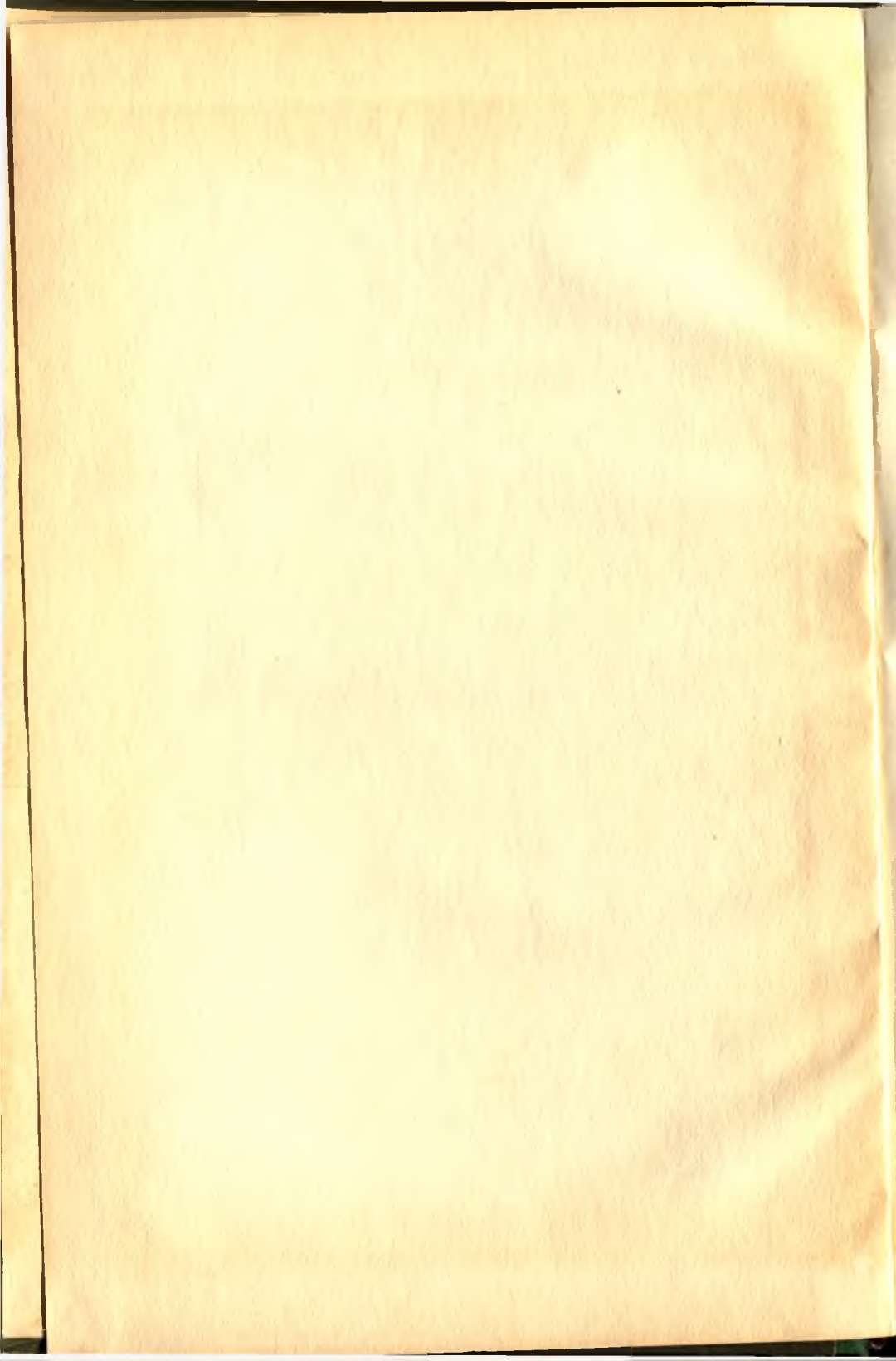




АКАДЕМИЯ НАУК СССР
ИНСТИТУТ ФИЗИОЛОГИИ им. И. П. ПАВЛОВА

РЕЧЬ
АРТИКУЛЯЦИЯ
И ВОСПРИЯТИЕ



ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА • 1965 • ЛЕНИНГРАД

Авторы:

Л. А. ЧИСТОВИЧ, В. А. КОЖЕВНИКОВ, В. В. АЛЯКРИПСКИЙ,
Л. В. БОНДАРКО, А. Г. ГОЛУЗИНА, Ю. А. КЛААС,
Ю. И. КУЗЬМИН, Д. М. ЛИСЕНКО, В. В. ЛЮБЛИНСКАЯ,
Н. А. ФЕДОРОВА, В. С. ШУПЛЯКОВ, Р. М. ШУПЛЯКОВА

Под общей редакцией

В. А. КОЖЕВНИКОВА и Л. А. ЧИСТОВИЧ

АННОТАЦИЯ

В книге излагаются результаты оригинального экспериментального исследования организации речевого процесса, выполненного коллективом физиологов, лингвистов, математиков и инженеров. Описываются новые методы исследования деятельности речеобразующего аппарата, основывающиеся на применении современной электронной техники. Выдвигаются новые теоретические представления относительно организации процессов восприятия речи. Результаты исследования рассматриваются в плане проблем речевого общения человека и электронных машин.

Издание рассчитано на физиологов нервной системы, исследователей в области биологической кибернетики, а также широкий круг специалистов, связанных с речью: акустиков, инженеров, лингвистов, психологов, педагогов, логопедов, невропатологов.

Г Л А В А 1

ФИЗИОЛОГИЯ РЕЧЕОБРАЗОВАНИЯ И ВОПРОСЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ И СИНТЕЗА РЕЧИ

Учитывая чрезвычайно быстрое развитие вычислительной техники и прогрессивное расширение области применения универсальных и специализированных электронных вычислительных и управляющих машин, легко предвидеть, что уже в ближайшие годы возникнет практическая необходимость в речевом общении между человеком и машиной.

До сих пор общение с машиной производится на языке, наиболее удобном для машины (обычно с помощью того или иного цифрового кода), но отнюдь не наиболее простом и удобном для человека. Естественно, что человек при этом работает малоэффективно, он тратит на передачу (или прием) сообщения значительно больше времени, чем ему понадобилось бы, если бы он просто сказал (или услышал) это сообщение. Результаты специальных работ, посвященных исследованию возможностей человека по приему информации (Pollack a. Ficks, 1954; Sumbly, Chambliss a. Pollack, 1958), не оставляют сомнения в том, что оптимальным решением проблемы связи человека с машиной являлась бы организация речевого общения между ними (Ликлайдер, 1960; Иванов и Шаумян, 1961; Сапожков, 1963).

Трудно переоценить увлекательность проблемы речевого общения человека с машиной для физиологов, лингвистов и психологов. Очевидно, что разработка устройств и алгоритмов для машинного распознавания и синтеза речи — это, по меньшей мере, создание одной из моделей распознавания и синтеза речи человеком. Для физиолога это означало бы попытку проникновения в ту область работы мозга, которая была определена И. П. Павловым как деятельность второй сигнальной системы человека (Павлов, 1951б).

Рассматривая речевое сообщение как условную реакцию и одновременно как условный сигнал, связанный с непосредствен-

ными раздражениями и действиями, И. П. Павлов фактически поставил вопрос о необходимости изучения механизмов перехода от непосредственных образов действительности к речевому сообщению и механизмов перехода от речевых образов к обозначаемым ими действиям и представлениям. На принятом в настоящее время языке эти задачи формулируются как исследование алгоритмов синтеза и распознавания речевого сообщения человеком.

Задача построения некоторого сложного действия на основании указаний, содержащихся в воспринимаемом речевом сообщении, принципиально близка к задаче перевода сообщения с одного языка на другой. Хорошо известно, что единицей речи, на уровне которой возможен переход от одного языка к другому, и от речевого сообщения к действию, является предложение (например, по предложениям: *зажги спичку, зажги лампу* необходимое действие может быть построено только на основании всего предложения в целом). Употребляя далее термин «сообщение», мы будем иметь в виду предложение или слово, выполняющее функцию предложения.

Если набор возможных сообщений и, соответственно, ответных действий резко ограничен, каждое сообщение может рассматриваться как отдельный условный сигнал, а каждое действие — как отдельная условная реакция. Тогда задача понимания сообщения может быть сведена к задаче опознавания образа этого сообщения, а задача построения действия может быть сведена к задаче выбора одного из действий, программа которого полностью заранее известна. Это означает, что для каждой пары, состоящей из сообщения и действия, может быть найдено соответствие — некоторый символ; выбор символа на основании услышанного сообщения будет при этом означать как опознавание сообщения, так и выбор действия.

Можно видеть, что при таком рассмотрении задача распознавания речевого сообщения ничем принципиально не отличается от задачи узнавания любого другого сигнала и целиком попадает в круг задач, исследуемых статистической теорией опознавания образов (Харкевич, 1959; Бэнгард, 1961; Браверман, 1962; Ковалевский, 1932; Sebestyen, 1931, 1932; Highleman, 1962).

Для того чтобы приобрести способность распознавать ограниченный набор речевых сообщений, распознающее устройство (человек, животное, машина) в принципе отнюдь не обязательно должно иметь какие-либо сведения о структуре языка, оно не обязано также членить сообщение на элементы (слова, фонемы). Ему необходимо и достаточно выработать образ сообщения, т. е. выделить набор полезных объективных признаков, по которым разные сообщения могут быть отличены друг от друга, и установить критерии, по которым должны приниматься решения. Если набор возможных сообщений очень мал, то полезные признаки могут быть довольно случайно выбранными, например длитель-

ность, громкость, какие-то особенности интонации, наличие шума в определенном месте сообщения и т. д. Полезным оказывается любой признак, по которому реализации одного сообщения отличаются от реализаций других сообщений.

Описываемый путь опознавания образа целого сообщения вполне себя оправдывает, когда дело идет об автоматическом распознавании ограниченного набора (до нескольких десятков) команд, состоящих из одного-двух слов (Цемель, 1961; Petrick a. Willett, 1960; Denes a. Mathews, 1960; Сапожков, 1963).

Вполне вероятно, что аналогичным образом осуществляется «понимание» речи животными и маленькими детьми на первых этапах овладения речью.

Однако, как только мы переходим от ограниченного набора сообщений к общему случаю, когда число возможных речевых сообщений можно принять равным, например, числу осмысленных предложений на данном языке, становится очевидным, что рассмотренный выше путь опознавания образа каждого отдельного сообщения явно не рационален. Для того чтобы хранить в памяти образы всех возможных предложений, распознающему устройству понадобился бы совершенно невероятный объем памяти. Кроме того, для того чтобы выработать образ предложения, его необходимо хотя бы один раз воспринять. Миллер, Галантер и Прибрам (Miller, Galanter, Pribram, 1960) приводят расчеты, показывающие, что, для того чтобы хотя бы один раз услышать все грамматически возможные английские фразы длиной до 20 слов, человеку пришлось бы слушать примерно по 3×10^{20} фразы в секунду, причем это продолжалось бы в течение 100 лет без перерывов.

Таким образом, предположение о том, что в мозгу человека хранится словарь всех возможных предложений, является совершенно неправдоподобным и должно быть отвергнуто. Отсюда необходимо следует, что в мозгу может храниться словарь только для единиц, существенно меньших, чем предложение, — словоформ, или слов, или каких-то других единиц, например основ слов и окончаний. Тогда, однако, необходимо допустить, что в мозгу должны также сохраняться и правила, по которым, используя данный словарь, можно построить предложение.

Вопрос о структуре словаря и правилах (алгоритме) синтеза предложения является основным вопросом современной теоретической лингвистики (Андреев и Зиндер, 1963).

Теоретические исследования показывают, что процесс синтеза предложения на любом языке осуществляется по принципу от общего к частному — от плана предложения, его структуры, к конкретизации отдельных частей этого предложения (модель «синтеза по непосредственным составляющим» — Хомский, 1962; Yngve, 1960). При синтезе предложений по такому принципу необходимо, чтобы синтезирующее устройство располагало ин-

формацией о предложении в целом еще до того, как предложение начнет синтезироваться. Иначе говоря, перевод не может быть пословным. Чтобы перевести с одного языка на другой, необходимо сначала понять смысл предложения в целом, и потом только строить согласно этому смыслу предложение на другом языке. Очевидно, что это в равной мере относится и к случаям, когда дело касается не перевода с одного языка на другой, а первичного построения предложения, описывающего некоторую ситуацию.

Введение операции установления смысла как промежуточного этапа между анализом ситуации и первичным синтезом предложения, между анализом и синтезом предложений при переводе с разных языков естественно приводит к представлению о некотором промежуточном абстрактном языке понятий. В работах по машинному переводу такой язык получил название языка-посредника. «При переводе с одного языка на другой должно быть передано значение текста, хотя полностью меняется способ выражения этого значения. . . . Чтобы можно было осуществить перевод с одного языка на другие, необходимо закодировать значения в виде чисто формальных символов и указания чисто формальных операций над этими символами. . . . Центральная проблема теории машинного перевода — создание языка-посредника, который представляет собой не что иное, как универсальный код, служащий для перевода с одних языков на другие» (Иванов и Шаумян, 1961, стр. 227).

Введение понятия языка-посредника позволяет формулировать проблему изучения речеобразования как исследование различных процессов и явлений, происходящих при переходе от изображения предложения на языке-посреднике к реализации этого предложения в форме акустического сигнала. Проблема изучения восприятия речи может быть сформулирована как исследование процессов и явлений, происходящих при обратном переходе — от акустического изображения предложения к его изображению на языке-посреднике. Очевидно, что обе проблемы являются теснейшим образом связанными. Очевидно также, что как процесс речеобразования, так и процесс восприятия включает в себя ряд этапов, причем явления, происходящие на разных этапах, относятся к компетенции различных специалистов и должны исследоваться различными методами.

В процессе речеобразования логично выделить три основных этапа. Первый из них может рассматриваться как собственно синтез предложения; результатом его должно быть построение в мозгу человека программы артикуляторных движений, соответствующей данному предложению. Следующий этап представляет собой реализацию этой программы, т. е. превращение ее в некоторую последовательность комплексов артикуляторных движений. Наконец, третий этап, по времени, естественно, совпадающий с артикуляторными движениями, — это образование звукового

сигнала. В литературе термин «речеобразование» употребляется в основном для описания последнего из этих процессов.

Современное состояние исследований по проблеме речеобразования (понимая ее в широком смысле) таково, что, с одной стороны, существуют чрезвычайно интересные предположения о возможной модели процесса синтеза предложения человеком (Yngve, 1960); с другой — достаточно подробно разработана акустическая теория речеобразования (Fant, 1960), описывающая математические зависимости между характеристиками акустического сигнала и состоянием речеобразующего аппарата. Однако совершенно неразработанной является та часть проблемы, которая целиком относится к компетенции физиологов. Полностью неизвестно, как организована артикуляторная программа, какова структура двигательных артикуляторных комплексов, как осуществляется управление артикуляторными органами в процессе речи. В статьях и руководствах, посвященных, согласно названиям, физиологии или механизмам речи (Жинкин, 1958; Lullies, 1953; Arnold, 1957; Kaplan, 1960, и др.), в лучшем случае излагаются данные об анатомии речеобразующих органов, иннервации речевых мышц и о нервных центрах, которые предположительно участвуют в управлении речью.

Отсутствие серьезных работ в области физиологии речи следует отнести за счет двух основных причин. Во-первых, физиология вообще до сих пор почти не занималась вопросами, касающимися организации сложных двигательных актов. Чрезвычайно интересные идеи Н. А. Бернштейна о принципах построения сложных движений (Бернштейн, 1947) получили признание только в самое последнее время, когда наметился «кибернетический подход» к проблеме изучения движений (Гельфанд, Гурфинкель, Цетлин, 1962; Алексеев, 1963; Miller, Galanter, Pribram, 1960). Вторую причину следует искать в отсутствии методов, которые позволили бы достаточно точно и надежно регистрировать такие быстрые и тонко дифференцированные движения, какими являются движения артикуляторных органов при речи.

Необходимость исследования физиологии артикуляторных движений определяется не только тем, что без такого исследования нельзя понять процесса речеобразования у человека. Другим, не менее веским основанием для проведения работ в этом направлении, является их важность для решения некоторых кардинальных вопросов теории автоматического распознавания речи.

Важнейшим теоретическим вопросом в проблеме автоматического распознавания речи является вопрос о единицах (элементах) распознавания. Как уже говорилось, конечным результатом распознавания речевого сообщения машиной должно быть «понимание» смысла этого сообщения. Если число подлежащих распознаванию сообщений достаточно велико, то «смысл» не может быть определен иначе, как методами анализа предложения, раз-

рабатываемыми в настоящее время специалистами по машинному переводу. Естественно, что звуковые колебания, являющиеся непосредственным входным сигналом, поступающим в распознающее устройство, не могут быть прямо подвергнуты лингвистическому смысловому анализу. Для того чтобы лингвистический анализ стал возможным, необходимо осуществить переход от элементов речи к смыслоразличительным элементам языка. Смыслоразличительные элементы языка по своей природе дискретны, а звуковой поток может быть непрерывным на протяжении всего предложения. Таким образом, первая задача, которая должна быть решена при автоматическом распознавании речи, состоит в преобразовании слитного речевого потока в последовательность дискретных символов, соответствующих смыслоразличительным элементам языка. В технике именно эта задача известна как задача автоматического распознавания речи. (Следующие этапы распознавания, включающие идентификацию последовательности смыслоразличительных единиц с элементами словаря, — словоформами, словами и т. д. — и синтаксический анализ предложения, как правило, не рассматриваются).

Согласно лингвистическим представлениям, минимальной смыслоразличительной единицей является фонема, которая, в свою очередь, может быть описана набором, «пучком» дифференциальных признаков (Jacobson a. Halle, 1956). Соответственно первая задача автоматического распознавания речи формулируется как преобразование речевого (акустического) потока в последовательность фонем, где каждая фонема представляется или в виде одного символа (число символов равно числу фонем), или в виде ряда параллельных символов, соответствующих дифференциальным признакам.

Многочисленные попытки решить эту задачу до сих пор не привели к желаемым результатам (см. обзоры: Цемель, 1961; Magill, 1961; Сапожков, 1963). Рассмотрение результатов ряда фонетических работ последнего времени (см. обзор в главе 6 настоящей монографии) заставляет считать, что одна из причин неудач кроется в тех примитивных представлениях о структуре речевого процесса, из которых исходили инженеры, занимающиеся автоматическим распознаванием. Суть этих представлений состоит в том, что речевой процесс рассматривается как последовательность произнесений фонем и что, следовательно, речевой поток, соответствующий предложению, можно разбить на такие участки, где каждый участок будет определяться одной и только одной фонемой, а число участков будет равно числу фонем в предложении.

И с физиологической, и с фонетической точек зрения реальнее допустить, что фонемы как смыслоразличительные элементы существуют только на уровне синтеза артикуляторной программы, соответствуя некоторым элементарным указаниям относительно

структуры комплексов артикуляторных движений, входящих в эту программу. В этом случае весь комплекс в целом может определяться несколькими фонемами сразу, и нет никаких оснований для поисков участков речевого потока, соответствующих одной фонеме. Очевидно, что единственный действенный способ выяснения этого вопроса состоит в исследовании реальной структуры (организации) речевого процесса.

Исходя из всего вышесказанного, мы поставили основной своей задачей попытку в какой-то мере разобраться в структуре артикуляторного процесса. Для проведения исследований в этом направлении понадобилась разработка новых методов регистрации артикуляторных явлений. Описанием результатов этой методической работы, оказавшейся достаточно сложной, начинается изложение материалов проводившихся исследований. Кроме результатов, касающихся артикуляции, в монографию включена специальная глава, посвященная вопросам восприятия речи человеком.

Г Л А В А 2

МЕТОДЫ ИССЛЕДОВАНИЯ ДЕЯТЕЛЬНОСТИ РЕЧЕОБРАЗУЮЩЕГО АППАРАТА

Необходимой предпосылкой для начала исследований организации процессов речеобразования является владение такими методами, которые позволили бы регистрировать артикуляторную динамику речи в достаточно полном виде. Понимание деятельности речеобразующего аппарата в целом и изучение взаимоотношений функционирования его различных элементов требуют обеспечения комплексной регистрации основных показателей физиологической деятельности речеобразующего тракта и возможности точного сопоставления с этими показателями акустической картины речевого потока. Необходимость получения количественного описания артикуляторных явлений и реализации опытных проверок предполагаемых моделей процесса речеобразования обязательно ведет к применению статистического подхода к оценке экспериментальных данных. Соответственно этому методы должны обеспечивать возможность получения этих данных в достаточном объеме и с достаточной точностью.

Известные по литературе (Щерба, 1912; Богородицкий, 1930; Артемов, 1956; Зиндер, 1960; Rousselot, 1901—1908; Lullies, 1953; Kaiser, 1957, и др.) методы регистрации артикуляторных показателей явно не удовлетворяют требованиям комплексной регистрации последних в динамике нормального потока речи.

В экспериментальной фонетике значительного совершенства достигли методы, позволяющие достаточно подробно изучать «статическую» сторону деятельности речеобразующих механизмов, т. е. особенности положения определенных элементов речевого тракта при произнесении отдельных звуков речи. Одно из ведущих мест здесь занимает рентгенография (Скалозуб, 1963; Fant, 1960). Ее методы дают возможность довольно точно изучать особенности положения отдельных элементов речевой трубки во

время фонации. Основные полученные данные касаются главным образом гласных звуков. В последние годы использовалась также и рентгенокинематография (Truby, 1959; Moll, 1960), с помощью которой возможно наблюдать в определенной мере и динамику речеобразующего тракта. Однако в связи с опасностью сколь угодно длительного облучения испытуемого, со сложностью аппаратуры и трудоемкостью обработки графического материала эти способы не могут быть признаны особенно перспективными.

Определенные данные об артикуляции могут быть получены также и с помощью обычных фотографических и кинематографических методов. Так, для изучения формы, которую принимает язык в некоторых случаях фонации, удалось произвести фотографирование непосредственно в ротовой полости (Бровченко, 1954). Для этого применялся сверхминиатюрный многообъективный фотоаппарат, разработанный для целей медицинской гастрографии. Удачным представляется метод (Fujimura, 1961b), позволявший получать данные, эквивалентные скоростному кинематографированию, с помощью стробоскопического освещения электронной лампой-вспышкой (240 кадров в сек.). Подробно изучались особенности речевых движений губ и их соотношения с акустической картиной произносимых односложных слов, синхронно регистрируемой на магнитной ленте. Но возможности и данных методов оказываются также довольно ограниченными.

Существенное значение для экспериментальной фонетики имеет так называемая палатография — получение отпечатков областей касания языком твердого неба при произнесении того или иного звука (Богородицкий, 1930; Ladefoged, 1957). Принцип этого по своему существу статического метода лег в основу разработанной нами и описываемой ниже «динамической палатографии».

Если особенности взаимного расположения отдельных анатомических образований речевого тракта при произнесении определенных звуков и могут определяться с достаточной точностью, то значительно хуже обстоит дело с изучением динамики непрерывно происходящих взаимосвязанных движений всех элементов речеобразующего механизма, в результате которых и возникает слитный поток членораздельной речи. В основе большинства «динамических» методов, до сегодняшнего дня применяющихся многими исследователями, лежит пневматическая техника регистрации, использующая барабанчики с резиновыми мембранами, воздушные трубки и кимограф с закопченной бумагой (Артемов, 1956; Dieth, 1950, Stetson, 1951).

Для записи движений губ и нижней челюсти используются системы рычагов, связанные с воспринимающими пневматическими капсулами (так называемыми капсулами Мареля). Последние через трубки соединяются с аналогично устроенными записывающими капсулами. Регистрация, как правило, ведется на закопченной

ленте, равномерно передвигаемой механизмом кимографа. На этой же ленте также с помощью пневматических систем могут регистрироваться потоки воздуха, выходящие при фонации из ротовой щели и носовых отверстий. С помощью устройств, напоминающих примитивные фонографы, здесь же может регистрироваться и акустическая картина речи. Запись ее при этом получается весьма неполноценной и требует кропотливой расшифровки.

Несомненные достоинства электрических методов измерения и регистрации различных явлений в последнее время начинают привлекать все большее внимание и специалистов, работающих в области экспериментальной фонетики. Однако, судя по литературе, различные исследователи использовали в основном лишь отдельные приборы для регистрации тех или иных явлений. Так, например, с помощью весьма чувствительного устройства — миниатюрных датчиков, устроенных по типу угольного микрофона, — удавалось регистрировать микродвижения языка при произнесении слов про себя — скрытую артикуляцию (Фонарев, 1956). Ряд исследователей сочетал методы электрической регистрации с применением «классических» воздушных и водяных спирометров и манометров (Draper, Ladefoged a. Whitteridge, 1959; Hoshiko, 1960). Однако и в данном случае далеко не всегда удавалось обеспечить достаточно безынерционную регистрацию, удовлетворительно отражающую динамику артикуляции, редко применялась одновременная регистрация более 2—4 процессов.

Интересным способом является регистрация биопотенциалов мышц — электромиография. В последние годы этот метод использовался многими исследователями для изучения речи (Кратин, 1955; Чаргейшвили, Тохадзе и Польшин, 1959; Draper, Ladefoged a. Whitteridge, 1959; Hoshiko, 1960; Lysaught, Rosov a. Harris, 1961). Нужно сказать, что и здесь имеется много трудностей, прежде всего связанных со сложностью, а часто и невозможностью раздельного отведения биопотенциалов от различных функционально, но тесно сплетающихся мышечных групп. Далеко не ясными являются также соотношения между электрической картиной биопотенциалов и механическими эффектами сокращения мышц.

Подводя итог краткому рассмотрению литературы, приходится сказать, что в настоящий момент по существу нет методов, позволяющих достаточно полно регистрировать динамику артикуляторной деятельности речеобразующего аппарата, сравнивая ее при этом с особенностями акустической картины потока речи. В связи с этим весьма существенную долю труда нам пришлось уделить разработке соответствующих методов и аппаратуры, описание которых и предпосылается изложению материалов исследования процессов образования и восприятия речи.

1. Датчики потоков воздуха из ротового и носовых отверстий

Разработка электрических методов точной регистрации движений воздуха оказалась неожиданно сложной задачей. Трудности связаны с рядом факторов и прежде всего — со слабостью потоков воздуха при фонации отдельных фонем, а также с кратковременностью толчков воздуха и резкими изменениями скоро-

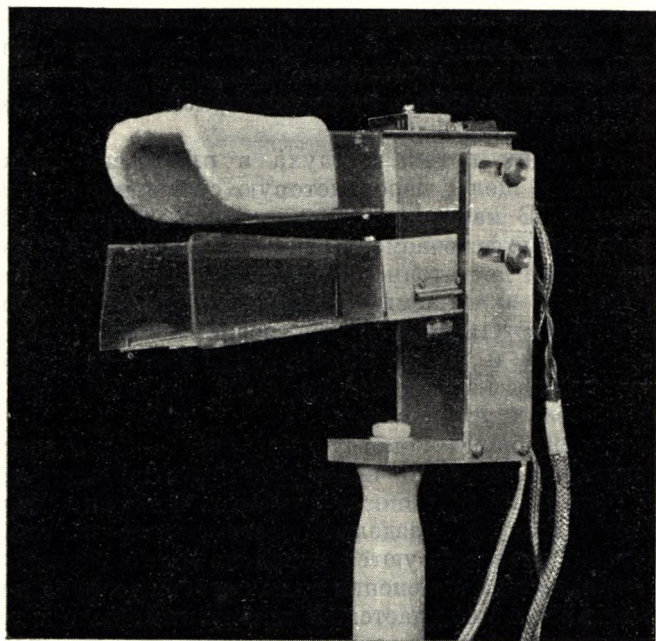


Рис. 2.1. Внешний вид рупорных фотоэлектрических датчиков для регистрации потоков воздуха из ротового и носовых отверстий.

стей потока, имеющими место при фонации некоторых звуков (например, взрывных), с наличием посторонних помех, возникающих из-за постоянных токов воздуха в результате движений испытуемого, практически неустраняемому легкому ветру в помещении и иным причинам.

Была сделана попытка использовать достаточно чувствительные емкостные датчики. Одной из обкладок конденсатора в данном случае являлась фольга, наклеенная на тонкую резиновую мембрану, прогибающуюся под влиянием потока воздуха. Второй обкладкой конденсатора служила металлическая пластина, помещавшаяся внутри чашечки, на которую натягивалась мембрана. Датчик включался в омически-емкостную мостовую схему, питав-

шующая током с частотой порядка 200 кгц. Существенными недостатками устройства оказались трудность балансировки, неустойчивость нуля и помехи, связанные с перемещением проводов схемы.

Практически более пригодными оказались фотоэлектрические датчики с рупорами, внешний вид которых показан на рис. 2.1. Потоки воздуха из ротовой щели и носовых отверстий поступают в два отдельных рупора, выходные отверстия которых закрыты тончайшими слюдяными прямоугольными пластинками, закрепленными с одной стороны. В спокойном положении пластинка прилегает к краям прямоугольного отверстия и является светопоглощающей преградой между миниатюрной электрической лампой и фотоэлементом.

При поступании потока воздуха в рупор пластинка изгибается, открывая щель, через которую свет лампочки попадает на фотоэлемент. В качестве последнего используются безынерционные вакуумные фотоэлементы типа СЦВ-2. Удобным оказалось также применение достаточно малоинерционных фотоспротивлений типа ФС-А4, включаемых в мостовую схему. Важным преимуществом при этом оказывается возможность питания моста переменным током и получение сигнала, который можно непосредственно подать на унифицированный усилительный и регистрирующий канал установки.

Датчики подобной конструкции отличаются достаточной чувствительностью и мало подвержены посторонним помехам. Благодаря малой инерционности датчиков временные характеристики потоков воздуха, отдельно поступающих изо рта и носа во время речи, регистрируются довольно четко. Однако точной картины, отражающей изменения силы потоков воздуха при фонации, получить все же не удастся. В связи с этим было желательно усовершенствовать датчики.

Перспективным с точки зрения чувствительности и малой инерционности представлялось использование электрических тепловых анемометров (Качурин, 1962). Были опробованы датчики в виде проволочек, нагреваемых током и изменяющих свое сопротивление благодаря охлаждению струей воздуха. Несмотря на применение достаточно тонких медных проволок (диаметром около 10 мк), не удалось добиться требующейся малой инерционности датчиков. Постоянные времена фронтов электрических сигналов, получающихся при подаче резких толчков воздуха, не могли быть сокращены менее чем до 80—50 мсек. В результате не удавалось обеспечить удовлетворительной регистрации быстрых изменений потоков воздуха, имеющих место при нормальных скоростях разговорной речи. Еще большая инерционность получалась при попытках использовать весьма подходящие с точки зрения чувствительности полупроводниковые микротермисторы (Шефтель, 1956).

Существенного уменьшения инерционности регистрации удалось добиться, применяя систему с обратной связью, изменяющей подогревный ток преобразователя (Партридж, 1961). Датчик (R_D) включен в схему моста, питаемого переменным током через трансформатор от усилителя мощности. Выходной сигнал с моста снимается на усилитель напряжения, имеющий резонансную характеристику с небольшой добротностью, и далее поступает на упомянутый усилитель мощности (рис. 2.2).

В замыкающейся таким образом системе с обратной связью возникают колебания с частотой, определяющейся настройкой

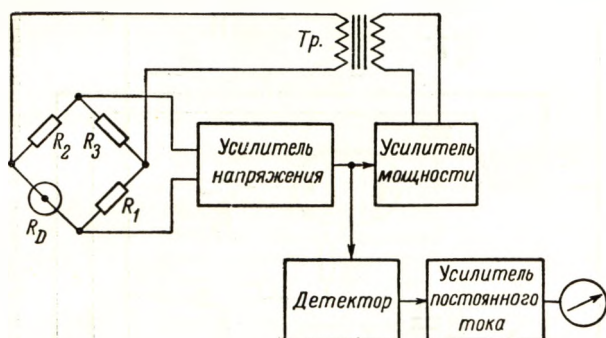


Рис. 2.2. Блок-схема малоинерционного термоанемометра.

Объяснения в тексте.

усилителя напряжения. Амплитуда колебаний устанавливается на некотором уровне благодаря стабилизирующему действию сопротивления R_D , значение которого изменяется с температурой. Поток воздуха, воздействующий на датчик R_D , охлаждает его в зависимости от линейной скорости воздуха (Туричин, 1951; Качурин, 1962). При этом сопротивление R_D изменяется так, что сигнал, снимаемый с моста на усилитель, увеличивается. Однако благодаря наличию обратной связи мощность питания моста возрастает, R_D дополнительно нагревается, и система снова стремится к равновесию моста. Как показывают расчеты (Партридж, 1961), инерционность регистрации в подобной системе уменьшается пропорционально усилению в петле обратной связи.

Выходным сигналом служит напряжение, снимаемое с усилителя в петле обратной связи. Оно детектируется, сглаживается и может дополнительно усиливаться для получения соответствующего масштаба регистрации. Описываемый электрический термоанемометр с обратной связью отличается от известных нам по литературе тем, что он работает как измерительная схема с самовозбуждением. Благодаря этому не требуется отдельного генера-

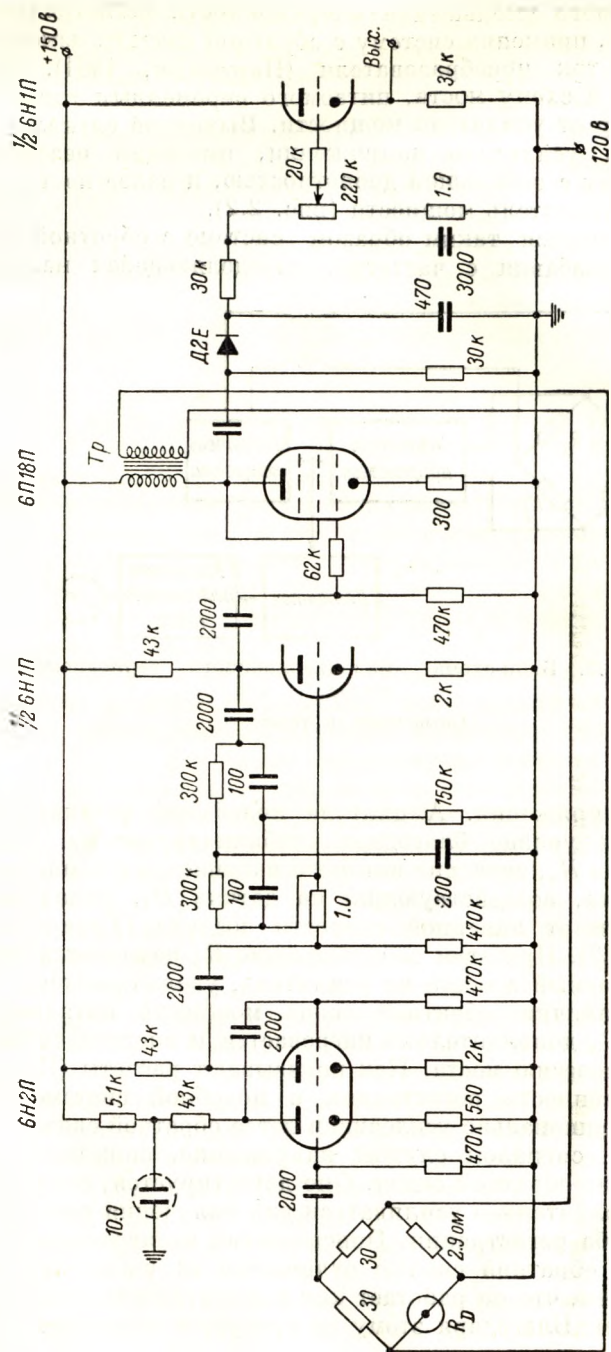


Рис. 2.3. Принципиальная схема малоинерционного термоанометра.

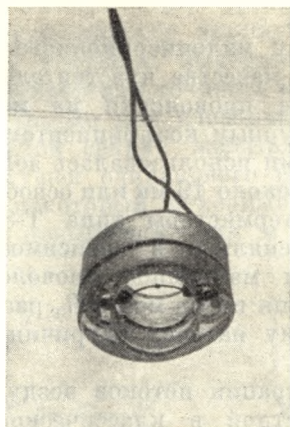


Рис. 2.4. Внешний вид датчика малоинерционного термоанемометра, регистрирующего скорость воздуха, выходящего из носового отверстия.

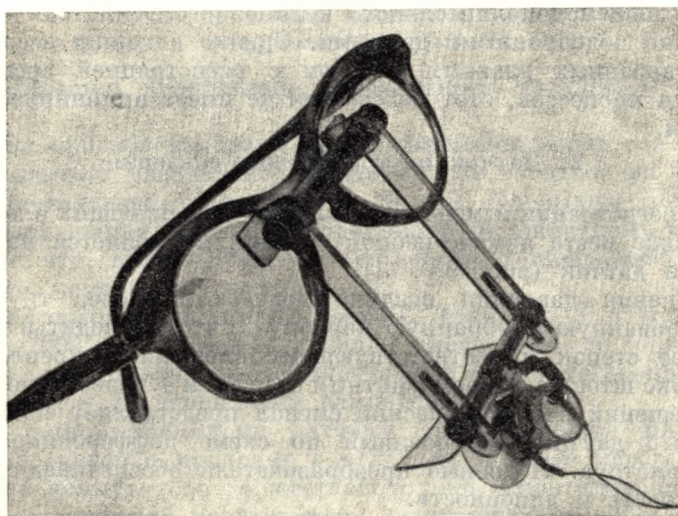


Рис. 2.5. Внешний вид датчика малоинерционного термоанемометра, регистрирующего скорость воздуха, выходящего из ротового отверстия.

тора, и все устройство при его достаточной несложности работает устойчиво и надежно.

Электронную схему малоинерционного термоанемометра показывает рис. 2.3. В качестве чувствительных элементов могут служить либо тонкие проволочки из металлов, обладающих достаточным температурным коэффициентом сопротивления, либо микротермисторы. Нами использовалась золоченая вольфрамовая проволока диаметром около 19 мк или освобожденные от стеклянной оболочки микротермисторы типа Т-8. Данные выходных трансформаторов изменялись в зависимости от того, применялся термистор или малоомное проволочное сопротивление. Величина сопротивления плеча моста R_1 рассчитывалась согласно допустимому исходному нагреву (Туричин, 1951; Нечаев, Удалов, 1961).

Датчик для регистрации потоков воздуха из носа оформлен в виде оливы, известной в классической экспериментальной фонетике (рис. 2.4). Датчик для регистрации воздушных потоков из ротовой щели заключен в небольшой рупор — амбушур и снабжен дополнительным козырьком, защищающим чувствительный элемент от попадания воздушной струи из носа. Датчик укрепляется на очковой оправе (рис. 2.5).

Следует отметить, что характеристика термоанемометра отличается значительной нелинейностью: по мере увеличения скорости потока чувствительность регистратора падает. Подобная неравномерность характеристики в принципе может быть исправлена введением дополнительного нелинейного усилителя с соответственно подобранными данными. Однако в наших исследованиях, связанных главным образом с регистрацией временной динамики процессов, этот недостаток не имеет принципиального значения.

2. Датчик внутриротового давления

Для регистрации изменений давления, возникающих в полости рта глубже места язычно-нёбной смычки, применяется манометрический датчик (рис. 2.6, А).

Изменения давления воздуха передаются через трубку 1 в гофрированную мембранную коробку 2, что приводит к прогибанию ее стенок и соответствующему смещению укрепленного на коробке штока с ферромагнитным стержнем 4. Преобразователь этих смещений в электрический сигнал представляет собою индуктивный датчик, построенный по схеме дифференциального трансформатора. Подобные преобразователи обеспечивают высокую точность и линейность.

На каждой из половин катушки 3 намотаны по две идентичных обмотки (рис. 2.6, Б). Две из них, включенные последовательно, присоединены к генератору переменного тока и возбуждают ЭДС одинаковой величины в двух других обмотках, расположен-

ных в тех же секциях катушки. Но эти снимающие напряжение катушки подключены навстречу друг другу. В результате в исходном положении ферромагнитного якоря, симметричном относительно обеих секций катушки, выходной сигнал имеет минимальное значение. Достаточная точность компенсации наводимых ЭДС обеспечивается только при идентичности обмоток. В связи с этим обмотки каждой секции желательно наматывать одновременно с двух катушек одинакового провода. Электрический нуль датчика (или некоторый исходный сигнал при желании регистрировать как повышения, так и понижения давления)

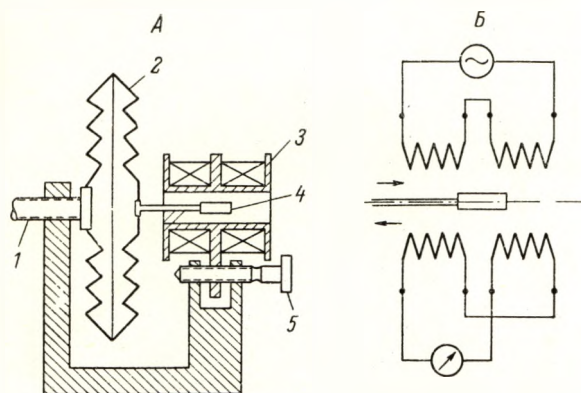


Рис. 2.6. Конструкция датчика внутриротового давления (А) и схема подключения обмоток (В).

Объяснения в тексте.

настраивается продольным перемещением катушки с помощью винта 5.

При прогибе мембраны якорь смещается вдоль оси катушки, что вызывает увеличение наводимой ЭДС в одной из обмоток и соответствующее уменьшение в другой. На выходе датчика появляется напряжение, пропорциональное линейному смещению якоря, что, в свою очередь, соответствует изменению давления внутри мембранной коробки.

Напряжение, питающее датчик, равно 1—2 в при частоте 1000—2000 гц. Выходные сигналы поступают на описываемый далее унифицированный усилительный канал установки и регистрируются на чернильном самописце.

К манометрическому датчику присоединяется пластиковая трубка, которая проходит в углу рта и далее между щекой и зубами испытуемого до жесткой трубки, вмонтированной в датчик динамической палатографии «искусственное небо» (рис. 2.7). Эта жесткая трубка огибает сзади верхние коренные зубы и открывается в полости рта по центральной линии недалеко от границы твердого и мягкого неба.

Для оценки инерционности регистрации в системе повышалось давление и отверстие датчика (или конца подсоединяемой трубки) резко открывалось. Характер выходного сигнала прослеживался на чернильном и электронном осциллографах. Постоянная времени фронта сигнала без дополнительной трубки не превышала 5 мсек. Короткая периодическая компонента сигнала составляла менее 0.1 от величины основной ступеньки. Присоединение к дат-

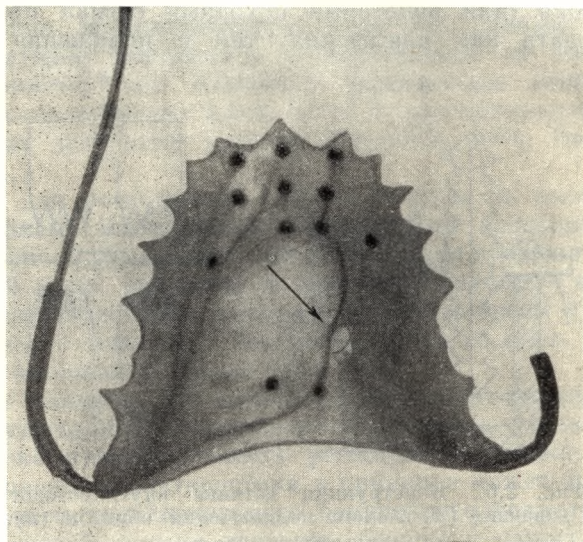


Рис. 2.7. Внешний вид датчика динамической палатографии «искусственное нёбо» со встроенной трубкой. Отверстие трубки указано стрелкой.

чику трубки длиной 400 мм и диаметром 3.5 мм увеличивало фронт сигнала не более чем до 8 мсек. Эти данные указывают на достаточное быстродействие всей системы регистрации.

Чувствительность системы при чернильной записи калибровалась с помощью простейшего водяного манометра и составляла обычно 30 мм водяного столба на 1 мм отклонения пера. Линейность электрического сигнала самого датчика оказалась весьма высокой, но она практически ограничивалась линейным диапазоном чернилопишущей системы (до 200—250 мм водяного столба).

3. Датчики для регистрации дыхательных движений

Для записи дыхательных движений используется широко известный метод пневмографии. Изменения периметра грудной клетки, а при необходимости и движения живота, регистрируются

с помощью датчиков сопротивления, выполненных в виде тонких резиновых трубок, наполненных проводящей средой (Хессе, 1962).

Применяющаяся модификация датчика отличается очень большой чувствительностью. Датчики изготавливаются из ниппельной резиновой трубки длиной около 50 см. Трубка плотно набивается очень тонко натертым порошком из чистого графита. Датчик обертывается вокруг тела в необходимом месте. Его исходное натяжение регулируется с помощью пояса с застёжками. Датчик включается в мост постоянного тока, питающийся напряжением 4—8 в. Выходной сигнал моста подается на усилитель постоянного тока с коэффициентом усиления, регулируемым в широких пределах, и далее на чернилопишущий 16-канальный осциллограф.

В установке имеются 4 идентичных моста, что позволяет одновременно регистрировать дыхательные движения на разных уровнях груди и живота. Чувствительность системы легко может быть доведена до столь высокого уровня, что наряду с дыхательными движениями четко могут отмечаться и сердечные толчки. С помощью подобного датчика можно регистрировать также и движения нижней челюсти. Для этого трубка повязывается вокруг головы таким образом, чтобы она проходила под подбородком.

4. Методы кинематографической и фотоосциллографической регистрации артикуляторных движений губ

Для подробного изучения сложной динамики артикуляции губ нижняя половина лица диктора кинематографировалась крупным планом анфас и в профиль. Для того чтобы оба эти изображения получить на одном кадре, было применено зеркало, которое ставилось под углом в 45° к лицу диктора и отражало изменение положения губ в профиль (рис. 2.8). Чтобы облегчить покадровые измерения артикуляторных движений, на губы диктора наносились точки с помощью туши. Покадровая обработка фильма производилась на диакопе Д-1; изображение кинокадра увеличивалось при этом в 10 раз.

Наличие двух изображений позволяет измерять движения губ в трех направлениях: вертикальном, горизонтальном и сагиттальном. В случаях, когда видны края зубов, измеряется расстояние между ними, а когда виден язык, то отмечается его форма.

Для контрольного сопоставления движений губ с невидимой артикуляцией языка, регистрация которой производилась с помощью методики динамической палатографии, а также со звуковыми параметрами речи, рядом с лицом диктора была установлена неоновая лампочка, периодические вспыхивания которой (3.5 раза в сек.) фотографировались одновременно с артикуляцией. Синхронно со вспыхиванием лампочки производилась

отметка на общей осциллограмме, где регистрировались в момент киносъемки сигналы динамической палатографии и параметры произносимых диктором звуков. В тех случаях, когда по видимым проявлениям артикуляции невозможно определить моменты перехода от согласного к гласному, и наоборот, такое сопоставление позволяет более точно устанавливать эти моменты.

Однако использование обычной киносъемки для регистрации артикуляторных движений губ имеет существенный недостаток, который заключается в том, что при скорости 24 кадра в 1 сек.

