к изменению порядковых отношений между выбросами по их величине (например, четвертый выброс становится бóльшим,

чем третий). Кроме того, использовалось еще одно ограничение, состоящее в том, что отклик модели на чистый тон должен характеризоваться только одним положительным выбросом.

В отношении вида весовых функций было принято, что для каждого из 240 элементов модели возбуждающая связь ограничена одним соответствующим ему периферическим частотным каналом (одним фильтром модели вычисления слухового спектра), а весовая функция тормозных связей имеет колоколообразную

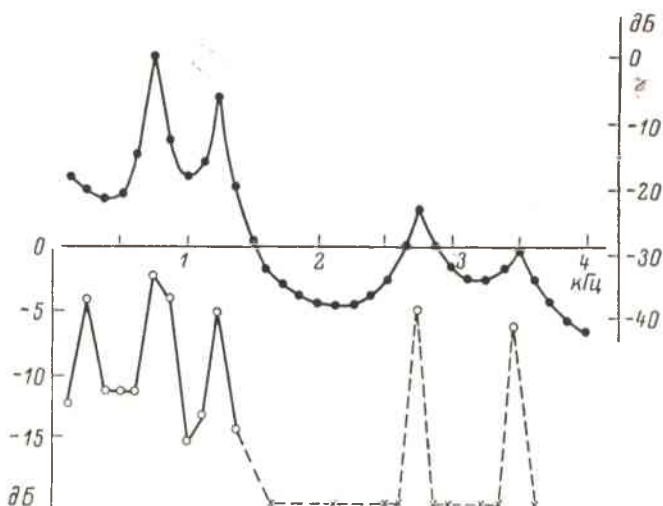


Рис. 10.14. Восприятие изменений интенсивности отдельных гармоник в спектре синтетического гласного [а]. По [298].

По оси абсцисс — частота гармоники; по оси ординат справа — относительный уровень интенсивности гармоник в спектре стандартного гласного, слева — пороговые уменьшения уровня тестируемой гармоники, при котором обнаруживается различие между стандартным и переменным гласными. Остальные обозначения см. в тексте.

форму (импульс Гаусса) и характеризуется двумя параметрами —  $\beta$  и  $A$  (рис. 10.15). Порог принимался равным нулю.

Оказалось, что наилучшее согласие с данными эксперимента Хирато и др. достигается при  $\beta=0.9$  Барк и  $A=-0.088$ , причем диапазоны допустимых значений параметров являются очень узкими. Отклик модели латерального торможения,  $g^*(z)$ , на слуховой спектр исходного гласного [а] при этих значениях параметров приведен на рис. 10.16. Можно видеть, что, кроме формантных максимумов, модель выделяет также максимумы, соответствующие пяти нижним гармоникам  $F_0$  ( $F_0=125$  Гц).

Основной вывод, вытекающий из пока еще весьма ограниченного опыта исследования моделей латерального торможения, состоит в следующем: 1) весовые функции должны быть достаточно узкими по шкале  $z$ , иначе модель не будет выделять тех неоднородностей на спектре гласных, которые обнаруживает человек;

2) при описании отклика модели приходится учитывать не только положение положительных выбросов на шкале  $z$ , но также хотя бы порядковые отношения между выбросами по их величине; без

этого модель не будет различать гласные стимулы, различаемые человеком.

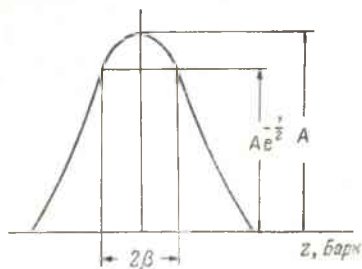


Рис. 10.15. Весовая функция тормозных связей, принятая в модели латерального торможения.

$A$  и  $\beta$  — параметры весовой функции.

Иначе говоря, совместимая с психоакустикой модель латерального торможения не может еще обеспечить чисто пространственного кодирования спектральной информации — кроме сведений о положении неоднородностей по шкале  $z$ , должна еще в какой-то мере присутствовать информация о величине, или «мощности»,

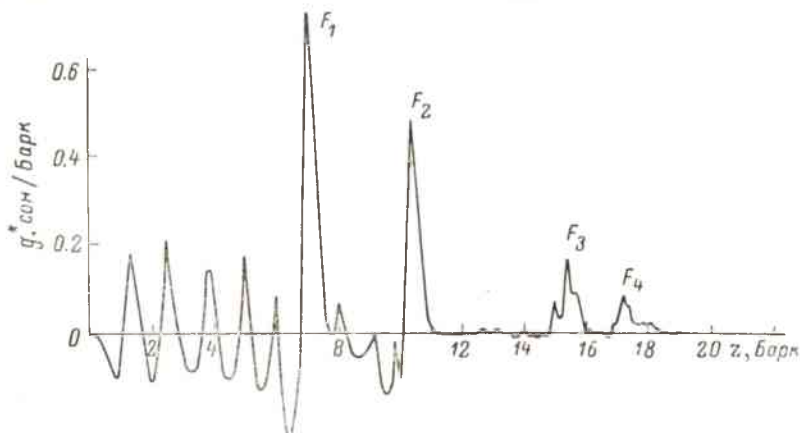


Рис. 10.16. Отклик модели латерального торможения на слуховой спектр синтетического гласного [а], приведенный на рис. 10.9.

По оси абсцисс — высота; по оси ординат — величина отклика.  $F_1$ — $F_4$  — формантные максимумы.

неоднородностей. Кроме того, число выделяемых неоднородностей оказывается в общем случае большим числа формант в гласном.

Так как это приходит в противоречие с формантной гипотезой восприятия гласных, приходится думать, что или выделение формант происходит каким-то другим способом, или сама формантная гипотеза является ошибочной и человек пользуется при фонемной классификации гласных более «богатым» описанием спектра стимула, чем это предполагается формантной гипотезой.



## 10.5. ЧИСЛО ПОЛЕЗНЫХ ПРИЗНАКОВ ГЛАСНЫХ

Известно, что при синтезе гласных [u], [o], [a] можно обойтись одним формантным контуром или группой из двух-трех соседних по частоте гармоник. Испытуемый будет достаточно уверенно идентифицировать эти стимулы с фонемами, хотя естественность гласного будет значительно снижена. Весь набор гласных может быть синтезирован с использованием двух формант [229]. Существенно, что значения  $F_2$  в двухформантном гласном должны в случае гласных [i], [e] существенно превышать по частоте значения  $F_2$  в естественном гласном [208, 253].

Одна из возможных интерпретаций этих фактов состоит в том, что человек использует в качестве признаков спектра естественного гласного не формантные максимумы, а что-то вроде «центров тяжести». В принципе такие признаки могли бы появиться, если бы разрешающая способность слухового анализатора была очень низкой (это предположение, как мы видели, исключается) или в модели латерального торможения использовались бы весьма широкие по  $z$  весовые функции. Эта интерпретация была очень отчетливо высказана в работах [206, 207]. Авторы использовали предположение, что, воспринимая четырехформантный (близкий к естественному) синтетический гласный, человек определяет его значения всего по двум признакам — двум «субъективным» формантам. По таким же двум признакам описывается при восприятии и двухформантный гласный. Если эта точка зрения верна, то, найдя пары, состоящие из максимально субъективно близких четырехформантного и двухформантного гласных, можно попытаться придумать такую модель (процедуру обработки сигналов), отклик которой на четырехформантный и двухформантный гласные будет одинаковым. Эта модель и будет, согласно принятому допущению, моделью восприятия гласного.

Однако априорно допустима и такая точка зрения, что в четырехформантном или естественном гласном человек выделяет больше признаков, чем реально есть в двухформантном или одноформантном гласном. Если перед человеком ставится задача подобрать к «богатому» признаками сигналу такой «бедный» признаками сигнал, чтобы оба звука были максимально похожи, ему, возможно, ничего не остается, кроме как игнорировать некоторые из признаков «богатого» сигнала, т. е. ориентироваться лишь на самые важные признаки. Главная трудность, с которой сталкивается эта вторая точка зрения, состоит в том, что она необходимо предполагает использование какой-то меры важности, или мощности, признака.

Экспериментальная процедура, при которой от испытуемого требуется подобрать к предположительно более «богатому» сигналу наиболее похожий на него более «бедный» сигнал, широко используется в психоакустике (подбор стационарного тона к тону, изменяющемуся во времени, подбор тона к полосовому шуму



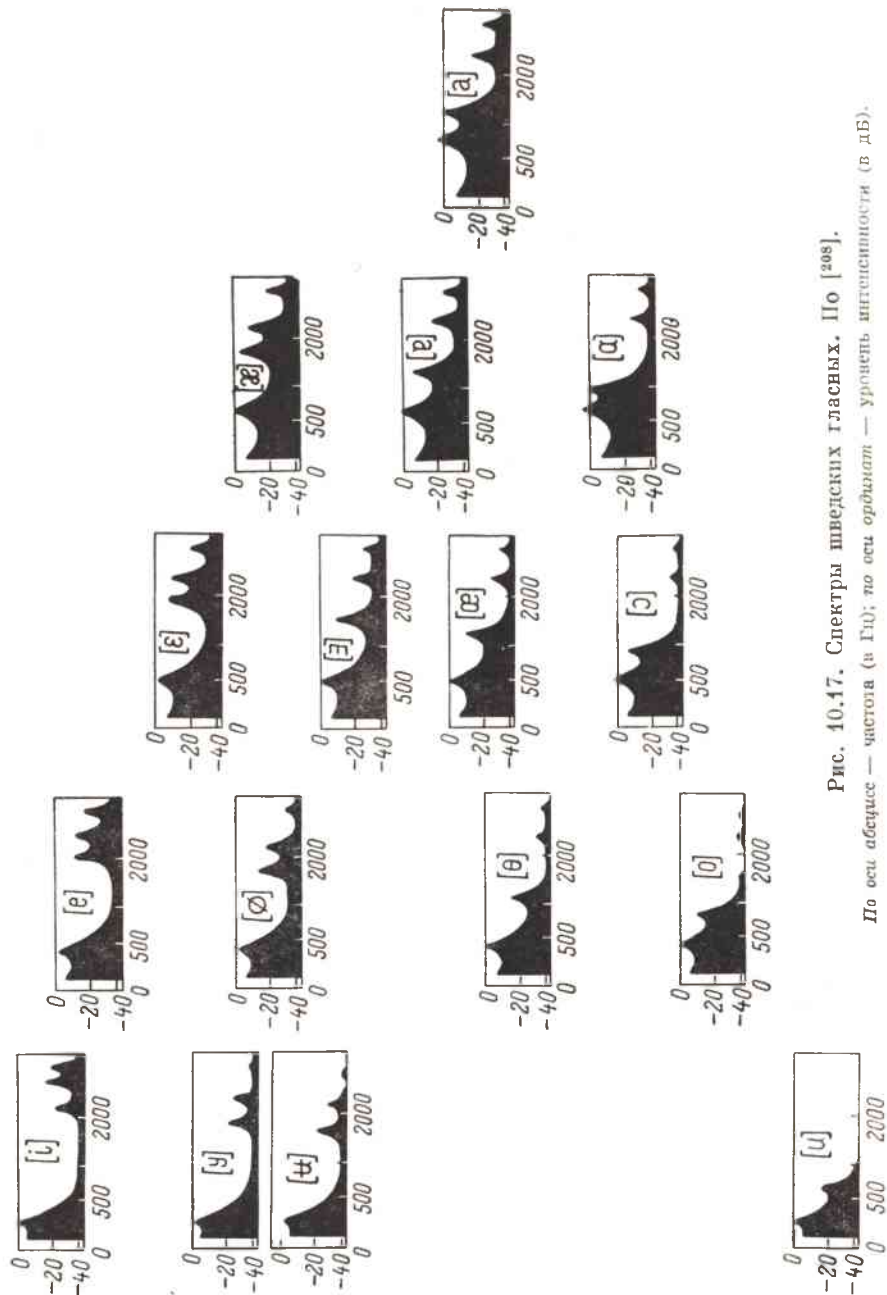


Рис. 10.17. Спектры шведских гласных. По [308].

По оси абсцисс — частота (в Гц); по оси ординат — уровень интенсивности (в дБ).

и т. д.). Основная информация, получаемая в таком эксперименте, содержится в гистограмме подобранных значений. Только в том случае, если распределение подобранных значений является одно-модальным, можно делать вывод о том, что человек основывается на некотором среднем из предположительных параметров исследуемого стимула. Наличие двух максимумов на распределении (соответствующих, например, начальному и конечному значениям частоты изменяющегося стимула, или частотам среза полосовых шумов) подтверждает гипотезу большого «богатства» исследуемого стимула, а относительная выраженность максимумов позволяет

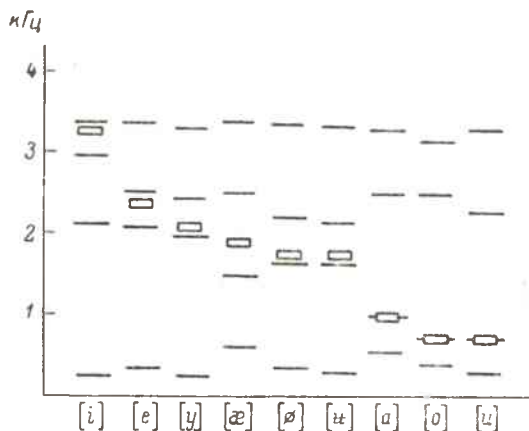


Рис. 10.18. Результаты подбора  $F'_2$  для достижения наибольшего сходства двухформантного гласного с четырехформантным. По [208].

Горизонтальными линиями показаны частоты формант в исходных четырехформантных гласных; прямоугольниками показаны средние значения  $F'_2$ .

примерно оценить относительную важность параметров. Такая характеристика как среднее из подобранных значений несет очень мало информации, так как не позволяет выбрать между гипотезами.

Во всех описываемых ниже экспериментах использовалась процедура подбора более «бедного» сигнала к предположительно более «богатому».

В уже упоминавшейся работе Карлсона и др. [208] применялись четырехформантные синтетические гласные, огибающие спектров которых приведены на рис. 10.17. Испытуемые сравнивали эти гласные с двухформантными гласными. Двухформантный гласный совпадал с четырехформантным по  $F_1$ . Задача испытуемого состояла в подборе  $F'_2$ , т. е. такого значения частоты второй форманты двухформантного гласного ( $F'_2$ ), при котором сравниваемые гласные оказывались наиболее близкими.

Средние из подобранных значений приведены на рис. 10.18. Можно видеть, что для [u], [o], [a] они хорошо совпадают с  $F'_2$  в четырехформантном гласном. В остальных случаях наблюдается отчетливое смещение среднего в сторону более высоких формант.

На основании этих данных авторы сделали вывод о том, что все три верхние форманты описываются при восприятии одним параметром, являющимся функцией или только частот этих формант (один вариант) или всего спектра в высокочастотной области (второй вариант). Так как авторы не приводят никаких данных относительно распределений устанавливаемых значений  $F'_2$ , вопрос об обоснованности этого вывода остается открытым. Сопоставив

рис. 10.17 и 10.18, легко заметить, что близость  $F'_2$  к тому или другому формантному максимуму, возможно, связана с амплиту-

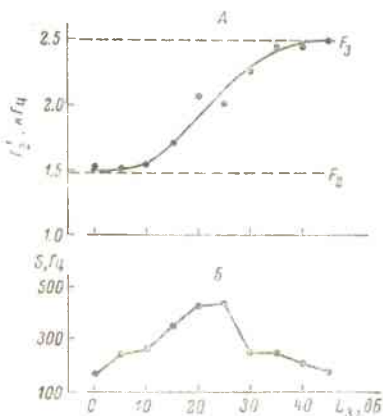


Рис. 10.19. Результаты подбора  $F'_2$  для достижения наибольшего сходства двухформантного гласного с трехформантным.

А — среднее значение  $F'_2$  (по оси ординат) как функция уровня интенсивности третьей форманты в трехформантном гласном; Б — стандартное отклонение (по оси ординат) подбираемых значений  $F'_2$ . По оси абсцисс — уровень интенсивности третьей форманты. Суммарные данные трех испытуемых.

дой этого максимума на спектре. Смещения  $F'_2$  и  $F_2$  не наблюдается для тех гласных, где относительные амплитуды верхних формант очень малы ([u], [o], [a]). В случае [i] третья и четвертая форманты больше по амплитуде, чем вторая, соответственно среднее  $F'_2$  находится между  $F_3$  и  $F_4$ . Для [e] амплитуда третьей форманты больше амплитуды второй и  $F'_2$  находится ближе к  $F_3$ ; для [y] амплитуда второй форманты больше, чем амплитуда третьей, и  $F'_2$  располагается ближе к  $F_2$ . Наконец, в случае [æ] амплитуды второй и третьей формант примерно равны и  $F'_2$  располагается примерно посередине между  $F_2$  и  $F_3$ .

Роль относительных амплитуд формант была показана в экспериментах [400] по подбору  $F'_2$  в двухформантном гласном для получения наибольшего сходства с трехформантным гласным.  $F_0$  и  $F_1$  в обоих стимулах совпадали ( $F_0=120$ ,  $F_1=300$  Гц). Частоты второй и третьей формант в трехформантном стимуле составляли 1480 и 2500 Гц соответственно. Переменным параметром трехформантного стимула являлся относительный уровень интенсивности третьей форманты. Он изменялся в диапазоне 45 дБ шагом до 5 дБ.

Рис. 10.19, А показывает, что при малых уровнях интенсивности третьей форманты  $F'_2$  совпадает с  $F_2$ , при больших уровнях

третьей форманты  $F'_2$  совпадает с  $F_3$ . Переходный участок, где  $F'_2$  находится между  $F_2$  и  $F_3$ , составляет примерно 25 дБ. Существование, что среднее квадратичное отклонение устанавливаемых значений  $F'_2$  (рис. 10.19, *Б*) резко возрастает на этом переходном участке. Это связано с тем, что на переходном участке испытуемые устанавливают  $F'_2$  в части случаев близко к  $F_2$ , а в части случаев — близко к  $F_3$  (см. рис. 10.20, где показаны распределения установленных значений  $F'_2$  для одного испытуемого).

Данные рис. 10.19 и 10.20 несовместимы с предположением, что  $F_2$  и  $F_3$  описываются всего одним обобщенным параметром. Необходимо принять, что в диапазоне уровней, по крайней мере не меньшем 10 дБ, человек располагает информацией о частотах и  $F_2$ , и  $F_3$ , т. е. обе форманты представлены как самостоятельные признаки гласного.

Отсутствие усреднения по  $z$  при восприятии гласных было показано в работе [217]. В роли предположительно более богатых признаками стимулов выступали синтетические двухформантные гласные. Для исключения гармонической структуры спектра использовалось возбуждение формантных фильтров от генератора шума. Частоты  $F_1$  и  $F_2$  изменялись в диапазоне 200—1100 Гц, разница формантных частот ( $F_2 - F_1$ ) составляла 100, 150 и 200 Гц. Всего в эксперименте было использовано 26 сочетаний формантных частот. Испытуемому предъявлялись следующие друг за другом пары стимулов, один из которых имел две форманты (частоты их устанавливались экспериментатором), второй стимул был одноформантным, а частота его форманты  $F'_1$  подбиралась испытуемым для получения наибольшего сходства стимулов в паре. Гистограммы подобранных частот  $F'_1$  представлены на рис. 10.21. Значения по оси абсцисс соответствуют нормированному расстоянию между величинами  $F'_1$  и частотами формант исходного стимула, 0 означает равенство  $F'_1$  с  $F_1$ , 1 — равенство  $F'_1$  с  $F_2$ . Гистограмма *А* построена для случая, когда расстояние между  $F_1$  и  $F_2$  больше критической полосы. Гистограмма *Б* — для случая, когда расстояние равно или меньше критической полосы. Можно видеть, что только вторая гистограмма характеризуется одним максимумом, находящимся на частоте, близкой к половине расстояния между формантами. Гистограмма *А* свидетельствует о том, что  $F'_1$  устанавливается испытуемыми на частотах, близких к  $F_1$  и  $F_2$ , т. е. что и  $F_1$ , и  $F_2$  выступают в роли самостоятельных признаков гласного.

Заметим, что расстояние между  $F_1$  и  $F_2$  в этих экспериментах (100—200 Гц) соответствует расстоянию между гармониками в спектре нормальных гласных, образуемых при голосовом источнике возбуждения.

Вопрос о том, определяет ли человек при восприятии некоторый суммарный максимум на слуховом спектре, соответствующий первой форманте в гласном с гармоническим спектром, или он

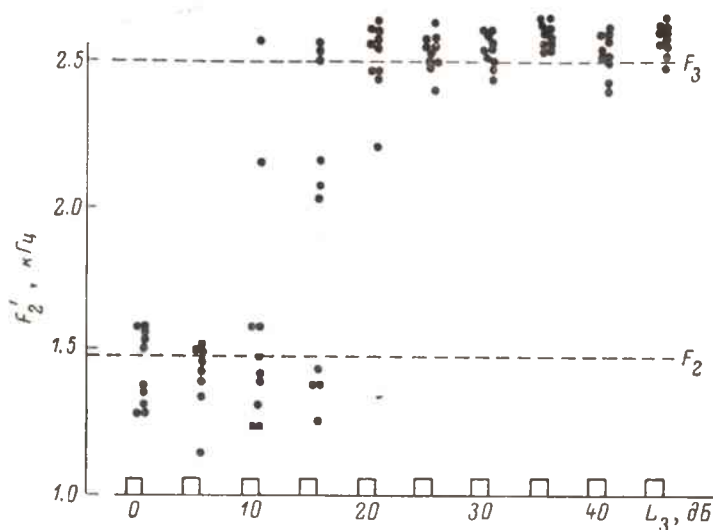


Рис. 10.20. Распределения значений  $F_2'$ , установленных одним испытуемым при подборе двухформантных гласных, максимально приближающихся к трехформантным гласным с различными уровнями интенсивности третьей форманты ( $L_3$ ).

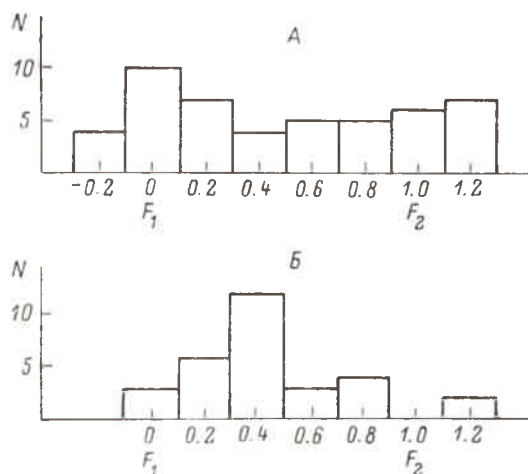


Рис. 10.21. Гистограммы значений  $F_1'$ , установленных при подборе одноформантного стимула, максимально приближающегося к двухформантному гласному.

А — разница  $F_1$  и  $F_2$  в исходных двухформантных стимулах превышает ширину критической полосы; Б — разница равна или меньше ширины критической полосы. Обозначения см. в тексте.

выделяет максимумы, соответствующие отдельным гармоникам, исследовался в работе [163].

В роли предположительно более «богатого» признаками сигнала выступал двухформантный гласный с фиксированным значением  $F_2=2300$  Гц и переменным значением  $F_1$ . Уравниваемый с ним стимул представлял собой сумму чистого тона и сигнала, соответствующего изолированной второй форманте с тем же значением  $F_2=2300$  Гц. Эксперимент проводился при нескольких значениях частоты основного тона в диапазоне от 80 до 200 Гц. Частота  $F_1$  двухформантного гласного менялась в диапазоне

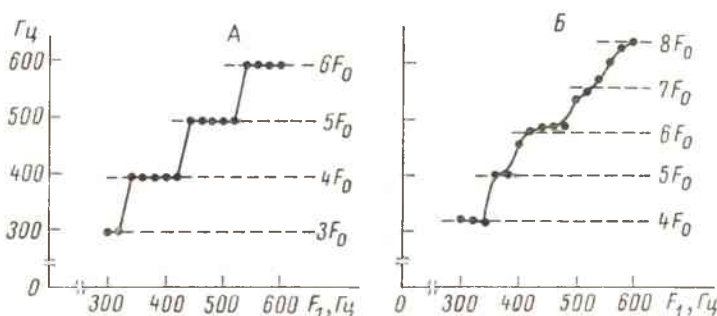


Рис. 10.22. Зависимость среднего значения подобранной частоты тона в сложном стимуле (по оси ординат) от частоты первой форманты в двухформантном гласном (по оси абсцисс). По [163].

А — частота основного тона ( $F_0$ ) в двухформантном гласном равна 100 Гц; Б —  $F_0=80$  Гц.

300—600 Гц ступенями по 20 Гц. Задача испытуемого состояла в подборе значений частоты тона ( $f$ ) во втором стимуле, при котором оба стимула в паре максимально совпадали по фонетическому качеству.

На рис. 10.22 приведены зависимости среднего значения установленной частоты тона от  $F_1$ , полученные при  $F_0=80$  и 100 Гц. Можно видеть, что при  $F_0=100$  Гц зависимость имеет отчетливо ступенчатый характер, т. е. испытуемый устанавливает частоту тона не на частоте форманты, а на частоте гармоники, наиболее близкой к  $F_1$  гласного. При  $F_0=80$  Гц кривая зависимости приобретает более плавный характер, однако анализ гистограмм установленных значений  $f$  показывает, что в нижней области  $F_1$  (300—400 Гц) по-прежнему имеются выраженные максимумы на частотах гармоник.

Таким образом, приходится сделать вывод, что человек действительно выделяет в слуховом спектре гласного стимула значительно большее число неоднородностей, чем этого хотелось бы с точки зрения наибольшей простоты схемы фонемной классификации гласных.



## 10.6. ПРОБЛЕМЫ ФОНЕМНОЙ ИНТЕРПРЕТАЦИИ СЛУХОВОГО ОПИСАНИЯ ГЛАСНЫХ

Если бы гипотеза двухформантного описания гласных оправдалась, разработка модели фонемной классификации гласных, очевидно, не представляла бы формальных трудностей. К сожалению, гипотеза не оправдывается. Приходится считать, что число неоднородностей в спектре гласного, обнаруживаемых человеком, не только больше двух, но, что особенно неприятно, различно в разных гласных и зависит от конкретных особенностей стимула — таких, например, как значение его основной частоты.

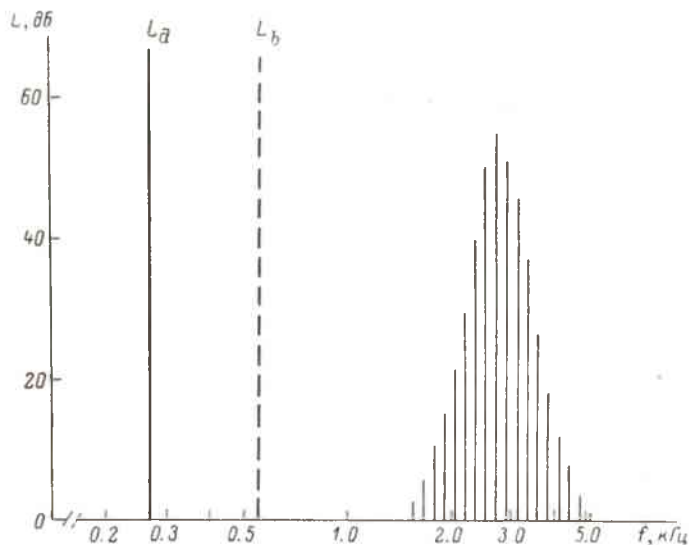


Рис. 10.23. Схематическое изображение спектра стимула, применявшегося в экспериментах по определению условий «равной значимости» нижних гармоник в стимуле. По [110].

Это значит, что мы никак не можем просто пронумеровать все неоднородности слева направо (как это делается с формантами) и считать соответствующие им значения  $z$  признаками гласного.

Экспериментальные данные, приведенные выше (раздел 10.5), заставляют думать, что человек учитывает что-то вроде «мощности» неоднородности, причем эта «мощность» связана с относительной величиной максимума на спектре. Вопрос о том, как оценивается «мощность», пока остается открытым.

Попытка найти экспериментальные подходы к решению вопроса была сделана в работах [106, 109, 110], где применялся гармонический сигнал, спектр которого схематически показан на рис. 10.23. Гармоника  $a$  всегда была ниже, а гармоника  $b$  — выше 400 Гц.

В зависимости от разности уровней гармоник ( $L_a, L_b$ ) стимул воспринимался как [i] (при  $L_a \gg L_b$ ) или как [e] (при  $L_a \ll L_b$ ). Задачей эксперимента было нахождение комбинаций  $L_a, L_b$ , соответствующих «равной значимости» гармоник (равная вероятность восприятия [i] и [e]), и построение частотного контура «равной значимости». Эксперимент проводился методом установившегося (испытуемый сам изменял  $L_a$  или  $L_b$ ); этим же методом определялись комбинации  $L_a, L_b$ , соответствующие равной громкости. Результаты показали, что контур «равной значимости» отличается от контура равной громкости в основном понижением «значимости» гармоники с повышением частоты. Не исключено, что это связано с угнетающим влиянием высокочастотной области спектра применявшегося стимула.

Выяснился также такой неприятный факт, что величина  $L_a - L_b$ , соответствующая «равной значимости», подвержена медленными флуктуациям, резко превышающим по величине размаха как флуктуации разности уровней при последовательных установках равной громкости, так и флуктуации порога обнаружения слабого спектрального максимума [110]. Это означает, что нужно искать какие-то другие экспериментальные методы исследования.

Необходимо заметить, что использование «мощности» для нумерации (ранжирования) неоднородностей кажется совместимым с экспериментальными данными лишь при таком дополнительном допущении, что шкала  $z$  разбита на два интервала (с границей около 1500 Гц) и что ранжирование неоднородностей осуществляется внутри этих интервалов. Необходимость этого допущения следует из двух фактов. Во-первых, очень значительные изменения относительной амплитуды первой форманты весьма мало влияют на восприятие таких гласных, как [i], [e], [y], [u] (см. раздел 10.3). Во-вторых, появление едва обнаруживаемой неоднородности на спектре звука в области выше 1500 Гц несовместимо с восприятием стимула как [u], [o] [108].

Итак, имеющиеся в настоящее время экспериментальные данные не оставляют, как кажется, сомнений в том, что обработка слухового спектра обеспечивает выделение неоднородностей на  $g(z)$  и что положение неоднородностей на шкале  $z$  является полезной информацией, используемой при распознавании фонем. Однако вопрос о том, по каким правилам происходит ранжирование неоднородностей, при каких условиях выделенная неоднородность определяет фонемное качество звука речи, а при каких относится, например, к индивидуальным характеристикам диктора, пока остается открытым.

## Глава 11

### СЛУХОВАЯ ОБРАБОТКА ОГИБАЮЩЕЙ СТИМУЛА

#### 11.1. ОБЩИЕ ПРЕДСТАВЛЕНИЯ О СЛУХОВОЙ ОБРАБОТКЕ ОГИБАЮЩЕЙ

Как показано в главе 8, плотность импульсации в  $i$ -том частотном канале периферической слуховой системы  $g_i(t)$  приближенно воспроизводит колебания на выходе  $i$ -того фильтра улитки, подвергнутые одно-полупериодному выпрямлению. При дальнейшей обработке сигнала на более высоких уровнях слуховой системы  $g_i(t)$  подвергается сглаживанию, благодаря чему выделяется огибающая сигнала (см. главу 9) в частотном канале.

Проблема обработки огибающей стимула рассматривалась в психоакустической литературе преимущественно с точки зрения характеристик сглаживания  $g_i(t)$  [280, 281]. Принимая, что  $g_i(t)$  пропускается через фильтр нижних частот, задача исследования формулируется как определение амплитудно-частотной характеристики этого фильтра или, если априорно допускается, что фильтр является инерционным звеном первого порядка, как определение его постоянной времени.

Легко заметить, что модель, состоящая только из фильтра нижних частот, не может воспроизвести способность человека обнаруживать изменения в огибающей стимула, например, обнаруживать наличие амплитудной модуляции. При сопоставлении экспериментальных данных с моделью молчаливо предполагается присутствие наблюдателя, который оценивает оциллограмму отклика фильтра нижних частот на  $g_i(t)$  и решает, присутствуют ли на ней колебания или неравномерности.

Естественно, что обработка  $g_i(t)$  в устройстве, моделирующем восприятие речи, должна быть полностью автоматической, т. е. модель должна включать не только сглаживание  $g_i(t)$ , но и извлечение из  $g_i(t)$  полезной информации.

Рассматривая психоакустические и электрофизиологические данные, можно выделить два типа полезных признаков  $g_i(t)$ ,

предполагающих, по всей вероятности, разные процедуры ее обработки.

Одна группа признаков связана с абсолютными значениями более или менее сильно сглаженной  $g_i(t)$ . Это «громкость» в частотной полосе. Предполагается, что суммирование по всем частотным полосам дает общую громкость звука (см. главу 10). Вопрос о том, сколько значений громкости приписывается при восприятии одному звуку — отрезку звукового потока — пока остается открытым. Допускаются значения громкости как в ее экстремальной точке, так и в моменты времени, привязанные к началу и окончанию звука [550]. Процедура измерения громкости пока остается неизвестной, однако несомненно, что она требует выделения особых точек на временной функции громкости и обеспечивает возможность запоминания на достаточно долгое время или самих выделенных значений, или каких-то результатов их преобразования. Последнее следует из способности человека сравнивать по громкости два звука, разделенных паузами значительной длительности.

В случае непрерывного амплитудно-модулированного сигнала восприятие его как меняющегося по громкости возможно лишь при низких частотах модуляции (не выше 10 Гц).

Вторая группа признаков, которая будет рассматриваться в настоящей главе, касается изменений огибающей  $g_i(t)$  во времени.

Существование нейронов, реагирующих только в моменты изменений огибающей  $g_i(t)$ , в частности в моменты начала и окончания звука, бесспорно доказано электрофизиологическими исследованиями (см. главу 9). Выделение человеком моментов начал и концов звуковых посылок также не вызывает сомнений (см. обсуждение в главах 3, 6).

Для выделения неравномерности в огибающей сигнала необходимо подвергнуть огибающую некоторой операции, аналогичной дифференцированию. Заметим, что это означает, что признаком является уже не амплитуда огибающей, а нечто близкое к скорости ее изменения. Однако выделить огибающую можно, лишь предварительно сгладив выпрямленный сигнал, т. е. отфильтровав несущую. Таким образом, система, выделяющая изменения огибающей, должна осуществлять нечто вроде полосовой фильтрации  $g_i(t)$ .

Для того чтобы система не только представляла изменения («производную») огибающей  $g_i(t)$  как самостоятельную переменную, но и обнаруживала бы неравномерности, необходимо допустить, что на выходе полосового фильтра находится пороговое устройство. При этом, естественно, возникает вопрос о том, реагирует ли система только на положительные значения «производной» или на значения обоих знаков, существен ли для нее только знак или важна также величина «производной», и т. д.

Психоакустические исследования восприятия изменений огибающей стимула находятся сейчас еще на самой начальной стадии, и о конструкции системы обработки огибающей можно пока

делать только очень грубые предположения. Примем, в качестве рабочей гипотезы, что в каждом частотном канале стоит своя система обработки (анализатор) огибающей, входным сигналом которой является  $g_i(t)$ . Состоит эта система из безынерционной нелинейности, одного или нескольких полосовых фильтров<sup>1</sup> и пороговых устройствах на их выходах. Рассмотрим, согласуются ли существующие данные с такой гипотезой.

### 11.1.1. СВИДЕТЕЛЬСТВА В ПОЛЬЗУ ДИФФЕРЕНЦИРОВАНИЯ

Для обнаружения факта изменения огибающей в стимуле, образованном из двух примыкающих друг к другу звуковых посылок с одинаковым спектром и разными уровнями интенсивности (рис. 11.1), можно в принципе использовать разные

процедуры обработки. Одна состоит в том, чтобы измерить абсолютные значения огибающей в моменты начала и конца стимула

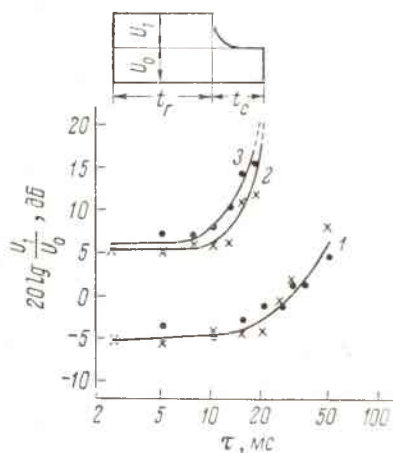


Рис. 11.1. Зависимость порогового перепада интенсивностей на соседних сегментах стимула от постоянной времени выключения. По [62].

По оси абсцисс — постоянная времени; по оси ординат — пороговый перепад интенсивностей. 1 — порог восприятия неопределенного согласного, 2 и 3 — пороги восприятия [ш] или [л]. Точки и крестики соответствуют разным испытуемым. В верхней части рисунка схематически показана огибающая применявшихся в эксперименте стимулов.

и проверить, совпадают ли эти значения. Эта процедура предполагается, когда говорится о сравнении по громкости. Другая процедура, основанная на полосовой фильтрации  $g_i(t)$ , позволяет обнаружить изменение огибающей как таковое. Существенно, что при этом обнаруживается не только факт изменения, но и момент времени, когда оно произошло.

<sup>1</sup> Гипотеза о том, что обработка огибающей осуществляется гребенкой полосовых фильтров, причем все фильтры имеют одинаковую абсолютную ширину полосы (6—8 Гц), была выдвинута Дубровским и Тумаркиной [67]. Авторы основывались на том обнаруженном ими факте, что в условиях модуляции сигнала суммой двух синусоидальных колебаний человек воспринимает «бienia» — изменения характера звука во времени — при одной и той же разности модулирующих частот, независимо от их абсолютных значений и от частот несущей. Однако этот факт может объясняться временными ограничениями процесса интерпретации звукового потока как последовательности (см. главу 6).



Если применяется полосовая фильтрация, естественно ожидать, что пороговые (едва обнаруживаемые) значения перепада в огибающей сигнала будут тем больше, чем меньше скорость изменения огибающей; в случае сравнения по громкости такой зависимости быть не должно.

В экспериментах [62, 214] использовался тот факт, что стимул с гласноподобным спектром и такой огибающей, как показано на рис. 11.1, воспринимается при достаточно большом  $U_1$  как сочетание ГС. В качестве несущей использовалась группа из 6 гармоник основной частоты 180 Гц; амплитуды гармоник были подобраны так, чтобы стимул максимально приближался к гласному [a]. Постоянными параметрами стимула были  $t_r=210$  мс,  $t_c=250$  мс и  $U_0$ , соответствующее уровню ощущения стимула 70 дБ. Огибающая стимула убывала от  $U_0+U_1$  до  $U_0$  по экспоненте. Экспериментатор задавал постоянную времени этой экспоненты  $\tau$ , испытуемый подбирал пороговое значение  $U_1$ . Измерения были сделаны для двух критериев. Критерий 1 соответствовал появлению любого согласного ([l], [v], [w]), критерий 2 — появлению [m] или [n].

Результаты эксперимента приведены на рис. 11.1. Можно видеть, что при малых значениях  $\tau$  устанавливаемое значение  $U_1$  не зависит от  $\tau$ . При больших значениях  $\tau$  увеличение  $\tau$  приводит к увеличению  $U_1$ . Такая зависимость явно говорит в пользу полосового фильтра. Из того что  $U_1$  начинает увеличиваться только при  $\tau > 10$  мс можно заключить, что сглаживание в этом случае довольно велико. Приведенные данные указывают также на то, что человек обнаруживает изменения огибающей отрицательного знака.

#### 11.1.2. СВИДЕТЕЛЬСТВА ПРОТИВ ОДНОГО ФИЛЬТРА

Предположение, что в обработке огибающей участвуют два различных механизма, вытекает из экспериментов Грина [282]. В этих экспериментах применялись послышки тона частотой 1, 2 и 4 кГц с огибающими, показанными в верхней части рис. 11.2. Один стимул обозначался как тихий-громкий (ТГ), другой — как громкий-тихий (ГТ).

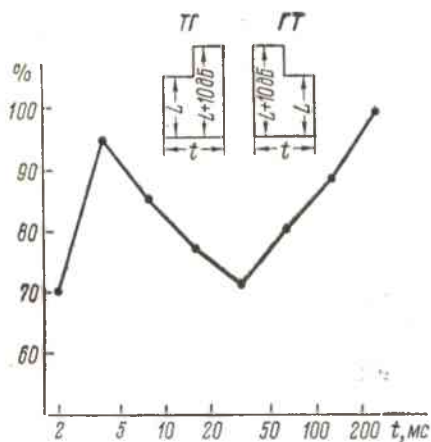
Стимулы предъявлялись парами со случайным порядком следования внутри пары. Испытуемый сообщал, какой из стимулов в паре ТГ, а какой — ГТ. Длительность стимулов  $t$  изменялась. Определялся процент правильных ответов.

Полученные данные показали, что для несущих 2 и 4 кГц пороговая длительность стимула (соответствующая 75% правильных ответов) лежит несколько ниже 2 мс. Для несущей 1 кГц она находится между 2 и 4 мс. Изменение уровня интенсивности стимула на 20 дБ не привело к изменению процента правильных ответов.



Данные для несущей 1 кГц приведены на рис. 11.2. Можно видеть, что начиная с 4 мс дальнейшее увеличение длительности стимула приводит не к улучшению, но к ухудшению различения. Вторичный подъем на психометрической функции начинается, когда длительность стимула становится больше 32 мс, т. е. когда длительность первого отрезка превышает 16 мс. Как было показано в главе 6, это соответствует появлению «согласного» — обнаружению неравномерности начала стимула.

При длительностях стимула, равных 2—4 мс, различение основывается на каких-то качественных особенностях звучания



воспринимаемого единичного щелчка. Для обнаружения этих качественных особенностей требуется значительная тренировка.

В работе [404] также исследовалось различение форм огибаю-

Рис. 11.2. Зависимость правильного различения сигналов  $TГ$  и  $ГТ$  от их длительности.

По оси абсцисс — длительность стимула; по оси ординат — процент правильных ответов. В верхней части рисунка показаны огибающие сигналов.

щей коротких посылок тона и шума. Использовались три формы огибающей: куполообразная и две треугольных. Сравнимые послышки имели одинаковую несущую частоту, равную энергии и равную пиковую интенсивность. Оказалось, что минимальная длительность посылок (определенная как длительность прямоугольной посылки, равной тестируемым по энергии и пиковой интенсивности), при которой возможно различение форм огибающей, составляет примерно 2 мс, как и по данным Грина [282].

Очень малые времена получаются и в экспериментах по восприятию двух звуковых импульсов. Так, эффект обратной маскировки (повышение порога обнаружения звукового импульса в случае, когда за ним следует более интенсивный импульс — маскер) наблюдается, если интервал между импульсами не превышает 10 мс [155, 183, 239, 437, 448]. В работе [448] показано, что минимальный интервал, при котором возможно различение пар импульсов с уровнями  $L_1L_2$  и  $L_2L_1$ , составляет 2 мс.

## 11.2. АМПЛИТУДНО-ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМЫ СЛУХОВОЙ ОБРАБОТКИ ОГИБАЮЩЕЙ

В том случае, если обработка огибающей осуществляется с помощью одного полосового фильтра, определение параметров этого фильтра не представляет трудностей. Достаточно определить экспериментальную зависимость порогового коэффициента синусоидальной амплитудной модуляции от частоты модуляции и рассматривать эту зависимость как амплитудно-частотную характеристику фильтра, полученную пороговым методом.

Однако, во-первых, не исключено, что на низких частотах модуляции человек обнаруживает модуляцию или путем слежения за громкостью, или, например, путем сравнения значений громкости через определенные интервалы времени. Наличие такого дополнительного механизма, естественно, не позволит обнаружить низкочастотный склон амплитудно-частотной характеристики фильтра. К интересующему нас фильтру будет относиться только высокочастотный склон зависимости порогового коэффициента модуляции от частоты.

Во-вторых, возможно, что система обработки огибающей состоит из нескольких полосовых фильтров. Тогда, измеряя зависимость порогового коэффициента модуляции от частоты, в лучшем случае можно получить только высокочастотный склон для самого высокочастотного из этих фильтров.

Допустим, что выходные сигналы разных фильтров анализатора огибающей соответствуют разным субъективным признакам сигнала (хрипlosti, [г]-образности). Тогда можно воспользоваться тем, что при изменении частоты модуляции характер восприятия отчетливо меняется. При средних частотах модуляция воспринимается как [г]-образный звук, при высоких — как хрипlostь в случае тональной несущей и как жужжание в случае шума. Следовательно, можно определить зависимости порогового коэффициента модуляции от частоты модуляции для разных критериев. Если наше допущение было верным, мы для разных критериев получим зависимости разной формы, соответствующие амплитудно-частотным характеристикам разных фильтров. Если же, например, в системе обработки огибающей стоит только один фильтр, а разные критерии соответствуют просто разным значениям порога на выходе, должны получиться зависимости одинаковой формы, но сдвинутые по оси коэффициентов модуляции.

Используя разные критерии, можно определить амплитудно-частотные характеристики и другим методом — измеряя коэффициент модуляции, соответствующий равной «мощности» определенного признака звука при разных частотах модуляции.