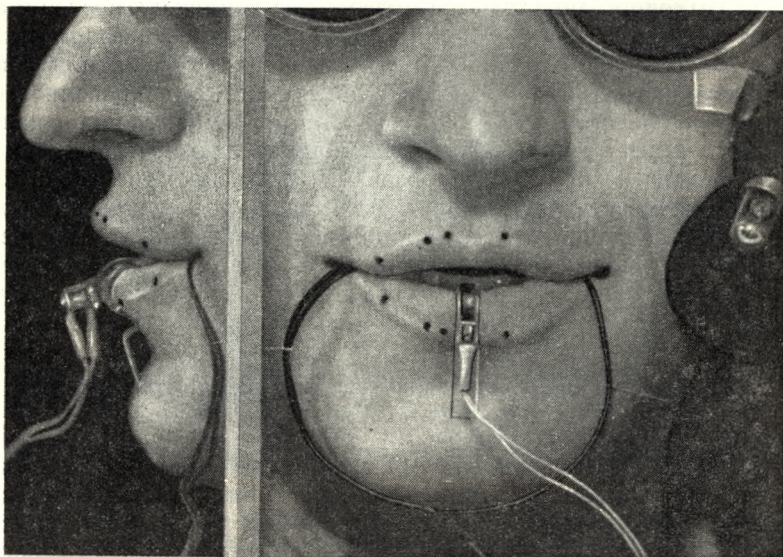


Рис. 2.8. Кадр из фильма, снятого одновременно с осциллографической регистрацией артикуляторных показателей.

Справа — мигающая неоновая лампочка для последующего сопоставления во времени видимой артикуляции и осциллограммы.

положение губ фиксируется через каждые 41 мсек., а реальная экспозиция уменьшается еще наполовину или более (в зависимости от угла открытия обтюратора, различного в разных типах кинокамер). Такая прерывистая киносъемка не может зафиксировать всю динамику артикуляторных движений губ, которые, по данным Фуджимура (Fujimura, 1961b), могут существенно изменить свое положение уже за несколько миллисекунд. Этот автор применил стробоскопическую технику освещения губ, позволившую производить съемку со скоростью 240 кадров в секунду, и показал, что начальное открытие губ происходит за 4—5 мсек., а через 40 мсек. раскрытие достигает уже 10 мм. Вторым недо-

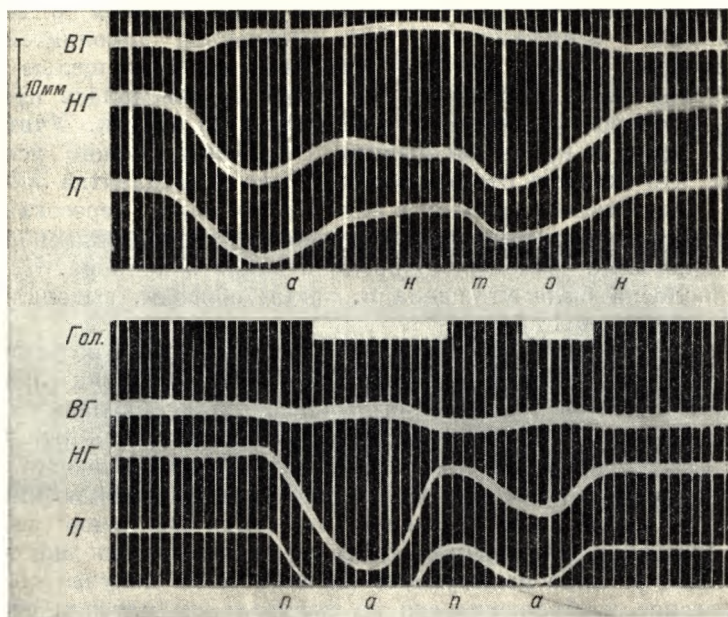
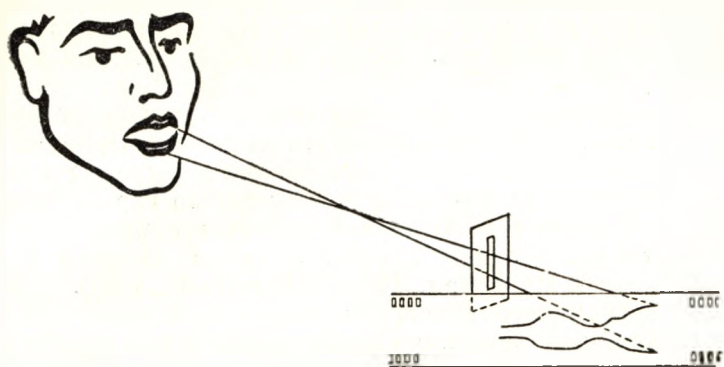


Рис. 2.9. Фотоосциллографическая регистрация вертикального размера ротовой щели.

На осциллограммах: движения верхней губы (ВГ), нижней губы (НГ) и подбородка (П) при произнесении слов *Антон* и *папа*. Гол. — отметка голоса (сигнала ларингофона); вертикальные линии — отметки времени.

статком киносъемки, даже ускоренной, является необходимость покадрового измерения положения губ, что требует большого труда, затрудняет получение статистического материала и может сопровождаться ошибками. Весьма целесообразно при изучении динамики движения губ пользоваться методикой, которая позволила бы регистрировать движения непрерывно и сразу получать осциллограммы, характеризующие эти движения. Принцип такой методики заключается в фотографировании движущегося объекта через узкую щель на фотопленку, непрерывно перемещающуюся с постоянной скоростью. Подобная методика применялась ранее для регистрации ширины зрачка, а также нистагма (Глезер, 1952, Левашов и Цукерман, 1961; Bellarminoff, 1885).

Для непрерывного фотографирования артикуляторных движений в какой-либо плоскости можно использовать фоторегистрирующее устройство с непрерывным движением фотопленки. В аппарате типа 1578/1, выпускаемом венгерской фирмой «Орион» и предназначенном для регистрации осциллограмм с экрана электронного осциллоскопа, был установлен дополнительный непрозрачный экран с узкой (0.1 мм) вертикальной щелью непосредственно перед фотопленкой. Съемка производилась при постоянной скорости движения фотопленки 30 мм/сек. Учитывая ширину щели и время экспозиции ($1/300$ сек.), можно условно считать, что скорость съемки объекта была эквивалентна 300 кадров в 1 сек. Отметка времени в виде тонких полос, пересекающих фотопленку в поперечном направлении, производилась через каждые 20 мсек. отметчиком времени данного прибора. Каждая пятая полоска была утолщена и, таким образом, выделяла отрезки времени по 100 мсек.

Для регистрации вертикального размера ротовой щели на губы испытуемого наклеивались две белые полоски липкого пластыря шириной в 1 мм. Третья полоска наклеивалась на подбородке с целью регистрации движений последнего. Это было важно для изучения вопроса о том, какой объем движения нижней губы обуславливается опусканием и подниманием нижней челюсти и какую часть составляет движение собственно нижней губы. Фоторегистрирующий аппарат устанавливался на таком расстоянии от лица испытуемого (23 см), чтобы получалось наиболее резкое изображение его на контрольном матовом стекле. Съемка производилась на фотопленку типа А-2.

Схему съемки и примеры осциллограмм показывает рис. 2.9. На нижней артикулограмме одновременно с записью движений губ и подбородка произведена отметка длительности звучания гласного *a* в ударном и безударном слогах слова *pāpa*. Эта отметка осуществлялась с помощью неоновой лампочки, зажигаемой выходным сигналом ларингофонного тракта. Одновременная запись движений губ и длительности звучания голоса позволяет производить сопоставление их по времени и получать более точ-

ное представление о временной характеристике механизма воспроизведения звуков речи.

С помощью методики непрерывной фотозаписи движений объекта в одной плоскости можно регистрировать также изменения горизонтального размера ротовой щели и, кроме того, динамику

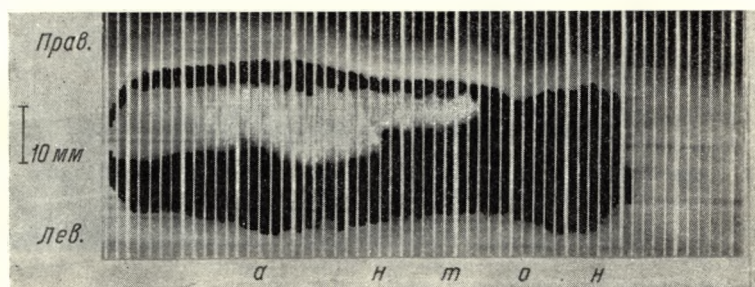
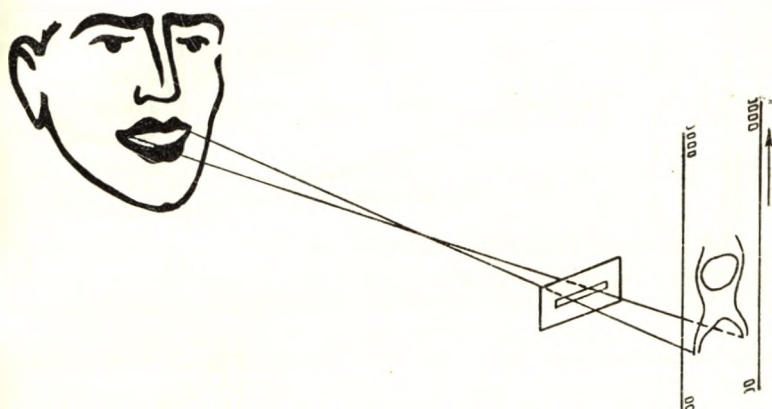


Рис. 2.10. Фотоосциллографическая регистрация горизонтального размера ротовой щели.

Прав. — правый, Лев. — левый углы рта. Отметка времени, как на рис. 2.9.

выпячивания губ при произнесении огубленных гласных. Для этого фоторегистрирующий аппарат нужно повернуть на 90° так, чтобы пленка перемещалась в вертикальном направлении. При записи изменений горизонтального размера ротовой щели губы диктора покрывались белым театральным гримом, а освещение подбиралось таким, чтобы при раскрытии губ меньше света попадало в полость рта и на зубы. Это увеличивало кон-

трастность между белыми углами рта и ротовой щелью. Схему и пример такой записи показывает рис. 2.10.

Для регистрации движений губ в сагитальном направлении (вперед и назад) к ним приклеивались с помощью липкого пла-

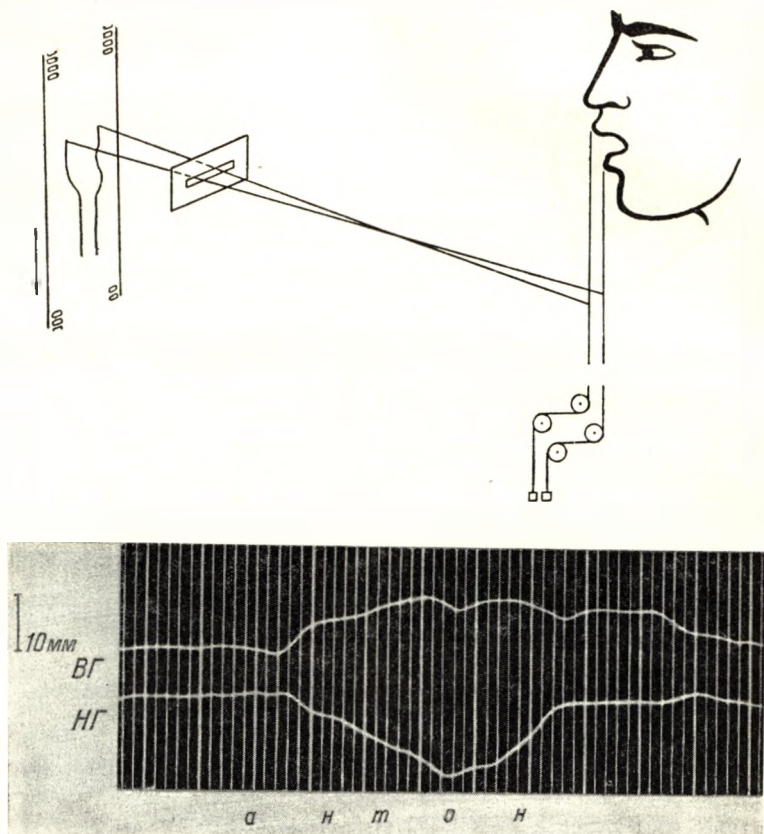


Рис. 2.11. Фотоосциллографическая регистрация движений верхней губы (ВГ) и нижней губы (НГ) в сагитальном направлении.

Отметка времени, как на рис. 2.9.

стыря толстые белые нитки, каждая из которых проходила через два ролика и несла на конце по небольшому грузу (5 г), чтобы обеспечить постоянство ее натяжения. Это несложное устройство предотвращало маятникообразные колебания нитей, но не мешало артикуляторным движениям. Смещение губ в сагитальной плоскости сопровождалось смещением ниток в той же плоскости. Фоторегистрирующий аппарат устанавливался сбоку и фикси-

ровал эти смещения ниток, обусловленные выпячиванием губ и обратным их движением. Схема движений губ в сагиттальном направлении и пример записи показаны на рис. 2.11.

Получаемые осциллограммы отпечатывались в увеличенном виде или копировались на прозрачную миллиметровую бумагу, с помощью диаскопа Д-1 и затем обрабатывались. Осциллограммы слов дают возможность получить данные о характере движений губ в трех плоскостях, величине этих движений и их временной характеристике. Эти данные обрабатывались далее статистически, и по ним вычислялась скорость движения губ (первая производная), а также ускорение этих движений (вторая производная).

5. Электрические методы регистрации артикуляторных движений губ

Фотографические методы записи движений губ не дают возможности использовать все те большие преимущества регистрации и автоматизации измерений, которые возникают в том случае, если изучаемые процессы предварительно преобразуются в электрические сигналы. В связи с этим были предприняты попытки использовать различные методы, дающие возможность получать электрические сигналы, с достаточной точностью говорящие о движениях губ во время речи.

Прежде всего мы обратились к записи мышечных биопотенциалов (так называемых электромиограмм — ЭМГ). Электроды диаметром около 5 мм располагались в разных точках кожи около ротового отверстия и на слизистой внутренних поверхностей губ. Регистрация ЭМГ производилась с помощью чернилопишущего электроэнцефалографа фирмы «Кайзер». Артефакты, связанные с перемещениями электродов, неизбежными при речевых движениях, могли быть существенно подавлены путем резкого ограничения усиления низких частот. Для этого на входе усилителей ставились дифференцирующие RC -цепочки с постоянной времени порядка 0.005 сек. Одновременные записи ЭМГ от нескольких точек производились при различных биполярных и монополярных отведениях. В последнем случае индифферентные электроды располагались на мочках ушей.

Выяснилось, что электромиографическим методом особенно речевых движений губ не удастся зарегистрировать с достаточной точностью и надежностью. Так, например, всегда наблюдалось существенное расхождение между началом и концом всплеск биопотенциалов, отражающих мышечное напряжение, и моментами реального открывания и закрывания ротовой щели. Оказалось весьма трудным, а часто и практически невозможным раздельно регистрировать работу различных мышечных групп, тесно переплетающихся в губах, но несущих существенно различные функции. Однако, весьма полезные данные могли все же

быть получены с помощью ЭМГ при оценке работы антагонистических групп мышц, обеспечивающих движения губ в различных слоговых конструкциях (см. далее).

Для изучения точных временных характеристик губных артикуляций оказалось невозможным ограничиться электромиографическими записями. Задача же разработки методов электрической регистрации, непосредственно отражающей характер перемещения губ, оказывается весьма непростой. Это связывается прежде всего с большой подвижностью и сложным характером движений губ, с наличием обязательных при речи движений нижней челюсти и сопутствующих движений всей головы.

Были опробованы различные приемы регистрации. Так, в оптическую систему, похожую на описанное выше устройство для непрерывной фоторегистрации движений губ, в качестве преобразователя оптического сигнала в электрический был введен фотоэлектронный умножитель ФЭУ-17.

Довольно легко мог быть получен четкий сигнал об изменении вертикального размера ротовой щели. Однако при этом в чрезвычайной степени сказывалась нестабильность положения исходной линии в связи с тем, что на записи резко отражались даже самые малые движения головы. Достаточного успеха не обеспечивала даже практически явно неудобная фиксация положения головы с помощью специальных опор.

Были также сделаны попытки использовать различные датчики электрических сигналов, укрепляющиеся на губах. Так, опробовались индуктивные датчики в виде двух миниатюрных катушек с ферромагнитными стержневыми сердечниками. Катушки наклеивались на те точки губы, расстояние между которыми должно измеряться. Через одну из катушек пропусклся переменный ток с частотой в несколько десятков килогерц, с другой катушки снималась наводимая на нее ЭДС. Величина этого сигнала изменялась при изменениях расстояния между катушками.

К сожалению, подобная система регистрации отличается рядом существенных недостатков. Прежде всего следует отметить резко нелинейную зависимость величины снимаемого электрического сигнала от расстояния между катушками и сильную зависимость этого сигнала от изменения взаимного направления осей катушек. Последний фактор имеет решающее значение, так как реальные движения губ отличаются сложностью, и перемещения датчиков происходят одновременно в нескольких плоскостях.

По аналогичным причинам пришлось отказаться и от привлекающего своей принципиальной простотой бесконтактного метода измерения механических перемещений, основывающегося на использовании эффекта Холла (Богомолов, 1961).

Миниатюрный полупроводниковый датчик Холла располагался в одной точке губы или в какой-либо неподвижной точке лица, например на носу. В другой точке, смещения которой

должны были измеряться, располагался миниатюрный ($4 \times 4 \times 2$ мм) постоянный магнит, поле которого охватывало датчик Холла. Изменения магнитного поля, происходящие в связи с взаимными перемещениями магнита и датчика, вызывали изменения ЭДС Холла, снимаемой с датчика. Данная система также обладала упомянутыми недостатками (нелинейностью и зависимостью снимаемого сигнала от вращательных движений магнита). Достаточно точной регистрации добиться не удалось.

Была сделана попытка использовать эффект, заключающийся в изменении емкости между двумя взаимно смещающимися металлическими пластинками. Измерения удобно производить согласно схеме, изображенной на рис. 2.12. Переменное напряжение U_r от генератора (при частоте 10 кГц или более) подводится к одной из пластинок (1). Вторая пластинка (2) соединяется через катодный повторитель со входом усилителя. Экран входного кабеля соединяется с катодом катодного повторителя, что дает возможность в существенной степени нейтрализовать емкость кабеля, протягивающегося от испытуемого к аппаратуре.

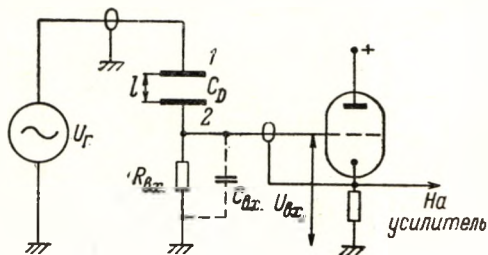


Рис. 2.12. Схема для регистрации движений губ при помощи емкостного датчика.

Объяснения в тексте.

Емкость между пластинками (C_d) и сложное входное сопротивление ($R_{вх.}$ и $C_{вх.}$) усилительной системы могут рассматриваться как делитель напряжения U_r . Изменение расстояния между пластинками (1) ведет к изменению емкости C_d , что, в свою очередь, отражается на величине снимаемого напряжения $U_{вх.}$ В наших опытах напряжение U_r выбиралось порядка 10 в, при этом исходная величина $U_{вх.}$ достигала нескольких десятков милливольт.

Одна из пластинок укреплялась на губе, другая на расстоянии нескольких сантиметров перед ней на картонном шаблоне, вырезанном по профилю испытуемого. К сожалению, и в данной системе проявлялись указанные ранее недостатки.

Наиболее четкие результаты при попытках получить количественные характеристики губных артикуляций были достигнуты при регистрации изменений электрического сопротивления участка губ, находящегося между двумя электродами. Измерение сопротивления производится на переменном токе звуковой или ультразвуковой частоты подобно методу, известному в физиологии под названием электрическая плетизмография (Москаленко и Науменко, 1956; Орлов, 1961).

Каждый из электродов в виде серебряного кружка диаметром около 2 мм располагается на тонкой плексигласовой полоске. Последняя изогнута в виде незамкнутого кольца и может быть одета на губу. Пара таких электродов располагается, например, на нижней губе симметрично относительно средней линии. При выпячивании губы вперед электроды сближаются, а само тело губы утолщается. В результате действия этих факторов сопротивление между электродами уменьшается, и сложное по своему характеру движение губы регистрируется однозначно.

Схему измерений показывает рис. 2.13. В качестве источника переменного тока используется генератор ЗГ-12. Объект вклю-

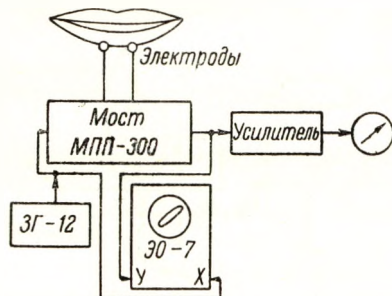


Рис. 2.13. Блок-схема установки для регистрации губных артикуляций путем измерения электрического сопротивления.

Объяснения в тексте.

чается в мост полных проводимостей типа МПП-300, позволяющий измерять сопротивления, обладающие как активной, так и реактивной компонентами. Процедура исходной компенсации моста существенно облегчается с помощью осциллоскопа (ЗО-7), используемого как индикатор баланса моста и по активной, и по реактивной компонентам. Измерения проводятся на частоте около 20 кГц, что дает возможность использовать унифицированный усилительный и регистрирующий канал, описываемый далее в разделе 8 главы 2.

Метод, основывающийся на измерении электрического сопротивления ткани, оказался наиболее подходящим при исследовании временных соотношений движений губ, относящихся к артикуляции гласного и согласного в слове.

Значительно легче можно было осуществить достаточно точную регистрацию более простого, но фонетически важного явления — моментов открывания и закрывания ротовой щели. Для этой цели было изготовлено и опробовано несколько вариантов электрических контактных датчиков. На верхнюю и нижнюю губы на изолирующей подкладке наклеивались кусочки тонкой фольги, при соприкосновении которых замыкалась электрическая цепь. Была опробована конструкция датчика в виде двух легких контактирующих между собою при сжатии губ рычагов, укрепляемых на очковой оправе. Однако эти датчики, а также и некоторые другие модификации не отличались достаточным удобством и надежностью.

Датчик более удачной конструкции, обеспечивающий точную регистрацию моментов смыкания и размыкания губ, практически

не изменяющий произношения и не мешающий испытуемому, представляет собою охватывающую нижнюю губу слегка пружинящую круглую скобку из плексигласа (рис. 2.14). На скобке укрепляется миниатюрный игловидный золотой контакт. Сверху над ним располагается тонкая стальная пружина, изогнутая по форме скобки. Достаточно малый контактный зазор и гибкость пружины обеспечивают соприкосновение контактов при самом

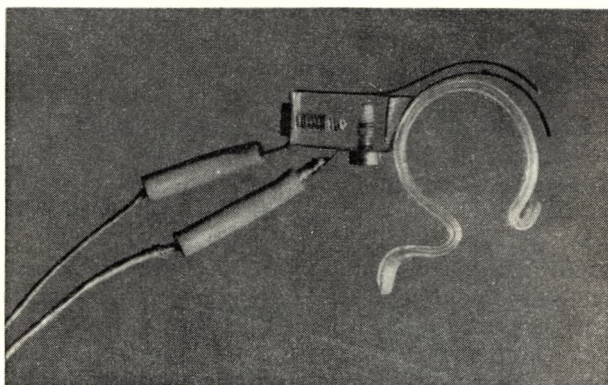


Рис. 2.14. Внешний вид контактного датчика, регистрирующего смыкание губ.

легком нажатии верхней губы на датчик. Контакты замыкают цепь, подающую сигнал переменного тока от генератора синусоидальных колебаний на вход унифицированного усилительного канала.

Несмотря на предельную простоту метода, надежное получение четкого электрического сигнала при смыкании контактов оказалось не совсем простой задачей. Так, регистрация могла нарушаться при попадании в датчик слюны. Устранить этот дефект удалось путем регулярного промывания датчика с последующим смазыванием контактных поверхностей легким часовым маслом и включением на вход усилителя низкоомного сопротивления (100 ом).

6. Динамическая палатография

Довольно сложной задачей оказалась разработка методики регистрации движений языка. Основные требования сводились к следующему: регистрация должна была осуществляться непрерывно, по возможности не связывать движений языка и не вызывать искажений акустического сигнала.

Из существующих методик, дающих возможность исследовать динамику движений языка в процессе связной речи, практиче-

ское применение нашла только кинорентгенография. Однако ряд особенностей этого метода делает его до некоторой степени уникальным. Во-первых, кинорентгенография не позволяет производить длительные измерения. Во-вторых, применение этого метода ограничено сложностью необходимой аппаратуры, длительностью и трудоемкостью процедуры получения данных.

На протяжении последних лет рядом исследователей делались попытки приспособить для указанных целей электромиографию (Кратин, 1955; Бассин и Бейн, 1957; Чаргейшвили, Тохадзе, Польшин, 1959). Однако эти попытки до сих пор не дали положительных результатов. Это становится понятным, если учесть сложную анатомию языка, глубокое залегание ряда приводящих его в движение мышц, трудность крепления электродов и т. п.

Таким образом, до последнего времени не существовало достаточно простой и надежной методики, которая позволяла бы непрерывно получать сведения о положении языка в процессе связной речи.

Положения языка при артикуляции согласных исследованы фонетиками преимущественно посредством метода палатографии. Сущность его заключается в следующем. В рот испытуемого вкладывается покрытая краской или припудренная тальком пластмассовая пластинка, точно повторяющая контуры твердого нёба. При произнесении звука на пластинке остается отпечаток языка, который может быть срисован или сфотографирован. При другой модификации этого метода, получившей название прямой палатографии, пластинка не применяется, а краска наносится непосредственно на нёбо испытуемого. После произнесения звука нёбо фотографируется через зеркало (Жренти, 1960; Witting, 1953; Abercrombie, 1957; Ladefoged, 1957).

Полученные любым из этих способов палатограммы дают возможность определить положение языка и до некоторой степени его конфигурацию при артикуляции большинства согласных и ряда гласных звуков. Однако этот метод не позволяет получить более одной палатограммы на протяжении произнесения любой длительности. Тем самым он не может быть использован для исследования динамики изменений положения языка при связной речи.

Первая попытка сделать палатографию непрерывной принадлежит Стетсону (Stetson, Hudgins a. Moses, 1940; Stetson, 1951). Он использовал нёбные пластинки, в которые были вделаны небольшие резиновые баллоны, связанные трубками с капсулами Маррея. Один такой баллон располагался у переднего, а другой — у заднего края нёба. При артикуляции переднеязычных и заднеязычных смычных согласных баллоны сдавливались языком, что вызывало отклонение писчика, укрепленного на капсуле Маррея. При помощи этой методики сравнительно точно измерялись временные характеристики взрывных согласных, однако метод лишен других преимуществ классической палатографии.

Поскольку палатография дает возможность довольно точно судить о положении языка при артикуляции большинства согласных и в то же время не связана с необходимостью крепить на языке какие-либо датчики, мы решили сконструировать многоканальный электрический датчик, в определенной мере напоминающий пневматический датчик Стетсона. После ряда попыток была разработана методика непрерывной или динамической палато-

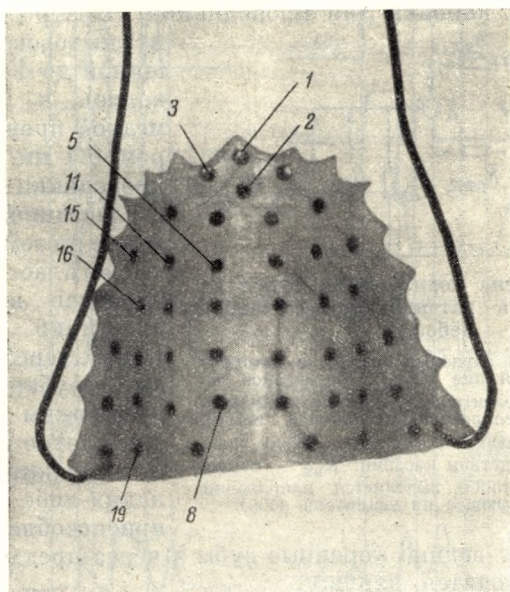


Рис. 2.15. Внешний вид датчика динамической палатографии «искусственное нёбо».

Номер — электроды, *расположенные в типичных точках, упоминаемых в тексте.

графии, осуществляемой при помощи датчика «искусственное нёбо» (Кузьмин, 1963).

Датчик представляет собой тонкую пластмассовую пластинку, на поверхности которой размещены электроды (рис. 2.15). На электроды, расположенные на одной половине пластинки, подается переменное напряжение; на электродах, расположенных на другой половине, сигнал возникает в те моменты времени, когда язык, являющийся проводящей средой, касается соответствующей зоны пластинки. Датчик подключается к описываемым далее измерительным каналам, которые осуществляют усиление, детектирование и сглаживание получаемых сигналов, после чего они регистрируются на многоканальном чернильном самописце,

а также могут измеряться с помощью ряда электронных устройств. Схема подключения одной из пар электродов датчика приведена на рис. 2.16.

Нёбные пластинки изготавливались нами индивидуально для каждого испытуемого из пластмассы АКР-7, применяемой в стоматологии. Технология их изготовления была такой же, как и при изготовлении нёбных пластинок при зубном протезировании. В готовых пластинках бором просверливались гнезда для электродов и канавки для проводников. Электроды вырубались

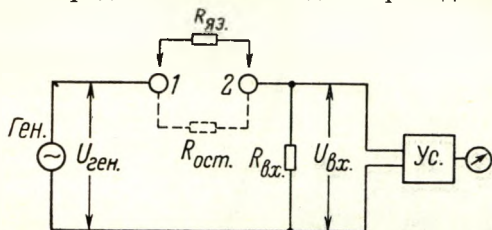


Рис. 2.16. Схема подключения одной из пар электродов датчика «искусственное нёбо».

Ген. — генератор звуковых частот, с которого снимается напряжение $U_{ген.}$, подаваемое на электрод 1; 2 — электрод, с которого снимается сигнал; $R_{яз.}$ — сопротивление тканей языка; $R_{ост.}$ — остаточное сопротивление между электродами при отсутствии касания; $R_{вх.}$ — сопротивление, на котором выделяется напряжение ($U_{вх.}$), поступающее на усилитель (Ус.).

из листового серебра толщиной до 1 мм и припаивались к лакированной медной проволоке диаметром 0.1 мм. Электроды и проводники крепились к пластинке зубным цементом, а потом покрывались той же пластмассой. Концы проводов выводились из пластинки в задних ее углах через хлорвиниловые трубки. Изготовленный таким образом датчик плотно прилегает к нёбу и держится без помощи каких-либо фиксирующих приспособлений. Проводники

огинают задние коренные зубы и через преддверие и отверстие рта выводятся наружу.

При использовании датчиков у испытуемых первое время наблюдается несколько повышенное слюноотделение и ухудшение артикуляции. Однако они довольно быстро привыкают к ношению датчика и начинают правильно артикулировать уже через несколько часов тренировки. Речь при говорении с датчиком на слух ничем не отличается от их речи без датчика. Произведенное на одном из испытуемых контрольное определение разборчивости речи показало, что ношение датчика практически не ухудшает произношения. Слоговая разборчивость при зачитывании стандартных слоговых таблиц с датчиком составила 87.5% при 89.9% слоговой разборчивости речи в тех же условиях без датчика.

Описанные датчики были изготовлены для шести испытуемых. Типичные динамические палатограммы слов приведены на рис. 2.17. Каждый из регистрируемых параметров соответствует определенной точке твердого нёба (нумерация упоминаемых точек дана на рис. 2.15). Наличие сигнала соответствует наличию контакта языка с нёбом в данной точке, а отсутствие сигнала — отсут-

ствию контакта. Таким образом, конфигурация зоны контакта может быть восстановлена по наличной комбинации сигналов на выходе датчика в любой заданный момент времени.

Изложенная методика дает возможность непрерывно осуществлять регистрацию положений языка при артикуляции большинства согласных, а также гласных высокого подъема в процессе

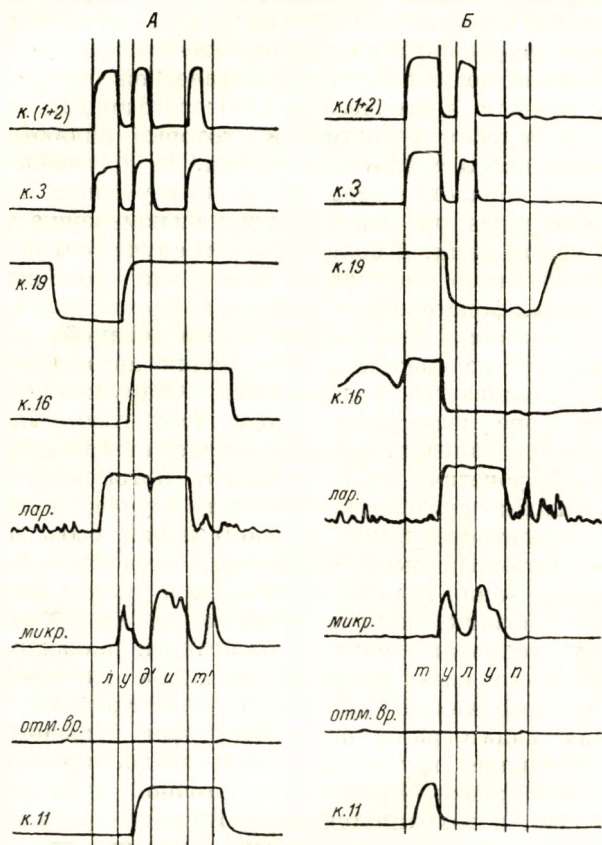


Рис. 2.17. Осциллограммы слов *лудить* (А) и *тулуп* (Б).

связной речи. При этом объективному исследованию становятся доступными временная структура согласных и особенности их артикуляции в разных фонетических положениях. Кроме того, регистрируемые сигналы могут быть использованы как опорные точки при исследовании более общих характеристик процесса речеобразования. Так, методика оказалась полезной при осуществлении ряда психофизических экспериментов, связанных

с изучением механизмов речеобразования и восприятия речи человеком (Чистович, Клаас, Кузьмин, 1962; Чистович, Бондарко, 1963).

Так как в ходе подобных исследований возникает необходимость обработки большого количества данных, ручная обработка материала оказывается довольно трудоемкой, а измерения временных характеристик процесса не во всех случаях могут быть осуществлены визуально с достаточной точностью. В силу этого была исследована возможность автоматической обработки получаемых при помощи динамической палатографии данных и определены такие условия регистрации, которые являются с этой точки зрения наиболее благоприятными. В простейшем случае автоматическая обработка сводится к обнаружению начала или конца сигналов в разных каналах датчика и измерению временных интервалов между ними. В более сложном случае сигналы, между которыми необходимо произвести измерение, должны соответствовать сложным фонетическим явлениям и формироваться устройством в соответствии с заложенной в него логикой.

Очевидно, что возможность применения автоматической обработки обуславливается надежностью обнаружения полезных сигналов на фоне возможных помех, точностью обнаружения начала и конца сигналов и степенью соответствия регистрируемого сигнала тому фонетическому явлению, которое исследуется.

Для того чтобы определить, во всех ли случаях полезный сигнал (наличие контакта языка с нёбом) может быть обнаружен в процессе автоматической обработки, следовало установить пределы варьирования величины сигналов и уровня помех. В качестве помех, возникающих в процессе регистрации, следует рассматривать возможное наличие остаточной проводимости при отсутствии смычки языка с нёбом, вызванное главным образом увлажнением поверхности датчика. При визуальной обработке динамических палатограмм в подавляющем большинстве случаев не представляет труда определить, когда в том или ином канале имеется сигнал. Величина остаточной проводимости относительно мала и сохраняется на протяжении регистрации довольно стабильно. Однако можно было ожидать, что некоторые изменения величины остаточной проводимости в ходе регистрации имеют место. Кроме того, величина полезного сигнала также может варьировать по ряду причин: из-за окисления электродов, из-за изменения степени влажности в полости рта и разной плотности смыкания языка с нёбом. Очевидно, что при больших изменениях отношения сигнала к шуму автоматическое обнаружение сигналов не могло бы быть надежно осуществлено. Можно думать, что условия детектирования сигналов могут быть улучшены путем соответствующего подбора величины подаваемого на датчик напряжения. Эту величину можно было в определенных пределах произвольно изменять, но поскольку характер зависимости от-

ношения сигнала к шуму от величины напряжения не был ясен, оптимальная величина напряжения могла быть установлена только экспериментально.

Вопрос о точности определения начала и конца сигналов непосредственно связан с предыдущим. Так как автоматической обработке должны подвергаться сигналы, полученные в результате сглаживания высокочастотной несущей, частота несущей оказывается тем дополнительным фактором, который определяет

возможную точность измерений. Чем выше частота несущей, тем меньшей может быть постоянная времени сглаживания и тем точнее может быть определен момент образования смычки. Однако применение высоких частот ограничивается наличием емкостных связей между проводниками и, следовательно, возникновением дополнительных помех. Надлежало выбрать такую частоту несущей, которая позволяла бы сделать постоянную времени сглаживания минимальной при отсутствии дополнительных помех.

Для определения пределов варьирования остаточной проводимости были произведены измерения величины напряжения на электродах датчика к.1 и к.19 при наличии и отсутствии смычки на протяжении сравнительно длительной записи. При

этих измерениях был использован один испытуемый, так как интересовавшие нас в данном случае процессы связаны с конструкцией датчика и условиями регистрации и, очевидно, не зависят от диктора. Измерение осуществлялось при помощи лампового вольтметра, подключенного на выходе датчика. Замеры производились через каждые 4 мин. на протяжении часа.

Оказалось, что проводимость между отводящими электродами и электродами, на которые подается напряжение, при наличии смычки значительно варьирует, но всегда остается примерно на один порядок выше остаточной проводимости (рис. 2.18). Заслуживает внимания, что проводимость при отсутствии смычки мало зависит от расстояния между электродами. Последнее существенно, так как позволяет применять для автоматического детектирования сигналов в разных каналах идентичные схемы.

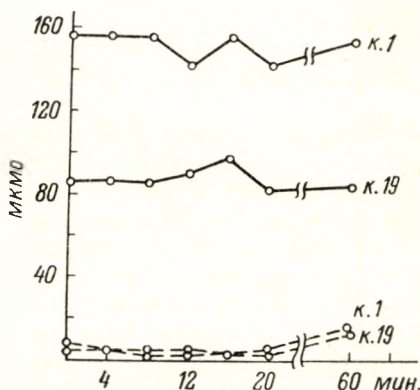


Рис. 2.18. Изменение проводимости между электродами датчика «искусственное небо» в ходе длительного опыта.

Сплошные кривые — проводимость для к.1 и к.19 при наличии смычки языка с небом; прерывистые кривые — проводимость для тех же электродов при отсутствии смычки. По оси абсцисс — время (в мин.), по оси ординат — проводимость (в мкмо).

Приведенные величины были получены при напряжении, равном 250 мв. Чтобы более точно определить величину напряжения, обеспечивающую наиболее выгодное отношение сигнала к шуму, были проведены дополнительные исследования. При различных напряжениях, подаваемых на датчик, измерялись величина сигналов и уровень помех для 3 разных электродов датчика при ряде произнесений двумя дикторами одного и того же слога *та*. Подводимое напряжение менялось ступенями от 50 мв до 1 в.

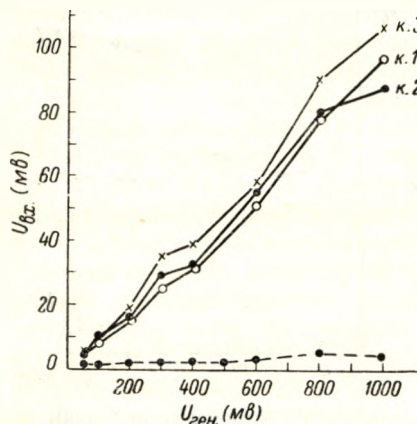


Рис. 2.19. Зависимость величины напряжения $U_{вх.}$, снимаемого с разных электродов датчика (по оси ординат), от величины напряжения $U_{ген.}$, подводимого к датчику (по оси абсцисс).

Сплошные кривые — напряжения на электродах к.2, к.3 и к.15 при касании языком пёба; прерывистая кривая — предельные значения $U_{вх.}$ на тех же электродах при отсутствии касания.

$т', д, д', с, с', з, з', ц, ч', н, н', л, л'$. Всего были исследованы четыре диктора. Звуки произносились в разных фонетических положениях в составе двухсложных слов в ударных и безударных слогах. Напряжение на входе датчика во всех случаях было 600 мв.

Полученные результаты показали следующее:

а) средняя величина напряжения на к.3 при произнесении указанных звуков варьировала у разных дикторов от 46 до 128 мв;

б) существенной разницы в величине сигналов при артикуляции согласных в ударных и безударных слогах не было обнаружено;

в) в среднем величина сигнала при артикуляции мягких согласных была больше величины сигнала при артикуляции твердых согласных;

Оказалось, что уровень шумов во всем исследованном диапазоне входных напряжений остается относительно низким, а интенсивность сигналов возрастает пропорционально увеличению входного напряжения (рис. 2.19). Поэтому условия обнаружения сигналов по мере увеличения подаваемого на датчик напряжения существенно улучшаются. Поскольку при применении указанных напряжений дикторы не испытывали никаких неприятных ощущений, вызываемых током, оказалось возможным использовать в дальнейшем напряжения до 600—1000 мв.

Для определения пределов варьирования величины полезных сигналов при помощи той же методики были определены интенсивности сигналов на электроде к.3 при произнесении разными дикторами звуков *т*,

г) величина сигналов значительно варьировала от произнесения к произнесению, но ни разу не была меньше 27 мв,¹ а уровень помех, измерявшийся на протяжении эксперимента, ни разу не превышал 6 мв.

Таким образом, поскольку уровень шумов остается относительно устойчивым и значительно более низким, чем возможные значения сигналов, автоматическое обнаружение сигналов может быть осуществлено сравнительно простыми техническими средствами, а именно при помощи амплитудного селектора с фиксированным уровнем селекции. При этом, поскольку в качестве полезного сигнала рассматривается факт наличия или отсутствия смыкания языка с нёбом, изменчивость абсолютных значений сигналов не является существенной.

Наиболее сложной задачей является установление наличия щели при артикуляции целевых переднеязычных c, c', z, z' . Для различения палатограмм, характерных для этой группы звуков, от палатограмм, характерных для смычных переднеязычных, были выбраны точки, обозначенные на рис. 2.15 номерами 1 и 2. В ходе предварительных записей было установлено, что при артикуляции смычных по крайней мере в одной из этих точек регистрируется смыкание. При артикуляции целевых сигнал на обоих электродах отсутствует. Тем не менее, так как ширина щели при артикуляции c и z измеряется несколькими миллиметрами, надлежало проверить, не сопровождается ли артикуляция целевых таким повышением наводок на указанных электродах, которое сделало бы обнаружение щели автоматическими средствами невозможным. Чтобы проверить это, были произведены измерения величины сигнала на электродах $\kappa.1$ и $\kappa.2$ при артикуляции звуков c, c', z, z' в составе 16 разных слов. Всего эти звуки были произнесены тремя дикторами 144 раза. Во всех случаях, кроме одного, сигнал на электродах $\kappa.1$ и $\kappa.2$ не превышал уровня помех, имеющих место в том случае, когда рот испытуемого был раскрыт.

Эти данные позволяют утверждать, что влияние близко расположенных электродов друг на друга практически не имеет места. Следовательно, конфигурация зоны контакта может быть установлена с большой степенью точности.

Для определения оптимальной частоты несущей были произведены осциллографические записи сигналов, получаемых с датчика при разной частоте подаваемого на датчик тока. Для этого был использован осциллограф типа Н-102, снабженный специальным усилителем. Полученные осциллограммы показали, что в диапазоне частот по крайней мере до 8000 гц увеличение уровня шумов не имеет места (рис. 2.20) и для питания датчика могла

¹ Исключение представлял звук $л$, который в измеряемой точке в ряде случаев вовсе не регистрировался.

быть выбрана частота порядка нескольких тысяч герц. Как правило, нами использовалась частота 2000 гц при постоянной времени сглаживания порядка 5 мсек. Так как время нарастания интенсивности реального сигнала, как это видно на осциллограмме, исчисляется несколькими миллисекундами, применение указанной постоянной времени сглаживания не приводило к существенному ухудшению условий детектирования сигналов.

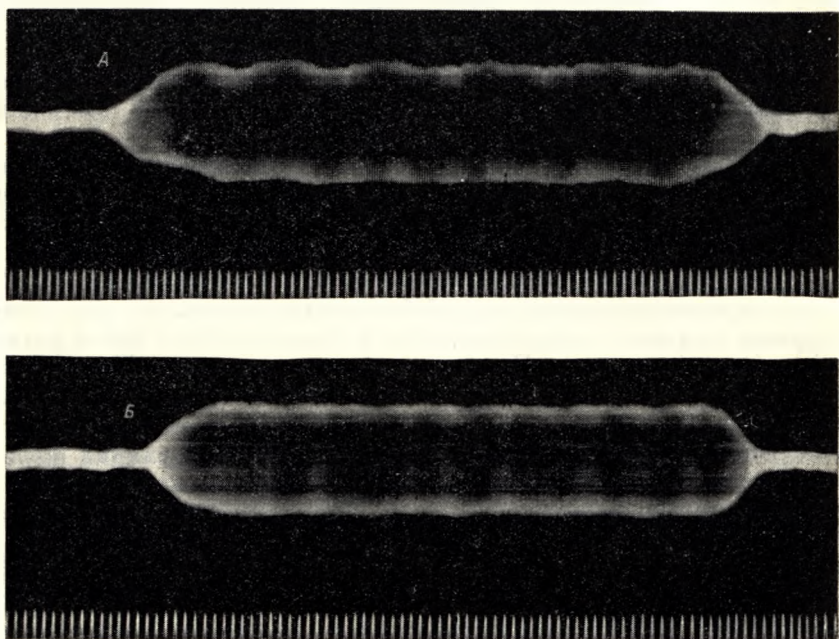


Рис. 2.20. Осциллограммы переменного напряжения (U_{vx}) на электроде к.11 при произнесении слога *ля* при частоте подводимого напряжения 2000 гц (А) и при частоте 6000 гц (Б). Отметка времени 2 мсек.

Одним из наиболее сложных вопросов является определение необходимого количества электродов и их размещение на поверхности датчика. Разумеется, в тех случаях, когда желательно определить наиболее точную конфигурацию зоны контакта, количество электродов должно быть максимальным. При этом, как показали записи, расположение контактов, на которые подается напряжение, и контактов, с которых снимается сигнал, не обязательно должно быть симметричным. Фактически напряжение может подаваться на любую точку слизистой рта, так что вся поверхность датчика может быть использована для размещения «полезных» электродов. Это обстоятельство существенно так как

в некоторых случаях возможны асимметричные положения языка, например при исследовании дефективной речи.

В большинстве же экспериментальных условий оказывается необходимым использовать только те электроды, которые позволяют различать некоторые типичные рисунки палатограмм. В частности, при исследовании потока речи измерение временных интервалов между сигналами, возникающими в разных каналах датчика, имеет смысл лишь в том случае, если эти сигналы являются достаточно надежными признаками определенных фонетических явлений. Возможность выделения таких признаков не является очевидной. Как известно из данных, полученных методом статических палатограмм, зоны контакта при произнесении одного и того же звука значительно варьируют от диктора к диктору (Богородицкий, 1930) и от произнесения к произнесению у одного и того же диктора (Stetson, Hudgins a. Moses, 1940). Следовало, таким образом, выяснить, может ли быть определен некоторый ограниченный набор точек, по наличию или отсутствию контакта в которых можно различать характерные для артикуляции разных звуков палатограммы.

Для решения этого вопроса при помощи классической методики были сняты статические палатограммы всех русских согласных звуков при их произнесении одним из дикторов. Далее путем сопоставления полученных палатограмм были выделены такие зоны на поверхности нёба, на протяжении которых контакт постоянно имеет место при артикуляции звуков одной группы и отсутствует при артикуляции звуков другой группы. При изготовлении датчиков электроды располагались именно в этих зонах.

Возможность выделения таких зон сама по себе не является ответом на поставленный вопрос. Так как при получении статических палатограмм звуки могут произноситься только в определенных фонетических положениях, а число полученных палатограмм было ограниченным, можно было думать, что в связной речи рисунки палатограмм варьируют в гораздо больших пределах. Дальнейшее исследование проводилось путем регистрации динамических палатограмм. Для этого дикторы зачитывали таблицы слов, в которых исследуемые звуки встречались в разных фонетических положениях. Темп произнесения специально не регулировался и приблизительно соответствовал средней скорости чтения текста. Приведем некоторые данные, полученные при графической обработке полученных записей.

При произнесении звуков *с, с', з, з'* тремя дикторами наличие щели было правильно определено в 279 из 282 случаев, т. е. только в 1% случаев на одном из срединных электродов (*ж.1* или *ж.2*) ошибочно был обнаружен сигнал. При произнесении звуков *т, т', д, д', л, л', н, н'* сигнал на том или другом из срединных электродов (или на обоих сразу) был зарегистрирован в 815 из 828 слу-

чаев.¹ Наличие сигнала на *к.3*, расположенном в зоне, характерной для артикуляции переднеязычных, было определено правильно во всех случаях артикуляции этих звуков, как смычных, так и щелевых.

Наличие контакта в точке *к.3* определялось, кроме того, в ряде случаев при артикуляции мягких заднеязычных *к', г', х'*. Однако, как показали записи, при артикуляции этих звуков достаточно надежно обнаруживается смыкание в точке *к.8*. Так, из 71 случая артикуляции *к', г', х'* двумя дикторами контакт в этой точке во всех случаях имел место. При артикуляции этими же дикторами переднеязычных *т, т', д, д', с, с', з, з', л, л', н, н'* наличие контакта в точке *к.8* ни разу не было обнаружено (всего 354 случая). Таким образом, если в качестве признака переднеязычных рассматривать сложное явление — наличие контакта *к.3* при отсутствии контакта в точке *к.8* — различение характерных для этих звуков палатограмм может быть достаточно надежным.

В преобладающем числе случаев артикуляции звука *ч* контакт в точке *к.3* также имеет место. Однако наиболее характерным признаком палатограмм при артикуляции этого звука является наличие контакта в точке *к.5*. При произнесении твердых переднеязычных (126 случаев) контакт в этой точке ни разу не регистрировался.

Аналогичные данные были получены в условиях автоматического детектирования сигналов в процессе артикуляции переднеязычных звуков в составе разных слов пятью дикторами. В этих экспериментах было установлено:

1) наличие контакта в точке *к.3* при артикуляции *т, т', д, д', з, з', с, с', л, л', н, н', ц* обнаруживается в 99.4% (1771 измерение);

2) смыкание по крайней мере в одной из двух точек — *к.1* или *к.2* при артикуляции всех этих звуков, кроме *с, с', з, з'*, обнаруживается в 97% (2195 измерений), причем большинство ошибок имеет место при артикуляции *л*;

3) наличие щели при артикуляции *с, с', з, з'* обнаруживается более чем в 99% (в 142 случаях из 143);

4) смыкание в точке *к.5* при артикуляции *ч* было зарегистрировано в 143 из 150 случаев произнесения этого звука. При произнесении *т, л, л', н* смыкание в этой точке ни разу не регистрировалось, однако при артикуляции мягких переднеязычных *т'* и *д'* контакт в точке *к.5* имел место в 15 из 242 измерений.

Наиболее сложной задачей оказалось определение признака, характерного для артикуляции мягких переднеязычных. Как известно, мягкие звуки отличаются от твердых с тем же местом артикуляции более широкой зоной контакта. Из этих соображе-

¹ Из 13 случаев, когда контакта в этих точках не было обнаружено, 11 имели место у одного и того же диктора.

ний в качестве признака мягкости был выбран сигнал с *к.11*, расположенного несколько выше зоны контакта, характерной для артикуляции твердых *т*, *д*, *с*, *з*, *ц*. Как оказалось, контакт в этой точке действительно имеет место во всех случаях артикуляции мягких *т'*, *д'*, *с'*, *з'*, *н'*.

Однако в значительном числе случаев артикуляции соответствующих твердых согласных смыкание в точке *к.11* также имеет место. Если сместить электрод еще выше, так чтобы твердые звуки не регистрировались, регистрация мягких тоже оказывается недостаточно надежной. Таким образом, различить палатограммы, характерные для мягких и твердых переднеязычных, по наличной комбинации сигналов в определенных точках не представляется возможным. Иными словами, зоны контакта при артикуляции этих звуков оказываются в ряде случаев одинаковыми. Тем не менее, как можно видеть на примере *т* и *т'* (см. рис. 2.18), даже при одинаковой комбинации сигналов временное их соотношение при артикуляции твердых и мягких не является идентичным. При артикуляции твердых звуков зона контакта к моменту размыкания смычки и образования взрыва сужается — контакт в точке *к.11* размыкается приблизительно одновременно или несколько раньше контакта в точке *к.3*. При артикуляции мягких контакт в точке *к.11* сохраняется и на протяжении взрыва. Следовательно, ширина зоны контакта на протяжении смычки не является фонетически существенной и не регулируется в процессе артикуляции достаточно точно, однако к моменту образования взрыва зона контакта приобретает конфигурацию, характерную именно для данного звука (мягкого или твердого). Сказанное хорошо согласуется с фонетическими данными о том, что информация о мягкости содержится в последней фазе согласного и в переходе к последующему гласному.

Аналогичная ситуация имеет место при артикуляции *ч* и мягких *т'* и *д'*. Как было сказано, в ряде случаев артикуляции *т'* и *д'* имеет место контакт в точке *к.5*, характерной для артикуляции *ч*. Однако к концу размыкания контакта в точке *к.3* создается такая комбинация сигналов, которая характерна именно для данного звука (рис. 2.21).

Сходство между палатограммами разных звуков, наблюдаемое при применении статической палатографии, является в определенной мере артефактом, так как подобная палатограмма является результатом наложения друг на друга зон контакта на протяжении всего произнесения. Если же учесть изменение зоны контакта языка с нёбом во времени, оказывается возможным описать характерные для разных звуков конфигурации палатограмм более однозначно. Очевидно, что фонетически существенным событием в этом случае следует считать не наличие того или иного сигнала (контакта в определенной точке), а осуществление некоторого более сложного явления. Так, признаком мягкости для передне-

язычных оказалось возможным считать наличие *к.11* после замыкания *к.3*.

Метод динамической палатографии дает возможность регистрировать положение языка при артикуляции согласных и гласных

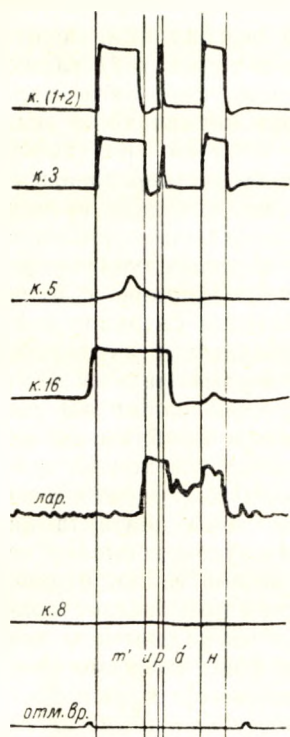


Рис. 2. 21. Осциллограмма слова *тирэн*.

высокого подъема непрерывно в процессе связной речи. Таким образом, объективному исследованию становится доступной временная структура движений на протяжении отдельных звуков, а также характер изменения артикуляции в зависимости от контекста и темпа произнесения. Как показали проведенные измерения, уровень помех в ходе регистрации остается довольно низким, что дает возможность достаточно надежно осуществлять обнаружение сигналов автоматически. Количество одновременно регистрируемых точек может быть довольно большим, однако в ряде экспериментов можно ограничиться сравнительно небольшим количеством точек, по наличию или отсутствию контакта в которых можно достаточно надежно определить возникновение определенных, фонетически существенных явлений. Поскольку применение методики практически не снижает разборчивость речи, методика может быть использована для выделения участков акустической картины речи, соответствующих определенным артикуляторным явлениям. В частности, получаемые в ходе регистрации электрические сигналы могут быть исполь-

зованы для сегментации потока речи на отрезки, строго соответствующие артикуляторным положениям. Полученные таким образом отрезки акустического потока могут быть исследованы в психофизических экспериментах. В ряде случаев для обнаружения некоторых признаков существенной является не наличная комбинация сигналов, а их последовательность во времени. В таких случаях автоматическое устройство, производящее сегментацию, должно осуществлять определенную логическую обработку сигналов.

7. Микрофонный и ларингофонный тракты

Нормальная акустическая картина речи регистрируется с помощью достаточно качественного электродинамического микрофона типа 82-А-2, подключаемого к усилительному тракту, час-

точная характеристика которого равномерна от 30 гц до 15 кгц. Уровни сигналов регулируются с помощью магазинов затухания типа МЗУ.

При необходимости в тракт может вводиться система многозвенных *LC*-фильтров, состоящая из наборов фильтров низких и высоких частот с большой крутизной спада. Граничные точки в обоих наборах равны 125, 250, 500, 750 и 1000 гц; затухание вне полосы пропускания более 70 дб.

Спектральный анализ звукового материала производился с помощью полосового анализатора параллельного действия СЗЧ и динамического спектрографа (типа системы видеографа с параллельным анализом, Сапожков, 1963).

В ряде опытов отдельные третьоктавные фильтры спектрометра звуковых частот СЗЧ включались непосредственно в акустический тракт. Это давало возможность существенно исказить речевой сигнал, сохраняя основную ритмическую структуру фразы (см. далее).

Для оценки динамики движения мягкого нёба, образующего собою носоглоточную перегородку, использовалось иногда применяемое фониатрами (Левидов, 1939) «выслушивание» речевого сигнала через носовые ходы. В этом случае вводимая в носовое отверстие резиновая трубка соединялась с приемной трубкой акустического зонда типа ЗА-4.

Для оценки работы голосовых связок используется контактный микрофон, располагаемый на шее в области гортани. Применение обычных ларингофонов оказалось неудовлетворительным в связи с резким западением их отдачи в области низких частот, т. е. в той области, где располагаются интересующие нас частоты основного тона голоса.

Удовлетворительные результаты получаются с помощью электромагнитного телефона ТГ-7М, на мембрану которого напаян миниатюрный пеллот. Пуговка последнего касается тела испытуемого. Такой ларингофон помещается в мягком резиновом амбушюре и привязывается на шею испытуемому.

Сигналы ларингофона усиливаются высококачественным измерительным усилителем, обычно используемым для электрофизиологических исследований (полоса частот от 3 гц до 20 кгц, уровень шумов порядка единиц микровольт). В тракт могут включаться упоминавшиеся ранее акустические фильтры.

8. Электронные устройства для обработки сигналов о деятельности речеобразующего аппарата

Согласно экспериментальным задачам, сигналы описанных выше датчиков должны одновременно регистрироваться и использоваться для измерений различных артикуляторных показателей. Выходные сигналы большинства датчиков лежат в об-

ласти звуковых частот. Это дает возможность использовать одинаковые каналы для усиления и регистрации различных явлений.

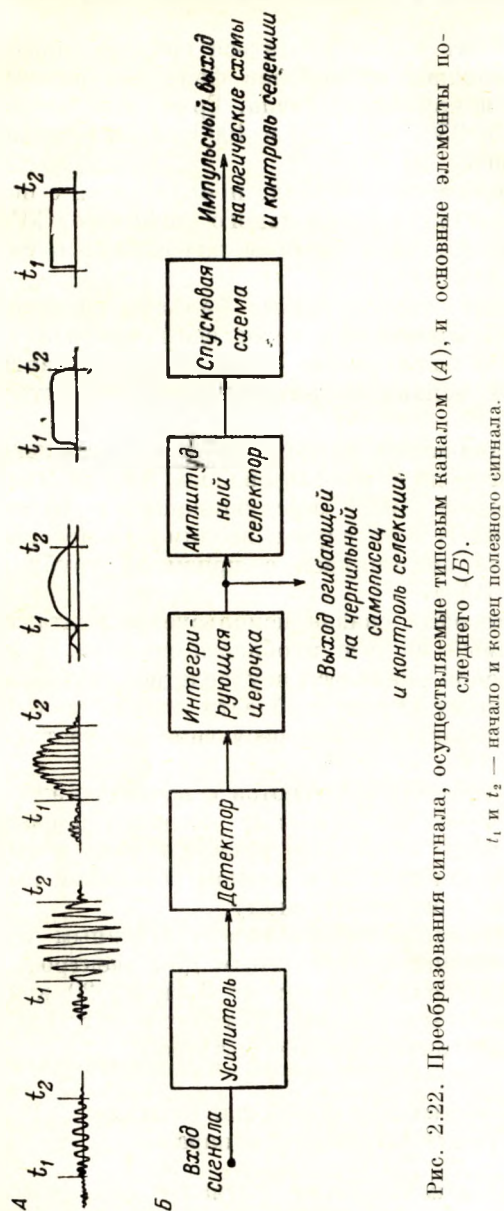


Рис. 2.22. Преобразования сигнала, осуществляемые типовым каналом (А), и основные элементы по-

работы каждого канала (рис. 2.22) сводится к получению двух выходных сигналов: во-первых, огибающей входного переменного напряжения и, во-вторых, прямоугольного импульса, создаваемого формирующей схемой в том случае, если входное напряжение превосходит некоторый заранее устанавливаемый пороговый уровень. Цель последней операции состоит в том, чтобы преобразовать сложное входное напряжение в сигнал двоичной формы, говорящий только о факте наличия или отсутствия измеряемого параметра.

Установка содержит систему унифицированных каналов, логических схем и контрольных устройств. Одновременно могут использоваться 10 и более идентичных каналов. Существенную часть установки составляют регистрирующие устройства, обладающие различными характеристиками.

Схема каждого из каналов установки (рис. 2.23 и 2.24) состоит из следующих последовательных каскадов: предварительного усилителя, фазоинвертера, двухполупериодного детектора, сглаживающей (интегрирующей) цепочки, амплитудного селектора

и выходной спусковой (формирующей прямоугольный импульс) схемы. На вход канала подаются сигналы с датчиков, которые

могут рассматриваться как амплитудно-модулированные колебания, так как датчики питаются переменным напряжением звуковой частоты (как правило 2000 гц). Амплитуда сигнала обычно не превышает 150 мв. Коэффициент усиления предварительного усилителя вместе с фазоинвертером, выбранный в соответствии с условиями работы на линейном участке характеристики, составляет около 300. Полоса пропускемых частот простирается от 20 гц до 20 кГц.

Для того чтобы устранить многократное срабатывание выходной формирующей схемы, необходимо по возможности уменьшить

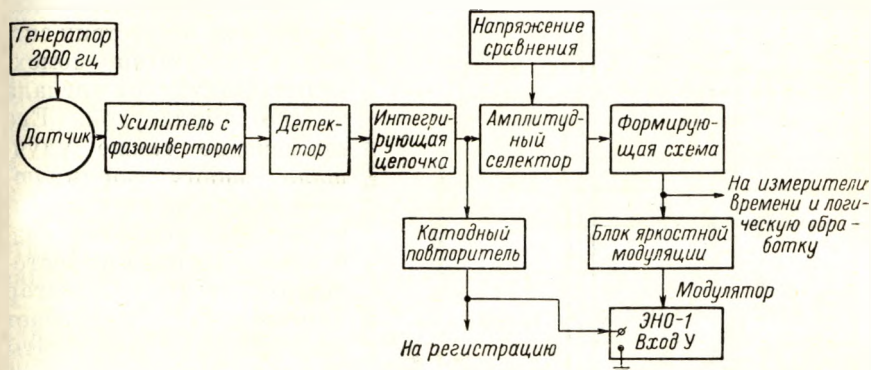


Рис. 2.23. Блок-схема типового усилительного канала и устройства контроля селекции.

коэффициент пульсации несущей частоты на выходе интегрирующей цепочки. Этот коэффициент можно уменьшить, увеличивая несущую частоту или увеличивая постоянную времени интегрирования. Но увеличение несущей частоты влечет за собой также увеличение величины наводок за счет паразитных емкостей между каналами, и особенно между отдельными каналами, питаемыми от датчика «искусственное небо». Увеличение же постоянной времени интегрирования влечет за собой возрастание ошибок при временных измерениях. Применение двухполупериодного детектирования при некотором усложнении схемы снижает коэффициент пульсаций до 2 раз по сравнению с однополупериодным детектированием. Поэтому после предварительного усилителя стоит самобалансирующийся фазоинвертер, обеспечивающий не только хороший коэффициент симметрии, но и существенное усиление. Детектор собран на кристаллических диодах типа Д-2-Ж.

В качестве сглаживающей интегрирующей цепочки применяется однозвенная RC -цепочка с постоянной времени порядка 5—20 мсек. Получаемая на выходе интегрирующей цепочки огибающая сигнала через согласующие делители и катодный повто-

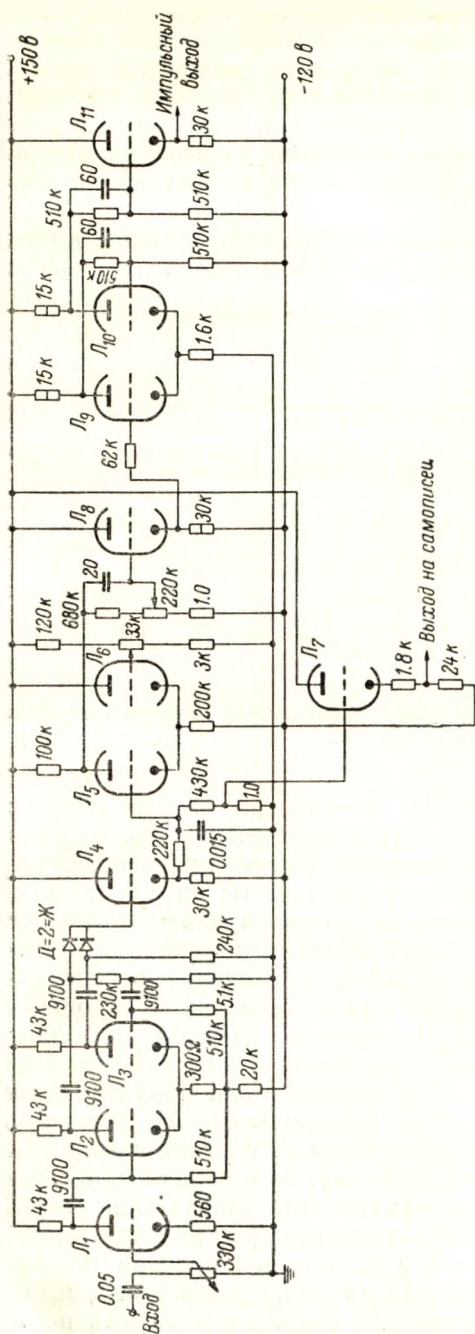


Рис. 2.24. Принципиальная схема типового усилительного канала.

ритель подается по постоянному току на оконечный каскад чернильного самописца, а также поступает на вход амплитудного селектора.

Применение амплитудного селектора было вызвано необходимостью обеспечить автоматическое «принятие решения» о наличии или отсутствии сигнала в присутствии помех, отличающихся от сигнала только амплитудой. Его применение также улучшало работу формирующей схемы за счет дополнительного усиления разностного сигнала по постоянному току. Селектор значительно уменьшает влияние на результаты временных измерений гистерезиса, присущего формирующим спусковым схемам. В качестве амплитудного селектора применен дифференциальный усилитель с катодной связью, на один вход которого подается огибающая сигнала, а на другой — напряжение сравнения, снимаемое с делителя. Каскадом усиливается разность подаваемых на входы напряжений.

Величина напряжения сравнения определяет уровень селекции и может быть в случае необходимости изменена. Практически это напряжение является уровнем, превышение которого передается на последующую формирующую

схему. Окончательное формирование сигнала в прямоугольный импульс производится так называемым триггером Шмидта. Установка уровня селекции играет очень важную роль в автоматическом принятии решения о наличии или отсутствии сигнала, а также в устранении различных помех. Поэтому безусловно необходимым является такой контроль уровня селекции, при котором имелась бы возможность визуально оценивать величину сигнала, помех и уровня селекции в процессе эксперимента.

Блок-схема устройства, позволяющего осуществлять такой контроль, представлена на рис. 2.23. На вход осциллоскопа ЭНО-1 подается огибающая сигнала в регулируемом канале.

Сформированный из этого сигнала прямоугольный положительный импульс подается по постоянному току на сетку мультивибратора, схема которого представлена на рис. 2.25. Все время, пока импульс отсутствует, лампа L_1 оказывается запертой нижним уровнем (-25 в) сигнала. При срабатывании спусковой схемы уровень повышается до $+10$ в, в результате чего мультивибратор начинает генерировать с частотой около 50 кГц. Сигналы с выхода этого мультивибратора подаются в цепь яркостной модуляции осциллоскопа, в результате чего на его экране (рис. 2.26) оказывается подсвеченной та часть огибающей сигнала, которая была сформирована в прямоугольный импульс. Отчетливо оказывается виден сам сигнал с помехами, а также и установленный уровень селекции. Выходные напряжения огибающих во всех каналах, а также сформированные прямоугольные импульсы подаются на переключатели, так что имеется возможность быстро проконтролировать и установить уровень селекции во всех каналах.

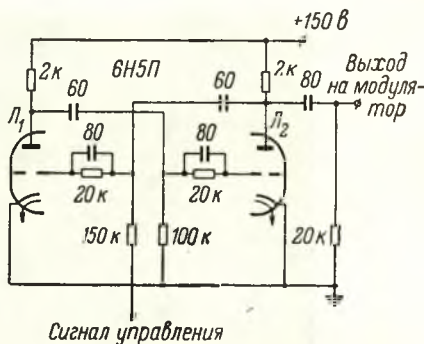


Рис. 2.25. Схема управляемого мультивибратора, используемого для яркостной модуляции луча электронного осциллоскопа в устройстве контроля селекции.

Сформированный выходной схемой канала импульс может быть непосредственно подан в схемы измерения временных интервалов, в логические схемы, а также может быть использован для целей управления регистрирующей аппаратурой и акустическим сепаратором (см. далее).

Проведение целого ряда временных измерений артикуляторных параметров, а также и выделение с помощью акустического сепаратора определенных участков акустической картины речи существенно облегчаются благодаря применению электронных

логических схем. В последние как бы закладываются таблицы, определяющие комплексы исходных простых признаков, характеризующие искомые более сложные явления. Сформированные вышеуказанным способом прямоугольные импульсные сигналы, поступающие на логическую схему, могут иметь три значения для искомого сложного явления: подтверждающее, отрицающее и безразличное. При выполнении условия, заложенного в логической схеме, на ее выходе возникает сигнал, с помощью которого может быть отмечен факт наличия исходного сложного явления,

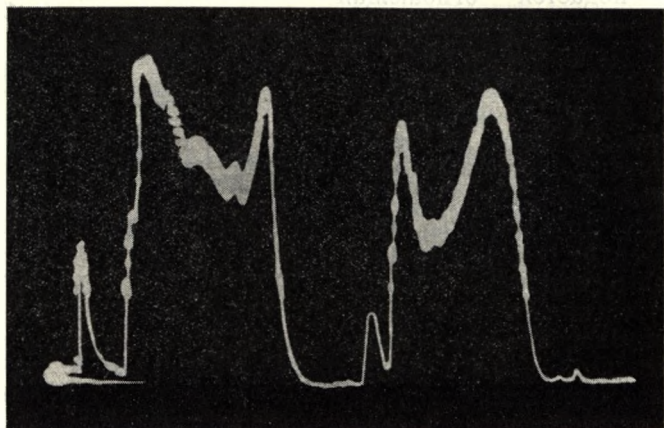


Рис. 2.26. Осциллограмма огибающей сигнала ларингофонного тракта, снятая с экрана электронного осциллоскопа. Увеличение яркости луча отмечает уровень селекции.

а также запущен электронный ключ — сепаратор, через который пройдет определенный отрезок речевого звукового сигнала.

В установке используются функциональные элементы, выполняющие следующие логические и преобразовательные операции:

1) операцию запоминания импульса выполняет триггер с положительным и отрицательным запуском;

2) логическую операцию И и ИЛИ выполняют элементы, построенные на кристаллических диодах;

3) операцию задержки импульса на некоторое наперед заданное время выполняют одностабильные мультивибраторы с анодно-сеточной связью;

4) логическую операцию НЕ выполняют элементы, построенные по схеме однокаскадного усилителя в ключевом режиме;

5) в качестве формирующих схем используются триггеры Шмидта.

Все основные электронные элементы, используемые в установке, выполнены в виде унифицированных одно-, двух- и трехламповых блочных конструкций (рис. 2.27), включаемых с помощью октальных разъемов в общие панели-стенды. На выходе всех функциональных элементов, как правило, имеется катодный

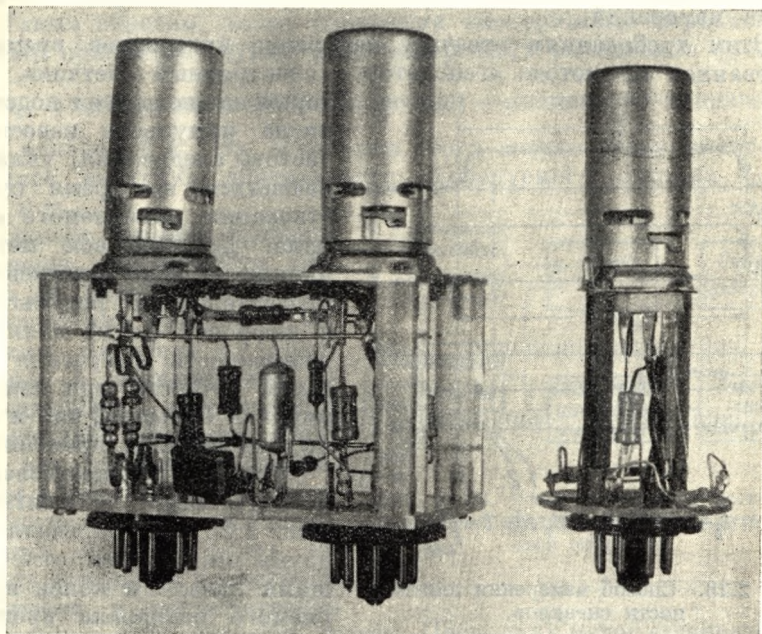


Рис. 2. 27. Внешний вид типовых конструкций блочных функциональных элементов.

повторитель. В результате этого объединение элементов в разнообразные сочетания не ведет к опасности возникновения паразитной генерации и различных наводок.

Питание основных элементов производится от электронных стабилизаторов на $+150$ в, -120 в и -25 в (уровень фиксации двоичных сигналов). Однородный монтаж цепей питания на стендах позволяет легко перестраивать используемые наборы блоков и коммутацию между ними, что дает возможность оперативно изменять характер измерительных задач, решаемых всей установкой.

9. Электронные измерители времени, автоматизирующие обработку данных

В задачи целого ряда экспериментов по изучению деятельности речевого аппарата входит оценка многократных измерений длительностей различных процессов. Наличие разброса величины

повторно измеряемых интервалов времени существенно осложняет обработку получаемых данных. Автоматизация таких измерений может существенно облегчить и ускорить рациональное сжатие огромных объемов информации, получаемых в экспериментах. Автоматизировать измерения помогает использование методов, обеспечивающих непосредственное цифровое выражение результатов измерений.

Этим требованиям отвечает измерение интервалов времени, основанное на методе «генератор — электронный счетчик», ко-

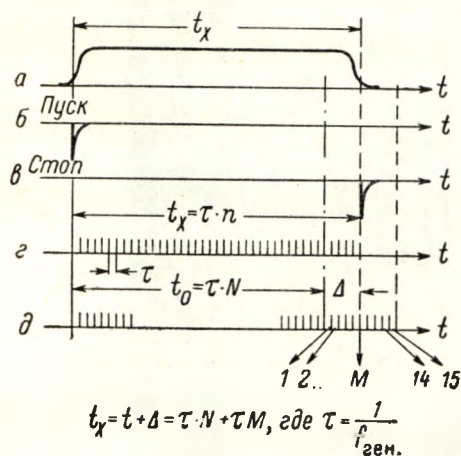


Рис. 2.28. Способ измерения длительности сигналов.

а — измеряемый сигнал; б — импульс, отмечающий начало сигнала; в — импульс, отмечающий конец сигнала; г — тактовые импульсы, укладываемые в интервале времени сигнала; д — фиксирование области флюктуации измеряемого сигнала набором накопительных счетчиков. Остальные объяснения в тексте.

Тактовых импульсов, а n — число импульсов, уложившихся в измеряемом интервале. Таким образом, измерение времени сводится к определению числа n .

Измерение средних величин длительностей сигналов является одной из самых простых и в то же время самых необходимых операций в каждом эксперименте. Устройство, блок-схема которого приведена на рис. 2.29, позволяет существенно упростить обработку данных при определении этих средних.

Импульсы, отмечающие начало и конец измеряемого интервала времени, регулируют работу входного триггера, который, в свою очередь, на время, равное длительности измеряемого интервала, открывает схему И. Отсчет числа тактовых импульсов, уложившихся в измеряемом интервале, производится с помощью

который заключается в подсчете числа импульсов известной частоты следования, укладываемых во время существования измеряемого сигнала. Достоинством метода наряду с непосредственным представлением результатов в цифровой форме является возможность получения высокой точности при измерениях как больших, так и малых интервалов времени.

Время существования исследуемого сигнала (рис. 2.28, а) определяется двумя импульсами, отмечающими начало и конец измеряемого интервала времени (пуск и стоп на рис. 2.28, б и в). Если длительность исследуемого сигнала обозначить через t_x (рис. 2.28, а), то в каждом конкретном случае $t_x = \tau \cdot n$, где τ — период

счетчика суммы импульсов. Для этого используется пересчетный прибор типа ПС-100 (четыре декатронных декады и одна быстродействующая триггерная декада). При повторении измерений показания на пересчетном приборе суммируются. Одновременно в качестве счетчика числа измерений работает электромеханический быстродействующий цифровой индикатор типа СБ-1-100.

Таким образом, после окончания эксперимента пересчетный прибор показывает сумму всех измеренных интервалов времени

$\sum_{i=1}^k n_i$, а электромеханический счетчик — количество произведенных измерений k . На основании этого можно сразу вычислить среднее значение длительности исследуемого сигнала $t_{\text{ср.}} = \frac{1}{k} \sum_{i=1}^k n_i$.

Однако знанием среднего значения длительности сигнала не исчерпывается характеристика измеряемой величины. Значения n , получаемые в эксперименте по определению длительности какого-либо процесса, обычно, как правило, флуктуируют в некотором интервале. Характеристикой этих флуктуаций может служить частота попадания измеряемой величины в соответственно выбранные статистические классы. Поскольку метод «генератор — электронный счетчик» дает результат измерений непосредственно в цифровой форме, то имеется возможность легко получить статистическое распределение длительности исследуемого сигнала с помощью набора цифровых накопительных счетчиков. Естественно принять при этом величину статистического класса равной одному периоду следования тактовых импульсов. Тогда каждое возможное значение n , варьирующее от измерения к измерению, попав в соответствующий статистический класс, фиксируется его цифровым индикатором.

Если каждому из возможных значений n , начиная от 1, приписать свой накопительный индикатор, то уже для обеспечения точности измерений порядка 1% потребуется более 100 запоминающих индикаторов с таким же числом управляющих ими электронных схем. Однако в целом ряде случаев область флуктуации значений n составляет лишь небольшую часть от величины его среднего значения. На основании этого возникает возможность ограничиться значительно меньшим количеством индикаторов, сохраняя в то же время высокую разрешающую способность измерений.

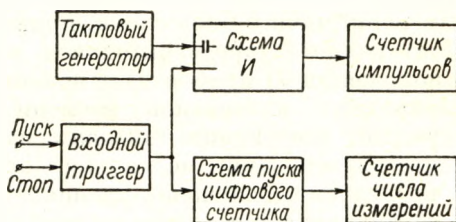


Рис. 2. 29. Блок-схема устройства для измерения средних величин длительностей сигналов.

Этот прием можно осуществить следующим образом: из общего числа импульсов, попадающих в измеряемый интервал времени, как бы вычитается число N , заведомо меньшее минимального из ожидаемых значений n , а запоминающими индикаторами фиксируются лишь флюктуации измеряемого интервала от числа $(N+1)$ до $(N+15)$. Таким образом, при многократном повторении измерений на каждом накопительном цифровом индикаторе регистрируется число случаев, соответствующих некоторому одному статистическому классу. Длительность измеряемого сигнала, соответствующая каждому конкретному классу, равна $t_x = \tau N + \tau M$ (рис. 2.28).

По принципу, описанному выше, был разработан и изготовлен электронный измеритель времени, автоматизирующий статистическую обработку данных. Проследим процесс производства измерений по его блок-схеме, приведенной на рис. 2.30.

В режиме «предварительное измерение» (переключатель Π_1 в положении 1) прибор дает возможность производить измерения конкретных интервалов времени с точностью, определяемой периодом следования тактовых импульсов. В каждом эксперименте на основании нескольких отсчетов числа n , уложившихся в исследуемом сигнале, оценивается его средняя длительность и устанавливается значение числа N .

Триггерный электронный ключ $Tr.к.1$ открывается на время t_x импульсами, соответствующими началу и концу измеряемого интервала времени. Число тактовых импульсов n , заполнивших исследуемый интервал, подсчитывается электронным счетчиком $СЧ-1$ и фиксируется неоновыми лампочками. Перед началом каждого измерения производится ручной сброс показаний предыдущего измерения.

В режиме «измерение статистического распределения» электронный измеритель времени производит распределение измеряемых сигналов по соответствующим статистическим классам. Цикл работы прибора состоит в следующем (рис. 2.30, переключатель Π_1 в положении 3). Пусковым импульсом, отмечающим начало измеряемого сигнала, открывается триггерный ключ $Tr.к.1$. Счетчик $СЧ-1$ производит отсчет числа тактовых импульсов согласно предварительно установленному на нем числу тактов N , после чего через диодный ключ $Д.К.2$ и триггерный ключ $Tr.к.2$ автоматически включается счетчик $СЧ-2$. Последний на отрезке времени τ через дешифратор-распределитель последовательно один за другим подключает к схеме совпадений накопительные индикаторы. Импульс, соответствующий концу измеряемого сигнала (поступивший на клемму *стоп* на рис. 2.30), фиксируется на том индикаторе, в течение времени подключения которого он пришел.

Электронная схема «сброса» автоматически возвращает счетчики $СЧ-1$ и $СЧ-2$ после окончания измерения в исходное состоя-

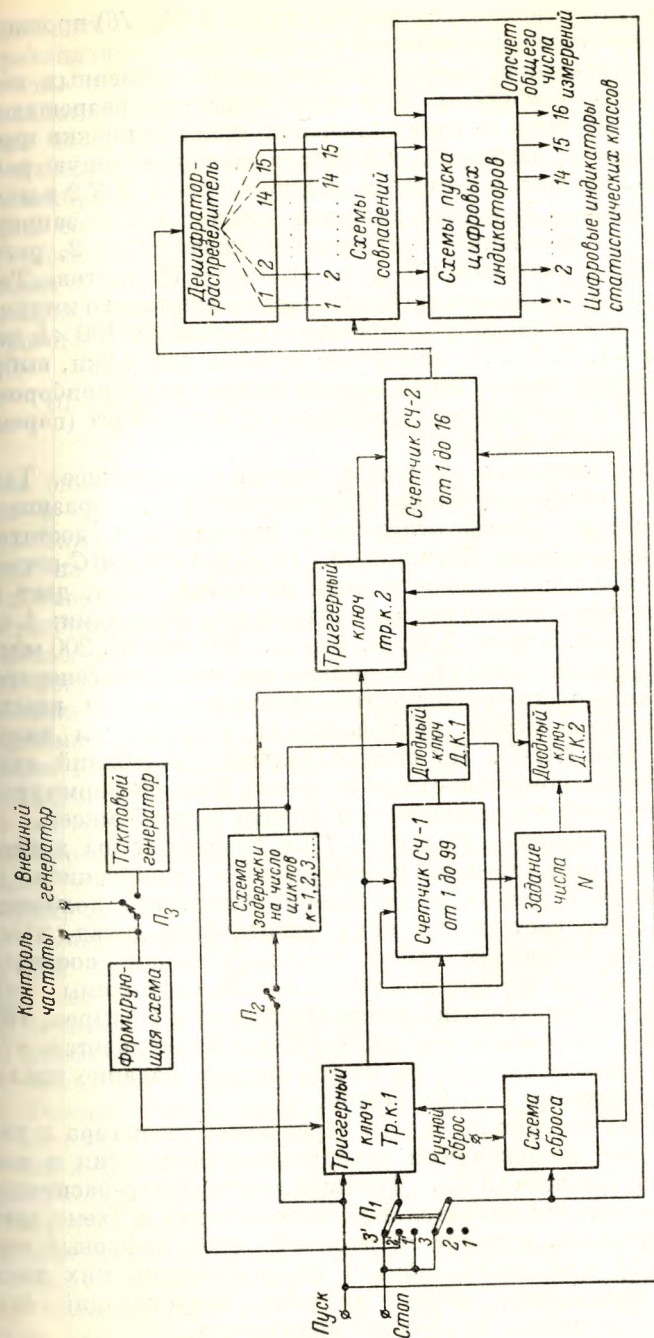


Рис. 2.30. Блок-схема электронного измерителя времени, автоматизирующего статистическую обработку данных.

ние. Отдельный цифровой индикатор (рис. 2.30, 16) производит подсчет общего числа измеренных интервалов.

При статистической оценке длительностей временных интервалов с малым разбросом возможно повысить разрешающую способность измерения за счет включения схемы задержки циклов (переключатель P_2 замкнут), обеспечивающей повторную работу счетчика $СЧ-1$ за счет перекрытия диодного ключа $Д.К.2$ в момент окончания цикла. Запертый диодный ключ $Д.К.2$ запрещает соответственно включение триггерного ключа $Тр.к.2$, регулирующего работу счетчика $СЧ-2$, на $(100 \times k + N)$ тактов. Таким образом, от схемы задержки требуется выработка такого импульса, длительность которого лишь несколько превышает $100 \times k$ число тактов (где $k=1, 2, 3, \dots$). Измерение времени задержки, выбранной в конкретном опыте, легко производится самим прибором на счетчике $СЧ-1$ в режиме «предварительное измерение» (переключатель P_1 в положении 2).

Устройство основных элементов прибора следующее. Тактовый генератор выполнен в виде RC -генератора с Γ -образным четырехполосником. Схема отличается простотой и достаточно высокой стабильностью (Криксунов, 1958). Набор RC -цепочек, подключаемых в цепь положительной обратной связи, дает возможность получать колебания со следующими периодами: 1 сек., 100 мсек., 10 мсек., 5 мсек., 2 мсек., 1 мсек., 500 мксек., 200 мксек., 100 мксек. Предусмотрено также включение внешнего генератора, что дает возможность производить измерения и при иных t . Контроль частоты может производиться измерителем частоты (например, типа ИЧ-7). Из синусоидальных колебаний задающего генератора с помощью схемы триггера Шмидта формируются тактовые импульсы с длительностью фронта менее 5 мсек.

Триггерные ключи $Тр.к.1$ и $Тр.к.2$ состоят из диодного ключа-вентилля, открываемого и запираемого напряжением, снимаемым с анода триггера и характеризующим его положения.

Счетчик $СЧ-1$ состоит из двух десятичных декад, каждая из которых представляет собой последовательное соединение двоичной триггерной пересчетной схемы и кольцевой схемы с пятью состояниями устойчивого равновесия (Петрович и Козырев, 1954). Выбор любого показания счетчика $СЧ-1$ обеспечивается с помощью двух переключателей на 10 положений. Задание числа N может варьировать от 1 до 99 тактов.

Счетчик $СЧ-2$ состоит из одного пересчетного триггера и кольцевой схемы с 8 положениями устойчивого равновесия и имеет коэффициент пересчета 2×8 . Диодный дешифратор-распределитель обеспечивает последовательное подключение к схеме совпадений (на интервалы времени t) пусковых схем цифровых индикаторов. Схема совпадений собрана на кристаллических диодах и в случае совпадения выдает импульс, запускающий схему пуска соответствующего цифрового индикатора.

Накопительные цифровые индикаторы представляют собой четырехзначные электромагнитные счетчики типа РС-2, которые включены в аноды реостатно-емкостной спусковой схемы одностабильного мультивибратора. Работа их контролируется загоранием неоновых лампочек.

В качестве схемы задержки циклов применен одностабильный мультивибратор, длительность посылок которого регулируется набором сопротивлений и конденсаторов с плавным перекрытием от 10 мсек. до 1 сек.

Схема сброса производит возвращение всех элементов прибора в исходное положение через 5 мсек. после прихода импульса, отмечающего конец исследуемого сигнала.

10. Комплексная регистрация артикуляторных и акустических параметров речи

Все описанные выше датчики и электронные устройства собраны в единую экспериментальную установку, позволяющую осуществлять комплексную регистрацию основных артикуляторных параметров речи наряду с ее акустической картиной. Как показывает блок-схема рис. 2.31, все узлы установки могут быть разбиты на следующие группы устройств: датчики, электронные схемы для усиления и предварительной обработки сигналов, схемы логической обработки информации, регистраторы и автоматические измерители (Кожевников и Шупляков, 1962).

Система применяющихся датчиков дает возможность одновременно и непрерывно получать в виде электрических сигналов следующие показатели:

- 1) внутриротовое давление;
- 2) скорость движения воздуха из ротового отверстия и из носовых отверстий (используются малоинерционные электроанемометры);
- 3) толчки воздуха, выходящего из ротового и из носовых отверстий при фонации (используются фотоэлектрические датчики с амбушюрами);
- 4) дыхательные движения на разных уровнях груди и живота;
- 5) движения нижней челюсти;
- 6) смыкание и размыкание губ;
- 7) явление огубленности (электроплетизмографическое измерение выпячивания и иных движений губ);
- 8) динамическую палатограмму (возможна одновременная регистрация динамики касания языком 10 и более отдельных зон твердого нёба);
- 9) колебания голосовых связок (отмечаемое с помощью ларингофона наличие голоса при фонации);
- 10) звуковую картину речи при отведении сигнала через носовые отверстия;

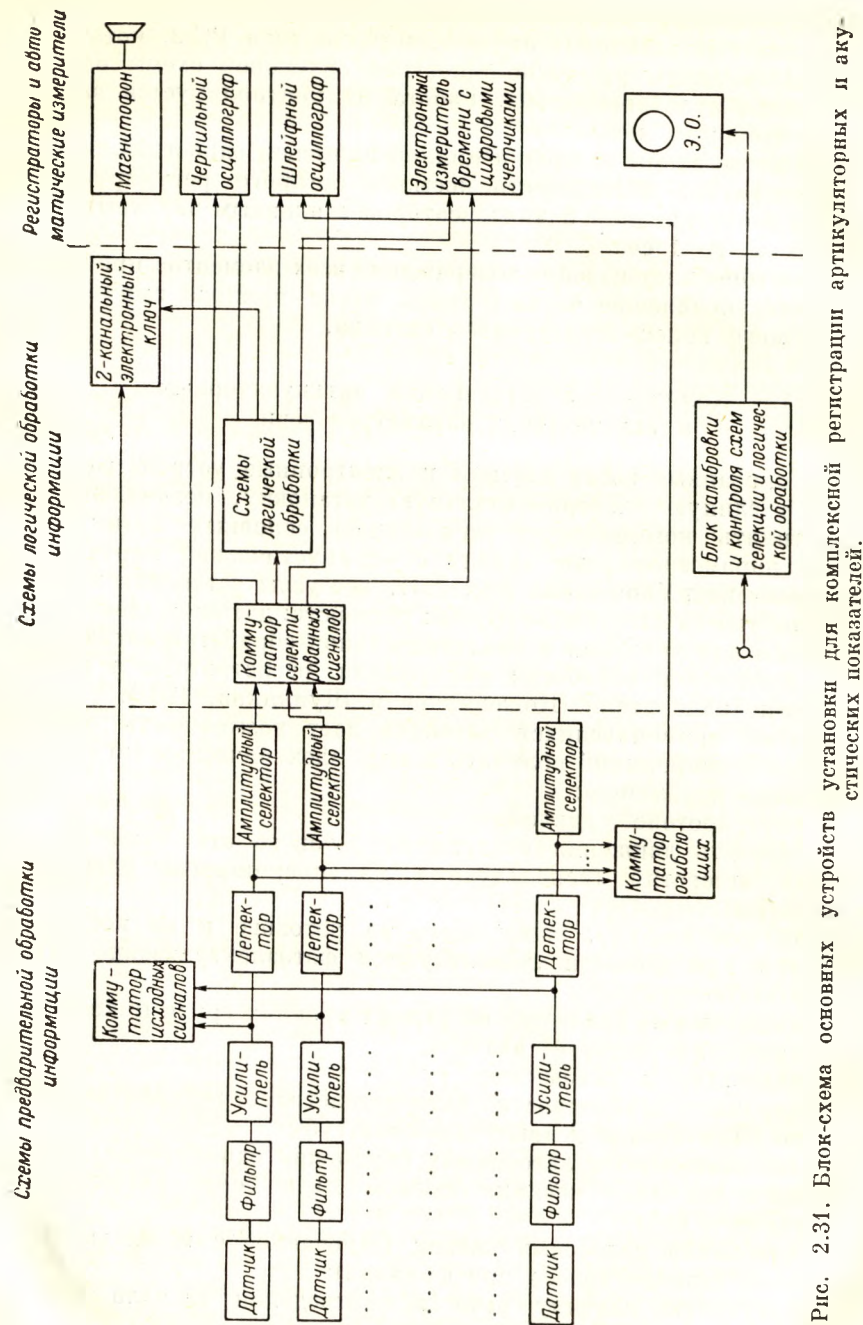


Рис. 2.31. Блок-схема основных устройств установки для комплексной регистрации артикуляторных и акустических показателей.

11) нормальную звуковую картину речи.

Фотография испытуемой с основными датчиками приводится на рис. 2.32.

Помимо этих показателей, возможна также регистрация ЭМГ различных мышц, производимая с помощью типичных электрофизиологических усилителей используемого в установке 16-ка-



Рис. 2.32. Испытуемая с основными датчиками артикуляторных параметров.