Заметим, что при определении амплитудно-частотных характеристик мы пренебрегаем нелинейностью системы обработки огибающей и считаем пороги, стоящие на выходах фильтров, постоянными.

### 11.2.1. АМПЛИТУДНО-ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ, ОПРЕДЕЛЕННЫЕ ПО ПОРОГАМ ОБНАРУЖЕНИЯ МОДУЛЯЦИИ ДЛЯ РАЗНЫХ КРИТЕРИЕВ

Наиболее удобным сигналом для исследования интересующих нас характеристик является синусоидально-модулированный белый шум. Дело в том, что в случае тональной несущей амплитудная модуляция может обнаруживаться за счет изменений в спектре сигнала.

В работе [145] были получены зависимости порогового коэффициента модуляции от частоты модуляции как для тонов, так и для шума. Все эти зависимости имеют минимум в районе 4 Гц, затем пороговые коэффициенты модуляции увеличиваются при повышении частоты. Для тональных несущих при достаточно высоких частотах модуляции пороговые коэффициенты модуляции

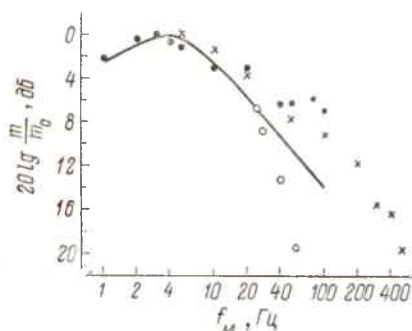


Рис. 11.3. Порог восприятия амплитудной модуляции шума и порог восприятия [г] как функция частоты модуляции.

По оси абсцисс — частота модуляции; по оси ординат — отношение порогового коэффициента модуляции к минимальному пороговому коэффициенту модуляции. Точки — порог восприятия амплитудной модуляции [145], крестики — порог восприятия амплитудной модуляции (данные И. Чистович и Мушников), кружки — порог восприятия [г], кривая — порог восприятия амплитудной модуляции [145]. Для порога восприятия [г] за  $t_0$  принята та же величина, что и для порога восприятия модуляции по данным И. Чистович и Мушников.

начинают опять убывать, для шума же зависимость оказывается монотонной. Объясняется это различие тем, что в случае тональной несущей при достаточно высоких частотах модуляции начинают восприниматься боковые составляющие спектра; в случае шума эти составляющие замаскированы.

Кривая [145] для шума (рис. 11.3) оказалась очень близкой к зависимости порога восприятия хриплости амплитудно-модулированного тона от частоты модуляции, полученной в работе [502]. Поскольку в работе [501] показано, что восприятие хриплости определяется только наличием флюктуаций огибающей в частотном канале, а спектральные изменения для восприятия хриплости незначительны, можно заключить, что так же обстоит дело и при восприятии амплитудной модуляции шума. Поэтому в дальнейшем мы будем пользоваться в основном данными, полученными на шумовых сигналах.

На рис. 11.3 показаны зависимости порогового коэффициента модуляции от частоты модуляции для белого шума, полученные

в работе [56] и в экспериментах И. Чистович и Мушникова. В работе [56] использовался шум с уровнем звукового давления 76 дБ, в экспериментах И. Чистович и Мушникова — с уровнем ощущения 40 дБ.

Из рис. 11.3 видно, что данные, полученные в этих работах, очень близки друг к другу и существенно отличаются от результатов [145], показанных на рис. 11.3 непрерывной кривой. Это различие объясняется, возможно, тем, что, как показано в работах [56, 105], наклон высокочастотной части зависимости порогового коэффициента модуляции от частоты уменьшается с тренировкой испытуемых; возможно также, что в работе [145] разные испытуемые пользовались разными критериями обнаружения модуляции.

В экспериментах И. Чистович и Мушникова была сделана попытка измерить на тех же самых испытуемых пороговые коэффициенты модуляции и для второго критерия — появления [г]. Испытуемые уверенно пользовались этим критерием, разброс данных оказался небольшим. Результаты показаны на рис. 11.3.

Из рис. 11.3 видно, что при использовании разных критериев получаются различные амплитудно-частотные характеристики. Для порога обнаружения модуляции спад характеристики в сторону высоких частот близок к 3 дБ/окт. Для порога обнаружения [г] спад характеристики в сторону высоких частот равен примерно 11 дБ/окт.

Следовательно, нельзя думать, что разные критерии соответствуют просто разным порогам на выходе фильтра системы обработки огибающей. Видимо, правильным было предположение, что в систему обработки огибающей входит несколько фильтров.

#### 11.2.2. АМПЛИТУДНО-ЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ, ПОЛУЧЕННЫЕ МЕТОДОМ УРАВНИВАНИЯ «МОЩНОСТИ» КРИТЕРИЯ

При определении амплитудно-частотных характеристик фильтров огибающей этим методом принимается, что каждому критерию соответствует свой фильтр и что стимулы равны по «мощности» данного критерия (по величине хриплости или [г]), если равны амплитуды выходных сигналов фильтра.

Данные по уравниванию величины хриплости были получены в работах [500, 514]. В этих работах использовалась следующая методика: сравнивались две посылки амплитудно-модулированного тона при одинаковой частоте несущей  $f_n$ ; частота модуляции первой посылки  $f_m$  задавалась экспериментатором, коэффициент модуляции  $m_1=1$ ; частота модуляции второй посылки выбиралась так, чтобы обеспечить максимальную хриплость (по данным [500]), частота модуляции, дающая максимальную хриплость, растет с увеличением  $f_n$  до 2 кГц, при  $f_n \geq 2$  кГц эта частота постоянна и равна 75 Гц). Коэффициент модуляции второй по-

сылки  $m_2$  устанавливался испытуемым так, чтобы посылки были равны по хрипlosti. На рис. 11.4 приведена зависимость  $20 \lg(1/m_2)$  от  $f_m$ , полученная в работе [514] для  $f_n = 2$  кГц. Заметим, что эта зависимость не является, строго говоря, амплитудно-частотной характеристикой «фильтра хрипlosti», так как субъективная величина хрипlosti нелинейно зависит от коэффициента модуляции  $m$  входного сигнала (по данным [500] величина хрипlosti пропорциональна  $m^2$ , по данным [514] —  $m^{1.5}$ ). Однако вносимые этой нелинейностью искажения невелики, и можно считать,

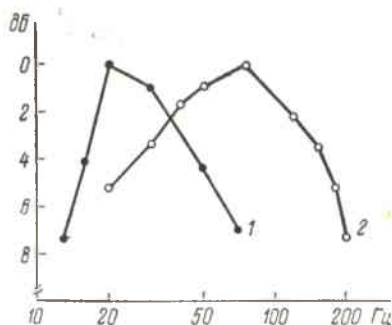


Рис. 11.4. Амплитудно-частотные характеристики «фильтра [г]» и «фильтра хрипlosti».

По оси абсцисс — частота модуляции; по оси ординат —  $20 \lg(m_2/m_{2min})$  для кривой 1,  $20 \lg(1/m_2)$  для кривой 2. Разные кривые получены разными методами: 1 — уравнивание «мощности» [г] (данные Жуковой, Столяровой и И. Чистович), 2 — уравнивание величины хрипlosti [514].

что зависимость с достаточной степенью приближения отражает амплитудно-частотную характеристику «фильтра хрипlosti».

Данные по уравниванию «мощности» [г] были получены в эксперименте Жуковой, Столяровой и И. Чистович. В этом эксперименте стимул представлял собой две посылки белого шума, разделенные паузой. Средняя часть каждой посылки модулировалась пятью периодами синусоиды. Частота модуляции первой посылки — 30 Гц, коэффициент модуляции  $m_1 = 0.4$ . Частота модуляции второй посылки  $f_m$  задавалась экспериментатором, испытуемый подбирает такой коэффициент модуляции второй посылки  $m_2$ , чтобы посылки были равны по «мощности» [г]. Результаты приведены на рис. 11.4. Из рисунка видно, что при использовании метода равной «мощности» критерия получаются, как и предполагалось, два полосовых фильтра, причем «фильтр [г]» оказывается существенно более низкочастотным, чем «фильтр хрипlosti».

### 11.2.3. ВЫХОДНЫЕ ПОРОГОВЫЕ УСТРОЙСТВА СИСТЕМЫ ОБРАБОТКИ ОГИБАЮЩЕЙ

При обсуждении данных по обнаружению амплитудной модуляции мы предполагали, что модуляция обнаруживается, если выходной сигнал некоторого полосового фильтра превосходит постоянный порог. Однако естественно ожидать, что значения порога случайно флуктуируют. В таком случае описанные выше зависимости порогового коэффициента модуляции от частоты будут отражать свойства не только фильтра, но и порогового



устройства. Если критерием обнаружения модуляции является хотя бы одно пересечение порога, то пороговый коэффициент модуляции должен уменьшаться при увеличении числа периодов модулирующего сигнала (так как вероятность пересечения будет возрастать за счет флуктуаций порога).

Зависимость порогового коэффициента модуляции от числа периодов модулирующей синусоиды была получена в работе [419]. Сигнал представлял собой посылку белого шума длительностью 800 мс. Начало и конец посылки были стационарными, середина модулировалась несколькими периодами синусоиды. Число периодов  $n$  и частота модуляции  $f_m$  задавались экспериментатором, испытуемый устанавливал такой коэффициент модуляции  $m$ , при котором он начинал слышать неравномерность в сигнале. Длительность первого стационарного участка стимула была равна 200 мс (для случая  $n=14$  и  $f_m=30$  Гц — 100 мс). Уровень ощущения сигнала был равен 60 дБ. На рис. 11.5 показаны полученные зависимости порогового коэффициента модуляции от числа периодов модулирующей синусоиды при постоянной частоте модуляции. Эти зависимости хорошо аппроксимируются прямыми с наклоном 3 дБ/окт. Следовательно, пороговый коэффициент модуляции обратно пропорционален корню квадратному из числа периодов при любой фиксированной частоте модуляции. Зависимость эта не меняет характера даже при очень больших длительностях  $T$  модулированного участка сигнала ( $T=460$  мс при  $n=14$ ,  $f_m=30$  Гц).

Вообще говоря, пороговый коэффициент модуляции может уменьшаться при увеличении числа периодов и за счет переходного процесса в фильтре. Однако из данных, приведенных в работе [52], следует, что длительность переходного процесса не должна превышать 100 мс. В работе [52] сигнал представлял собой посылку белого шума длительностью 600 мс. Внутри посылки (через 200 мс от ее начала) помещались две паузы. Длительность одной из пауз постоянна,  $t_0=1.6$  мс. Экспериментатор задавал интервал между паузами ( $t$ ). Испытуемый должен был установить такую длительность другой паузы ( $t_1$ ), при которой он начинал слышать неравномерность в сигнале. Оказалось, что при увеличении  $t$  примерно до 25 мс  $t_1$  растет, затем при  $t=35$  мс наблюдается ее небольшой минимум, при  $t > 100$  мс  $t_1$  остается постоянной.

Следовательно, уменьшение порогового коэффициента модуляции в работе [419] не может быть вызвано только переходным процессом в фильтре, так как это уменьшение наблюдается до  $T=460$  мс. Видимо, действительно, вероятность превышения порога выходным сигналом фильтра увеличивается при увеличении числа периодов модулирующего сигнала либо из-за случайного характера использованного в экспериментах сигнала (белый шум), либо из-за флуктуаций порога.

На рис. 11.6 показаны зависимости порогового коэффициента модуляции от частоты при постоянном числе периодов модулирующей



щей синусоиды [419]. Можно видеть, что эти зависимости неплохо аппроксимируются прямыми с наклоном 6 дБ/окт. Из рис. 11.5 и 11.6 следует, что весь набор полученных в работе [419] данных описывается формулой  $m = a f_m / \sqrt{n}$ , где  $a = \text{const}$ .

Если бы зависимость порогового коэффициента модуляции от частоты измерялась при постоянной длительности  $T$  модулированного участка вместо постоянного числа периодов, то эта формула должна была бы иметь вид  $m = a \sqrt{f_m / T}$ , т. е. наклон зависимости в сторону высоких частот должен был бы быть равен 3 дБ/окт. Заметим, что наклон зависимостей порогового коэффициента модуляции от частоты для случая длительной модуляции действительно близок к 3 дБ/окт (см. рис. 11.3). Таким образом, данные [419] не противоречат другим данным по пороговым коэффициентам модуляции, если допустить, что система обработки огибающей включает в себя флюктуирующие пороги.

Интересно сравнить полученные в работе [419] зависимости порогового коэффициента модуляции от числа периодов с данными Фогеля [514] для надпороговых величин хриплости. В эксперименте Фогеля уравнивались по хриплости два сигнала с одинаковой частотой несущей (2 кГц), одинаковой частотой модуляции ( $f_m$ ) и примерно одинаковой громкостью. Разница между сигналами заключалась в том, что стандарт имел значительную длительность, соответствующую не менее чем 20 периодам модуляции, а длительность теста изменялась так, чтобы он содержал разное (целое) число периодов модуляции ( $n_t$ ). Коэффициент модуляции теста был постоянным ( $m_t = 0.8$ ). Испытуемый подбирал такой коэффициент модуляции стандарта  $m_c$ , чтобы тест и стандарт были равны по хриплости.

Результаты показаны на рис. 11.7. Можно видеть, что  $m_c$  практически не зависит от  $n_t$ , слабая зависимость наблюдается только при  $f_m = 150$  Гц. Таким образом, если предположение о существовании флюктуирующих порогов и верно, эти флюктуации не сказываются при измерении надпороговых величин хриплости.

Перейдем теперь к вопросу о том, сколько порогов стоит на выходе каждого фильтра анализатора огибающей. Поскольку человек обнаруживает как возрастание, так и убывание огибающей сигнала, естественно допустить, что выходные сигналы фильтров сравниваются с двумя порогами — положительным и отрицательным. Проверить это допущение можно, исследуя восприятие сигналов, модулированных положительным и отрицательным полупериодами синусоиды. Если в системе обработки огибающей стоят пороги только одного знака, то сигналы, модулированные положительным и отрицательным полупериодами, должны либо вообще не различаться человеком, либо различаться только по «мощности» неравномерности в сигнале; значения коэффициента модуляции, соответствующие порогу обнаружения неравномерности в сиг-

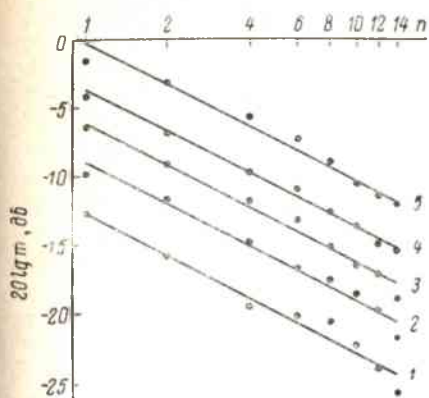


Рис. 11.5. Зависимость порога восприятия амплитудной модуляции от числа периодов модулирующей синусоиды. По [419].

По оси абсцисс — число периодов; по оси ординат — пороговый коэффициент модуляции. Параметр — частота модуляции: 1 — 30, 2 — 50, 3 — 70, 4 — 100, 5 — 150 Гц.

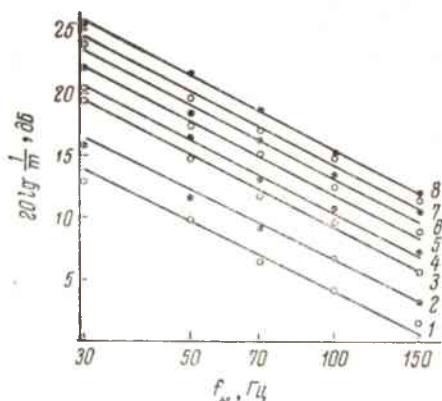


Рис. 11.6. Зависимость порога восприятия амплитудной модуляции от частоты модуляции. По [419].

По оси абсцисс — частота модуляции; по оси ординат — величина, обратная пороговому коэффициенту модуляции. Параметр — число периодов модулирующей синусоиды: 1 — 1, 2 — 2, 3 — 4, 4 — 6, 5 — 8, 6 — 10, 7 — 12, 8 — 14.

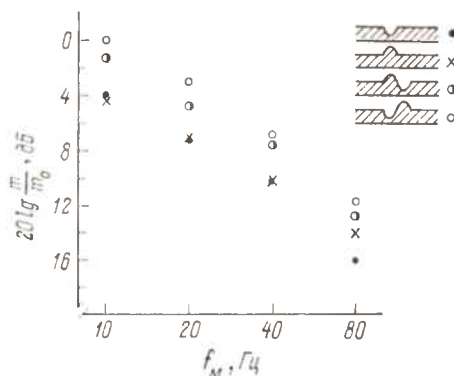
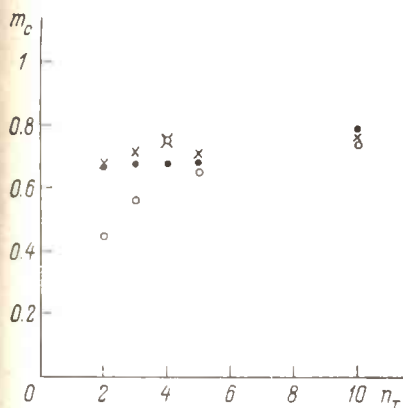


Рис. 11.7. Коэффициент модуляции сигнала-стандарта как функция от числа периодов ( $n_T$ ) синусоиды, модулирующей сигнал-тест. По [614].

По оси абсцисс — число периодов; по оси ординат — коэффициент модуляции стандарта, соответствующий равной хрипlosti теста и стандарта. Параметр — частота модуляции: точки — 30, крестики — 100, кружки — 150 Гц.

Рис. 11.8. Порог восприятия амплитудной модуляции как функция частоты модулирующей синусоиды.

По оси абсцисс — частота модулирующей синусоиды; по оси ординат — отношение порогового коэффициента модуляции к минимальному пороговому коэффициенту модуляции. Обгибающие стимулов показаны в верхней части рисунка.

нале, должны в таком случае оказаться существенно различными для положительных и отрицательных полупериодов.

Восприятие сигналов с такими огибающими исследовалось в экспериментах И. Чистович и Мушников. Стимул в этих экспериментах представлял собой непрерывный белый шум. Использовались четыре варианта огибающей, показанные на рис. 11.8 вверху. Переменная составляющая огибающей подавалась периодически, примерно раз в 1.5 с. Специальный эксперимент показал, что испытуемые безошибочно различают стимулы, модулированные положительным и отрицательным полупериодами (при коэффициенте модуляции  $m=0.8$ ). В основном эксперименте были получены зависимости порогового коэффициента модуляции от частоты модуляции для всех четырех форм огибающей. Экспериментатор задавал частоту модулирующей синусоиды, испытуемый устанавливал такой коэффициент модуляции, при котором он начинал слышать неравномерность в сигнале.

На рис. 11.8 приведены результаты, полученные при уровне ощущения шума  $L=40$  дБ. Пороговый коэффициент модуляции представлен в дБ от его минимального значения ( $m_0=0.138$  при  $f_m=10$  Гц).

Из рис. 11.8 видно, что при  $f_m < 80$  Гц пороговые коэффициенты модуляции для положительного и отрицательного полупериодов практически одинаковы. Это подтверждает предположение, что на выходах фильтров огибающей стоят и положительные, и отрицательные пороги, причем равные по абсолютной величине.

Из рис. 11.8 видно также, что пороговый коэффициент модуляции для половины периода оказался примерно на 3 дБ больше, чем для целого периода. Этот факт, как и приведенные выше данные [419], объясняется, вероятно, случайными флюктуациями порогов.

### 11.3. СВЯЗЬ АНАЛИЗАТОРА ОГИБАЮЩЕЙ С ЧАСТОТНЫМ КАНАЛОМ ПЕРИФЕРИЧЕСКОЙ СЛУХОВОЙ СИСТЕМЫ

В разделе 11.1 было высказано предположение, что существует набор анализаторов огибающей, причем каждый из них производит обработку информации в локальном участке спектра, т. е. входным сигналом  $i$ -того анализатора является  $g_i(t)$  — плотность импульсации в  $i$ -том частотном канале периферической слуховой системы. Это предположение основано на электрофизиологических данных (см. главу 9), а также на том факте, что восприятие хриплости определяется флюктуациями огибающей сигнала в частотном канале [501].

В таком случае на восприятии амплитудно-модулированных сигналов должно сказываться сглаживание этих сигналов фильтрами улитки. Естественно ожидать, что влияние фильтров улитки будет тем сильнее, чем ниже частота несущей, так как ширина

полосы фильтров улитки уменьшается с понижением характеристической (резонансной) частоты (см. главу 7). При достаточно высоких частотах несущей, когда ширина полосы фильтров улитки существенно превышает частоту среза фильтров анализатора огибающей, восприятие амплитудно-модулированных сигналов должно определяться только характеристиками фильтров анализатора огибающей. Следовательно, в тех случаях, когда сигнал обрабатывается «фильтром хрипlosti», влияние фильтров улитки должно распространяться на более высокие частоты несущей, чем для «фильтра [г]».

Результаты психоакустических экспериментов по восприятию амплитудно-модулированных сигналов вполне согласуются с приведенными рассуждениями. Исследование восприятия хрипlosti синусоидально-модулированных тонов показало, что как частота модуляции, соответствующая максимальной хрипlosti сигнала, так и максимальная частота модуляции, при которой еще воспринимается хриплость, увеличиваются с увеличением частоты несущей до 2 кГц [500, 501].

С другой стороны, в главе 6 показано, что минимальный интервал времени между началом стимула и ступенчатым приращением, при котором воспринимается неравномерность начала стимула (или начальный согласный), зависит от частоты несущей всего до 500 Гц.

В работе [62] исследовалась зависимость длительности минимальной воспринимаемой паузы от частоты несущей. Стимулами являлись посылки тона длительностью 600 мс с уровнем ощущения 40 дБ. Пауза переменной длительности находилась примерно на середине посылки. Испытуемые использовали следующий критерий обнаружения паузы: пауза считалась существующей, если смещение ее внутри стимула приводило к перемещению удара с одного отрезка звука на другой.

Испытуемые отмечали, что эффект перемещения удара наблюдался тогда, когда пауза воспринималась как согласный [г]. При исчезновении [г] сохранялось восприятие щелчка, действующего на фоне непрерывного тона (в дальнейшем мы будем называть паузу, воспринимаемую как [г], [г]-паузой). Полученные данные приведены на рис. 11.9. Можно видеть, что значение пороговой паузы отчетливо возрастает на частотах ниже 500 Гц.

Тот факт, что [г]-пауза обнаруживается только в случае существования паузы в сигнале в частотном канале (а не включения сигнала в одном канале и выключения в другом), демонстрируется экспериментом Лесогор. В этом эксперименте сигнал представлял собой последовательность из двух тональных посылок длительностью по 300 мс, разделенных паузой. Частота первой посылки  $f_0$  задавалась экспериментатором, задачей испытуемого было установить нижнюю ( $f_n$ ) и верхнюю ( $f_v$ ) граничные частоты второй посылки, при которых [г]-пауза еще воспринимается. Длительность паузы испытуемый мог регулировать.

На рис. 11.10 показаны полученные зависимости  $f_v - f_0$  и  $f_0 - f_n$  от  $f_0$ . Для сравнения приведена величина критической полосы [145]. Можно видеть, что [г]-пауза воспринимается только в том случае, если разность частот первой и второй посылки не превышает величины критической полосы.

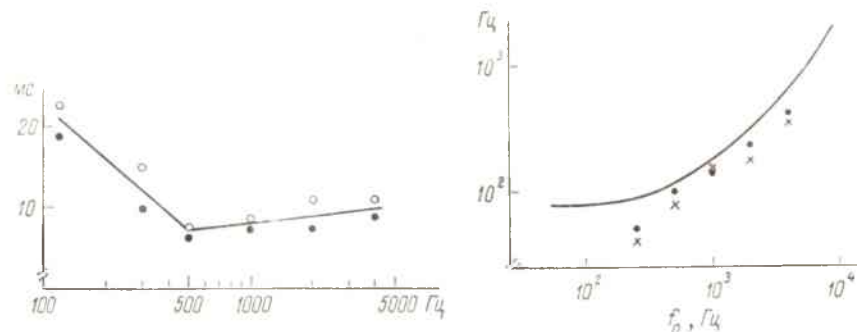


Рис. 11.9. Зависимость минимальной длительности [г]-паузы в тональном сигнале от частоты тона. По [62].

По оси абсцисс — частота тона; по оси ординат — минимальная длительность [г]-паузы. Кружки и точки соответствуют разным испытуемым.

Рис. 11.10. Восприятие [г]-паузы в стимуле, образованном из двух тональных посылок с разными частотами.

По оси абсцисс — частота первой посылки; по оси ординат — абсолютная величина разности частот первой и второй посылок. Точки —  $(f_v - f_0)$ , крестики —  $(f_0 - f_n)$ . Кривая — зависимость от частоты ширины критической полосы [145]. Остальные обозначения см. в тексте.

#### 11.4. ВХОДНАЯ НЕЛИНЕЙНОСТЬ СИСТЕМЫ ОБРАБОТКИ ОГИБАЮЩЕЙ

Стремясь к тому, чтобы устройство, моделирующее обработку огибающей в частотном канале слуховой системы, было максимально простым, естественно предположить, что фильтры анализатора огибающей линейны. Логично также принять допущение, что пороговые устройства на выходе фильтров анализатора срабатывают при некотором фиксированном значении выходного сигнала фильтра, не зависящем от уровня интенсивности входного звукового сигнала. В таком случае появляется возможность определить амплитудную характеристику преобразователя, расположенного на входе анализатора огибающей, воспользовавшись для этого экспериментальной зависимостью порогового коэффициента модуляции от уровня ощущения стимула.

Если фильтры огибающей являются полосовыми (не пропускают постоянной составляющей), а амплитудная характеристика преобразователя линейна, то амплитуда сигнала на выходе фильтра будет пропорциональна амплитуде переменной составляющей модулирующего сигнала  $A_m$  и не будет зависеть от величины его постоянной составляющей  $A_c$  (т. е. от уровня ощущения стимула,



так как он обычно определяется при  $A_m=0$ ). Это значит, что для получения пороговой амплитуды сигнала на выходе фильтра (для обнаружения модуляции) при любых значениях  $A_c$  потребуется одно и то же значение  $A_m$ , т. е. пороговый коэффициент модуляции  $m=A_m/A_c$  окажется обратно пропорциональным  $A_c$ .

Если преобразователь обладает нелинейной (по типу компрессии) амплитудной характеристикой, то пороговое значение  $A_m$  будет возрастать с увеличением  $A_c$ . В случае логарифмической характеристики отношение  $A_m/A_c$  будет постоянным, т. е. пороговый коэффициент модуляции не будет зависеть от уровня ощущения стимула.

Трудность использования экспериментальных данных для определения амплитудной характеристики преобразователя связана с тем, что для чистых тонов и для белого шума наблюдаются существенно различные зависимости. Это заставляет думать, что по крайней мере в одном из этих случаев экспериментальная зависимость отражает, кроме характеристики преобразователя, еще какой-то дополнительный эффект и соответственно не является «истинной».

Различие проявляется как в экспериментах по определению порогового коэффициента модуляции (при низких частотах модуляции), так и в экспериментах по определению дифференциального порога по интенсивности (сравнение двух стимулов, разделенных паузой). Для белого шума пороговый коэффициент модуляции (относительный дифференциальный порог) уменьшается при увеличении уровня ощущения стимула примерно до 20 дБ, при больших уровнях пороговый коэффициент модуляции сохраняется постоянным. Для чистого тона уменьшение порогового коэффициента модуляции (относительного дифференциального порога) наблюдается при увеличении уровня ощущения стимула во всем диапазоне уровней [145, 379, 397, 513].

Предположение, что осложненной дополнительными эффектами является зависимость, полученная для шума, исходило из того факта, что на огибающей шума имеются случайные флуктуации, мощность которых возрастает с увеличением уровня интенсивности [379, 380]. Допустим, что входной преобразователь обладает линейной амплитудной характеристикой. Тогда для эффективного различения модулированного и немодулированного шумов необходимо изменять значение порога на выходе фильтра соответственно уровню ощущения стимула. Такая подстройка не нужна в случае тона.

Экспериментальное подтверждение получила за последние годы та точка зрения, что «осложненной» является, наоборот, зависимость, полученная для тональных сигналов [197, 397, 513, 539, 542]. Существо дела сводится к тому, что, обнаруживая амплитудную модуляцию тона (при низких частотах модуляции), человек ориентируется на изменения в тембре звука, т. е. основывается не только на изменениях огибающей  $g_s(t)$  в частотном канале,

настроенном на данный тон, но и на изменении формы  $g(z)$  — распределения плотности импульсации по оси частот. Эти изменения  $g(z)$  обусловлены как нелинейностью фильтров улитки (см. главу 7), так и, при больших уровнях стимула, нелинейными искажениями сигнала в более периферических звеньях системы.

Из сказанного следует, что и шум, и тон не являются идеальными стимулами для определения амплитудных характеристик преобразователя. Требования к идеальному стимулу состоят в том, чтобы он вызывал изменения огибающей  $g(t)$  только в одном частотном канале и чтобы случайные флуктуации огибающей отсутствовали.

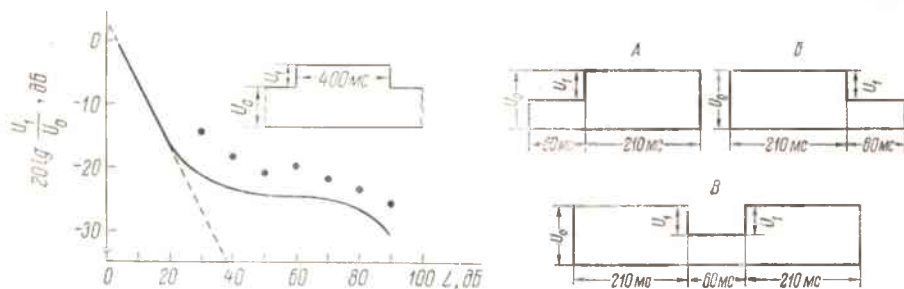


Рис. 11.11. Порог обнаружения перегулярности в стимуле как функция уровня ощущения стимула.

По оси абсцисс — уровень ощущения стимула; по оси ординат — пороговое изменение огибающей стимула. Сплошная кривая соответствует стимулу с огибающей, показанной в верхней части рисунка [204]; точки — стимулу В рис. 11.12; штриховая прямая соответствует условию  $U_1 = \text{const}$ .

Рис. 11.12. Огибающие стимулов.

Обозначения см. в тексте.

Удовлетворяющий этим требованиям стимул был применен в эксперименте Кэмпбелла и Лэски по определению порога обнаружения приращения огибающей [204].

Стимул был образован из трех составляющих: 700, 990 и 1300 Гц. Уровень тона 990 Гц был на 3 дБ выше, чем уровень остальных тонов; амплитуда этого тона увеличивалась ступенькой на величину  $U_1$  (рис. 11.11). Независимой переменной являлась величина пьедестала  $U_0$ ; определялось значение  $U_1$ , соответствующее порогу обнаружения неравномерности в стимуле. Результаты эксперимента показаны на рис. 11.11 непрерывной кривой. Из рис. 11.11 видно, что в области уровней ощущения примерно до 20 дБ экспериментальная зависимость совпадает с прямой  $20 \lg (U_1/U_0) = a - 20 \lg U_0$ , где  $a = \text{const}$ . Эта прямая соответствует условию  $U_1 = \text{const}$ . Как говорилось выше, такая зависимость должна наблюдаться, если амплитудная характеристика входного преобразователя является линейной.

При увеличении уровня ощущения (в пределах до 70 дБ) экспериментальная зависимость стремится к некоторому постоянному значению  $20 \lg (U_1/U_0)$ . Следовательно, в этом диапазоне характеристика преобразователя близка к логарифмической, что соответствует данным, полученным ранее для шума. При уровне ощущения более 70 дБ значение  $20 \lg (U_1/U_0)$  снова уменьшается.

Стимулы со сложным спектром использовались также в экспериментах Жуковой. В этих экспериментах исследовались зависимости порогов обнаружения перепадов амплитуды огибающей от уровня ощущения стимула для стимулов с тремя типами огибающей, схематически показанными на рис. 11.12. Спектр стимулов

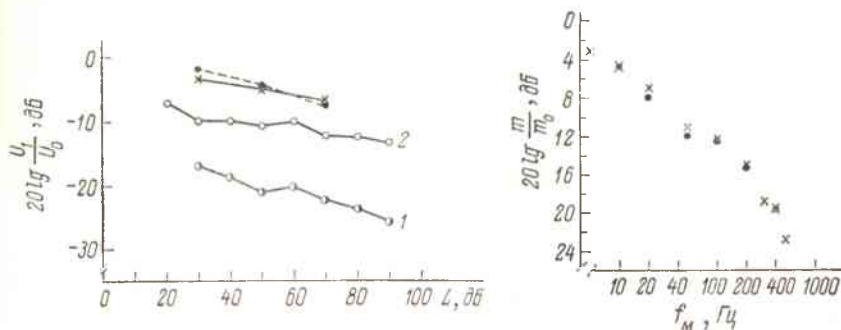


Рис. 11.13. Порог обнаружения нерегулярности в стимуле и порог восприятия согласного как функция уровня ощущения для стимулов рис. 11.12.

По оси абсцисс — уровень ощущения сегмента стимула с большей интенсивностью; по оси ординат — пороговая величина изменения в огибающей стимула. 1 — порог обнаружения нерегулярности в стимуле В, 2 — порог восприятия согласного в стимуле В, точки — порог восприятия согласного в стимуле В, крестики — порог восприятия согласного в стимуле А.

Рис. 11.14. Зависимость порога восприятия модуляции шума от частоты модуляции.

По оси абсцисс — частота модуляции; по оси ординат — отношение порогового коэффициента модуляции к минимальному пороговому коэффициенту модуляции ( $m_0=0.055$ ). Крестики соответствуют уровню ощущения стимула 40 дБ, точки — 15 дБ. Зависимость для уровня ощущения 40 дБ сдвинута на 3.2 дБ вниз по оси ординат.

состоял из первых шести гармоник основной частоты 180 Гц, амплитуды гармоник были подобраны так, чтобы стимул максимально приближался к гласному [а]. Величина  $U_0$  задавалась экспериментатором, испытуемый мог изменять  $U_1$ .

Для подробного исследования зависимости  $U_1$  от  $U_0$  использовались стимулы типа В. Определялись значения  $U_1$ , соответствующие порогу обнаружения неравномерности в гласном и порогу появления согласного. Для стимулов типа А и В определялся только порог появления согласного.

На рис. 11.11 точками показаны полученные Жуковой пороги обнаружения неравномерности в гласном. Можно видеть, что точки практически идеально воспроизводят кривую Кэмпбелла

и Лэски, если ее сдвинуть на 4.5 дБ вверх. Необходимость сдвига возникает, вероятно, за счет того, что в эксперименте Жуковой длительность сегмента с измененной амплитудой составляла 60 мс, а в эксперименте Кэмпбелла и Лэски — 400 мс. На рис. 11.13 приведены полученные в работе Жуковой значения порога восприятия согласного для стимулов типа *А*, *Б* и *В* (рис. 11.12). Можно видеть, что порог восприятия согласного в стимуле *В* (кривая 2) примерно на 10 дБ выше порога обнаружения нерегулярности в гласном для этого же стимула (кривая 1). Вместе с тем вид зависимости оказывается достаточно близким.

Различий в порогах восприятия согласного для стимулов *А* и *Б* не обнаруживается при уровнях ощущения, равных 50—70 дБ. При уровне 30 дБ порог достоверно выше для стимула *Б* (отрицательное приращение), однако разница мала по абсолютной величине. Существенно, что пороги для стимулов *А* и *Б* оказываются значительно большими, чем пороги восприятия согласного в стимуле *В*. Возможно, что этот эффект объясняется тем, что для восприятия согласного в стимуле *В* достаточно обнаружения хотя бы одного из двух явлений — отрицательного или положительного скачка огибающей (см. раздел 11.2.3).

Вывод о том, что в области средних уровней интенсивности амплитудная характеристика преобразователя должна быть близка к логарифмической, следует также из экспериментов по сравнению хрипlosti двух посылок амплитудно-модулированного тона с сильно различающимися (на 20 и 40 дБ) уровнями интенсивности [502]. Подчеркнем, что в данном случае применялись большие значения коэффициента модуляции и, согласно принятой интерпретации, сравнивались существенно надпороговые сигналы на выходе «фильтра хрипlosti». Методика эксперимента состояла в следующем. Посылки предъявлялись парами, уровни звукового давления посылок в паре различались на 20 и 40 дБ. Коэффициент модуляции тона с большим уровнем задавался случайным образом в пределах от 0.2 до 1, коэффициент модуляции тона с меньшим уровнем был равен 1. Испытуемые определяли, первая или вторая посылка звучит более хрипло. Из распределения ответов были определены коэффициенты модуляции тонов большей интенсивности, дающие такую же хрипlostь, как более слабые тоны с  $m=1$ . Частота модуляции была равна 70 Гц, частота несущей — 1 и 4 кГц. Оказалось, что результаты существенно зависят от порядка предъявления посылок, но даже в худшем случае различие в уровне сигналов на 40 дБ соответствует различию коэффициентов модуляции не более чем на 6 дБ (в лучшем случае — на 1.4 дБ).

Так как система обработки огибающей заведомо включает нелинейные преобразования и принятое пока допущение о линейности фильтров является условным, более исчерпывающей характеристикой системы было бы семейство амплитудно-частотных характеристик фильтров огибающей, снятых при разных уровнях стимула.

В экспериментах И. Чистович и Мушниковы были получены зависимости порогового коэффициента модуляции от частоты для уровней ощущения шума  $L=40$  и  $15$  дБ. Результаты приведены на рис. 11.14 в децибелах от минимального коэффициента модуляции при  $L=40$  дБ ( $m_0=0.055$  при  $f_m=5$  Гц), причем зависимость для  $L=40$  дБ сдвинута на 3.2 дБ вниз по оси ординат. Легко видеть, что зависимости совпадают практически идеально.

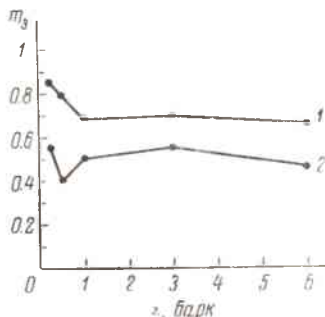
### 11.5. ОБЪЕДИНЕНИЕ ИНФОРМАЦИИ ОБ ОГИБАЮЩЕЙ В ЧАСТОТНЫХ КАНАЛАХ

Вопрос о том, как взаимодействуют системы обработки огибающей в разных частотных каналах, пока очень мало изучен. Очевидно, происходит нечто вроде суммирования сигналов, отражающих неравномерности в огибающей в частотных каналах.

Это следует из данных по пороговым [г]-паузам (см. 11.3). Величина пороговой [г]-паузы оказалась больше для чистых тонов, чем для гласного [а] и для белого шума [62]. Ясно, что в слу-

Рис. 11.15. Коэффициент модуляции ( $m_3$ ) тона, равного по хрипlosti сумме двух тестирующих амплитудно-модулированных тонов. По [502].

По оси абсцисс — расстояние между тестирующими тонами по шкале Барк; по оси ординат — коэффициент модуляции. 1 — модулирующие сигналы в фазе, 2 — в противофазе.



чае гласного и шума неравномерность в огибающей имеется в большем числе каналов, чем в случае тона. Кроме того, величина пороговой паузы несколько увеличивается при увеличении частоты несущей выше 500 Гц (рис. 11.9). Это может объясняться тем, что при более высоких несущих действует меньшее число каналов.

Необходимость объединения информации по частотным каналам подтверждается и данными по восприятию хрипlosti.

Для изучения взаимодействия частотных каналов при определении хрипlosti Терхардт [502] провел эксперимент с двумя амплитудно-модулированными тонами. Сигнал представлял собой сумму двух синусоидально модулированных тонов с частотами  $f_1$  и  $f_2$ . В качестве стандарта применялся тон с частотой  $f_3$  и коэффициентом модуляции  $m_3$ . Частота модуляции всех трех тонов составляла 40 Гц. Коэффициенты модуляции тонов  $f_1$  и  $f_2$  подбирались так, чтобы хриплость каждого из этих тонов, предъявленного отдельно, была равна хрипlosti стандарта с коэффициентом



модуляции, равным 0.5. Затем испытуемому предъявлялась сумма тонов  $f_1$  и  $f_2$  с подобранными таким образом равными хрипlostями; коэффициент модуляции  $m_3$  стандарта подбирался так, чтобы хрипlostь стандарта была равна хрипlostи суммы двух тонов.

На рис. 11.15 показаны полученные зависимости  $m_3$  от расстояния по частоте между  $f_1$  и  $f_2$  в Барках для случаев, когда сигналы, модулирующие тоны комплекса, подавались в фазе и в противофазе. Из рисунка видно, что, если расстояние между частотами комплекса превышает критическую полосу, хрипlostь суммы тонов равна хрипlostи каждого тона в отдельности, когда модулирующие сигналы подаются в противофазе, и хрипlostь суммы больше хрипlostи одного тона, если модулирующие сигналы в фазе. Отсюда следует, что при определении общей хрипlostи некоторым образом складываются отклонения одного знака в каждом канале.

#### 11.6. ФУНКЦИОНАЛЬНАЯ МОДЕЛЬ ОЦЕНКИ ХРИПЛОСТИ

Построить полную функциональную модель системы обработки огибающей пока не представляется возможным из-за недостатка экспериментальных данных. Остается неясным, сколько фильтров (два или больше) должна содержать модель каждого частотного канала, не полностью известны параметры этих фильт-

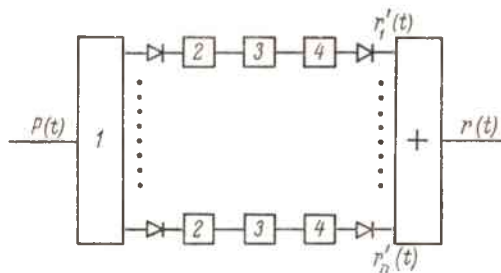


Рис. 11.16. Блок-схема модели оценки хрипlosti амплитудно-модулированного сигнала.

Обозначения см. в тексте.

ров. Чрезвычайно мало известно об объединении информации по каналам.

Первой попыткой смоделировать часть системы обработки огибающей является предложенная Фогелем [514] функциональная модель оценки хрипlosti амплитудно-модулированного сигнала, блок-схема которой показана на рис. 11.16. В модели блок 1 представляет собой набор 96 параллельных линейных LC-фильтров, моделирующих фильтры улитки. Полоса пропускания каждого фильтра примерно равна 1 Барк. Спад амплитудно-частотной

характеристики в сторону высоких частот крутой, в сторону низких — более пологий. Координатная характеристика отражается (приближенно) неравнобедренным треугольником. Набор из 96 фильтров перекрывает всю область от 0 до 24 Барк таким образом, что расстояние между частотами максимумов соседних фильтров равняется примерно 0.25 Барк. В каждом канале сигнал подвергается однополупериодному выпрямлению и затем сглаживается интегрирующей цепочкой с постоянной времени 1.4 мс (блок 2). Блок 3 осуществляет логарифмирование сигнала. Блок 4 представляет собой полосовой фильтр, частота максимума амплитудно-частотной характеристики которого находится в районе 90 Гц, а наклоны амплитудно-частотной характеристики в сторону низких и высоких частот равны соответственно 12 и 6 дБ/окт. Выходной сигнал полосового фильтра подается на однополупериодный выпрямитель. Выпрямленные сигналы  $r'_i(t)$  суммируются по всем каналам. Величина хрипlosti оценивается по максимуму выходной величины  $r(t)$ .

#### 11.7. РАЗЛИЧЕНИЕ ФОРМ НЕРАВНОМЕРНОСТЕЙ

Результаты, приведенные в данной главе и в главе 6, показывают, что человек не только обнаруживает однократные неравномерности в огибающей сигнала, но и разделяет их на некоторые классы форм. Так, положительные и отрицательные ступенчатые приращения и [r]-пауза относятся, видимо, к трем разным классам. Пока совершенно неясно, по каким признакам происходит эта классификация. Однако есть данные, указывающие на то, что важную роль здесь играет скорость изменения огибающей стимула.

Исследование различения шумовых сигналов с разным передним фронтом проводилось в работе Поллака [427] и Гершуни и др. [41]. Результаты, полученные в обеих работах, являются достаточно близкими. Очень интересными для анализа являются данные Гершуни и др. Авторы применяли стимулы с экспоненциальным нарастанием (значение  $\tau$  изменялось) и провели два рода экспериментов. В одном из них от испытуемых требовалось, используя фонетические ассоциации, осуществлять классификацию стимулов. Выяснилось, что испытуемые выделяют три класса стимулов, соответствующие резкому началу, без резкого начала и плавному нарастанию. Полученные функции идентификации приведены на графике Б рис. 11.17.

Далее производилось измерение дифференциального порога по  $\tau$ . Значение  $\tau_1$  (стандарта) было фиксированным, определялось значение  $\tau_2$  переменного стимула ( $\tau_2 > \tau_1$ ), при котором испытуемый различал стимулы в 75% случаев. Индивидуальные данные трех испытуемых, участвовавших в экспериментах, приведены на графике А рис. 11.17.