нального электроэнцефалографа. В связи с высокой симметрией усилителей дополнительной экранировки испытуемого при этом не требуется, но ЭМГ невозможно регистрировать одновременно с динамической палатографией в связи с подачей в данном случае напряжения звуковой частоты на электроды, которых касается язык. Впрочем, ценою некоторого усложнения электронных схем может быть преодолено и это ограничение.

Одновременно с записью электрических сигналов возможно вести и кинематографическую съемку лица или фотоосциллографи-

ческую регистрацию движений губ. Сопоставление ЭМГ и результатов съемки легко осуществлять на основании синхронизирующих отметок.

Электрические сигналы, отражающие упомянутые выше показатели, разнятся друг от друга в весьма большой степени. Так, сигналы о дыхательных движениях имеют в своем частотном составе компоненты, начинающиеся от малых долей герца, звуковая же картина речи практически простирается до 8—10 кГц.

Необходимость достаточно точной одновременной регистрации всего комплекса сигналов и при этом получения весьма значительных объемов экспериментальных данных (последнее, естественно, вытекает из статистического характера исследований) ставит сложные и противоречивые требования к технике измерения и регистрации информации.

В установку включены регистрирующие устройства, существенно отличающиеся своими характеристиками. Центральным регистрирующим прибором является 16-канальный чернильный самописец — электроэнцефалограф датской фирмы «Кайзер». Сигналы, усиленные, выпрямленные и сглаженные электронными схемами предварительной обработки информации, подаются на оконечные усилители самописца. Полоса регистрируемых частот при этом простирается от 0 до 80—90 Гц. На бумажной ленте, которая может двигаться со скоростями от 5 мм в сек. до 120 мм в сек., одновременно регистрируются до 17 процессов (включая отметку времени). На этом приборе с большим удобством регистрируются основные показатели артикуляторной динамики — сигналы динамической палатографии, сигналы о смыкании губ, движениях воздуха и др.

Вместе со всем комплексом артикуляторных показателей на общей ленте записываются и акустические сигналы в виде огибающих. Если же имеется необходимость регистрации исходных осциллограмм звуковых сигналов и точного сопоставления с ними артикуляторных параметров, то тогда дополнительно может производиться регистрация с помощью 8-шлейфного осциллографа Н-102. Осциллограф снабжен комплектом специально разработанных оконечных усилителей тока. 4 усилителя постоянного тока позволяют регистрировать сигналы на чувствительных, но относительно низкочастотных шлейфах (IV или V тип). Регистрация же акустических явлений производится с помощью 4 усилителей переменного тока. Усилители имеют трансформаторные выходы, с помощью которых легко обеспечивается получение достаточно больших токов через низкоомные петли шлейфных гальванометров. В усилителях введена частотная коррекция, в результате при использовании шлейфов II типа регистрируются частоты от 50 Гц до 10 кГц.

Использование многоканальных осциллографов удобно с точки зрения возможности комплексной регистрации многих показате-

телей одновременно на одной ленте. Однако точность измерения амплитуд (динамический диапазон регистратора) при этом получается довольно ограниченной. В особенной степени это касается 16-канального чернильного самописца. В связи с этим в установке предусматривается параллельно с основной осциллограммой возможность дополнительной записи отдельных процессов на регистраторах с большим динамическим диапазоном.

Так, для получения точной оценки относительных громкостей звуковых сигналов используется обладающий логарифмической шкалой самописец уровня электрических колебаний типа Н-110. Такой же низкочастотный процесс, как дыхательные движения, удобно регистрировать на электронном потенциометре ЭПП-09. Записи, получаемые на его широкой шкале (270 мм), отличаются большой точностью и содержанием мелких подробностей пнеймограммы.

Акустический речевой материал регистрируется с помощью качественных магнитофонов МЭЗ-28. При этом, помимо записи полной звуковой картины, может использоваться также и довольно широко применяемый акустиками и фонетиками метод «временной сепарации» (Дукельский, 1962). Метод сводится к тому, что с помощью электронного ключа (или сепаратора) выделяются некоторые определенные отрезки исходного акустического сигнала, которые далее могут анализироваться и использоваться в психофизических экспериментах.

Существенным требованием к ключам, используемым при исследовании акустических сигналов, является обеспечение достаточно большого динамического диапазона, определяемого в основном разницей между уровнем сигналов, неискаженно передаваемых ключом, и уровнем помех, связанных с работой самого переключающего устройства.

Как правило, подобные звуковые послышки создаются с помощью электронных переключателей, основным элементом которых служит запертая усилительная лампа, отпираемая на требующийся интервал времени и пропускающая электрический сигнал. Существенные осложнения возникают в связи с резким изменением — броском анодного тока лампы при ее отпирании и запираании. Обычным путем ослабления возникающего при этом так называемого щелчка включения является применение симметричных балансных схем. В таких схемах симметричные сигналы, в том числе и изменения анодного тока, одновременно происходящие в лампах обоих плеч, в значительной мере подавляются. Однако отношение между полезным сигналом и щелчком включения не удается получить большим 30—35 дб.

Специальными мерами — применением подбираемых высокоточных деталей и сложной балансировкой схем — это отношение можно дополнительно увеличить на 10—20 дб., однако и в этом

подключался магазин затухания МЗУ, напряжение с которого подавалось на динамический телефон ТД-6. Введением соответствующих затуханий определялись на слух пороговые уровни для тональной посылки, неискаженно передаваемой ключом, и для щелчка включения, имеющего место при отключенном генераторе. При крутых фронтах звуковой посылки разность этих уровней оказалась примерно равной 60 дБ. При увеличении постоянной времени нарастания фронта посылки эта разность увеличивается и при R_1C_1 , равной 14 мсек., превосходит 80 дБ.

Импульсы управления электронным ключом берутся от схем предварительной обработки сигналов или от логических устройств. В первом случае это дает возможность выделять участки акустической картины, по времени соответствующие наличию определенного артикуляторного параметра. Во втором же случае выделяются участки, в течение которых весь комплекс сигналов от датчиков удовлетворяет условиям, заложенным в логической схеме.

Возможность выделения схемами предварительной и логической обработки из всего потока сигналов определенных участков, удовлетворяющих некоторым задаваемым условиям, позволяет также автоматизировать ряд измерений временных свойств артикуляторной картины речи. Разработанные для этого электронные измерители с непосредственным цифровым отсчетом описаны выше.

Оценивая в целом основные возможности установки, можно сказать, что разработанные датчики и электронные устройства позволяют:

- 1) изучать динамику и соотношения основных показателей артикуляторной и акустической картин речи;

- 2) с помощью чернильной регистрации и цифровых электронных измерителей легко получать значительный экспериментальный материал, пригодный для дальнейшей статистической обработки;

- 3) со значительной точностью измерять длительности отдельных артикуляторных явлений, а также временные характеристики их взаимоотношений между собой и с акустическими сигналами;

- 4) обеспечивать выделение из акустического потока речи отдельных участков, строго соответствующих по времени определенным артикуляторным явлениям.

11. Основные возможности использования комплексной регистрации артикуляторных и акустических параметров речи

Основной задачей при использовании комплексной регистрации артикуляторных движений было получение возможности членения непрерывного речевого потока на дискретные элементы, соответствующие произносимым звукам, в первую очередь согласным. Известно, что различение согласных фонем в русском языке

основывается на следующих противопоставлениях (Богородицкий, 1930; Матусевич, 1948; Зиндер, 1960):

- 1) по месту образования: губные (губно-зубные), переднеязычные, среднеязычные, заднеязычные;
- 2) по способу образования: смычные, щелевые, дрожащие;
- 3) по участию голоса: глухие, звонкие;
- 4) по участию носового резонатора: чистые, носовые;
- 5) по положению средней части спинки языка по отношению к твердому нёбу: твердые, мягкие.

Любой согласный звук русского языка может быть охарактеризован перечисленными признаками. Естественно, что выделение согласных в потоке речи возможно лишь с их учетом. Описываемая система регистрации предусматривает в первую очередь именно эти характеристики и может быть использована для членения речевого потока на отрезки, соответствующие фонемам. В большинстве исследований, содержащихся в данной книге, как раз и используются эти необходимые признаки.

Рассмотрим соотношение между перечисленными признаками согласных и теми параметрами, которые записываются комплексной системой регистрации артикуляторных движений.

Место образования согласных определяется по сигналам ряда датчиков. Губные согласные регистрируются при помощи контактного датчика, обеспечивающего возможность точно определять момент смыкания и размыкания губ. Устройство датчика дает возможность регистрировать даже сравнительно слабые движения губ, когда полного закрывания ротовой щели нет, например при произнесении щелевых (губно-зубных) согласных.

Язычные согласные регистрируются при помощи датчика «искусственное нёбо», расположение контактов на котором делает возможным различение переднеязычных, среднеязычных и заднеязычных. Образцы записи переднеязычных согласных *т*, *л* приводятся на рис. 2.17 и 2.21, на которых видно, что начало и конец переднеязычных согласных на осциллограммах определяются моментами включения и выключения сигналов *к.1+2* и *к.3*.

Среднеязычные согласные, характеризующиеся более отодвинутой назад по сравнению с переднеязычными зоной касания языка с нёбом, могут быть выделены по признаку наличия сигнала на линии *к.16* и *к.19* при отсутствии сигнала на линии *к.3*.

Нужно особенно отметить, что в группу среднеязычных при этом попадают согласные *ш*, *ж*, *ч*, артикуляция которых обычно описывается как переднеязычная; это объясняется тем, что контакт передней части языка с нёбом при их образовании происходит в той же зоне твердого нёба, что и при образовании среднеязычного *й*. Разницу между регистрацией переднеязычных и среднеязычных демонстрируют рисунки 2.34 и 2.35.

На рис. 2.35 приведена осциллограмма слова *село*; во время произнесения *с'* регистрируется наличие касания языка с нёбом

в точке, характерной для образования переднеязычных (к.3). На рис. 2.35 осциллограмма слова *дебош* показывает, что при произнесении *ш* контакта между языком и нёбом в этой области нет, а имеется контакт только в более задних областях твердого неба (к.16 и к.19).

Учитывая все сказанное, следует сделать вывод, что комплексная запись может дать сведения о месте артикуляции следующих

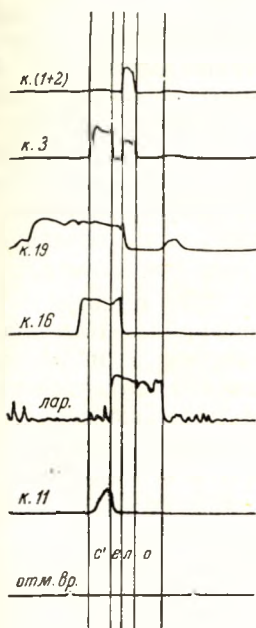


Рис. 2.34. Осциллограмма слова *село*.

Объяснения в тексте.

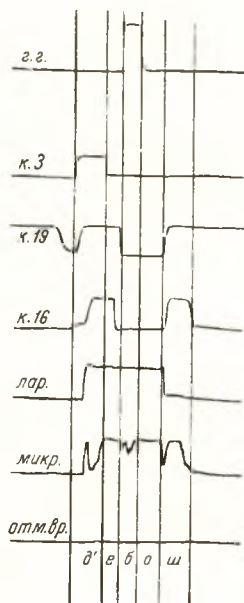


Рис. 2.35. Осциллограмма слова *дебош*.

Объяснения в тексте.

групп согласных: губных согласных *п, б, м, в, ф* и соответствующих мягких; переднеязычных согласных *т, д, н, с, з, л, р* и соответствующих мягких, а также *ц*; среднеязычных согласных *ж, ш, ж, ч*.

О способе образования согласных (т. е. о том, образуются ли они в результате полной смычки губ или языка с нёбом, или же артикулирующие органы образуют щель) в ряде случаев можно судить по расположению регистрируемых контактов: так, при образовании смычных переднеязычных обязательно имеется сигнал в точке к.(1+2), а при образовании соответствующих щелевых этого сигнала нет. Разница между смычными и щелевыми губными согласными также может быть зарегистрирована. Если нужно при записи отличать губные смычные от щелевых, то

контактный датчик, регистрирующий смыкание губ, сдвигают от середины нижней губы вправо или влево; в таком случае сигнал с этого датчика появляется только при артикуляции смычных согласных.

Когда этих сведений оказывается недостаточно для того, чтобы определить согласный по способу образования (например, при артикуляции *л* всегда регистрируется сигнал в точке *к.(1+2)*, так как кончик языка контактирует с передней частью твердого нёба, по щель все-таки имеется по бокам от места смычки), исчерпывающие сведения дают показания датчиков потоков воздуха из ротового отверстия (рупорный датчик и датчик малоинерционного термоанемометра). Отсутствие сигнала такого датчика свидетельствует о смычном характере согласного.

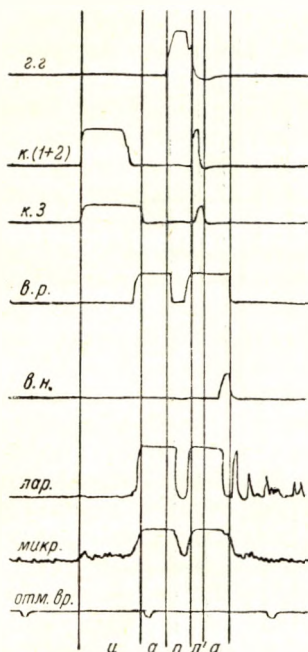


Рис. 2.36. Осциллограмма слова *цапля*.

Объяснения в тексте.

На рис. 2.36 приведена осциллограмма слова *цапля*. Как видно из рисунка, согласные *ц* и *л'* характеризуются наличием контактов языка с нёбом в одной и той же зоне; о щелевом характере *л'* можно судить по тому, что с самого начала его артикуляции регистрируется ток воздуха через рот.

Такие согласные, артикуляция которых включает и смычку, и щель (аффрикаты), регистрируются следующим образом: с начала регистрации контакта в точке, характерной для данного согласного, сигнала о наличии воздушной струи нет; здесь, хотя еще отмечается наличие контакта, начинает поступать сигнал датчика потоков воздуха из ротового отверстия; момент включения

этого сигнала определяет конец смычки и начало щелевой фазы аффрикаты. На том же рис. 2.36 регистрируется таким образом аффриката *ц*.

Имплозивный характер согласного (т. е. отсутствие резкого выхода воздуха изо рта при размыкании смычки неносовых согласных) определяется на осциллограммах по отсутствию показаний датчиков потоков воздуха изо рта после окончания смыкания губ или контакта языка с нёбом. На рис 2.37 приводятся осциллограммы взрывного и имплозивного *т* в слове *ат*. При артикуляции взрывного *т* после размыкания *к.(1+2)* кривая *в.р.* резко поднимается вверх, сигнализируя о выходе воздуха изо рта; при

артикуляции импловзивного сразу после размыкания *к.(1+2)* воздух через рот не проходит.

Имеющие специфический способ образования дрожащие согласные характеризуются многократным смыканием и размыканием контакта. Видно, что на рис. 2.38 дрожащий характер согласного *р* в слове *сырой* вызывает двухкратное включение и выключение сигнала *к.(1+2)* и *к.3* за время артикуляции этого звука.

Таким образом, фонетическая классификация по способу образования однозначно соотносится с показаниями следующих

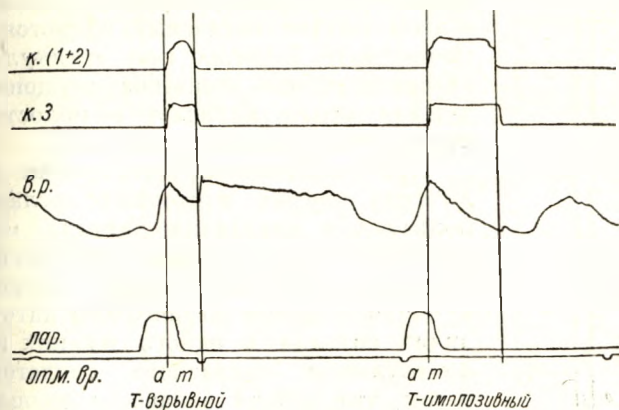


Рис. 2.37. Осциллограмма слога *ат* со взрывным и импловзивным *т*.

Объяснения в тексте.

датчиков: датчика, регистрирующего смыкание губ, датчика «искусственное нёбо» и датчиков воздушных потоков.

Участие голоса, т. е. работа голосовых связок при артикуляции согласных, регистрируется при помощи ларингофона. Это дает возможность не только определять глухость или звонкость согласных, но и получать сведения о начале и конце работы голосовых связок. В связи с возможностью параллельной записи других артикуляторных движений это приобретает большое значение, особенно при изучении координации движений в процессе образования речи.

В качестве иллюстрации новых возможностей приведем частный пример относительно временной последовательности начала смычки и включения голоса при образовании звонких смычных согласных. Традиционное описание звонких смычных основывается на представлении о том, что начало смычки и включение голоса по времени совпадают; поэтому при измерениях длительности звонких смычных обычно измеряют тот участок согласного, где работают голосовые связки. Однако действительные отноше-

ния гораздо более сложны: для начальных звонких согласных характерно, что контакт языка с нёбом *к.1* начинается гораздо раньше, чем включаются голосовые связки (рис. 2.39, *лар.*, слово *Даня*).

Комплексная запись дает возможность гораздо более точно, чем все существующие методы регистрации, судить об особенностях звонкости и глухости согласных.

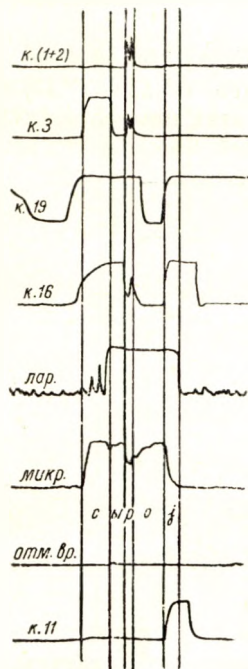


Рис. 2.38. Осциллограмма слова *сырой*.

Объяснения в тексте.

Участие носового резонатора определяет назализованность звука. При произнесении согласного нёбная занавеска может быть поднята и закрывать проход в полость носа; в этом случае согласный является чистым (неносовым). Если же при артикуляции согласного нёбная занавеска опущена, то воздушная струя попадает в полость носа и образуются носовые согласные.

При комплексной регистрации для различения чистых и носовых согласных используются показания датчика малоинерционного термоанемометра для регистрации токов воздуха через носовые отверстия или рупорного фотоэлектрического датчика. Наличие сигнала с одного из них во время артикуляции согласного свидетельствует о том, что нёбная занавеска опущена, воздушная струя проходит через носовую полость, согласный является носовым. На рис. 2.40 приведена осциллограмма слова *налим*, где имеются два носовых согласных — *н* и *м*, во время артикуляции которых отмечается сигнал с рупорного фотоэлектрического датчика (*в.н. руп.*).

Положение средней части спинки языка по отношению к твердому нёбу для русских согласных является весьма существенным,

так как на этом основано различие между твердыми и мягкими согласными. Датчик «искусственное нёбо» фиксирует эти дополнительные движения контактами в соответствующей части нёба. На рис. 2.41 приведены осциллограммы слогов *л'а*, *л'у*, *л'и*, *ла*, *лу*; при произнесении первых трех слогов имеется контакт языка с твердым нёбом, о чем можно судить по наличию сигнала *к.11*, которого нет при произнесении соответствующих твердых.

Все сказанное дает основания утверждать, что система комплексной регистрации артикуляторных движений позволяет произвести членение записанного речевого потока по всем признакам, необходимым для характеристики русских согласных звуков. Существуют однозначные отношения между показаниями приме-

няющихся датчиков и принципами фонетической классификации звуков речи.

Однако возможности такой системы регистрации этим не ограничиваются. Из сказанного ранее очевидно, что, регистрируя основные признаки артикуляции, мы можем параллельно наблюдать и такие явления, как модификация звуков в потоке речи, так как эта модификация часто является следствием или запаздывания, или более раннего включения некоторых артикуляторных

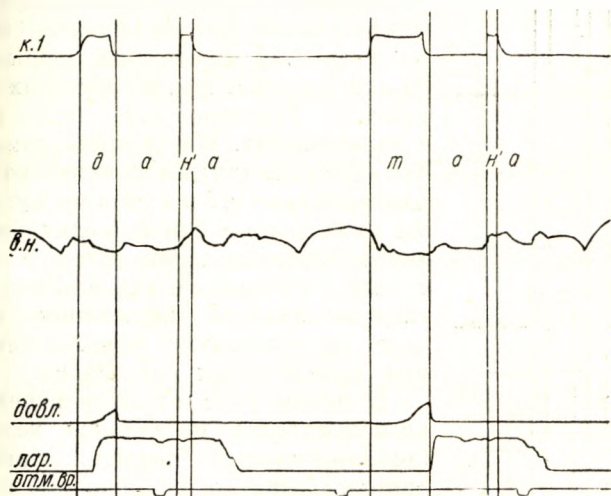


Рис. 2.39. Осциллограмма слов Дая и Таня.

Объяснения в тексте.

движений, характерных для соседних звуков. Кроме того, описываемая методика регистрации дает возможность фиксировать такие артикуляторные явления, которые не являются необходимыми для различения звуков речи, но имеют большое значение при их фонетической характеристике и в фонетической литературе собственно и называются дополнительными артикуляциями. Их изучение имеет большое значение при исследовании вопросов организации последовательности артикуляторных движений.

Использование комплексной регистрации артикуляторных движений позволяет одновременно с перечисленными уже признаками записывать следующие особенности артикуляции:

- 1) изменение артикуляции согласного в зависимости от качества окружающих гласных;
- 2) аспирацию и аффрикатизацию согласных;
- 3) озвончение и оглушение согласных в потоке речи;
- 4) назализацию чистых гласных и согласных в соседстве с носовыми;

5) огубленность согласных в соседстве с огубленными гласными.

Рассмотрим более подробно эти возможности.

Изменение артикуляции согласного в зависимости от качества окружающих гласных наблюдается довольно часто. Известно, что место образования согласного может существенно сдвигаться: перед гласными переднего ряда быть более передним, перед гласными заднего ряда — более задним. При комплексной регистрации

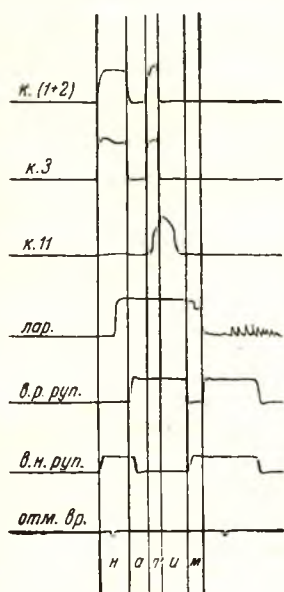


Рис. 2.40. Осциллограмма слова *налим*.

Объяснения в тексте.

параметров такие явления могут быть отмечены порядком включения контактов: на рис. 2.42 приведены осциллограммы слогов *та*, *та*, *ту*, на которых регистрируются отдельно к.1, к.2 и сигнал с ларингофона. На осциллограмме видно, что при артикуляции *та* включаются почти одновременно к.1 и к.2; при артикуляции *та* включение к. 2 происходит с некоторым запаздыванием, а при артикуляции *ту* — с еще большим запаздыванием. Эта последовательность включения контактов дает представление о положении языка при артикуляции согласного.

В целом ряде случаев может быть зарегистрировано и явление веляризации (дополнительный подъем задней части спинки языка по направлению к мягкому нёбу). При этом возникают контакты языка с задней областью твердого нёба.

Аспирация и аффрикатизация согласных возникает в потоке речи в определенных фонетических условиях. Согласные аспирируются, если после размыкания смычки гласный, следующий за согласным, начинается не сразу, а ему предшествует

глухой выдох, длительность которого может быть различной. При регистрации артикуляторных движений описываемыми здесь датчиками это явление может быть охарактеризовано как некая задержка между размыканием смычки и включением сигнала с ларингофона, превышающая по длительности обычный интервал, характерный для взрывных непридыхательных согласных (рис. 2, 4), появление слабого сигнала с ларингофонного датчика во время артикуляции глухого согласного может свидетельствовать о том, что шум придыхания образуется именно в области носоглотки.

Аффрикатизация согласных, в результате которой смычный согласный приобретает в конце звучания щелевой характер, также может быть зарегистрирована: на рис. 2.43 видно, что при произнесении *а́тя* и *а́ца* после размыкания к.(1+2) все еще со-

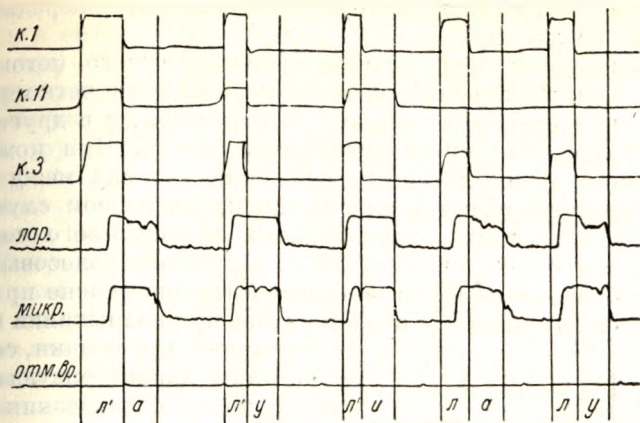


Рис. 2.41. Осциллограммы слогов с мягким и твердым л.
Объяснения в тексте.

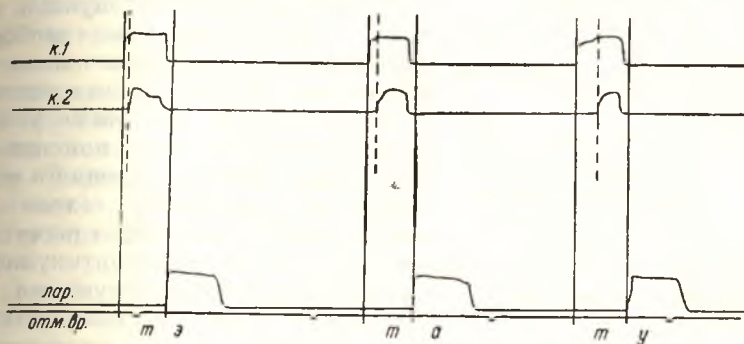


Рис. 2.42. Осциллограммы слогов тэ, та, ту.
Объяснения в тексте.

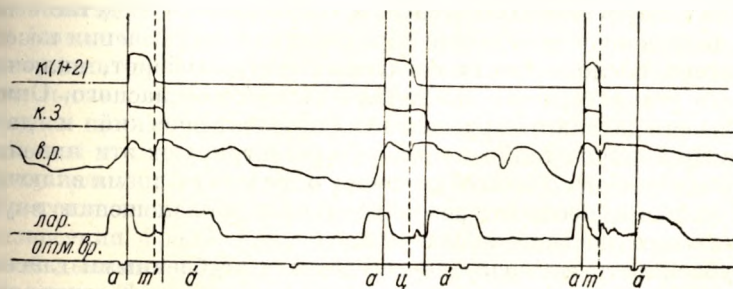


Рис. 2.43. Осциллограммы слогов с т, ц и придыхательным т.
Объяснения в тексте.

храняется $\kappa.3$, наличие которого характеризует переднеязычные щелевые согласные.

Традиционные способы регистрации речевого потока (кинограммы, осциллограммы) не дают возможности четко различать аспирацию и аффрикатизацию, так как в том, и в другом случае места образования шума установить нельзя. При комплексной регистрации артикуляторных движений разница между аспирацией и аффрикатизацией видна, так как в первом случае после размыкания контакта, характерного для смычного согласного, фиксируется только ток воздуха до включения голосовых связок, а место образования шума не может быть определено при помощи имеющихся датчиков. Во втором случае при размыкании контакта, характерного для смычки, сохраняется контакт, характеризующий близкий по месту образования щелевой согласный.

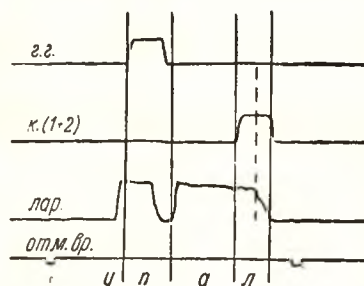


Рис. 2.44. Осциллограмма слова *упал*.

Объяснения в тексте.

Озвончение и оглушение согласных в потоке речи объясняются или задержкой выключения голоса при переходе от звонкого звука к глухому (озвончение), или, наоборот, преждевременным выключением его еще во время артикуляции согласных (оглушение). Эти явления могут быть зарегистрированы при помощи ларингофона. Так как сведения о включении и выключении голоса при комплексной регистрации поступают

одновременно со сведениями о других компонентах артикуляции, всегда можно не только установить сам факт оглушения или озвончения, но и получить временные характеристики этих явлений. На рис. 2.44 можно увидеть озвончение *п* и оглушение *л* в слове *упал* и определить длительность озвонченной или оглушенной фазы звука.

Назализация чистых гласных и согласных в соседстве с носовыми возникает в результате запаздывания выключения носового резонатора после носовых согласных или, наоборот, включения носового резонатора еще до начала носового согласного. Описанные выше датчики — рупорный фотоэлектрический и датчик малоинерционного термоанемометра регистрируют эти явления и дают возможность достаточно точно определять время включения и выключения носового резонатора при произношении звуков, окружающих носовой согласный.

Огубленность согласных в соседстве с огубленными гласными в сочетаниях типа *ту*, *сту* и т. д. заключается в выпячивании и округлении губ, которые начинаются до начала гласного еще во время произнесения согласного. Однако, хотя это было замечено

фонетиками еще очень давно, количественных измерений не проводилось в связи с отсутствием необходимых датчиков. Используемые при комплексной записи электроплетизмографические электроды, располагаемые на нижней губе испытуемого, регистрируют изменение конфигурации нижней губы параллельно с регистрацией других параметров другими датчиками (работа голосовых связок, участие носового резонатора и т. д.). Такая запись дает возможность количественных определений явления лабиализации согласных в потоке речи. На рис. 2.45 приведены осциллограммы слов *дну* и *дна*; на верхней линии фиксируется движение нижней губы. Кроме этого, на осциллограмме записаны движения небной занавески (*в.н.*), работа голосовых связок (*лар.*) и огибающая акустического сигнала с микрофона.

Осциллограмма позволяет судить о времени начала огубленности по отношению к другим параметрам: включению ларингофона на *д*, включению носового резонатора на *н*. Рядом приведена контрольная осциллограмма слова *дна*, где явление огубленности отсутствует.

Для изучения механизмов речеобразования немаловажное значение имеют данные о дыхании во время артикуляции: необходимо иметь представление о соотношенности дыхательных движений с другими артикуляторными движениями как при образовании звуков, так и при организации более крупных отрезков речи (синтагма, фраза, период). Используемый при комплексной регистрации датчик дыхательных движений является очень чувствительным; возможность длительной записи имеет большое значение при изучении фразового дыхания и других явлений, связанных с членением речевого потока. На рис. 2.46 приведена осциллограмма, на которой одновременно с движениями языка, губ и голосовых связок записаны дыхательные движения. При произнесении фразы *Тоня топила баню* имеется резкое падение кривой дыхания, совпадающее с концом интонационно выделяемого слова *Тбня* (отмечено стрелкой).

Аналогичный датчик, как уже было сказано ранее, может регистрировать движения нижней челюсти, по которым можно

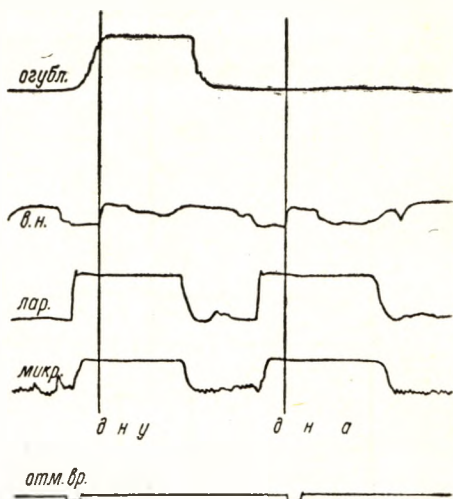


Рис. 2.45. Осциллограммы слов *дну* и *дна*.

Объяснения в тексте.

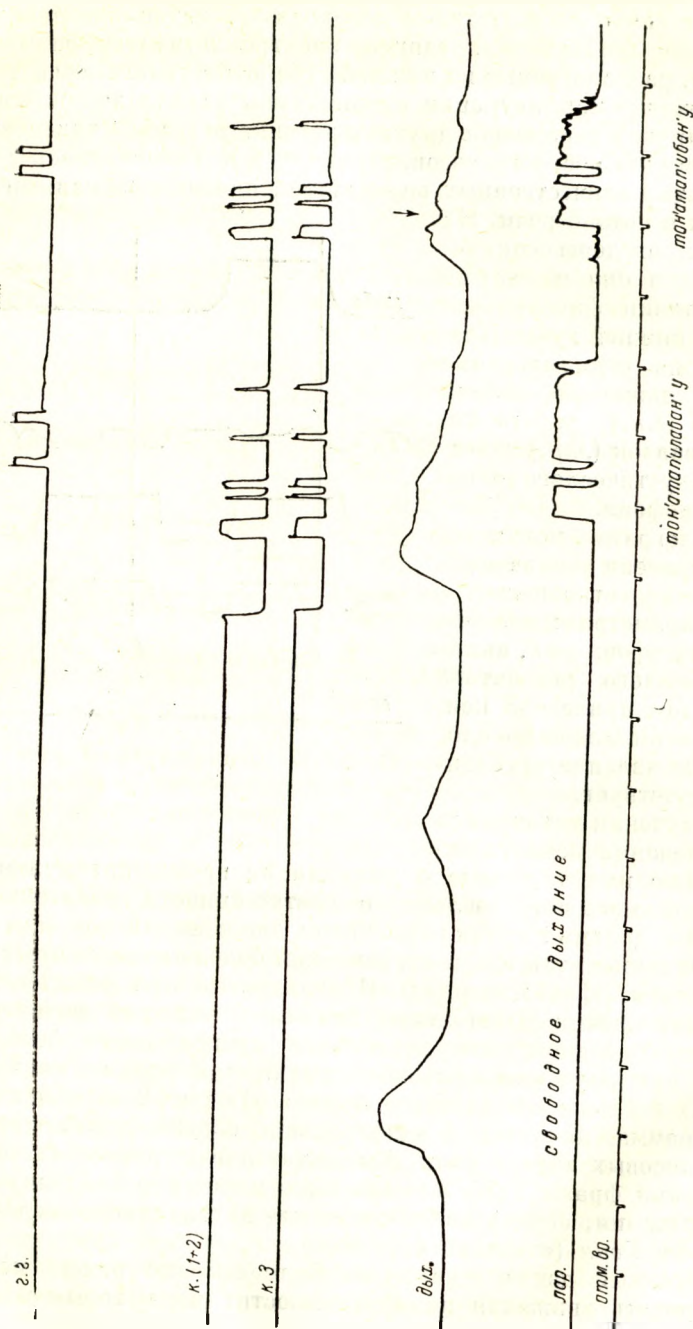


Рис. 2.46. Осциллограммы фраз Тоня топала баню и Тоня, топил баню. Объяснения в тексте.

судить о степени открытости гласных. На рис. 2.47 приведены осциллограммы слогов *та*, *ту*, *то*, *тэ*, *ти*, на которых различие в степени опускания челюсти видно очень хорошо.

С проблемами дыхания и размыкания смычки тесно связан вопрос о внутриротовом давлении при артикуляции. Существующие в экспериментальной фонетике представления о роли внутриротового давления при образовании разных согласных довольно разноречивы. Предлагаемая регистрация дает возможность проследить, существует ли связь между размыканием смычки согласных и определенным значением внутриротового давления. Предварительные записи показывают, что такая связь как будто существует (см. рис. 2.39). На осциллограмме видно, что размы-

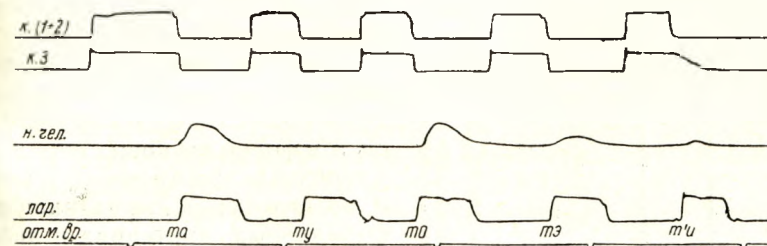


Рис. 2.47. Осциллограммы слогов с гласными *а*, *у*, *о*, *э*, *и*.

кание смычки *к.1* происходит примерно в момент достижения максимума внутриротового давления; кроме того, осциллограмма свидетельствует о разном уровне внутриротового давления для глухих и звонких согласных.

Для целого ряда фонетических исследований представляет интерес определение акустической характеристики той или иной фазы артикуляции. Исследования подобного рода, проводившиеся до сих пор, основывались на анализе акустической стороны речи: при помощи, например, сепаратора, выделялось некое звучание, которое, по мнению исследователя, соответствовало той или иной фазе артикуляции (звонкая смычка, взрыв и т. д.). Методика комплексной регистрации артикуляторных параметров позволяет идти от артикуляции к звучанию: имеется электронный ключ, управляемый сигналами, поступающими с датчиков, и вырезающий из акустического потока участки, соответствующие заданной артикуляции. Так, например, если нужно выделить из звучания слова *туман* безударный гласный *у*, то поступающий на электронный ключ сигнал о включении ларингофонного датчика открывает ключ (определяет начало гласного), а сигнал о начале губной смычки закрывает ключ, т. е. заканчивает поступление сигнала.

Параллельная запись на осциллограмму не только сигналов о включении и выключении контактов, но и сигнала с выхода

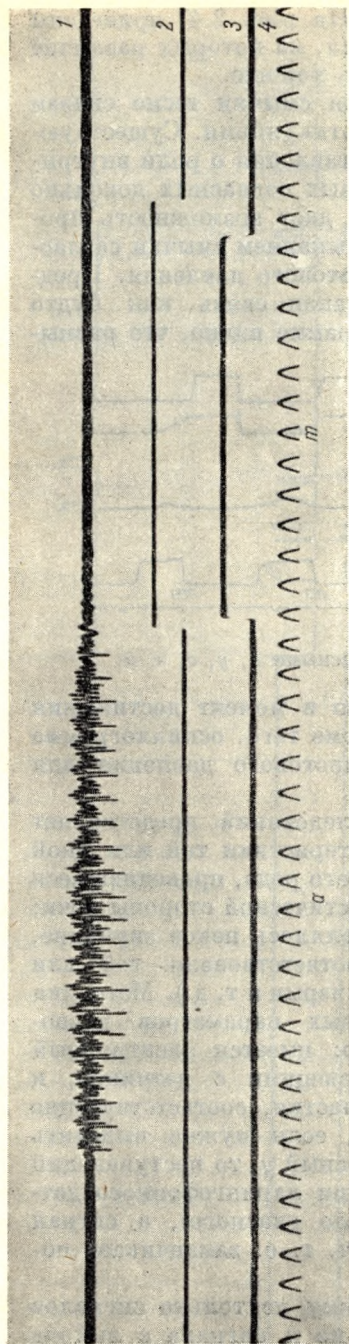


Рис. 2.48. Осциллограмма слога *am*, записанная на шлейфном осциллографе.

1 — осциллограмма акустического сигнала; 2 и 3 — отметки касания языком нёба в точках *к.3* и *к.(1+2)*, полученные с помощью схем амплитудной селекции; 4 — отметка времени 0,01 сек.

микрофонного тракта дает возможность установить строгие критерии деления осциллографических записей звуков речи (рис. 2.48). Определение границы между гласными и следующим согласным осложняется тем, что в слоге *am* начало глухой смычки согласного *m* озвончается; одновременная регистрация сигналов с датчика «искусственное нёбо» (две нижних кривых) дает сведения о реальном начале артикуляции согласного.

Понятно, что такая методика членения речевого потока (которое может также производиться и в форме звуковых отрезков с помощью электронного ключа и магнитофона), основанная на формальных критериях и автоматически осуществляемая электронными устройствами, дает более достоверные результаты, чем те, которые получаются при членении акустического потока речи на основании слуховых впечатлений экспериментатора (сепаратор, для выделения участков звучания на слух, магнитофон с вращающимися головками и т. д.).

Из сказанного ясно, что основные артикуляторные явления, связанные с согласными, а отчасти и с гласными, предлагаемая система комплексной регистрации фиксирует с достаточной полнотой и надежностью. Записи, полученные при помощи этой системы, могут быть использованы как для определения качественных характеристик звуков в потоке речи, так и для различных количественных измерений: длительности отдельных звуков или отдель-

ных фаз артикуляций, временных отношений между разными артикуляторными движениями и т. д. Ряд примеров представляет материал настоящей книги.

При любом количественном описании того или иного явления, чтобы оценить полученные данные, необходимо располагать большим материалом, достаточным для статистической обработки. Определение длительностей явлений, интересующих исследователя, может быть проведено вручную, по осциллограммам, однако такой способ мало пригоден при проведении большого статистического исследования. В этом отношении принципиально новый шаг вперед представляет методика автоматического измерения временных интервалов, описанная подробно ранее. Используя эту методику, исследователь получает в свое распоряжение большой статистический материал, при этом автоматизированно могут определяться не только длительности самых разнообразных по характеру интервалов (от части отдельного звука до синтагмы или фразы), но и ряд их статистических характеристик.

В заключение нужно сказать, что описанные выше возможности получения фонетических характеристик звуков речи при использовании комплексной регистрации артикуляторных и акустических параметров далеко не ограничивают круг применения этой методики.

Г Л А В А 3

ОРГАНИЗАЦИЯ ВРЕМЕННОЙ ПРОГРАММЫ СИНТАГМЫ

1. Определение понятия артикуляторной программы. Синтагма как единица артикуляции

Вопрос о том, что такое артикуляторная программа и чему (словам, предложениям, частям предложений) она должна соответствовать, является весьма сложным и требует предварительного рассмотрения некоторых теоретических представлений относительно процесса синтеза предложений. Весьма простая и правдоподобная модель синтеза предложения человеком была предложена Ингве (Yngve, 1960). Мы будем рассматривать эту модель, считая, что она является тем автоматом, сигналы на выходе которого используются для построения программ артикуляторных движений.

Модель Ингве состоит из грамматики, содержащей правила конкретного языка, и механизма, который является общим и должен работать с грамматикой любого языка. Грамматика содержится в долговременной памяти и состоит из конечного набора правил типа

$$\begin{array}{lll}
 A = B & B = M & M = X \\
 A = B + C & B = M + N & M = X + Y \\
 A = B + C + D & B = M + N + O & M = X + Y + Z \\
 \dots & \dots & \dots
 \end{array}$$

Символ в левой части выражения соответствует названию грамматической конструкции, символы в правой части выражения соответствуют непосредственным составляющим этой конструкции. Они, в свою очередь, могут быть или символами других конструкций, или словами. Например, пусть символ A означает предложение, тогда символы B и C означают части предложения, где B — группа подлежащего, а C — группа сказуемого.

Те выражения, в левой части которых стоит B , показывают, из каких частей может в данном языке состоять группа подлежащего.

Механизм, изображенный на рис. 3.1, действует следующим образом. Он начинает работу с того, что в вычислительный блок поступает символ, соответствующий предложению. Пусть это будет символ A . Он передается на выход и одновременно посылается в долговременную память. В ответ на него из долговременной памяти поступает расшифровка, т. е. символы, образующие правую часть одного из выражений, в которое A входит в качестве левой части. Первый (левый) из группы этих символов поступает в вычислительный блок, а остальные поступают в блок оперативной памяти. Если в ответ на символ A было выбрано правило $A = B + C$, то B поступает в вычислительный блок, а C — в оперативную память. Теперь начинается расшифровка символа B (C пока запоминается в нерасшифрованном виде). Для этого теперь уже B передается на выход и одновременно поступает в долговременную память. В долговременной памяти выбирается одно из выражений, в которое символ B входит в качестве левой части. Например, $B = M + N$. После этого процедура повторяется снова. M поступает в вычислительный блок и затем расшифровывается, N поступает в оперативную память. Можно видеть, что теперь в оперативной памяти нужно хранить уже два символа: N и C .

Так продолжается до тех пор, пока символ, находящийся в вычислительном блоке, не окажется словом. Символ слова передается только на выход, а в вычислительный блок поступает теперь символ из оперативной памяти, причем именно тот, который был последним в нее помещен.

Работу автомата Ингве можно популярно пояснить следующим образом. Допустим, что человек хочет сказать какое-то предложение на заданном языке, причем ему совсем не важно, будет ли смысл в этом предложении, нужно только, чтобы оно было грамматически правильным. Тогда он сначала обращается к учебнику грамматики и выясняет, какие структуры предложений возможны в данном языке. Оказывается, например, что предложение может состоять из группы подлежащего на первом месте и группы сказуемого — на втором. Теперь он выясняет, из чего может состоять группа подлежащего, и получает ответ, что в нее может, например, входить определение женского рода в именительном падеже и существительное женского

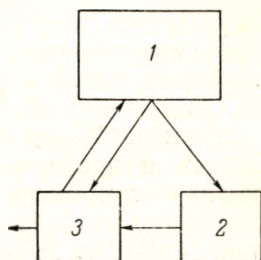


Рис. 3.1. Блок-схема механизма, осуществляющего синтез предложений. (Модель по Yngve, 1960).

1 — блок долговременной памяти; 2 — блок оперативной памяти; 3 — вычислительный блок.

рода в том же падеже. После этого он ищет какое-либо прилагательное женского рода в именительном падеже. Ему случайно попадаетея, допустим, слово *зеленая*. Теперь он ищет существительное женского рода опять же в именительном падеже и попадает на слово *девочка*. Таким образом, он синтезирует сочетание слов *зеленая девочка*, которое, будучи весьма странным, окажется тем не менее вполне русским сочетанием, грамматически правильным.

Ингве показывает, что если в набор грамматических правил включить такие, как $B = D + \dots + N$, то предлагаемая им модель оказывается способной генерировать любые правильные английские предложения. При этом она обладает тем свойством, что образует слова в предложении одно за другим и в том же порядке по времени (от первого слова к последнему), как это делает человек.

Дальше Ингве делает существенный новый шаг в развитии модели и допускает, что объем оперативной памяти автомата составляет примерно 7 символов, что соответствует, по данным экспериментальной психологии, объему оперативной памяти человека (Miller, 1956). Рассматривая различные типы конструкций английских предложений, Ингве убеждается, во-первых, в том, что максимальный объем оперативной памяти, требующийся для синтеза обычных английских предложений, действительно не превышает семи символов. Во-вторых, Ингве устанавливает, что в английском языке существуют «приспособления» (способы), используемые для уменьшения объема оперативной памяти, требующейся для синтеза предложений. Эти результаты являются серьезным доводом в пользу модели Ингве.

Рассмотрим теперь подробнее, какие символы образуются на выходе автомата Ингве и содержат ли они всю информацию, необходимую для построения артикуляторной программы. Первое существенное для нас положение состоит в том, что символы слов, образуемые на выходе автомата, соответствуют словоформам. Это значит, что если слово — существительное, то оно дано уже в нужном числе и падеже, если слово — глагол, то оно дано в нужном лице, числе и времени. Следовательно, символ слова, образуемый на выходе автомата, содержит полную информацию о той последовательности фонем, которой может быть описано данное слово. Для того чтобы превратить этот символ в последовательность фонем, достаточно знать этот, и только этот, символ.

Второе очень интересное положение состоит в том, что, кроме символов слов, на выходе автомата образуются также символы структур предложения. Хорошо известно, что, для того чтобы правильно признать предложение (часть предложения) на любом языке, недостаточно знать произношение изолированных слов этого языка. Необходимо знать также правила произнесения определенных структур, например вопросительной или повествовательной. Эти правила касаются управления так называемыми

просодическими характеристиками речи — ритмикой, громкостью, интонацией. Естественно, что для того чтобы нужные просодические характеристики были созданы при произнесении, они должны быть заданы в артикуляторной программе. Для этого указания о них, в свою очередь, должны быть поданы на вход устройства, реализующего эту артикуляторную программу. Очевидно, что символы структур, образуемые на выходе автомата Ингве, могли бы служить в качестве таких указаний.

Обратимся теперь к вопросу о том, сколько последовательных символов на выходе автомата Ингве должно использоваться для построения одной артикуляторной программы. Так как указания о просодических характеристиках относятся не к отдельному слову, но к определенным структурам, очевидно, что артикуляторная программа должна включать не отдельные слова, но всю группу слов, относимых к данной структуре. К сожалению, эти рассуждения выявляют только, что последовательность символов, образуемых на выходе автомата, синтезирующего предложения, должна разбиваться на группы символов, и что первым символом в группе должен быть символ структуры. По каким признакам должна проходить разбивка на группы, какие структуры могут входить в одну артикуляторную программу, а какие не могут, остается неясным.

Прежде чем предполагать, что процесс группирования символов для целей построения артикуляторной программы осуществляется устройством, следующим за автоматом (стоящим на его выходе), посмотрим, не обладает ли сам автомат такими свойствами, которые заставили бы его вырабатывать символы отдельными сериями.

Сам Ингве этого вопроса не рассматривает. Он пишет, что автомат работает циклично, вырабатывая по одному выходному символу за каждый цикл. Работа автомата продолжается непрерывно до тех пор, пока все предложение не будет синтезировано.

Однако автомат Ингве не является полной моделью синтеза предложения. Автор сам предупреждает, что он не рассматривает вопроса о механизмах, с помощью которых можно построить не какое угодно грамматически правильное предложение данного языка, но предложение со строго определенным смыслом. Говоря более формально, Ингве принимает как данное, что при направлении символа конструкции для поиска в долговременной памяти из нее выбирается одно, и только одно, строго определенное частное значение этой конструкции.

Если в долговременной памяти автомата имеются правила типа

$$\begin{aligned} A &= B, \\ A &= B + C, \\ A &= B + C + D, \end{aligned}$$

то, для того чтобы в ответ на A образовать $B+C$, необходимо не только получить символ A , но и нужно одновременно располагать указаниями о том, какое из перечисленных в таблице правил должно в данный момент использоваться. Интуитивно понятно, что выбор одного из грамматически возможных правил может определяться только смыслом будущего предложения. Однако если так, то мы можем отвлечься от весьма сложного вопроса о природе и способах записи смысла в мозгу человека и чисто формально определить смысл как такие указания, в результате которых из всей таблицы правил выделяется (например, подчеркивается) только одно определенное правило. Эти указания будем дальше называть смысловыми.

Очень интересной особенностью автомата Ингве является то, что смысловые указания могут поступать последовательно. В то время, когда символ A поступает в долговременную память, туда же должны быть посланы указания о том, как именно этот символ должен быть расшифрован. Указаний о том, как должен расшифровываться символ B , в это время еще не требуется. Они станут необходимыми только тогда, когда в долговременную память будет послан символ B . Иначе говоря, это означает, что синтез предложения может происходить без того, чтобы все детали смысла этого предложения были заранее полностью известны. Это очень хорошо соответствует интуитивным представлениям о том, что, начиная говорить, человек еще не знает в деталях того, что он скажет (знает это только в самом общем виде).

Итак, автомат Ингве допускает, чтобы смысловые указания поступали последовательными порциями. Можно далее показать, что только в том случае, если поступление смысловых указаний будет последовательным, автомат Ингве будет способен образовать сложное предложение заданного смысла.

Для синтеза строго определенного предложения необходимо каждую конструкцию расшифровывать единственным образом. Это значит, что в таблице правил для данной конструкции может быть одновременно подчеркнуто только одно из них. Отсюда следует, что смысл целого сложносочиненного предложения не может быть задан одновременно, так как это означало бы, что в таблицах для каких-то конструкций были бы одновременно подчеркнуты два (или более) правила.

Введем теперь представление о единицах смысла. Единицей смысла будем считать набор таких смысловых указаний, которые являются совместимыми и могут задаваться одновременно. Легко обнаруживается, что совместимыми являются такие смысловые указания, которые относятся к правилам для разных конструкций. Одновременные сведения о том, что $A=B+C$, $B=M+N$, а $N=X$, не могут мешать друг другу, так как конструкции расшифровываются последовательно, и каждая конструкция расшифровывается однозначно.

Несовместимыми должны быть смысловые указания, относящиеся к правилам для одной конструкции. Указания о том, что $A=B+C$ и $A=D$, не могут задаваться одновременно и, следовательно, не могут принадлежать к одной и той же единице смысла. Итак, дополнив автомат Ингве блоком, посылающим смысловые указания, мы должны допустить, что в работе автомата в целом должны обнаруживаться дополнительные более крупные смысловые циклы. Блок смысловых указаний должен менять свое состояние, переходя от одной смысловой единицы к другой, причем эта смена состояний должна быть синхронизирована с работой механизма, описанного Ингве. Естественно считать, что в последовательности символов, образуемых на выходе автомата, одной смысловой единице должна соответствовать серия символов, не содержащая повторяющихся конструкций.

Эти выводы являются для нас весьма важными, так как они позволяют предположить, что группировка выходных символов автомата, синтезирующего предложения, необходимая для построения артикуляторной программы, не требует какого-то дополнительного механизма. Эта группировка должна иметься уже на выходе самого автомата благодаря наличию смысловых циклов в его работе. Если это предположение правильно, то одна артикуляторная программа должна соответствовать одной смысловой единице.

В лингвистике существует такое интересное понятие, как синтагма. Для определения ее, с одной стороны, прибегают к понятию смысла. Она описывается как предложение или часть предложения, выделенная по смыслу. С другой стороны, синтагма явно связана с артикуляцией и обязательно произносится на одном выдохе. Синтагмы разделяются друг от друга паузами, тогда как слова внутри каждой из них произносятся слитно. Средняя ее длина в свободной речи составляет семь слогов (Kaiser, 1939). Если допустить, что элементами синтагмы являются слоги, то длина ее согласуется с объемом оперативной памяти человека.

Учитывая эти интересные особенности синтагмы, естественно предположить, что она и является той смысловой и одновременно артикуляторной единицей, необходимость существования которой мы пытались ранее выяснить чисто логическим путем.

Если синтагма является смысловой единицей, она не должна содержать повторяющихся конструкций, а предложение, содержащее повторяющиеся конструкции, должно состоять более чем из одной синтагмы. На первый взгляд кажется, что это действительно оправдывается, однако исследование этого вопроса представляет собой чисто лингвистическую задачу, лежащую вне пределов нашей компетенции.

Для нас гораздо более интересна вторая сторона вопроса, а именно предположение о том, что одна синтагма соответствует

одной артикуляторной программе, а для произнесения предложения, состоящего из нескольких синтагм, соответственно используется несколько последовательных артикуляторных программ. Важность этого вопроса для нас вполне очевидна, так как, не определив, чему в речевом потоке соответствует артикуляторная программа, естественно нельзя исследовать структуру этой программы.

Поэтому наши эксперименты были начаты с попытки выяснить, соответствует ли предложению, состоящему из двух синтагм, одна артикуляторная программа или же для его произнесения используются две программы.

При выборе метода исследования были использованы следующие соображения. Известно, что синтагмы разделяются паузами. Допустимо предположить различные механизмы возникновения паузы. Один из них может состоять в том, что пауза представляет собой такой же элемент в артикуляторной программе, как и все остальные комплексы артикуляторных движений, осуществляемых на протяжении синтагм. Если пауза задана в программе, то нет никаких оснований ожидать, что флюктуации длительности паузы между синтагмами при повторных произнесениях предложения будут превышать флюктуации длительностей синтагм. Вторым механизмом может состоять в том, что пауза отражает остановку в работе автомата, связанную со сменой артикуляторных программ. В этом случае длительность паузы между синтагмами не должна быть заданной, и естественно ожидать, что она будет флюктуировать больше, чем длительность самих синтагм.

Для проверки этого предположения было исследовано на одном дикторе произнесение двух простых фраз, состоящих из двух синтагм: *Эй, мальчик, Дима тут гуляет?* и *Эй, мальчик, тут гуляет детский садик?* Первая фраза произносилась с логическим ударением на слове *гуляет*, вторая на слове *тут*.

Фразы произносились в составе таблицы, состоящей из восьми различных фраз. Вся таблица была повторена 40 раз подряд. Испытуемому давалась инструкция придерживаться при всех произнесениях одного и того же темпа речи. Обработка осциллограмм артикуляции состояла в измерении средней длительности (M) и среднего квадратического отклонения (σ) для каждой из синтагм и для интервала между синтагмами.

Полученные данные показали, что пауза между синтагмами отнюдь не является обязательной. В первой из фраз отсутствие паузы наблюдалось в 3 случаях из 40, во второй — в 9 случаях.

Наиболее характерной оказалась колоссальная изменчивость интервала между концом первой и началом второй синтагмы. При этом длительности самих синтагм колебались в весьма небольших пределах. Полученные данные приведены ниже.

	Фраза № 1		Фраза № 2	
	Средняя длитель- ность, в мсек. (М)	Коэффи- циент вариации $\left(\frac{\sigma}{M}\right)$	Средняя длитель- ность, в мсек. (М)	Коэффи- циент вариации $\left(\frac{\sigma}{M}\right)$
Первая синтагма	630	0.047	620	0.052
Вторая синтагма	1000	0.029	1470	0.035
Интервал между синтагмами . .	350	0.610	160	0.875

Полученные данные с несомненностью показывают, что исследование временной организации последовательности артикуляторных движений может производиться только в пределах синтагмы. Переход от одной синтагмы к другой не является определенным по времени.

Этот результат позволил нам принять синтагму в качестве артикуляторной единицы. Естественно, что она может быть и очень малой по величине, состоять из одного односложного слова, например *да* или *нет*. Соответственно, задача дальнейшего исследования была определена как изучение организации артикуляторных движений внутри синтагмы.

С помощью описанных в предыдущих главах методов измерений нельзя было регистрировать такие характеристики, как интонационный рисунок и распределение мышечных усилий внутри синтагмы. Зато можно было достаточно надежно измерять временные отношения как между различными комплексами артикуляторных движений, так и между отдельными движениями внутри этих комплексов. Поэтому вся работа была построена на исследовании временной организации артикуляции. Это представлялось рациональным и с других точек зрения. Дело в том, что вопрос о временной организации движений является наиболее сложным в теоретическом отношении. Поэтому казалось, что если удастся в какой-то мере в нем разобраться, то исследование остальных вопросов будет сопряжено уже со значительно меньшими трудностями.

2. Влияние темпа речи на длительность различных элементов синтагмы

Любая синтагма представляет собой ритмически организованную последовательность движений. Характер ритмической организации определяется прежде всего положением ударений внутри слов, входящих в синтагму, а также положением логического ударения. Ритмическая организация выражается в том, что звуки, находящиеся на разном месте относительно словесного ударения, имеют различную длительность; звуки, входящие в слово, на которое падает логическое ударение, имеют большую длительность, чем соответствующие им звуки других слов синтагмы. Известно также, что абсолютная длительность звуков

принципиально не является существенной, так как в огромной степени зависит от темпа речи.

К сожалению, вопрос о том, что такое темп речи и каким образом он связан с принятым в фонетике понятием стиль речи, до сих пор совершенно неясен. В лингвистических работах, затрагивающих эти вопросы (Щерба, 1912, 1915; Зиндер, 1960; Passy, 1887; Jones, 1909; Jespersen, 1926; Fonagy a. Magdics, 1960; Hildebrandt, 1961; Blankinship, 1962, и др.), дело сводится, как правило, к констатации существования таких явлений и к предупреждениям о том, что темп речи влияет на длительность звуков речи. В. Чернышев (1915) сделал попытку найти причины изменения темпа речи, однако результаты, к которым он пришел, весьма неутешительны. По его мнению, темп речи зависит от содержания высказывания, от пола и возраста говорящего, от его воли, темперамента, от состояния его чувств и мыслей, от интонации речи и ее синтаксического строения. Л. В. Бондарко, Л. А. Вербицкая и Л. Р. Зиндер (1959) специально исследовали длительность звуков речи при произнесении в полном и в разговорном стиле. Однако полученные данные также не дают возможности ответить на интересующий нас вопрос о влиянии темпа речи на временной рисунок синтагмы.

В поисках метода исследования этого вопроса мы исходили из следующих соображений. Если темп речи может непрерывно меняться в достаточно широких пределах (может принимать любое значение внутри этих пределов), то нереально допустить, что для каждого возможного темпа строится своя отдельная артикуляторная программа. Это просто было бы трудно сделать, и, кроме того, это было бы явно неэкономным. Значительно естественнее сделать другое предположение, а именно допустить, что темп вообще никак не фигурирует в артикуляторной программе, он относится уже к механизмам осуществления программы, упрощенно говоря, характеризует скорость реализации программы.

Но если мы встанем на вторую точку зрения, то временная организация синтагмы должна характеризоваться именно такими отношениями внутри нее, которые не зависят от темпа, являясь инвариантными при изменениях последнего. Так как абсолютные значения длительности элементов синтагмы явно не сохраняются постоянными, естественно предположить, что инвариантными могут быть только относительные длительности.¹

Такой ход рассуждений приводит к новой постановке вопроса. Если не рассматривать изменения темпа как помеху, которая искажает данные, то они, наоборот, должны служить способом выявления константных временных отношений в речи,

¹ Постоянства относительных длительностей следует ожидать, исходя из того факта, что слух человека является чувствительным именно к относительным длительностям звуковых сигналов (Чистович, 1959; Stott, 1935; Henry, 1948; Small a. Campbell, 1962; Creelman, 1962).

остающихся скрытыми, когда темп является постоянным. Заставив человека произвольно в широких пределах менять темп речи, можно выяснить, что именно во временной картине синтагмы остается постоянным. Это даст возможность разделить временные свойства синтагмы, которые обусловлены программой, и особенности, которые связаны с процессами осуществления программы.

Конкретной задачей экспериментов, описываемых в настоящей главе, являлось выяснение влияния темпа речи на относительную длительность слов в синтагме, относительную длительность слогов в слове и относительную длительность звуков речи в слоге.

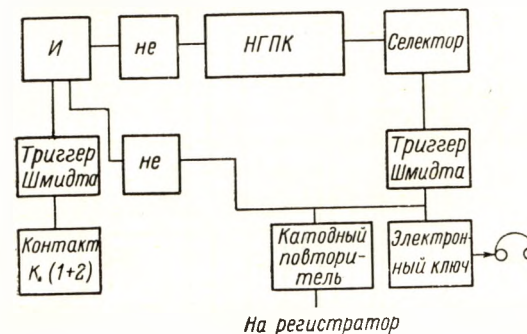


Рис. 3.2. Блок-схема установки, использовавшейся для задания длительности фразы.
Объяснения в тексте.

Исследование проводилось на примере простой фразы, состоящей из открытых слогов (*Тоня топила баню*). Фраза произносилась в повествовательном стиле с тремя различными логическими ударениями (на *Тбня*, *топила* и *бáню*). В опытах участвовали два испытуемых (мужчина и женщина), которые должны были произносить фразу в семи различных темпах речи (общая длительность фразы задавалась 0.8, 1.0, 1.4, 1.8, 2.2, 2.6 и 3.0 сек.). Испытуемый Ю. К., ввиду индивидуальной склонности к медленной речи, затруднялся произнести фразу за 0.8 сек. Для него имеются данные только для 6 темпов. Задание темпа производилось с помощью схемы, изображенной на рис. 3.2.

В момент начала фразы испытуемый, произнося звук *т*, замыкает не́бные контакты в точке *к. (1+2)*. Возникающий при этом сигнал через формирующие блоки запускает генератор типа НГПК, который выдает треугольный импульс заданной длительности. Задний фронт этого импульса используется для включения короткого звукового сигнала, который через головные телефоны подается испытуемому, указывая на окончание заданного интервала. Испытуемый должен регулировать свой темп речи таким образом, чтобы конец последнего звука во фразе совпадал

с началом этого сигнала. Схема обеспечивает подачу достаточно стабильных импульсов. Отклонения длительности интервала не превышают 1%.

Важным являлось то обстоятельство, что испытуемому подавался сигнал только о конце интервала, начало которого он задавал сам началом произнесения фразы. Этим достигалось более естественное ее произнесение.

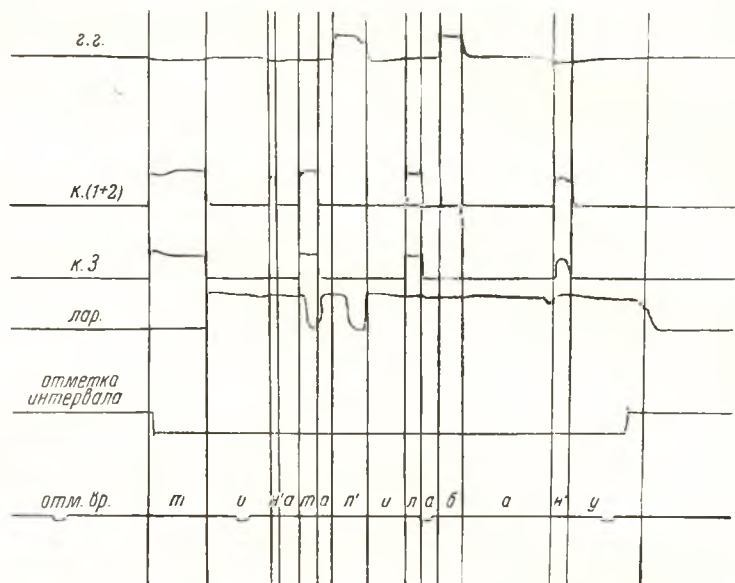


Рис. 3.3. Осциллограмма фразы *Тоня топила баню*, произнесенной диктором Л. Ч.

У испытуемых регистрировались артикуляторные движения по методике, описанной в предыдущей главе: контакт между губами *г.г.*, контакт языка с искусственным нёбом в точках *к.(1+2)* и *к.3*, а также наличие или отсутствие голоса (сигнал с ларингофона *лар.*).

Кроме того, на осциллограмме отмечались начало и конец задаваемого интервала времени, а также тот короткий звуковой сигнал, который подавался в наушники испытуемого в конце этого интервала. Скорость движения бумаги равнялась 60 мм/сек. Для статистической обработки было использовано 1260 осциллограмм. Пример осциллограммы фразы, произнесенной одним из испытуемых при заданной длительности 2.6 сек., приведен на рис. 3.3.

При каждой заданной длительности фразы и данном логическом ударении записывалось 25—30 повторных произнесений.

Длительность первых произнесений, как правило, существенно отличалась от задаваемой — испытуемый производил поиск нужного темпа речи.

Обработка материала на первом этапе заключалась в том, что для всех фраз, сказанных при заданном интервале, измерялась длительность фразы. Строился график, показывающий изменение длительности при последовательных произнесениях (рис. 3.4). Затем выбиралось пять произнесений, характеризующихся одинаковой длительностью, ближе всего соответствующей заданной длительности. Они подвергались дальнейшей обработке, которая заключалась в следующем.

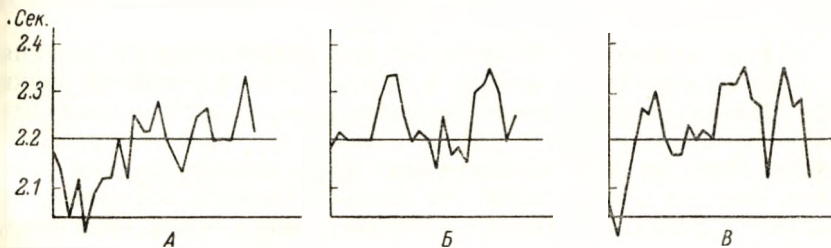


Рис. 3.4. Изменения длительности фразы при повторных произнесениях.

А — при логическом ударении на слове *Тоня*; Б — при логическом ударении на слове *баню*. По оси абсцисс — последовательные произнесения; по оси ординат — время (в сек.). Прямой линией указана задаваемая длительность.

За начало отсчета принимался конец контакта $\kappa.(1+2)$ (конец *Т* в *Тбня*), и далее с помощью точной измерительной линейки определялись расстояния (в мм) до следующих явлений на осциллограмме:

- | | |
|--|---|
| 1) начало $\kappa.(1+2)$ (начало <i>т</i>), | 8) начало $\kappa.3$ (начало <i>л</i>), |
| 2) начало $\kappa.(1+2)$ (начало <i>н</i>), | 9) конец $\kappa.3$ (конец <i>л</i>), |
| 3) конец $\kappa.(1+2)$ (конец <i>н</i>), | 10) начало <i>г.г.</i> (начало <i>б</i>), |
| 4) начало $\kappa.(1+2)$ (начало <i>т</i>), | 11) конец <i>г.г.</i> (конец <i>б</i>). |
| 5) конец $\kappa.(1+2)$ (конец <i>т</i>), | 12) начало $\kappa.(1+2)$ (начало <i>н</i>), |
| 6) начало <i>г.г.</i> (начало <i>п</i>), | 13) конец $\kappa.(1+2)$ (конец <i>н</i>), |
| 7) конец <i>г.г.</i> (конец <i>п</i>), | 14) конец <i>голоса</i> (конец фразы). |

Полученные значения расстояний записывались с точностью 0.5 мм (это соответствовало по времени 8.3 мсек.) в таблицы, строки которых соответствовали разным осциллограммам, а столбцы — разным явлениям на осциллограмме.

Разность между значениями, записанными в двух соседних клетках одной строки, была принята за длительность звука речи (в мм). Длительность согласного при этом определялась как

расстояние между концом и началом данного контакта, длительность гласного — как расстояние между началом данного контакта и концом предыдущего.

Средние из пяти значений длительности звука речи оценивались как характерные для данной длительности фразы. Кроме того, для каждой длительности фразы находились средние длительности слогов и слов. Необходимо сказать несколько слов по поводу принятого нами условного определения согласного как фазы смычки артикуляторных органов, а гласного — как фазы раскрытия. Такое определение удобно в том отношении, что оно является достаточно формальным. В главах, касающихся вопросов организации слога, будет показано, что оно имеет также и определенные теоретические основания.

В фонетической литературе не существует единого мнения по вопросу о методах членения артикуляторного потока на звуки речи. Членя речевой поток на звуки речи, фонетики пытаются найти компромисс между двумя взаимоисключающими требованиями. Одно из них состоит в том, чтобы считать звуком речи такой участок артикуляции, на котором имеются все признаки, требуемые фонемной принадлежностью данного звука. Например, участком артикуляторного потока, соответствующим согласному *н*, должен быть такой участок, на протяжении которого опущена нёбная занавеска, кончик языка поднят и прижат к верхним зубам, имеется фонация. Второе требование состоит в том, чтобы разделить речевой поток на участки, непосредственно примыкающие друг к другу (один звук речи должен непосредственно следовать за другим звуком речи). Так как различные артикуляторные органы работают не строго одновременно и, кроме того, обладают определенной инерционностью, естественно, что эти два требования принципиально нельзя совместить друг с другом. В результате в фонетике появляются логически парадоксальные понятия, такие, как, например, «озвонченный участок глухого согласного».

Очевидно, что научное решение вопроса нереально до тех пор, пока не определено, каким образом происходит управление речью и что представляют собой элементарные артикуляторные комплексы. Пока это не выяснено, желательно пользоваться наиболее простыми критериями для членения, которые можно легко формально определить.

На основании данных эксперимента можно сделать вывод о том, что относительная длительность слов во фразе не обнаруживает каких-либо закономерных изменений при изменении темпа речи (общей длительности фразы). Полученные результаты приведены на рис. 3.5. Из графиков следует, что относительная длительность слов не остается строго постоянной, она несколько колеблется от одной длительности фразы к другой. Однако сопоставление всех графиков показывает, что флюктуации явно слу-

чайны в отношении общей длительности фразы. Это заставляет считать, что они связаны с какими-то явлениями, не имеющими прямого отношения к темпу речи.

Одной из причин флюктуаций может быть недостаточно точное соблюдение диктором заданного ему логического ударения.

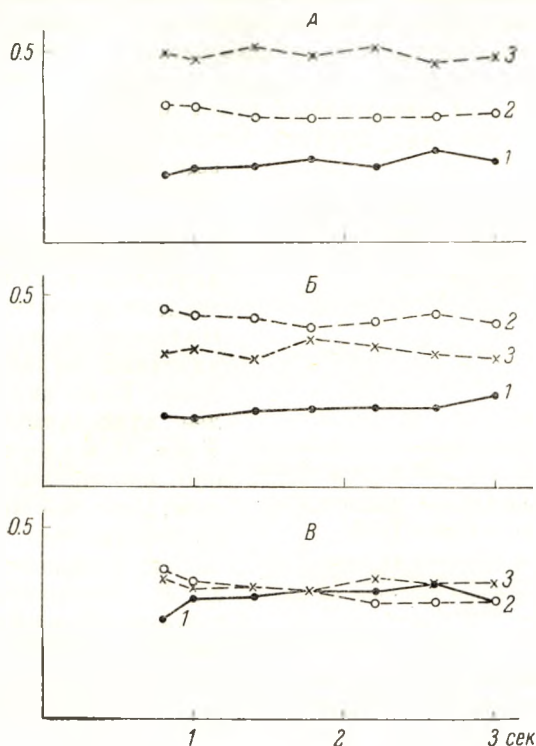


Рис. 3.5. Зависимость относительной длительности слов во фразах с различным логическим ударением от темпа речи.

А — при ударении на слове баню; Б — при ударении на слове топил; В — при ударении на слове Тоня. 1 — слово Тоня; 2 — слово топил; 3 — слово баню. По оси абсцисс — общая длительность фразы (в сек.); по оси ординат — относительная длительность слова.

Логическое ударение автоматически осуществляется при естественном произнесении фразы в речевом потоке. Однако при многократных повторениях одной и той же фразы становится трудным его придерживаться, особенно если человек не обладает актерскими данными и не умеет говорить «с выражением». Диктор Ю. К. вообще хуже справлялся с логическим ударением — относительные длительности слов во фразах довольно мало отличались друг от друга. У диктора Л. Ч. разница между фразами

с различным логическим ударением была выражена значительно более отчетливо, причем относительная длительность слов флюктуировала существенно меньше. Если дело действительно в несоблюдении логического ударения, то наблюдавшиеся нами флюктуации относительных длительностей слов должны быть отнесены не за счет неточностей осуществления артикуляторной программы, а за счет недочетов ее планирования (синтеза).

Из графиков следует, что слово, на которое падает логическое ударение, не обязательно имеет большую относительную длительность, чем два другие. Очевидно, здесь вмешивается

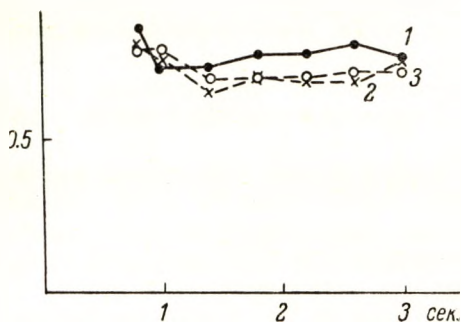


Рис. 3.6. Относительная длительность слога *то-* в слове *Тоня*.

1 — при логическом ударении на слове *Тоня*;
2 — при логическом ударении на слове *топила*;
3 — при логическом ударении на слове *баню*.

еще одна особенность распределения длительности фразы между ее элементами. Дело в том, что относительная длительность слова отчетливо зависит от его места во фразе. Из графиков видно, что при ударении на слове *топила*, слово *баню* оказывалось более длительным, чем слово *Тоня*. При ударении на *Тоня* оно оказывалось или равным, или меньшим по длительности, чем слово *баню*.

Это совпадает с данными Л. В. Златоустовой (1962), которая объясняет описанное явление тем, что неударные

гласные в конечном положении во фразе почти в два раза длиннее, чем в остальных положениях (в начале и в середине фразы). Полученные нами данные интересны тем, что они свидетельствуют об обусловленности различия в длительности слов самой программой артикуляции (явление наблюдается при всех темпах речи).

Распределение длительности слова между слогами показано на рис. 3.6, 3.7 и 3.8. Для двусложных слов (*Тоня* и *баню*) приведены кривые относительной длительности только для одного слога (ударного), для второго слога они являются их зеркальным отражением. Данные, касающиеся относительной длительности слогов в словах *Тоня* и *топила* при общей длительности фразы от 0.8 до 1.4 сек. для испытуемого Л. Ч. и от 1.0 до 1.4 сек. для испытуемого Ю. К., нельзя считать абсолютно достоверными по следующим причинам. При произнесении фразы в столь быстром темпе редукция конечного гласного в слове *Тоня* доходила до его полного исчезновения. В этом случае на осциллограмме нельзя было разделить звуки *н* из *Тоня* и *т* из *топила*. Условно длительность каждого из этих звуков принималась за половину их суммарной длительности. Относительная ошибка, связанная

с такой вольностью в трактовке осциллограммы, очевидно, мала при измерении относительной длительности слова, на относительную же длительность слога в слове она оказывает существенное влияние.

Относительная длительность слогов в слове не зависит от темпа речи (см. рис. 3.6—3.8). Относительная длительность несколько флюктуирует от одной серии произнесений к другой, однако эти флюктуации несомненно имеют случайный характер и почти во всех случаях невелики.

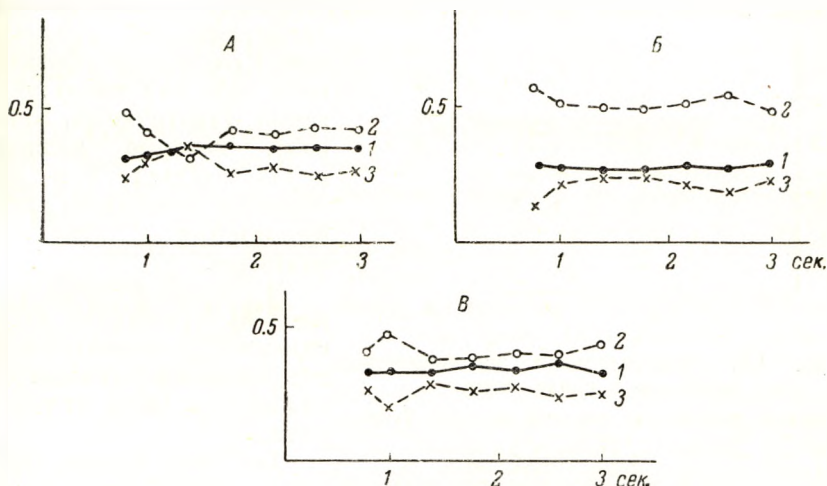


Рис. 3.7. Относительная длительность слогов в слове *топила*.

А — при логическом ударении на слове *баню*; Б — при логическом ударении на слове *топила*; В — при логическом ударении на слове *Тоня*. 1 — слог *та*; 2 — слог *пи*; 3 — слог *ла*. По оси абсцисс — общая длительность фразы (в сек.); по оси ординат — отношение длительности слога к длительности слова *топила*.

Относительная длительность слогов в слове определяется не только характером самого слова (положением в нем словесного ударения), но и структурой фразы в целом — местом слова в синтагме, наличием или отсутствием логического ударения на данном слове.

Значение места слова в синтагме отчетливо выступает при сравнении рис. 3.6 (слово *Тоня*) и рис. 3.8 (слово *баню*). Можно видеть, что ударный слог в слове *Тоня* значительно превышает по длительности безударный слог. В слове *баню* оба слога имеют примерно одинаковую длительность.

Влияние логического ударения на относительную длительность слогов ярко выражено у диктора Л. Ч. Если слово находится под логическим ударением, то относительная длительность ударного слога возрастает. Это явление наблюдалось также

Л. В. Златоустовой (1962). У диктора Ю. К., который, как уже говорилось, значительно хуже воспроизводил логическое ударение, это явление почти не выражено.

Зависимость относительной длительности слогов внутри слова от свойств синтагмы (места слова в синтагме, положения логического ударения) подтверждает представление о том, что артикуляторная программа синтагмы не может рассматриваться как простая последовательность словесных подпрограмм. Это заставляет предположить, что слово, как и фонема, является единицей логического характера, действующей на уровне синтеза

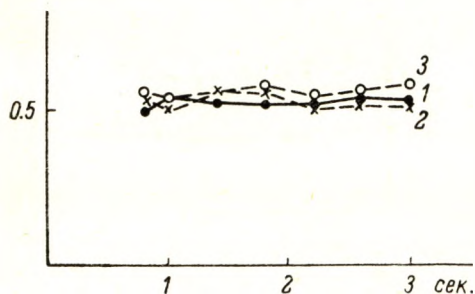


Рис. 3.8. Относительная длительность слога *ба-* в слове *баню*.

1 — при логическом ударении на слове *Тоня*;
2 — при логическом ударении на слове *топила*;
3 — при логическом ударении на слове *баню*.
По оси абсцисс — общая длительность фразы (в сек.); по оси ординат — отношение длительности слога *ба-* к длительности слова *баню*.

лись значительными, был использован такой интегральный показатель, как отношение суммы всех согласных интервалов к сумме всех гласных интервалов. При увеличении длительности фразы наблюдается явное уменьшение доли времени, занимаемого в ней согласными (рис. 3.9).

Полученная закономерность согласуется с фонетическими наблюдениями о том, что длительность гласного в значительно большей мере подвержена изменению, чем длительность согласного (Щерба, 1912; Зиндер, 1960; Jespersen, 1926).

Изменение относительной длительности гласного в слове при изменении темпа речи может быть обусловлено тем, что скорость изменения длительности при изменении темпа различна для гласных и согласных. Однако возможно и другое предположение: длительность согласного является ограниченной: согласный не может быть как меньше, так и больше некоторых величин (Stetson, 1951). Для проверки подобного предположения для каждого согласного были построены кривые зависимости их абсолютной длительности от заданной длительности фразы. В большинстве

артикуляторной программы. Его можно представить себе как набор указаний (инструкций, правил), которые должны использоваться при построении программы артикуляторных движений синтагмы.

В экспериментах выявилась отчетливая зависимость относительной длительности согласного и гласного в слове от темпа речи.

Так как флюктуации относительной длительности звуков речи в индивидуальных слогах оказа-

случаев кривые имели форму, подтверждающую это предположение (рис. 3.10).

Можно видеть, что смычка *n* в слове *баню* не увеличивается при продолжительности фразы от 0.8 до 1.4 сек., затем растет, и рост прекращается в интервале от 2.6 до 3.0 сек.

Длительность *n* в слове *топила* и *б* в слове *баню* остается постоянной при продолжительности фразы от 0.8 до 1.0 сек., затем растет и перестает увеличиваться в интервале 2.2—3.0 сек. (рис. 3.10).

Увеличение длительности гласного может, вероятно, происходить без ограничений. Из фонетики известно также, что гласный может полностью исчезнуть при быстром темпе речи. Мы наблюдали полную редукцию гласного в слоге *-ня* (*Тоня*) при заданной длительности фразы 0.8 и 1.0 сек. Зная длительность первого

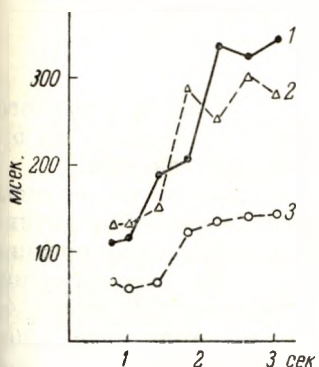


Рис. 3.10. Зависимость абсолютной длительности согласных от общей длительности фразы.

1 — длительность *б* из слова *баню*; 2 — длительность *п* из слова *топила*; 3 — длительность *н* из слова *баню*. По оси абсцисс — общая длительность фразы (в сек.); по оси ординат — длительность согласных (в мсек.).

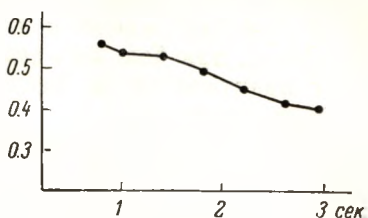


Рис. 3.9. Зависимость относительной длительности согласных от общей длительности фразы.

По оси абсцисс — общая длительность фразы; по оси ординат — отношение суммы длительностей всех согласных к сумме длительностей всех гласных.

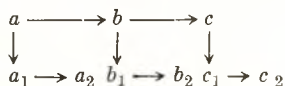
слога и отношение длительности ударного слога к длительности неударного (оно определялось по данным, полученным при других темпах речи), легко было рассчитать, какова должна была быть длительность безударного слога в тех случаях, когда мы наблюдали исчезновение гласного. Подсчеты показали, что она должна была составлять 45—60 мсек. Рассмотрение же всего материала показало, что минимальная длительность согласного (смычки согласного) составляет от 30 до 60 мсек., в зависимости от характера согласного. Как видно, эти величины практически совпадают. Это заставляет считать, что в основе редукции гласных лежит весьма простое явление. Она возникает, когда заданный в программе интервал между слогами оказывается слишком малым для того, чтобы при быстром темпе речи на этом интервале могло осуществиться как смыкание, так и

размыкание органов, артикулирующих согласный. В результате для гласного просто не остается времени.

Итак, полученные результаты показали, что изменение темпа речи не влияет на относительные длительности слоговых и более крупных (словесных) интервалов внутри синтагмы. Вместе с тем изменения темпа существенно изменяют относительные длительности согласного и гласного внутри слога. При быстром темпе речи гласный может полностью исчезнуть (полная редукция гласного), при больших замедлениях темпа длительность согласного практически не меняется и удлинение слога происходит за счет гласного.

Следовательно, если мы рассматриваем синтагму как последовательность звуков речи, мы не можем обнаружить какого-либо постоянства в ее временном рисунке. Однако если мы переходим к слогам и рассматриваем синтагму как последовательность слогов, то ее ритмический рисунок оказывается инвариантным, полностью независимым от темпа речи. Отсюда естественно сделать вывод, что в программе синтагмы ритмически организованными являются слоговые команды, т. е. команды, вызывающие осуществление всего слогового комплекса движений. Развертывание слога в последовательность звуков речи происходит уже по каким-то собственным законам.

Этот вывод можно наглядно представить в виде следующей схемы:



где a , b , c — слоговые команды, a_1 , b_1 , c_1 — те части слогового комплекса движений, которые соответствуют согласным, a_2 , b_2 , c_2 — те части комплекса, которые соответствуют гласным.

Из такого представления, естественно, возникают две группы вопросов. Одна из них касается организации (выработки) ритмической последовательности слоговых команд, а другая — организации движений внутри слогового комплекса. Их рассмотрению посвящены последующие разделы настоящей главы и глава 4.

Следует еще раз остановиться на вопросах темпа речи. Очевидно, что рассуждения о физиологических механизмах темпа следует отложить, по крайней мере пока не будут составлены некоторые представления о том, как могут образовываться в нервной системе ритмические последовательности управляющих команд. Пока мы можем чисто формально определить темп, как скорость следования слоговых команд, как коэффициент, на который умножаются заданные в программе интервалы между слоговыми командами. Необходимо подчеркнуть, что скорость следования команд отнюдь не должна находиться в простых отношениях с физической скоростью отдельных движений. Смещение этих понятий (Stetson, Hudgins a. Moses, 1940) может существенно осложнить понимание вопроса.

Придя к выводу, что темп существует как самостоятельный параметр, определяющий осуществление артикуляторной программы, мы попытались выяснить, насколько произвольно человек может управлять этим параметром. В частности, нас интересовало, насколько постоянен темп у данного испытуемого и

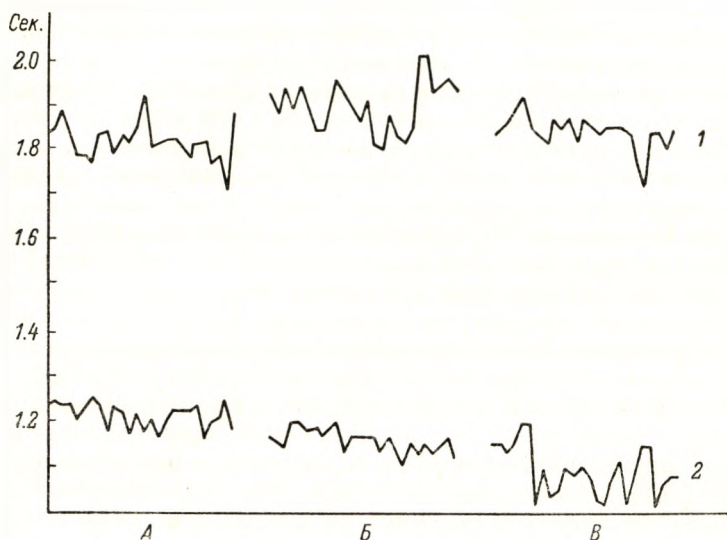


Рис. 3.11. Длительность фразы при повторных произнесениях в разные опытные дни (А, Б, В), разделенные интервалами времени около одного месяца.

1 — при произнесении в полном стиле; 2 — при произнесении в разговорном стиле. По оси абсцисс — последовательные произнесения; по оси ординат — длительность фраз (в сек.).

существуют ли у человека достаточно определенные стандарты относительно того, что такое быстрый и что такое медленный темп. Для выяснения этого у двух испытуемых несколько раз на протяжении нескольких месяцев были записаны серии повторных произнесений той же фразы *Тоня топила баню*. Результаты одного из дикторов приведены на рис. 3.11.

Большое однообразие длительности фразы в разные дни позволило считать, что человек обладает достаточно четкими внутренними стандартами для темпа и способен без специальных извне заданных стандартов удерживать его на постоянном уровне.

3. Флюктуация длительности слогов и более крупных отрезков синтагмы

Результаты, описанные в предыдущем разделе, позволили сделать вывод, что ритмическая структура синтагмы является инвариантной (не зависит от темпа речи) только на уровне слога-

вых и более крупных элементов. Это заставляет считать, что ритмически организованными являются только последовательности слоговых команд; отдельные движения внутри слога, обеспечивающие переход от согласного к гласному, подчиняются своим собственным внутрислоговым законам. В настоящем разделе будет сделана попытка рассмотреть, каким образом может создаваться ритмическая последовательность слоговых команд.

В исследованиях высшей нервной деятельности последовательность движений обычно рассматривается как цепочка, в которой каждое предыдущее движение (или его результат) является сигналом для последующего движения (обзор работ см.: Воронин, 1952). Аналогичное предположение относительно организации последовательности речевых движений было высказано Фэрбенксом (Fairbanks, 1955; Fairbanks a. Guttman, 1958). Он рассматривает речь как последовательность элементарных речевых комплексов. Аfferентные импульсы, возникающие при произнесении каждого предыдущего комплекса (слуховые, тактильные, проприоцептивные) рассматриваются как сигналы, непосредственно вызывающие осуществление следующего комплекса (более подробно эта гипотеза будет рассмотрена в главе 4).

Рассматривая слог как элементарный артикуляторный комплекс (определение, принятое в фонетике, см.: Зиндер, 1960; Stetson, 1951), можно предположить, что артикуляция каждого нового слога является реакцией на сигналы, возникшие при артикуляции предыдущего. Однако такая формулировка требует некоторого уточнения, так как допускает различные толкования.

Одно из толкований высказывания «первый слог является сигналом следующего» состоит в том, что характер первого слога определяет характер следующего слога. Иначе говоря, предполагается, что программа осуществлена по типу матриц переходов, в которых указано, что от такого-то слога (скажем, *то-* в *Тоня*) нужно переходить к такому-то слогу (*-ня*). Очевидно, что работа по такому принципу предполагает необходимость опознания (идентификации) каждого сказанного слога самим говорящим. Если слог будет произнесен или опознан неправильно, то переход к следующему слогу не произойдет или начнутся неправильные переходы. Кроме того, такая система может удовлетворительно работать только в том случае, если элементы (слоги) в последовательности не повторяются. С этой точки зрения очевидно, что вся фраза в целом несомненно не может быть описана одной таблицей переходов, так как повторение во фразах одних и тех же слогов может быть достаточно частым. Поэтому, если думать о матрицах переходов, то их можно в лучшем случае допустить только для слов. Число слогов в словах, как правило, достаточно мало, и повторные логи редки.

Описание последовательности слогов матрицей переходов само по себе еще ничего не говорит о том, чем должно задаваться начало

слога. Предыдущий слог может указывать только направление перехода, моменты переходов могут определяться каким-то внешним устройством, например некоторым тактовым генератором.

Другой, более интересный для нас смысл высказывания «первый слог является сигналом второго» состоит в том, что возникающая при произнесении первого слога афферентная импульсация вызывает с определенным скрытым периодом комплекс движений второго слога. При этом можно допустить, что характер как первого, так и второго слога является для данного механизма несущественным. Какой именно слог осуществится после какого, может определяться как описанной выше матрицей, так и различными другими способами записи программы. Так, если для каждого слова имеется своя подпрограмма и если нервная система обладает возможностью счета слогов в слове, то можно допустить, что слоги в слове просто пронумерованы. Иначе говоря, в программе указано, что первый слог должен быть, например, *то-*, второй *-ни*, третий *-ла*. Роль афферентной импульсации, возникающей при артикуляторных движениях, может в такой схеме свестись только к автоматическому вызову следующего слога.

Такое предположение кажется привлекательным по ряду причин. Во-первых, оказывается не нужным искать какой-то внешний источник, вызывающий движения. Источником данного движения может быть импульсация, вызванная предыдущим. Во-вторых, интервал между слоговыми командами может тогда рассматриваться как скрытый период реакции. Это значит, что ритмический рисунок синтагмы мог бы быть задан в виде последовательности изменений возбудимости нервных центров. Это может приводить к соответствующему удлинению или укорочению скрытых периодов.

Возможным возражением против такой гипотезы является слишком малая величина временных интервалов между началами соседних слогов. Длительность неударного слога при быстрой речи может быть меньше 100 мсек. (случай редукции гласных). Скрытый период простой двигательной реакции на звуковой сигнал находится, как правило, в пределах 110—150 мсек. (Chocholle, 1942a, 1942b; Rao a. Nayak, 1956; Raab, 1962; см. также: Вудвортс, 1950). Скрытые периоды такой чрезвычайно быстрой реакции, как повторение слышимой речи, также, как правило, превышают 100 мсек. (Чистович, Клаас, 1962; Чистович, Клаас, Кузьмин, 1962). Вопрос о возможных задержках реакции на проприоцептивный сигнал рассматривался Рачем (1960). Его интересовал вопрос о возможности коррекции произвольного движения на основании проприоцептивной импульсации, получаемой в процессе осуществления этого движения. Рач пришел к выводу, что задержка между появлением проприоцептивной

импульсации и реакцией на эту импульсацию (моторной коррекцией) должна составлять около 100 мсек.