Можно видеть, что зависимость  $\tau_2$  от  $\tau_1$  имеет отчетливо ступенчатый вид, свидетельствующий о категориальном характере различения. Изменение  $\tau_1$  в некоторых пределах (разных у разных испытуемых) не изменяет значения  $\tau_2$ ; когда стимул выйдет за эти пределы,  $\tau_2$  изменяется скачком и затем опять сохраняется неизменным при некотором интервале значений  $\tau_1$ . Число категорий, как и по данным идентификации, оказывается равным трем. Только внутри третьей категории ( $\tau_1 > 50$  мс) имеет место непрерывная зависимость между  $\tau_2$  и  $\tau_1$ , которую следовало бы

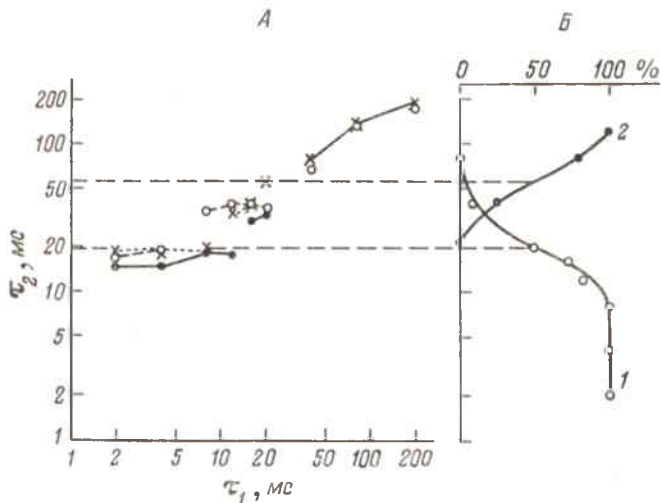


Рис. 11.17. Классификация (А) и различение (В) шумов с разной постоянной времени нарастания.

На А: по оси абсцисс — постоянная времени нарастания стандарта; по оси ординат — постоянная времени нарастания тестового стимула, при которой он отличается от стандарта в 75% случаев; точки, кружки и крестики соответствуют разным испытуемым. На В: по оси абсцисс — постоянная времени нарастания стимула; по оси ординат — процент отнесения стимула к определенному классу; 1 — процент отнесения стимула к первому классу (резкое начало), 2 — к третьему классу (плавное нарастание).

ожидать, если бы человек пользовался, например, такой характеристикой стимула, как начальное значение громкости.

Важно отметить, что скорость нарастания переднего фронта шума имеет существенное значение для различения взрывных и щелевых согласных. Для получения взрывного согласного необходимы крутой передний фронт и малая длительность, для получения щелевого — медленное нарастание и большая длительность [9].

По какому бы принципу ни осуществлялась классификация форм огибающей сигнала, во всяком случае приходится допустить существование некоторого набора детекторов форм. Эти детекторы должны некоторым образом взаимодействовать, что следует из результатов, полученных в работе [102]. В работе использовалось дихотическое предъявление [а]-подобного стимула с разными ва-

риантами огибающей, схематически показанными на рис. 11.18. Обозначения правое ухо и левое ухо являются условными, так как никакого намека на доминантность одного уха в этой работе получено не было.  $L_A$  и  $L_B$  — уровни ощущения стимулов.

В предварительных экспериментах было показано, что стимулы типа *I* воспринимаются как [aga] или [a], вероятность их восприятия как [aga] является монотонно возрастающей функцией  $L_B^*$ , где  $L_B^* = L_B$  при  $L_A \leq C$  и  $L_B^* = L_B - 0.6(L_A - C)$  при  $L_A > C$ . Стимулы типа *II* воспринимаются как [ala], [a]—[a] или [a]; вероятность их восприятия как двух гласных, разделенных паузой или согласным [l], также является монотонно возрастающей

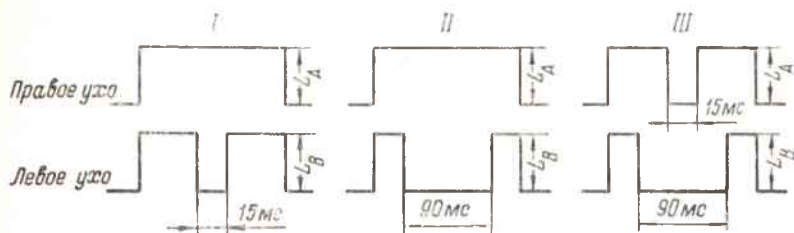


Рис. 11.18. Типы (I—III) огибающих дихотически предъявляемых стимулов. По [102].

Обозначения см. в тексте.

щей функцией  $L_B^*$ . Существенно, что в стимулах типа *I* испытуемый воспринимал только [r]-паузу; в стимулах типа *II* он ее никогда не слышал. В основных экспериментах использовались стимулы типа *III*.

На рис. 11.19 приведены результаты опыта, когда стимул *A* с паузой, равной 15 мс, имел высокий уровень интенсивности ( $L_A = 61$  дБ), а уровень стимула *B* с паузой в 90 мс менялся в пределах от 0.5 до 53 дБ. Испытуемым разрешались ответы [aga] и [ala]. На рисунке крестиками показана зависимость вероятности ответа [ala] от  $L_B^*$ . Точками показаны результаты предварительного эксперимента со стимулом *II*. Можно видеть, что данные обоих экспериментов совпадают. Это говорит о том, что испытуемые воспринимают [r]-паузу только в тех случаях, когда интенсивность стимула с большой (90 мс) паузой слишком мала, чтобы эта большая пауза могла быть обнаружена.

На рис. 11.20 приведены результаты опыта, где стимул с 90-миллисекундной паузой имел постоянный уровень, равный 35 дБ. Мы назовем его сейчас стимулом *A*. Стимул *B* с паузой 15 мс изменялся в пределах от 18 до 61 дБ. Испытуемым разрешались ответы [a]—[a], [ala], [ara]. Крестиками показана зависимость вероятности ответа [aga] от  $L_B^*$ . Можно видеть, что вероятность не превышает 0.2. Для сравнения точками приведены данные эксперимента со стимулом типа *I* рис. 11.18, когда стимул *A* является сплошной

посылкой. Можно видеть, что в этих условиях вероятность ответа [ara] превышает 0.9. Таким образом, приходится считать, что обнаружение большой паузы или, вернее, неравномерностей, соответствующих переднему и заднему фронту ступеньки, запрещает обнаружение [г]-паузы.

В простейших случаях (например, для различения положительных и отрицательных ступенек и [г]-паузы) роль детекторов форм могут играть фильтры огибающей с положительными и отрицательными порогами, если дополнить эту систему некоторым

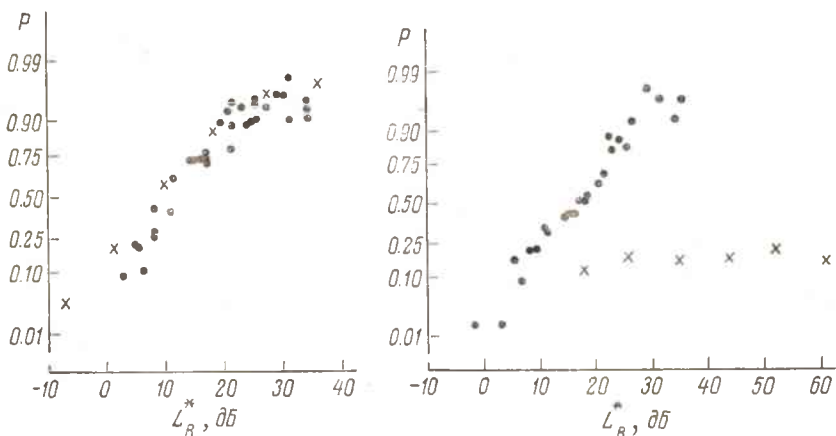


Рис. 11.19. Относительная частота восприятия [л] в стимулах типа II и III (рис. 11.18) как функция уровня интенсивности стимула. По [102].

По оси абсцисс — уровень интенсивности; по оси ординат — относительная частота ответа [ala]. Крестики — стимул типа III, точки — стимул типа II.

Рис. 11.20. Относительная частота восприятия [г] в стимулах типа I и III (рис. 11.18) как функция уровня интенсивности стимула. По [102].

По оси абсцисс — уровень интенсивности; по оси ординат — относительная частота ответа [aga]. Крестики — стимул типа III, точки — стимул типа I.

логическим устройством. Пусть система состоит из двух фильтров — «фильтра [г]» и «фильтра хрипlosti». В таком случае положительное ступенчатое приращение соответствует превышению положительных порогов на выходе обоих фильтров, отрицательное — превышению отрицательных порогов; [г]-пауза соответствует превышению отрицательного порога на выходе «фильтра [г]» и превышению обоих порогов на выходе «фильтра хрипlosti».

За счет переходных процессов в фильтрах могут происходить некоторые взаимодействия неравномерностей, относящихся к разным классам, например, ухудшение обнаружения паузы вблизи переднего фронта стимула [286].

В работе [286] сигнал представлял собой посылку шума длительностью 500 мс с паузой внутри посылки. Пауза помещалась



на разных интервалах времени от начала или от конца посылки. Если пауза была близка к началу посылки, то фиксированным (установленным экспериментатором) был интервал от начала стимула до начала паузы. Если пауза была близка к концу посылки, то фиксированным был интервал между концом паузы и концом стимула. Задача испытуемого состояла в установке порогового значения длительности паузы ( $t_n$ ), при которой он ее обнару-

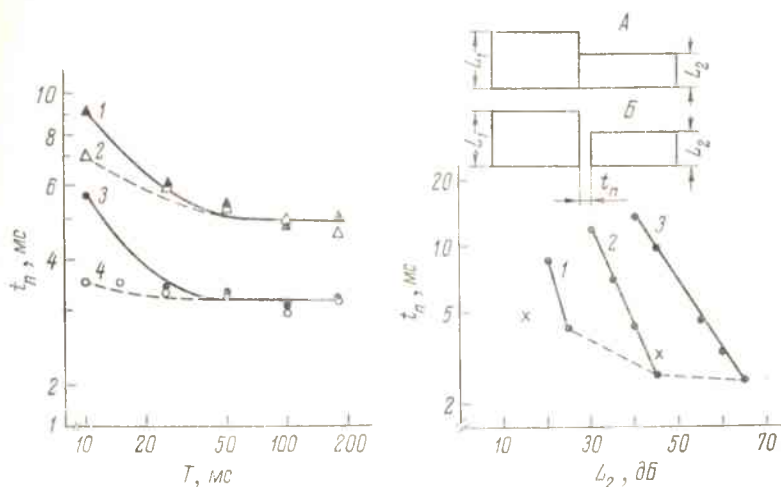


Рис. 11.21. Зависимость порога обнаружения паузы от ее положения в шумовом стимуле. По [286].

По оси абсцисс — длительность интервала от начала стимула до начала паузы (1 и 3) или от конца паузы до конца стимула (2 и 4); по оси ординат — пороговая длительность паузы. 1 и 2 — при уровне ощущения стимула 15 дБ, 3 и 4 — 46 дБ. Остальные обозначения см. в тексте.

Рис. 11.22. Минимальная длительность паузы, соответствующая порогу различения стимулов А и В, как функция уровня ощущения стимула на слабом сегменте.

По оси абсцисс — уровень ощущения стимула; по оси ординат — минимальная длительность паузы. В верхней части рисунка показаны огибающие стимулов. 1 —  $L_1=25$  дБ, 2 —  $L_1=45$  дБ, 3 —  $L_1=65$  дБ. Крестики — порог обнаружения паузы, находящийся вдаль от краев стимула [286].

руживает. Измерения проводились при двух значениях уровня ощущения стимула: 15 и 46 дБ. Полученные данные (средние по обоим испытуемым значения  $t_n$ ) приведены на рис. 11.21. Можно видеть, что увеличение  $t_n$  с приближением паузы к началу стимула выражено очень отчетливо и наблюдается при обоих уровнях интенсивности стимула. Приближение паузы к концу стимула также ухудшает ее обнаружение. Эффект оказывается статистически значимым при обоих уровнях интенсивности стимула, однако он количественно очень мал при уровне 46 дБ.

Более интересная ситуация, при которой приходится допу-

стить какие-то логические операции для различения форм, исследовалась в работе Пломпа [424]. Вверху на рис. 11.22 схематически изображены применявшиеся в работе стимулы. В работе определялось пороговое значение паузы ( $t_{\text{п}}$ ), при которой испытуемые отличали стимул  $A$  от стимула  $B$ . Измерения проводились при трех значениях  $L_1$ ; исследовалась зависимость  $t_{\text{п}}$  от  $L_2$ .

На рис. 11.22 приведены данные для тех комбинаций  $L_1L_2$ , где обнаруживаемая пауза была еще достаточно малой и можно быть уверенным, что различение основывалось именно на восприятии паузы. Можно видеть, что  $t_{\text{п}}$  отчетливо возрастает при увеличении перепада интенсивностей, причем скорость относительного изменения  $t_{\text{п}}$  при увеличении перепада (уменьшении  $L_2$ ) зависит от начального уровня интенсивности ( $L_1$ ). При  $L_1=45$  дБ наклон прямой (скорость изменения  $t_{\text{п}}$ ) больше, чем при  $L_1=65$  дБ.

Из данных Пломпа следует также, что пороговая длительность паузы возрастает при уменьшении  $L_1=L_2$ . Это согласуется с данными Герена и Жуковой [286]; полученные в их экспериментах значения  $t_{\text{п}}$  (для случая, когда пауза находилась вдали от краев стимула) приведены на рис. 11.22 крестиками. Данные Пломпа можно объяснить, предположив, что сигнал относится к классу  $B$  в том случае, если на выходе «фильтра хриплости» превышаются и положительный, и отрицательный пороги; в остальных случаях сигнал относится к классу  $A$ .

Конечно, высказанные в этом разделе соображения о детекторах форм сугубо приближительны. Экспериментальных данных пока слишком мало, чтобы построить сколько-нибудь пригодную модель классификации форм огибающей. Кроме того, такую модель едва ли можно построить не имея хорошей модели системы обработки огибающей.

## СЛУХОВОЕ ОПИСАНИЕ СИГНАЛОВ С ИЗМЕНЯЮЩИМСЯ ВО ВРЕМЕНИ СПЕКТРОМ

Существующие в настоящее время данные относительно слухового анализа сигналов с изменяющимся спектром касаются или синусоидальных тонов, модулированных по частоте (ЧМ-тоны), или звуков со сложным спектром, у которых один из спектральных максимумов изменяет свое положение на шкале частот. Функцию, описывающую зависимость от времени частоты спектрального максимума исходного сигнала, обозначим  $f(t)$ .

Естественно исходить из того, что слуховая обработка звуковых сигналов основана на спектральном анализе, т. е. преобразовании исходного акустического сигнала в пространственно-временное распределение плотности импульсации  $g(z, t)$ . Будем считать, что последующей необходимой процедурой является определение или текущего положения спектрального максимума на шкале  $z$  (переход к  $z^*(t)$ ;  $z^*$  — координата спектрального максимума), или каких-то параметров, отражающих особенности изменения частоты максимума во времени.

Такое представление отражает существо явлений несколько упрощенно, так как не исключено, что слуховое измерение частоты может быть основано также и на анализе периодичности  $g(t)$  в частотных каналах системы [276, 396, 460].

Насколько подробно высшие отделы слуховой системы могут «отслеживать» временной ход частоты спектрального максимума, зависит от инерционности спектрального анализа и тех процессов, которые приводят к определению положения спектрального максимума на шкале  $z$ . Качественные данные, полученные в экспериментах по восприятию быстрых тональных последовательностей [234] и по определению маскировки короткого стационарного тона ЧМ-сигналом [295, 449], показывают, что переходные процессы этих преобразований не очень длительны.

В первой цитируемой работе испытуемые идентифицировали (после предварительной тренировки) шесть стимулов, представляю-

щих собой последовательность из трех примыкающих друг к другу тональных посылок равной длительности, образующих разные мелодические контуры, но занимающие одно и то же среднее на весь стимул положение по частоте (1000 Гц). Оказалось, что неслучайные ответы наблюдаются уже в том случае, когда длительность каждой из посылок составляет всего 4—6 мс. (Пороговая длительность зависит от некоторых факторов, в том числе и от расстояния между соседними звуками по шкале частот; здесь указана та величина, которая получена при достаточно большом разнесении частот —  $1/3$  октавы).

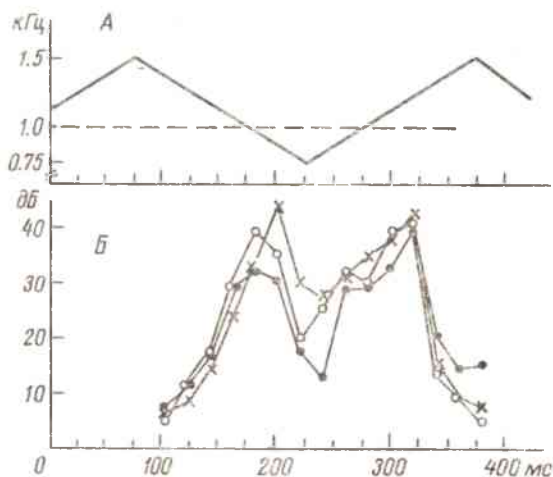


Рис. 12.1. Маскировка короткой посылки тона частотно-модулированным сигналом. По [449].

На А — изменение частоты маскира во времени. По оси абсцисс — время; по оси ординат — частота; штрихами указана частота маскируемого тона (1000 Гц). На Б — кривые маскировки посылки тона при разном положении его во времени относительно начала маскира. По оси абсцисс — время задержки начала тональной посылки; по оси ординат — разность порогов обнаружения тона при маскировании и в тишине; разные кривые — данные разных испытуемых.

Кривые одновременной маскировки короткой посылки (20 мс) стационарного тона непрерывным ЧМ-сигналом были получены в экспериментах Ронкена [449] при 2 условиях: 1) тестовый сигнал варьировал по частоте, но имел фиксированное положение относительно периода изменения частоты маскира; 2) тестовый сигнал имел постоянную частоту, равную центральной частоте маскира, а его положение относительно периода ЧМ-маскира изменялось ступенями. Результаты одного из экспериментов, выполненных в соответствии со вторым условием для периода модуляции маскира 300 мс, приведены на рис. 12.1. Как видно из рисунка, максимумы маскировки приходятся только на те положения тестового сигнала, при которых его частота совпадает с мгновенной частотой маскира. Объясняя инерционность всех процессов, приводящих

к обнаружению тона на фоне маскера низкочастотной фильтрацией огибающей  $g(t)$ , автор находит, что постоянная времени интегрирования не должна превышать 10 мс.

Свидетельством малой инерционности спектрального слухового анализа могут служить также результаты опытов [285] по определению остаточной маскировки. Маскером служила посылка тона длительностью 50 мс с частотой, линейно изменяющейся во времени от 1000 до 1500 Гц (или в обратном направлении). Тестовая посылка стационарного тона длительностью 20 мс следовала через 5 мс после окончания маскера и могла иметь разные значения частоты. В этих условиях наиболее выраженная маскировка имела место только на частоте, близкой к конечной частоте маскера.

Таким образом, приведенные данные дают основание считать, что инерционность спектрального преобразования исходного сигнала, обусловленная переходными процессами слухового анализатора, ограничивается длительностями порядка нескольких миллисекунд (не более 10).

Будем условно считать функцию  $f(t)$  достаточно медленной, если она задана на интервалах, больших длительности переходных процессов, и при этом не будем различать  $f(t)$  и  $z^*(t)$ , называя и ту и другую функцией временным контуром стимула.

Определив таким образом понятие контура  $f(t)$ , будем рассматривать проблему восприятия изменяющихся по частоте сигналов как задачу определения способов слуховой обработки контура и выяснения признаков, по которым он описывается.

### 12.1. ПРЕДПОЛОЖЕНИЯ ОБ ОПИСАНИИ КОНТУРА ИЗМЕНЕНИЯ ЧАСТОТЫ

Среди современных исследователей речи [483] распространено предположение о том, что на некотором этапе слуховой обработки контур  $f(t)$  может быть представлен полностью — в виде последовательности отсчетов, производимых через равномерные интервалы времени (см. обсуждение в главе 6). В таком виде информация о контуре сохраняется некоторое время в непосредственной, слуховой памяти и считывается устройствами, осуществляющими дальнейшую специфическую обработку, в частности детекторами фонетических признаков.

В противоположность этой гипотезе выдвигается другое предположение, в принципе отвергающее возможность полного, поточечного описания. Предполагается обработка стимула, в результате которой слуховое описание контура представляет собой набор его характерных признаков, таких, например, как местоположение «особых точек» (начало, конец, экстремумы) по координате  $z$  и некоторые характеристики изменений (например, направление и скорость). Функция определения характерных признаков контура возлагается на специальные детекторы.



В работах [48. 98. 100. 371], предпринятых для выбора между этими гипотезами, использовались следующие соображения. При описании  $f(t)$  поточечным способом расстояние между двумя параллельно сдвинутыми друг относительно друга по частоте контурами является возрастающей функцией от числа отсчетов, т. е. от длительности сигнала (следует из определения метрик, употребляющихся при поточечном описании функций [143]).

При втором способе описания предполагается, что набор признаков для контуров одинаковой формы будет одним и тем же, независимо от длительности, и потому длительность не должна влиять на субъективное расстояние. Как было показано в работе [99], при численном шкалировании высотных интервалов, образуемых двумя стационарными тонами, субъективное расстояние есть функция не только частотного различия стимулов, но и их длительности: при изменении последней в пределах от 5 до 100 мс субъективное расстояние для звуков с одинаковым различием по частоте монотонно увеличивается. Этот эффект может свидетельствовать в пользу гипотезы полного описания, однако для ее утверждения требуется, чтобы подобным свойством обладали расстояния для звуков с изменяющейся во времени частотой. Насколько выполнимо это условие, показали результаты следующих экспериментов, проводившихся также с помощью метода численного шкалирования субъективных расстояний.

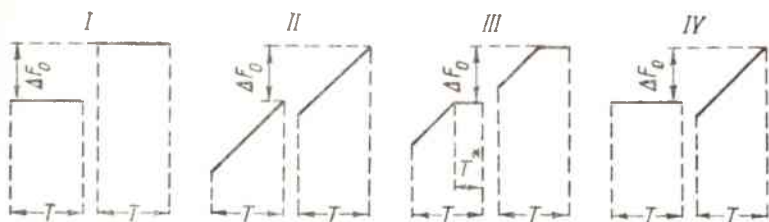
Стимулы — пары следующих друг за другом синтезированных гласноподобных звуков — оценивались слушателями по степени их различия, создаваемого либо за счет частоты основного тона, либо за счет частоты второй форманты (в каждом опыте изменялся только один параметр). В каждой паре, подлежащей оценке, звуки были одинаковы по всем параметрам и различались только параллельным сдвигом контуров друг относительно друга. Примеры схематического изображения контуров частоты основного тона звуков в парах и результаты экспериментов по шкалированию таких стимулов представлены на рис. 12.2.

Средние оценки, характеризующие субъективное расстояние при сравнении стационарных звуков (рис. 12.2, *Б*, стимулы *I* и *Ia*), а также в том случае, когда только один из звуков стационарный (стимулы *IV*), увеличиваются с удлинением стимулов при одной и той же разности частот основного тона. Для звуков с изменяющейся частотой основного тона субъективное расстояние либо не зависит от длительности, либо зависит в гораздо меньшей степени, чем для стационарных. При этом для стимулов типа *III* в тех случаях, когда эта зависимость проявляется, оценки положительно коррелируют только с длительностью постоянной части стимула.

Как следует из принятых допущений, зависимость субъективного расстояния от длительности должна наблюдаться для изменяющихся по частоте сигналов с любым законом изменения. Настоящие результаты не согласуются с этим условием и дают основание

полагать, что, оценивая различия между звуками, испытуемые пользовались не полным их описанием, а некоторым набором специальных признаков контура. К этому же выводу приводят ре-

А



Б

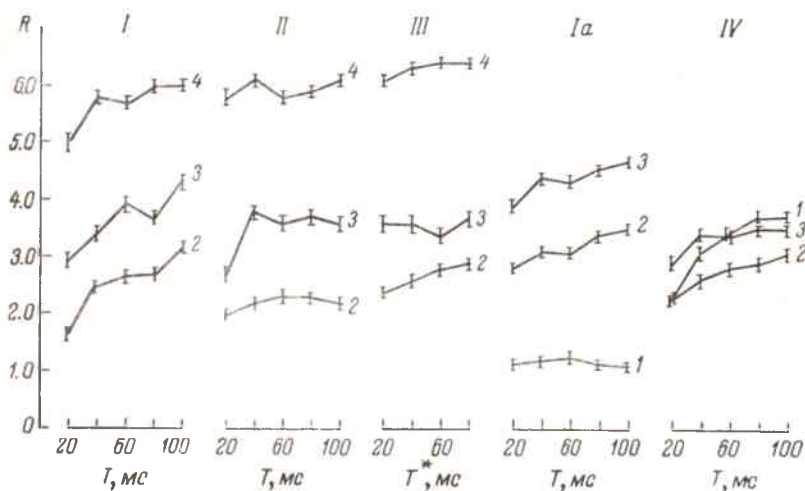


Рис. 12.2. Результаты опытов по шкалированию субъективных расстояний между синтетическими гласными. По [100].

На А — временные контуры частоты основного тона для разных типов стимулов. На Б — зависимость от длительности средних оценок различия разных типов стимулов по высоте. По оси абсцисс — длительность посылки (для стимулов типа III — длительность стационарного участка); по оси ординат — средние оценки в баллах. Параметр (кривых —  $\Delta F_0$ : 1 — 0, 2 — 50, 3 — 100, 4 — 200 Гц. Вертикальные штрихи у каждой точки обозначают доверительные интервалы (критерий значимости 0.05). Римские цифры обозначают, к какому типу стимулов относятся кривые (I и Ia — данные разных опытов).

зультаты экспериментов по шкалированию тембральных различий звуков с разными контурами частоты второй форманты [98].

Полученные в работе [100] данные позволили сделать также некоторое предположение о природе признаков контура частоты основного тона, на основании которых оценивались различия между звуками. Анализируя те отличия, которые проявляются в оценках высотных интервалов стимулов типа I и IV (кривые Ia и IV на рис. 12.2, Б), удалось показать, что требуемыми свой-

ствами должны обладать расстояния, если они определяются на пространстве двух признаков, один из которых характеризует положение стимула на шкале частот, другой есть некоторый эквивалент изменения частоты на протяжении звука.

Эффект увеличения субъективного расстояния при увеличении длительности стационарных стимулов, по-видимому, имеет отношение к известному явлению улучшения точности различения частоты при больших длительностях, что проявляется при исследовании дифференциальных порогов [101, 395, 463]. Оба эффекта наблюдаются при одном и том же диапазоне длительностей сравниваемых стимулов.

## 12.2. ВЫДЕЛЕНИЕ «ОСОБЫХ ТОЧЕК» В КОНТУРЕ $f(t)$

Измерение слухом частоты спектральных максимумов (частоты формант) при восприятии фонетического качества речевых сигналов можно считать экспериментально установленным фактом. Устанавливается он на примере восприятия стационарных стимулов. В отношении к сигналам с изменяющимся спектром это положение требует более конкретной формулировки: в какие моменты времени слуховая система осуществляет измерение частоты, чем определяется задание моментов измерения и какова природа измеряемых величин (мгновенные отсчеты или средние значения и т. д.).

### 12.2.1. ВОСПРИЯТИЕ РЕЧЕПОДОБНЫХ ЗВУКОВ С ОДНОНАПРАВЛЕННЫМ КОНТУРОМ

В главе 6 при обсуждении проблемы сегментации приводились данные о том, что твердость или мягкость [r], создаваемого кратковременным прерыванием гласноподобного стимула, определяется значением  $F_2$  в момент начала второго гласного. Этот результат качественно согласуется с данными, полученными ранее Деркачом [230], который исследовал восприятие твердости-мягкости согласных [t] и [s] в сочетаниях ГСГ. На рис. 12.3 приведена схема синтеза параметров для 7 стимулов: интервалы включения голосового и шумового источников возбуждения, контур второй форманты и контур первой форманты. Частота формант на участке *аб* соответствует гласному [i], на участке *вг* — гласному [a]. Начало переходов второго гласного обозначено точками (его форманты для стимулов 1 и 2, а также 6 и 7 начинаются из общих точек). Цифры под стрелками обозначают процент ответов, в которых согласный опознан как мягкий. Видно, что признак мягкости распознается почти полностью (в 90%) начиная со стимула 6, т. е. только тогда, когда начальная точка переходов  $F_1$  и  $F_2$  гласного [a] соответствует частотам, характерным для [i].

Экспериментальное доказательство использования краевых значений  $F_2$  для фонемной классификации гласных русскими слушателями было получено Люблинской и Слепокуровой при исследовании восприятия синтетических гласноподобных звуков дифтонгоидного характера.

Сочетание двух разных гласных, а также дифтонги в потоке речи реализуются в виде непрерывно изменяющегося спектра: зависимости частот формант от времени образуют плавные траек-

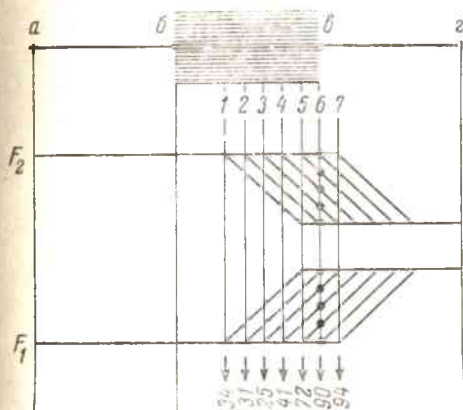
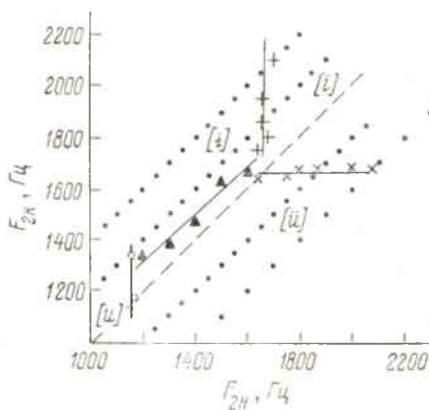


Рис. 12.3. Схема параметров синтетических ГСГ-слов при исследовании восприятия признака мягкости согласных. По [230].

Объяснения в тексте.

Рис. 12.4. Результаты опытов по идентификации синтетических дифтонгоидных гласных звуков.

По оси абсцисс — частота второй форманты в начале звука; по оси ординат — частота второй форманты в конце звука. Сплошные линии — фонемные границы, точки — параметры стимулов, штриховая линия соответствует условию  $F_{2н} = F_{2к}$ .



тории, при этом формантные максимумы перемещаются из положения, характерного для первого по времени гласного, ко второму. В естественной речи, произносимой в нормальном темпе, начальные и конечные положения оформляются в виде стационарных участков. При быстром произношении стационарные участки могут отсутствовать, и остается, практически, только участок однонаправленного изменения частот формант. Известные работы, в которых исследуются признаки восприятия дифтонгов [271, 273, 274], показывают, что стационарные участки роли не играют.

В опытах Люблинской и Слепокуровой в качестве стимулов использовались синтетические гласноподобные звуки с частотой основного тона 125 Гц, частотой первой форманты 300 Гц и длительностью 80 мс. Частота второй форманты либо линейно возрастала, либо убывала, либо оставалась постоянной. В изменяющихся сигналах начальное и конечное значения частоты фиксиро-

вались в течение 10 мс, образуя стационарные участки. Изменение частоты осуществлялось в середине звука на интервале в 60 мс на величину  $\Delta F_2 = \pm 200, \pm 400$  Гц. (Величина частотного перепада  $\Delta F_2$  определялась как разность частот начального  $F_{2н}$  и конечного  $F_{2к}$  участков). При фиксированном перепаде  $\Delta F_2$  положение звуков по шкале частот второй форманты варьировало в широком диапазоне:  $F_{2н} = 1000 \div 2400$  Гц, ступенями через 50 или 100 Гц (сетка значений параметров стимулов приведена на рис. 12.4). Испытуемые идентифицировали предъявляемые им звуки с одним из 4 гласных: [i], [u], [i] или [ü]. Последний звук представляет собой аллофон гласного [u], встречающийся в сочетаниях после мягких согласных, который воспринимается носителями русского языка как подкласс звуков, отличный от собственно [u] после твердых согласных [32].

На рис. 12.4 представлены границы между указанными категориями, определенные по матрице ответов по критерию равновероятного опознавания. Основное значение для вопроса о признаках траекторий имеет форма границ между гласными [i], [i] и [ü]. Образуя смежные области классов, они отделяются довольно простыми по форме границами: между [i] и [i] — это прямая, параллельная оси  $F_{2к}$ , между [ü] и [i] — прямая, параллельная оси  $F_{2н}$ .

В обоих случаях граница описывается только одним числом, и числа эти очень близки друг к другу: 1670 и 1660 Гц соответственно. Малые различия этих значений позволяют их считать случайными реализациями одной и той же величины. Существенно, что в случае опознавания [i] и [i], у которых  $F_{2н} < F_{2к}$ , граница устанавливается по значению  $F_{2н}$ , в случае [i] и [ü], у которых  $F_{2к} < F_{2н}$ , граница устанавливается по  $F_{2к}$ , т. е. граница определяется по наименьшему из параметров.

Из приведенных данных следует, что обработка звуков с однонаправленным изменением формантного контура в условиях фоновой идентификации должна включать в себя измерение частоты в начальной и конечной точках контура и процедуру принятия решения, основанную на сравнении измеренных значений с некоторым фиксированным порогом  $\Pi$ .

В качестве одного из возможных алгоритмов можно предположить следующий:

1) стимул  $S$  определяется как [i], ( $S \in [i]$ ), если  $(z_1^* > \Pi) \wedge (z_2^* > \Pi)$ ; здесь  $z_1^*$  и  $z_2^*$  обозначают результат слухового измерения частоты 2-й форманты в начале и конце звука;

2)  $S \in [i]$ , если  $(z_1^* < \Pi) \wedge (z_1^* < z_2^*)$ ;

3)  $S \in [\bar{u}]$ , если  $(z_2^* < \Pi) \wedge (z_2^* < z_1^*)$ .

Можно видеть, что обработка сигналов в данной ситуации, кроме измерения отсчетов и сравнения их с порогом, должна включать в себя или сравнение  $z_2^*$  и  $z_1^*$ , или прямое измерение направления изменения частоты (о способности слуховой системы выделять этот параметр переменного звука будет сказано ниже).



Доводом в пользу того, что однонаправленный частотный контур описывается в слухе своими краевыми точками, может служить также результат работ [96, 97], в которых вопрос о признаках контура частоты основного тона  $F_0(t)$  исследовался с помощью метода имитации. Синтетические гласные звуки с линейно возрастающей частотой основного тона предъявлялись испытуемым, задачей которых было имитировать интонацию стимулов. Определяющая роль краевых точек контура  $F_0(t)$  устанавливалась по тому факту, что изменение  $F_0$  в ответном сигнале, как правило, начиналось и кончалось участками с постоянной частотой, равной в среднем начальной и конечной частотам исходного контура соответственно. При этом в исходном стимуле стационарных участков могло и не быть.

Таким образом, приведенные экспериментальные факты дают возможность утверждать, что при восприятии коротких сигналов с однонаправленным частотным контуром слуховыми признаками контура могут служить результаты измерения частоты в начале и конце звука. Важно, что на основании этих данных мы удостоверяемся в способности слуховой системы измерять абсолютные значения параметра в некоторые определенные моменты времени, привязанные к входному стимулу.

#### 12.2.2. ИЗМЕРЕНИЕ ЧАСТОТЫ В ЭКСТРЕМАЛЬНЫХ ТОЧКАХ КОНТУРА $f(t)$

До сих пор речь шла о сигналах с самым простым контуром  $f(t)$ , где частота изменяется только в одном направлении. Более сложным случаем представляется восприятие звуков, у которых форма траектории включает смену направления изменения. Интуитивно понятно, что два сигнала, начало и конец которых имеют одинаковые значения по частоте, но изменение частоты внутри звука у одного из них осуществляется по выпуклой, а у другого по вогнутой траектории, должны восприниматься по-разному. Подобная форма траектории, например, присуща формантным контурам разных гласных звуков между одинаковыми согласными в потоке речи [359, 480, 484, 486].

Насколько известно, была сделана только одна попытка исследовать восприятие звуков с формантным контуром такой формы [361]. В указанной работе использовались синтетические стимулы с тремя различными формами траекторий изменения частоты  $F_2$  и  $F_3$ . 1-й тип стимулов — стационарные гласные (Г), 2-й тип — звукосочетания [w] Г [w] с выпуклой формой контуров формант (рис. 12.5, А), 3-й тип — звукосочетания [j] Г [j] с вогнутой формой контуров (рис. 12.5, Б). Начальные, конечные и экстремальные участки оформлялись в виде 20-миллисекундных стационарных интервалов. Параметры стимулов ( $F_2$  и  $F_3$ ) приведены на рис. 12.6. Кроме того, все стимулы имели два варианта временного масштаба: в случае медленного варианта переменная часть стимула

была равна 200 мс, в случае быстрого — 100 мс. Каждый из типов стимулов (всего 6) предъявлялся в виде отдельного теста.

Испытуемые (10 американцев) должны были классифицировать все звуки на две категории гласных (в слоге или в изолированном предъявлении) — [u] и [i]. Задача эксперимента состояла в том, чтобы путем сравнения положения фонемных границ, полученных для разных типов стимулов, проверить предположение о том, что восприятие фонетического качества определяется частотой фор-

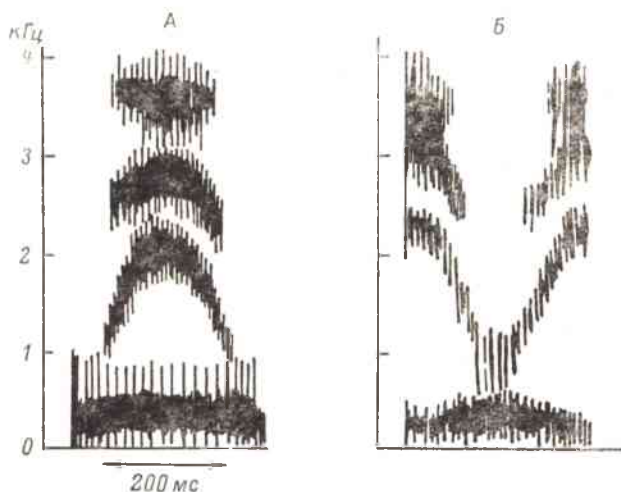


Рис. 12.5. Динамические спектрограммы синтетических слогов. По [361].  
По оси абсцисс — время; по оси ординат — частоты формант. На А — слоги [w] Γ [w];  
на Б — слоги [j] Γ [j].

мант в экстремальной точке контура. Если предположение верно, положение фонемной границы по  $F_2$  и  $F_3$  должно быть одинаковым для всех трех типов стимулов.

Результат получился следующий: средние для всех испытуемых значения границы (для обеих длительностей стимулов) для стационарных гласных и [j] Γ [j]-звукосочетаний действительно довольно похожи: для стационарных гласных  $F_2 = 1644$  Гц,  $F_3 = 2648$  Гц, для [j] Γ [j] —  $F_2 = 1634$  Гц,  $F_3 = 2632$  Гц (данные вычислены по таблицам, приведенным в статье). В то же время при классификации [w] Γ [w]-стимулов совпадения границ с изолированными гласным в ответах большинства испытуемых (для 7 из 10) не наблюдается (рис. 12.7). Средние значения границы в этом случае существенно ниже:  $F_2 = 1474$  Гц,  $F_3 = 2553$  Гц.

Совпадение границ для изолированных гласных и слогов типа [j] Γ [j] можно рассматривать как доказательство использования человеком при распознавании гласных значений частоты формант в экстремальной точке.

Результаты обсуждаемой работы интересным образом переключаются с данными опытов Люблинской и Слепокуровой по восприятию дифтонгоидных гласных. Фонемная граница по параметру  $F_2$  между [i] и [u] для стимулов типа [j] Г [j] и между [i] и [ü] для стимулов с однонаправленным нисходящим контуром описывается практически одним и тем же числом, только в первом случае оно соответствует экстремальной, а во втором — конечной точке кон-

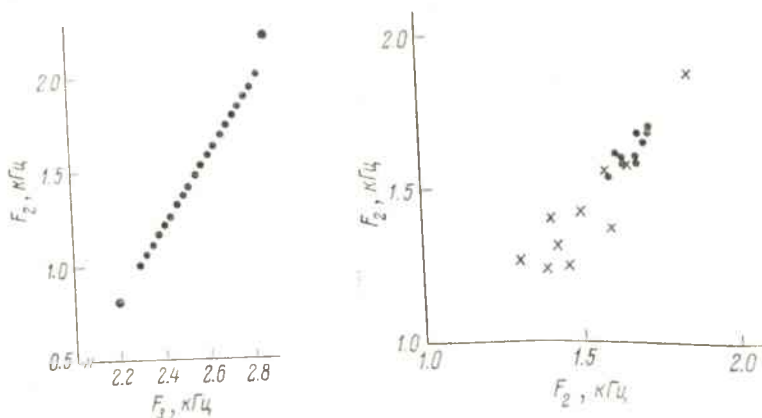


Рис. 12.6. Параметры синтетических стимулов, использованных в опытах по восприятию слогов [w] Г [w] и [j] Г [j]. По [361].

По оси абсцисс — частота третьей форманты; по оси ординат — частота второй форманты. Маленькие точки соответствуют значениям  $F_2$  и  $F_3$  на стационарном участке гласного; большие точки показывают значения  $F_2$  и  $F_3$  на участках согласного (нижняя — для [w]; верхняя — для [j]).

Рис. 12.7. Положение границы между [i] и [u] по частоте второй форманты ( $F_2$ ) в изолированных гласных (точки) и в слогах [w] Г [w] (крестики) у разных испытуемых. По [361].

Каждое значение на графике — результаты одного испытуемого при идентификации стимулов с двумя разными длительностями. По оси абсцисс — границы для гласных длительностью 200 мс; по оси ординат — для гласных длительностью 100 мс.

тура. Если такое совпадение не случайно, то этот факт указывает на возможность рассматривать конечное и экстремальное значения частоты как одну и ту же «особую точку» контура. Т. е. возможно, что момент для измерения (считывания) частоты максимума в обоих случаях задается по некоторому общему правилу и результат измерения одинаковым образом используется при фонетическом анализе. Насколько реально такое предположение, решить можно только в результате специального исследования правил выделения особых точек в контуре.

В связи с высказанным предположением кажется возможным предложить следующее объяснение наблюдаемому смещению фонемной границы для [w] Г [w]-стимулов по отношению к границе изолированного гласного.

Как показывают опыты по восприятию дифтонгоидных гласных, для стимулов с формой траектории формантных контуров, при которой начальные значения частоты  $F_2$  меньше конечных, а применительно к данному случаю меньше экстремальной частоты, для русских испытуемых существует не два, а три класса возможных решений о гласных: [u], [i] и [i]. При этом граница по  $F_2$  между [u] и [i] расположена ниже, чем граница между [i] и [i]. Поскольку есть основание считать (см. главу 4), что число воспринимаемых категорий гласных — психологических фонем — в любом языке больше числа лингвистических фонем и соответствует примерно числу вариантов, то возможно, что носители американского варианта английского языка, так же как и русские, различают три категории в наборе стимулов типа [w] Г [w]. При выборе решений относительно [u] испытуемые могли ориентироваться либо на одну, либо на другую границу, что в результате и должно приводить к такому сдвигу средней границы, который наблюдается в опытах. Поддержкой данной версии может служить большой разброс ответов испытуемых, регистрируемый при идентификации стимулов типа [w] Г [w].

### 12.2.3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ ПО ВОСПРИЯТИЮ ОДНОНАПРАВЛЕННЫХ КОНТУРОВ В ТОНАЛЬНЫХ СИГНАЛАХ

В ряде психоакустических работ, исследующих восприятие изменяющихся по частоте сигналов [184, 198, 295, 402, 403], использовалась следующая простая логика рассуждений. Известно, что изолированный стационарный сигнал воспринимается человеком с определенной высотой. Очевидно, что сигнал с изменяющейся во времени частотой (при ограниченной области изменения) также должен характеризоваться какой-то высотой. Ее можно экспериментально определить, заставив человека подобрать такое значение частоты стационарного тона ( $F_T$ ), при котором его высота совпадает с высотой переменного стимула (стимул с изменяющейся частотой будем называть в дальнейшем глайдом). Статистические параметры распределения значений  $F_T$ , которые испытуемый устанавливает при многократном выполнении этой задачи, могут служить в качестве характеристики восприятия глайда.

Следует сразу же сказать, что все авторы, использующие подобный метод, отмечают, что сравнение глайда и тона по высоте составляет довольно трудную задачу для испытуемых. В работе Набелека и др. [402] из 6 человек, приглашенных на опыты, двое вообще не могли справиться с этой задачей и были отстранены от участия в основных экспериментах. Ответы испытуемых в опытах Бейли [184] имели весьма разнообразный характер и сопровождались большим разбросом.