Таким образом, наименьшая длительность слога близка к предельным минимальным величинам скрытых периодов реакций кортикального уровня.

Альтернативная гипотеза образования ритмической последовательности слоговых команд заключается в том, что команды генерируются в силу собственной ритмической активности каких-то нервных центров. Иначе говоря, допускается существование некоторого генератора сложного ритма, схема которого не включает элементов, находящихся вне центральной нервной системы. С точки зрения существующих нейрофизиологических данных, это также является вероятным. Хорошо известна собственная ритмическая активность дыхательного центра (см. обзор работ: Сергиевский и Иванов, 1931).

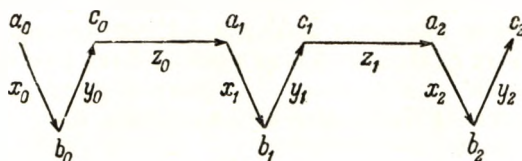
Шеррингтоном и его учениками было показано, что правильные ритмические движения, образующие чесательный рефлекс и рефлекс шага, не нарушаются при полной деафферентации конечностей и, следовательно, определяются автоматической ритмической активностью самих центров спинного мозга (Крид, Денни-Броун, Икклс, Лидделл, Шеррингтон, 1935).

В настоящей работе мы попытались выяснить, какая из двух гипотез является более вероятной, используя для этого статистический анализ флюктуаций длительностей различных отрезков синтагмы.

Способ, избранный для проверки гипотез, основывался на достаточно очевидном явлении моторных неточностей исполнения движений. Хорошо известно, что при повторном вызывании одного и того же рефлексорного движения происходят флюктуации как скрытого периода этого движения, так и силы мышечного сокращения. Колебания скрытого периода обычно связывают с колебаниями состояния возбудимости промежуточных нейронов и, соответственно, с различиями во времени суммации (Рач, 1960; Eccles, 1939). Флюктуации силы мышечного сокращения обуславливаются изменениями числа возбуждаемых мотонейронов, что наблюдается и в случае простейшего моносинаптического рефлекса (Hunt, 1955; Rall a. Hunt, 1956; см. обзор: Hunt a. Perl, 1960).

К сожалению, нам не удалось найти нейрофизиологических данных, которые количественно характеризовали бы флюктуации задержки между центральной командой к двигательному органу и началом движения. Флюктуации скрытых периодов движения при электрическом раздражении моторной области коры (Chang, Ruch, Ward, 1947) слишком велики и явно не могут использоваться для характеристики явления в нормальных условиях двигательного поведения.

Если принять гипотезу о том, что команда начать следующий слог выдается в ответ на афферентную импульсацию, свидетельствующую о начале предыдущего слога (гипотеза 1), то речевой процесс можно изобразить следующей схемой:



где a_0, a_1, a_2 — моменты поступления слоговых команд, b_0, b_1, b_2 — моменты начала соответствующих движений, c_0, c_1, c_2 — моменты поступления в нервную систему афферентной импульсации, свидетельствующей о начале движения.

Обозначим временной интервал $a_i b_i$ ($i=0, 1, \dots$), т. е. длительность задержки исполнения команды, через x_i , интервал $b_i c_i$ ($i=0, 1, \dots$), т. е. время, за которое нервная система получает афферентную импульсацию о начале движения, через y_i , а интервал $c_i a_{i+1}$ ($i=0, 1, \dots$), т. е. скрытый период появления следующей команды, через z_i . Эти интервалы естественно считать случайными.

Согласно схеме, интервал между началом движений первого и началом движений второго слога равняется

$$b_0 b_1 = y_0 + z_0 + x_1;$$

интервал между началом второго и началом третьего слога —

$$b_1 b_2 = y_1 + z_1 + x_2.$$

Так как величины y_0, z_0, x_1 , согласно принятой гипотезе об организации речевого процесса, естественно считать независимыми, то дисперсия интервала $b_0 b_1$ равна

$$D(b_0 b_1) = D y_0 + D z_0 + D x_1;$$

аналогично

$$D(b_1 b_2) = D y_1 + D z_1 + D x_2.$$

Рассмотрим теперь интервал между началом первого и началом третьего слога. Он равен сумме двух вышеупомянутых интервалов

$$b_0 b_2 = b_0 b_1 + b_1 b_2 = y_0 + z_0 + x_1 + y_1 + z_1 + x_2.$$

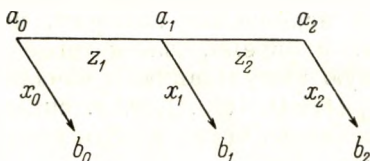
Написанные справа случайные интервалы, очевидно, попарно независимы между собой, за исключением пары z_0 и z_1 . Если темп речи не меняется, то z_0 и z_1 также должны быть независимыми. В случае изменения темпа они должны быть положительно коррелированными. Из этого следует, что дисперсия интервала

b_0b_2 должна быть больше или равна сумме дисперсий интервалов b_0b_1 и b_1b_2 :

$$D(b_0b_2) \geq D(b_0b_1) + D(b_1b_2).$$

Таким образом, если предполагаемая схема организации речевого процесса является правильной, то дисперсия суммы n последовательных слогов во фразе должна быть больше или равна сумме дисперсий составляющих слогов, а длительности соседних слогов должны быть или независимыми, или положительно коррелированными.

Рассмотрим теперь другое предположение, состоящее в том, что заданными являются непосредственно сами интервалы между слоговыми командами и афферентная импульсация не влияет на начало следующего слога (гипотеза 2). Эту гипотезу можно изобразить следующей схемой:



В этой схеме, как и в предыдущей, a_0, a_1, a_2 — моменты поступления слоговых команд, b_0, b_1, b_2 — моменты начала соответствующих движений, x_0, x_1, x_2 — задержки начала движения относительно момента поступления слоговой команды; z_1 и z_2 соответствуют интервалам между командами, их также естественно считать случайными величинами.

Согласно этой схеме, интервалы

$$b_0b_1 = z_1 + x_1 - x_0,$$

$$b_1b_2 = z_2 + x_2 - x_1,$$

$$b_0b_2 = z_1 + z_2 + x_2 - x_0.$$

Соответственно дисперсии этих интервалов равны

$$D(b_0b_1) = Dz_1 + Dx_1 + Dx_0,$$

$$D(b_1b_2) = Dz_2 + Dx_2 + Dx_1,$$

$$D(b_0b_2) = D(z_1 + z_2) + Dx_2 + Dx_0.$$

Разность между суммой дисперсий соседних слогов и дисперсией интервала между началом первого и началом третьего слога будет равна

$$D(b_0b_1) + D(b_1b_2) - D(b_0b_2) = 2Dx_1 + Dz_1 + Dz_2 - D(z_1 + z_2).$$

В случае, если темп сохраняется постоянным, т. е. z_1 и z_2 являются независимыми,

$$D(z_1 + z_2) = Dz_1 + Dz_2.$$

Следовательно, дисперсия интервала между началом первого и началом третьего слога должна быть на $2Dx_1$ меньше суммы дисперсий слоговых интервалов, т. е. должен получиться результат, невозможный в случае первой рассматривавшейся схемы. Если темп меняется, то z_1 и z_2 положительно коррелированы, и результат, который должен получиться, будет определяться размахом этих изменений.

Таким образом, методом сравнения дисперсий слоговых и более крупных отрезков фразы принципиально возможно выбрать одну из двух рассматриваемых гипотез; однако для этого необходимо, чтобы изменения темпа или отсутствовали, или были достаточно малыми. Другое необходимое условие состоит в том, чтобы дисперсия ошибки измерения начала движения была достаточно малой. Подробнее необходимость этого условия будет рассмотрена дальше.

Для экспериментальной проверки описанных выше предположений были использованы осциллограммы многократно повторенных произнесений одной и той же фразы *Тоня топил баню*. Фраза произносилась двумя дикторами, теми же, что участвовали в описанных ранее опытах по изменению темпа. Каждый диктор повторял фразу 150—200 раз в быстром темпе и столько же раз в медленном темпе. На всех полученных осциллограммах было первоначально измерено по одному интервалу (от конца смычки *т* в *Тоня* до начала смычки *н* в *баню*). Затем были построены графики, показывающие изменение данного интервала от произнесения к произнесению.

На каждом графике отыскивался относительно стационарный участок, и из него выбиралось для дальнейшей обработки 100 фраз, в которых не было каких-либо дефектов регистрации. Первичная обработка отобранных 400 фраз состояла в измерении интервалов (в мм) от размыкания смычки *т* в *Тоня* до всех остальных изменений в артикуляции, состоящих или в замыкании, или в размыкании соответствующего контакта. За окончание фразы принималось падение сигнала на ларингофоне. Измерения проводились с точностью до 0.5 мм (8.3 мсек).

Вся дальнейшая статистическая обработка данных проводилась на электронной вычислительной машине Урал-1 по специально составленной программе. Статистическая обработка данных включала следующее:

- 1) определение среднего, дисперсии и средней квадратической ошибки длительности каждого звука речи в данной фразе, а также коэффициента корреляции между длительностями соседних звуков;

2) определение среднего, дисперсии и средней квадратической ошибки интервала, включающего два соседних звука речи (сог-

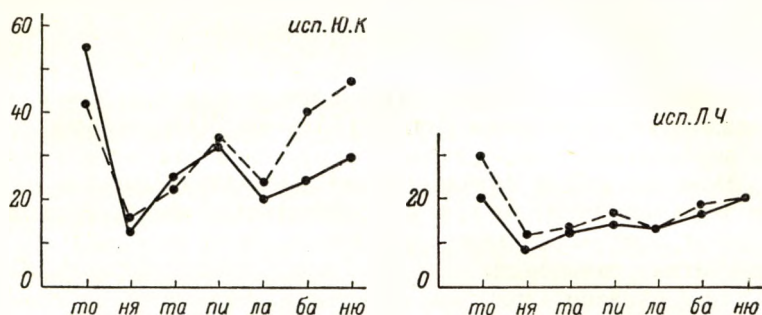


Рис. 3.12. Средние квадратические отклонения длительностей слогов (по оси ординат, в мсек.) во фразе *Тоня топила баню*, произнесенной испытуемыми Ю. К. и Л. Ч. в быстром (сплошные кривые) и медленном (прерывистые кривые) темпе.

ласный—гласный или гласный—согласный), а также коэффициента корреляции между длительностями двух соседних двухзвучных интервалов;

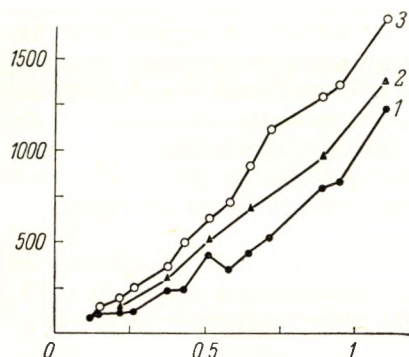


Рис. 3.13. Дисперсии общих интервалов (1), суммы дисперсий длительностей слогов и звука *о* (2) и суммы дисперсий длительностей звуков, входящих в общий интервал (3) для фразы *Тоня топила баню*, произнесенной в быстром темпе (по оси ординат, в мсек.²). (По оси абсцисс — длительность соответствующего общего интервала, в сек.).

3) определение средних, дисперсий и средних квадратических ошибок следующих интервалов: первый интервал состоял из *о* в *Тоня*, второй включал 2 звука (*он*), третий 3 (*оня*) и т. д.; последний интервал включал длительность всей фразы (без смычки начального *т*). Эти интервалы будем дальше для краткости называть общими.

Значения средних квадратических отклонений длительностей слогов приведены на рис. 3.12. У испытуемого Л. Ч. средние квадратические отклонения сравнительно малы (за исключением первого и последнего слога во фразе), около 10—15 мсек. У испытуемого Ю. К. средние квадратические отклонения необоснованно ве-

лики, достигая 40 мсек. Видимо, он недостаточно точно соблюдал темп речи или характер логического ударения, что, есте-

ственно, должно было привести к большому разбросу данных. Это вынуждает нас подходить к этим данным с осторожностью, и все дальнейшие выводы будут основываться преимущественно на результатах испытуемого Л. Ч.

Как уже говорилось, основной задачей являлось сравнение дисперсии крупных отрезков фразы с суммой дисперсий составляющих ее элементов (словов). На рис. 3.13 приведены дисперсии общих интервалов и суммы дисперсий звуков и слогов, входящих в этот интервал, при быстром темпе произнесения фразы испытуемым Л. Ч. Аналогичная картина была получена при медленном темпе произнесения.

Можно видеть, что сумма дисперсий звуков значительно превышает дисперсию общего интервала, в который входят эти звуки. Этого следовало ожидать: исходя из данных, приведенных в предыдущем разделе, мы уже вынуждены были прийти к выводу, что начало следующего слога не связано с моментом перехода от согласного к гласному в предыдущем слоге. В соответствии с этим оказалось, что длительности звуков, входящих в один слог, отрицательно коррелированы. Приводим соответствующие коэффициенты корреляции; выделены те значения коэффициентов корреляции, которые являются существенными (с вероятностью 0.95).

Испытуемый	Темп	То	ня	то	ни	ла	ба	ню
Л. Ч.	Быстро	-0.13	-0.36	-0.13	-0.23	-0.40	-0.26	-0.02
Л. Ч.	Медленно	-0.05	-0.31	-0.11	-0.21	-0.17	-0.33	-0.23
Ю. К.	Быстро	0.29	-0.57	-0.20	-0.29	-0.07	-0.39	0.10
Ю. К.	Медленно	-0.01	-0.25	0.08	-0.17	-0.06	-0.08	0.01

Приводим также значения коэффициентов корреляции между соседними звуками, входящими в различные слоги. Можно видеть, что и в этом случае корреляция является отрицательной.

Испытуемый	Темп	он	ят	он	ил	аб	ан
Л. Ч.	Быстро	-0.22	-0.41	-0.54	-0.51	-0.72	-0.30
Л. Ч.	Медленно	0.08	-0.06	-0.49	-0.23	-0.23	0.17
Ю. К.	Быстро	-0.21	-0.64	-0.63	-0.22	-0.47	0.05
Ю. К.	Медленно	-0.34	0.12	-0.10	-0.08	-0.30	0.02

Очевидно, начало смычки согласного не является сигналом, вызывающим ее прекращение, — случайные флюктуации начала смычки относительно запускающей команды не влияют на момент времени, когда произойдет размыкание. Эти соотношения существенны для понимания организации движений внутри слога, которая будет рассматриваться в главе 4.

В настоящее время нас интересует вопрос, является ли начало артикуляции предыдущего слога сигналом, вызывающим следующий слог. Как уже говорилось, в этом случае должны выполняться следующие условия: 1) дисперсия общего интервала не может быть меньше суммы дисперсий слогов, входящих в этот интервал, 2) между соседними слогами не может быть отрицательной корреляции. Рис. 3.13 показывает, что первое условие не выполняется.

Не выполняется также и второе условие:

Испытуе- мый	Темп	То-ня	ня-то	то-ти	ти-ла	ла-ба	ба-ню
Л. Ч.	Быстро	-0.04	-0.04	-0.27	-0.58	-0.63	-0.04
Л. Ч.	Медленно	0.06	-0.17	0.01	-0.15	-0.28	-0.03
Ю. К.	Быстро	0.42	-0.16	-0.56	-0.28	-0.39	0.09
Ю. К.	Медленно	-0.03	0.34	0.16	0.01	0.06	-0.08

Здесь в столбцах помещены коэффициенты корреляции между указанными слогами.

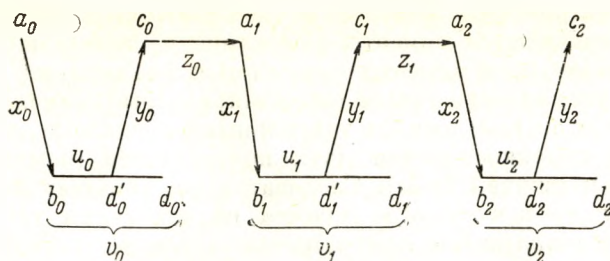
Таким образом, полученные данные свидетельствуют скорее в пользу второй гипотезы и заставляют предполагать, что каждая следующая слоговая команда задается автоматически: для выработки ее не требуется использования афферентной импульсации, возникающей при осуществлении предыдущего слога.

Прежде чем обсуждать, можно ли считать этот вывод окончательным, сопоставим полученные данные с описанными в литературе.

Нам не удалось найти данных, касающихся флюктуаций длительности крупных отрезков синтагмы, однако в некоторых работах приводятся данные о флюктуациях длительности отдельных звуков речи (Kaiser, 1939; Stetson, 1951). Приводимые величины свидетельствуют о том, что относительная ошибка длительности одного и того же звука речи при постоянном темпе произнесения составляет 10—20%. Это вполне согласуется с тем, что наблюдалось на нашем материале. Существенно, что при такой большой относительной ошибке длительности отдельных звуков речи относительная ошибка длительности всей синтагмы составляет всего около 3%.

Возвращаясь к основному интересующему нас вопросу, мы вынуждены признать, что полученные результаты еще не позволяют полностью отвергнуть первую гипотезу. Дело в том, что такие же результаты могут получиться, если момент, принимаемый в эксперименте за начало слога, и момент, принимаемый нервной системой за начало слога, не совпадают и дисперсия ошибки измерения начала движения в эксперименте достаточно велика.

Действительно, если имеет место последний случай, его можно изобразить следующей схемой



На этой схеме a, b, c, x, y, z обозначают то же, что и на стр. 101. d_0, d_1, d_2 — моменты замыкания смычки, принимаемые за начало слога в эксперименте; d'_0, d'_1, d'_2 — моменты, принимаемые нервной системой за начало слога; u_0, u_1, u_2 — ошибки, совершаемые нервной системой при измерении начала слога; v_0, v_1, v_2 — ошибки, совершаемые при измерении начала слога в эксперименте.

Обозначим интервалы $d_0 d_1$ и $d_1 d_2$, принимаемые в эксперименте за длительности первого и второго слогов, через α_1 и α_2 соответственно.

Тогда:

$$\alpha_1 = d_0 d_1 = y_0 + z_0 + x_1 + v_1 + u_0 - v_0,$$

$$\alpha_2 = d_1 d_2 = y_1 + z_1 + x_2 + v_2 + u_1 - v_1,$$

$$\alpha_1 + \alpha_2 = d_0 d_2 = d_0 d_1 + d_1 d_2.$$

Естественно предположить, что интервалы $x_0, x_1, x_2, y_0, y_1, y_2$ попарно независимы между собой и не зависят от интервалов $z_0, z_1, u_0, u_1, u_2, v_0, v_1, v_2$.

Предположим, что интервал от момента начала движения до момента замыкания смычки согласного (v_i) и интервал от момента начала движения до момента, принимаемого нервной системой за начало движения (u_i), не зависят от темпа произнесения фразы, т. е. все интервалы u_0, u_1, u_2 , и v_0, v_1, v_2 попарно независимы между собой и не зависят от z_0, z_1 . (Если это предположение не выполняется, то выкладки принимают довольно сложный вид, и такой случай поэтому рассматриваться не будет). Тогда ковариация между интервалами α_1 и α_2 будет

$$R(\alpha_1, \alpha_2) = R(z_0, z_1) - Dv_1.$$

Если теперь окажется, что дисперсия ошибки измерения начала произнесения слога

$$Dv_1 \geq R(z_0, z_1),$$

то этим и можно будет объяснить невыполнение условий 1 и 2, необходимых для первой гипотезы.

Рассмотрим, какой эффект движения измеряется в эксперименте и какие эффекты могла бы измерять нервная система. С по-

мощью электрических методов в эксперименте с большой точностью определяются моменты замыкания и размыкания смычки языка с нёбом (или моменты смыкания и размыкания губ). Это соответствует данным о начале движения, которые получал бы человек, если бы он пользовался тактильными сигналами, возникающими при соприкосновении двух поверхностей (языка и нёба и верхней и нижней губы). Следовательно, гипотеза о том, что команда следующего слога задается на основании тактильного сигнала, возникающего при замыкании соответствующего контакта при произнесении согласного предыдущего слога, должна быть сразу же отвергнута. Однако человек может измерять начало слога также на основании акустического сигнала. Подъем языка к нёбу и сближение губ вызывают изменение частот формант предшествующих гласных, что воспринимается человеком как начало согласного (Дукельский, 1962; Ohman, 1962). Этот эффект возникает по времени раньше, чем регистрируемое нами начало смычки.

Чтобы проверить, являются ли акустические изменения сигналом начала следующего слога, был поставлен специальный опыт, заключающийся в том, что испытуемый Л. Ч. 14 раз повторил фразу *Тоня топил баню* беззвучно, совершая только движения губами и языком и не прекращая нормального дыхания. В этих условиях артикуляторная картина фразы оказалась измененной, в частности отношение длительности согласного к гласному в слове резко изменилось в пользу согласного. Однако точность осуществления фразы осталась той же самой. Отношение среднего квадратического отклонения общего интервала от конца смычки *т* в *Тоня* до окончания *н* в *баню* к абсолютной длительности интервала находится в пределах 1.7—3.6%, то совпадает с теми величинами, которые наблюдались при естественной громкой речи (3.1% для быстрого темпа и 2.6% для медленного). Это означает, что устранение обратной слуховой связи не только не исключает возможности произнесения (Ladefoged, 1958), но и не уменьшает точности воспроизведения ритмической последовательности слогов.

Таким образом, приходится исключить предположение, что акустические изменения, связанные с началом слога, используются нервной системой как необходимые сигналы для выработки команды начала следующего слога. Остается предположить, что команда начать следующий слог является или ответом на проприоцептивную импульсацию, возникающую при произнесении предыдущего слога (первая гипотеза), или задается независимо от исполнения движений предыдущего слога (вторая гипотеза).

Была сделана попытка оценить, каковы должны быть дисперсии скрытых периодов начала слога при условии, если верной является первая гипотеза. Вычисления производились на основании того, что для первой гипотезы сумма дисперсий скрытых

периодов слоговых движений равна дисперсии общего интервала за вычетом дисперсии интервала от конца смычки *т* в *Тоня* до начала произнесения слога *-ня* и дисперсии ошибки измерения начала слога, следующего за рассматриваемым общим интервалом.

Оказалось, что верхние границы для средних квадратических ошибок скрытых периодов начала слога (для слогов *-ни*, *-ла*, *ба-*, *-ню* и обоих темпов произнесения) — порядка 10—20 мсек.

Если учесть большое число синапсов в дуге проприоцептивного рефлекса, проходящей через кору головного мозга (Рач, 1960), а также неизбежные моторные ошибки и ошибки измерения начала слога первой системой, эти величины кажутся неправдоподобно малыми.

Таким образом, исходя из полученных данных, кажется возможным предпочесть вторую гипотезу и считать, что проприоцептивная импульсация также не является необходимым сигналом для выработки следующей слоговой команды. Однако для окончательного решения вопроса необходимы дополнительные эксперименты, выполненные какими-то иными методами.

4. Влияние фонетического состава на временные характеристики слова (синтагмы)

В предыдущих главах мы пытались рассмотреть различные возможности организации ритмической программы одной конкретной синтагмы. Однако хорошо известно, что человек может воспроизвести ритмический рисунок слова или фразы, используя для этого совсем не те движения (комплексы движений), которые входят в оригинальное слово или фразу. Не обязательно также, чтобы движения были артикуляторными — ритмический рисунок может быть воспроизведен постукиванием пальца, нажиманием на педаль и т. д. Это заставляет предположить, что ритмический рисунок слова (фразы) существует как самостоятельный признак, который может быть общим для большой группы слов (фраз). Если ритмический рисунок действительно существует как нечто самостоятельное, естественно далее допустить, что в нервной системе существуют специальные образования, генерирующие сложные ритмы, используемые для вызывания (пуска) различных конкретных движений.

В лингвистической литературе широко используется понятие ритмического рисунка слова, хотя строгого определения этого понятия, насколько нам известно, не существует. Ритмически однородными считаются слова, имеющие одинаковое количество слогов и одинаковое расположение ударного слога (Гвоздев, 1949). При этом конкретный звуковой состав и количество входящих в слово звуков не принимаются во внимание (*ска́та* — *па́па*, *па́талок* — *нага́ток*).

Из такого понимания ритмической тождественности исходят, видимо, практика и теория стихосложения. В стихотворные ритмы укладываются, как правило, именно слова с одинаковым местом ударения и одинаковым числом слогов, выступая в стихотворной речи как сравниваемые по ритму ряды (Томашевский, 1929). О несущественности для ритмики стихотворной речи конкретного звукового состава слов говорит и тот факт, что любой стихотворный текст может быть передан последовательностью однотипных по звуковому составу ударных и неударных слогов:

та-та́-та-та́-та-та́-та-та́-та́
мой-дя-дя-са-мых-чест-ных-пра-вил.

Однако то, что человек определяет целый ряд фраз как ритмически однородные, еще не может служить окончательным доказательством, что эти фразы действительно образуются одинаково и что в них используется одна и та же ритмическая последовательность слоговых команд. Прямым доказательством было бы установление физического тождества временных характеристик фраз, определяемых как ритмически одинаковые.

Говоря о возможном совпадении временных характеристик фраз, различных по фонетическому составу, мы можем иметь в виду только длительности слогов и более крупных элементов речи. Бесспорно доказано, что практически все звуки речи отличаются друг от друга по длительности. Так, глухие согласные длиннее, чем звонкие, щелевые длиннее, чем смычные, язычные согласные короче, чем губные, мягкие длиннее, чем твердые (Златоустова, 1962; Kaiser, 1934, и др.).

Различия в длительности согласных, отличающихся по месту образования, обычно связывают с различиями в физиологических свойствах работающих при их артикуляции мышечных групп (Kaiser, 1934).

С другой стороны, как было показано ранее, ритмически организованными следует считать только слоговые команды. Распределение артикуляторных движений внутри слога должно быть отнесено за счет частных условий (возможностей) реализации слоговых команд, определяющихся физическими и физиологическими особенностями самих исполнительных органов.

Чтобы выяснить, одинаков ли интервал между слоговыми командами в словах, определяемых как ритмически идентичные, но имеющих различный фонетический состав, была проведена серия измерений. Были подобраны пары слов, отличающихся друг от друга только по одному согласному, находящемуся в середине слова в интервокальном положении (*добить—допить*). Предполагалось, что если ритмическая программа обоих слов одинакова, то интервал между начальным и конечным согласным в обоих словах также должен быть одинаковым. Разница в дли-

тельности интервокальных согласных (смычка *п'* должна быть длиннее смычки *б'*) должна сопровождаться компенсаторным изменением длительности соседних гласных. В опытах участвовали два диктора. Каждый прочел 20 таблиц, составленных из следующих слов: *добить, допить, это дом, это том, забыл, забил, кататься, котятся, это тын, это сын, это пол, это тол, это кол.*¹

Чтобы избежать влияния положения слова, давался различный порядок их следования в разных таблицах. Слова в парах отличались по глухости или звонкости согласного (*это дом — это том, заил — забил*), по твердости или мягкости его (*забыл — забил, кататься — котятся*), по месту образования (*это пол — это тол — это кол*) и по способу образования согласного (*это тын — это сын*).

Полученные результаты не подтвердили нашего предположения: слова, отличающиеся по длительности согласного, во всех случаях отличались также по длительности интервала, включающего этот согласный, и примыкающие к нему гласные. Приводим средние значения разности интервалов для соответствующих пар слов (в мсек.).²

Глухие— звонкие		Мягкие— твердые		Щелевые— смычные		Место артикуляции		
<i>т—д</i>	<i>п—б</i>	<i>б'—б</i>	<i>т'—т</i>	<i>с—ш</i>	<i>п—п</i>	<i>п—к</i>	<i>т—к</i>	
<i>это</i>	<i>до-</i>	<i>за-</i>	<i>ко-</i>	<i>это</i>	<i>это</i>	<i>это</i>	<i>это</i>	
<i>том—</i>	<i>пить—</i>	<i>бил—</i>	<i>тят—</i>	<i>сын—</i>	<i>пол—</i>	<i>пол—</i>	<i>тол—</i>	
<i>это</i>	<i>до-</i>	<i>за-</i>	<i>ся—</i>	<i>это</i>	<i>это</i>	<i>это</i>	<i>это</i>	
<i>дом</i>	<i>бить</i>	<i>был</i>	<i>ка-</i>	<i>тын</i>	<i>тол</i>	<i>кол</i>	<i>кол</i>	
			<i>мать-</i>					
			<i>ся</i>					
Диктор								
Ю. К.	22.5 *	27.2 *	31.3 *	32.4 *	8	9.5	27 *	17.5 *
Л. Б.	13.3	21.7 *	5	—	—	28.3 *	38.3 *	10

Только в том случае, когда слова отличались друг от друга глухостью или звонкостью согласного, достаточно отчетливо наблюдался предполагавшийся эффект компенсации различий в длительности смычки различиями в длительности соседних гласных, причем эта компенсация не была полной. Было обнаружено, что гласный (интервал до момента начала смычки согласного) перед глухим согласным несколько короче, чем перед звонким, а длительность смычки глухого согласного существенно больше длительности смычки звонкого; данные, касающиеся длительности второго гласного, не являются отчетливыми (рис. 3.14).

¹ Сочетания *это кол* и т. д. представляли в произнесении наших дикторов одно фонетическое слово.

² Знаком «звездочка» (*) отмечены те разности, которые являются существенными ($P > 0.95$).

Меньшая длительность гласного, предшествующего глухому согласному, описана Л. В. Златоустовой (1962) для русской речи. Абсолютные значения разности длительностей примерно такие же, что и наблюдавшиеся нами. Сходные различия в длительности (18.2 мсек.) были обнаружены в испанской речи (Zimmerman a. Sapon, 1958). Для английской речи было найдено, что длительность гласного, предшествующего звонкому согласному, очень сильно увеличена по сравнению с длительностью того же гласного перед глухим согласным (Heffner, 1940; House a. Fairbanks, 1953; Zimmerman a. Sapon, 1958; Delattre, 1962). По данным

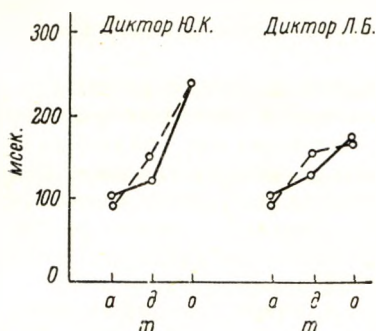


Рис. 3.14. Длительность согласного и примыкающих к нему гласных в словах, различающихся глухотой или звонкостью согласного.

Сплошная линия соответствует словам это дом, прерывистая — словам это том.

Очевидно, что разница в длительности гласного порядка 10—20 мсек. не может служить фонематическим признаком. Естественно думать, что она является побочным эффектом различий в артикуляции глухих и звонких согласных. Нам представляется, что причина эффекта весьма элементарна. Движения губ или языка, совершаемые при смычке глухих согласных, производятся с большей силой. Соответственно скорость движения при глухих согласных является большей, и движущийся орган достигает конечного положения (смычки) за меньший интервал времени. Данные, свидетельствующие в пользу такого механизма явления, приводятся в главе 5 настоящей работы.

Попытки более подробно проанализировать различия, связанные со способом, местом образования, твердостью или мягкостью согласного, не привели к содержательным результатам. Это объясняется прежде всего тем, что различия в длительностях весьма незначительны и для их выявления требуется большое число наблюдений.

Итак, первая серия экспериментов не дала доказательств того, что слова, включающие различные звуки, образуются применением одной и той же ритмической последовательности слоговых команд. Напротив, полученный материал заставляет счи-

тать, что интервал между слоговыми командами меняется в зависимости от того, какие артикуляторные движения (звуки) используются в составе слога. Однако нет еще основания отвергнуть предположение, что в этих словах используется одна и та же исходная ритмическая программа.

Если артикуляция происходит с участием обратных связей, вполне естественно допустить, что осуществление одной и той же программы в какой-то мере видоизменяется в зависимости от конкретных движений, производимых в процессе реализации этой программы.

Подобная точка зрения вполне совместима с существующими нейрофизиологическими данными. Исследованиями Шерринг-

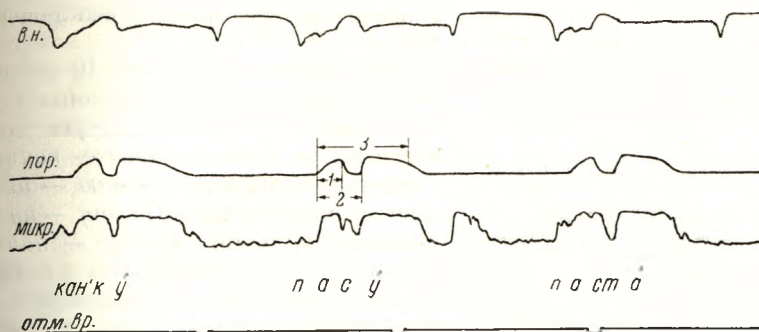


Рис. 3.15. Осциллограмма слов *коньку*, *пасу* и *поста* при «говорении» с закрытым ртом.

Объяснения в тексте.

тона и его учеников было установлено, что правильные ритмические движения спинального животного сохраняются при полной деафферентации конечностей. Отсюда был сделан следующий вполне естественный вывод: «Нервный механизм, лежащий в основе ритмичности чесательного рефлекса и рефлекса шага, является, по-видимому, центральным спинно-мозговым механизмом; для осуществления этих рефлексов не требуется ритмического постуления импульсов извне» (Крид, Денни-Броун, Икклс, Лидделл, Шеррингтон, 1935, стр. 141). Иначе говоря, ритмическая последовательность команд может вырабатываться автоматически за счет работы генератора, полностью находящегося внутри центральной нервной системы. Вместе с тем столь же убедительно было показано, что проприоцептивная импульсация изменяет, регулирует этот ритм.

Очевидно, что лучшим способом проверки предположения об идентичности ритмической программы для слов с разным фонетическим составом являлась бы полная деафферентация артикуляторных органов и исследование речи в этих условиях.

Понятно, что такой эксперимент невыполним. Поэтому мы пошли по другому пути, искусственно заставив человека «говорить» в таких условиях, когда артикуляторные органы должны быть неподвижными.

Испытуемым предлагалось читать таблицы слов с закрытым ртом и прижатым к нёбу языком. В этих условиях единственной работающей артикуляторной системой была система голосообразования. При чтении таблиц наблюдалась ритмически организованная фонация, которая и регистрировалась. Пример осциллограммы приведен на рис. 3.15. Измерялись интервалы от начала

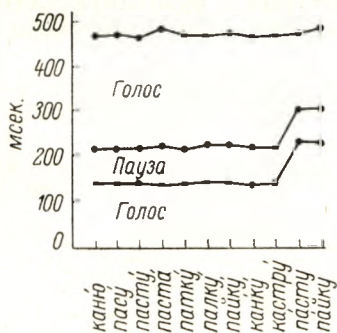


Рис. 3.16. Временные характеристики фонации, осуществляемой при «говорении» различных слов с закрытым ртом.

Объяснения в тексте.

голоса до его первого выключения (1), от начала голоса до его следующего включения (2) и от начала голоса до повторного выключения (3).

Материалом служили 10 таблиц, состоявших из 11 двусложных слов, отличающихся друг от друга количеством согласных (*коню* — *коньку*), качеством согласных (*коню* — *пасу*), местом слогораздела (*посту* — *полку*) и местом ударения (*пайку* — *пайку*). В опыте участвовали два диктора. Полученные данные показали, что в таких условиях все слова таблицы, имеющие одинаковое место ударения, производятся совершенно идентично, т. е. все измеренные интервалы имеют равную длительность, вне

зависимости от качества или количества входящих в слово согласных. Это особенно отчетливо выявилося в данных первого испытуемого: картина произнесения совершенно стандартна, различия между отдельными словами явно случайны, малы по величине и не являются статистически достоверными (рис. 3.16). Данные другого испытуемого являются принципиально аналогичными, однако флюктуации параметров от слова к слову и при повторных произнесениях одного и того же слова оказались несколько большими.

Сравнительное исследование нормального чтения тех же таблиц показало, что влияние качества согласного, находящегося в середине слова, на общую длительность слова¹ весьма мало. При общей длительности слова порядка 600—700 мсек. различия между такими словами как *посту*, *потку*, *полку*, *коньку*, *пайку*

¹ Под общей длительностью слова в данном эксперименте подразумевался интервал от первого включения голоса до его окончательного выключения. Так как первый согласный во всех словах был глухим, эта величина соответствует истинной длительности слова за вычетом первого согласного.

не превышали 20—25 мсек. Зато весьма значительными и статистически существенными оказались различия в длительности слов, имеющих разное количество согласных (*коню* — *коньку*, *пасу* — *посту*):

	«Говорение» с закрытым ртом			Естественная речь		
	<i>Посту</i>	<i>Пасу</i>	Раз- ность	<i>Посту</i>	<i>Пасу</i>	Раз- ность
Диктор Л. Ч. . .	465	473	— 8	714	630	+84
Диктор Л. Б. . .	647	657	—10	703	633	+70

При естественной речи наличие в слове дополнительного согласного отчетливо увеличивает длительность всего слова. Этого не происходит при «говорении» с закрытым ртом — длина слова остается той же самой (уменьшение длительности на 8—10 мсек. не является статистически достоверным).

Таким образом, можно признать, что исходная ритмическая программа для слов, считающихся ритмически совпадающими, действительно является одной и той же. Кроме того, приходится предположить, что артикуляция дополнительного согласного в слове задерживает осуществление программы, тормозит переход к следующей слоговой команде. Это явление представляется чрезвычайно интересным, однако прежде чем переходить к его обсуждению остановимся еще на некоторых особенностях обнаруженных при исследовании «говорения» с закрытым ртом.

Первая особенность состояла в том, что при произнесении всех двусложных слов происходило обязательное прекращение фонации в середине слова, хотя при естественной речи этого не бывает (например *коню*). Во-вторых, длительность интервала между включением голоса и его выключением в середине слова не зависела от характера слогов (*по-сту*, *пол-ку*). Чем объясняются эти явления — пока совершенно непонятно. Казалось бы, исключение деятельности части органов, участвующих в артикуляции, не должно сказываться на работе остальных артикулирующих органов. Однако тем не менее это явно происходит. Может быть, из анализа этого явления возникнут интересные сведения о том, каким образом информация о наличном состоянии артикуляторных органов используется для задания следующего комплекса движений. Ряд данных свидетельствует, что эта информация играет очень существенную роль (Лурия, 1962, Ladefoged, 1958), однако как это осуществляется, до сих пор совсем не выяснено.

Еще одно любопытное явление состояло в том, что слова, различающиеся по месту ударения, совпадали друг с другом по общей длительности, отличаясь, конечно, по распределению длительностей внутри элементов слова (см. *пасту́—па́сту*, *пайкú—пáйку*, рис. 3.16). Это было затем специально проверено на естественном произнесении ряда звукосочетаний (рис. 3.17).

Вернемся теперь к рассмотрению вопроса, каким образом может происходить удлинение слова при наличии в слоге дополнительного согласного или при замене короткого согласного на более длинный. Предполагается, что исходная ритмическая программа слова остается одной и той же.

Одна из причин удлинения слова может быть вполне элементарной, и, если можно так выразиться, периферической по своему происхождению.

Чтобы возникло артикуляторное движение, необходимо не только поступление соответствующей импульсации из высших

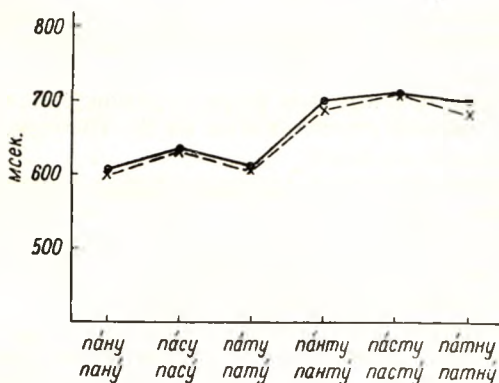


Рис. 3.17. Общая длительность слов с ударением на первом (сплошная кривая) и на втором (пунктирная кривая) слоге.

отделов центральной нервной системы, нужно также, чтобы возбудимость мотонейронов была на достаточном уровне. Из нейрофизиологических работ известно, что состояние мотонейронов изменяется на протяжении движения, что, в частности, связано с изменением проприоцептивной импульсации (см. обзор Hunt a. Perl, 1960).

Логично допустить, например, что мотонейроны, иннервирующие мышцы, смыкающие губы (m. orbicularis oris), должны изменять свою возбу-

димость на протяжении фазы раскрытия губ, причем возбудимость должна нарастать с увеличением раскрытия.

Известно (см. обзор Рач, 1960), что импульсация, поступающая из моторной области коры, вызывает движение не сразу же, но с определенной временной задержкой, скрытым периодом. Часть этой задержки приходится на время суммации в эффекторных центрах (вставочных нейронах). Так как время суммации должно зависеть от исходного состояния возбудимости мотонейронов, естественно ожидать, что задержка между началом центральной импульсации и началом движения будет зависеть от фазы движения, на которую попадет начало центральной импульсации. Если команда, требующая смыкания губ, придется на начальную фазу раскрытия губ, то запаздывание движения смыкания относительно команды будет наибольшим. Если команда придется в то время, когда раскрытие уже достаточно велико, задержка между командой и движением должна быть уменьшена.

Очевидно, что в слоге типа ССГ артикуляция гласного начинается позже, чем в слоге СГ. Если интервалы между слоговыми

командами в последовательностях ССГСГ и СГСГ являются одинаковыми, то условия для реализации второй слоговой команды должны быть различными. Они будут менее благоприятными в случае ССГСГ, так как вторая команда совпадет по времени с более ранней фазой артикуляции гласного. В результате артикуляция второго слога должна запаздывать.

В случае еще более длинных сочетаний должно наблюдаться запаздывание также третьего, четвертого и последующих слогов. Однако при этом величина запаздывания должна постепенно уменьшаться. Соответственно разница в длительности между последовательностями ССГСГСГ... СГ и СГСГСГ... СГ должна стремиться к нулю при увеличении числа слогов.

Для того чтобы проверить, существует ли этот эффект, были исследованы произнесения двух пар фраз: *слово пасту надо произносить четко* (1а) — *слово пасту надо произносить четко* (1б) и *Дина сбежала на улицу* (2а) — *Динка сбежала на улицу* (2б). Фразы были наговорены двумя дикторами. Каждый повторил их 20 раз в следующем порядке: 2б, 1а, 2а, 1б. Приводим длительности соответствующих слов и фраз (в мсек.) (при вычислении длительности слов *Динка* и *Дина* первый согласный не учитывался):

	<i>Пасту</i>	<i>Пасту</i>	Разность <i>пасту—пасту</i>	Фраза <i>1а</i>	Фраза <i>1б</i>	Разность <i>1б—1а</i>
Диктор Л. Ч. . .	401	455	54	2253	2303	50
Диктор Ю. К. . .	521	616	95	2500	2566	66
	<i>Дина</i>	<i>Динка</i>	Разность <i>Динка—Дина</i>	Фраза <i>2а</i>	Фраза <i>2б</i>	Разность <i>2б—2а</i>
Диктор Л. Ч. . .	261	308	47	1496	1553	57
Диктор Ю. К. . .	355	421	66	1655	1716	61

Оказалось, что никакого выравнивания длительностей фраз не происходит — разница между фразами совпадает с разницей между словами, по которым различаются эти фразы.

Таким образом, приходится признать, что изменение длительности слова, вносимое дополнительным согласным, имеет отнюдь не периферический характер. Очевидно, оно является результатом того, что все следующие за данным слогом слоговые команды сдвигаются по времени на определенную величину.

Мы не имеем пока оснований утверждать, что это относится также и к изменению длительности слова за счет разницы в длительностях отдельных звуков речи. Как уже говорилось, в этих случаях различия между разными словами весьма невелики, и для проверки, сохраняются ли они во фразе, требуется очень большой статистический материал. Мы надеемся выполнить такую работу в дальнейшем.

Итак, было получено два рода фактов. Во-первых, оказалось, что слова, определяемые как совпадающие по ритмическому рисунку, действительно имеют одну и ту же исходную ритмическую программу. Во-вторых, выяснилось, что произнесение в слоге двух согласных вместо одного видоизменяет реализацию этой программы, задерживая все следующие слоговые движения на определенное время. На основании первого факта можно сделать вывод, что ритмический рисунок действительно существует как некоторый самостоятельный признак слова (фразы) и что, следовательно, необходимо предположить наличие в нервной системе специальных схем, обеспечивающих генерирование сложных ритмических последовательностей. Как уже говорилось, наличие в нервной системе собственных генераторов простых ритмов очень хорошо известно и не вызывает ни малейших сомнений.

Прежде чем переходить к обсуждению этого вывода и его возможных следствий, упомянем результаты еще одного эксперимента, которые также свидетельствуют, что ритмический рисунок слова существует как отдельная самостоятельная часть артикуляторной программы.

Было случайно замечено, что если просить человека как можно быстрее произносить подряд любые звуко сочетания, то он, как правило, образует «слова», совпадающие друг с другом по ритмическому рисунку. Этот эффект был специально исследован Н. В. Персианцевой (личное сообщение). Испытуемому давалась инструкция говорить как можно быстрее все что угодно и задавалось слово, с которого он должен начать. Произносимые слова записывались на магнитную пленку и затем анализировались с точки зрения их сходства с начальным словом. Одновременная запись артикуляторных движений показала, что слова произносились без пауз (имелись только дыхательные паузы), т. е. ни о каком намеренном придумывании не могло быть речи. О характере материала дает представление набор сочетаний, образованных вслед за начальным словом *протокол*: *протокбл, такабл, матабл, фатабл, ватабл, ратабл, магабл, фатабл, бататбл, микатбл, каратбл, бататбл, валатбл, санатбл, вератбл, каратбл, дарабл, паракбн, каматан, маратан*. Анализ данных показал, что ритмический рисунок слова (число слогов и место ударения) никогда не искажается до самого конца произнесения. Слова сразу же упрощаются с точки зрения структуры составляющих слогов — сочетания ССГ и СССР заменяются на сочетания СГ. Мягкие согласные встречаются гораздо реже, чем твердые; если заданное слово включало мягкий согласный, то следующие за ним слова уже его не содержат. Смена согласных происходит очень легко и как будто бы достаточно случайно.

Данные показывают, что хотя от человека требовалось случайное поведение, он фактически все время оставался в пределах одной и той же ритмической последовательности, изменяя лишь

ее заполнение. Это позволяет предположить, что ритм является основой, скелетом слова и, возможно, играет очень существенную роль как в организации слов в словаре в мозгу человека, так и в процессах опознавания слов.

Если ритмическая картина слова существует как отдельный признак, возможно допустить, что для синтеза (записи) артикуляторной программы слова используются два отдельных блока (системы). В одном из блоков записываются указания только о том, когда нужно совершать движения. Во втором блоке содержатся перечисление необходимых движений и указание их последовательности.

Сделав эти допущения, естественно пойти дальше и предположить, что работа блока, обеспечивающего временной рисунок состоит в выработке ритмической последовательности импульсов, которые не имеют конкретных адресов. Они распространяются достаточно диффузно, направляясь «до востребования», и возбуждают только восприимчивые к ним элементы.