Рассмотрим результаты работы Набелека и др. [402], в которой наиболее подробно исследовалось восприятие переменных по частоте тонов с использованием указанной методики.

В первой серии экспериментов использовались стимулы (тоны), у которых частота изменялась во времени по линейному закону вверх или вниз непрерывно от начала,  $F_n$ , до конца посылки,  $F_k$  (100%-ные глайды), в различных диапазонах ( $|\Delta F| = |F_k - F_n| = 20 \div 1100$  Гц) при одной и той же средней частоте ( $F_{cp} = 715$  Гц) и при разной длительности стимулов, от 12 до 250 мс.

Для всех стимулов распределения значений  $F_t$  (общее число ответов на каждый стимул равно примерно 20 при участии 4 испытуемых) имели одномодальную форму (рис. 12.9). Их средние значения ( $\bar{F}_t$ ), приведенные в работе [402] в виде таблицы, указаны на рис. 12.8 точками на вертикальных стрелках. Несколько точек

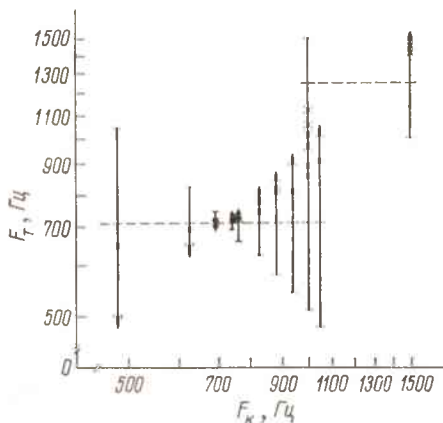


Рис. 12.8. Результаты опытов по уравниванию высоты непрерывного глайда и постоянного тона.

По оси абсцисс — частота в конце глайда  $F_k$ ; по оси ординат — частота тона. Остальные обозначения см. в тексте.

(крестиков) на одной стрелке соответствуют результатам, полученным при разных длительностях глайдов. Данные эти специально не выделены, так как не было обнаружено регулярного влияния длительности на  $F_t$ . Длина и направление стрелок соответствуют величине и знаку  $\Delta F$ , а их положение по оси абсцисс соответствует  $F_k$ . На этом же рисунке в его верхней части приведены данные работы Хейнца и др. [295], в которой применялись аналогичные стимулы длительностью 20 и 50 мс при  $\Delta F = 500$  Гц. Средние значения обозначены крестиками. Штриховые горизонтальные линии для обеих групп данных указывают среднюю частоту глайдов ( $F_{cp}$ ).

Как видно из рисунка, в большинстве случаев  $F_t$  близка к конечной частоте глайдов, причем близость эта в большей степени свойственна данным Хейнца и др., но и в результатах Набелека и др. выражена вполне отчетливо. Существенно, что  $F_t$  не равна  $F_{cp}$  даже для самых коротких глайдов и самых малых  $\Delta F$ . Авторы [402] показывают, что степень отличия  $F_t$  от средней частоты тем больше, чем больше  $\Delta F$ : соблюдается примерное постоянство отношения  $(F_t - F_{cp})/\Delta F$ .



Данные аналогичного характера были получены ранее в опытах Брэди и др. [198], использовавших в качестве переменных стимулов звуки гармонического спектра с одним максимумом, частотное положение которого изменялось во времени. Стимулы создавались путем импульсного возбуждения (с частотой следования импульсов 100 Гц) резонансного контура с перестраиваемой средней частотой. Они предъявлялись испытуемым в паре с сигналами со стационарным спектром, в которых можно было изменять частоту максимума, добиваясь близкого сходства обоих стимулов по тембру. В результате оказалось, что устанавливаемая частота в среднем равна конечной частоте спектрального максимума переменного сигнала.

Рассмотренные данные допускают вывод о том, что при установлении сходства между глайдом и стационарным тоном человек использует в качестве признака глайда положение спектрального максимума в момент окончания звука. Другими словами, характеристикой контура  $f(t)$  служит результат измерения слухом его конечной частоты  $z_k^*$ . Измерение конечной частоты сопровождается усреднением частоты на некотором временном интервале. Судя по тому, что даже для коротких (12 мс) глайдов  $F_t \neq F_{cp}$ , интервал усреднения не очень велик (не превышает  $10 \div 20$  мс).

Во второй серии экспериментов Набелека и др. использовались стимулы, контур  $f(t)$  которых соответствовал комбинации отрезков с постоянной и изменяющейся частотой, как показано на рис. 12.9. Присутствие в контуре стационарных участков изменяет характер распределений  $F_t$  следующим образом:

1) если  $\Delta F < \Delta F_I$  (где  $\Delta F_I$  — некоторая критическая величина), то независимо от соотношения длительностей стационарных участков и переходов распределение  $F_t$  имеет одномодальную форму; удлинение одного стационарного отрезка за счет другого приводит к сдвигу  $F_t$  в направлении более длинного отрезка (рис. 12.9, А);

2) если  $\Delta F > \Delta F_{II}$  (где  $\Delta F_{II} > \Delta F_I$ ), то в зависимости от соотношения длительностей перехода и стационарных участков распределение  $F_t$  имеет либо один, либо два максимума. Последнее наблюдается лишь при наличии двух стационарных участков по краям контура (рис. 12.9, В), и распределения концентрируются в областях  $F_{II}$  и  $F_k$ .

Определенные в работе значения  $\Delta F_I$  и  $\Delta F_{II}$  для стимулов с двумя стационарными отрезками равной длительности и коротким (5% от общей длительности) переходом (контур ступенчатой формы) в зависимости от общей длительности стимулов приведены на рис. 12.10 для случая, когда  $F_{cp} = 715$  Гц. Подобная зависимость была определена также для стимулов с  $F_{cp} = 2000$  Гц. Оказалось, что значение  $\Delta F_{II}$  при длительностях, больших 40 мс, примерно соответствует величинам критических полос на этих частотах [145].

Можно утверждать, что для тех стимулов, для которых наблюдалось двухмодальное распределение, частотный контур ха-

рактируется не менее чем двумя признаками — значениями  $z_n^*$  и  $z_k^*$ . Случаи, в которых наблюдается одномодальное распределение  $F_T$ , допускают две интерпретации. Либо слух выделяет и измеряет в контуре только одно значение частоты, которое однозначно характеризует высоту стимула, как и для стационар-

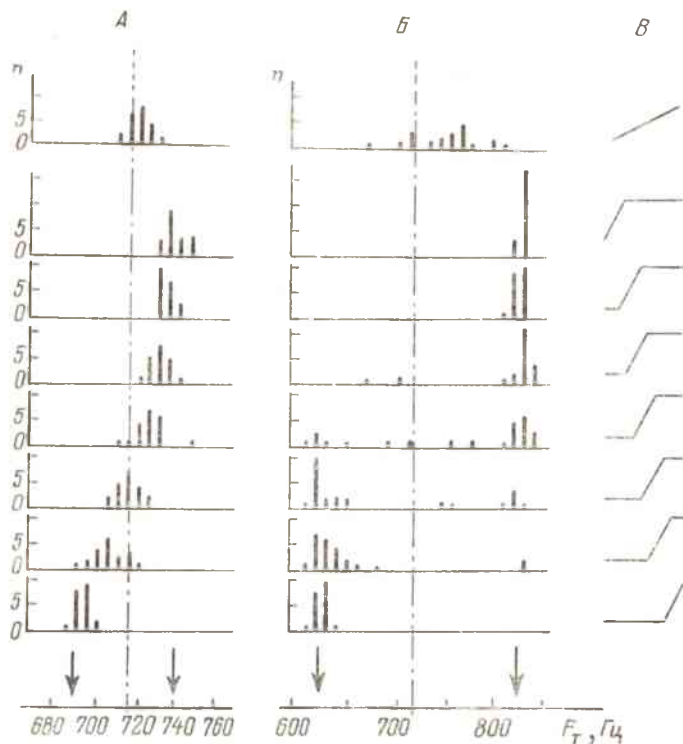


Рис. 12.9. Результаты опытов по уравниванию высот постоянного тона и стимулов с разной формой контуров изменения частоты. По [402].

На А и Б — распределения частот постоянного тона, равного по высоте переменным стимулам длительностью 40 мс; по оси абсцисс — частота, стрелками отмечены начальные и конечные значения частоты контура, штрихпунктирными линиями указаны средние значения частоты контура; по оси ординат — число случаев установления данной частоты  $F_T$ .  
На В — форма контуров переменных стимулов.

ного тона, либо он выделяет два значения,  $z_n^*$  и  $z_k^*$ , но при сравнении такого глайда с чистым тоном человек по каким-то правилам ориентируется только на одно из них или на некоторую функцию от обоих.

Первая из этих интерпретаций кажется сомнительной по следующим соображениям. Стимулы, у которых короткий переход предшествует стационарному участку, при уравнивании их по высоте с чистым тоном характеризуются одномодальным распределением со средним значением, равным частоте стационарного

участка. Пример этого рода данных наблюдается на рис. 12.9, *Б* для контура подобного типа. Из этого должно следовать, что слух выделяет в таком контуре только одну конечную точку. В то же время, как можно видеть по данным о восприятии мягкости синтетических речевых сигналов, контур второй форманты которых имеет аналогичную форму, кроме частоты стационарного участка, слухом измеряется и начальная частота перехода.

Против однозначности высоты глайдов с малым диапазоном изменения частоты (меньшим  $\Delta F_I$ ) свидетельствует эффект увели-

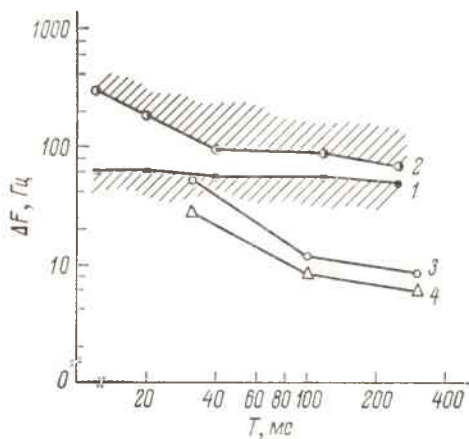


Рис. 12.10. Условия обнаружения разницы между начальными и конечными значениями частоты ( $\Delta F$ ) в изменяющихся сигналах.

По оси абсцисс — длительность стимулов; по оси ординат — разность между начальным и конечным значениями частоты. 1 — критическое значение  $\Delta F_I$ , при котором наблюдается одномодальная форма распределений значений частоты постоянного тона  $F_T$  (см. рис. 12.9); 2 — критическое значение  $\Delta F_{II}$ , при котором распределения  $F_T$  имеют двухмодальную форму. По [402]. 3 и 4 — пороговые значения  $\Delta F$ , при которых обнаруживаются различия краевых частот в стимулах с непрерывной (3) или ступенчатой (4) формой контура. По [509].

чения разброса (стандартного отклонения) значений  $F_T$ , наблюдавшийся Набелеком и др. для 100%-х глайдов при увеличении их длительности (для фиксированного  $\Delta F$ ). Заметим, что для восприятия стационарных тонов в аналогичной ситуации — при уравнивании двух тонов по высоте — характерна обратная закономерность [101, 395].

Учитывая эти факты, а также то, что процедура уравнивания высоты стационарного и переменного сигналов является трудной задачей для человека, допускающей сложные правила ее решения, следует скорее всего думать, что вторая интерпретация является более правильной.

Поддержкой этой точки зрения могут служить также результаты работ [473, 509], в которых исследовались свойства слухо-

вого обнаружения изменений частоты в сигналах, изменяющихся по законам, близким к тем, что и в стимулах Набелека и др. В указанных экспериментах испытуемым предъявлялась последовательность двух стимулов — стационарный тон и глайд, — при случайном порядке их следования. Управляя частотой  $F_t$  постоянного тона, испытуемые сначала добивались наибольшего сходства обоих сигналов по высоте, а затем должны были ответить, в каком из двух стимулов ощущается изменение.

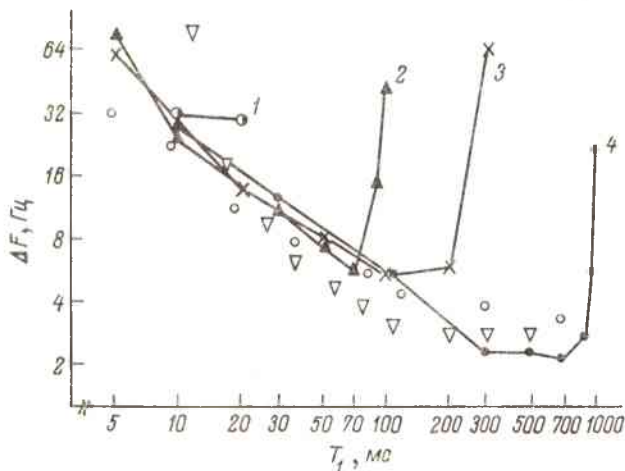


Рис. 12.11. Зависимость порога обнаружения разницы краевых значений частоты в изменяющемся сигнале от длительности начального стационарного участка. По [509].

По оси абсцисс — длительность стационарного участка в начале стимула со ступенчатым контуром  $f(t)$ ; по оси ординат — пороговая разность краевых значений частоты. Параметр кривых — полная длительность стимула ( $T$ ): 1 — 32, 2 — 102, 3 — 302, 4 — 1000 мс. На графике приведены пороги обнаружения различий по частоте для пары стационарных тонов в зависимости от длительности. По [101] — кружки и [510] — незачерненные треугольники.

Экспериментатор устанавливал разные значения перепада частоты на переходах ( $\Delta F$ ) при постоянном значении  $F_n = 1000$  Гц. По ответам испытуемых определялись пороги обнаружения изменений частоты —  $\Delta F$ , при которых изменения правильно обнаруживались в 75% случаев. Были определены пороги в зависимости от общей длительности стимулов ( $T$ ), от длительности перехода ( $T_2$ ) и от длительности первого и второго стационарных участков ( $T_1$  и  $T_3$ ) в стимулах со ступенчатым контуром.

На рис. 12.10 приведены пороговые значения  $\Delta F$ , определенные для стимулов с непрерывными и ступенчатыми контурами. В последнем случае  $T_2 = 1/3 T$  при  $T_1 = T_3$ . Видно, что все значения порогов  $\Delta F$  лежат существенно ниже кривой  $\Delta F_1(T)$ . Из этого следует, что в случаях, когда из результатов опытов по установлению сходства глайда и стационарного тона зависи-

мость  $f(t)$  должна предположительно описываться только одним значением  $z^*$ , на самом деле слух выделяет по крайней мере два последовательных значения  $z_i$  и  $z_{i+1}$ , не равных друг другу. Обнаружение изменений можно трактовать как результат сравнения этих величин друг с другом, при котором решающая система регистрирует их неравенство. В цитируемых работах были получены также данные о том, что пороги обнаружения изменений частоты для стимулов со ступенчатой формой контура (рис. 12.9, В) уменьшаются при увеличении длительности крайних стационарных отрезков (рис. 12.11). При этом оказывается, что характер зависимости от  $T_1$  и значения порогов  $\Delta F$  очень хорошо совпадают с известными данными о порогах обнаружения различий по частоте, определяемых в условиях сравнения двух последовательных стационарных стимулов. На основании этого факта можно сделать вывод о том, что обнаружение изменений в данном случае обуславливается сравнением частоты начального и конечного участков контура, т. е. что  $z_i = z_n^*$  и  $z_{i+1} = z_k^*$ .

Таким образом, рассмотренные на протяжении трех предыдущих разделов данные заставляют предполагать, что в процессе обработки сигналов с изменяющимся во времени спектром слуховая система выделяет и измеряет частоту спектрального максимума в начале и в конце отрезка сигнала, а также частоту в экстремальных точках контура  $f(t)$ .

Определение моментов измерения краевых значений частоты, вероятнее всего, связано с процессом сегментации, осуществляемой в слухе при анализе огибающей интенсивности сигнала, а измерение самих значений параметра сопровождается усреднением на некотором интервале времени.

Для дальнейшего понимания слухового описания контура  $f(t)$  наиболее важными представляются вопрос о механизме выделения экстремумов и выяснение величины интервалов, в течение которых измеряется частота.

### 12.3. СЛУХОВОЙ АНАЛИЗ «ПРОИЗВОДНОЙ» ЧАСТОТНОГО КОНТУРА

Специалисты по психоакустике уже давно высказывают предположение о том, что скорость изменения спектральных составляющих в звуках типа «глиссандо» имеет свой слуховой коррелят — субъективную переменную, поддающуюся количественной оценке [464]. Экспериментальные данные, обсуждаемые в настоящем разделе, свидетельствуют о том, что человек в своих решениях об акустических сигналах с переменным спектром действительно основывается на данных о направлении или величине скорости изменения частоты спектральных максимумов.

Есть основания считать, что выделение такого признака сиг-



нала, как скорость изменения частоты, нельзя объяснить действием какого-то одного, общего для всех случаев механизма. Поддержкой этому могут служить нейрофизиологические данные работы [318].

Допустимы по крайней мере два разных способа слухового измерения скорости изменения частоты.

1. Поскольку слух выделяет и измеряет краевые значения частоты в отдельной посылке с однонаправленным контуром  $f(t)$  и может измерять ее длительность, то эти три величины могут служить исходными параметрами для определения диапазона, направления и скорости изменения частоты. В данном случае непрерывно изменяющийся сигнал подвергается такой же обработке, как последовательность из двух стационарных посылок. Как можно было видеть по предыдущему разделу, данные по обнаружению непрерывных изменений частоты в тональных стимулах [473, 509] интерпретируются именно с точки зрения такого способа анализа.

2. Второй способ предполагает в слуховой системе возможность вырабатывать некоторым аналоговым путем сигнал, пропорциональный производной частотного контура,  $\dot{f}(t)$ , независимо от измерения координат спектральных максимумов в краевых точках. Для этого предполагается наличие специализированных устройств, выполняющих действие, аналогичное дифференцированию. Входной информацией для них должен служить либо некоторый аналоговый сигнал, монотонно связанный с мгновенным значением частоты (координаты) максимума (функцией  $z^*(t)$ ), либо непосредственно огибающая слухового спектра  $g(z, t)$ .

В пользу предположения о выделении слухом аналоговой переменной, близкой по смыслу к производной  $\dot{f}(t)$ , говорят результаты нейрофизиологических исследований (см. главу 9), которые показывают, что реакции отдельных нейронов центральной слуховой системы систематически изменяются в зависимости от величины и направления скорости изменения частоты тонов. Этим устанавливается возможность осуществления в слухе так называемого измерительного преобразования для  $\dot{f}(t)$ . По типам наблюдаемых нейронных ответов можно предполагать, что слух может измерять знак и величину производной  $\dot{f}(t)$  и длительность участка, где производная не равна нулю, или фиксировать момент появления такого участка.

Естественно ожидать, что в результате психоакустических исследований должно быть выяснено, в какой мере при слуховом восприятии реализуется возможность таких преобразований и каковы реальные количественные характеристики анализа производной  $\dot{f}(t)$ . К сожалению, существующие в настоящее время данные могут ответить только на очень немногие вопросы в этом плане.

### 12.3.1. ВОСПРИЯТИЕ НАПРАВЛЕНИЯ ИЗМЕНЕНИЯ СПЕКТРА В РЕЧЕВЫХ СИГНАЛАХ

Экспериментальные данные, свидетельствующие о том, что человек использует направление изменения частоты спектрального максимума в качестве полезного признака при распознавании фонем, были получены при исследовании восприятия формантных переходов — быстрых однонаправленных изменений частот формант на начальном или конечном участках гласного [220, 228, 349, 351, 369, 411, 481].

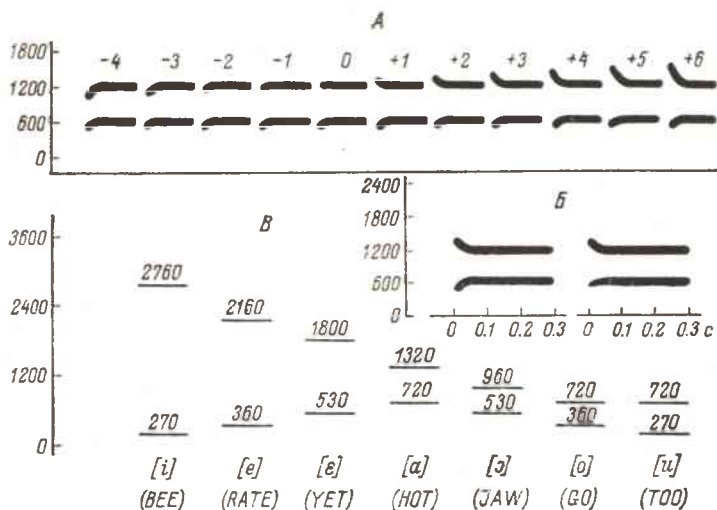


Рис. 12.12. Характеристики стимулов, применявшихся в опытах по исследованию восприятия формантных переходов. По [351].

По оси ординат — частота формант (в Гц). На А — варианты переходов второй форманты ( $F_2$ ); на Б — два варианта переходов первой форманты ( $F_1$ ); на В — комбинации значений  $F_1$  и  $F_2$  на стационарном участке сигнала и соответствующие им гласные.

При прослушивании естественных гласных звуков, выделенных из окружения согласных (обычно исключаются участки сигналов, предшествующие быстрому нарастанию амплитуды, если исследуются начальные переходы, или непосредственно следующие за спадом, если исследуются конечные), оказывается, что такие сигналы воспринимаются как слоги, в которых хорошо распознается признак места образования согласного [95, 171, 465].

Различение согласных по начальным переходам формант подробно исследовалось в работах сотрудников Хаскинских лабораторий на синтетических стимулах, параметры которых схематически представлены на рис. 12.12 (из работы [351]). Спектр стимулов имел двухформантную структуру, на рис. 12.12, В показаны значения  $F_1$  и  $F_2$  на стационарном участке. Наличие перехода первой форманты определяет главным образом слогоподобный

арактер звука, а направление перехода имеет значение для различения согласных по признаку глухости-звонкости (см. главу 6).

Место образования согласных зависит от характера перехода второй форманты, варианты которого приведены на рис. 12.12, А. Все предъявляемые стимулы испытуемые должны были идентифицировать как слоги, начинающиеся со звонкого согласного [b], [d] или [g] (левый вариант перехода  $F_1$  на рис. 12.12, Б), либо с глухого согласного [p], [t], [k] (при правом варианте перехода  $F_1$ ). Результаты идентификации представлены на рис. 12.13. Можно убедиться, что при возрастании  $F_2$  на переходе в качестве начального согласного выбирается губной звук [b] или [p], при обратном направлении — язычный [d], [t] или [g], [k].<sup>1</sup>

Если на основании направления переходов  $F_2$  человек отличает группу губных согласных от язычных, то, как показывает ряд экспериментов [293, 298, 366, 382], для различения язычных согласных между собой он использует характер перехода третьей форманты.

Интересно, что человек способен различать направление изменения спектрального максимума независимо от того, с помощью периодического или шумового возбуждения образован этот максимум. В экспериментах Кожевникова и др. [81] в качестве стимулов применялись последовательности, состоящие из короткой посылки полосового шума и следующего за ней гласноподобного звука с гармоническим спектром, характеризующимся одним максимумом.

Фонемная граница между [p] и [t], определенная как среднее значение центральной частоты полосового шума  $F_{ш}$ , совпадает с частотой спектрального максимума гласного  $F_r$  (рис. 12.14) — факт, свидетельствующий о том, что в этих опытах фонемная классификация основывается на направлении различия между  $F_{ш}$  и  $F_r$ . Существенно, что наблюдаемые при этом небольшие отклонения границы от  $F_r$  носят закономерный характер, который говорит о том, что порог различения направления зависит от значений частоты центрального максимума на участке гласного.

На рис. 12.15, А приведены величины относительных отклонений границ  $((F_r - F_{ш})/F_r$  в %) в зависимости от частоты максимума гласного. Подобная закономерность проявляется также при анализе данных Либермана и др. [351] о различении согласных по признаку губной—язычный на основании начальных переходов  $F_2$  (рис. 12.15, Б). При составлении этого графика анализировались кривые идентификации [p] или [b], при-

<sup>1</sup> Сотрудники Хаскинских лабораторий попытались объяснить полученные данные с помощью чрезвычайно сложной гипотезы «локуса» [296], которая, однако, имеет смысл лишь в плане выбора параметров переходов формант при синтезе согласных и не имеет реальных оснований рассматриваться как один из способов слуховой обработки (описание гипотезы «локуса» и критику ее см. [159]).

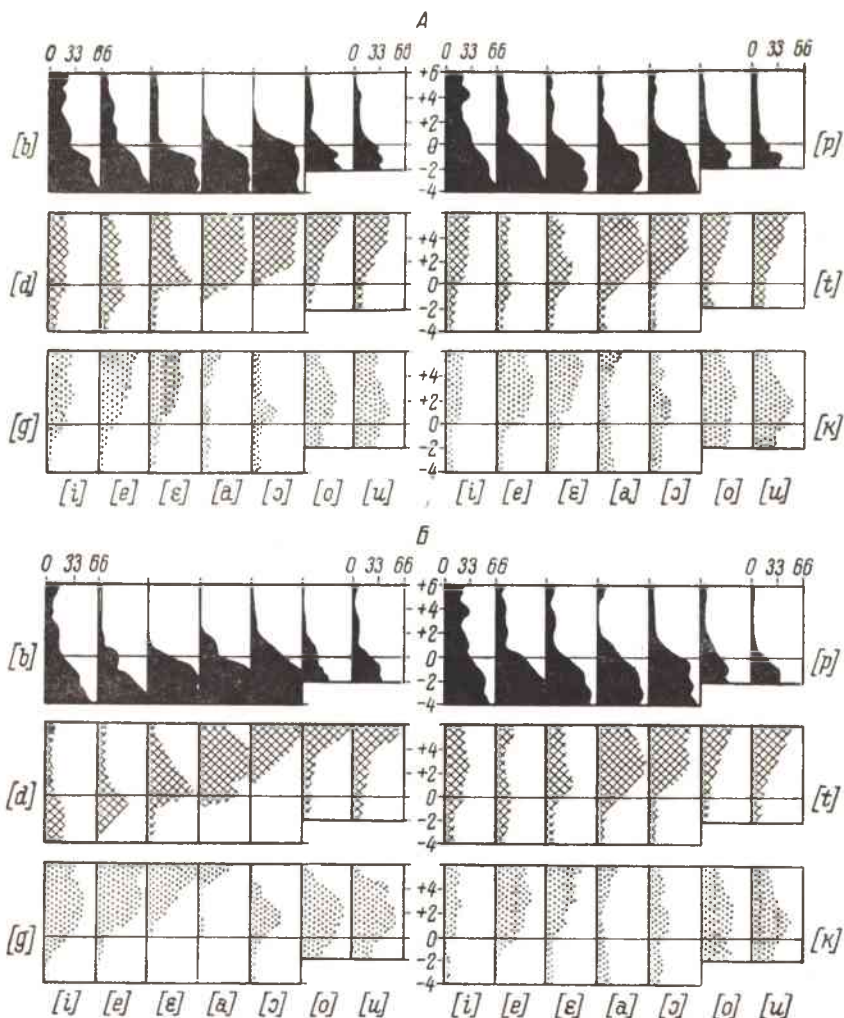


Рис. 12.13. Результаты идентификации стимулов, показанных на рис. 12.12.

На А — для стимулов с явно выраженным переходом  $F_1$ . На Б — для стимулов с мало выраженным переходом  $F_1$ . Левые части рисунков соответствуют случаю, когда испытуемым было разрешено пользоваться в ответах согласными [b], [d], [g], правые части — согласными [p], [t], [k]. Каждый столбец (три расположенных друг под другом прямоугольника) соответствует определенному гласному, фонетическое обозначение которого приведено под столбцом. Внутри каждого прямоугольника показана зависимость между значением перехода второй форманты (цифры по вертикали в центре рисунка) и числом идентификаций стимула (цифры по горизонтали) с тем согласным, который указан слева или справа у соответствующей строки. Так, рельеф в верхнем левом прямоугольнике (строка [b] столбец [i] на А) показывает, что при значении перехода, равном -4, испытуемые опознают согласный как [b] в 66 случаях; при повышении значения перехода второй форманты количество опознаний [b] уменьшается.

веденные на рис. 12.13. Определялась разность между начальным и стационарным значениями  $F_2$ , соответствующая 50% распознавания [p] или [b], отнесенная к  $F_2$  на стационарном участке.

Направление изменения формантных переходов в конце гласного также определяет признак места образования согласных [95]. При этом если человек слушает изолированные слоги, то конечные согласные опознаются по переходам более точно, чем начальные [465]. Однако правила различения согласных на конце проти-

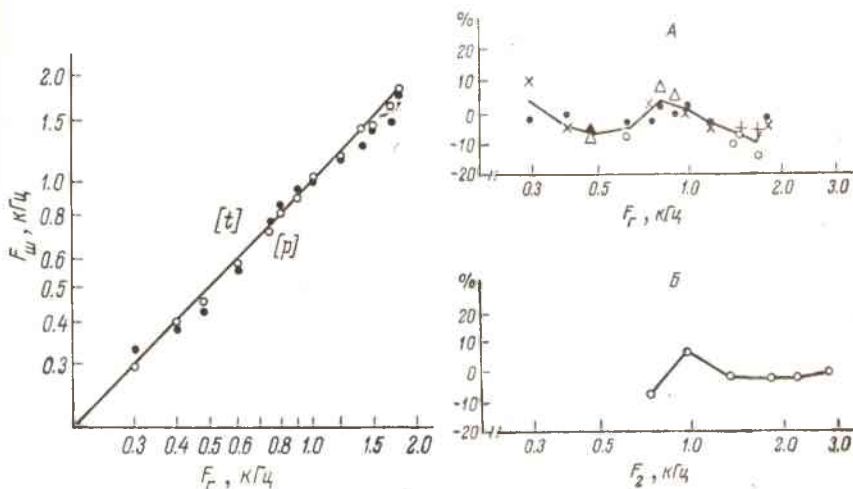


Рис. 12.14. Граница между [p] и [t]. По [81].

По оси абсцисс — частота спектрального максимума гласного; по оси ординат — средняя частота полосового шума. Точки и кружки — данные разных испытуемых.

Рис. 12.15. Отклонения средних значений фонетических границ при восприятии признака места артикуляции согласных. По [81].

На А — отклонение среднего значения границы между [p] и [t] от частоты спектрального максимума гласного; по оси абсцисс — частота максимума в спектре гласного; по оси ординат — отклонение среднего значения границы от  $F_2$ ; данные 3 испытуемых. На Б — величина разности между значениями границы «губные-язычные согласные» и значениями частоты второй форманты на стационарном участке гласного (данные [381]). По оси абсцисс — частота второй форманты; по оси ординат — значение  $(F_{2H} - F_2)/F_2$ .

воположны тем, которые наблюдаются в начале гласного: так, губные звуки характеризуются понижением частоты, а язычные — повышением.

Направление изменения спектра на переходах выступает как специфический признак сигнала, а сам факт однонаправленного изменения, очевидно, выделяется слуховой системой как отдельное событие. Эта точка зрения иллюстрируется в следующем опыте Рэнда [438].

Рис. 12.16 показывает схематические характеристики стимулов и способ их предъявления. На одно ухо, например на правое, подавался сигнал, состоящий из двух сегментов: первый сегмент характеризовался спектром с одним максимумом в области пер-



вой форманты гласного, спектр второго имел структуру, характерную для трехформантного стационарного гласного [a] — схема 1 на рис. 12.16, А (такой звук сам по себе воспринимается как [ba]). На другое ухо — левое — поступала короткая посылка звука, спектр которой соответствовал начальным переходам  $F_2$  и  $F_3$ , характерным для [ba] — 2, [da] — 3 и [ga] — 4.

В этих условиях испытуемые слышали в правом ухе совершенно отчетливо полный слог с согласным, соответствующим направлению переходов  $F_2$  и  $F_3$ , и дополнительно еще в левом ухе — короткий верещащий звук. При разных направлениях переходов  $F_2$  и  $F_3$  различение согласных по признаку места образования осуществлялось с вероятностью 100%. Распознавание было воз-

можным даже в том случае, когда нарушалась синхронность поступления обеих частей: посылку, содержащую переход, можно было сдвигать по отношению к началу сигнала на другом ухе в пределах  $\pm 25$  мс.

Однако фонетический признак места образования согласного

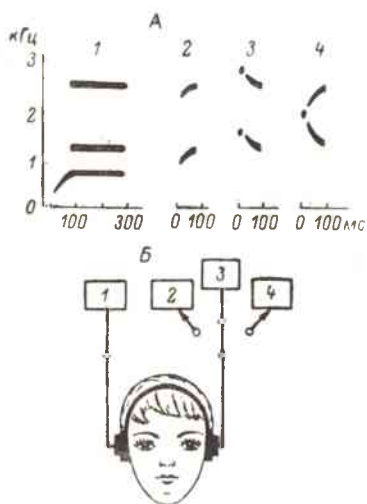


Рис. 12.16. Схематическое изображение параметров стимулов (А) и способ их предъявления (Б) при дихотическом восприятии формантных переходов. По [438].

По оси абсцисс — время; по оси ординат — частоты формант. Остальные обозначения см. в тексте.

определяется не только таким простым событием, как направление изменения спектра. Требуется еще наличие других событий, таких, например, как начало и конец гласного. Известно, что если человеку предъявлять изолированный отрезок сигнала с такими характеристиками, как на переходах, то он вообще не воспринимается как речевой [326].

Таким образом, существующие экспериментальные данные заставляют считать, что факт определения слухом направления временного изменения спектра не вызывает сомнения. Это обстоятельство ставит исследователей непосредственно перед задачей выяснения конкретного способа его слухового измерения.

Можно указать пока только на одну работу Стивенса [483] как на первую попытку моделирования механизма определения направления изменения спектра на переходах. В основе модели Стивенса лежит механизм сравнения энергии в соседних частотных полосах, производимого через малые (10 мс) интервалы времени. Эту функцию выполняет система детекторов трех типов,

один из которых должен срабатывать на такой сигнал, в котором концентрация энергии в низкочастотной полосе следует за концентрацией энергии в высокочастотной, другой — на сигнал с обратным направлением спектра и третий — на сигнал с расходящимся спектром.

Насколько верна модель Стивенса, сказать пока трудно: слишком мало сведений мы имеем о реальном слуховом анализе временных изменений спектра. Для дальнейшего его количественного исследования наиболее удобным методическим приемом кажется использование синтетических речеподобных стимулов, содержащих информацию о признаке места образования согласных.

### 12.3.2. ВОСПРИЯТИЕ СКОРОСТИ ИЗМЕНЕНИЯ ЧАСТОТЫ ТОНА

Психоакустические данные, показывающие, что скорость изменения частоты тона может служить различительным признаком сигналов, были получены в работах Поллака [429], Набелека и Хирша [401].

В обеих работах в качестве стимулов использовались отдельные посылки тонов, закон изменения частоты которых указан

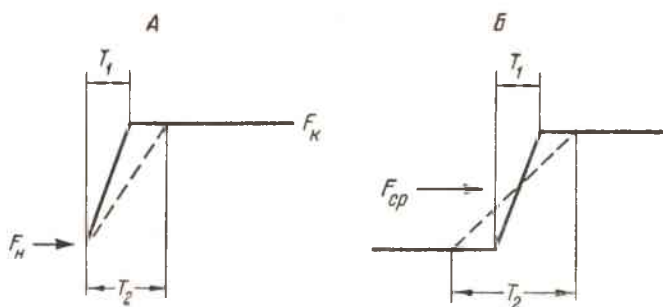


Рис. 12.17. Схема параметров стимулов, используемых при измерении порогов различения скорости изменения частоты.

На А — по [401], на Б — по [429].  $T_1$  — длительность перехода стандартного стимула;  $T_2$  — длительность перехода переменного стимула;  $F_n$ ,  $F_k$ ,  $F_{cp}$  — начальное, конечное и среднее значения частоты соответственно.

на рис. 12.17. В экспериментах Поллака контур  $f(t)$  имел изменяющийся участок — переход — в центре посылки (рис. 12.17, Б), в экспериментах Набелека и Хирша переходом начинался сигнал (рис. 12.17, А).

Предъявляемые для сравнения стандартный и переменный стимулы (использовался классический метод парного сравнения) имели фиксированные значения  $F_n$  и  $F_k$ , а различия их по скорости перехода задавались за счет длительности перехода переменного стимула ( $T_2$ ). Дифференциальный порог определялся по

значению  $\Delta T = |T_2 - T_1|$ , при котором испытуемые в 75% случаев обнаруживали различие стимулов. По этому значению можно затем вычислить порог различения по скорости:  $\Delta \dot{f} = \Delta F / \Delta T$ .

При такой процедуре не исключена возможность для испытуемых различать стимулы по длительности переходов, и вопрос о том, пользуется ли человек этой возможностью, может быть решен при анализе величин дифференциальных порогов по одному и другому параметрам ( $\Delta T$  и  $\Delta \dot{f}$ ).

В опытах Поллака использовались стимулы только с возрастающим направлением переходов, параметры которых варьиро-

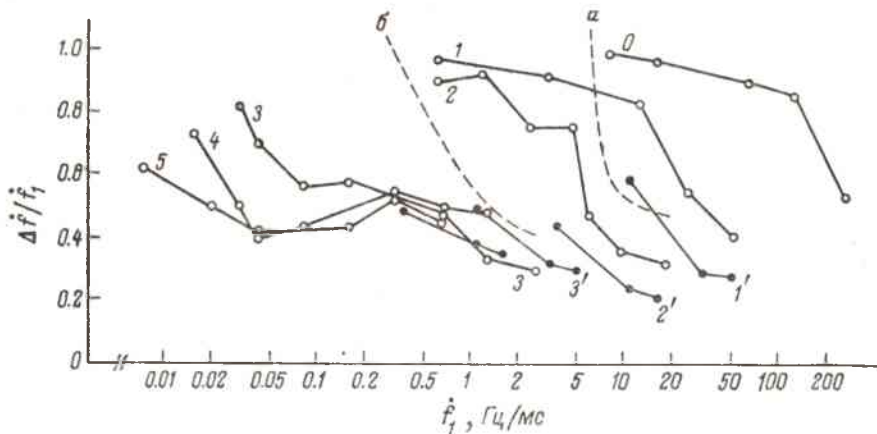


Рис. 12.18. Пороги различения скорости изменения частоты тонов.

По оси абсцисс — скорость изменения частоты на переходах в тестовом стимуле; по оси ординат — относительный порог различения скорости изменения частоты. Параметр кривых (сплошные линии) — длительность переходов ( $T_1$ ): 0 — 2 мс, 1 и 1' — 10, 2 — 27, 2' — 30, 3 — 100, 3' — 200, 4 — 400, 5 — 870 мс. Штриховыми линиями обозначены относительные пороги различения длительности переходов,  $\Delta T/T_1$ : α — для  $T_1 = 27$  мс, β — для  $T_1 = 200$  мс. Остальные обозначения см. в тексте.

вали в широких пределах:  $\Delta F = 4 \div 512$  Гц,  $T_1 = 0,1 \div 870$  мс. Средняя частота стимулов была одна и та же ( $F_{\text{ср}} = 707$  Гц), длительность стационарных участков равна 1000 мс у каждого.

В опытах Набелека и Хирша [401] положение стимулов на шкале частот, характеризующееся более высокой частотой ( $F_{\text{к}}$  для переходов с возрастающим направлением и  $F_{\text{н}}$  для переходов с убывающим), соответствовало 250, 1000 и 4000 Гц. В каждом положении диапазон изменения частоты на переходах принимал три значения:  $F_{\text{к}}/F_{\text{н}} = 2:1$ ,  $3:2$  и  $9:3$  (при направлении другого знака — соответственно обратные отношения). Длительность переходов  $T_1 = 10 \div 300$  мс при соотношении ее с длительностью стационарного участка 1:4.

На рис. 12.18 представлены зависимости относительных порогов различения скорости изменения частоты  $\Delta \dot{f}/\dot{f}_1$  от  $f_1$ , при

построении которых использованы графические данные из работы Поллака (кружки) и табличные данные работы Набелека и Хирша (точки) для стимулов из области 1000 Гц. Учитывая разницу в методиках, можно считать, что соответствие данных обеих работ достаточно хорошее.

Прежде чем обсуждать характер данных, остановимся на вопросе о том, не могло ли различение стимулов определяться в данном случае длительностью переходов. Для этого сравним между собой значения относительных порогов различения скорости  $\Delta f/f_1$  и длительности  $\Delta T/T_1$ , считая эти величины характеристиками относительной погрешности измерения слухом исходных параметров  $f$  и  $T$ . На рис. 12.18 штриховыми линиями указаны зависимости  $\Delta T/T_1$  от  $f_1$  для стимулов с  $T_1=27$  мс и 200 мс, по данным Поллака. Видно, что с уменьшением значения  $f_1$  относительные пороги различения по длительности становятся очень большими и существенно превышают относительные пороги различения скорости. Там, где значения  $\Delta T/T_1$  приближаются к 1, можно считать, что информация о длительности стимулов практически отсутствует (длительность не измеряется).

Кроме того, даже минимальные значения относительных порогов различения длительности переходов, полученные как в экспериментах Поллака, так и Набелека и Хирша, оказались существенно большими, чем известные значения порогов различения длительности отдельных посылок [173] или паузы между тонами разной частоты [408]. К тому же в последнем случае зависимость порогов от разности частот, окружающих паузу, имеет противоположный характер, чем соответствующая зависимость для переходов.

На этом основании можно считать, что различение стимулов в подавляющем большинстве случаев было основано на слуховом измерении скорости изменения частоты на переходах и приведенные кривые отражают свойства системы этого измерения. Судя по тому, что зависимость порогов  $\Delta f/f_1$  от  $f_1$  представляет собой семейство кривых, параметром которых выступает длительность переходов, можно сказать, что субъективная переменная  $f^*(t)$  — субъективная скорость — не есть отображение мгновенного (пикового) значения  $f(t)$ . В противном случае зависимость описывалась бы одной общей кривой.

Конкретный характер влияния длительности переходов на пороги различения скорости можно проследить по графикам на рисунках 12.19 и 12.20, на которых представлены зависимости  $\Delta f/f_1$  от  $T_1$ . Они позволяют выделить два диапазона значений  $T_1$ , в пределах которых пороги уменьшаются с увеличением длительности переходов. Первый диапазон соответствует длительностям примерно до 30 мс, второй — до 200 мс. Различие их сказывается в том, что уменьшение порогов в пределах первого происходит с более медленной скоростью, чем в пределах второго. Кроме того, в пределах диапазона малых длительностей более медленные

переходы (переходы с малым значением  $\Delta F$ ) практически не различаются по скорости.

Можно предположить, что в характере различения коротких переходов проявляются инерционные свойства либо спектрального анализатора, обуславливающие сглаживание исходного контура  $f(t)$ , либо дифференцирующего звена системы анализа его производной  $f'(t)$ . Обе причины должны приводить к тому, что при одной и той же исходной скорости стимулов отклик системы ее измерения в случае коротких переходов будет меньше по амплитуде, чем в случае длинных. При наличии порога чувствительности

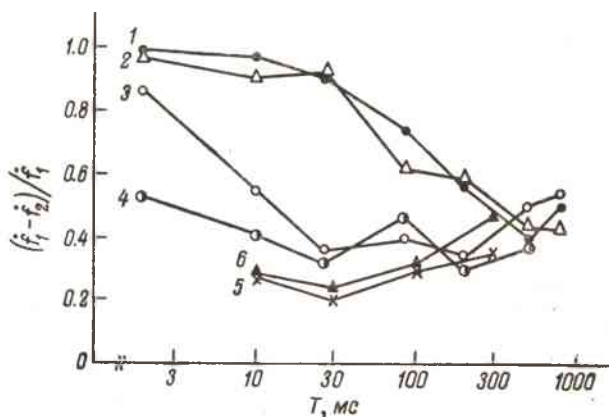


Рис. 12.19. Зависимость порогов различения скорости изменения частоты тона от длительности переходов.

По оси абсцисс — длительность перехода стандартного стимула; по оси ординат — относительный порог различения скорости изменения частоты. Параметр кривых — диапазон изменения частоты на переходе ( $\Delta F$ ): 1 — 16, 2 — 32, 3 — 256, 4 — 512 Гц (данные [429]); 5 — 500, 6 — 334 Гц (данные [401]).

измерителя медленные и короткие переходы вообще не должны давать никакого сигнала на его выходе и не должны различаться по скорости.

Что касается эффекта увеличения различимости более длинных переходов в пределах диапазона длительностей до 200 мс, то его можно было бы объяснить, например, тем, что измерение скорости сопровождается усреднением во времени субъективной производной  $f^*(t)$  одного знака.

Однако в этом случае допустима также и другая интерпретация. Подобного эффекта накопления информации во времени можно ожидать, если изменения частоты «отслеживаются» с помощью временного сканирования мгновенных значений функции  $f(t)$  [429]. Допущение такого механизма анализа применительно к длинным переходам не лишено оснований, так как обычно предполагаемый интервал сканирования составляет  $200 \div 250$  мс.