Управление с помощью адресации импульсов «до востребования» проанализировано на примере диффузной нервной системы беспозвоночных (Horridge, 1961). В исследованиях человека и высших позвоночных животных долгое время основное внимание привлекали факты, свидетельствующие о достаточно строгой локализации двигательных функций в центральной нервной системе. Большим числом исследователей было показано, что при раздражении строго определенных участков моторной области коры головного мозга вызываются также строго определенные и тонко дифференцированные двигательные реакции (см. обзоры: Шеррингтон, 1935; Орбели, 1938; Рач, 1960). Это заставляло считать, что импульсация всегда распространяется по специализированным путям, адресуется к строго определенному получателю.

Однако существует ряд фактов, описываемых обычно под общим названием явления иррадиации, свидетельствующих, что импульсы, которые, казалось бы, должны направляться к строго определенному органу, на самом деле могут возбуждать самые различные группы мышц. Это явление было описано К. П. Кунстман и Л. А. Орбели (1924). Деафферентировав заднюю конечность собаки, авторы обнаружили, что через некоторое время после операции «проявились явления широкой иррадиации возбуждения в центральной нервной системе. Деафферентированная конечность, лишенная контроля над собой, представляла собой как бы флюгер, который... принимал участие во всех двигательных актах соответствующими движениями». На основании этих данных Л. А. Орбели (1949) сделал вывод, «что безграничная иррадиация возбуждения имеет место всегда и что наша мозговая масса по существу своему является диффузно-проводящей системой и сохраняет в скрытой форме свойства диффузной нервной системы».

Интерес к диффузным связям внутри нервной системы резко возрос за последние годы в связи с изучением функций ретикулярной формации.

Итак, представление об адресации «до востребования» в нервной системе человека с физиологической точки зрения является достаточно правдоподобным.

Однако чтобы импульсы, адресованные «до востребования», привели к возникновению строго определенных движений, необходимы очень точные указания в артикуляторной программе, какие именно нервные элементы должны быть к ним восприимчивы в данный момент времени.

Кроме того, должна быть также указана последовательность смены этих элементов, так как следующий элемент может быть возбужден только после того, как был возбужден предыдущий.

На основании исследований А. Р. Лурия (1947, 1962), приходится думать, что нервные элементы, определяющие комплекс движений слога (фонемы), расположены в коре головного мозга. Отсюда приходится предположить, что импульсы генератора сложного ритма (где бы он ни находился) должны поступать в кору

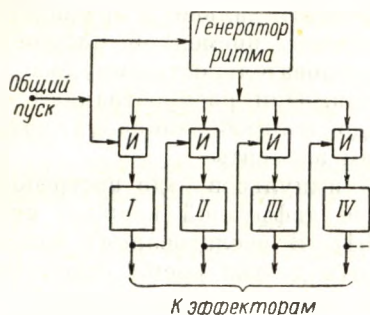


Рис. 3.18. Схема образования слова.

Объяснения в тексте.

головного мозга и вызывать возбуждение соответствующих элементов.

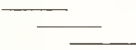
Простейшая логическая схема, которая, казалось бы, могла обеспечить генерирование слова, изображена на рис. 3.18. В этой схеме «общий пуск» обозначает поступление команды о начале произнесения слова, все слоговые элементы (I—IV), которые должны участвовать в его произнесении, последовательно соединены друг с другом через схемы совпадений (И). Для возбуждения второго и каждого следующего элемента необходимы два условия: возбуждение предшествующего ему элемента и поступление импульсов от генератора ритма. От каждого слогового элемента направляются командные импульсы к эффекторам. То, что элементы в данной схеме соответствуют слогам, а не фонемам, не имеет принципиального значения. Можно представить несколько более сложную схему, в которой все элементы, соответствующие фонемам каждого данного слога, будут соединены параллельно и, следовательно, будут возбуждаться импульсом от одного и того же слогового элемента. Обсуждение этого вопроса не имеет смысла, пока не определено, как организован

слог. Вопрос же об организации слога будет рассмотрен в следующих главах.

Мы пока ничего не можем сказать о том, где локализуется предполагаемый генератор ритмов. Возможно, что он должен быть каким-то образом связан с мозжечком (известно, что одним из проявлений мозжечковых нарушений является так называемая скандированная речь). Остановимся только на вопросе, какова может быть логическая схема, используемая для генерации сложной ритмической последовательности.

Пользуясь грубыми техническими аналогиями, можно представить себе, что для создания схемы используется набор элементов временной задержки. Каждый элемент после поступления импульса на его вход через определенный интервал времени образует импульс на выходе. Если имеется набор таких элементов, то возможно образовать нужную последовательность импульсов по крайней мере двумя способами. В одном случае все элементы могут быть соединены параллельно, все они будут возбуждаться одним начальным импульсом, одновременно поданным на все входы. Естественно, что при этом сигнал, появляющийся на выходе каждого отдельного элемента, не будет зависеть от того, появился ли сигнал на выходе другого элемента. В другом случае элементы могут быть соединены последовательно, выходной сигнал каждого предыдущего элемента будет служить входным сигналом для следующего. Так как наши данные показали, что введение дополнительного согласного в слог приводит к сдвигу по времени всех следующих слогов фразы, приходится предположить, что более вероятным является последовательное соединение.

В заключение следует подчеркнуть, что реальная схема организации артикуляции несомненно должна быть значительно более сложной. Однако нам кажется, что упрощенно схематическое рассмотрение явлений представляется полезным хотя бы для того, чтобы лучше выявить существующие сложности.



Г Л А В А 4

ОРГАНИЗАЦИЯ СЛОГА

Для того чтобы команды, образующие ритмическую последовательность, вызывали именно те движения, которые должны быть в данном слове, необходимо, чтобы к моменту пускового импульса уже было установлено, какой именно слог следует произнести. Заранее очевидно, что смена состояний нервных центров должна происходить с частотой не меньшей, чем частота слогов. Однако движения внутри самого слога также являются последовательными. Естественно, возникает вопрос, необходима ли для создания внутрислоговой последовательности движений также последовательность указаний о движениях из высших управляющих уровней, или последовательность движений может получаться автоматически в силу особенностей организации самих исполнителей и непосредственно управляющих ими центров.

Исследованиями Шеррингтона (Крид, Денни-Броун, Икклс, Лидделл и Шеррингтон, 1935) было показано, что последовательность движений может возникать за счет цепных проприоцептивных рефлексов, т. е. в силу того, что антагонистические мышцы, управляющие ими спинальные центры и реагирующие на движения проприоцепторы образуют систему с обратными связями, обладающую автоколебательными свойствами (Випер, 1958). Однократное воздействие извне на такую систему может приводить при определенных условиях к образованию ритмической последовательности движений. Понятно, что воздействием, внешним по отношению к системе, в равной мере может считаться как раздражение из внешней среды, так и импульсация, поступившая из высших управляющих отделов нервной системы.

1. Слог типа СГ

Рассмотрим сначала комплекс движений, осуществляемых при произнесении элементарного слога типа СГ, и постараемся выяснить, какие чисто физиологические отношения должны

существовать между отдельными движениями этого комплекса. Возьмем в качестве примера слог *ту* и перечислим движения, которые требуются для артикуляции *т* и *у*. К сожалению, за недостатком физиологических данных придется ограничиться весьма неполным перечнем:

<i>т</i>	<i>у</i>
Подъем кончика языка до соприкосновения с нёбом и зубами.	Опускание кончика языка.
	Подъем спинки языка и оттягивание его назад.
	Выпячивание губ (движения огубленности).
Подъем нёбной занавески.	Сведение голосовых связок.
Разведение голосовых связок.	Выдох.
Выдох.	

Можно видеть, что некоторые из движений, требуемых гласными и согласными, по чисто физическим причинам не могут осуществляться одновременно.

К таким движениям относятся подъем (*т*) и опускание (*у*) кончика языка, сведение и разведение голосовых связок. Важно, что осуществление этих движений происходит в результате работы систем, образованных мышцами-антагонистами. Подробное описание анатомо-физиологических отношений между мышцами, участвующими в речеобразовании, содержится в статье Арнольда (Arnold, 1957). Электромиографическое исследование деятельности мышц гортани (Faaborg-Andersen, 1957) показало, что усиление электрической активности мышц, приводящих голосовые связки, происходит параллельно с угнетением активности отводящих мышц. Это свидетельствует о том, что центральная импульсация, требующая голосообразования, действует не изолированно на мотонейроны приводящих мышц, а на целостную физиологическую систему, координирующую работу антагонистов. Как уже говорилось, подобные системы способны при определенных условиях генерировать периодическую последовательность движений в ответ на однократное воздействие.

Антагонистические отношения легко обнаруживаются при электромиографическом исследовании мышц, обеспечивающих смыкание и раскрытие губной щели. Параллельная запись биопотенциалов с *m. orbicularis oris* и *m. depressor labii inferioris* при произнесении звукосочетания *а пана* и *а пата* показала, что одновременно с появлением активности *m. orbicularis oris* (смыкание губ при *п*) наблюдается торможение активности *m. depressor labii inferioris*. Примеры осциллограмм приведены на рис. 4.1.

Обратимся снова к перечню артикуляторных движений. Можно видеть, что в нем указан также ряд движений, которые мы условно назовем одиночными. Они характеризуются тем, что определены

только в отношении одного из элементов слога (гласного или согласного) и записаны лишь в одном из столбцов таблицы. В правилах артикуляции другого элемента не фигурирует ни это движение, ни движение, его исключающее. К таким одиночным движениям в системе русского языка относятся, в частности, опускание нёбной занавески (назализация) и выпячивание губ (огубленность). Именно одиночные движения наиболее интересны с точки зрения выяснения того, задается ли весь комплекс арти-

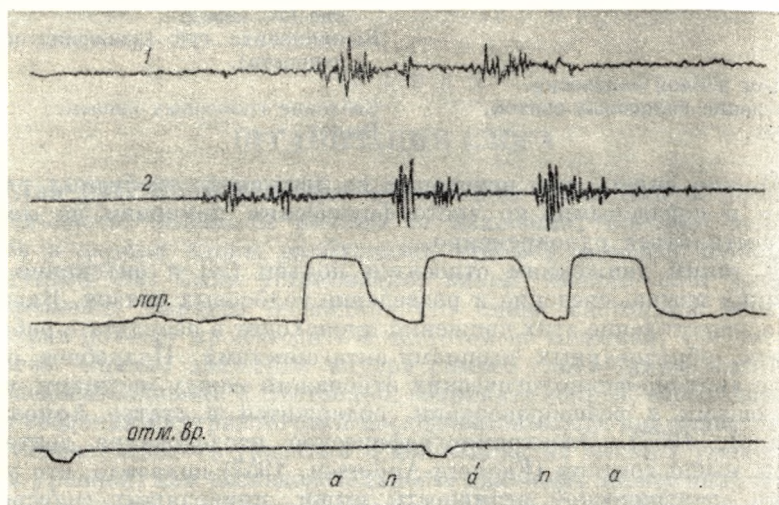


Рис. 4.1. ЭМГ *m. orbicularis oris* (1) и *m. depressor labii inferioris* (2) при произнесении звуко сочетания *a n a n a*.

куляции слога одновременно, или же слог образуется в результате последовательных указаний из высших управляющих уровней. Антагонистические движения не являются полезными для выяснения этого вопроса, так как, во-первых, они физически не могут осуществляться одновременно, и, во-вторых, потому что и при одновременном задании этих движений они могут реализоваться как последовательные (Орбели, 1938).

Тот факт, что согласный, непосредственно предшествующий огубленному гласному, сам является огубленным, известен очень давно и упоминается во всех фонетических руководствах (Богородицкий, 1909; Брок, 1910; Зиндер, 1960). Фуджимура (Fujimura, 1961a) исследовал движения губ и нижней челюсти во время произнесения интервокальных язычных согласных методом киносъемки при стробоскопическом освещении с частотой 60 кадров в секунду. Трое испытуемых произносили английские слова (*veto*, *echo* и т. д.), в которых первый и второй гласный требовали

существенно различной губной артикуляции. Результаты показали, что во время смычки язычного согласного конфигурация губ изменяется соответственно требованиям следующего гласного. Изменение конфигурации губ может начинаться еще до начала согласного.

Таким образом, если основываться на движениях губ, то следует признать, что движения, требуемые гласным, задаются одновременно с движениями предшествующего согласного. Представляется интересным выяснить, пользуясь литературными данными, распространяется ли это также и на другие компоненты артикуляции.

В языках, где нет фонематического противопоставления твердых и мягких согласных, «знание» артикуляторными органами следующего гласного во время произнесения согласного отчетливо проявляется в зависимости артикуляции *k* и *g* от характера гласного. Это выражается, в частности, в том, что спектральные характеристики взрыва этих согласных оказываются различными при разных гласных (Halle, Hughes, Radley, 1957).

Лейдфоджед (Ladefoged, 1957) методом исследования палатограмм установил, что конфигурация контакта языка с нёбом при произнесении согласных существенным образом зависит от характера следующего гласного.

Менцерат и де Лацерда (Menzerath u. de Lacerda, 1933) нашли, что опускание нижней челюсти, характерное для артикуляции гласного, может начинаться на самом начальном участке согласного (согласный осуществляется в фазе опускания челюсти).

Фуджимура (Fujimura, 1961b) при кинематографическом исследовании артикуляции губных согласных обнаружил, что опускание нижней челюсти начинается задолго до момента размыкания губной смычки.

Таким образом, уже на основании имеющихся литературных данных можно считать, что в слогах типа СГ все движения гласного, не противоречащие артикуляции согласного, начинаются с самого начала слога. Иначе говоря, кажется возможным сформулировать правило, состоящее в том, что все движения, требуемые слогом СГ, осуществляются одновременно, за исключением тех движений, которые являются антагонистическими.

Интересно отметить, что случай, когда два одновременно задаваемые комплекса движений частично реализуются как последовательные, может быть создан экспериментально. Такой пример был подробно разобран Л. А. Орбели (1938) при обсуждении возможных физиологических механизмов образования ритмических последовательностей движений.

Если на спинальное животное одновременно подействовать двумя раздражениями, одно из которых рефлекторно вызывает комплекс движений, включающий разгибание конечности, а другое — комплекс движений, включающий сгибание конечности,

животное в ряде случаев начинает совершать альтернирующие, попеременные сгибательные и разгибательные движения. По мнению Л. А. Орбели, этот эффект является следствием периодической смены торможения и возбуждения в центрах спинного мозга и не требует для своего возникновения участия проприоцептивной импульсации.

Мы не беремся утверждать, что аналогия между слогом СГ и двумя одновременно вызываемыми частично антагонистическими рефлексам имеет принципиальный характер, для нас пока существенно, что доказана физиологическая возможность последовательной реализации одновременно заданных антагонистических движений.

2. Сложные слоговые конструкции. Исследование движений огубленности

Слоги типа СГ составляют сравнительно небольшую часть всех слогов, используемых в речи. Считается, что в русском языке возможны следующие слоговые конструкции: Г, СГ, ССГ, СССГ, ГС, ГСС, ГССС.

Нерешенным, однако, является вопрос, когда какая из этих конструкций реально имеет место, на какие слог следует разделить слово, имеющее, например, вид СГССССГСС (слово *конструкт*). Если исходить только из принятых возможными слогов, то его можно разделить на два слога четырьмя различными способами. Чтобы избежать неопределенности, вводятся специальные правила слогаделения, например, правило возрастающей звучности (Jespersen, 1904; Аванесов, 1956), которые, однако, не являются ни общепризнанными, ни достаточно теоретически обоснованными (обсуждение вопросов слогаделения см.: Щерба, 1955; Аванесов, 1956; Зиндер, 1960).

Отсутствие однозначного решения вопроса заставляет, казалось бы, допустить, что все более сложные конструкции, чем слог СГ, уже не являются единицами в артикуляторной программе человека, и что слогов во фразе столько, сколько имеется согласных (согласный представляет собой редуцированную форму от СГ). Однако против этого говорит тот факт, что любой человек легко воспроизводит стихотворную фразу, включающую сложные слоговые конструкции, ритмической последовательностью слогов СГ, число которых соответствует числу гласных в оригинальной фразе, но отнюдь не числу согласных. Кроме того, против этого говорит и наблюдавшийся нами факт совпадения временных рисунков слогов СГСГ и СГССГ при произнесении их с участием только одной фонации (с закрытым ртом). Эти факты заставляют думать, что импульсы ритма действительно возбуждают сложные слоговые комплексы, которые, включая один, и только один, гласный, могут содержать разное число согласных.

Если слог, включающий несколько согласных, задается набором одновременных команд, относящихся ко всем элементам этого слога, естественно ожидать, что те из движений, которые не противоречат друг другу, будут осуществляться одновременно. Некоторые данные, позволяющие предполагать, что это на самом деле является правилом, приводятся в работе Траби (Truby, 1959). Пользуясь методом кинорентгена, автор обнаружил, что в сочетаниях типа *ple* движения языка, соответствующие артикуляции *l*,

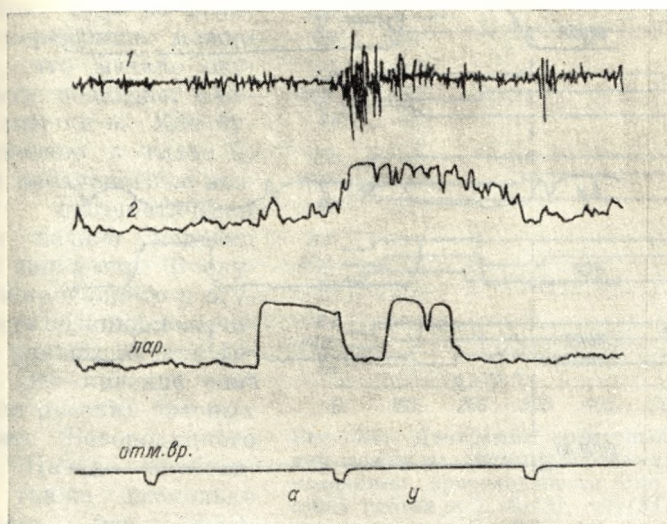


Рис. 4.2. ЭМГ *m. orbicularis oris* (1) и ее огибающая (2) при произнесении гласных *a* и *y*.

осуществляются одновременно с движениями губ, артикулирующих *p*. Автор приходит к выводу, что согласный комплекс вначале является одновременным и только потом становится последовательным.

Так как движения огубленности в русском языке используются только для различения гласных, они не противоречат движениям согласных и принципиально могут сочетаться с артикуляцией любого согласного.

Поэтому, если слог типа ССГ или СССГ задается как одновременная пачка указаний артикуляторным органам, естественно ожидать, что огубленность должна начинаться одновременно с началом артикуляции первого согласного.

Для проверки этого предположения были исследованы движения огубленности в различного рода звукосочетаниях и осмысленных словах.

К сожалению, попытки найти надежный метод электрической или фотоэлектрической регистрации движений губ при огубленности не привели к удовлетворительным результатам. Поэтому мы вынуждены были остановиться на регистрации токов действия мышц (*m. orbicularis oris*) и на косвенной оценке механического эффекта движения губ путем измерения сопротивления между электродами, располагаемыми на губах (описание методики см. в главе 2).

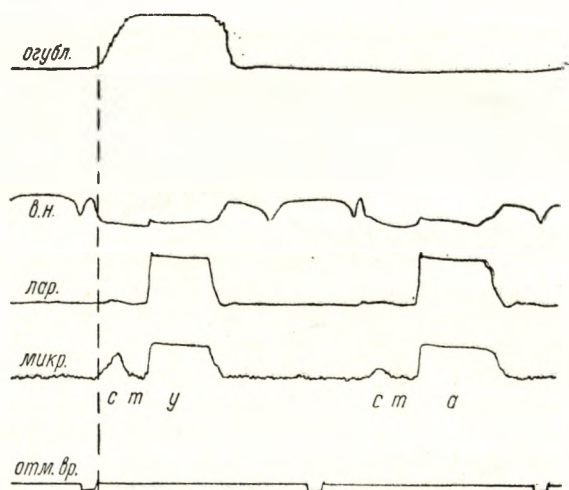


Рис. 4.3. Осциллограмма слогов *сту* и *ста*.

Использование этих методов в связи с наличием электрических артефактов не допускало одновременной регистрации динамической палатограммы. Мы могли параллельно с огубленностью записывать только наличие или отсутствие голоса (сигнал с ларингофона), скорость тока воздуха через нос и сигнал с микрофона. Этими параметрами мы и пользовались для установления моментов начала согласных.

На рис. 4.2 показан эффект огубленности, зарегистрированный электромиографическим способом. На рис. 4.3 показан тот же эффект, зарегистрированный в виде изменения сопротивления между электродами, расположенными на нижней губе. Сравнение осциллограмм показывает, что определение момента начала эффекта является более простым, если пользоваться измерением сопротивления. Поэтому основные измерения были сделаны с помощью последнего метода.

Исследование было начато со слогов *у*, *ту*, *сту*, *нту*, *дну*. Один диктор повторил эти сочетания 11 раз. Полученные результаты (средние из 11 произнесений) представлены на рис. 4.4.

Из графика следует, что в слогах *сту*, *нту*, *дну* огубленность начинается практически одновременно с началом первого согласного. В сочетании *сту* интервал между началом огубленности и началом подъема нёбной занавески (уменьшение скорости тока воздуха через нос) составляет 0.0 ± 9.7 мсек. В сочетаниях *нту* и *дну* интервал между началом огубленности и началом голоса составляет для *нту* 20.0 ± 7.8 мсек., для *дну* 15.0 ± 11.2 мсек. В случае *ту* огубленность начинается раньше начала подъема нёбной занавески на $62 \pm$

± 12 мсек. Судя по величине опережения, можно думать, что начало огубленности совпадает с началом смычки *т*. Как будет показано в главе 5, смычка начального *т* возникает систематически раньше начала подъема нёбной занавески. В случае изолированного *у* огубленность возникает значительно раньше начала фонации. Это явление описано для русских гласных в работе Богородицкого (1909). Начало огубленности также несколько опережает (на $28.8 \pm$

± 9.5 мсек.) прекращение тока воздуха через нос, свидетельствующее о смыкании голосовых связок. По данным Шимизу (Shimizu, 1961), интервал времени от начала смыкания голосовых связок до начала фонации составляет 80—150 мсек. в зависимости от характера гласного. Существенная разница во времени между появлением электрической активности мышц гортани и началом фонации была обнаружена Катсуки (Katsuki, 1950) и Фаборг-Андерсеном (Faaborg-Andersen, 1957).

Установив, что эффект огубленности начинается одновременно с первым из согласных в слоге ССГ и, следовательно, относится ко всему слогу в целом, мы могли далее принять, что начало огубленности является объективным показателем начала огубленного слога.

Это давало возможность экспериментально исследовать вопрос о слоговой границе. Предположим, что мы имеем дело с двумя типами звукосочетаний: $СГС_1/C_2Г$ и $СГ/C_1C_2Г$, где первый гласный является не огубленным (*а*), а второй — огубленным (*у*). Логично ожидать, что эти два типа звукосочетаний будут отли-

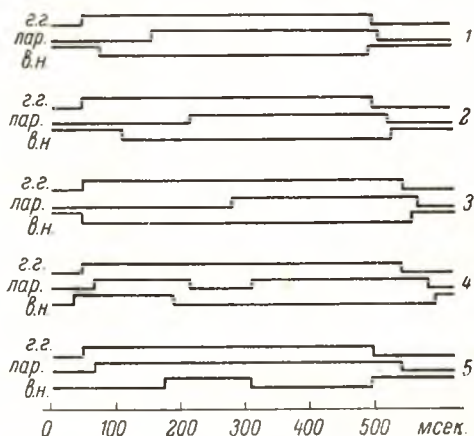


Рис. 4.4. Диаграммы временных отношений между различными артикуляторными явлениями, происходящими при произнесении слогов *у(1)*, *ту(2)*, *сту(3)*, *нту(4)*, *дну(5)*.

чаться друг от друга по задержкам начала огубленности от начала звукосочетания. В первом случае задержка должна быть большей, так как C_1 относится к первому слогу и, следовательно, не может быть огубленным. Во втором случае задержка должна быть меньшей, так как C_1 относится к огубленному второму слогу. Если наше допущение о том, что огубленность относится ко всему слогу в целом, является правильным, то задержки огубленности в сочетаниях типа $СГ/С_1C_2Г$ должны быть равны задержкам огубленности в сочетаниях типа $СГ/С_1Г$.

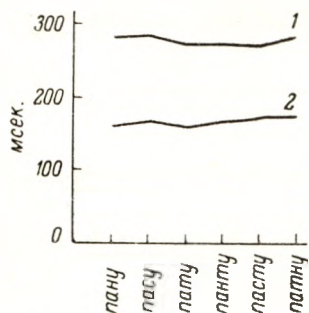


Рис. 4.5. Интервал времени между началом первого гласного в слове (начало фонации) и началом огубленности при ударе на первом (1) и на втором (2) слоге.

Первая проверка этих предположений проводилась на материале бессмысленных звукосочетаний: *пану, пасу, пату* (тип $СГ/С_1Г$), *пасту, пнтну* (тип $СГ/С_1C_2Г$) и *панту* (тип $СГC_1/C_2Г$). Двое дикторов читали таблицы, составленные из указанных звукосочетаний. Каждое звукосочетание фигурировало в таблицах в двух вариантах: с ударением на первом и на втором слоге. Каждый вариант повторялся в таблицах по 10 раз.

При обработке материала за начало звукосочетания было принято начало *а* в *па* (появление фонации). Были вычислены средние величины задержек начала огубленности от начала *а* (рис. 4.5). Из графика следует, что величина задержки не зависит от типа звукосочетания ($СГ/С_1Г$, $СГ/С_1C_2Г$ или $СГC_1/C_2Г$). Существенным оказывается только местоположение ударения. Если ударение падает на первый слог, задержка огубленности возрастает.

Так как задержки огубленности в сочетаниях $СГ/С_1Г$ совпадают с задержками огубленности в сочетаниях $СГ/С_1C_2Г$, можно утверждать, что момент поступления слоговой команды второго слога не зависит от того, один или два согласных входят в этот слог. Отсутствие увеличения задержки огубленности в сочетаниях, где, согласно лингвистическим критериям, слоговая граница проходит между согласными ($СГC_1/C_2Г$), допускает, очевидно, только одно толкование. А именно, приходится предположить, что принятое лингвистами слоговое деление не относится к артикуляторному уровню. Все сочетания $СГC_1C_2Г$ в артикуляторной программе представляют собой два слога, где первый слог кончается гласным ($СГ/С_1C_2Г$). Так как такое заключение является весьма существенным (из него вытекает ряд важных следствий), необходимо было проверить, подтверждается ли полученный нами факт на материале произнесения осмысленных слов.

Для этой цели сравнивались задержки начала огубленности от начала первого гласного в словах следующих двух групп: 1) *па|су по|сту по|тку пол|ку пай|ку*, 2) *ко|ню конь|ку*.

В опытах участвовали два других диктора. Возможность сравнения их с дикторами, произносившими звукосочетания, обеспечивалась тем, что в обеих таблицах фигурировали слова *пасу* и *посту*. Средние величины задержек огубленности от начала первого гласного в мсек. были следующими:

	Диктор Ю. К.	Диктор М. Л.
<i>па сѹ</i>	111	258
<i>по стѹ</i>	111	255
<i>по ткѹ</i>	113	175
<i>пол кѹ</i>	117	163
<i>пай кѹ</i>	120	215
<i>ко нѹ</i>	101	163
<i>конь кѹ</i>	90	206

Можно видеть, что данные диктора Ю. К. оказались достаточно определенными. Они с несомненностью свидетельствуют, что положение слогораздела не имеет значения. Величины диктора М. Л. очень сильно колеблются, в силу чего их нельзя считать доказательными. Большой разброс данных обусловлен недостаточным совершенством методики регистрации эффекта огубленности (см. главу 2).

Итак, полученные данные позволили принять гипотезу о том, что конструкция слога на артикуляторном уровне является достаточно простой; слог оканчивается на гласный, группа согласных, предшествующих гласному, относится к тому же слогу, что и этот гласный. В следующих разделах настоящей главы будут описаны результаты проверки этой гипотезы другими методами.

Считая появление огубленности началом слога, мы пришли к выводу, что задержка второго слога относительно начала первого одна и та же вне зависимости от того, сколько согласных входит во второй слог и каков его конкретный фонетический состав. Это соответствовало разобранному в третьей главе представлению, что момент посылки слоговой команды не зависит от конкретной характеристики слога.

Приняв на основании эффекта огубленности, что слоговые команды второго слога в сочетаниях СГСГ и СГССГ поступают через один и тот же интервал времени от начала первого слога, мы могли теперь определить, есть ли разница в задержках начала реализации согласного в этих сочетаниях. Это можно было сделать весьма просто, используя пары звукосочетаний (*пасу — пасту*, *пату — патну*, *пану — панту*) и подсчитав задержки между началом фонации (началом *а*) и началом какого-либо из эффектов, характеризующих произнесение согласного. В качестве таких эффектов было выбрано прекращение фонации (начало *с* и начало *т*)

и увеличение скорости тока воздуха через нос (начало *н*). Полученные данные приведены ниже в форме разностей (в мсек.) между задержкой согласного в сочетании СГСГ и задержкой его в сочетании СГССГ:

Дикторы	<i>пасу— пасту</i>	<i>пáсу— пáсту</i>	<i>пату— патну</i>	<i>пáту— пáтну</i>	<i>пану— панту</i>	<i>пáну— пáнтну</i>
Л. Б. . . .	12	17	17	18	25	41
Л. Ч. . . .	12	21	2	35	20	12
Ю. К. . . .	6					
М. Л. . . .	17					

Из этих данных следует, что артикуляция первого согласного в слоге ССГ начинается немного раньше, чем артикуляция согласного в слоге СГ, однако разница эта настолько мала, что вполне может быть отнесена за счет различий в условиях реализации команды, посланной в одно и то же время.

3. Сложные слоговые конструкции. Временные отношения внутри комплекса согласных

Если группа ССГ представляет собой единый слоговой комплекс, осуществляемый в результате применения одной пусковой слоговой команды, естественно поставить вопрос, какова организация движений внутри этого комплекса. Очевидно, что до того как пусковая команда подействует, все три описывающие слог фонемы должны быть заранее известны, т. е. должны существовать как одновременные указания относительно конкретной фонетической структуры этого комплекса. Тогда должны быть какие-то специальные механизмы, обеспечивающие последовательность согласных внутри слога при одновременном указании этих согласных высшими управляющими уровнями. Пытаясь разобраться в этом вопросе, мы обратились к исследованию временных отношений между губными и язычными смычками в сочетаниях из губных и язычных согласных. Так как движения губ и языка не являются антагонистическими движениями, никаких физических препятствий к их одновременному осуществлению, естественно, не должно быть.

Исследовались следующие сочетания согласных; *пт, пс, пн, бд, бн, мн, мт, мд, мс, тп, тм, дб, дм, нм, нп, нб, сп, см*. Была составлена таблица из 35 фраз, в которых эти сочетания встречались как внутри слова (следовали за ударным гласным или предшествовали ударному гласному), так и на стыке двух слов (например, *раб долга*). Вся таблица содержала 41 интересующее нас сочетание. Таблица была прочитана по 4 раза двумя дикторами (два раза в разговорном и два раза в полном стиле). Регистрировались следующие параметры: контакты губ, контакты языка (*к. (1+2), к. 3*), скорость тока воздуха через нос, наличие или от-

существование голоса. Пример осциллограммы фразы приведен на рис. 4.6.

Основная часть обработки материала заключалась в измерении длительности смычки первого согласного и в измерении интервала времени от начала смычки первого согласного до начала смычки второго согласного. В случае язычных смычных согласных измерение производилось по $\kappa.(1+2)$, в случае согласного c — по $\kappa.3$.

Результаты измерений показали, что, как правило, смычка второго согласного начинается на территории смычки первого (рис. 4.6, г.г. и $\kappa.(1+2)$ при сочетании $бн$). При разговорном стиле речи это явление наблюдалось в 136 случаях из 161 (84%), при речи в полном стиле — в 129 из 162 (79%). Существенной разницы между дикторами обнаружить не удалось.

Переслаивание смычек губных и язычных согласных было впервые обнаружено при исследовании русской речи (Kaiser, 1929). С помощью пневмографической записи было найдено, что во многих случаях нет никаких признаков взрыва между двумя согласными, общая длительность суммарной смычки двух согласных может быть существенно меньше по величине суммы длительностей смычек каждого из согласных, произносимых в интервокальном положении.

Таким образом, можно утверждать, что эффекторные органы уже располагают сведениями о втором согласном в то время, когда осуществляется смычка первого. Кроме того, из полученных данных следует, что в правилах координации движений слога не содержится требования задержки смычки второго согласного до тех пор, пока первая смычка не кончится. Далее была сделана попытка выяснить, являются ли смычки последовательными в том смысле, что смычка второго согласного не может начаться раньше, чем смычка первого. Ответ на этот вопрос дает гистограмма ин-

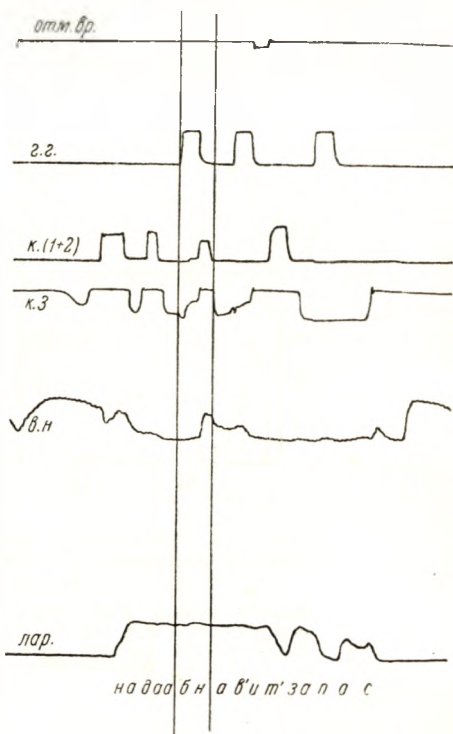


Рис. 4.6. Осциллограмма фразы *Надо обновить запас.*

тервалов между началами смычек (рис. 4.7). Можно видеть, что на всем материале не наблюдалось ни одного случая, когда смычка второго согласного опережала бы смычку первого или возникала бы одновременно с ней. При рассмотрении гистограмм обращают на себя внимание еще два факта: во-первых, оказывается, что задержки второй смычки относительно первой могут быть очень маленькими — составлять всего 10—20 мсек. Во-вторых, характер гистограмм показывает, что распределение величин задержек имеет отчетливо несимметричный характер — распределение огра-

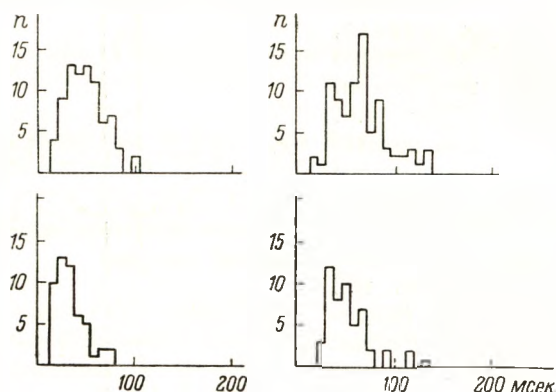


Рис. 4.7. Гистограммы задержек начала смычки второго согласного относительно начала смычки первого.

Слева — данные, полученные при разговорном стиле речи, справа — при полном стиле, сверху — при последовательности губные-язычные, снизу — при последовательности язычные-губные.

ничено снизу, со стороны малых задержек. Это заставляет думать, что между движениями двух согласных имеется рефлекторная связь — второе движение разрешается только после того, как начато первое. Так как величины задержек весьма малы, очевидно, что такая рефлекторная связь, если она действительно есть, должна осуществляться на сравнительно низком уровне центральной нервной системы. Логично было предположить, что смычка второго согласного не должна осуществляться до тех пор, пока не будет достигнуто состояние, полностью характеризующее первый согласный как фонему. В частности, смычка согласного, следующего за *м*, не должна возникать раньше, чем осуществится опускание небной занавески, смычка согласного, следующего за глухим согласным, не должна возникать раньше, чем будет прекращена фонация. Проверка этого предположения показала, что в сочетаниях, начинающихся с глухого согласного, смычка

второго согласного может начинаться раньше или одновременно с прекращением фонации (уменьшением сигнала с ларингофона). Например, смычка губ при произнесении *n* в сочетании *tn* начинается на территории озвонченного участка *m* (рис. 4.8). При разговорном стиле речи это наблюдалось в 23 случаях из 68 (34%), при произнесении в полном стиле это наблюдалось в 9 случаях из 68 (13%). В сочетаниях, начинающихся с носовых согласных, смычка второго согласного во всех без исключения случаях возникала уже после того, как нёбная занавеска опускалась (скорость тока воздуха через нос начинала расти). Примеры осциллограмм приведены на рис. 4.9. Полученная разница достаточно понятна, так как прекращение фонации при глухих согласных, следующих за гласными, как правило, несколько запаздывает относительно начала смычки, тогда как опускание нёбной занавески при носовых согласных в большинстве случаев несколько опережает смычку. Во всяком случае данные, полученные с сочетаниями, начинающимися с глухих согласных, свидетельствуют, что переход к смычке второго согласного может происходить до того, как осуществится весь комплекс артикуляций первого согласного.

Так как смычка второго согласного начиналась в то время, когда весь комплекс артикуляций первого согласного еще только разворачивался, можно было ожидать нарушения обычных временных отношений между смычкой второго согласного и остальными компонентами его артикуляции (подниманием или опусканием нёбной занавески, включением или выключением голоса). Измерения показали, что поднятие нёбной занавески при *m*, *n*,

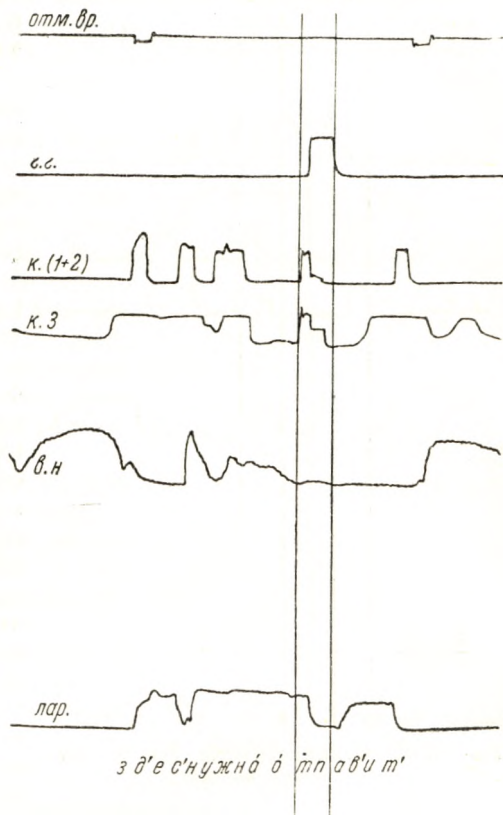


Рис. 4.8. Осциллограмма фразы *Здесь нужна отповедь.*

д, б, следующими за м, н, запаздывает относительно начала смычки в среднем на 50 мсек. (среднее из 72 случаев).

Задержка прекращения фонации относительно начала смычки в п, т, следующих за м, н, составила в среднем 59 мсек. (среднее из 38 случаев).

Опускание нёбной занавески при м, н, следующими за неносовыми согласными, запаздывало относительно начала смычки в среднем на 27 мсек. (52 случая).

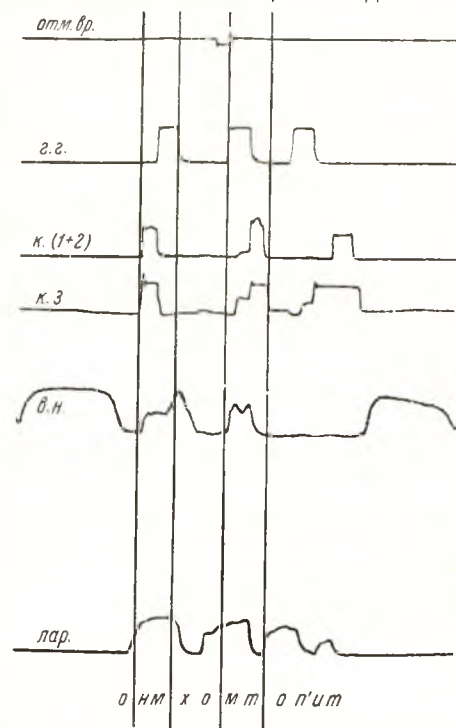


Рис. 4.9. Осциллограмма фразы *Он мгом топил.*

Опускание нёбной занавески при м и н в интервокальном положении опережало смычку в среднем на 40—50 мсек. Начало фонации при м, н, следующими за глухими согласными, запаздывало относительно смычки в среднем на 33 мсек. (29 случаев).

Таким образом, все полученные данные свидетельствуют о некотором запаздывании движений голосовых связок и нёбной занавески относительно начала смычки по сравнению с теми условиями, когда согласный следует за гласным. Очевидно, что причина такого запаздывания может быть весьма простой: каждый артикуляторный орган обладает инерционностью и не может изменять своего состояния с частотой, большей некоторого

предельного значения. Попытки определить максимальные возможные частоты для разных артикуляторных органов предпринимались рядом исследователей (Stetson, 1951; Kaiser, 1934; Hudgins а. Stetson, 1937). Было найдено, что для разных артикуляторных органов эти частоты различны, но максимальная не превышает 7—9 в секунду. Из этих данных можно сделать вывод, что минимальный интервал времени между двумя противоположно направленными движениями одного и того же органа не может быть меньше 55—70 мсек. Однако интервал между двумя смычками может быть значительно меньше этой величины (10—20 мсек.), так как в этом случае дело идет о движениях различных органов.

Итак, весь описанный до сих пор материал показал, что движения, требуемые согласными, осуществляются внутри слога последовательно. Сначала осуществляется комплекс движений, требуемых первым согласным, затем — комплекс движений, требуемых вторым согласным. Наименьшие значения интервала между согласными могут быть получены, если судить по движениям, не являющимся антагонистическими (смычки губ и языка). Представляло интерес выяснить, нет ли еще каких-нибудь признаков артикуляции второго согласного, которые проявлялись бы раньше возникновения смычки. Если такой признак существует, то можно с уверенностью утверждать, что оба согласных заданы параллельно и что последовательность в движениях есть результат последовательности в исполнении, но не последовательности заданий.

Эффект, к которому мы обратились, хорошо известен фонетикам — это явление озвончения глухого согласного при условии, если за ним следует шумный звонкий согласный. Озвончения не происходит, если следующий согласный — сонант или *в* (Аванесов, 1956). На нашем материале мы наблюдали этот эффект на примере следующих фраз:

это род болезни	}	озвончение всегда было
он раб долга		
это род маски	}	озвончения ни разу не было
он раб науки		

Прежде всего необходимо было проверить, не обусловлено ли это явление тем, что задержка *б*, *д* относительно *т*, *п* меньше задержки *м*, *н* и что *б*, *д* просто начинаются в то время, пока фонация еще не успела прекратиться (*п*, *т* еще не успели стать глухими). Измерения величин задержек между смычками согласных в сочетаниях не подтвердили этого предположения. Были получены следующие величины: *раб науки* — 75 мсек., *раб долга* — 72 мсек., *род маски* — 60 мсек., *род болезни* — 50 мсек.

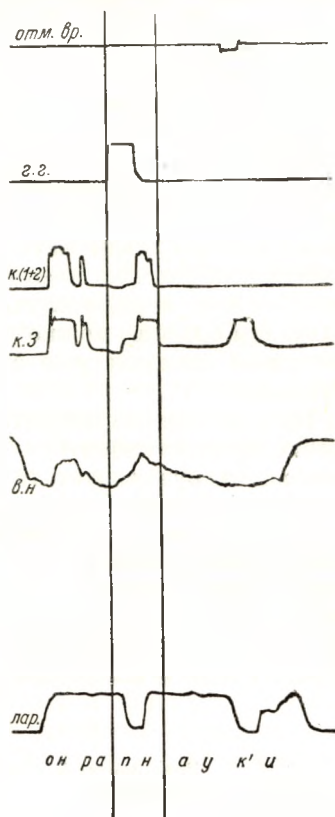


Рис. 4.10. Осциллограмма фразы *Он раб науки*.

Можно видеть, что различия очень малы, они не являются статистически существенными.

Кроме того, было обнаружено, что прекращение фонации в сочетаниях *пн*, *тм* иногда происходит на фоне появления некоторых признаков артикуляции носового согласного. Например, уменьшение сигнала с ларингофона начинается, когда скорость воздуха через нос уже начала расти (рис. 4.10).

Придя к выводу, что прекращение или непрекращение фонации при первом согласном есть результат знания второго согласного, мы можем утверждать, что эти знания в сочетании *пд* (*раб долга*) должны были быть получены не позже того момента времени, когда фонация прекратилась бы, если бы это было сочетание *пн*. Так как в ряде случаев голос прекращается одновременно с началом смычки, можно прийти к выводу, что сведения о втором согласном имеются уже с самого начала произнесения первого согласного.

Прежде чем переходить к обсуждению полученных данных, необходимо остановиться еще на одном вопросе, а именно, на выяснении, существует ли какая-либо зависимость во временных отношениях между согласными от положения слоговой границы и от того, относятся ли согласные к одному слову или принадлежат разным словам.

Приводим средние величины задержек (в мсек.) между началами смычек первого и второго согласного (сочетания *пт*, *тп*, *бд*, *дб*, *мн*, *нм*) в условиях, когда оба они находятся в середине одного слова и когда они принадлежат двум различным словам:

	Одно слово		Два слова
	ГССГ	ГССГ	ГССГ
Разговорный стиль . . .	50	45	44
Полный стиль	68	68	68

Очевидно, что величина задержки смычки второго согласного не зависит от того, находятся ли оба согласных в одном слове или принадлежат разным словам.

Приводим средние величины задержки (в мсек.) смычки второго согласного для сочетаний, согласные которых по существующим лингвистическим критериям или принадлежат одному слогу, или входят в разные слоги.

	Один слог				Два слога	
	Губной-языч- ный		Язычный-губ- ной		Губной- язычный	Язычный- губной
	<i>пт</i> , <i>пс</i> , <i>пн</i>	<i>бд</i> , <i>бн</i>	<i>тп</i> , <i>тм</i>	<i>дб</i> , <i>дм</i>	<i>мн</i> , <i>мт</i> , <i>мб</i> , <i>мс</i>	<i>нм</i> , <i>нт</i> , <i>нб</i>
Разговорный стиль . . .	50	49	35	39	47	33
Полный стиль	74	63	60	50	60	51

Эти данные демонстрируют эффект, который являлся бы парадоксальным, если бы интервал между смычками зависел от поло-

жения слогораздела. Можно видеть, что задержки минимальны в тех случаях, когда между согласными должна проходить слоговая граница.

Однако если допустить, что артикуляторная программа человека построена по типу открытых слогов — любой согласный после гласного означает начало следующей слоговой группы, то наблюдаемая тенденция к укорочению задержки в сочетаниях, начинающихся с носовых согласных, может быть легко понята. Дело в том, что смычки *м*, *н*, как известно, несколько короче смычек остальных смычных согласных (Златоустова, 1954). Можно думать, что это в значительной мере должно быть отнесено за счет меньшей силы мышечного сокращения и соответственно большего запаздывания начала смычки относительно начала движения (обсуждение этого вопроса см. в главе 5 настоящей работы). Если второй согласный разрешается началом первого согласного, то наблюдаемое укорочение задержки второго согласного в сочетаниях, начинающихся с носовых согласных, является в некотором смысле артефактом измерения. Мы должны были бы измерять начало движения, а вместо этого измеряли конечный эффект (смычку), который в случае носовых согласных больше запаздывал относительно момента начала движения.