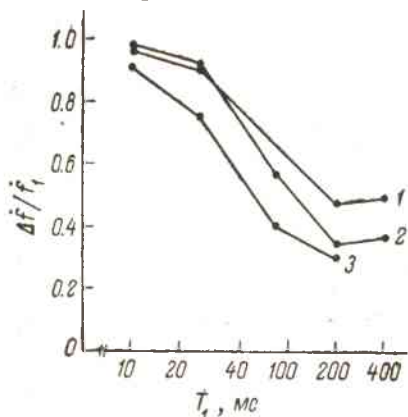


Другим фактором, влияющим на различие скорости переходов, является диапазон изменения частоты  $\Delta F$  (или значение исходной скорости). Особенно заметно влияние  $\Delta F$  в характере зависимости порогов от  $T_1$ , приведенных на рис. 12.19. Все кривые четко разделяются на две группы в зависимости от параметра  $\Delta F$ : одну группу составляют кривые со сравнительно малыми значениями  $\Delta F$ , другую — с большими.

Границу между группами указать точно нельзя, так как,

Рис. 12.20. Зависимость порога различения скорости изменения частоты тона от длительности перехода.

По оси абсцисс — длительность перехода стандартного стимула; по оси ординат — относительный порог различения скорости изменения частоты. Параметры кривых — диапазон значений  $f'$  (по данным [425]): 1 —  $0.6+0.64$ ; 2 —  $1.2+1.5$ ; 3 —  $2.4+3.2$  Гц/мс.



если судить по результатам Поллака, кривая с параметром  $\Delta F = 128$  Гц попадает в первую группу, тогда как кривая с параметром  $\Delta F = 111$  Гц, соответствующая данным Набелека и Хирша, попадает во вторую. Проявление подобного эффекта связано, очевидно, с тем, что преобразование  $f(t) \rightarrow z^*(t)$  сопровождается усреднением частотного контура по пространственному параметру. Как было показано в главе 10, интегрирование энергии спектральных составляющих в пределах некоторой полосы частот является экспериментально подтвержденным фактом.

Совершенно очевидно, однако, что все предположения о свойствах системы анализа «производной» частотного контура, которые высказаны здесь на основании рассмотренных двух работ, имеют проблематичный характер и требуют более основательной экспериментальной проверки.

### 12.3.3. РОЛЬ СКОРОСТИ ИЗМЕНЕНИЯ СПЕКТРА ПРИ ВОСПРИЯТИИ РЕЧЕВЫХ СИГНАЛОВ

Насколько известно, попытки исследовать восприятие скорости изменения спектра речевых сигналов были приняты всего в двух случаях.

В работе [352] авторам удалось показать, что в зависимости от того, насколько быстро изменяется частота формант на начальном переходе гласного в синтетических слогах, первый звук идентифицируется испытуемыми или как согласный, или как полугласный, или как гласный. В опытах использовались сиг-

налы с двухформантным спектром, временная форма которого схематически представлена на рис. 12.21.

При повышении частоты  $F_2$  на переходах начальный сегмент стимула воспринимается как губной звук: [b], [w] или [u], при понижении  $F_2$  — как язычный: [g], [j] или [i]. Варьируя длительность переходов  $T_n$  в пределах от 10 до 300 мс, при постоянных диапазонах изменения  $F_1$  и  $F_2$  ( $\Delta F_1 = 360$  Гц,  $|\Delta F_2| = 720$  Гц) и неиз-



Рис. 12.21. Схема синтеза стимулов, использованных при исследовании восприятия скорости изменения частоты формант. По [352].

По оси абсцисс — время; по оси ординат — частота,  $T_n$  — длительность формантного перехода. I и II — разные типы формантных переходов.

менном спектре стационарного участка гласного [e], авторы получили следующие значения фонемных границ.

Для губных:

$$[b] - [w]: T_n = 60 \div 80 \text{ мс } (\dot{F}_1 = 4.5 \div 6 \text{ Гц/мс}, \dot{F}_2 = 9 \div 12 \text{ Гц/мс}), \\ [w] - [u]: T_n = 150 \div 200 \text{ мс } (\dot{F}_1 = 1.8 \div 2.4 \text{ Гц/мс}, \dot{F}_2 = 3.6 \div 4.8 \text{ Гц/мс}).$$

Для язычных:

$$[g] - [j]: T_n = 95 \text{ мс } (\dot{F}_1 = 3.8 \text{ Гц/мс}, \dot{F}_2 = 7.6 \text{ Гц/мс}), \\ [j] - [i]: T_n = 210 \div 240 \text{ мс } (\dot{F}_1 = 1.5 \div 1.7 \text{ Гц/мс}, \dot{F}_2 = 3 \div 3.4 \text{ Гц/мс}).$$

Граничные значения  $T_n$  определены по кривым идентификации, соответствующие им значения скоростей вычислены по соотношениям  $\dot{F}_1 = \Delta F_1 / T_n$ ,  $\dot{F}_2 = \Delta F_2 / T_n$ .

Интересно, что хотя ответы испытуемых группируются по классам достаточно компактно, границы имеют неоднозначный характер: для губных звуков все значения меньше, чем для язычных. Кроме того, в тех случаях, когда испытуемым был предъявлен такой набор стимулов, который допускал идентификацию только по двум классам — согласный или полугласный, — границы оказались сдвинутыми влево:

$$[b] - [w]: T_n = 40 \text{ мс } (\dot{F}_1 = 9 \text{ Гц/мс}, \dot{F}_2 = 18 \text{ Гц/мс}), \\ [g] - [j]: T_n = 50 \div 60 \text{ мс } (\dot{F}_1 = 6 \div 7.2 \text{ Гц/мс}, \dot{F}_2 = 12 \div 14.4 \text{ Гц/мс}).$$

Судя по материалам предыдущего раздела, такие же скорости изменения частоты в тональных сигналах поддаются измерению в слухе, и результаты этого измерения могут использоваться в процедуре различения стимулов как исходная информация.

Однако авторам работы [352] не удалось получить экспериментального подтверждения того, что идентификация речевых стимулов в их экспериментах обуславливалась измерением именно скорости изменения частоты формант, а не длительности переходов.

В двух других работах на эту тему [496, 497] исследовался вопрос о роли изменений частоты первой форманты для различения между собой сочетаний [aba], [awa], [aua] и [aaa]. Характер временного контура  $F_1(t)$  и его параметры схематически представлены на рис. 12.22. Остальные три форманты оставались неизменными на протяжении стимула, общая длительность которого составляла 500 мс. Вариации скорости переходов осуществлялись путем изменения  $F_1$  на среднем участке, а также за счет смещения положения моментов  $t_1$  и  $t_2$  (так что могла меняться и длительность участка между ними). Длительности

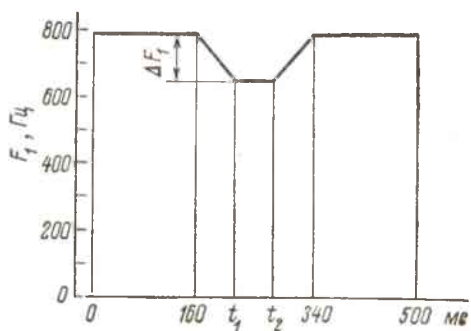


Рис. 12.22. Схема изменения параметров контура  $F_1(t)$  синтетических стимулов типа ГСГ. По [496].

По оси абсцисс — время; по оси ординат — частота первой форманты. Остальные обозначения см. в тексте.

начального и конечного стационарных участков гласного сохранялись неизменными и равными друг другу.

При анализе и интерпретации результатов опытов автор исходил из предположения о том, что классификация стимулов основывается на измерении двух параметров: диапазона и скорости изменения  $F_1$  на переходах ( $\Delta F_1$  и  $\dot{F}_1$ ). В соответствии с этим он рассматривает фонемные границы на плоскости этих двух параметров и предлагает такую аппроксимацию границ, как показано на рис. 12.23.

Судя по характеру приведенных данных, имеется некоторая тенденция к зависимости границ от параметра  $\dot{F}_1$ . В большей степени она выражена для категорий [b] и [w], но в целом проявляется довольно слабо. Причина этого, возможно, кроется в том, что в качестве исходного параметра здесь используются мгновенные значения исходной скорости  $\dot{F}_1$ . Действительно, автору [497] удастся добиться большей выраженности эффекта влияния скорости, определяя ее не по идеальному контуру  $F_1(t)$ , а по преобразованному с помощью некоторого формального соотношения, включающего интегрирование  $F_1(t)$  по времени. Кроме скорости изменения частоты  $F_1$ , на различение сигналов влияет также другой параметр, в качестве которого автор предлагает считать диапазон измене-

ния  $\Delta F$  на переходе. Однако вероятнее предполагать, что вторым признаком служит абсолютное значение  $F_1$  на среднем участке, что в наибольшей степени относится к распознаванию стимулов [aaa] и [aau].

Следует сказать, что все три обсуждаемые работы по восприятию речевых сигналов обладают одним недостатком методического характера. Известно, что изменение спектра в области первой

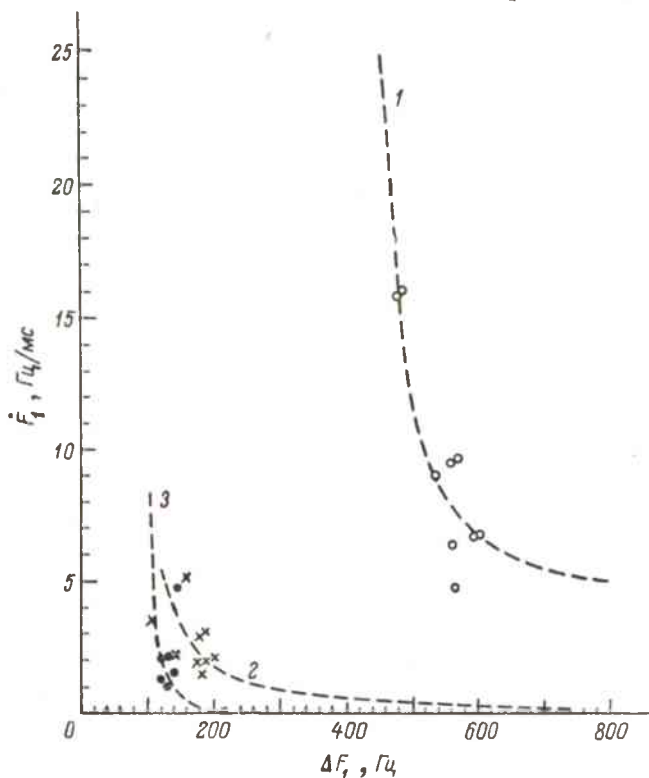


Рис. 12.23. Фонемные границы в плоскости параметров «перепад частоты — скорость изменения частоты» контура  $F_1(t)$ . По [496].

По оси абсцисс — величина частотного перепада; по оси ординат — скорость изменения  $F_1$  на переходах. Кривые — аппроксимация границ по экспериментальным данным; 1 (кружки) — между [b] и [w]; 2 (крестики) — между [w] и [u]; 3 (точки) — между [u] и [a].

форманты приводит к перепаду общей огибающей интенсивности, наличие которого само по себе может обуславливать различие категорий гласных и согласных (см. главу 11). То обстоятельство, что изменение огибающей никак не контролировалось и не учитывалось и все результаты интерпретировались только с точки зрения изменений спектра, мешает правильному пониманию роли последних.

Таким образом, предположение, что скорость изменения спектра в речевом сигнале служит полезным признаком для распознавания таких фонемных категорий, как согласные—полугласные—гласные, пока остается предположением, хотя и вероятным, но требующим экспериментального подтверждения.

Как уже упоминалось выше, в экспериментах по различению стимулов с изменяющимся спектром в качестве возможного различительного признака подозревается длительность переходов. Этим подразумевается, что слуховая система каким-то образом умеет выделять в потоке участок с изменяющимся в одном направлении спектром и измерять его длительность. Иначе говоря, предполагается возможность такого анализа формы контура  $f(t)$ , в результате которого контур разбивается на однородные по некоторому признаку отрезки или на нем выделяются моменты (метки), соответствующие изменению некоторых параметров, например смене знака производной.

С другой стороны, неоднократно высказывалось предположение [159, 482] о том, что быстрые изменения спектра на коротких интервалах, например на переходах между взрывными согласными и гласными, могут вызывать в слуховой системе только одномоментный отклик. Если было бы возможно зафиксировать момент возникновения такого события в потоке, оно могло бы выполнять сегментирующую функцию.

При постановке экспериментов, направленных на исследование восприятия скорости изменения спектра, следует учитывать тот факт, что этот параметр может выполнять двойную функцию в процессе слуховой обработки входного сигнала — служить характеристикой изменения спектра как некоторого события и быть меткой сегментации акустического потока.

#### 12.3.4. ГИПОТЕЗА О ДЕТЕКТОРАХ ВРЕМЕННЫХ ИЗМЕНЕНИЙ ЧАСТОТЫ В СЛУХЕ

Факты, рассмотренные в трех предыдущих разделах, делают достаточно правдоподобным предположение о том, что слуховая система измеряет скорость и знак изменения спектрального максимума.

Вполне вероятно также, что основу одного из вариантов этого измерения составляет аналоговое преобразование, близкое по смыслу к дифференцированию частотного контура  $f(t)$ . Из сказанного следует, что схема модели слухового анализа должна предусматривать соответствующие звенья, формирующие и передающие информацию о производной  $f(t)$  и осуществляющие ее измерение.

За последнее время широкое распространение получила гипотеза о том, что слуховой анализ временных изменений частоты звуковых сигналов осуществляется с помощью специальных устройств — детекторов изменений, — расположенных, вероятнее



всего, в высших отделах слуховой системы (см. главу 9). Следует заметить, что при обсуждении этой гипотезы в литературе [318, 318, 408] допускаются неоднозначные интерпретации понятия детектора изменений. Некоторые авторы подразумевают под этим любое устройство, осуществляющее измерение скорости изменения частоты, независимо от того, какой способ измерения лежит в его основе — дискретный или аналоговый (см. вступление к разделу 12.3). Нам представляется более правомерным использовать понятие «детектор изменений» пусть в условном, но в более ограниченном смысле: только применительно к аналоговому способу анализа производной частотного контура.

Исследование вопроса о детекторах изменений находится пока что в стадии накопления сведений феноменологического характера и проводится, как правило, нейрофизиологическими методами. Известно только, что детекторы располагаются в разных отделах слухового пути и могут выполнять различные функции. Одни из них обнаруживают изменение частоты независимо от направления, другие избирательно реагируют на знак производной, третьи передают информацию о величине скорости и т. д.

Поскольку неизвестно, как такие элементы взаимодействуют между собой и каким образом они участвуют в общей процедуре обработки звуковых сигналов, мы имеем право говорить только о некотором общем канале для слухового анализа производной  $f(t)$ .

Принципиально важным кажется вопрос о том, что служит входной информацией для этого канала. До сих пор мы пользовались формальным допущением, что ее роль выполняет функция  $f(t)$ , описывающая частотный контур сигнала (или его субъективный вариант  $z^*(t)$ ). Такое представление заимствует технический принцип определения производной формантного контура.

Технический способ предполагает двухэтапную процедуру, в которой выделяется аналоговый сигнал, повторяющий во времени форму исходного частотного контура, например напряжение, пропорциональное частоте форманты. Затем этот сигнал дифференцируется также аналоговым способом, и результат дифференцирования может оцениваться либо по пиковому значению производной, либо каким-то другим путем.

В слуховой системе принципиально допустим подобный анализ производной, однако более вероятно, что слух решает эту задачу другим способом, исключая необходимость начального этапа выделения аналога частотного контура.

С одной стороны, эффекты типа латерального торможения (см. главу 10) указывают на возможность сравнения аналоговым способом выходов соседних по  $z$  частотных каналов (суммирование выходов с разными знаками). При этом могут суммироваться, очевидно, не сами огибающие, а их производные (о фактах дифференцирования огибающих в частотных каналах см. главу 11). Легко показать, что сигнал, который явится результатом такого

суммирования, должен содержать информацию о направлении и величине скорости изменения спектра.

Исходя из того факта, что суммирование выходов частотных каналов возможно только при сравнительно близком их соседстве, следует ожидать, что обнаружение изменений по такому способу также должно происходить в узкой области спектра, т. е. что канал анализа спектральных изменений должен иметь строго локальный характер. Объединение информации об изменении всего спектра, по-видимому, должно быть задачей следующего этапа обработки.

Высказанное предположение в какой-то степени поддерживается результатами работы [316], представляющей собой пока что единственный пример психоакустического исследования канала слухового детектирования изменений частоты.

При измерении порогов обнаружения временных изменений частоты в ЧМ-тонах с несущей частотой 250 Гц и с низкой частотой модуляции (1÷80 Гц) авторы наблюдали эффект адаптации, проявляющейся следующим образом. Пороги (пороговые девиации), измеренные в условиях предъявления стимула сразу же за длительным (12 с) адаптирующим ЧМ-тоном с такими же, как у стимула, или близкими значениями несущей и модулирующей частоты, но с гораздо большей девиацией, существенно превышают (максимум в 3 раза) пороги для изолированно предъявляемого стимула.

Как известно [89], эффект адаптации по отношению к некоторому признаку воспринимаемого стимула рассматривается как свидетельство действия специального детектора, настроенного на обнаружение данного признака (или канала, в котором этот признак выделяется).

Среди целого ряда свойств, которыми характеризуется эффект адаптации (он имеет центральную природу), особый интерес вызывают следующие два:

1) при использовании разных значений несущей частоты в стимуле и адаптирующем тоне существенное увеличение порогов наблюдается только при условии не слишком большой разности этих частот (примерно  $\pm 100$  Гц); 2) используемый в качестве адаптирующего сигнала АМ-тон с такой же частотой модуляции, что и ЧМ-тон, не вызывает увеличения порогов обнаружения изменений частоты последнего.

Эти свойства адаптации позволяют думать, что устройство для обнаружения изменений частоты извлекает информацию только из некоторой ограниченной области спектра и что его работа не зависит от работы тех устройств, которые обнаруживают неравномерности огибающей в частотных каналах.

Учитывая разнообразный характер временных изменений спектра в реальном речевом сигнале, можно предполагать, что канал слухового анализа изменений включает в себя систему детекторов с различными свойствами и с разной специализацией по способу их при-

менения. В частности, важным вопросом представляется участие детекторов изменения частоты в процессе сегментации речевого потока.

Таким образом, все рассмотренные в данной главе экспериментальные факты позволяют считать, что слуховой анализ изменения частоты выполняется комплексом устройств, различающихся функционально, имеющих дело с различными параметрами стимула и играющими различную роль в процессе общей обработки сигналов.

Исследования в направлении изучения механизма анализа временных изменений частоты представляются в настоящий момент наиболее актуальными для решения проблемы слуховой обработки речевых сигналов.

## Глава 13

### ПЕРСПЕКТИВЫ РАЗРАБОТКИ ТЕОРИИ ВОСПРИЯТИЯ РЕЧИ

Главной проблемой физиологии слуха и психоакустики на протяжении всей истории развития этих дисциплин было исследование улитки как анализатора звуковых частот. Психологически понятным следствием такой концентрации интересов явилось то, что слуховая система в целом стала трактоваться как спектральный анализатор.

При описании отклика на сигнал спектрального анализатора представляется естественным рассматривать и входной сигнал, и отклик на него как непрерывные функции времени. По аналогии это было перенесено на ощущение, которое также рассматривается классической психоакустикой как непрерывная функция времени. В частности, такие субъективные признаки звука, как его высота, громкость, тембр обычно определяются как текущие, в той или иной мере сглаженные по времени, характеристики отклика анализатора и соответственно характеристики текущего состояния звукового воздействия на входе системы.

Аналогичные представления были положены в основу теорий (моделей) восприятия речи. Практически всеми авторами принималось, что фонетическая интерпретация стимула базируется на непрерывной (или производимой с достаточно высокой частотой) классификации текущих значений отклика спектрального анализатора.

Это нашло отражение в структуре систем, предназначенных для автоматического распознавания речи. Фонемная классификация речевого сигнала, осуществляемая в этих системах, состоит в большинстве случаев из двух этапов. Выходом первичного классификатора, применяемого на первом этапе, является или текущее «фонемное состояние», или оценка текущего состояния речеобразующей системы диктора.

Фонемное состояние соответствует обозначению категории (класса) спектров, в которую попадает значение спектра сигнала, существующее в данный момент времени. Число возможных

категорий спектров обычно принимается равным или несколько большим числа фонем, принятых в лингвистике для описания данного языка. Пока спектр сигнала находится в пределах одной категории, фонемное состояние на выходе классификатора сохраняется неизменным; когда спектр сигнала перемещается в другую категорию, классификатор вырабатывает новое фонемное состояние.

Оценка текущего состояния речеобразующей системы также основывается на текущих (мгновенных) значениях спектральных или некоторых других признаков сигнала. Диагностированное состояние описывается в терминах значений параметров речеобразующей системы. Таким образом, и в этом случае непрерывная последовательность состояний речевого сигнала на входе системы представляется в виде непрерывной последовательности состояний на выходе классификатора.

Изображением сигнала, используемым на втором этапе классификации, является запомненная (записанная) последовательность выходных состояний первичного классификатора.

Если выходом первичного классификатора являются фонемные состояния, то на втором этапе классификации применяются правила, по которым последовательность отличающихся друг от друга фонемных состояний на некотором временном отрезке заменяется одним фонемным состоянием, приписанным всему отрезку в целом. При разработке этих правил второго этапа классификации исследователи стремятся к тому, чтобы число таких окончательных «фонемных отрезков» в слове соответствовало числу фонем в слове.

Красной нитью в этих подходах проходит идея состояний. Фонемы рассматриваются как состояния; процесс распознавания основывается на классификации состояний речевого сигнала.

Факты, говорящие о том, что эти модельные представления являются ошибочными, постепенно накапливались в двух независимых областях исследования.

Электрофизиологические эксперименты показали, что подавляющее большинство нейронов центральных отделов слуховой системы реагируют лишь в моменты изменений звукового воздействия и, следовательно, передают информацию не о текущих значениях отклика периферического спектрального анализатора, но об изменениях этого отклика.

Исследования восприятия синтетических речеподобных сигналов показали, что для указания связи между динамической спектрограммой стимула и фонемной категорией, в которую относит этот стимул слушатель, необходимо располагать некоторой промежуточной системой понятий, в основе которой лежит понятие локализованного во времени акустического события, т. е. изменения звука. Производными понятиями в этой системе являются такие понятия, как последовательность событий, отрезок процесса, длительность отрезка, значения процесса в мо-



менты начала или конца отрезка и т. д. Существенно, что с введением понятия акустического события появляется возможность использования качественных временных отношений типа «одновременно», «следует за», «вскоре» и ряда других.

Результаты психоакустических экспериментов сейчас уже не оставляют сомнений в том, что человек действительно располагает некоторым словарем акустических событий, причем эти события качественно соответствуют тем классам изменений отклика периферического спектрального анализатора, избирательная чувствительность к которым была обнаружена у нейронов центральных отделов слуховой системы.

Использование человеком системы понятий, связанной с понятием акустического события, следует из данных, относящихся как к фонемному распознаванию, так и к восприятию просодической информации.

Очевидно, что алгоритмы обработки речевого сигнала в модели восприятия речи должны использовать ту же систему понятий и отношений, которой пользуется человек. Чтобы упомянутые выше понятия (последовательность событий, отрезок и т. д.) могли быть сформированы, модель прежде всего должна быть наделена способностью к обнаружению и опознаванию некоторого набора акустических событий. Иначе говоря, в модель восприятия речи должен быть включен блок, воспроизводящий в какой-то мере центральную слуховую обработку информации и обеспечивающий выделение и классификацию изменений отклика периферического спектрального анализатора.

Включение в модель дополнительного блока промежуточной обработки информации означает, что модель становится более сложной, чем предполагалось ранее, и что перед исследователями возникает ряд новых трудных задач, связанных с определением структуры и параметров этого блока.

Рассмотрим, следует ли из этого, что создание теории восприятия речи отодвигается в необозримое будущее.

Ответ нейрофизиологов на этот вопрос должен быть, по всей вероятности, положительным. Сама идея о том, что центральную слуховую систему можно рассматривать как вспомогательный промежуточный блок в модели восприятия речи, должна вызывать возмущение у нейрофизиологов слуха. Огромное количество нейронов, образующих центральный слуховой путь, разнообразие их поведения, наличие нескольких анатомических уровней со сложными связями как внутри каждого уровня, так и между уровнями — все это приводит к убеждению, что слуховая система является необычайно сложным устройством, выполняющим такую изысканную обработку информации, какую сейчас не могут даже вообразить математики и инженеры.

Можно, однако, назвать несколько допущений, при которых, если они верны, сложность слуховой системы окажется скорее кажущейся, чем действительной. В качестве таких допущений

приведем следующие: 1) нейроны организованы в упорядоченные группы — нейронные каналы, число которых весьма ограничено; 2) сигналом, передаваемым по каналу (используемым соответствующим приемником), является такая суммарная по всем нейронам переменная, как плотность импัลсации в канале; 3) использование нескольких анатомических уровней обусловлено не столько принципиальными, сколько «техническими» причинами, например трудностью получить требуемые характеристики преобразователя при использовании только одного звена.

Легко заметить, что в основе этих допущений лежит еретическая для нейрофизиолога мысль о том, что большое число нейронов и сложность связей между ними являются средством преодоления того, что нейрон представляет собой не слишком удобный элемент для выполнения тех преобразований звуковой информации, которые нужны организму.

Верны или ошибочны приведенные допущения, пока сказать нельзя. Во всяком случае ясно, что они являются весьма рискованными. Тем не менее исследователь, пытающийся уже сегодня разрабатывать модель центральной слуховой системы, вынужден делать именно такие или близкие к ним допущения.

Обратимся теперь к тем специалистам, которые непосредственно заинтересованы в скорейшей разработке функциональной модели центральной слуховой системы.

Исследователи, занимающиеся поиском алгоритмов фонетической интерпретации речевого потока, делают ли они это путем исследования восприятия речи или путем изучения акустических характеристик естественной речи, необходимо задаются тем или иным описанием речевого сигнала. Без этого они, естественно, просто не могут работать. Для этих специалистов переход к описанию сигнала в терминах акустических событий означал бы возможность использования в правилах интерпретации удобных понятий и интуитивно понятных типов отношений.

На разработку совершенной модели центральной слуховой системы потребуется не одно десятилетие. Ждать столько времени исследователи восприятия речи просто не могут. Став на точку зрения этих специалистов, разумно на первых этапах работы резко упростить проблему и определить блок центральной слуховой системы в модели восприятия речи как специализированное техническое устройство, обеспечивающее обнаружение и опознавание определенных классов изменений в отклике спектрального анализатора.

Практически это означает, во-первых, что на разрабатываемую модель блока сразу накладывается множество ограничений (реализуемость современными техническими средствами, согласованность с остальными блоками общей модели восприятия речи и т. д.), во-вторых, задачей исследования становится получение ответов на вопросы, возникающие у конструктора этого блока. Эксперимент, естественно, остается психоакустическим или электро-

физиологическим. Предвзятыми становятся цели (выбор характера) эксперимента и интерпретация получаемых результатов.

После этих предварительных замечаний обратимся к вопросу о том, какие правдоподобные предположения о структуре интересующего нас блока можно сделать на основании уже имеющихся данных.

И электрофизиологические, и психоакустические данные указывают на то, что обработка отклика спектрального анализатора, обеспечивающая обнаружение изменений в нем, осуществляется в частотных каналах слуховой системы. Иначе говоря, обнаруживаются локальные по спектру изменения отклика, которые можно рассматривать как частичные события. Данные указывают на наличие нескольких классов частичных событий. Так, приходится думать, что выделяются по меньшей мере четыре класса изменений в огибающей, отличающиеся по знаку и по скорости. Обнаруживаемые сдвиги спектральных максимумов по шкале частот разделяются по меньшей мере на два класса по знаку изменения.

Допущение такой структуры приводит к выделению двух групп вопросов, которые могут исследоваться в какой-то мере параллельно. Одна из них касается структуры и параметров самих обнаружителей частичных событий. Вторая группа вопросов касается правил построения полных событий из частичных.

Для фонетической интерпретации, очевидно, используются уже полные события типа начала или конца «звука», положительного или отрицательного формантного перехода и т. д. Естественно предположить, что полные события получаются путем какого-то объединения частичных событий по частотным диапазонам (группам частотных каналов) и по времени. При этом очевидно, что некоторые классы частичных событий могут быть совместимыми, а некоторые — взаимоисключающими. Вопрос о структуре и параметрах обнаружителей частичных событий можно и полезно исследовать как электрофизиологическими, так и психоакустическими методами.

Проблема объединения частичных событий, вероятно, доступна пока лишь психоакустическому исследованию. Предложенные за последние годы методы «адаптации детекторов признаков» и методы дихотического предъявления признаков кажутся в данном случае весьма перспективными.

Обратимся теперь еще к одному вопросу, имеющему принципиальное значение. Очевидно, что наличие системы, обнаруживающей события — изменения в отклике анализатора, — не исключает возможности существования параллельной системы, передающей сведения о текущем, возможно усредненном по времени, состоянии отклика. Современные электрофизиологические данные не согласуются с представлением о том, что такая система находится в центральных отделах классического слухового пути,

но классический слуховой путь не является единственным путем, по которому поступает информация.

Так как обнаружение изменений в отклике наверняка связано с целым рядом нелинейных преобразований, очевидно, что восстановление по реакциям обнаружителей частичных событий абсолютных значений отклика должно быть сопряжено с большими трудностями.

С этой точки зрения кажется разумным предположить, что существуют две параллельные системы обработки информации и что сигналы, возникающие в моменты обнаружения акустических событий, играют роль импульсов опроса, позволяющих считывать значения состояния отклика в заданные этими импульсами моменты времени. Приведенные в книге данные об использовании при фоновой классификации значений спектра сигнала в моменты начала и конца отрезка кажутся наиболее совместимыми с этим предположением.

Представление о том, что система располагает информацией о текущих значениях отклика анализатора, но использует при интерпретации сигнала значения отклика в определенные моменты времени, естественно, сильно отличается от представлений классической психоакустики. Если это представление верно, то исследование восприятия абсолютных значений сигнала может успешно вестись лишь после того, как будут определены моменты отсчета, т. е. будет уже разработана модель обнаружения событий.

Все вопросы, которые сейчас обсуждались, касаются слуховой обработки речевого сигнала. Логично ожидать, что правила, основывающиеся на сведениях о том, что обрабатываемый сигнал является именно речевым и что он принадлежит некоторому определенному языку, должны использоваться на следующем уровне обработки — при фонетической интерпретации сигнала.

Ни у кого из исследователей речи не вызывает сомнений тот факт, что изучение этого уровня не может осуществляться только методами исследования восприятия синтетических речеподобных сигналов. Правила, выявленные при исследовании восприятия, обязательно должны проверяться с точки зрения их эффективности для естественной речи. Такая проверка требует реализации этих правил в системе для автоматического распознавания речи.

С другой стороны, алгоритмы фонетической интерпретации, придуманные для систем автоматического распознавания и оказавшиеся эффективными, могут служить в качестве исходных гипотез при исследовании восприятия.

Иначе говоря, исследование правил фонетической интерпретации речевого сигнала требует реальной кооперации работников в области автоматического распознавания речи и изучения восприятия ее человеком. Условием осуществимости такой кооперации является использование одинакового описания сигнала на входе фонетического интерпретатора теми и другими исследователями,

т. е. принятие ими одинаковой модели предварительной «слуховой» обработки сигнала. Если предварительная обработка осуществляется принципиально различающимися устройствами, то правила, касающиеся последующей обработки выходных сигналов этих устройств, также, естественно, окажутся разными и, возможно, несравнимыми.

Таким образом, проблема изучения центральной слуховой обработки речевого сигнала является сейчас важнейшей частью исследований восприятия речи и требует максимальной концентрации усилий для быстрого создания хотя бы приблизительно удовлетворяющей модели.



## ЛИТЕРАТУРА

1. Анализ, синтез и распознавание речи человеком и машиной. Библиографический указатель литературы за 1970—1973 гг. Львов, 1974.
2. Апресян Ю. Д. Идеи и методы современной структурной лингвистики. М., 1966.
3. Арутюнян Э. А. О физиологических механизмах реализации логических ударений. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 18—30.
4. Архипов Г. Б. О влиянии темпа речи на аудирование. — Учен. зап. 1-го Моск. гос. пед. ин-та иностр. языков, 1968, № 44, с. 23—33.
5. Барышникова К. К., Гайдучик С. М. О просодемах вопроса и утверждения. — Zeitschr. für Phonetik, 1969, Bd 22, S. 324—334.
6. Башкина Б. М., Лобанов Б. М. Восприятие русской интонации односложной синтетической фразы. — Тез. докл. Всесоюз. науч. конференции «Анализ и синтез как взаимообусловленные методы экспериментально-фонетических исследований речи», Минск, 1972, с. 22—24.
7. Башкина Б. М., Лобанов Б. М. Анализ и синтез просодических характеристик двусложного слова. — Тез. докл. Всесоюз. науч. конференции «Анализ и синтез как взаимообусловленные методы экспериментально-фонетических исследований речи», Минск, 1972, с. 25—27.
8. Бекеш Г., Розенблит В. А. Механика улитки. — В кн.: Экспериментальная психология. М., 1963, т. 2, с. 682—724.
9. Белозерова Л. Б. Синтез аффрикат грузинского литературного языка. — В кн.: Вопросы анализа речи. Тбилиси, 1973, с. 115—129.
10. Богданов Б. В. Восприятие коротких отрезков гласных звуков. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 49—57.
11. Богос И. И., Деркач М. Ф. Иерархические модели восприятия речи. — Вестн. Львовск. гос. ун-та, Серия биологическая, 1974, вып. 8, с. 3—10.
12. Бондаренко В. П. Модель слуховой системы. — Изв. Томск. политехн. ин-та, 1973, вып. 202, с. 38—40.
13. Бондарко Л. В. Осциллографический анализ речи. Л., 1965.
14. Бондарко Л. В., Вербицкая Л. А. О фонетических характеристиках заударных флексий в современном русском языке. — Вopr. языкознания, 1973, № 1, с. 37—49.
15. Бондарко Л. В., Вербицкая Л. А., Зиндер Л. Р. Акустическая характеристика безударности. — В кн.: Структурная типология языков. М., 1966, с. 56—64.
16. Бондарко Л. В., Вербицкая Л. А., Зиндер Л. Р., Павлова Л. П. Различаемые звуковые единицы русской речи. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 165—179.
17. Бондарко Л. В., Вербицкая Л. А., Щербакова Л. П. Об определении места ударения в слове. — Изв. АН СССР, Серия литературы и языка, 1973, т. 32, вып. 2, с. 141—153.

18. Бондарко Л. В., Загоруйко Н. Г., Кожевников В. А., Молчанов А. П., Чистович Л. А. Модель восприятия речи человеком. Новосибирск, 1968.
19. Бондарко Л. В., Зиндер Л. Р. Исследование фонетики. — В кн.: Основы теории речевой деятельности. М., 1974, с. 145—160.
20. Бондарко Л. В., Зиндер Л. Р., Светозарова Н. Д. Разграничение слов в потоке речи. — Вopr. языкознания, 1968, № 2, с. 68—81.
21. Бондарко Л. В., Кукольник Л. Е., Павлова Л. П., Светозарова Н. Д., Штерн А. С. Восприятие фонем в слогах различных типов. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 182—192.
22. Бровченко Т. А. К вопросу о взаимосвязи акустических характеристик слога. — В кн.: Вопросы фонологии и фонетики. М., 1971, ч. 1, с. 53—59.
23. Вардапетян Г. А. Характеристики реакций нейронов слуховой коры кошки. — В кн.: Механизмы слуха. Л., 1967, с. 74—90.
24. Вартамян И. А. Импульсная активность нейронов заднего двухолмия на звуковые сигналы, модулированные по амплитуде. — Физиол. журн. СССР, 1969, т. 55, с. 273—282.
25. Вартамян И. А. Возбуждающие и тормозные частотные зоны ответов нейронов задних холмов. — Нейрофизиология, 1971, т. 3, с. 369—378.
26. Вартамян И. А. Реакции центральных слуховых нейронов на сигналы с различным направлением частотной модуляции. — Нейрофизиология, 1974, т. 6, с. 237—245.
27. Вартамян И. А., Гершун Г. В. Реакция нейронов задних холмов на амплитудно-модулированные сигналы при изменении несущей частоты и ритма модуляции. — Нейрофизиология, 1972, т. 4, с. 12—21.
28. Величко В. М., Загоруйко Н. Г. Автоматическое распознавание 200 устных команд. — В кн.: Вычислительные системы. Новосибирск, 1969, вып. 37, с. 73—76.
29. Венцов А. В. О работе голосовых связок при глухих смычных интервокальных согласных. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 62—72.
30. Венцов А. В. О сокращенном описании изменений частоты основного тона в речи. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 132—141.
31. Венцов А. В. Влияние длительности согласного и гласного на восприятие ударения в русском языке. — Тез. докл. 8-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 3, с. 78—80.
32. Вербичкая Л. А. Звуковые единицы русской речи и их соотношение с оттенками и фонемами. — Дисс. на соиск. учен. степени канд. филологич. наук. Л., 1965.
33. Винцук Т. К. Распознавание слов устной речи методами динамического программирования. — Кибернетика, 1968, № 1, с. 81—88.
34. Винцук Т. К. Конструктивный подход к распознаванию речи. Применение математического программирования в распознавании слуховых образов. — Труды 4-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Киев, 1969, с. 21—41.
35. Вокoderная телефония. Ред. А. А. Пирогов, М., 1974.
36. Володка З. М., Молошная Т. К., Николаева Т. М. Опыт описания русского языка в его письменной форме. М., 1964.
37. Волошин Г. Я. Методы описания и распознавания речевых сигналов. — В кн.: Распознавание слуховых образов. Новосибирск, 1970, с. 205—270.
38. Высоцкий Г. Я., Рудный Б. Н., Трунин-Донской В. Н. Особенности автоматического распознавания типовых фраз слитной

- речи (при использовании ограниченного словаря). — Труды 7-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Алма-Ата, 1973, с. 4—7.
39. Галунов В. И., Манеров В. Х. Связь между психофизиологическим состоянием говорящего и характеристиками речевого сигнала. — Тез. докл. 8-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 4, с. 46—48.
  40. Галунов В. И., Чистович Л. А. О связи моторной теории с общей проблемой распознавания речи. — Акуст. журн., 1965, т. 11, с. 417—426.
  41. Гершуни Г. В., Авакян Р. В., Бару А. В., Миронova Л. М. О различении звуков с разными временами нарастания. — В кн.: Эволюция функций. М.—Л., 1964, с. 66—75.
  42. Гершуни Г. В., Вартанян И. А. Реакции центральных слуховых нейронов на звуки с различной временной структурой. — Физиол. журн. СССР, 1972, т. 58, с. 988—999.
  43. Глезер В. Д., Дудкин К. Н., Куперман А. М., Леушина Л. И., Невская А. А., Подвигин Н. Ф., Праздникова Н. В. Зрительное опознание и его нейрофизиологические механизмы. Л., 1975.
  44. Голубцов С. В. О применении ЭЦВМ для исследования методов автоматического распознавания речевых элементов. — Труды 4-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Киев, 1969, с. 73—86.
  45. Голубцов С. В. Синтез речи. — Труды 4-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Киев, 1969, с. 107—129.
  46. Голубцов С. В. Задачи и перспективы распознавания речи. — Тез. докл. 6-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Таллин, 1972, с. 64—73.
  47. Голузина А. Г. Шкалирование субъективных расстояний между синтетическими гласными. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 155—160.
  48. Голузина А. Г., Люблинская В. В., Рохтла М. К., Чистович Л. А. Шкалирование субъективных расстояний как метод исследования слухового восприятия. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 96—110.
  49. Гольдбурт С. Н. Нейродинамика слуховой системы человека. Л., 1964.
  50. Горелик А. Л., Скрипкин В. А. Некоторые вопросы построения систем распознавания. М., 1974.
  51. Горский С. М., Каспранский Р. Р., Ваймер А. Ф. О выражении границы фонетических слов при естественном членении речевого потока. — Учен. зап. Горьковск. пед. ин-та иностр. языков, 1971, т. 46, с. 39—47.
  52. Гранстром М. П., Жуков С. Я., Жукова М. Г., Кожевников В. А., Мушников В. Н., Родionoв В. Д., Чистович И. А., Чистович Л. А. К изучению особенностей восприятия ритмических речеподобных сигналов: длительность минимальной паузы и минимальной звуковой посылки. — В кн.: Проблемы переработки информации в сенсорных системах организма. Л., 1975, с. 81—90.
  53. Гумецкий Р. Я., Деркач М. Ф., Ермолова Е. Н. Сравнительный анализ распознавания слов автоматом и человеком по их динамическим спектрограммам. — Вестн. Львовск. гос. ун-та, Серия биологическая, 1974, вып. 8, с. 87—96.
  54. Девис Г. Периферическое кодирование слуховой информации. — В кн.: Теория связи в сенсорных системах. М., 1964, с. 232—250.

55. Дөрках М. Ф., Гумецкий Р. Я., Мишин Л., Оверченко М., Чабан М. Восприятие речи в распознающих моделях. Львов, 1971.
56. Дубровский Н. А., Тумаркина Л. Н. Исследование восприятия человеком амплитудно-модулированных шумов. — Акуст. журн., 1967, т. 13, с. 51—59.
57. Дубровский Н. А., Тумаркина Л. Н. Исследование слухового механизма восприятия модуляций. Амплитудная модуляция двумя тонами. — Труды Акустического ин-та, 1970, вып. 12, с. 48—53.
58. Дудкин К. Н., Гауэльман В. Е. Модель переработки сигналов в нейронных ансамблях типа гломерулы наружного колеччатого тела зрительной системы. — Биофизика, 1973, т. 18, с. 715—723.
59. Дукельский Н. И. Принципы сегментации речевого потока. М.—Л., 1962.
60. Жинкин Н. И. Восприятие ударения в словах русского языка. — Изв. Академии пед. наук РСФСР, 1954, т. 54, с. 7—83.
61. Жолковский А. К., Леонтьева Н. Н., Мартынов Ю. С. О принципиальном использовании смысла при машинном переводе. — В кн.: Машинный перевод. Труды Ин-та точной механики и вычислительной техники, вып. 2, М., 1961, с. 17—46.
62. Жуков С. Я., Жукова М. Г., Чистович Л. А. Некоторые новые представления относительно слухового анализа звукового потока. — Акуст. журн., 1974, т. 20, с. 386—392.
63. Жуков С. Я., Кожевников В. А., Кузьмин Ю. И., Чистович Л. А. Образование слога при амплитудной модуляции гласноподобного звука. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов, № 2. Объединенный науч. совет «Физиология человека и животных» АН СССР, Информационные материалы (вып. 32), Л., 1971, с. 36—44.
64. Жуков С. Я., Лисенко Д. М. Марки сегментации и их роль в интерпретации формантного контура. — Тез. докл. 8-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 2, с. 38—41.
65. Жукова М. Г. Восприятие длительности звонких согласных. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 92—100.
66. Жукова М. Г. Некоторые характеристики восприятия глухих и звонких начальных согласных. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов, № 2. Объединенный науч. совет «Физиология человека и животных» АН СССР, Информационные материалы (вып. 32), Л., 1971, с. 45—51.
67. Жукова М. Г. Восприятие паузы между синтетическими гласными. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 120—121.
68. Загоруйко Н. Г. Структура проблемы распознавания слуховых образов и методы ее решения. — В кн.: Распознавание слуховых образов. Новосибирск, 1970, с. 3—54.
69. Загоруйко Н. Г. Методы распознавания и их применение. М., 1972.
70. Зимняя И. А., Фомичев В. А. Исследование одного из акустических стимулов, вызывающих восприятие речевой интонации вопроса. — Вопр. психол., 1964, № 5, с. 73—81.
71. Зиндер Л. Р. Русские артикуляционные таблицы. — Труды Военной Краснознаменной академии связи им. С. М. Буденного, 1951, № 29—30, с. 31—46.
72. Зиндер Л. Р. Общая фонетика. Л., 1960.
73. Зиндер Л. Р. Влияние темпа речи на образование отдельных звуков. — Учен. зап. Лен. гос. ун-та, Серия филологических наук, 1964, вып. 69, т. 325, с. 3—27.
74. Златоустова Л. В. Фонетическая природа русского словесного ударения. — Учен. зап. Казанск. гос. ун-та, 1956, т. 116, с. 3—32.

75. Исследование речи. Труды Хаскинской лаборатории. Новосибирск, 1967.
76. Казюкова Н. Длительность звуков как средство эфазы. — Сб. научн. работ 1-го Моск. гос. пед. ин-та иностр. языков, 1969, № 16, с. 77—86.
77. Карницкая Э. Г. О расчетах слуховых эквивалентов спектров стационарных сигналов. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 167—181.
78. Катарская А. Р. Увеличение длительности звуков слова как средство его эмфатического ударения. — Учен. зап. 1-го Моск. гос. пед. ин-та иностр. языков, 1971, т. 60, с. 388—393.
79. Клас Ю. А. О членении на слова предложений, лишенных лексической информации. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 193—205.
80. Книппер А. В., Петров О. А. Опыт опознавания небольшого набора слов. — В кн.: Опознавание образов. М., 1968, с. 155—165.
81. Кожевников В. А., Рейтблат Л. Е., Чистович Л. А. Знак изменения частоты спектрального максимума как различительный признак места образования согласного. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 74—82.
82. Кривнова О. Ф. Некоторые особенности интонационной системы русского языка (на основе экспериментального исследования). — Дисс. на соиск. учен. степени канд. филологич. наук. М., 1969.
83. Кривнова О. Ф., Венцов А. В. Об интонационном членении некоторых типов предложений русского языка. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 161—172.
84. Крылов И. Н. К вопросу о регуляции темпа простых ритмических движений. — В кн.: Управление движениями. Л., 1970, с. 38—49.
85. Кузьмин Ю. И., Лисенко Д. М. Фонетическая интерпретация стимулов с резкими изменениями интенсивности. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 82—91.
86. Кузьмин Ю. И., Лисенко Д. М. Контекстуальные влияния в экспериментах по определению фонемных границ. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 142—149.
87. Лабутин В. К., Молчанов А. П. Модели механизмов слуха. М., 1973.
88. Линдгрен Н. Машинное распознавание человеческого языка. — Зарубежная радиоэлектроника, 1966, № 3, с. 55—89; № 4, с. 63—89; № 5, с. 56—76.
89. Линдсей П., Норман Д. Переработка информации у человека. М., 1974.
90. Лисенко Д. М. О членении человеком непрерывного потока речи на слова. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 180—192.
91. Лисенко Д. М. Влияние контекста на восприятие гласных. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов, № 2. Объединенный научн. совет «Физиология человека и животных» АН СССР, Информационные материалы (вып. 32), Л., 1971, с. 52—59.
92. Лисенко Д. М., Федорова Н. А. К вопросу о временных ограничениях слуховой сегментации. Тез. докл. 8-го Всесоюзн. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 2, с. 49—51.
93. Лисенко Д. М., Федорова Н. А., Чистович Л. А., Бороздин А. Н., Жуков С. Я., Кожевников В. А. Влияние временной структуры синхронизирующего речевого стимула на временную структуру произносимого звуко сочетания. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 33—55.



94. Л ю б и н с к и й И. А., Я х н о В. П. Исследование и моделирование механизмов восприятия ритмических сигналов. — Тез. докл. 8-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 1, с. 48—50.
95. Л ю б л и н с к а я В. В. Распознавание артикуляторных признаков смычных согласных по переходу от гласного к согласному. — Акуст. журн., 1966, т. 12, с. 213—221.
96. Л ю б л и н с к а я В. В. Воспроизведение простых контуров изменения частоты основного тона звуков. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 66—73.
97. Л ю б л и н с к а я В. В. Воспроизведение голосом изменения частоты основного тона звука. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов, № 2. Объединенный науч. совет «Физиология человека и животных» АН СССР, Информационные материалы (вып. 32), Л., 1971, с. 92—98.
98. Л ю б л и н с к а я В. В. Воспринимаемые параметры формантного контура. — Тез. докл. 8-го Всесоюз. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 2, с. 74—76.
99. Л ю б л и н с к а я В. В., Г о л у з и н а А. Г. Влияние длительности тональных сигналов на величину субъективного расстояния. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 111—117.
100. Л ю б л и н с к а я В. В., М и к е л ь В., Ш е й к и н а О. Шкалирование субъективных расстояний между звуками с изменяющейся частотой основного тона. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 142—155.
101. Л я н Ч ж и - а н ь, Ч и с т о в и ч Л. А. Дифференциальные пороги по частоте в зависимости от длительности тональных посылок. — Акуст. журн., 1960, т. 6, с. 81—86.
102. М а й е р с Т. Ф., Ж у к о в а М. Г., Ч и с т о в и ч Л. А., М у ш н и к о в В. Н. Применение метода дихотической стимуляции для изучения процедур слуховой сегментации. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 86—119.
103. М е л ь ч у к И. А. Опыт теории лингвистических моделей «смысл → текст». М., 1974.
104. М е л ь ч у к И. А., Р а в и ч Р. Д. Автоматический перевод. 1949—1963. Критико-библиографический справочник. М., 1967.
105. М о р о з о в В. П., Ч е р н и г о в с к а я Т. В. Об избирательной чувствительности слуха человека к амплитудной модуляции речи. — Журн. эволюц. биологии и физиологии, 1975, т. 11, с. 468—473.
106. М у ш н и к о в В. Н., Ч и с т о в и ч Л. А. Метод экспериментального исследования роли громкостей составляющих гласного в его распознавании. — Акуст. журн., 1971, т. 17, с. 405—411.
107. М у ш н и к о в В. Н., Ч и с т о в и ч Л. А. О слуховом описании гласного. I. Признаки, различающие [i] и [e]. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 5—10.
108. М у ш н и к о в В. Н., Ч и с т о в и ч Л. А. О слуховом описании гласного. II. Обнаружение второй форманты в синтетическом гласном. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 11—19.
109. М у ш н и к о в В. Н., Ч и с т о в и ч Л. А. Об экспериментальной проверке полосной гипотезы восприятия гласных. — Акуст. журн., 1973, т. 19, с. 386—392.
110. М у ш н и к о в В. Н., Ч и с т о в и ч Л. А. О флюктуациях фонемной границы. — Акуст. журн., 1973, т. 19, с. 571—579.
111. М я с н и к о в Л. Л. Объективное распознавание звуков речи. — Журн. технич. физики, 1943, т. 13, с. 109—115.
112. М я с н и к о в Л. Л. Звуки речи и их объективное распознавание. — Вестн. Лен. гос. ун-та, 1946, № 3, с. 3—13.

113. П а н о в В. М. Русская фонетика. М., 1967.
114. П о з и н Н. В. Моделирование нейронных структур. М., 1970.
115. П у х и р ь Н. В., Щ е р б а к о в а Л. П. Восприятие и фонемная классификация человеком фонетически неопределенных гласных. — Тез. докл. 8-го Всесоюзн. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 2, с. 77—80.
116. Р а д и о н о в а Е. Л. Функциональная характеристика нейронов кохлеарных ядер и слуховая функция. Л., 1971.
117. Р а т н и к о в а Г. И. Строение путей и центров слуховой системы. — В кн.: Руководство по физиологии. Физиология сенсорных систем. Л., 1972, ч. 2, с. 192—208.
118. Р е ф о р м а т с к и й А. А. Из истории отечественной фонологии. М., 1970.
119. Р у д н ы й Б. Н., Т р у н и н - Д о н с к о й В. Н. Современное состояние работ по речевому управлению. — В кн.: Речевое управление. М., 1972, с. 3—16.
120. С а м о й л о в а И. К. Зависимость величины маскировки коротких тональных сигналов от временного интервала между маскируемым и маскирующим звуками. — Биофизика, 1959, т. 4, с. 550—558.
121. С а п о ж к о в М. А. Речевой сигнал в кибернетике и связи. М., 1963.
122. С в е т о з а р о в а Н. Д., Г е й л ь м а н Н. И., З ы к о в а М. А. Об использовании синтезированной речи при анализе интонации (на материале русского языка). — Тез. докл. Всесоюзн. науч. конференции «Анализ и синтез как взаимообусловленные методы экспериментально-фонетических исследований речи», Минск, 1972, с. 140—142.
123. С в е т о з а р о в а Н. Д., Щ е р б а к о в а Л. П. Роль изменения частоты основного тона в восприятии ударения в изолированных словах и предложениях. — Тез. докл. 6-го Всесоюзн. семинара «Автоматическое распознавание слуховых образов», Таллин, 1972, с. 164—167.
124. С е б е с т и а н Г. С. Процессы принятия решений при распознавании образов. Киев, 1965.
125. С е л е з н е в а И. С. Восприятие словесного ударения. — Учен. зап. Моск. гос. пед. ин-та иностр. языков, 1960, т. 18, с. 359—389.
126. С е р к о в Ф. Н., С т о р о ж у к В. М. Реакция нейронов слуховой коры на звуковые раздражения. — Нейрофизиология, 1969, т. 1, с. 147—157.
127. С е р к о в Ф. Н., Я н о в с к и й Е. Ш. О реакциях нейронов слуховой коры на парные звуковые щелчки. — Нейрофизиология, 1970, т. 2, с. 227—234.
128. С е р к о в Ф. Н., Я н о в с к и й Е. Ш., Т а л ь н о в А. Н. О характеристике тормозящих нейронов коры мозга. — Нейрофизиология, 1974, т. 6, с. 119—126.
129. С л е п о к у р о в а Н. А. О процедуре распознавания стационарных гласных. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 156—166.
130. С т ю а р т Д. Л., К е с с о н Д. М. Модель, дающая зрительное представление нейронной активности в слуховом анализаторе при восприятии речи. — В кн.: Кибернетические проблемы бионики. Анализ биологических прототипов. М., 1971, с. 261—274.
131. Т е м о в В. Л. Модель для описания результатов психоакустических экспериментов со стационарными сигналами. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 36—49.
132. Т р у б е ц к о й Н. С. Основы фонологии. М., 1960.
133. Т у р б о в и ч И. Т., Ю р к о в Е. Ф., Г и т и с В. Г. Детерминированный метод принятия решения при опознании образов (метод эталонов). — В кн.: Опознание образов. М., 1968, с. 40—66.
134. Ф а н т Г. Акустическая теория речеобразования. М., 1964.
135. Ф а н т Г. Анализ и синтез речи. Новосибирск, 1970.

136. Федорова Н. А. Восприятие удара в простой звуковой последовательности. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 110—122.
137. Фланаган Д. Анализ, синтез и восприятие речи. М., 1968.
138. Фланаган, Кокер, Рабинер, Шефер, Умеда. Синтезированная речь для ЦВМ. — Зарубежная радиоэлектроника, 1971, № 10, с. 45—81.
139. Фрейдлин А. А. О роли критических полос в обработке информации слуховой системой человека. — Акуст. журн., 1968, т. 14, с. 321—336.
140. Френкс Л. Теория сигналов. М., 1974.
141. Хайретдинова А. Г. Алгоритмы автоматического выделения словесного удара в повествовательных фразах. — Тез. докл. 8-го Всесоюзн. семинара «Автоматическое распознавание слуховых образов», Львов, 1974, ч. 3, с. 93—97.
142. Харкевич А. А. Спектры и анализ. М., 1957.
143. Харкевич А. А. Теория информации. Опознавание образов. Избранные труды. М., 1973, т. 3.
144. Холшевников В. Е. Основы стиховедения. Л., 1962.
145. Цвикер Э., Фельдткеллер Р. Ухо как приемник информации. М., 1971.
146. Цемель Г. И. Автоматическое (объективное) распознавание звуков речи. — Зарубежная радиоэлектроника, 1961, № 4, с. 52—73.
147. Цемель Г. И. Опознавание речевых сигналов. М., 1971.
148. Цеплитис Л. К. Анализ речевой интонации. Рига, 1974.
149. Чистович Л. А. Частотные характеристики эффекта маскировки. — Биофизика, 1957, т. 2, с. 743—755.
150. Чистович Л. А. Различение интервала времени между двумя короткими звуковыми импульсами. — Акуст. журн., 1959, т. 5, с. 480—484.
151. Чистович Л. А. Изменение основной частоты голоса как различительный признак согласных. — Акуст. журн., 1968, т. 14, с. 449—456.
152. Чистович Л. А. К вопросу о процедуре распознавания фонем. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 123—136.
153. Чистович Л. А. Громкость. — В кн.: Руководство по физиологии. Физиология сенсорных систем. Л., 1972, ч. 2, с. 345—356.
154. Чистович Л. А., Жукова М. Г., Малинникова Т. Г., Кожевников В. А., Бороздин А. Н. Имитация и восприятие длительности изолированных гласных. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 128—156.
155. Чистович Л. А., Иванова В. А. Взаимная маскировка коротких звуковых импульсов. — Биофизика, 1959, т. 4, с. 170—179.
156. Чистович Л. А., Клаас Ю. А. К анализу скрытого периода «произвольной» реакции на звуковые сигналы. — Физиол. журн. СССР, 1962, т. 48, с. 899—906.
157. Чистович Л. А., Клаас Ю. А., Алексин Р. О. О значении имитации для распознавания звуковых последовательностей. — Вопр. психологии, 1961, № 5, с. 173—182.
158. Чистович Л. А., Кожевников В. А. Восприятие речи. — В кн.: Руководство по физиологии. Физиология сенсорных систем. Л., 1972, ч. 2, с. 427—514.
159. Чистович Л. А., Кожевников В. А., Алякринский В. В., Бондарко Л. В., Голузина А. Г., Клаас Ю. А., Кузьмин Ю. И., Лисенко Д. М., Люблинская В. В., Федорова Н. А., Шупляков В. С., Шуплякова Р. М. Речь. Артикуляция и восприятие. М.—Л., 1965.
160. Чистович Л. А., Кожевников В. А., Лесогор Л. В. Исследование периферической слуховой адаптации в психоакустическом эксперименте. — Физиол. журн. СССР, 1972, т. 58, с. 1543—1547.

161. Чистович Л. А., Лисенко Д. М., Фёдорова Н. А. Управление речеобразованием в процессе синхронизации с периодически повторяющимся внешним речевым стимулом. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов. Л., 1972, вып. 3, с. 56—85.
162. Чистович Л. А., Чистович И. А., Лесогор Л. В., Гранстрем М. П. Функциональная модель частотного канала периферического слухового анализатора. — Физиол. журн. СССР, 1972, т. 58, с. 1548—1557.
163. Чистович Л. А., Шупляков В. С. Слуховое измерение первой форманты. — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 19—28.
164. Шенгели Г. А. Тракта́т о русском стихе. М., 1923.
165. Шупляков В. С. О тональной высоте звуков [s] и [ʃ]. — В кн.: Механизмы речеобразования и восприятия сложных звуков. М.—Л., 1966, с. 87—95.
166. Шупляков В. С. Слуховой анализ стационарных шумных согласных. — Дисс. на соиск. учен. степени канд. биологич. наук. Л., 1967.
167. Шупляков В. С. Акустический признак восприятия мягкости стационарных шумных согласных [s] и [ʃ]. — Материалы 6-й Всесоюз. акустической конференции, М., 1968, 3-III-10.
168. Шупляков В. С. Аппроксимация кривых одновременной маскировки слуха. — В кн.: Сенсорные системы. Вопросы теории и методов исследования восприятия речевых сигналов, № 2. Объединенный науч. совет «Физиология человека и животных» АН СССР, Информационные материалы (вып. 32), Л., 1971, с. 5—20.
169. Шупляков В. С., Фант Г., Серпа-Лейтао А. Акустический признак твердых и мягких согласных русского языка в связанной речи (спектрографическое исследование). — В кн.: Анализ речевых сигналов человеком. Л., 1971, с. 57—66.
170. Щерба Л. В. Фонетика французского языка. М.—Л., 1937.
171. Щербакова Л. П. Об использовании избыточных признаков при идентификации фонем. — Автореф. дисс. на соиск. учен. степени канд. филологич. наук. Л., 1971.
172. Aaronson D. Temporal factors in perception and short-term memory. — Psychol. Bull., 1967, v. 67, p. 130—144.
173. Abel S. M. Duration discrimination of noise and tone bursts. — J. Acoust. Soc. Amer., 1972, v. 51, p. 1219—1223.
174. Abeles M., Goldstein M. H. Functional architecture in cat primary auditory cortex: columnar organization and organization according to depth. — J. Neurophysiol., 1970, v. 33, p. 172—187.
175. Abeles M., Goldstein M. H. Responses of single units in the primary auditory cortex of the cat to tones and to tone pairs. — Brain. Res., 1972, v. 42, p. 337—352.
176. Abramson A. S. On matching the speech signal to mechanisms of production. — Haskins Lab. Status Rep. on Speech Research, N. Y., 1967, SR-10, p. 2.1—2.36.
177. Abramson A. S., Lisker L. Discriminability along the voicing continuum: cross-language tests. — Haskins Lab. Status Rep. on Speech Research, N. Y., 1967, SR-11, p. 17—22.
178. Abramson A. S., Lisker L. Voice timing: cross-language experiments in identification and discrimination. — Haskins Lab. Status Rep. on Speech Research, N. Y., 1968, SR-13/14, p. 49—63.
179. A des A. E. Bilateral component in speech perception? — J. Acoust. Soc. Amer., 1974, v. 56, p. 610—616.
180. Aitkin I. M., Dunlop C. W. Interplay of excitation and inhibition in the cat medial geniculate body. — J. Neurophysiol., 1968, v. 31, p. 44—62.

181. Atal B. S., Hanauer S. L. Speech analysis and synthesis by linear prediction of the speech wave. — J. Acoust. Soc. Amer., 1971, v. 50, p. 637—655.
182. Atal B. S., Schroeder M. R. Predictive coding of speech signals. — Rep. of the 6th Internat. Congr. on Acoustics, Tokyo, 1968, p. C-13—C-16.
183. Babkoff H., Sutton S. Monaural temporal masking of transients. — J. Acoust. Soc. Amer., 1968, v. 44, p. 1373—1378.
184. Bailey P. J. Perception of changing fundamental frequency in complex tones. — Speech Synthesis and Perception. The Psychological Lab. Univ. of Cambridge Progr. Rep., 1972, № 6, p. 13—23.
185. Bailey P. J. Perceptual adaptation for acoustical features in speech. — Preprints of the Speech Communication Seminar, Stockholm, 1974, v. 3, p. 47—53.
186. Bauch H. Die Schwingungsform der Basilarmembran bei Erregung durch Impulse und Geräusche, gemessen an einem elektrischen Modell des Innenohres. — Frequenz, 1956, Bd 10, S. 222—234.
187. Bekesy G. Über die nichtlinearen Verzerrungen des Ohres. — Annales Physik., 1934, Bd 20, S. 809—827.
188. Bekesy G. Über die Resonanzkurve und die Abklingzeit der verschiedenen Stellen der Schneckentrennwand. — Akust. Zeitschr., 1943, Bd 8, S. 66—76.
189. Bekesy G. Experiments in hearing. N. Y., Toronto, London, 1960.
190. Bell C. G., Fujisaki H., Heinz J. M., Stevens K. N., House A. S. Reduction of speech spectra by analysis-by-synthesis techniques. — J. Acoust. Soc. Amer., 1961, v. 33, p. 1725—1736.
191. Bismark G. Vorschlag für ein einfaches Verfahren zur Klassifikation stationären Sprachschall. — Acustica, 1973, v. 28, p. 186—188.
192. Bogdanski D. F., Galambos R. Studies of the auditory system with implanted electrodes. — In: Neural mechanisms of auditory and vestibular system. Springfield, Illinois, 1960, p. 143—148.
193. Bogert B. P. A network to represent the inner ear. — Bell Lab. Record, 1950, v. 28, p. 481—485.
194. Bond Z. S. The perception of obstruent clusters. — Proceed. of the 7th Internat. Congr. of Phonetic Sciences, The Hague, Paris, 1972, p. 465—469.
195. Bondarko L. V. The syllable structure of speech and distinctive features of phonemes. — Phonetica, 1969, v. 20, p. 1—40.
196. Boomer D. S., Dittman A. T. Hesitation pauses and juncture pauses in speech. — Language and Speech, 1962, v. 5, p. 215—220.
197. Bos C. E., de Boer E. Masking and discrimination. — J. Acoust. Soc. Amer., 1966, v. 39, p. 708—715.
198. Brady P. T., House A. S., Stevens K. N. Perception of sounds characterized by a rapidly changing resonant frequency. — J. Acoust. Soc. Amer., 1961, v. 33, p. 1357—1362.
199. Bregman A. S., Campbell J. Primary auditory stream segregation and perception of order in rapid sequences of tones. — J. Exper. Psychol., 1971, v. 89, p. 244—249.
200. Brown W. S., McGlone R. E. Aerodynamic and acoustic study of stress in sentence production. — J. Acoust. Soc. Amer., 1974, v. 56, p. 971—974.
201. Brugge J. F., Anderson D. J., Hind J. E., Rose J. E. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. — J. Neurophysiol., 1969, v. 32, p. 386—401.
202. Bürck W., Kotowski P., Lichte H. Die Hörbarkeit von Laufzeitdifferenzen. — Elektr. Nachrichten Technik, 1935, Bd 12, S. 355—362.
203. Burlet H. M. de. Handbuch der Vergleichenden Anatomie der Wirbeltiere. Berlin, 1934, Bd 2.



204. Campbell R. A., Lasky E. Z. Masker level and sinusoidal-signal detection. — *J. Acoust. Soc. Amer.*, 1967, v. 42, p. 972—976.
205. Carlson R., Erikson Y., Granström B., Lindblom B., Rapp K. Neutral and emphatic stress patterns in Swedish. — Preprints of the Speech Communication Seminar, Stockholm, 1974, v. 2, p. 209—218.
206. Carlson R., Fant G., Granström B. Two-formant models, pitch and vowel perception. — *Acustica*, 1974, v. 31, p. 360—362.
207. Carlson R., Fant G., Granström B. Two-formant models, pitch and vowel perception. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1975, p. 55—82.
208. Carlson R., Granström B., Fant G. Some studies concerning perception of isolated vowels. — *Speech Transmiss. Lab. Quart. Progr. and Status Rep.*, Stockholm, 1970, № 2/3, p. 19—35.
209. Carroll J. D., Chang J. J. Analyses of individual differences in multi-dimensional scaling via an N-way generalization of Eckart-Young decomposition. — *Psychometrika*, 1970, v. 35, p. 283—319.
210. Centmayer K. Interrelations of vowel perception and linguistic context. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1975, p. 143—152.
211. Chistovich I. A., Granstrem M. P., Kozhevnikov V. A., Lesogor L. W., Shupljakov V. S., Taljasin P. A., Tjul'kov W. A. A functional model of signal processing in the peripheral auditory system. — *Acustica*, 1974, v. 31, p. 349—353.
212. Chistovich L. A. Auditory processing of speech stimuli — evidences from psychoacoustics and neurophysiology. *Proceed. of the 7th Internat. Congr. on Acoustics*, Budapest, 1971, v. 1, p. 27—42.
213. Chistovich L. A., Fant G., Serpa-Leitao A. de, Tjertlund P. Mimicking of synthetic vowels. — *Speech Transmiss. Lab. Quart. Progr. and Status Rep.*, Stockholm, 1966, № 2, p. 1—18.
214. Chistovich L. A., Fyodorova N. A., Lissenko D. M., Zhukova M. G. Auditory segmentation of acoustic flow and its possible role in speech processing. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1975, p. 221—232.
215. Chomsky N., Halle M. *The sound pattern of English*. N. Y., 1968.
216. Christov Ph. Experiment for changing the envelope form of vowels. — *Proceed. of the 5th Internat. Congr. on Acoustics*, Liege, 1962, AII.
217. Cohen A. Formant discrimination in the auditory system. — Preprints of the Speech Communication Seminar, Stockholm, 1974, v. 3, p. 111—116.
218. Coker C. H., Umeda N. Subphonemic variations in American English. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1975, p. 539—564.
219. Collier R., t'Hart J. The role of intonation in speech perception. — Preprints of the Symposium on Dynamic Aspects of Speech Percept., Eindhoven, The Netherlands, 1975.
220. Cooper F. S., Delattre P. C., Liberman A. M., Borst J. M., Gerstman L. J. Some experiments on the perception of synthetic speech sounds. — *J. Acoust. Soc. Amer.*, 1952, v. 24, p. 597—606.
221. Cooper F. S., Liberman A. M., Borst J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. — *Proceed. National. Acad. Sci.*, 1951, v. 37, p. 318—325.
222. Cooper W. E. Adaptation of phonetic feature analyzers for place of articulation. — *J. Acoust. Soc. Amer.*, 1974, v. 56, p. 617—627.
223. Creelman C. D. Human discrimination of auditory duration. — *J. Acoust. Soc. Amer.*, 1962, v. 34, p. 582—593.

224. Dallos P. Low-frequency auditory characteristics: species dependence. — J. Acoust. Soc. Amer., 1970, v. 48, p. 489—499.
225. Dallos P., Cheatham M. A., Ferraro J. Cochlear mechanics, nonlinearities and cochlear potentials. — J. Acoust. Soc. Amer., 1974, v. 55, p. 597—605.
226. Dankbaar W. A. The pattern of stapedial vibration. — J. Acoust. Soc. Amer., 1970, v. 48, p. 1021—1022.
227. Davis H. The cocktail hour before the serious banquet. — In: Basic mechanisms in hearing. N. Y., 1973, p. 259—271.
228. Delattre P. C., Liberman A. M., Cooper F. S. Acoustic loci and transitional cues for consonants. — J. Acoust. Soc. Amer., 1955, v. 27, p. 769—773.
229. Delattre P. C., Liberman A. M., Cooper F. S., Gerstman L. J. An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. — Word, 1952, v. 3, p. 195—210.
230. Derkach M. F. Acoustical cues of softness in Russian syllables and their application in the automatical speech recognition. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1975, p. 349—358.
231. Diehl R. L. Effect of selective adaptation on the identification of speech sounds. — J. Acoust. Soc. Amer., 1974, v. 56, S52(A).
232. Diestel H.-G. Akustische Messungen an einem mechanischen Modell des Innenohres. — Acustica, 1954, v. 4, p. 421—431.
233. Divenyi P. L., Danner W. F. Temporal-order discrimination in nonspeech sequences presented to the left or the right ear. — J. Acoust. Soc. Amer., 1974, v. 56, p. S43.
234. Divenyi P. L., Hirsh I. J. Identification of temporal order in three-tone sequences. — J. Acoust. Soc. Amer., 1974, v. 56, p. 144—151.
235. Djupesland G., Zwislocki J. J. Sound pressure distribution in the outer ear. — Acta Oto-laryngol., 1973, v. 75, p. 350—352.
236. Dorman M. F., Cutting J. E., Raphael L. J. Identification of vowel order: concatenated versus formant-connected sequences. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1974, SR-37/38, p. 217—224.
237. Dorman M. F., Kewley-Port D., Brady S., Turvey M. T. Two processes in vowel recognition: inferences from studies of backward masking. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1974, SR-37/38, p. 233—253.
238. Dowling W. J. Rhythmic fission and perceptual organization. — J. Acoust. Soc. Amer., 1968, v. 44, p. 369(A).
239. Duifhuis H. Consequences of peripheral frequency selectivity for nonsimultaneous masking. — J. Acoust. Soc. Amer., 1973, v. 54, p. 1471—1488.
240. Efron R. The relationship between the duration of a stimulus and the duration of a perception. — Neuropsychologia, 1970, v. 8, p. 37—55.
241. Egan J. P., Schulman A. I., Greenberg G. Z. Operating characteristics determined by binary decisions and by rating. — J. Acoust. Soc. Amer., 1959, v. 31, p. 768—773.
242. Eimas P. D. The relation between identification and discrimination along speech and non-speech continua. — Language and Speech, 1968, v. 6, p. 206—217.
243. Eimas P. D., Cooper W. E., Corbit J. D. Some properties of linguistic feature detectors. — Perception and Psychophysics, 1973, v. 13, p. 247—252.
244. Elenius Y., Rapp K. Acoustic correlates of neutral and emphatic stress. — Ann. Progr. Rep. Inst. of Linguistics, Univ. of Stockholm, 1974, v. 1, p. 25.

245. Elliott L. L. Backward and forward masking of probe tones of different frequencies. — J. Acoust. Soc. Amer., 1962, v. 34, p. 1116—1117.
246. Evans E. F. The frequency response and other properties of single fibers in the guinea-pig cochlear nerve. — J. Physiol., 1972, v. 226, p. 263—287.
247. Evans E. F. Neural processing for the detection of acoustic patterns and for sound localization. — The Neurosciences Third Study Program, MIT Press, 1974, p. 131—145.
248. Evans E. F. Auditory frequency selectivity and the cochlear nerve. — In: Psychophysical models and physiological facts in hearing. Berlin, Heidelberg, N. Y., 1974, p. 118—129.
249. Evans E. F., Nelson P. G. The responses of single neurons in the cochlear nucleus of the cat as a function of their location and the anaesthetic state. — Exp. Brain Res., 1973, v. 17, p. 402—427.
250. Evans E. F., Nelson P. G. On the functional relationship between the dorsal and ventral divisions of the cochlear nucleus of the cat. — Exp. Brain Res., 1973, v. 17, p. 428—442.
251. Evans E. F., Whitfield I. C. Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. — J. Physiol., 1964, v. 171, p. 476—493.
252. Evans E. F., Wilson J. P. The frequency selectivity of the cochlea. — In: Basic mechanisms in hearing. N. Y., London, 1973, p. 519—554.
253. Fant G. Acoustic theory and synthesis of speech with applications to Swedish. — Ericsson Technics, 1959, № 1, p. 3—108.
254. Fant G. Descriptive analysis of the acoustic aspects of speech. — Logos, 1962, v. 5, p. 3—17.
255. Fant G. Automatic recognition and speech research. — Speech Transmiss. Lab. Quart. Progr. and Status Rep., Stockholm, 1970, № 1, p. 16—31.
256. Fant G. Key-note speech. — IEEE Symposium on Speech Recognition, Pittsburgh, 1974, p. 1—5.
257. Fastl H. Über Tonhöhenempfindungen bei Raushen. — Acustica, 1971, v. 25, p. 350—354.
258. Flanagan J. L. Models for approximating basilar membrane displacement. — Bell System Techn. J., 1962, v. 41, p. 959—1009.
259. Flanagan J. L. Focal points in speech communication research. — IEEE Transact. on Communication Technology, 1971, v. COM-19, p. 1006—1015.
260. Flanagan J. L. Speech analysis, synthesis and perception. Berlin, Heidelberg, N. Y., 1972.
261. Flanagan J. L. Voices of men and machines. — J. Acoust. Soc. Amer., 1972, v. 51, p. 1375—1387.
262. Fletcher H. On the dynamics of the cochlea. — J. Acoust. Soc. Amer., 1951, v. 23, p. 637—645.
263. Fry D. B. Duration and intensity as physical correlates of linguistic stress. — J. Acoust. Soc. Amer., 1953, v. 27, p. 765—768.
264. Fry D. B. Experiments in the perception of stress. — Language and Speech, 1958, v. 1, p. 126—152.
265. Fry D. B., Abramson A. S., Eimas P. D., Liberman A. M. The identification and discrimination of synthetic vowels. — Language and Speech, 1962, v. 5, p. 171—189.
266. Fujisaki H., Kawashima T. The influence of various factors on the identification and discrimination of synthetic speech sounds. — Rep. of the 6th Internat. Congr. on Acoustics, Tokyo, 1968, v. 2, p. B-95—B-98.
267. Fujisaki H., Nakamura K., Imoto T. Auditory perception of duration of speech and non-speech stimuli. — Ann. Bull. of Research Inst. of Logopedics and Phoniatrics Univ. of Tokyo, 1973, № 7, p. 45—64.

268. Fujisaki H., Sudo H. Models for the word and sentence pitch contours of Japanese. — Ann. Rep., Engrs. Research Inst., Faculty of Engrs., Univ. of Tokyo, 1970, v. 29, p. 215—221.
269. Gacek R. R. Neuroanatomy of the auditory system. — In: Foundations of modern auditory theory. N. Y., London, 1972, p. 239—262.
270. Gårding E. Internal juncture in Swedish. Lund, 1967.
271. Gay T. A perceptual study of American-English diphthongs. — Language and Speech, 1970, v. 13, p. 65—88.
272. Geisler C. D., Hubbard A. E. New boundary conditions and results for the Peterson-Bogert model of the cochlea. — J. Acoust. Soc. Amer., 1972, v. 52, p. 1629—1634.
273. Gerber S. E. Perception of segmented diphthongs. — Proceed. of the 7th Internat. Congr. of Phonetic Sciences, The Hague, Paris, 1972, p. 479—493.
274. Gerber S. E. Categorical perception of segmented diphthongs. — Preprints of the Speech Communication Seminar, Stockholm, 1974, v. 3, p. 131—136.
275. Goldstein J. L. Auditory nonlinearity. — J. Acoust. Soc. Amer., 1967, v. 41, p. 676—689.
276. Goldstein J. L. An optimum processor theory for the central formation of the pitch of complex tones. — J. Acoust. Soc. Amer., 1973, v. 54, p. 1469—1516.
277. Goldstein J. L., Kiang N. Y.-S. Neural correlates of the aural combination tone  $2f_1 - f_2$ . — Proceed. IEEE, 1968, v. 56, p. 981—992.
278. Goldstein M. H., Abeles M., Daly R. L., McIntosh J. Functional architecture in cat primary auditory cortex: tonotopic organization. — J. Neurophysiol., 1970, v. 33, p. 188—197.
279. Goldstein M. H., Hall J. L., Butterfield B. O. Single unit activity in primary auditory cortex of unanaesthetized cats. — J. Acoust. Soc. Amer., 1968, v. 43, p. 444—455.
280. Green D. M. Temporal auditory acuity. — Psychol. Rev., 1971, v. 78, p. 540—551.
281. Green D. M. Minimum integration time. — In: Basic mechanisms in hearing. N. Y., London, 1973, p. 829—843.
282. Green D. M. Temporal acuity as a function of frequency. — J. Acoust. Soc. Amer., 1973, v. 54, p. 373—379.
283. Gresser J.-Y., Mercier G. Exemple de reconnaissance automatique de la parole. — Communication et Electronique, 1971, № 32, p. 48—63.
284. Gruber J., Boerger G. Binaurale Verdeckungspegeldifferenzen (BMLD) und Vor- und Rückwärtsverdeckung. — Proceed. of the 7th Internat. Congr. on Acoustics, Budapest, 1971, v. 3, p. 457—460.
285. Guerin B. Simulation du conduit vocal. — Bull. de l'Inst. de Phonétique de Grenoble, 1972, v. 1, p. 11—24.
286. Guerin B., Zhukova M. G. Gap detection as a function of the gap temporal location within a noise burst. — Bull. de l'Inst. de Phonétique de Grenoble, 1975, v. 4.
287. Guinan J. J. Jr., Peake W. T. Middle ear characteristics of anaesthetized cats. — J. Acoust. Soc. Amer., 1967, v. 41, p. 1237—1261.
288. Gutknecht C. A note on the role of pitch as an element of the accent within synthetic pairs of syllables. — Proceed. of the 7th Internat. Congr. of Phonetic Sciences, The Hague, Paris, 1972, p. 890—894.
289. Haavel R. Experiments on the perception of emphasis and question. — In: Estonian Papers in Phonetics. Tallin, 1974, p. 33—41.
290. Haavel R. Temporal characteristics of the pitch contour. — In: Estonian Papers in Phonetics. Tallin, 1974, p. 42—49.
291. Haavel R. On the significance on some parameters of a pitch contour. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1975, p. 511—520.



292. Haggard M., Ambler S., Callow M. Pitch as a voicing cue. — J. Acoust. Soc. Amer., 1970, v. 47, p. 613—617.
293. Harris K. S., Hoffman H. S., Liberman A. M., Dellatre P. C., Cooper F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. — J. Acoust. Soc. Amer., 1958, v. 30, p. 122—126.
294. Hart J., Cohen A. Intonation by rule: a perceptual quest. — Instituut voor Perceptie Onderzoek Ann. Progr. Rep., Eindhoven, Holland, 1973, Man. № 242/11.
295. Heinz J. M., Lindblom B. E. F., Lindqvist J. Patterns of residual masking for sounds with speech-like characteristics. Conference on Speech Communication and Processing, Boston, 1967, p. 246—251.
296. Hirato H., Kakusho O., Kato K., Kobayashi T. Harmonic components and vowel perception. — Rep. of the 6th Internat. Congr. on Acoustics, Tokyo, 1968, v. 2, p. B-79—B-82.
297. Hirsch J. J. Auditory perception of temporal order. — J. Acoust. Soc. Amer., 1959, v. 31, p. 759—767.
298. Hoffman H. S. Study of some cues in the perception of the voiced stop consonants. — J. Acoust. Soc. Amer., 1958, v. 30, p. 1035—1041.
299. Holmes J. N., Mattingly I. G., Searme J. N. Speech synthesis by rule. — Language and Speech, 1964, v. 7, p. 127—143.
300. Houtgast T. Psychophysical evidence for lateral inhibition in hearing. — J. Acoust. Soc. Amer., 1972, v. 51, p. 1885—1894.
301. Houtgast T. Auditory analysis of vowel-like sounds. — Acustica, 1974, v. 31, p. 320—324.
302. Houtgast T. Lateral suppression in hearing. Soesterberg, Holland, 1974.
303. Huggins A. W. F. On the perception of temporal phenomena in speech. — J. Acoust. Soc. Amer., 1972, v. 51, p. 1279—1290.
304. Hurme P. T. Oudon kielen matkimisesta: Ihmisen kyvystä ja rajoi-tuksista imitoida oudon kielen lauseita. — V. Fonetiikan Paperit, Hel-singin Fonetiikan Laitoksen Julkaisuja, 1975.
305. Jakobson R., Fant G., Halle M. Preliminaries to speech analysis. The distinctive features and their correlates. — Massachusetts Inst. of Technology, Acoustics Lab. Techn. Rep., 1952, № 13, p. 19—58.
306. Janota P. Personal characteristics of speech. — Rozprawy Českoslo-venské Akademie Věd, 1967, v. 77, p. 1—72.
307. Jassem W., Morton J., Steffen-Batog M. The per-ception of stress in synthetic speech-like stimuli by Polish listeners. — In: Speech Analysis and Synthesis. Warsaw, 1968, p. 289—308.
308. Johnstone B. M., Boyle A. J. F. Basilar membrane vibrations examined with the Mössbauer technique. — Science, 1967, v. 158, p. 389—390.
309. Johnstone B. M., Taylor K. J., Boyle A. J. Mechanics of the guinea-pig cochlea. — J. Acoust. Soc. Amer., 1970, v. 47, p. 504—509.
310. Johnstone B. M., Yates G. K. Basilar membrane tuning curves in the guinea pig. — J. Acoust. Soc. Amer., 1974, v. 55, p. 584—587.
311. Kaiser L. Some properties of speech muscles and the influence there-of on language. — Arch. Neerlandaises de Phonétique Experimental, 1934, v. 10, p. 121—133.
312. Karcevskij S. Sur la Phonologia de la Phrase. — In: A Prague School Reader in Linguistics. Bloomington, 1964, p. 206—252.
313. Karnickaya E. G., Mushnikov V. N., Slepokurova N. A., Zhukov S. Ja. Auditory processing of steady-state vowels. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1973, p. 37—53.
314. Katwijk A. van. The perception of stress. — Instituut voor Perceptie Onderzoek Ann. Progr. Rep., Eindhoven, Holland, 1969, № 4, p. 69—73.



315. Katwijk A. van, Govaert G. A. Prominence as a function of the location of pitch movement. — *Instituut voor Perceptie Onderzoek Ann. Progr. Rep.*, Eindhoven, Holland, 1967, № 2, p. 115—117.
316. Kay R. H., Matthews D. R. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. — *J. Physiol.*, 1973, v. 225, p. 657—678.
317. Keidel W. D. Information processing in the higher parts of the auditory pathways. — In: *Psychophysical models and physiological facts in hearing*. Berlin, Heidelberg, N. Y., 1974.
318. Kelly J. B., Whitfield I. C. Effects of auditory cortical lesions on discriminations of rising and falling frequency-modulated tones. — *J. Neurophysiol.*, 1971, v. 34, p. 802—816.
319. Kiang N. Y.-S., Moxon E. C. Tails of tuning curves of auditory nerve fibers. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 620—630.
320. Kiang N. Y.-S., Pfeiffer R. R., Warr W. B., Backus A. S. N. Stimulus coding in the cochlear nucleus. — *Ann. Otol. Rhinol. Laryngol.*, 1963, v. 54, p. 463—485.
321. Kiang N. Y.-S., Watanabe T., Thomas E. C., Clark L. F. Discharge patterns of single fibers in the cat's auditory nerve. — *Mass. Inst. Technol. Techn. Rep.*, 1965, Monograph № 35.
322. Kim D. O., Molnar C. E., Pfeiffer R. R. A system of nonlinear differential equations modelling basilar-membrane motion. — *J. Acoust. Soc. Amer.*, 1973, v. 54, p. 1517—1529.
323. Klatt D. H. Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 8—16.
324. Klatt D. H. On the design of understanding systems. — *Preprints of the Speech Communication Seminar*, Stockholm, 1974, v. 3, p. 277—289.
325. Klatt D. H., Peterson G. H. Reexamination of a model of the cochlea. — *J. Acoust. Soc. Amer.*, 1966, v. 40, p. 54—61.
326. Klatt D. H., Shattuck S. R. Perception of brief stimuli that resemble rapid formant transitions. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1973, p. 293—391.
327. Klatt D. H., Stevens K. N. On the automatic recognition of continuous speech: implications from a spectrogram-reading experiment. — *IEEE Transact. on Audio and Electroacoust.*, 1973, v. AU—21, p. 210—227.
328. Klein W., Plomp R., Pols L. C. W. Vowel spectra, vowel spaces and vowel identification. — *J. Acoust. Soc. Amer.*, 1970, v. 48, p. 999—1009.
329. Klump R. G., Webster J. C. Intelligibility of time-compressed speech. — *J. Acoust. Soc. Amer.*, 1961, v. 33, p. 265—267.
330. Koenig W., Dunn H. K., Eacy L. J. The sound spectrograph. — *J. Acoust. Soc. Amer.*, 1946, v. 18, p. 19—49.
331. Koerber K. C., Pfeiffer R. R., Warr W. B., Kiang N. Y.-S. Spontaneous spike discharges from single units in the cochlear nucleus after destruction of the cochlea. — *J. Exper. Neurol.*, 1966, v. 16, p. 119—130.
332. Kohlöffel L. U. E. A study of basilar membrane vibrations. I. Fuzziness-detection: a new method for the analysis of microvibrations with laser light. — *Acustica*, 1972, v. 27, p. 49—65.
333. Kohlöffel L. U. E. A study of basilar membrane vibrations. II. The vibratory amplitude and phase pattern along the basilar membrane (post-mortem). — *Acustica*, 1972, v. 27, p. 66—81.
334. Kohlöffel L. U. E. A study of basilar membrane vibrations. III. The basilar membrane frequency response curve in the living guinea pig. — *Acustica*, 1972, v. 27, p. 82—89.
335. Kohlöffel L. U. E. Observations of the mechanical disturbances along the basilar membrane with laser illumination. — In: *Basic mechanisms in hearing*. N. Y., London, 1973, p. 95—117.

336. Krivnova O. F. Some notes on melody component of intonation as a cue in text segmentation. — Preprints of the Symposium on Auditory Analysis and Perception of Speech, Leningrad, 1973.
337. Kudela K. A study of the optimal formant frequency values of Polish vowels using synthetic speech. — In: Speech analysis and synthesis. Warsaw, 1970, v. 2, p. 219—238.
338. Labutin V. K. Analog of auditory nerve response to acoustical signals and its relation to some psychoacoustic phenomena. — *Acustica*, 1974, v. 31, p. 342—345.
339. Lane H. The motor theory of speech perception: a critical review. — *Psychol. Rev.*, 1963, v. 72, p. 275—309.
340. Lea W. A. An approach to syntactic recognition without phonemics. — *IEEE Transact. on Audio and Electroacoust.*, 1973, v. 21, p. 249—258.
341. Lea W. A., Medress M. F., Skinner T. E. Use of syntactic segmentation and stressed syllable location in phonemic recognition. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 356 (A).
342. Lehiste I. An acoustic-phonetic study of internal open juncture. — *Phonetica*, 1960, v. 5, Suppl.
343. Lehiste I. Juncture. — *Proceed. of the 5th Internat. Congr. of Phonetic Sciences*, Basel, 1965, p. 172—200.
344. Lehiste I. Readings in acoustic phonetics. Cambridge, Mass., 1967.
345. Lehiste I. Suprasegmentals. Cambridge, Mass., London, 1970.
346. Lehiste I. The units of speech perception. — In: *Speech and cortical functioning*. N. Y., London, 1972, p. 187—237.
347. Lehiste I. Phonetic disambiguation of syntactic ambiguity. — *Glossa*, 1973, v. 7, p. 107—122.
348. Leshowitz B., Hanzl R. Auditory pattern discrimination in the absence of spectral cues. — *J. Acoust. Soc. Amer.*, 1972, v. 52, p. 166 (A).
349. Liberman A. M. Some results of research on speech perception. — *J. Acoust. Soc. Amer.*, 1957, v. 29, p. 117—123.
350. Liberman A. M., Cooper F. S., Harris K. S., MacNeillage P. F. A motor theory of speech perception. — *Proceed. of the Speech Communication Seminar*, Stockholm, 1962, Session D. 3, p. 1—10.
351. Liberman A. M., Delattre P. C., Cooper F. S., Gerstman L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. — *Psychol. Monogr.*, 1954, v. 68, p. 1—13.
352. Liberman A. M., Delattre P. C., Gerstman L. J., Cooper F. S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. — *J. Exper. Psychol.*, 1956, v. 52, p. 127—137.
353. Liberman A. M., Harris K. S., Hoffman H. S., Griffith B. C. The discrimination of speech sound within and across phoneme boundaries. — *J. Exper. Psychol.*, 1957, v. 54, p. 358—368.
354. Liberman Ph. Some acoustic correlates of word stress in American English. — *J. Acoust. Soc. Amer.*, 1960, v. 32, p. 451—454.
355. Liberman Ph. On the acoustic basis of the perception of intonation by linguistics. — *Word*, 1965, v. 21, p. 40—54.
356. Liberman Ph. *Intonation, perception and language*. Cambridge, Mass., 1967.
357. Lilienkrantz J. The OVE III speech synthesizer. — *Speech Transmiss. Lab. Quart. Progr. and Status Rep.*, Stockholm, 1967, № 2/3, p. 76—81.
358. Lindblom B. Sona-Graph measurements. — *Speech Transmiss. Lab. Quart. Progr. and Status Rep.*, Stockholm, 1964, № 3, p. 3—5.
359. Lindblom B. Spectrographic study of vowel reduction. — *J. Acoust. Soc. Amer.*, 1963, v. 35, p. 1773—1781.
360. Lindblom B. Motor control mechanisms. — *Preprints of the Speech Communication Seminar*, Stockholm, 1974, Chairman review of papers in session 2, p. 1—14.

361. Lindblom B., Studdert-Kennedy M. On the role of formant transitions in vowel recognition. — J. Acoust. Soc. Amer., 1967, v. 42, p. 830—843.
362. Lindblom B., Svensson S.-G. Interaction entre facteurs segmentaux et non segmentaux dans la reconnaissance de la parole. — Journées d'Etudes sur la Parole, Lannion, 1972, p. 199—213.
363. Lindner G. Veränderung der Beurteilung synthetischer Vokale unter dem Einfluss des Sukzessivkontrastes. — Zeitschr. für Phonetik, 1966, Bd 19, S. 287—307.
364. Lindner G. Akustische Tauschungen. — Zeitschr. für Psychologie, 1967, Bd 173, S. 248—267.
365. Lindqvist J., Pauli S. The role of relative spectrum levels in vowel perception. — Rep. of the 6th Internat. Congr. on Acoustics, Tokyo, 1968, v. 2, p. B-91—B-94.
366. Lisker L. Minimal cues for separating (w, r, l, y) in intervocalic position. — Word, 1957, v. 13, p. 257—267.
367. Lisker L., Abramson A. S. Cross-language study of voicing in initial stops. — J. Acoust. Soc. Amer., 1963, v. 35, p. 1889—1890.
368. Lisker L., Abramson A. S. A cross-language study of voicing in initial stops: acoustical measurements. — Word, 1964, v. 20, p. 384—422.
369. Lisker L., Abramson A. S. The voicing dimension: some experiments in comparative phonetics. — Haskins Lab. Status Rep. on Speech Research, N. Y., 1967, SR-11, p. 9—15.
370. Löfqvist A. Some phonetic correlates of emphatic stress in Swedish. — Phonetic Lab., Dept. of Gen. Linguistics, Lund Univ., Working Papers, 1965, № 10, p. 105—115.
371. Lublinskaja V. V., Mikiel W. Scaling of pitch intervals between vowel-like sounds with constant and variable fundamental frequency. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1973, p. 233—242.
372. MacNeilage P. F. Speech physiology. — In: Speech and cortical functioning. N. Y., London, 1972, p. 1—72.
373. Maeda Sh. A characterization of fundamental frequency contours of speech. — Mass. Inst. Technol. Res. Lab. of Electronics, Quart. Progr. Rep., 1974, № 114, p. 193—211.
374. Maiwald D. Beziehungen zwischen Schallepektrum, Mithörschwelle und der Erregung des Gehörs. — Acustica, 1961, v. 18, p. 69—80.
375. Makhoul J. Linear prediction vs. analysis-by-synthesis. — Preprints of the Speech Communication Seminar, Stockholm, 1974, v. 1, p. 35—43.
376. Massaro D. W. Preperceptual images, processing time, and perceptual units in auditory perception. — Psychol. Rev., 1972, v. 79, p. 124—145.
377. Massaro D. W. Perceptual units in speech recognition. — J. Exper. Psychol., 1974, v. 102, p. 199—208.
378. Mattingly I. G., Liberman A. M., Syrdal A. K., Halwes T. Discrimination in speech and nonspeech modes. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1970, SR-21/22, p. 99—132.
379. McGill W. J. Neural counting mechanisms and energy detection in audition. — J. Mathemat. Psychol., 1967, v. 4, p. 351—376.
380. McGill W. J., Goldberg J. P. Pure-tone intensity discrimination and energy detection. — J. Acoust. Soc. Amer., 1968, v. 44, p. 576—581.
381. McRobert H., Bryan M. E., Tempest W. The acoustic stimulation of the middle ear muscles. — J. Sound. Vib., 1968, v. 7, p. 129—142.
382. Menon K. M. N., Rao P. V. S., Thomas R. B. Formant transitions and stop consonants perception in syllables. — Language and Speech, 1974, v. 17, p. 27—46.

383. Merzenich M. M., Knight P. L., Roth G. L. Orderly representation of the cochlea within primary auditory cortex in the cat. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 586—587.
384. Michon J. A. Studies on subjective duration. I. Differential sensitivity in the perception of repeated temporal intervals. — *Acta Psychologica*, 1964, v. 22, p. 441—450.
385. Miller G. A., Taylor W. Perception of repeated bursts of noise. — *J. Acoust. Soc. Amer.*, 1948, v. 20, p. 171—182.
386. Mol H., Uhlenbeck E. M. The linguistic relevance of intensity in stress. — *Lingua*, 1956, v. 5, p. 205—213.
387. Møller A. R. Transfer function of the middle ear. — *J. Acoust. Soc. Amer.*, 1963, v. 35, p. 1526—1534.
388. Møller A. R. Unit responses in the cochlear nucleus of the rat to sweep tones. — *Acta Physiol. Scand.*, 1969, v. 76, p. 503—512.
389. Møller A. R. Unit responses in the rat cochlear nucleus to tones of rapidly varying frequency and amplitude. — *Acta Physiol. Scand.*, 1971, v. 81, p. 540—556.
390. Møller A. R. The middle ear. — In: *Foundations of modern auditory theory*. N. Y., London, 1973, v. 2, p. 133—194.
391. Møller A. R. Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat. — *Acta Physiol. Scand.*, 1972, v. 86, p. 223—258.
392. Møller A. R. The acoustic middle ear muscle reflex. — In: *Handbook of sensory physiology*. Berlin, Heidelberg, N. Y., 1974, v. 1, p. 519—548.
393. Møller A. R. Coding of sounds with rapidly varying spectrum in the cochlear nucleus. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 631—640.
394. Møller A. R. Responses of units in the cochlear nucleus to sinusoidally amplitude-modulated tones. — *Exper. Neurol.*, 1974, v. 45, p. 104—117.
395. Moore B. C. J. Frequency difference limens for short-duration tones. — *J. Acoust. Soc. Amer.*, 1973, v. 54, p. 610—619.
396. Moore B. C. J. Frequency difference limens for narrow bands of noise. — *J. Acoust. Soc. Amer.*, 1973, v. 54, p. 888—896.
397. Moore B. C. J., Raab D. H. Pure-tone intensity discrimination: some experiments relating to the «near-miss» to Weber's law. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 1049—1054.
398. Morton J., Jassem W. Acoustic correlates of stress. — *Language and Speech*, 1965, v. 8, p. 159—181.
399. Mowbray G. H., Gebhard I. W., Byham C. L. Sensitivity to changes in the interruption of white noise. — *J. Acoust. Soc. Amer.*, 1956, v. 28, p. 106—110.
400. Mushnikov V. N., Slepokurova N. A., Zhukov S. Ya. On the auditory correlate of the second formant of vowels. — *Preprints of the 8th Internat. Congr. on Acoustics*, London, 1974, p. 323.
401. Nábelek I. V., Hirsh I. J. On the discrimination of frequency transitions. — *J. Acoust. Soc. Amer.*, 1969, v. 45, p. 1510—1519.
402. Nábelek I. V., Nábelek A. K., Hirsh I. J. Pitch of tone bursts of changing frequency. — *J. Acoust. Soc. Amer.*, 1970, v. 48, p. 536—553.
403. Nábelek I. V., Nábelek A. K., Hirsh I. J. Pitch of sound bursts with continuous or discontinuous change of frequency. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 1305—1312.
404. Nábelek I. V., Rozsypal A. J. Discrimination of the envelope curve for tone and noise bursts. — *J. Acoust. Soc. Amer.*, 1970, v. 48, p. 70 (A).
405. Neazelnitsky V. Measurements of sound pressure in the cochlea of anaesthetized cats. — In: *Psychophysical models and physiological facts in hearing*. Berlin, Heidelberg, N. Y., 1974, p. 45—55.
406. Nelson P. G., Erulkar S. D., Bryan J. S. Responses of units of the inferior colliculus to time-varying acoustic stimuli. — *J. Neurophysiol.*, 1966, v. 29, p. 834—860.



407. Newell A., Barnett J., Forgie J., Green C., Klatt D., Licklider J., Munson J., Reddy R., Woods W. Speech-understanding systems: Final report of a study group. — Published by Carnegie-Mellon Univ., Pittsburgh, Pennsylvania, 1974.
408. Noorden L. P. van. Temporal coherence in the perception of tone sequences. — Doctoral Thesis, Eindhoven, The Netherlands, 1975.
409. Nootboom S. G. Some timing factors in the production and perception of vowels. — Instituut voor Perceptie Onderzoek, Eindhoven, Holland, 1974. Manuscript № 221.
410. Nootboom S. G. Contextual variation and the perception of phonemic vowel length. — Preprints of the Speech Communication Seminar, Stockholm, 1974. v. 3, p. 149—153.
411. O'Connor J. D., Gerstman L. J., Liberman A. M., Delattre P. C., Cooper F. S. Acoustic cues for the perception of initial (w, j, r, l) in English. — Word, 1957. v. 13, p. 24—43.
412. O'Connor J. D., Tooley O. M. The perception of some words' boundaries. — In: In Honour of Daniel Jones. London, 1964. p. 171—177.
413. Oettinger R., Hauser H. An electrical network for the investigation of the mechanical vibrations of the inner ear. — Acustica, 1961. v. 11, p. 161—177.
414. Ohman S. Word and sentence intonation. A quantitative model. — Speech Transmiss. Lab. Quart. Progr. and Status Rep., Stockholm, 1967. № 2—3, p. 20—54.
415. Orr D. B., Friedman H. L., Williams J. C. Trainability of listening compression of speeded discourse. — J. Educ. Psychol., 1965. v. 56, p. 148—156.
416. Peterson G. E., Lehiste I. Duration of syllable nuclei in English. — J. Acoust. Soc. Amer., 1960. v. 32, p. 693—703.
417. Peterson J. L., Liden G. Some static characteristics of the stapedial muscle reflex. — Audiology, 1972. v. 11, p. 97—114.
418. Peterson L. C., Bogert B. P. A dynamical theory of the cochlea. — J. Acoust. Soc. Amer., 1950. v. 22, p. 369—381.
419. Petrov V., Sorin C., Tjul'kov W. Just audible amplitude modulation of noise as a function of the frequency and the number of cycles of modulating sine-wave. — Bull. de l'Inst. de Phonétique de Grenoble, 1973. v. 4.
420. Pfeiffer R. R. Classification of response patterns of spike discharges for units in the cochlear nucleus: tone-burst stimulation. — Exp. Brain Res., 1966. v. 1, p. 220—235.
421. Pickett J. M., Decker L. R. Time factors in the perception of a double consonant. — Language and Speech, 1960. v. 3, p. 11—17.
422. Pierce J. R. Whither speech recognition? — J. Acoust. Soc. Amer., 1969. v. 46, p. 1049—1051.
423. Pisoni D. B. Perceptual processing time for consonants and vowels. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1972. SR-31/32, p. 83—92.
424. Plopp R. Rate of decay of auditory sensation. — J. Acoust. Soc. Amer., 1964. v. 36, p. 277—282.
425. Plopp R. Auditory analysis and timbre perception. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1975. p. 7—22.
426. Pollack I. Auditory flutter. — Amer. J. Psychol., 1952. v. 65, p. 544—554.
427. Pollack I. Onset discrimination for white noise. — J. Acoust. Soc. Amer., 1963. v. 35, p. 607—609.
428. Pollack I. Number of pulses required for minimal pitch. — J. Acoust. Soc. Amer., 1967. v. 42, p. 895 (L).
429. Pollack I. Detection of rate changes of «auditory frequency». — J. Exper. Psychol., 1968. v. 77, p. 535—541.



430. Pollack I. Detection and relative discrimination of auditory «jitter». — J. Acoust. Soc. Amer., 1968, v. 43, p. 308—315.
431. Pols L. C. W., Kamp L. J. van der, Ploomp R. Perceptual and physical space of vowel sounds. — J. Acoust. Soc. Amer., 1969, v. 46, p. 458—467.
432. Pols L. C. W., Tromp H. R. C., Ploomp R. Frequency analysis of Dutch vowels from 50 male speakers. — J. Acoust. Soc. Amer., 1973, v. 53, p. 1093—1101.
433. Potapova R. K. Auditory estimate of syllable and vowel duration in sentences. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1973, p. 465—483.
434. Potter R. K., Kopp G. A., Green H. C. Visible speech. N. Y., 1947.
435. Price G. R. Correspondence between cochlear microphonic sensitivity and behavioural threshold in the cat. — J. Acoust. Soc. Amer., 1971, v. 49, p. 1899—1901.
436. Quancard J. P. Synthèse des travaux effectués en France dans le domaine de la reconnaissance de la parole. — Journées d'Etudes sur la Parole, Lannion, 1972, p. 347—345.
437. Raab D. H. Forward and backward masking between acoustic clicks. — J. Acoust. Soc. Amer., 1961, v. 33, p. 137—139.
438. Rand T. C. Dichotic release from masking for speech. — Haskins Lab. Status Rep. on Speech Research. New Haven, 1973, SR-33, p. 47—55.
439. Raphael L. J. Vowel duration as a cue to the perceptual separation of cognate sounds in American English. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1971, Suppl.
440. Raphael L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. — J. Acoust. Soc. Amer., 1972, v. 51, p. 1296—1303.
441. Reddy D. R. Speech recognition: prospects for the seventies. — Proceed. of IFIP Congr. 71, Ljubljana, Yugoslavia, 1974, p. I-5—I-13.
442. Reichardt W., Niese H. Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level. — J. Acoust. Soc. Amer., 1970, v. 47, p. 1083—1090.
443. Rhode W. S. Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. — J. Acoust. Soc. Amer., 1971, v. 49, p. 1218—1231.
444. Rhode W. S. An investigation of post-mortem cochlear mechanics using the Mössbauer effect. — In: Basic mechanisms in hearing. N. Y., London, 1973, p. 49—67.
445. Rhode W. S., Deisler C. D. Measurement of the amplitude and phase of vibration of the basilar membrane using the Mössbauer effect. — J. Acoust. Soc. Amer., 1970, v. 47, p. 60 (A).
446. Rhode W. S., Robles L. Evidence from Mössbauer experiments for nonlinear vibration in the cochlea. — J. Acoust. Soc. Amer., 1974, v. 55, p. 588—596.
447. Robles L., Rhode W. S. Nonlinear effects in the transient response of the basilar membrane. — In: Psychophysical models and physiological facts in hearing. Berlin, Heidelberg, N. Y., 1974, p. 43—55.
448. Ronken D. A. Monaural detection of a phase difference between clicks. — J. Acoust. Soc. Amer., 1970, v. 47, p. 1091—1099.
449. Ronken D. A. Masking produced by sinusoids of slowly changing frequency. — J. Acoust. Soc. Amer., 1973, v. 54, p. 905—915.
450. Rose J. E. Discharges of single fibers in the mammalian auditory nerve. — In: Frequency analysis and periodicity detection in hearing. Leiden, 1970, p. 176—188.
451. Rose J. E., Brugge J. F., Anderson D. J., Hind J. E. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. — J. Neurophysiol., 1967, v. 30, p. 769—793.

452. Rose J. E., Brugge J. F., Anderson D. J., Hind J. E. Patterns of activity in single auditory nerve fibers of the squirrel monkey. — In: *Hearing mechanisms in vertebrates*. London, 1968, p. 144—157.
453. Rose J. E., Galambos R., Hughes J. R. Microelectrode studies of the cochlear nuclei of the cat. — *Bull. Johns Hopkins Hosp.*, 1959, v. 104, p. 211—251.
454. Rose J. E., Greenwood D. D., Goldberg J. M., Hind J. E. Some discharge characteristics of single neurons in the inferior colliculus of the cat. I. Tonotopic organization, relation of spike-counts to tone intensity, and firing patterns of single elements. — *J. Neurophysiol.*, 1963, v. 26, p. 294—320.
455. Rupert A. L., Moushegian G. Neuronal responses of Kangaroo rat ventral cochlear nucleus to low frequency tones. — *Exper. Neurol.*, 1970, v. 26, p. 84—102.
456. Sachs M. B. Stimulus-response relation for auditory-nerve fibers: two-tone stimuli. — *J. Acoust. Soc. Amer.*, 1969, v. 45, p. 1025—1036.
457. Sachs M. B., Kiang N. Y.-S. Two-tone inhibition in auditory nerve fibers. — *J. Acoust. Soc. Amer.*, 1968, v. 43, p. 1120—1128.
458. Scholes R. J. Categorical responses to synthetic vocalic stimuli by speakers of various languages. — *Language and Speech*, 1967, v. 10, p. 252—282.
459. Schönfeld M. Einheitliche Darstellung der Theorien über das System Steigbügel Basilarmembran in Innenohr. — *Zeitschr. elektr. Inform. and Energietechn.*, 1973, Bd 3, S. 9—21.
460. Schouten J. F. The residue revised. — In: *Frequency analysis and periodicity detection in hearing*. Leiden, 1970, p. 54—58.
461. Schroeder M. R. An integrable model for the basilar membrane. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 429—434.
462. Schroeder M. R., Hall J. L. Model for mechanical to neural transduction in the auditory receptor. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 1055—1060.
463. Sekey A. Short-term auditory frequency discrimination. — *J. Acoust. Soc. Amer.*, 1963, v. 35, p. 682—690.
464. Sergeant R. L., Harris J. D. Sensitivity to unidirectional frequency modulation. — *J. Acoust. Soc. Amer.*, 1962, v. 34, p. 1625—1628.
465. Sharf D. J., Beiter R. C. Identification of consonants from formant transitions presented forward and backward. — *Language and Speech*, 1974, v. 17, p. 110—116.
466. Shepard R. W. Analysis of proximities as a technique for the study of information processing in man. — *Human Factors*, 1963, v. 5, p. 33—48.
467. Sigurd B. Maximum rate and minimal duration of repeated syllables. — *Language and Speech*, 1973, v. 16, p. 373—395.
468. Singh S. A step towards a theory of speech perception. — *Preprints of the Speech Communication Seminar*, Stockholm, 1974, v. 3, p. 55—66.
469. Small A. M. Pure tone masking. — *J. Acoust. Soc. Amer.*, 1959, v. 31, p. 1619—1625.
470. Small A. M., Daniloff R. G. Pitch of noise bands. — *J. Acoust. Soc. Amer.*, 1962, v. 41, p. 506—512.
471. Smith R. L. Short-term adaptation and incremental responses of single auditory nerve fibers. — *Lab. of Sensory Communication Syracuse Univ. Spec. Rep.*, N. Y., 1973.
472. Smoorenburg G. F. Audibility region of combination tones. — *J. Acoust. Soc. Amer.*, 1972, v. 52, p. 603—614.
473. Sone T., Tsumura T. Detection of frequency transition. — *Proceed. of the 7th Internat. Congr. on Acoustics*, Budapest, 1971, v. 3, p. 525—528.
474. Sorkin R. D. Extension of theory of signal detectability to matching procedures in psychoacoustics. — *J. Acoust. Soc. Amer.*, 1962, v. 34, p. 1745—1751.

475. Spoendlin H. The organization of the cochlear receptor. Basel, N. Y., 1966.
476. Spoendlin H. Structural basis of peripheral frequency analysis. In: Frequency analysis and periodicity detection in hearing. Leiden, 1970, p. 2—36.
477. Stein H. J. Das Absinken der Mithörschwelle nach dem Abschalten von weissem Rauschen. — *Acustica*, 1960, v. 10, p. 116—119.
478. Stevens J. C., Shickman G. M. The perception of repetition rate. — *J. Exper. Psychol.*, 1959, v. 58, p. 433—440.
479. Stevens K. N. Acoustical aspects of speech production. — In: Handbook of physiology. Washington, 1964, Sect. 3, Respiration, v. 1, p. 347—355.
480. Stevens K. N. On the relation between speech movements and speech perception. — *Proceed. of the 18th Internat. Congr. of Psychology, Moscow*, 1966, Symp. 23, p. 68—74.
481. Stevens K. N. Acoustic correlates of certain consonantal features. Conference on Speech Communication and Processing, Boston, 1967, p. 177—184.
482. Stevens K. N. The role of rapid spectrum changes in the production and perception of speech. — In: Form and substance. Copenhagen, 1971, p. 95—101.
483. Stevens K. N. The potential role of property detectors in the perception of consonants. — In: Auditory analysis and perception of speech. London, N. Y., San Francisco, 1975, p. 303—330.
484. Stevens K. N., House A. S. Perturbation of vowel articulations by consonantal context: an acoustical study. — *J. Speech and Hearing Res.*, 1963, v. 6, p. 111—128.
485. Stevens K. N., House A. S. Speech perception. — In: Foundations of modern auditory theory. N. Y., London, 1972, v. 2, p. 3—62.
486. Stevens K. N., House A. S., Paul A. P. Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic model of articulation. — *J. Acoust. Soc. Amer.*, 1966, v. 40, p. 123—132.
487. Stevens K. N., Kasowski S., Fant G. An electrical analog of the vocal tract. — *J. Acoust. Soc. Amer.*, 1953, v. 25, p. 734—742.
488. Stevens K. N., Klatt D. H. Role of formant transitions in the voiced-voiceless distinction for stops. — *J. Acoust. Soc. Amer.*, 1974, v. 55, p. 653—659.
489. Stevens K. N., Liberman A. M., Studdert-Kennedy M., Ohman S. Cross-language study of vowel perception. — *Language and Speech*, 1969, v. 12, p. 1—23.
490. Stewart J. L. A theory and physical model for cochlear mechanics. — *Acta Oto-laryngol.*, 1972, Suppl. 294.
491. Studdert-Kennedy M. The perception of speech. — Haskins Lab. Status Rep. on Speech Research, New Haven, 1970, SR-23, p. 15—48.
492. Studdert-Kennedy M., Liberman A. M., Stevens K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. — *J. Acoust. Soc. Amer.*, 1963, v. 35, p. 1900 (A).
493. Suen Ch. Y., Beddoes M. P. Discrimination of vowel sounds of very short duration. — *Perception and Psychophysics*, 1972, v. 11, p. 417—419.
494. Suga N. Recovery cycles and responses to frequency-modulated tone pulses in auditory neurons of echo-locating bats. — *J. Physiol.*, 1964, v. 175, p. 50—80.
495. Suga N. Responses of inferior collicular neurones of bats to tone bursts with different rise time. — *J. Physiol.*, 1971, v. 217, p. 159—177.
496. Suzuki H. Mutually complementary effect of rate and amount of formant transition in distinguishing vowel, semivowel, and stop consonants. — *Mass. Inst. Technol. Res. Lab. of Electronics, Quart. Progr. Rep.*, 1970, № 96, p. 164—172.

497. Suzuki H. Mutually complementary effect between the amount and rate of formant change in perception of speech-like sounds. — *Proceed. of the 7th Internat. Congr. on Acoustics, Budapest, 1971*, v. 3, p. 5—8.
498. Svensson S.-G. A preliminary study of the role of prosodic parameters in speech perception. — *Speech Transmiss. Lab. Quart. Progr. and Status Rep.*, Stockholm, 1971, № 2/3, p. 24—40.
499. Svensson S.-G. Prosody and grammar in speech perception. — *Monographs from the Inst. of Linguistics Univ. of Stockholm, Stockholm, 1974*, v. 2.
500. Terhardt E. Über akustische Rauigkeit und Schwankungsstärke. — *Acustica*, 1968, v. 20, p. 215—224.
501. Terhardt E. Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. — In: *Frequency analysis and periodicity detection in hearing*. Leiden, 1970, p. 278—287.
502. Terhardt E. On the perception of periodic sound fluctuations (roughness). — *Acustica*, 1974, v. 30, p. 201—213.
503. Thomas I. B., Cetti R. P., Chase P. W. Effect of silent intervals on the perception of temporal order for vowels. — *J. Acoust. Soc. Amer.*, 1971, v. 49, p. 85 (A).
504. Thomas I. B., Hill P. B., Carroll F. S., Garcia B. Temporal order in the perception of vowels. — *J. Acoust. Soc. Amer.*, 1970, v. 48, p. 1010—1013.
505. Tomisawa M., Kirikae I., Shitara T., Sato T., Fujisaki H. Identification of synthetic vowels in perceptive deafness. — *Rep. of the 6th Internat. Congr. on Acoustics, Tokyo, 1968*, v. 4, p. A-77—A-80.
506. Tonndorf J., Khanna S. M. Tympanic membrane vibrations in human cadaver ears studied by time-averaged holography. — *J. Acoust. Soc. Amer.*, 1972, v. 52, p. 1221—1233.
507. Trager G. L., Smith H. L. An outline of English structure. Ann Arbor, 1951.
508. Tsuchitani C., Boudreau J. C. Single unit analysis of cat superior olive S segment with tonal stimuli. — *J. Neurophysiol.*, 1966, v. 29, p. 684—699.
509. Tsumura T., Sone T., Nimura T. Auditory detection of frequency transition. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 17—26.
510. Turnbull W. W. Pitch discrimination as a function of tonal duration. — *J. Exper. Psychol.*, 1944, v. 34, p. 302—316.
511. Vaissiere J. On French prosody. — *Mass. Inst. Technol. Res. Lab. of Electronics Quart. Progr. Rep.*, 1974, № 114, p. 212—223.
512. Vaissiere J. A further note on French prosody. — *Mass. Inst. Technol. Res. Lab. of Electronics Progr. Rep.*, 1975, № 115, p. 251—261.
513. Viemeister N. F. Intensity discrimination of pulsed sinusoids: the effects of filtered noise. — *J. Acoust. Soc. Amer.*, 1972, v. 51, p. 1265—1269.
514. Vogel A. Ein gemeinsames Funktionsschema zur Beschreibung der Lautheit und der Rauigkeit. — *Diss. Technischen Univ. München*, 1974.
515. Vogten L. L. M. Pure-tone masking of a phase-locked tone burst. — *Instituut voor Perceptie Onderzoek Ann. Progr. Rep.*, Eindhoven, Holland, 1972, № 7, p. 5—16.
516. Vogten L. L. M. Pure tone masking; a new result from a new method. — In: *Psychophysical models and physiological facts in hearing*. Berlin, Heidelberg, N. Y., 1974, p. 142—155.
517. Wamond C. On the mechanism of hearing. — *Philips Res. Rep.*, 1962, № 1, p. 1—140.
518. Warren R. M. Relation of verbal transformations to other perceptual phenomena. — *Conf. of Pattern Recognition Inst. of Electrical Engineers, Teddington, England, 1968*, Conf. publ. № 42.
519. Warren R. M. Auditory temporal discrimination by trained listeners. — *Cogn. Psychol.*, 1974, v. 6, p. 237—256.



520. Warren R. M., Obusek C. J. Identification of temporal order within auditory sequences. — *Perception and Psychophysics*, 1972, v. 12, p. 86—90.
521. Warren R. M., Obusek C. J., Farmer R. M., Warren R. P. Auditory sequence: confusion of patterns other than speech or music. — *Science*, 1966, v. 164, p. 586—587.
522. Watson C. S., Rilling M. E., Bourbon W. T. Received-operating characteristics determined by a mechanical analog to the rating scale. — *J. Acoust. Soc. Amer.*, 1964, v. 36, p. 283—288.
523. Weiss T. F., Peake W. T. Cochlear potential response at the round window membrane of the cat. — A reply to the comment of G. R. Price. — *J. Acoust. Soc. Amer.*, 1972, v. 52, p. 1729—1734.
524. Weiss T. F., Peake W. T., Sohmer H. S. Intracochlear potential recorded with micropipets. III. Relation of cochlear microphonic potential to stapes velocity. — *J. Acoust. Soc. Amer.*, 1971, v. 50, p. 602—615.
525. Wells R. S. Immediate constituents. — *Language*, 1947, v. 23, p. 83—117.
526. Wever G. G., Vernon J. A., Rahm W. E., Strother W. F. Cochlear potentials in the cat in response to high-frequency sounds. — *Proceed. Nat. Acad. Sci.*, 1953, v. 44, p. 1087—1090.
527. Wickelgren W. A. Context-sensitive coding, associative memory, and serial order in (speech) behaviour. — *Psychol. Rev.*, 1969, v. 1, p. 1—15.
528. Wiener F. M., Ross D. A. The pressure distribution in the auditory canal in a progressive sound field. — *J. Acoust. Soc. Amer.*, 1946, v. 18, p. 401—408.
529. Willems Y. D., Allen J. The use of prosodic features to derive word and syllable boundaries. — *J. Acoust. Soc. Amer.*, 1973, v. 53, p. 355 (A).
530. Wilson J. P. Basilar membrane vibration data and their relation to theories of frequency analysis. — In: *Psychophysical models and physiological facts in hearing*. Berlin, Heidelberg, N. Y., 1974, p. 56—63.
531. Wilson J. P., Johnstone J. R. Basilar membrane correlates of the combination tone  $2f_1 - f_2$ . — *Nature*, 1971, v. 241, p. 206—207.
532. Wilson J. P., Johnstone J. R. Capacitive probe measures of basilar membrane vibration. — In: *Hearing theory* 1972, Symposium held at IPO, Eindhoven, Holland, 1972, p. 172—181.
533. Wood S. Speech tempo. — *Phonetics Lab. Lund Univ. Working Papers*, 1973, № 9, p. 99—147.
534. Woolsey C. N. Electrophysiological studies on thalamocortical relations in the auditory system. — In: *Unfinished tasks in the behavioural sciences*. Baltimore, 1964.
535. Zlatin M. A. Voicing contrast: perceptual and productive voice onset time characteristics of adults. — *J. Acoust. Soc. Amer.*, 1974, v. 56, p. 981—994.
536. Zlatoustova L. V. Rhythmic structure types in Russian speech. — In: *Auditory analysis and perception of speech*. London, N. Y., San Francisco, 1975, p. 477—483.
537. Zwerg G., Lipes R., Pierce J. R. The cochlear compromise. — *J. Acoust. Soc. Amer.*,
538. Zwicker E. Der ungewöhnliche Amplitudengang der nichtlinearen Verzerrungen des Ohres. — *Acustica*, 1955, v. 5, p. 67—74.
539. Zwicker E. Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs. — *Acustica*, 1956, v. 6, p. 356—381.
540. Zwicker E. Subdivision of the audible frequency range into critical bands. — *J. Acoust. Soc. Amer.*, 1961, v. 33, p. 248—249.
541. Zwicker E. Subjektive und objektive Dauer von Schallimpulsen und Schallpausen. — *Acustica*, 1969—1970, v. 22, p. 214—218.



542. Zwicker E. Masking and psychological excitations as consequences of the ear's frequency analysis. — In: Frequency analysis and periodicity detection in hearing. Leiden, 1970, p. 376—394.
543. Zwicker E. On a psychoacoustical equivalent of tuning curves. — In: Psychophysical models and physiological facts in hearing. Berlin, Heidelberg, N. Y., 1974, p. 132—139.
544. Zwicker E., Scharf B. A model of loudness summation. — Psychol. Rev., 1965, v. 72, p. 3—26.
545. Zwislocki-Mosicki J. Theorie der Schneckenmechanik. — Acta Oto-laryngol., 1948, Suppl. 72, p. 1—76.
546. Zwislocki J. Theory of the acoustical action of the cochlea. — J. Acoust. Soc. Amer., 1950, v. 22, p. 778—784.
547. Zwislocki J. Review of recent mathematical theories of cochlear dynamics. — J. Acoust. Soc. Amer., 1953, v. 25, p. 743—751.
548. Zwislocki J. Analysis of some auditory characteristics. — Lab. of Sensory Communication Syracuse Univ. Semiannual Res. Rep., Syracuse, N. Y., 1963.
549. Zwislocki J., Pirodda E., Rubin H. On some poststimulatory effects at the threshold of audibility. — J. Acoust. Soc. Amer., 1959, v. 31, p. 9—14.
550. Zwislocki J., Sokolich W. G. Loudness as a function of tone duration. — Lab. of Sensory Communication Syracuse Univ. Semiannual Res. Rep., Syracuse, N. Y., 1971, LSC-15.
551. Zwislocki J., Sokolich W. G. Neuro-mechanical frequency analysis in the cochlea. — In: Psychophysical models and physiological facts in hearing. Berlin, Heidelberg, N. Y., 1974, p. 107—117.

551

171

---

380

## ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Адаптация периферическая 209,  
211—214, 217, 220, 236
- Акустические события 140, 158, 160,  
162—164, 167, 170, 341, 342, 344
- Анализ речи, морфологический 5, 33,  
34, 36, 38, 39, 41, 92, 135, 138
- — синтаксический 5, 33, 34, 36,  
38, 74, 75, 92, 135
- — спектральный, см. Динами-  
ческая спектрограмма
- — технический 28—34
- Базиллярная мембрана 179—181, 183,  
190—192, 197—199, 207—208, 216
- — частотно-избирательные свой-  
ства 184, 185, 187—190, 195, 203,  
205, 206
- Внутреннее ухо 11, 172, 174, 176  
см. также Улитка
- Глайд 314—319
- Гласный (е) 21, 22, 24, 25, 27, 32,  
40, 53—55, 57—59, 61—63, 65,  
68—71, 77, 79, 82, 83, 85, 86,  
93, 95—97, 99, 105, 107—109,  
111, 113, 142
- — восприятие 58, 59, 61—68, 79—  
82, 84, 95—97, 101, 105, 133, 142,  
148, 157, 159, 248, 249, 260,  
264, 270, 271, 275, 312, 313, 322,  
331, 334
- — распознавание последователь-  
ностей 84, 142—146, 152, 167,  
247, 259, 312
- — синтетические 79—81, 83—85,  
105, 144, 146, 148, 154, 155, 159,  
257—260, 262, 263, 265, 267, 269,  
307, 311
- — ударные 49, 65, 95, 96, 121, 122,  
125, 127, 128, 131, 133, 135, 163
- Громкость 95—97, 132, 136, 251,  
255, 277—279, 298
- — плотность громкости 255—257
- Детекторы признаков согласных 87,  
89
- — — «селективная адаптация» 89
- — изменений частоты 335—337, 338
- — форм огибающей сигнала 298,  
300, 302
- Динамическая спектрограмма 22—  
25, 40, 51, 56, 61, 75, 76, 105,  
139, 142, 155, 157, 163, 164, 340
- Интонация фразовая 113, 114, 116,  
118—121, 124, 125—127, 130, 131,  
133, 134, 137
- Информация, грамматическая 47
- — лексическая 93
- — просодическая 12, 49, 63, 70,  
71, 91—94, 96, 101, 113, 131—  
138, 341
- — источники 92—94
- — параметры 132—134, 136
- — семантическая 47
- — фонемная 43, 49, 61, 68, 91, 93,  
104, 131, 132, 137, 138
- Контекстуальные влияния, восприя-  
тия 65, 84, 85
- — — эффект, ассимиляции 85, 86
- — — контраста 85—87
- — — речесоборования 65
- — — эффект коартикуляции 32,  
112, 113
- Контур, формантный 160, 266, 311,  
314, 336
- — частоты основного тона (мело-  
дический) 113—117, 119—121,  
123—131, 133, 134, 137, 304, 308
- — интонационная единица 115—  
117, 120
- Кортнев орган 177, 178, 192
- Критическая полоса 201, 202, 252,  
254, 271, 272, 290, 296, 316
- Латеральное торможение 239, 263
- Маскировка звуков 142, 145, 200,  
202, 203, 205, 211, 213, 251, 252, 303,  
304

- — обратная 145, 146, 280
- — одновременная 252, 304
- — остаточная 212, 213, 217, 219, 220, 305

Модель, восприятия речи 4, 6, 7, 35, 38, 39, 82, 339, 341, 342

- латерального торможения 226, 230, 236, 238, 263—267

- преобразования сигнала в периферической слуховой системе 213—221
- функциональная оценки хриплости 296
- функциональная периферической слуховой системы 208, 209, 222, 250

Морфема 38, 39, 41, 45, 71

Наружное ухо 172, 173, 190

Описание речевого сигнала 9, 10, 16, 35, 36, 39, 138, 139, 157, 158, 342

- — — методы 11
- — — параметрическое 9, 11, 12, 158
- — — полное 10
- — — спектральное 9, 10, 158
- — — фонемное 36, 39, 57, 75
- — — фонетическое 9, 11, 12, 138

Память, долговременная 36

- кратковременная (оперативная) 12, 41, 42, 45, 49—51, 94, 104, 109, 110, 113, 131, 132, 134—138, 146
- — объем 41—44
- — структура 94, 135, 136

— «сенсорная» 139, 152, 155, 157

Признаки, гласных 79, 84, 96, 138, 139, 247, 271, 274

- — гипотеза(ы), кодирования величинной возбуждения 248, 249, 262

- — — пространственного кодирования 248, 249

- — — формантная 249, 263, 266
- — — мелодического контура 114, 134, 306, 307, 311

- — — просодические 91—94, 96, 100, 103, 109, 113, 114, 131, 134, 135, 137, 139

- — — психологические (дифференциальные) 138

- — — согласных 74, 138, 139

Психоакустика 6, 90, 95, 107, 170, 250, 266, 267, 320, 339, 344

- психометрическая функция 112, 280

- решающие правила 54, 82, 84, 85, 92, 135—137, 159

- функция, идентификации 64, 68, 74, 75, 77—79, 81, 84, 88, 260, 297

- — различия 74, 75, 78, 79, 81, 89

Речевой сигнал 3, 5, 9, 11, 12, 20, 25—28, 31, 34—36, 38—40, 43, 44, 49, 50—58, 61, 70, 73, 74, 75, 82, 87, 90—95, 104—106, 113, 124, 131, 132, 137, 138, 140, 158, 168, 236, 318, 337, 340

- — описание, см. Описание речевого сигнала

- — параметры 92, 100, 136, 139

- — фонетическая интерпретация 6, 12, 34—36, 50, 52, 57, 138, 141, 170, 247, 339, 344

Речеобразование 12, 14, 16, 22, 25, 26, 31, 33, 40, 57, 58, 61, 65, 82, 101, 104, 132, 133

- акустические явления 12, 14, 16, 17, 19, 22

- аэродинамические процессы 12, 14—19

- физиология 12, 14—16, 19, 21, 22

Ритм (речевой) 43, 44, 49, 51—53, 92, 120, 141, 150, 151

Сегментация, речевого потока 11, 12, 30, 55, 56, 95, 109, 132—134, 152, 155, 170, 308, 338

- — — выделение начала и конца гласного 57, 61, 70, 109, 133, 155, 326

- — — на слова 40, 47

- — — словесная граница 91, 104, 110, 112, 113, 135

- — — — пограничный, аллофон 91, 112

- — — — — сигналы 40, 92, 104, 113

- — — — — признак 109, 110, 112, 113

Синтез речи 22, 25—28

Слог(и) 40—44, 46—49, 53—55, 57, 65, 69, 70, 76, 77, 85, 87—89, 92—94, 97, 98, 100, 116, 117, 119, 121, 135, 141, 142, 148—150, 157, 161, 162, 168, 322, 323, 325, 331

- — — длительность 44, 53, 100, 101, 105, 106, 109, 111, 141, 146, 151, 152, 158

- — — закрытый 49, 54, 69, 70, 109

- — — открытый 49, 54, 69, 70, 109

- — — условия обнаружения 146, 148, 152, 155, 161

- Слуховая обработка, огибающей стимула 258, 276, 278, 279, 281, 287, 292, 294, 295, 320
- — — выделение неравномерностей 251, 285, 287, 288, 293, 295, 297, 337
- — — сигналов с изменяющимся во времени спектром 303—320
- Слуховой, нерв 179, 181, 198, 199, 203, 207, 216, 219, 226, 232, 252
- — амплитудная характеристика нейронов 211—213
- — нейронный (частотный) канал 208—210, 213, 222, 223, 225, 247
- — частотно-пороговые характеристики 200—207
- — путь (восходящий) 226—228, 231—240, 242—245
- Слуховая система, тонотопическая организация 227, 228
- Слуховой спектр стимула 247, 250, 256, 257, 259, 263, 266
- — — выделение неоднородностей 262—264, 273—275
- — — процедура вычисления 254—257
- — — экспериментальное измерение 251—253
- Согласный(е) 25, 27, 32, 40, 49, 53, 55, 58, 59, 61, 65, 67—70, 74, 75, 77—79, 81, 85, 88, 101, 102, 109, 111—113, 133, 279, 298, 310, 322, 324, 325, 331, 333, 335
- — — восприятие 87, 89, 102, 110, 112, 133, 138, 149, 159, 161, 164, 165, 261, 278, 294, 308, 323, 327
- Среднее ухо 172, 174, 179, 182, 190
- Суммация, временная 223, 224
- — — пространственная 223, 224
- Темп речевой 65, 92, 104, 105, 107—110, 132, 133, 135, 136, 151, 158
- — — восприятие 105—109
- — — «мгновенный» 104, 106, 109
- — — мера 104, 106, 108, 109
- Теория, акустическая речеобразования 14, 16, 20, 25
- — — восприятия речи (моторная) 31, 32, 54, 73, 94, 245
- — — дифференциальных признаков 72, 73
- — — образования речевого сигнала 20
- — — слухового спектрального анализа 207
- Ударение, словесное 19, 43, 47—49, 94, 95, 97, 100, 103, 109, 114, 116, 121, 124, 135
- — — восприятие 95—97, 98, 100—103, 119—124
- — — признаки 96, 97, 103, 113, 120, 124, 131
- — — роль длительности гласного и согласного в слоге 97—103
- — — частота основного тона 101, 113—117, 119—126, 128, 129, 131, 133
- — — логическое (фразовое) 104, 124—126, 128, 130, 135
- — — восприятие 125, 126, 128, 129, 131
- — — типы 124, 131
- Улитка 174, 179—182, 184, 185, 187, 188, 190, 191, 193—197, 205—208, 232, 247, 250, 253, 276, 288, 289, 292, 296, 339
- — — потенциал микрофонный 182, 183, 188
- — — строение, 176, 177, 179, см. также Кортиев орган, Базиллярная мембрана
- Фонема 30—32, 34—36, 38—41, 45, 48, 49, 52—55, 63, 68—73, 77, 79—86, 89, 91, 92, 104, 126, 130, 140, 145, 146, 152, 154, 160, 164, 267, 340
- — — варианты 31, 71, 91, 93, 94, 135
- — — комбинаторные 40
- — — позиционные 40
- — — гласная 40, 55, 63, 68—70, 72, 152, 155
- — — граница 65, 77—79, 83—88, 90, 126, 160, 162, 163, 259, 260, 261, 312—314, 323, 332—334
- — — лингвистическая 71, 82, 314
- — — признаки, дифференциальные 32, 48, 49, 72, 73, 92
- — — психологические 73
- — — психологическая, аллофон 30, 71, 81, 82, 314
- — — распознавание 35, 54—56, 68, 275
- — — параллельное 54—56
- — — последовательное 54—56
- — — решение 68, 73—75, 79, 82, 84—86
- — — согласная 55, 69, 109, 141, 160
- Фонетическая последовательность 39, 41—43, 46, 49—53
- — — фонемный состав 41, 51, 52, 54
- — — ритмико-мелодический рисунок 51, 93, 119
- Форманта 22, 24, 27, 28, 55, 63, 75, 77, 80, 83, 84, 86, 87, 90, 92, 157, 160, 161, 249, 259—264, 266, 267, 269—275, 307, 310, 312, 313, 318, 322—326, 331—334, 336
- Формантный переход 61—63, 89, 113, 144, 163, 164, 322, 325, 326, 332, 333, 343

## ИМЕННОЙ УКАЗАТЕЛЬ

- Авакян Р. В. 162, 297; [41]  
 Алексин Р. О. 142, 168; [157]  
 Алякринский В. В. 15, 42, 49, 53, 69, 104, 124, 335; [159]  
 Адресян Ю. Д. 39; [2]  
 Арутюнян Э. А. 124, 133; [3]  
 Архипов Г. В. 107; [4]  
 Бару А. В. 297; [41]  
 Барышникова К. К. 114; [5]  
 Башкина Б. М. 97, 130, 131; [6, 7]  
 Белозерова Л. В. 298; [9]  
 Богданов Б. В. 146; [10]  
 Богос И. И. 34; [11]  
 Бондаренко В. П. 194; [12]  
 Бондарко Л. В. 15, 33, 40, 42, 47, 49, 53, 54, 61, 69, 81, 96, 104, 113, 124, 135, 335; [13-21, 159, 195]  
 Бороздин А. Н. 57-59, 153, 154; [93, 154]  
 Бровченко Т. А. 95; [22]  
 Ваймер А. Ф. 93; [51]  
 Вардапетян Г. А. 242; [23]  
 Вартамян И. А. 231, 232, 239, 240; [24-27, 42]  
 Величко В. М. 29; [28]  
 Венцов А. В. 53, 101, 105, 111, 115, 116, 133; [29-31, 83]  
 Вербицкая Л. А. 40, 81, 96, 135, 310; [14-17, 32]  
 Виндюк Т. К. 29, 34; [33, 34]  
 Волоцкая З. М. 36; [36]  
 Волошин Г. Я. 22; [37]  
 Высоцкий Г. Я. 34; [38]  
 Гайдучик С. М. 114; [5]  
 Галунов В. И. 31, 92; [39, 40]  
 Гаузелеман В. Е. 223, 224; [58]  
 Гейльман Н. И. 131, 137; [122]  
 Гершуни Г. В. 6, 239, 297; [27, 41, 42]  
 Гитис В. Г. 29; [133]  
 Глезер В. Д. 223; [43]  
 Голубцов С. В. 26, 28; [44-46]  
 Голузина А. Г. 15, 42, 49, 53, 69, 79, 80, 104, 124, 306, 335; [47, 48, 49, 159]  
 Гольдбурт С. Н. 150, 151; [49]  
 Горелик А. Л. 29; [50]  
 Горский С. М. 93; [51]  
 Гранстрем М. П. 147, 148, 159, 171, 213, 214, 216, 219, 285; [52, 162, 211]  
 Гумецкий Р. Я. 25, 34; [53, 55]  
 Деркач М. Ф. 25, 34, 160, 308, 309; [11, 53, 56, 230]  
 Дубровский Н. А. 278, 282, 283; [59, 57]  
 Дудкин К. Н. 223, 224; [43, 58]  
 Дукедьский Н. И. 55; [59]  
 Ермолова Е. Н. 25; [53]  
 Жиникин Н. И. 96, 101; [60]  
 Жолковский А. К. 36; [61]  
 Жуков С. Я. 57-59, 61, 84, 146, 148, 151, 155, 159, 160, 161, 251, 264, 270, 278, 279, 285, 289, 290, 295; [52, 62-64, 93, 313, 400]  
 Жукова М. Г. 147, 148, 153, 154, 159, 160, 161, 278, 279, 284, 285, 289, 290, 293-295, 298-302; [52, 62, 66-67, 102, 154, 214, 286]  
 Загоруйко Н. Г. 29, 33; [18, 28, 68, 69]  
 Зимняя И. А. 125, 129; [70]  
 Зиндер Л. Р. 40, 47, 54, 81, 96, 107, 113, 115; [15, 16, 19, 20, 71-73]  
 Златоустова Л. В. 96, 101, 120; [74, 536]  
 Зыкова М. А. 131, 137; [122]  
 Иванова В. А. 280; [155]  
 Казюкова Н. 104, 124; [76]  
 Карницкая Э. Г. 84, 251, 254, 255, 264; [77, 313]  
 Каспранский Р. Р. 93; [51]  
 Катанская А. Р. 104, 124; [78]  
 Клаас Ю. А. 15, 42, 47, 49, 53, 69, 104, 124, 142, 152, 154, 168, 335; [79, 156, 157, 159]  
 Книппер А. В. 29; [80]



- Рожевников В. А. 15, 33, 42, 49, 53, 57—59, 69, 81, 85, 104, 124, 146—148, 151, 153, 154, 159, 162, 171, 211—213, 219, 285, 323, 325, 335; [18, 52, 63, 81, 93, 154, 158—160, 211]
- Кривнова О. Ф. 114—116, 118, 119; [82, 83, 336]
- Крылов Н. П. 106, 107; [84]
- Кузьмин Ю. П. 15, 42, 49, 53, 69, 85, 104, 124, 146, 147, 151, 155, 156, 335; [63, 85, 86, 159]
- Кукольник Л. Е. 40; [21]
- Куперман А. М. 223; [43]
- Лабутина В. К. 194, 213, 217; [87, 338]
- Леонтьева Н. Н. 36; [61]
- Лесогор Л. В. 171, 211—214, 216, 219, 289; [160, 162, 211]
- Леушина Л. П. 223; [43]
- Лисенко Д. М. 15, 42—44, 46, 49, 53, 57—60, 69, 85, 87, 104, 113, 124, 149, 150, 155, 156, 279, 335; [64, 85, 86, 90—93, 159, 161, 214]
- Лобанов Б. М. 97, 130, 131; [6, 7]
- Люблинский П. А. 109; [94]
- Люблинская В. В. 15, 42, 49, 53, 69, 104, 124, 306, 307, 309, 311, 313, 322, 325, 335; [48, 96—100, 159, 371]
- Лян Чжи-ань 308, 318, 319; [101]
- Малининкова Т. Г. 153, 154; [154]
- Манеров В. Х. 92; [39]
- Мартемьянов Ю. С. 36; [61]
- Мельчук И. А. 36; [103, 104]
- Миронова Л. М. 297; [41]
- Мишин Л. 34; [55]
- Молодая Т. Н. 36; [36]
- Молчанов А. П. 33, 194; [18, 87]
- Морозов В. П. 283; [106]
- Мушников В. П. 83, 84, 147, 148, 159, 251, 261—264, 270, 274, 275, 283, 285, 288, 295, 298—300; [52, 102, 106—110, 313, 400]
- Мясников Л. Л. 28; [111, 112]
- Невская А. А. 223; [43]
- Николаева Т. М. 36; [36]
- Оверченко М. 34; [55]
- Павлова Л. П. 40, 81; [16, 21]
- Папов В. М. 112, 120; [113]
- Петров В. 285—288; [418]
- Петров О. А. 29; [80]
- Подвигин Н. Ф. 223; [43]
- Позин Н. В. 199, 222, 263; [114]
- Потанова Р. К. 51, 101; [433]
- Праздник Н. В. 223; [43]
- Пухирь Н. В. 85; [115]
- Равич Р. Д. 36; [104]
- Радинова Е. Л. 232, 234; [116]
- Ратникова Г. П. 227, 232; [117]
- Рейтблат Л. Е. 162, 323, 325; [81]
- Реформатский А. А. 71; [118]
- Родионов В. Д. 147, 148, 159, 204, 285; [52]
- Рохтла М. К. 306; [48]
- Рудный Б. П. 29, 34; [38, 119]
- Самойлова Н. К. 212, 217; [120]
- Сапожкова М. А. 10, 11, 22, 26; [121]
- Светозарова Н. Д. 40, 47, 113, 120, 122—124, 131, 137; [20, 21, 122, 123]
- Селезнева Н. С. 96, 101; [125]
- Серков Ф. П. 242—245; [126—128]
- Скрипкин В. А. 29; [50]
- Слепокурова Н. А. 83, 84, 251, 264, 270, 309, 313; [129, 313, 400]
- Столярова Э. Н. 284
- Сторожук В. М. 242, 244, 245; [126]
- Тальнов А. П. 245; [128]
- Талысин П. А. 171, 213, 219; [211]
- Темов В. Л. 251, 264; [131]
- Трубецкой Н. С. 41, 72, 73, 91, 104, 132; [132]
- Трунин—Донской В. П. 29, 34; [38, 119]
- Тумаркина Л. Н. 278, 282, 283; [56, 57]
- Турбович Н. Т. 29; [133]
- Тюльков В. А. 171, 213, 219, 285—288; [211, 249]
- Федорова Н. А. 15, 42—44, 49, 53, 57—60, 69, 95—97, 98, 100, 104, 124, 149, 150, 279, 335; [92, 93, 136, 159, 161, 214]
- Фомичев В. А. 125, 129; [70]
- Фрейдлин А. А. 202; [139]
- Хаавель Р. 125—126, 128—131, 133, 134; [289—291]
- Хайретдинова А. Г. 93; [141]
- Харкевич А. А. 11, 306; [142, 143]
- Холщевников В. Е. 53; [144]
- Цемель Г. П. 28, 34; [146, 147]
- Цеплитис Л. К. 93, 105; [148]
- Чабан М. 34; [55]
- Черниговская Т. В. 283; [105]
- Чистович Н. А. 147, 148, 159, 171, 213, 214, 216, 219, 283—285, 288, 295; [52, 162, 211]
- Чистович Л. А. 15, 31, 33, 42, 49, 53, 57—61, 69, 81, 83, 85, 100, 104, 105, 124, 134, 142, 146—148, 151—154, 159, 160, 162, 163,

- 168, 200, 203, 211—214, 216, 261—263, 274, 275, 278—280, 285, 289, 290, 295, 298—300, 306, 308, 318, 319, 323, 325, 335; [18, 40, 48, 52, 62, 63, 81, 93, 101, 102, 106, 107—110, 149—163, 212—214]
- Шейкина О. 306, 307; [100]  
 Шенгели Г. А. 47; [164]  
 Штерн А. С. 40; [21]  
 Шуляков В. С. 15, 42, 49, 53, 69, 104, 124, 160, 161, 163, 171, 197, 213, 219, 261, 262, 335; [159, 163, 165—169, 211]  
 Шулякова Р. М. 15, 42, 49, 53, 69, 104, 124, 335; [159]  
 Щерба Л. В. 39, 72, 115; [170]  
 Щербакова Л. И. 85, 120, 122—124, 131, 135, 322; [17, 115, 123, 171]  
 Юрков Е. Ф. 29; [133]  
 Яновский Е. И. 243—245; [127, 128]  
 Яхно В. П. 109; [94]  
 Aaronson D. 146; [172]  
 Abel S. M. 96, 103, 105, 329; [173]  
 Abeles M. 227, 228, 230, 240, 245; [174, 175, 278]  
 Abramson A. S. 75—79, 85, 87, 322; [176—178, 265, 367—369]  
 Ades A. E. 87; [179]  
 Aitkin I. M. 245; [180]  
 Allen J. 93; [529]  
 Ambler S. 134; [292]  
 Anderson D. J. 198, 209; [201, 451, 452]  
 Atal B. S. 31; [181, 182]  
 Babkoff H. 280; [183]  
 Backus A. S. N. 234; [320]  
 Bailey P. L. 87, 89, 90, 314; [184, 185]  
 Barnett J. 34; [407]  
 Bauch H. 194, 197; [186]  
 Beddoes M. P. 146; [493]  
 Beiter R. C. 322, 325; [465]  
 Bekesy G. 117, 174, 183—185, 187, 190, 192, 197; [8, 187—189]  
 Bell C. G. 31; [190]  
 Bismark G. 259; [191]  
 Boer E., de 291; [197]  
 Boerger G. 212; [284]  
 Bogdanski D. F. 240; [192]  
 Bogert B. P. 192, 194, 197; [193, 418]  
 Bond Z. S. 164, 165; [194]  
 Boomer D. S. 93; [196]  
 Borst J. M. 25, 26, 54, 322; [220, 221]  
 Bos C. E. 291; [197]  
 Boudreau J. C. 227; [508]  
 Bourbon W. T. 136; [522]  
 Boyle A. J. F. 97, 174, 175, 185; [308, 309]  
 Brady P. T. 314, 316; [198]  
 Brady S. 145; [237]  
 Bregman A. S. 168; [199]  
 Brown W. S. 124; [200]  
 Brugge J. F. 198, 209; [201, 451, 452]  
 Bryan J. S. 240; [406]  
 Bryan M. E. 174; [381]  
 Bürck W. 165, 166, 170; [202]  
 Burlet H. M. 176; [203]  
 Butterfield B. O. 240; [279]  
 Byham C. L. 106, 107; [399]  
 Callow M. 134; [292]  
 Campbell J. 168; [199]  
 Campbell R. A. 292—294; [204]  
 Carlson R. 83, 84, 104, 124, 259, 267—269; [205—208]  
 Carroll F. S. 142, 143, 153; [504]  
 Carroll J. D. 73; [209]  
 Centmayer K. 85, 86; [210]  
 Cetti R. P. 143, 153; [503]  
 Chang J. J. 73; [209]  
 Chase P. W. 143, 153; [503]  
 Cheatham M. A. 207; [225]  
 Chomsky N. 73, 82; [215]  
 Christov Ph. 159; [216]  
 Clark L. 211; [321]  
 Cohen A. 114, 271; [217, 294]  
 Coker C. H. 26, 90, 91, 112; [138, 218]  
 Collier R. 119; [219]  
 Cooper F. S. 25, 26, 31, 54, 55, 267, 322, 323, 325, 331, 333; [220, 221, 228, 229, 293, 350—352, 411]  
 Cooper W. E. 87—89; [222, 243]  
 Corbit J. D. 87—89; [243]  
 Creelman C. D. 96, 103, 105; [223]  
 Cutting J. E. 143, 144; [236]  
 Dallos P. 183, 188, 207; [224, 225]  
 Daly R. L. 227; [278]  
 Daniloff R. G. 262; [470]  
 Dankbaar W. A. 174; [226]  
 Danner W. F. 168, 170; [233]  
 Davis H. 34, 117; [54, 227]  
 Decker L. R. 109, 110; [421]  
 Deisler C. D. 174, 186; [445]  
 Delattre P. C. 25, 54, 55, 267, 322, 323, 325, 331, 333; [220, 228, 229, 293, 351, 352, 411]  
 Diehl R. L. 87; [231]  
 Diestel H.—G. 194; [232]  
 Dittman A. T. 93; [196]  
 Divenyi P. L. 168, 170, 303; [233, 234]  
 Djupesland G. 173; [235]  
 Dorman M. F. 143—145; [236, 237]  
 Dowling W. J. 168; [238]  
 Duifhuis H. 280; [239]

- Dunlop C. W. 245; [180]  
Dunn H. K. 11, 24; [330]
- Efron R. 153, 154; [240]  
Egan J. P. 136; [241]  
Eimas P. D. 79, 85, 87, 89; [242, 243, 265]  
Elenius Y. 104, 124; [244]  
Elliott L. L. 212; [245]  
Erikson Y. 104, 124; [205]  
Erulkar S. D. 240; [406]  
Evans E. F. 201—203, 205, 207, 227, 228, 232—235, 240, 241, 244; [246—262]  
Fant G. 20, 22, 25, 27, 33, 73, 83—85, 152, 154, 160, 257, 259, 267—269; [134, 135, 169, 206—208, 213, 253—256, 305, 487]  
Farmer R. M. 139, 168; [521]  
Fastl H. 262; [267]  
Feldtkeller R. 202, 251, 254—256, 282, 283, 290, 291, 316; [145]  
Ferraro J. 207; [225]  
Flanagan J. L. 10—12, 26, 33, 34, 172, 173, 183; [137, 138, 258—261]  
Fletcher H. 192; [262]  
Forgie J. 34; [407]  
Franks L. E. 31; [140]  
Friedman H. L. 146; [415]  
Fry D. B. 79, 85, 87, 97, 101, 113, 121; [263—265]  
Fujisaki H. 31, 64, 80, 81, 83, 96, 103, 105, 133, 141; [190, 266—268, 505]
- Gacek R. R. 227; [269]  
Galambos R. 227, 240; [192, 453]  
Garcia B. 142, 143, 153; [504]  
Gårding E. 91, 112, 113; [270]  
Gay T. 309; [271]  
Gebhard I. W. 106, 107; [399]  
Geisler C. D. 193, 194; [272]  
Gerber S. E. 309; [273, 274]  
Gerstman L. J. 25, 54, 55, 267, 322, 323, 325, 331, 333; [220, 229, 351, 352, 411]  
Goldberg J. M. 227; [454]  
Goldberg J. P. 291; [380]  
Goldstein J. L. 197, 207, 303; [275—277]  
Goldstein M. N. 227, 228, 230, 240; [174, 175, 278, 279]  
Govaert G. A. 124; [315]  
Gränstrom B. 83, 84, 104, 124, 259, 267—269; [205—208]  
Green C. 34; [407]  
Green D. M. 276, 279, 280; [280—282]  
Green H. C. 24; [434]  
Greenberg G. Z. 136; [241]  
Greenwood D. D. 227; [454]
- Gresser J.-Y. 34; [283]  
Griffith B. C. 79; [353]  
Gruber J. 212; [284]  
Guerin B. 27, 300—302; [285, 286]  
Guinan J. J. Jr. 174, 175; [287]  
Gutknecht C. 122; [288]
- Haggard M. 134; [292]  
Hall J. L. 213, 215, 240; [279, 462]  
Halle M. 73, 82; [215, 305]  
Halwes T. 79; [378]  
Hanauer S. L. 31; [181]  
Hanzi R. 168; [348]  
Harris J. D. 320; [464]  
Harris K. S. 31, 55, 79, 323; [293, 350, 353]  
't Hart J. 114, 119; [219, 294]  
Hauser H. 194; [413]  
Heinz J. M. 31, 303, 305, 314, 315, 323; [190, 295]  
Hill P. B. 142, 143, 153; [504]  
Hind J. E. 198, 209, 227; [201, 451, 452, 454]  
Hirato H. 257, 259, 264, 265; [296]  
Hirsh I. J. 165, 166, 303, 314, 315, 317, 319, 327—331; [234, 297, 401—404]  
Hoffman H. S. 79, 323; [293, 298, 353]  
Holmes J. N. 27; [299]  
House A. S. 31, 42, 90, 311, 314, 316; [190, 198, 484—486]  
Houtgast T. 252—254, 257, 258; [300—302]  
Hubbard A. E. 193, 194; [272]  
Huggins A. W. F. 53; [303]  
Hughes J. R. 227; [453]  
Hurme P. T. 43, 44; [304]
- Imoto T. 64, 96, 103, 105, 141; [297]
- Jakobson R. 73; [305]  
Janota P. 83, 84; [306]  
Jassem W. 96, 97, 121, 122; [307, 398]  
Johntone B. M. 97, 174, 175, 185, 187, 197; [308—310]  
Johnstone J. R. 186, 187; [531, 532]
- Kaiser L. 146; [311]  
Kakusho O. 257, 259, 264, 265; [296]  
Kamp L. J., van der 248, 249; [431]  
Karcevskij S. 118; [312]  
Kasowski S. 27; [487]  
Kasson J. M. 171, 194; [130]  
Kato K. 257, 259, 264, 265; [296]  
Katwijk A., van 96, 124; [314, 315]  
Kawashima T. 80, 81; [266]  
Kay R. H. 336, 337; [316]  
Keidel W. D. 241; [317]  
Kelly J. B. 321, 336; [318]  
Kewley-Port D. 145; [237]

- Khanna S. M. 175, 176; [506]  
 Kiang N. Y.-S. 171, 197, 202—204, 207, 211, 232, 234, 252; [277, 319—321, 331, 457]  
 Kim D. O. 197; [322]  
 Kirikae I. 83; [505]  
 Klatt D. H. 25, 33, 34, 121, 163, 193, 194, 326; [323—327, 407]  
 Klein W. 248, 249; [328]  
 Klump R. G. 146; [329]  
 Knight P. L. 227; [383]  
 Kobayashi T. 257, 259, 264, 265; [296]  
 Koenig W. 11, 24; [330]  
 Koerber K. C. 232; [331]  
 Kohlöffel L. U. E. 186, 187, 190; [332—335]  
 Kopp G. A. 24; [434]  
 Kotowski P. 165, 166, 170; [202]  
 Kudela K. 83, 84; [337]
- Lacy L. J. 11, 24; [330]  
 Lane H. 31; [339]  
 Lasky E. Z. 292—294; [204]  
 Lea W. A. 93; [340, 341]  
 Lehiste I. 25, 61, 91, 112, 113, 146; [342—347, 416]  
 Leshowitz B. 168; [348]  
 Liberman A. M. 25, 26, 31, 54, 55, 79, 82, 85, 87, 267, 322, 323, 325, 331, 333; [220, 221, 228, 229, 265, 293, 349—353, 378]  
 Lichte H. 165, 166, 170; [202]  
 Licklider J. 34; [407]  
 Liden G. 174; [417]  
 Lieberman Ph. 93, 95, 118; [354—356]  
 Liliencrants J. 26; [357]  
 Lindblom B. 25, 33, 93, 104, 124, 303, 305, 311—315, 323; [205, 205, 358—362]  
 Lindgren H. 28, 30; [58]  
 Lindner G. 83, 85; [363, 364]  
 Lindqvist J. 260, 303, 305, 314, 315, 323; [295, 365]  
 Lindsey P. 36, 41, 337; [59]  
 Lipes R. 179, 194; [537]  
 Lisker L. 76—78, 322, 323; [177, 178, 366—369]  
 Löfqvist A. 124; [370]
- MacNeilage P. F. 16, 31, 33, 55; [360, 372]  
 Maeda Sh. 115; [373]  
 Maiwald D. 255; [374]  
 Makhoul J. 31; [375]  
 Massaro D. W. 139, 141, 145; [376, 377]  
 Matthews D. R. 336, 337; [316]  
 Mattingly I. G. 27, 79; [299, 378]  
 McGill W. J. 291; [379, 380]  
 McGlone R. E. 124; [200]  
 McIntosh J. 227; [278]
- McRobert H. 174; [381]  
 Medress M. F. 93; [341]  
 Menon K. M. N. 323; [382]  
 Mercier G. 34; [283]  
 Merzenich M. M. 227; [383]  
 Michon J. A. 106, 107; [384]  
 Mikiel W. 306, 307; [100, 371]  
 Miller G. A. 106, 107; [385]  
 Moll H. 97; [386]  
 Möller A. R. 174, 175, 234, 236—238; [387—394]  
 Molnar C. E. 197; [322]  
 Moore B. C. J. 291, 303, 308, 318; [395—397]  
 Morton J. 96, 97, 121, 122; [307, 398]  
 Moushegian G. 232; [455]  
 Mowbray G. H. 106, 107; [399]  
 Moxon E. G. 171, 202—204; [319]  
 Munson J. 34; [407]
- Näbëlek A. K. 314, 315, 317—319; [402, 403]  
 Näbëlek I. V. 280, 314, 315, 317—319, 327—334; [401—404]  
 Nakamura K. 64, 96, 103, 105; [267]  
 Neazelnitsky V. 181, 205, 206; [405]  
 Nelson P. G. 227, 232—235, 240; [249, 250, 406]  
 Newell A. 34; [407]  
 Niese H. 95; [442]  
 Nimura T. 318, 319, 321; [509]  
 Noorden L. P., van. 329, 336; [408]  
 Nooteboom S. G. 63—65; [409, 410]
- Obusek C. J. 139, 168; [520, 521]  
 O'Connor J. D. 41, 322; [411, 412]  
 Oettinger R. 194; [413]  
 Öhman S. 79, 82, 133; [414, 489]  
 Orr D. B. 146; [415]
- Paul A. P. 311; [486]  
 Pauli S. 260; [365]  
 Peake W. T. 174, 175, 182, 183; [287, 523, 524]  
 Peterson G. H. 193, 194; [325]  
 Peterson G. E. 61; [416]  
 Peterson J. L. 174; [417]  
 Peterson L. C. 192, 194; [418]  
 Pfeiffer R. R. 197, 232, 234; [320, 322, 331, 420]  
 Pickett J. M. 109, 110; [421]  
 Pierce J. R. 33, 197, 194; [422, 537]  
 Piroda E. 212, 217; [549]  
 Pisoni D. B. 145; [423]  
 Plomp R. 248, 249, 259, 302; [328, 424, 425, 431, 432]  
 Pollack I. 106, 107, 109, 298, 327—331; [426—430]  
 Pols L. C. W. 248, 249; [329, 431, 432]

- Potter R. K. 24; [434]  
 Preece G. R. 183; [435]
- Quancard J. P. 33; [436]
- Raab D. H. 280, 291; [397, 437]  
 Rabiner L. R. 26; [138]  
 Rahm W. E. 183; [526]  
 Rand T. C. 325, 326; [438]  
 Rao P. V. S. 323; [382]  
 Raphael L. J. 65—68, 143, 144;  
 [236, 439, 440]  
 Rapp K. 104, 124; [205, 244]  
 Reddy D. R. 33; [441]  
 Reddy R. 34; [407]  
 Reichardt W. 95; [442]  
 Rilling M. E. 136; [522]  
 Rhode W. S. 174, 184, 186, 187, 189,  
 206; [413—447]  
 Robles L. 184, 186, 187; [446, 447]  
 Ronken D. A. 280, 303, 304; [448,  
 449]  
 Rose J. E. 198, 209, 210, 227; [201,  
 450—454]  
 Rosenblith W. A. 177; [8]  
 Ross D. A. 173; [528]  
 Roth G. L. 227; [383]  
 Rozsypal A. J. 280; [404]  
 Rubin H. 212, 217; [549]  
 Rupert A. L. 232; [455]
- Sachs M. B. 252; [459, 457]  
 Sato T. 83; [505]  
 Schafer R. W. 26; [138]  
 Scharf B. 252, 254; [544]  
 Scholes R. J. 83; [458]  
 Schönfeld M. 191, 193; [459]  
 Schouten J. F. 303; [460]  
 Schroeder M. R. 31, 194, 213, 215;  
 [182, 461, 462]  
 Schulman A. I. 136; [241]  
 Sebestyen S. 29; [124]  
 Sekey A. 308; [463]  
 Sergeant R. L. 320; [464]  
 Serpa-Leitao A., de 85, 152, 154, 160;  
 [169, 213]  
 Sharf D. J. 322, 325; [465]  
 Shattuck S. R. 326; [326]  
 Shearme J. N. 27; [299]  
 Shepard R. W. 73; [466]  
 Shickman G. M. 105; [478]  
 Shitara T. 83; [505]  
 Sigurd B. 107, 146; [467]  
 Singh S. 74; [468]  
 Skinner T. E. 93; [341]  
 Small A. M. 200, 262 [469, 470]  
 Smith H. L. 93, 118; [507]  
 Smith R. L. 211, 213, 217—219;  
 [471]  
 Smoorenburg G. F. 197, 207; [472]
- Sohmer H. S. 183; [524]  
 Sokolich W. G. 199, 277; [550, 551]  
 Sone T. 318, 319, 321; [473, 509]  
 Sorin Ch. 285—288; [419]  
 Sorkin R. D. 136; [474]  
 Spoendlin H. 177, 178, 198; [475, 476]  
 Steffen-Batog M. 96, 97, 122; [307]  
 Stein H. J. 217; [477]  
 Stevens J. C. 105; [478]  
 Stevens K. N. 17, 21, 25, 27, 31, 42,  
 79, 82, 85, 87, 90, 163, 305, 311,  
 314, 316, 322, 326, 327, 335;  
 [190, 198, 327, 479—489, 492]  
 Stewart J. L. 171, 194; [130, 490]  
 Strother W. E. 183; [526]  
 Studdert-Kennedy M. 79, 82, 85,  
 311—313; [361, 489, 491, 492]  
 Sudo H. 133; [268]  
 Suen Ch. Y. 146; [493]  
 Suga N. 239, 240; [494, 495]  
 Sutton S. 280; [183]  
 Suzuki H. 333, 334; [496, 497]  
 Svensson S.-G. 25, 93; [362, 498, 499]  
 Syrdal A. 79; [378]
- Taylor K. J. 174, 175; [309]  
 Taylor W. 106, 107; [385]  
 Tempest W. 174; [381]  
 Terhardt E. 282—284, 288, 289, 294,  
 295; [500—502]  
 Thomas E. C. 211; [321]  
 Thomas I. B. 142, 143, 153; [503, 504]  
 Thomas R. B. 323; [382]  
 Tjerlund P. 85, 152, 154; [213]  
 Tomisawa M. 83; [505]  
 Tonndorf J. 175, 176; [506]  
 Tooley O. M. 41; [412]  
 Trager G. L. 93, 118; [507]  
 Tromp H. R. C. 249; [432]  
 Tsuchitani C. 227; [508]  
 Tsumura T. 318, 319, 321; [473, 509]  
 Turnbull W. W. 319; [510]  
 Turvey M. T. 145; [237]
- Uhlenbeck E. M. 97; [386]  
 Umeda N. 26, 90, 91, 112; [138, 218]
- Vaissiere J. 114, 119; [511, 512]  
 Vernon J. A. 183; [526]  
 Viemeister N. F. 291; [513]  
 Vogel A. 283, 284, 286, 287, 296; [514]  
 Vogten L. L. M. 200, 203; [515, 516]
- Wansdrond C. 194, 197; [517]  
 Warr W. B. 232, 234; [320, 331]  
 Warren R. M. 139, 142, 168—170;  
 [518—521]  
 Warren R. P. 139, 168; [521]  
 Watanabe T. 211; [321]  
 Watson C. S. 136; [522]



- Webster J. C. 146; [329]  
 Weiss T. F. 182, 183; [523, 524]  
 Wells R. S. 118; [525]  
 Wever G. G. 183; [526]  
 Whitfield I. C. 240, 321, 336; [251, 318]  
 Wickelgren W. A. 82; [527]  
 Wiener F. M. 173; [528]  
 Willems Y. D. 93; [529]  
 Williams J. C. 146; [415]  
 Wilson J. P. 185—187, 201—203,  
     205, 207; [252, 530—532]  
 Wood S. 104, 107; [533]  
 Woods W. 34; [407]  
 Woolsey C. N. 227; [534]  
 Yates G. K. 185, 187, 197; [310]  
 Zlatin M. A. 163; [535]  
 Zweig G. 194, 197; [537]  
 Zwicker E. 96, 171, 197, 199, 200, 202,  
     203, 207, 251, 252, 254—256, 282,  
     283, 290, 291, 316; [145, 538—544]  
 Zwislocki J. J. (Zwislocki-Moscicki J.)  
     173, 177, 183, 191, 192, 194, 199,  
     212, 217, 277; [235, 545—551]

## О Г Л А В Л Е Н И Е

|                                                                                                                | Стр.   |
|----------------------------------------------------------------------------------------------------------------|--------|
| Предисловие . . . . .                                                                                          | 3      |
| <br><b>Р А З Д Е Л I. РЕЧЕВОЙ СИГНАЛ И ПРОБЛЕМЫ ЕГО ОПИСАНИЯ . .</b>                                           | <br>9  |
| <b>Г л а в а 1. Свойства речевого сигнала и некоторые вопросы автоматического распознавания речи . . . . .</b> | <br>9  |
| 1.1. Элементы теории речеобразования . . . . .                                                                 | 12     |
| 1.1.1. Классификация звуков речи . . . . .                                                                     | 14     |
| 1.1.2. Источники акустической энергии . . . . .                                                                | 16     |
| 1.1.3. Частотная фильтрация в речевом тракте . . . . .                                                         | 20     |
| 1.2. Анализ речевых сигналов . . . . .                                                                         | 22     |
| 1.3. Синтез речевых сигналов . . . . .                                                                         | 25     |
| 1.4. Автоматическое распознавание речи . . . . .                                                               | 28     |
| 1.4.1. Автоматическое распознавание речи и проблема распознавания образов . . . . .                            | 28     |
| 1.4.2. «Моторная теория» и «анализ с помощью синтеза» . . . . .                                                | 31     |
| 1.4.3. Иерархические системы автоматического распознавания речи . . . . .                                      | 33     |
| <br><b>Г л а в а 2. Об описании речевого сигнала после его фонетической интерпретации . . . . .</b>            | <br>35 |
| 2.1. Описание смысла и предшествующие ему описания сообщения . . . . .                                         | 36     |
| 2.2. Потери информации в устной речи . . . . .                                                                 | 39     |
| 2.3. Объем оперативной памяти на фонетические элементы . . . . .                                               | 41     |
| 2.4. Задача перехода от последовательности фонетических элементов к словам языка . . . . .                     | 44     |
| 2.5. Членение на «слова» бессмысленных речевых последовательностей и псевдопредложений . . . . .               | 46     |
| 2.6. Фонетический образ слова . . . . .                                                                        | 48     |
| 2.7. Требования к описанию фонетической последовательности . . . . .                                           | 49     |
| <br><b>Р А З Д Е Л II. ФОНЕТИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕЧЕВОГО СИГНАЛА . . . . .</b>                               | <br>50 |
| <b>Г л а в а 3. Природа фонетического образа . . . . .</b>                                                     | <br>50 |
| 3.1. Понятие фонетической последовательности . . . . .                                                         | 50     |
| 3.2. Ритмический рисунок фонетической последовательности . . . . .                                             | 52     |

|                                                                                                                                                        | Стр.   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| 3.3. Фонемный состав фонетической последовательности . . .                                                                                             | 54     |
| 3.3.1. Фонемная классификация текущих значений речевого сигнала . . . . .                                                                              | 54     |
| 3.3.2. Последовательное или параллельное распознавание фонем . . . . .                                                                                 | 55     |
| 3.4. Экспериментальные доказательства выделения отрезков гласных звуков в речевом сигнале . . . . .                                                    | 57     |
| 3.4.1. Выделение начала и конца гласного в речевом сигнале (опыты по синхронизации) . . . . .                                                          | 57     |
| 3.4.2. Гласный звук и формантные переходы . . . . .                                                                                                    | 61     |
| 3.4.3. Длительность гласного звука как фонематический признак . . . . .                                                                                | 63     |
| 3.4.4. Длительность гласного как признак для идентификации глухости-звонкости последующего согласного . . . . .                                        | 65     |
| 3.5. Разные интерпретации понятия «фонетический элемент» . . . . .                                                                                     | 68     |
| <br>Г л а в а 4. Лингвистические и психологические фонемы . . . . .                                                                                    | <br>71 |
| 4.1. Лингвистические представления о фонемах . . . . .                                                                                                 | 71     |
| 4.2. Психологические признаки фонем . . . . .                                                                                                          | 73     |
| 4.3. Фонема как запоминаемое сокращенное описание сигнала . . . . .                                                                                    | 74     |
| 4.3.1. Согласные . . . . .                                                                                                                             | 74     |
| 4.3.2. Гласные . . . . .                                                                                                                               | 79     |
| 4.4. Набор психологических фонем . . . . .                                                                                                             | 81     |
| 4.5. Фонемная классификация . . . . .                                                                                                                  | 82     |
| 4.5.1. Фонемные границы стационарных гласных в пространстве частот формант . . . . .                                                                   | 83     |
| 4.5.2. Область неопределенности и контекстуальные влияния при восприятии гласных . . . . .                                                             | 84     |
| 4.5.3. Детекторы признаков согласных . . . . .                                                                                                         | 87     |
| <br>Г л а в а 5. Восприятие просодической информации . . . . .                                                                                         | <br>91 |
| 5.1. Интенсивность и длительность как самостоятельные источники просодической информации. Данные исследований восприятия словесного ударения . . . . . | 94     |
| 5.2. Длительности разных сегментов слога — разные просодические признаки . . . . .                                                                     | 100    |
| 5.2.1. Влияние длительности гласного и согласного в слоге на восприятие словесного ударения . . . . .                                                  | 101    |
| 5.2.2. Речевой темп . . . . .                                                                                                                          | 104    |
| 5.2.3. Длительность согласного как признак словесной границы . . . . .                                                                                 | 109    |
| 5.3. Частота основного тона и просодические признаки речи . . . . .                                                                                    | 113    |
| 5.3.1. Мелодический контур фразы. Мелодические изменения во фразе и синтаксические отношения в языке . . . . .                                         | 114    |
| 5.3.2. Частота основного тона и признаки ритмической структуры высказывания. Восприятие словесного ударения . . . . .                                  | 119    |
| 5.3.3. Использование мелодических изменений для выделения слов во фразе. Логическое ударение . . . . .                                                 | 124    |
| 5.4. Об измерении просодических признаков в непрерывном речевом потоке и правилах принятия решений о них . . . . .                                     | 131    |

|                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------|-----|
| <b>Глава 6. Временная организация процесса фонетической интерпретации</b>                           | 138 |
| 6.1. Слог как элемент потока                                                                        | 141 |
| 6.1.1. Распознавание последовательностей гласных. Обратная маскировка гласных                       | 142 |
| 6.1.2. Минимальная длительность слога, необходимая для его обнаружения                              | 146 |
| 6.1.3. Проблема конца гласного                                                                      | 152 |
| 6.2. Описание сигнала на входе фонетического интерпретатора                                         | 157 |
| 6.2.1. Обнаружение [r] русскими слушателями                                                         | 159 |
| 6.2.2. Восприятие начальных звонких и глухих смычных согласных русскими слушателями                 | 161 |
| 6.2.3. Различение начальных звонких и глухих смычных согласных носителями английского языка         | 163 |
| 6.2.4. Распознавание порядка следования фонем в сочетаниях согласных                                | 164 |
| 6.2.5. Временной порог обнаружения последовательности                                               | 165 |
| 6.2.6. Распознавание порядка следования звуковых посылок в быстрых перечевых последовательностях    | 167 |
| 6.3. Реальность акустических событий                                                                | 170 |
| <b>РАЗДЕЛ III. СЛУХОВАЯ ОБРАБОТКА РЕЧЕВОГО СИГНАЛА</b>                                              | 171 |
| <b>Глава 7. Спектральный слуховой анализ и его моделирование</b>                                    | 171 |
| 7.1. Система наружного уха                                                                          | 172 |
| 7.2. Система среднего уха                                                                           | 174 |
| 7.3. Система улитки внутреннего уха                                                                 | 176 |
| 7.3.1. Амплитудно-частотная характеристика базальной части системы улитки                           | 181 |
| 7.3.2. Характеристики гидромеханической части системы улитки                                        | 184 |
| 7.3.3. Моделирование улитки как спектрального анализатора                                           | 191 |
| 7.4. Характеристики «фильтра улитки» при измерении прямыми методами                                 | 197 |
| <b>Глава 8. Представление сигнала в «частотном канале» слухового нерва</b>                          | 208 |
| 8.1. Воспроизведение формы колебаний                                                                | 209 |
| 8.2. Эффект периферической адаптации                                                                | 209 |
| 8.3. Модели преобразования сигнала в периферической слуховой системе                                | 213 |
| 8.3.1. Модели с переменным коэффициентом передачи                                                   | 214 |
| 8.3.2. Модели с переменным порогом                                                                  | 219 |
| <b>Глава 9. Принципы обработки речевого сигнала центральными отделами слуховой системы</b>          | 222 |
| 9.1. Качественная классификация возможных признаков сигнала                                         | 222 |
| 9.2. Тонотопическая организация слуховой системы. Данные о локальном по спектру характере признаков | 226 |
| 9.3. Природа признаков: текущие значения амплитуды (энергии) в полосе или особенности и события     | 232 |

|                                                                                                                           | Стр.       |
|---------------------------------------------------------------------------------------------------------------------------|------------|
| 9.4. Центральные слуховые нейроны и «детекторы образов»                                                                   | 241        |
| 9.5. Значение данных электрофизиологии слуха для теории восприятия речи . . . . .                                         | 245        |
| <b>Глава 10. Обработка «слухового спектра» стимула . . . . .</b>                                                          | <b>247</b> |
| 10.1. Гипотезы о природе признаков гласных . . . . .                                                                      | 247        |
| 10.1.1. Кодирование величины возбуждения . . . . .                                                                        | 248        |
| 10.1.2. Пространственное кодирование . . . . .                                                                            | 249        |
| 10.2. Определение слухового спектра стимула . . . . .                                                                     | 250        |
| 10.2.1. Экспериментальные методы измерения эквивалентного спектра . . . . .                                               | 251        |
| 10.2.2. Процедура вычисления слухового спектра . . . . .                                                                  | 254        |
| 10.2.3. Примеры слуховых спектров гласных . . . . .                                                                       | 257        |
| 10.3. Природа полезных признаков спектра . . . . .                                                                        | 259        |
| 10.3.1. Локальный характер признаков . . . . .                                                                            | 259        |
| 10.3.2. Неоднородность на спектре как фонетический признак . . . . .                                                      | 262        |
| 10.4. Выделение неоднородностей на слуховом спектре моделью латерального торможения . . . . .                             | 263        |
| 10.5. Число полезных признаков гласных . . . . .                                                                          | 267        |
| 10.6. Проблемы фонемной интерпретации слухового описания гласных . . . . .                                                | 274        |
| <b>Глава 11. Слуховая обработка огибающей стимула . . . . .</b>                                                           | <b>276</b> |
| 11.1. Общие представления о слуховой обработке огибающей                                                                  | 276        |
| 11.1.1. Свидетельства в пользу дифференцирования . . . . .                                                                | 278        |
| 11.1.2. Свидетельства против одного фильтра . . . . .                                                                     | 279        |
| 11.2. Амплитудно-частотные характеристики системы слуховой обработки огибающей . . . . .                                  | 281        |
| 11.2.1. Амплитудно-частотные характеристики, определенные по порогам обнаружения модуляции для разных критериев . . . . . | 282        |
| 11.2.2. Амплитудно-частотные характеристики, полученные методом уравнивания «мощности» критерия . . . . .                 | 283        |
| 11.2.3. Выходные пороговые устройства системы обработки огибающей . . . . .                                               | 284        |
| 11.3. Связь анализатора огибающей с частотным каналом периферической слуховой системы . . . . .                           | 288        |
| 11.4. Входная нелинейность системы обработки огибающей                                                                    | 290        |
| 11.5. Объединение информации об огибающей в частотных каналах . . . . .                                                   | 295        |
| 11.6. Функциональная модель оценки хриплости . . . . .                                                                    | 296        |
| 11.7. Различение формы неравномерностей . . . . .                                                                         | 297        |
| <b>Глава 12. Слуховое описание сигналов с изменяющимся во времени спектром . . . . .</b>                                  | <b>303</b> |
| 12.1. Предположения об описании контура изменения частоты                                                                 | 305        |
| 12.2. Выделение «особых точек» в контуре $f(t)$ . . . . .                                                                 | 308        |
| 12.2.1. Восприятие реченподобных звуков с односторонним контуром . . . . .                                                | 308        |
| 12.2.2. Измерение частоты в экстремальных точках контура $f(t)$ . . . . .                                                 | 311        |
| 12.2.3. Результаты экспериментов по восприятию односторонних контуров в тональных сигналах . . . . .                      | 314        |
| 12.3. Слуховой анализ «производной» частотного контура . . . . .                                                          | 320        |



|                                                                                   | Стр.       |
|-----------------------------------------------------------------------------------|------------|
| 12.3.1. Восприятие направления изменения спектра в речевых сигналах . . . . .     | 322        |
| 12.3.2. Восприятие скорости изменения частоты тона . . . . .                      | 327        |
| 12.3.3. Роль скорости изменения спектра при восприятии речевых сигналов . . . . . | 331        |
| 12.3.4. Гипотеза о детекторах временных изменений частоты в слухе . . . . .       | 335        |
| <b>Г л а в а 13. Перспективы разработки теории восприятия речи . . .</b>          | <b>339</b> |
| Литература . . . . .                                                              | 346        |
| Предметный указатель . . . . .                                                    | 372        |
| Именной указатель . . . . .                                                       | 375        |

**ФИЗИОЛОГИЯ РЕЧИ.  
ВОСПРИЯТИЕ РЕЧИ ЧЕЛОВЕКОМ**

*В серии*

**«РУКОВОДСТВО ПО ФИЗИОЛОГИИ»**

*Утверждено к печати*

*Научным советом по комплексным проблемам физиологии  
человека и животных Академии наук СССР*

Редактор издательства *С. И. Налбандян*

Художник *М. И. Разулевич*

Технический редактор *О. А. Мокеева*

Корректоры *С. В. Добрянская, Л. А. Привалова*  
и *Г. И. Суворова*

Сдано в набор 11/III 1976 г. Подписано к печати 15/VI 1976 г.  
Формат 60×90<sup>1</sup>/<sub>16</sub>. Бумага № 1. Печ. л. 24<sup>1</sup>/<sub>4</sub> = 24,25 усл. печ. л.  
Уч.-изд. л. 26,38. Изд. № 6180. Тип. зак № 1018. М-14940. Тираж 7000.  
Цена 2 руб. 08 коп.

Ленинградское отделение издательства «Наука»  
199164, Ленинград, В-164, Менделеевская линия, д. 1

---

1-я тип. издательства «Наука»  
199034, Ленинград, В-34, 9 линия, д. 12

## ИЗДАТЕЛЬСТВО «НАУКА»

*В серии*

### «РУКОВОДСТВО ПО ФИЗИОЛОГИИ»

ВЫШЛИ СЛЕДУЮЩИЕ КНИГИ:

- Физиология системы крови (1968)
- Общая и частная физиология нервной системы (1969)
- Физиология труда, мышечной деятельности и спорта (1969)
- Физиология высшей нервной деятельности. Ч. I (1970)
- Физиология высшей нервной деятельности. Ч. II (1971)
- Физиология сенсорных систем. Ч. I. Физиология зрения (1971)
- Физиология сенсорных систем. Ч. II (1972)
- Физиология сенсорных систем. Ч. III. О. Б. Ильяевский. Физиология механорецепторов (1975)
- Физиология почки (1972)
- Клиническая нейрофизиология (1972)
- Физиология дыхания (1973)
- Физиология лактации. Авторы: П. П. Грачев и В. А. Галанцев (1973)
- Физиология пищеварения (1974)
- Общая физиология возбудимых мембран. Автор: Б. А. Ходоров (1975)
- Возрастная физиология (1975)
- Физиология движений (1976)

В БЛИЖАЙШЕЕ ВРЕМЯ НАМЕЧЕНО ВЫПУСТИТЬ:

- Физиология всасывания
- Физиология кровообращения. Физиология сердца
- Физиология гистогематических барьеров
- Экологическая физиология человека. Адаптация человека к экстремальным условиям среды
- Физиология сельскохозяйственных животных

Готовящиеся к изданию книги можно предварительно заказать в магазинах «Академкниги»

Для получения книг почтой заказы просим направлять по адресу:  
117464, Москва, В-464, Мичуринский пр., 12, магазин «Книга — почтой»  
Центральной конторы «Академкнига»  
197110, Ленинград, П-110, Петрозаводская ул., 7, магазин «Книга — почтой»  
Северо-Западной конторы «Академкнига» или в ближайший магазин  
«Академкнига»