Суммируя все приведенные данные относительно организации движений в сочетаниях типа ССГ, следует прийти к выводу, что все сочетание задается на каком-то уровне управления как одно-временная пачка указаний к эффекторным центрам. Это действительно вне зависимости от качества согласных в сочетании, а также от того, относятся ли оба согласных к одному слову или к двум разным словам. Однако будучи заданными одновременно, движения осуществляются как последовательные. Это не может быть отнесено за счет врожденных анатомио-физиологических координационных отношений между мышцами-антагонистами, которые, вероятно, обеспечивают последовательность движений согласного и гласного в элементарных слогах типа СГ. Приходится предположить, что здесь играют роль уже выработанные координационные отношения такого типа, что движения первого согласного рефлекторно стимулируют движения второго согласного.

Ряд фактов заставляет думать, что сигналом, стимулирующим (разрешающим) движения второго согласного, является или проприоцептивная импульсация, возникающая при движениях губ (языка), производимых при артикуляции первого согласного, или непосредственно возбуждение эффекторных центров этих движений. К таким фактам относятся малая величина задержки второго согласного, возможность возникновения смычки второго согласного до того как состояние, требуемое первым согласным, будет достигнуто, и уменьшение задержки между смычками в сочетаниях, начинающихся с *н*, *м*. Эти факты несовместимы с представлением о том, что сигналами, разрешающими следующий со-

гласный, могут быть звуковые или тактильные эффекты, возникающие при артикуляции первого согласного.

Какова реальная физиологическая схема связей между движениями согласных, пока остается неясным. В качестве простейшего предположения можно допустить, что элементы, ответственные за оба согласных, одновременно получают импульсацию из высших управляющих уровней, однако возбуждение элемента, соответствующего второму согласному, является подпороговым. Для того чтобы возбуждение стало надпороговым, вызвало внешний эффект, необходим дополнительный толчок, создаваемый импульсацией, возникающей при артикуляции первого согласного. Нужно подчеркнуть, что эти соображения являются чисто спекулятивными и приведены только для того, чтобы показать, что можно придумать кажущуюся физиологической схему, которая могла бы обеспечить образование последовательности согласных за счет внутренних связей между движениями. Оставаясь в пределах формального описания речевых явлений, мы можем пока только утверждать, что, во-первых, сочетание ССГ представляет собой артикуляторную единицу и, во-вторых, что эта единица должна включать выработанные отношения между движениями и в этом смысле должна быть более сложной, чем слог типа СГ.

Как будет показано в следующем разделе, сочетания ССГ при определенных условиях распадаются на элементарные слоги типа СГ.

4. Нарушения артикуляции в условиях задержанной обратной слуховой связи

В предыдущих разделах работы были изложены некоторые предварительные рабочие гипотезы об организации речевого процесса, созданные на основании исследования временных отношений в нормальной речи. Очевидно, что одним из наилучших и, возможно, наиболее реальных способов проверки этих гипотез является исследование патологии речи. Однако количественное экспериментальное исследование нарушений речи на клиническом материале сопряжено с весьма значительными трудностями как принципиального, так и организационного и чисто технического характера. С методической точки зрения значительно более удобным представляется исследование экспериментальной патологии речи — создание временных нарушений речи у здоровых людей и количественный анализ этих нарушений. Этим приемом мы и попытались воспользоваться для исследования наиболее интересующего нас вопроса о структуре элементов речевой программы человека.

Ли (Lee, 1950, 1951) описал чрезвычайно интересный эффект «искусственного заикания», возникающего в условиях, когда человек говорит, слыша свою собственную речь с задержкой

по времени. Методика вызывания этого эффекта весьма проста: речь диктора записывается на магнитную пленку и тут же воспроизводится, поступая, после соответствующего усиления, на телефоны, находящиеся на ушах диктора. Изменяя расстояние между записывающей и снимающей головками магнитофона, можно устанавливать требуемую величину временной задержки сигнала.

За последнее десятилетие выполнено огромное число работ, посвященных исследованию этого эффекта (специальная библиография вопроса опубликована: Chase, Sutton, First, 1959), однако физиологические механизмы эффекта до сих пор остаются неясными. Установлено, что, кроме заикания (повторения каких-то отрезков речи), эффект может выражаться в резком удлинении времени произнесения и увеличении громкости речи (Black, 1951; Atkinson, 1953; Spilka, 1954, и др.). Выраженность эффекта сильно варьирует от человека к человеку (Beaumont a. Foss, 1957). Эффект тем более выражен, чем большей является интенсивность задержанного сигнала (Atkinson, 1954; Peters, 1954; Butler a. Galloway, 1957; Chase, Guilfoyle, 1962) и зависит от величины задержки. Наиболее эффективной оказывается задержка порядка 200 мсек. (Black, 1951; Fairbanks, 1955; Butler a. Galloway, 1957; Spuehler, 1962).

За последние годы было установлено, что эффект не является специфическим для речи. Близкие по характеру нарушения наблюдаются при выстукивании ритмов в условиях задержанной обратной слуховой связи (Chase, Harvey, Standfast, Rapin a. Sutton, 1961; Ruhm, Cooper, 1962, и др.). Были обнаружены также существенные нарушения письма в условиях, когда испытуемый видит то, что он пишет, с задержкой по времени (van Bergeijk a. David, 1959).

Очевидно, что анализ механизмов нарушений речи при задержанной обратной слуховой связи вряд ли может быть осуществлен до того, как будут выяснены основные принципы организации речевого процесса. Поэтому в настоящей работе мы поставили перед собой значительно более скромную задачу, чем поиски механизмов явления. Нас интересовал весьма узкий вопрос о том, какие элементы речи повторяются при искусственном заикании.

Вопрос об элементах речи, повторяемых при искусственном заикании, исследовался в работе Фэрбенкса и Гуттмана (Fairbanks a. Guttman, 1958). Речь испытуемых записывалась на магнитную пленку, многократно прослушивалась и транскрибировалась фонетическими знаками. Затем производился тщательный анализ записей — выявлялись ошибки, замены и повторения. На основании рассмотрения фонетического состава повторяемых отрезков авторы пришли к выводу, что эти отрезки не совпадают ни с одним из элементов речи, выделяемых лингвистами, и что, следовательно, элементы речевой программы человека построены

по какому-то иному, до сих пор неясному принципу. Так как этот вывод полностью расходится с теми представлениями, которые обсуждались в предыдущих разделах настоящей работы, мы считали необходимым повторить исследование Фэрбенкса и Гуттмана и еще раз проанализировать характер наблюдаемых нарушений.

С точки зрения методических возможностей мы находились в значительно лучших условиях, чем Фэрбенкс и Гуттман, так как могли регистрировать непосредственно артикуляторные параметры речи. Это не только позволяло легко расшифровать, что именно было произнесено испытуемым, но и давало возможность использовать информацию о временных характеристиках процесса.

Так как интересующим нас эффектом являлось заикание, мы постарались создать условия, наиболее благоприятные для проявления этого эффекта. К ним относятся: большая громкость задержанного акустического сигнала (Butler a. Galloway, 1957; Chase, Guilfoyle, 1962) и требование от испытуемого быстрого темпа речи (Beaumont a. Foss, 1957).

Желая проверить описанный Фэрбенксом и Гуттманом факт зависимости величины повторяемого отрезка от величины задержки акустического сигнала, мы исследовали достаточно широкий диапазон задержек (100, 210, 300 и 470 мсек.) и использовали во всех случаях один и тот же речевой материал (10 фраз, состоящих из 14—25 звуков речи). Испытуемый повторял каждую фразу по два раза в нормальных условиях и по два раза при каждой из задержек. Порядок изменения величин задержек от одной серии фраз к другой был различным у разных испытуемых. Фразы были подобраны таким образом, чтобы каждый из согласных можно было надежно расшифровать, используя минимальное число артикуляторных параметров. Основную часть согласных во фразах составляли переднеязычные (щелевые и смычные) и губные (смычные) согласные.

Учитывая данные о частых ошибках по признаку глухости—звонкости при произношении в условиях задержанного акустического сигнала (Fairbanks a. Guttman, 1958), мы при подборе материала избегали соседства согласных, отличающихся друг от друга только по этому признаку.

В опытах принимали участие пять испытуемых. Каждый из них с соответствующими датчиками произносил заранее выученные фразы перед электродинамическим микрофоном 82А-2, расположенным на уровне рта на расстоянии 15 см. Сигнал с микрофона поступал на предварительный усилитель и затем на вход усилителя записи первого магнитофона. Для получения небольших (100, 210 мсек.) задержек использовался один магнитофон МЭЗ-28, изменение задержки достигалось путем изменения скорости движения пленки. Для получения больших задержек ис-

пользовались системы из двух и трех магнитофонов. Сигнал с выхода усилителя воспроизведения последнего из магнитофонов поступал на магазин затухания МЗУ и два головных телефона ТД-6. Напряжение на зажимах телефонов поддерживалось одинаковым при всех задержках и равнялось напряжению на частоте 1 кгц, при котором звуковое давление, развиваемое телефоном, составляет 115 дб над $2 \cdot 10^{-4}$ мкб. Параметры артикуляционных движений регистрировались на ленте чернилопишущего осциллографа. Во всех случаях регистрировались следующие параметры: контакт между губами, голос (сигнал с ларингофона), контакты языка с искусственным нёбом в точках к.(1+2) и к.3.

Результаты записей давали возможность расшифровать все встречавшиеся во фразах согласные, однако они не позволяли определить характер гласных. Поэтому гласные определялись на слух при прослушивании магнитных записей речи испытуемых.

Для сравнения полученных нами данных с литературными и сравнительной оценки выраженности нарушений речи у разных испытуемых мы воспользовались такой интегральной характеристикой, как время произнесения. Мерой его служило среднее значение отношения длительности произнесения фразы к числу фонем во фразе.

Средние данные по всем испытуемым (рис. 4.11, 6) не противоречат результатам других авторов, наблюдавших максимальное изменение времени произнесения в области задержек от 150 до 300 мсек. (Fairbanks, 1955; Spuehler, 1962). Вместе с тем испытуемые очень значительно отличаются друг от друга как по выраженности эффекта, так и по значению задержки, на которой эффект максимален. Можно видеть, что у одного испытуемого (рис. 4.11, 5) удлинения произнесения почти не было. Поэтому весь дальнейший анализ проводился на материале только четырех испытуемых.

Рассмотрение осциллограмм артикуляции показало, что наиболее ярко выраженным эффектом, наблюдавшимся при всех значениях задержек, являлось многократное смыкание и размы-

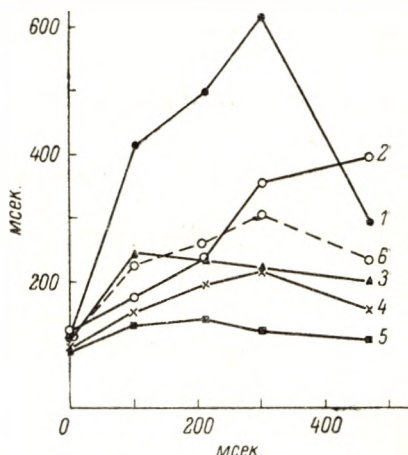


Рис. 4.11. Зависимость среднего времени произнесения (по оси ординат, в мсек.) от величины задержки акустического сигнала (по оси абсцисс, в мсек.).

1—5 — данные отдельных испытуемых, 6 — усредненные данные по всей группе испытуемых.

Части фраз, повторно произнесенные дикторами при различных задержках обратного акустического сигнала

№	Задержки (в мсек.)											
	100		210		300		470					
	Дикторы											
	Л. Б.	М. Л.	Л. Ч.	Ю. К.	Л. Б.	М. Л.	Л. Ч.	Ю. К.	Л. Б.	М. Л.	Л. Ч.	Ю. К.
1								тно	ном л'ем'е ном л'ем'е н'исан' т'оп'ые		палат	
2								п'лы п'л'ин сту			надап'и	п'лы сто
3											т'оп	
4	сам										ст'е н'ту бал'ту сн'а	пас
5								п'ло			н'и'л'и н'и'л'и н'и'л'и	
6								ску			н'и'л'ота н'и'л'ит капу пусту паса паса пас	пас
7								м'па			л'ит' паса'л'и сала'и	
8	лка							был			сн'е паса ата	пас
9								сн'е сн'е сн'у пат'			сн'е паса ата	н'ит
10								ата			заснул заснь	н'ит

Фэрбенкс рассматривает речевой процесс как последовательность (цепочку) некоторых элементарных артикуляционных комплексов (единиц речи), где каждый следующий комплекс разрешается и начинается осуществляться при получении сведений о выполнении предыдущего комплекса движений. Иначе говоря, каждый следующий комплекс рассматривается как реакция, возникающая в ответ на тактильные, проприоцептивные и звуковые сигналы, сопровождающие осуществление предыдущего комплекса. Если мы обозначим артикуляторные комплексы как A , B , а соответствующие им тактильные, проприоцептивные и звуковые сигналы как $a_1, a_2, a_3, b_1, b_2, b_3$, то комплекс B явится реакцией на a_1, a_2, a_3 , т. е.

$$\left. \begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} \right\} \rightarrow B.$$

Тогда нормальную речь можно представить в виде последовательности

$$A \left. \begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} \right\} \rightarrow B \left. \begin{matrix} b_1 \\ b_2 \\ b_3 \end{matrix} \right\} \rightarrow C \dots$$

Очевидно, что при задержке акустического сигнала может возникнуть сложная ситуация, когда одновременно с тактильными и проприоцептивными сигналами от комплекса B будет существовать акустический сигнал от комплекса A .

$$A \left. \begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix} \right\} \rightarrow B \left. \begin{matrix} b_1 \\ b_2 \\ b_3 \end{matrix} \right\}$$

Исход такой ситуации может быть различен в зависимости от относительных весов b_1, b_2, a_3 . Если a_3 «сильнее» чем b_1, b_2 , то может возникнуть повторение комплекса B и будет образована последовательность ABB^1 . Существенно, однако, что от комплекса B^1 может быть переход только к комплексу C , так как одновременно с тактильным и проприоцептивным сигналами от комплекса B^1 должен существовать акустический сигнал от комплекса B , который имеет совпадающее с ними значение — сигнализирует переход к C . Таким образом, критической особенностью гипотезы Фэрбенкса является то, что она допускает не более чем двукратные повторения одного и того же артикуляторного комплекса. Фэрбенкс и Гуттман пишут, что они наблюдали более чем двукратные повторения, которые не согласуются с гипотезой, однако, так как

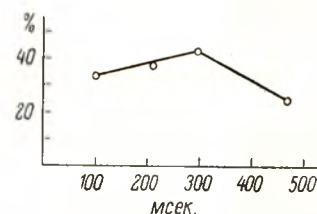


Рис. 4.13. Доля согласных (слов), превращенных в серии смыканий и размыканий (по оси ординат, в %) при разном времени задержки акустического сигнала (по оси абсцисс, в мсек.).

подобные случаи встречались достаточно редко, они сочли возможным считать это нетипичным.

Обработка нашего материала показала, что свыше 50% всех серий заиканий образованы из более чем двукратных повторений, а число повторений в серии может быть очень большим, достигая 16 (рис. 4.14). Очевидно, что малая выраженность многократных повторений в материале Фэрбенкса и Гуттмана обусловлена просто малой громкостью акустического сигнала.

Итак, первый вывод, в котором мы вынуждены разойтись с Фэрбенксом и Гуттманом, состоит в том, что многократные повторения комплекса движений, формально совпадающего со слогом СГ, являются весьма типичными для искусственного заикания.

Однако так как этот эффект не предусмотрен гипотезой Фэрбенкса, мы подробнее рассмотрим его несколько позже, а пока что обратимся к случаям повторения более крупных отрезков речи (см. табл. 1). Вопрос заключается в следующем: возможно ли предположить, как это делают Фэрбенкс и Гуттман, что такие отрезки являются элементами речевой программы человека? Понятие элемента может иметь смысл только в том случае, если ему приписаны определенные свойства, такие как например константность, неделимость, определенный принцип организации составляющих его единиц, наличие определенных отношений с более крупными единицами.

Рассмотрение повторяемые отрезки являются настолько разнообразными, что не удовлетворяют ни одному мыслимому требованию. Они не только находятся в случайных отношениях со словами, не только разные дикторы повторяют разные куски из одного и того же слова, но и один и тот же диктор при одной и той же

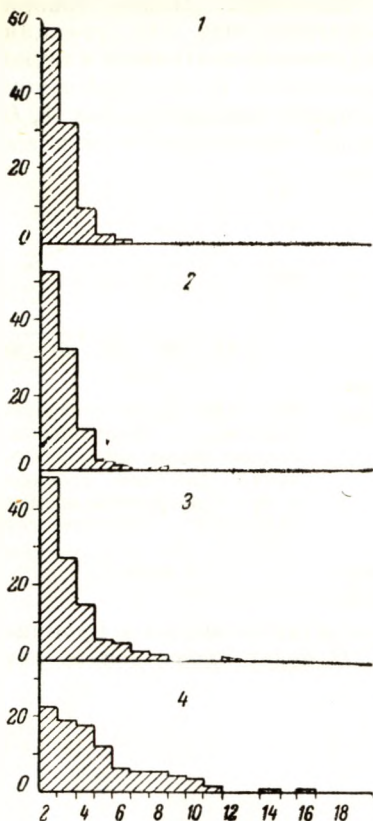


Рис. 4.14. Относительные частоты встречаемости (по оси ординат) серий, состоящих из разного числа смыканий и размыканий (по оси абсцисс) у дикторов: Ю. К. (1), М. Л. (2), Л. Ч. (3), Л. Б. (4).

задержке акустического сигнала обращается с одним и тем же словом по-разному. Так, диктор Л. Б. повторяет из слова *паса-лить* то *пасали-*, то *сали-*. Диктор Л. Ч. повторяет из того же слова то *паса-*, то *пас-*. Он же повторяет из слова *капусту* то *капу-*, то *-пусту*, то повторяет целое слово *заснул*, то только часть этого слова (*зас-*). Такая случайность в делении одного и того же слова на разные куски является, как нам кажется, убедительным доводом против гипотезы Фэрбенкса и Гуттмана и заставляет считать, что большие отрезки речи, повторяемые при искусственном заикании, ни в коей мере не позволяют судить об элементах, из которых построена речевая программа.

Таким образом, мы вынуждены пока полностью отказаться от гипотезы Фэрбенкса и Гуттмана и попытаться проанализировать полученные данные с другой точки зрения.

Как уже говорилось, основной наблюдавшийся нами эффект заключался в превращении одиночного слога СГ или одиночного согласного в серию открытых слогов типа СГ.

В предыдущей главе были приведены соображения в пользу того, что последовательность согласный—гласный в слоге СГ образуется за счет автоматической деятельности систем (нервных центров), управляющих мышцами-антагонистами губ, языка, голосовых связок. КRIDOM, Денни-Броуном, Икклсом, Лидделлом, Шеррингтоном (1935) было показано, что подобные системы способны длительно поддерживать ритмические движения при условии, если искусственно исключены тормозящие влияния со стороны высших управляющих нервных центров. Естественно, возникает предположение, не имеем ли мы в нашем случае дело с аналогичным явлением, не приводит ли задержанный и усиленный акустический сигнал к временному исключению управляющих влияний и выявлению собственной активности систем, ведающих мышцами-антагонистами.

Очевидно, что если мы имеем дело с аналогичным феноменом, то период колебательных смыкательно-размыкательных движений (губ, языка, голосовых связок) должен определяться свойствами самих нервных центров и не должен зависеть от времени задержки акустического сигнала. Обработка материала подтвердила, что период смыкательно-размыкательных движений действительно не зависит от задержки акустического сигнала (рис. 4.15). Обращает на себя внимание также примерное совпадение длительностей

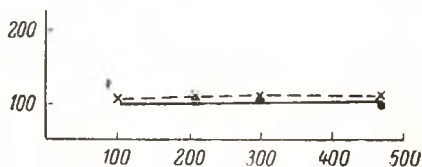


Рис. 4.15. Средние длительности фаз смыкания контактов (сплошная линия) и фаз размыкания (пунктирная линия) при разных задержках акустического сигнала (по оси абсцисс, в мсек.; по оси ординат — длительности фаз смыкания и размыкания, в мсек.).

фаз смыкания и размыкания, что является отнюдь не характерным для нормальной речи. Эти факты заставляют считать, что акустический сигнал непосредственно не участвует в процессах, вызывающих каждое отдельное движение в серии. Нужно сказать, что в пользу этого вывода говорит также то, что задержка акустического сигнала, нарушая произношение речевого материала, отнюдь не мешает повторениям одного и того же открытого слога (Chase, 1958).

Если при искусственном заикании возникают собственные колебания в нескольких параллельно работающих системах (центры антагонистических мышц языка, нёбной занавески, голосовых

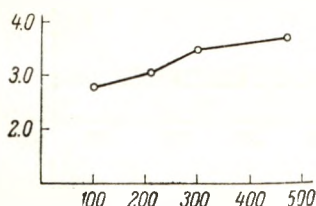


Рис. 4.16. Среднее число смыкательно-размыкательных движений в серии заиканий (по оси ординат) при разных задержках акустического сигнала (по оси абсцисс, в мсек.).

связок), то следует ожидать появления десинхронизации параллельных движений. На нашем материале очень отчетливо выявлялись нарушения обычной координации между фонацией и движениями губ и языка. Фэрбенкс и Гуттман также указывают, что ошибки глухости-звонкости являются наиболее частыми ошибками при задержанной обратной слуховой связи.

Для того чтобы продолжать рассмотрение материала, необходимо ввести понятие структуры слога СГ. Примем, что структура описывается перечнем органов, которые должны принимать

участие в артикуляции, и указанием набора правил, связывающих движения этих органов (например, поднять язык и поднять нёбную занавеску и прекратить фонацию и т. п.). Последовательность движений внутри слога СГ не указывается, так как она должна определяться уже собственными свойствами систем управления антагонистическими мышцами. Если определить фонему как элементарный набор правил (инструкций) относительно требуемого комплекса движений, то структура слога СГ исчерпывающе описывается одновременным указанием согласной и гласной фонемы. Естественно, что структура слога не меняется, если набор правил движений остается постоянным. С этой точки зрения, серия заиканий *та-та-та*, будучи последовательностью движений, не является последовательностью структур движений: на всем протяжении серии сохраняется одна и та же структура.

Введя представление о структуре, можно описать как нормально сказанную фразу, так и фразу, образуемую в условиях задержки акустического сигнала, в виде определенной последовательности структур. Как уже говорилось, период движений, совершаемых на протяжении существования каждой структуры, не зависит от величины задержки акустического сигнала. Естественно поставить вопрос, влияет ли величина задержки на ско-

рость смены структур. Для получения ориентировочных данных по этому вопросу было подсчитано среднее число смыкательно-размыкательных движений в сериях, образованных при различных задержках. При подсчете учитывались только те случаи, когда было не менее двух повторений, т. е. когда эффект нарушения речи заведомо имел место. Оказалось, что длина серии (время существования структуры) возрастает с увеличением задержки (рис. 4.16). Этот эффект не проявляется при оценке суммарной длительности фразы (см. рис. 4.11), так как удлинение задержки приводит также к уменьшению вероятности совершить ошибку (см. рис. 4.13).

Полученные факты заставляют считать, что существо эффекта задержки акустического сигнала состоит не в вызове повторных движений, а в задержке перехода от одной структуры движений к другой их структуре. Действительно, при задержке акустического сигнала человек может произвольно или заикаться, или растягивать речь. Если он торопится — возникает заикание, если старается говорить медленно — речь растягивается. Однако что бы он ни делал, смена структур замедляется, время, занятое произнесением фразы, увеличивается.

Возвращаясь к определению структуры, мы должны установить, в каких отношениях она находится с традиционно выделяемыми лингвистическими элементами. Данные настоящей работы показывают, что при задержанном акустическом сигнале не задерживается только переход от согласного к гласному (от смыкания к размыканию), т. е. в любом случае сохраняется элементарная модель слога СГ. Для проверки, каким образом реализуются более сложные последовательности (СГС, ССГ, СГСС и т. д.), был специально подобран материал, содержащий сочетания этих типов. Оказалось, что произнесение подобных сочетаний при задержанном акустическом сигнале приводит к образованию новых слогов:

$$\begin{aligned} \text{СГС} &= \text{СГ} + \text{СЪ} \\ \text{ССГ} &= \text{СЪ} + \text{СГ} \\ \text{СГСС} &= \text{СГ} + \text{СЪ} + \text{СЪ} \end{aligned}$$

где Ъ — редуцированный гласный (заднего ряда после твердых согласных и переднего ряда после мягких).

Особый интерес представляет характер образования последовательностей ССГ в случаях, когда слоговая граница, определяемая по упомянутым ранее критериям (Аванесов, 1956), или находится перед группой согласных, или разделяет их. Для проверки был подобран материал с необходимыми сочетаниями (*письмо, цапли, полотенно, ватник, заснул; султан, подарки, винтик, замполит* и т. д.). Оказалось, что принципиальной разницы между сочетаниями того и другого типа нет: *письмо* = *пи* + *съ* + *мо*, *султан* = *су* + *лъ* + *та* + *нъ*.

Наконец, интересно было установить, существует ли какая-либо связь между характером описываемых структур и морфологическими отношениями. Известно, что с лингвистической точки зрения характер слога зависит часто от того, где проходит морфологическая граница; в связи с существованием таких мнений были подобраны фразы типа *мои подарки*, *свернул под арку* и т. д. При задержанной обратной слуховой связи произнесение последовательностей того и другого типа ничем не отличалось друг от друга, подчиняясь общей закономерности — тенденции к образованию простых открытых слогов типа СГ.

Нужно особенно отметить, что вновь образованные слоги ничем не отличаются от других слогов. Дело не только в том, что фаза размыкания в серии движений, заменяющей согласный, имеет ту же длительность, что и в серии движений, заменяющей открытый слог. Важно, что размыкание смычки сопровождается включением голоса, если согласный был глухим. Это означает, что образованный комплекс движений действительно соответствует открытому слогу.

Итак, чисто формально описывая последовательности структур для фраз, произнесенных в условиях задержки обратной слуховой связи, необходимо пользоваться алфавитом элементарных структур, соответствующих слогам типа СГ.

То, что гласный должен входить в один двигательный комплекс с предшествующим согласным, вполне понятно. Понятным является также отделение конечного согласного в закрытом слоге. Результаты, изложенные в предыдущих разделах, привели нас к выводу, что каждый новый согласный после гласного означает начало новой слоговой группы.

Неожиданным, с точки зрения данных, полученных на нормальной речи, оказалось распадение слоговой группы ССГ на два открытых слога (*ца-нъ-ли*, *па-ла-тъ-но* и т. д.). Как было ранее установлено, вся группа ССГ фигурирует как одна единица на высших уровнях управления речью. Движения гласного начинаются в нормальной речи одновременно с началом движений первого согласного, первый согласный оказывается глухим или звонким в зависимости от того, каким будет второй согласный. Вывод, который был сделан на основании материалов, приведенных в предыдущих разделах, состоял в том, что согласные в группе ССГ последовательны не в отношении их задания, но в отношении исполнения. Так как временная задержка между смычками согласных в ССГ может быть очень малой (10—20 мсек.), единственное казавшееся возможным предположение состояло в том, что в схеме артикуляции подобных сочетаний действует правило, по которому второй согласный не разрешается, пока не поступила импульсация, связанная с началом движений первого согласного.

Следовательно, распад сочетаний ССГ на структуры СГ должен вызываться не задержками в поступлении указаний из выс-

ших управляющих уровней, но нарушениями в осуществлении рефлекторных связей между комплексами движений.

В наборе исследованных фраз имелась одна, в которой последний глухой согласный одного слова должен был при нормальной речи озвончаться за счет первого согласного следующего слова. Рассмотрение данных, полученных в условиях искусственного заикания, показало, что озвончение всегда имело место, хотя озвонченный согласный и превращался в серию СГ. Таким образом, можно прийти к выводу, что в момент артикуляции этого согласного уже поступили указания о следующем согласном, невозможным же оказывается только осуществление перехода к движениям второго согласного.

Мы пока ничего не знаем о том, в результате каких физиологических механизмов нарушается переход от артикуляции одного согласного к артикуляции второго согласного. Возможно, что это происходит потому, что каким-то образом маскируется импульсация, возникшая при движениях первого согласного. Однако вполне допустимы и другие предположения. Пока для нас существенным является только тот факт, что при условии, когда такой переход задержан, комплекс артикуляции согласного автоматически завершается артикуляцией гласного (размыканием смычки, включением голоса), хотя никакого гласного не наблюдается, если речь осуществляется в нормальных условиях. Это заставляет считать, что основными элементами речи являются простейшие артикуляторные комплексы типа СГ, а более сложные сочетания типа ССГ, СССГ представляют собой не что иное, как группы этих простых комплексов, организованные таким образом, что следующий комплекс начинается раньше, чем успевает закончиться предыдущий.

Если принять для простоты, что каждый слог состоит только из смыкательно-размыкательных движений одного какого-то органа, то можно представить себе весьма простую схему, которая должна быть способной генерировать слоговую группу типа ССГ. В этой схеме (рис. 4.17) одна из систем (№ 1) осуществляет смыкание и размыкание губ, другая система (№ 2) осуществляет подъем и опускание языка. Обе системы одновременно получают извне возбуждающие сигналы, однако сигнал, поступающий на систему № 1, достаточен для ее возбуждения, а сигнал, поступающий на систему № 2, является подпороговым. Сигнал, возникающий на выходе блока a_1 , поступает параллельно на входы блоков a_2 и b_1 и вызывает как размыкание губ, так и осуществление смычки

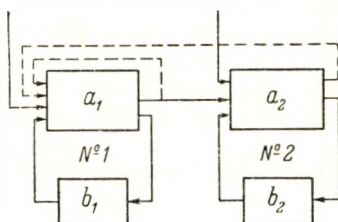


Рис. 4.17. Схема образования слоговой группы ССГ.

Объяснения в тексте.

языка. Для того чтобы не осуществлялось дополнительных движений губ параллельно с движениями языка, следует или предположить наличие тормозной связи от системы № 2 к системе № 1 (пунктирные линии), или же допустить, что сигнал с выхода блока a_1 действует как тормозный на вход того же самого блока.

Если сигнал с выхода блока a_1 по каким-то причинам не попадет на вход блока a_2 , то должно возникнуть заикание — автоколебание в системе № 1.

Приведенная схема, очень далекая, очевидно, от реальной физиологической схемы организации артикуляторных движений, полезна в том отношении, что она позволяет представить себе, как принципиально может происходить иерархическое управление речью.

До сих пор мы говорили о нарушениях артикуляции внутри-слоговых групп типа ССГ. Однако при искусственном заикании задерживается также переход от одной слоговой группы к следующей слоговой группе. Как было показано в главе 3, интервал времени между слоговыми группами определяется ритмической организацией слова (фразы).

Представляло интерес выяснить, сохраняется ли ритмическая организация при искусственном заикании.

При обработке материалов обнаружилось, что последний глас-

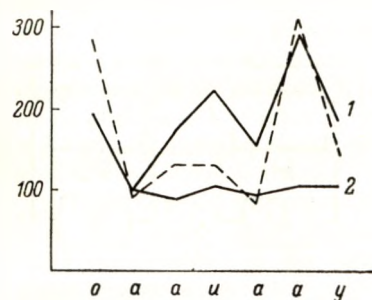


Рис. 4.18. Средние длительности гласных (по оси ординат, в мсек.) во фразе *Тоня топила баню*, произнесенной в нормальных условиях (пунктирная кривая) и при задержке акустического сигнала (сплошные кривые).

1 — длительность последних гласных в сериях заиканий; 2 — длительность непоследних гласных.

ный в серии заиканий бывает длиннее остальных гласных и значительно варьирует по длительности в разных фонетических ситуациях. Поэтому при подсчете длительности фаз размыкания при разных задержках акустического сигнала последние гласные серии не принимались во внимание (см. рис. 4.15). Эти наблюдения привели к предположению, что ритмическая структура фразы находит какое-то отражение в длительности последних слогов (гласных) в серии заиканий. Для проверки этого предположения были поставлены дополнительные эксперименты на одном испытуемом, наиболее склонном к заиканию. Испытуемый повторил по 20 раз при задержках 100 и 300 мсек. фразу *Тоня топила баню*. Для этой фразы уже имелись статистические данные, характеризующие произнесение ее этим испытуемым в нормальных условиях.

Оказалось, что фразовый рисунок средних длительностей гласных, последних в серии (рис. 4.18, 1), достаточно хорошо воспроизводит фразовый рисунок длительностей гласных при нормальном произнесении. Средняя длительность остальных гласных в серии (рис. 4.18, 2) оказывается весьма постоянной и не зависит от длительности соответствующих им гласных в нормально сказанной фразе.

Полученные результаты заставляют думать, что ритмический рисунок строится как последовательность значений интервалов времени между началом последнего согласного в предыдущей слоговой группе и следующей слоговой командой. Начало предыдущего движения (слога СГ), очевидно, выступает как опорный момент времени, от которого отсчитывается интервал до следующей команды. К сожалению, сколько-нибудь содержательные рассуждения на эту тему невозможны до тех пор, пока не будут сделаны хотя бы самые предварительные предположения, что же такое время в центральной нервной системе и как оно может быть указано в программе движений.

В заключение следует остановиться на эффекте повторения крупных отрезков фразы (см. таблицу). Анализ материала заключался в попытке установить роль акустического сигнала в этих повторениях. Пользуясь осциллограммами и зная величину задержки, можно было определить, что слышал испытуемый перед тем, как он осуществил возврат к уже сказанному (т. е. начал повторять). Для 37 случаев возврата были составлены анкеты, отвечающие на следующие вопросы:

1. Что человек сказал перед тем, как возвратился назад.
- 1а. Что он услышал перед тем, как возвратился.
2. Что он должен был бы сказать, продолжая без возврата (после 1).
- 2а. Что он должен был бы сказать, продолжая услышанное (после 1а).
3. Что он сказал.

Приводим пример такой анкеты для фразы *Над полотном ле-
тели цапли*:

1. Сказал *натпалатно*.
- 1а. Услышал *натпала*.
2. Продолжение сказанного *м'ет'ел'и*.
- 2а. Продолжение услышанного *тном*.
3. Сказал (возвратясь) *тном*.

Оказалось, что в 29 случаях из 37 то, что человек сказал, возвратившись к уже сказанному, было непосредственным продолжением того, что он к этому моменту услышал.

Тот факт, что человек с минимальными задержками может продолжить фразу или слово, начало которых сказано другим человеком, известен очень давно. На этом, в частности, построен ассоциативный эксперимент. Эффект очень отчетливо проявляется в экспериментах с текущим повторением речи диктора. В начале

фразы речь испытуемого отстает от речи диктора, однако очень скоро оба человека начинают говорить одновременно, или испытуемый даже опережает диктора (Чистович, Алякринский, Абульян, 1960). Для описания явления используются различные понятия: предполагается, что в мозгу человека существуют временные связи между элементами слова (фразы), что человек способен предсказывать будущие значения речевого сигнала, что человек использует избыточность, существующую в языке.

Неоднократно высказывались соображения, что процесс речевого восприятия всегда происходит с участием внутренней речи — слушатель является одновременно говорящим, образуя то же сообщение, что и диктор. Приняв эту точку зрения, можно предположить, что человек, которому подается некоторое речевое сообщение, не повторяет его вслух или потому, что не слушает его, или потому, что возбудимость речевых органов понижена — речь является скрытой.

Обращаясь к нашим экспериментам, можно видеть, что оба эти условия не могли соблюдаться. Громкость речевого сигнала в телефонах была очень большой, так что игнорировать его было невозможно. Возбудимость речевых органов, естественно, была достаточно высокой, так как испытуемый обязан был говорить.

Если приведенные соображения являются основательными, то эффект насильственного продолжения слышимой речи должен наблюдаться также в тех условиях, когда испытуемый слышит громкую речь другого человека, говорящего то же самое, что и он. Мы попробовали вызвать этот эффект, сначала записав фразы на магнитную пленку и затем предъявив их диктору, произносящему те же фразы. Из эксперимента ничего не вышло, так как диктор стал говорить синхронно с тем, что он слышал.

Очевидно, что обсуждая эти явления, мы попадаем в значительно более сложную область, чем та, которая нас пока интересует. На настоящем этапе исследования наиболее существенным результатом экспериментов с задержкой обратного акустического сигнала является факт превращения сложного слога ССГ в последовательность элементарных открытых слогов типа СГ. Это позволяет считать, что простейшим и основным артикуляторным комплексом является слог СГ. Более сложный слог типа ССГ представляет собой систему из этих простейших комплексов, построенную таким образом, что второй из комплексов может осуществляться частично параллельно с осуществлением первого.

Следующей задачей исследования должно быть выяснение организации отдельных артикуляторных движений (явлений) внутри простейшего слогового комплекса. Некоторые данные по этому вопросу будут приведены в следующей главе.

Г Л А В А 5

ВЗАИМООТНОШЕНИЯ МЕЖДУ ЭЛЕМЕНТАМИ АРТИКУЛЯТОРНЫХ КОМПЛЕКСОВ

1. Элементы артикуляции и дифференциальные признаки

Классическая фонетика рассматривала речь как последовательность дискретных элементов — звуков речи. Звуки речи описывались в терминах определенных состояний артикуляторного тракта. Из фонетики эти представления перешли к широкому кругу специалистов, занимающихся в настоящее время вопросами автоматической обработки речевых сигналов. Однако при объективном исследовании речевого процесса было обнаружено, что речь является потоком практически непрерывных изменений, принципы членения которого на дискретные элементы оказались неясными. Таким образом, возникла необходимость уточнить, что такое звуки речи и в каком смысле можно говорить о речи как последовательности дискретных элементов.

На протяжении последних нескольких лет сотрудниками Хаскинских лабораторий были получены экспериментальные данные, свидетельствующие, что классическое описание звука речи в виде набора взаимно независимых признаков может быть сохранено, если в качестве коррелятов фонематических признаков рассматривать не состояния тех или иных отделов артикуляторного тракта, а активность определенных мышечных групп. В частности, при миографической регистрации движений губ во время произнесения английских губных согласных *p*, *b* и *m* было обнаружено, что сокращения мышц во всех этих случаях являются одинаковыми по длительности и интенсивности и, следовательно, не зависят от того, какие другие движения совершаются одновременно. Аналогичным образом изменения мышечной активности в области нёбной занавески оказались однотипными как при произнесении глухого *p*, так и при произнесении звонкого *b* (Lysaught, Rosov, Harris, 1961; Harris, Schvey, Lysaught, 1962).

Далее при помощи аналогичной методики были получены данные, свидетельствующие, что мышечная активность в области угла рта при произнесении f остается инвариантной по амплитуде и длительности вне зависимости от того, произносится ли данный звук в конце слова или в окружении других согласных (Mac-Neilage, 1963).

Основываясь на этих данных, указанные авторы приходят к выводу, что наблюдаемые изменения конфигурации речевого тракта являются результатом относительно стандартных двигательных актов (артикуляторных жестов), характер которых не зависит ни от сопутствующих, ни от предшествующих и последующих движений. Таким образом, фонема оказывается перечислением набора независимых движений, а поток речи — результатом реализации последовательности фонем. Следовательно, представляется возможным описать процесс речеобразования предельно простым и экономным образом.

Понятно, что изложенная точка зрения имеет принципиальное значение как для понимания процессов синтеза и восприятия речи человеком, так и для разработки процедур автоматического анализа и синтеза речи. Однако экспериментальные данные, которые позволили указанным авторам сделать эти выводы, немногочисленны и, видимо, противоречат ряду фонетических представлений. В частности, фактически общепризнанным является мнение, что глухие согласные произносятся при большем напряжении артикуляторного тракта, чем звонкие и назальные согласные с тем же местом образования. Кроме того, считается, что длительность глухих согласных больше, чем длительность соответствующих звонких (Зиндер, 1960).

Однако мнение фонетистов о более сильной артикуляции глухих согласных основано преимущественно на субъективных ощущениях самих говорящих, а данные относительно различной длительности глухих и звонких согласных были получены главным образом посредством визуальной обработки осциллограмм акустического сигнала и пневмограмм (Златоустова, 1954). Экспериментальных данных о длительности самих движений, насколько нам известно, в фонетической литературе не существует.

Поэтому, чтобы установить, действительно ли указанные соотношения имеют место, нами было произведено исследование артикуляции глухих, звонких и назальных согласных при помощи описанной выше регистрационной методики. Набор исследованных звуков состоял из губных согласных p , b и m и язычных согласных t , d и n . Эти звуки были произнесены в одинаковых фонетических положениях в составе двухсложных слов, следовавших в случайном порядке. На протяжении записи непрерывно осуществлялась регистрация смыкания губ, смыкания языка с передним участком твердого нёба (x_3), сигнала с ларингофонного датчика и скорости тока воздуха через нос. По изме-

нению последнего параметра мы могли косвенно судить об изменении положения нёбной занавески. Такие записи были произведены на двух испытуемых.

Первым диктором все исследованные звуки были произнесены в интервокальном положении в составе слов типа *тапа*, *таба*, *тапа* и т. п. по 30 раз каждый. При обработке полученных осциллограмм была измерена длительность губных или нёбно-язычных смычек и длительность предшествующих гласных. За начало гласного принималось размыкание смычки первого согласного. Средняя длительность смычек оказалась наибольшей при произнесении глухих *п* и *т* и наименьшей при произнесении назальных *м* и *н*. Разница между звонкими согласными *б* и *д*, с одной стороны, и назальными согласными *м* и *н* — с другой, оказалась незначительной (5—3 мсек.). Таким образом, можно было думать, что движения губ и языка зависят от звонкости или глухости согласного и существенно не зависят от признака назальности—неназальности. Однако при сопоставлении длительностей предшествующих гласных стало очевидным, что последнее предположение не соответствует действительности. Как можно видеть на графиках, представляющих результаты усреднения 30 осциллограмм (рис. 5.1), укорочение звонких согласных *б* и *д* по сравнению с парными им глухими происходит преимущественно за счет запаздывания момента образования смычки. Размыкание смычек звонких и глухих согласных осуществлялось практически одновременно. При артикуляции назальных согласных картина оказывается несколько иной. Предшествующий гласный удлиняется в этом случае больше, чем при артикуляции соответствующих звонких согласных. Иными словами, признак назальности изменяет начальную фазу артикуляции согласного в том же направлении, что и звонкость. Однако размыкание смычки при *м* и *н* осуществляется позднее, чем при *б* и *д*. Вследствие этого при простом сопоставлении длительностей смычек различие в артикуляции этих групп согласных в значительной мере маскируется.

Близкие соотношения были обнаружены при произнесении этим же диктором глухих, звонких и назальных согласных в начале слова. В этом случае исследованные звуки произносились в составе слов *папа*, *бапа*, *напа* и т. п. О запаздывании начала

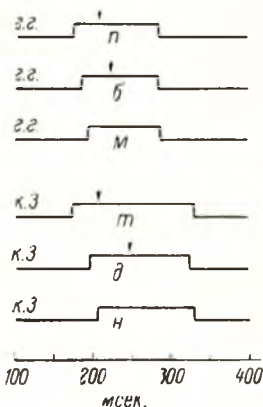


Рис. 5.1. Время начала и конца (в мсек.) губных и нёбно-язычных смычек относительно начала предшествующего гласного при произнесении звуков *п*, *б*, *м*, и *т*, *д*, *н* в словах *тапа*, *таба*, *тапа* и т. п.

Стрелками указано начало уменьшения скорости тока воздуха через нос.

смычки мы могли судить по укорочению интервала от начала смычки до начала голоса. Как видно на графиках, каждый из которых является результатом усреднения 25 осциллограмм, длительность указанного интервала при произнесении *м* и *н* меньше, чем при произнесении *б* и *д* (рис. 5.2). Однако, как и в предыдущем случае, размыкание смычки при *м* и *н* происходит позже, чем при *б* и *д*. Вследствие этого укорочение длительности

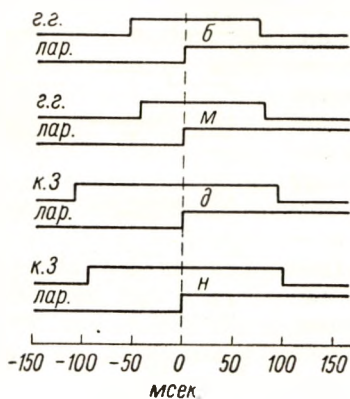


Рис. 5.2. Длительность губных и нёбно-язычных смычек и их положение относительно начала голоса при произнесении согласных *б*, *м*, *д*, *н* в начале слов *ба́та*, *ма́та*, *да́та*, *на́та*. Время (по оси абсцисс, в мсек.) отсчитано вправо и влево от начала голоса.

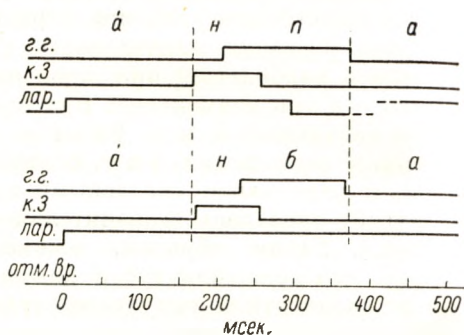


Рис. 5.3. Усредненные осциллограммы слов *анна* и *анба*, совмещенные по началу первого гласного.

назальных согласных оказывается менее значительным. Вторым диктором взрывные согласные *п*, *б*, *т*, и *д* были произнесены в двух фонетических положениях — в интервокальном и после согласных *м* и *н*. При произнесении указанных звуков в интервокальном положении им предшествовал согласный *н* и гласный *а* (слова *на́па*, *на́ба*, *на́та* и *на́да*). При записи артикуляции этих согласных в сочетаниях с *м* и *н* были использованы слова *а́нна*, *а́нба*, *а́нта*, *а́нда*, *а́мпа* и т. п. Каждое из 12 исследованных слов было произнесено от 25 до 42 раз.

Как и в предыдущем опыте, на полученных осциллограммах были измерены длительность согласных и длительность интервала от начала предшествующего гласного до начала смычки взрывного согласного. В словах типа *анпа* и т. п. длительность взрывных согласных могла быть измерена только в тех случаях, когда им предшествовал согласный с другим местом образования.

Как показали измерения, при произнесении глухих и звонких согласных во всех исследованных положениях между ними сохраняются те же самые различия, которые были обнаружены

у первого диктора. Укорочение смычки звонких *б* и *д* происходило преимущественно за счет задержки начала их возникновения. Размыкание смычки при артикуляции *б* и *д* происходило лишь немного раньше, чем при артикуляции *п* и *т* (рис. 5.3).

Таким образом, при артикуляции глухих, звонких и назальных согласных во всех исследованных положениях у обоих дикторов наблюдались одни и те же явления. Во всех случаях имело место укорочение смычек звонких и назальных согласных по сравнению с длительностью смычек глухих согласных с тем же местом образования. Это укорочение сопровождалось удлинением предшествующих гласных, т. е. являлось результатом запаздывания начала смычки. Размыкание смычки при назальных согласных происходило заметно позже размыкания смычки при звонких согласных.

Следовательно, можно утверждать, что движения губ и языка при произнесении согласных не являются стандартными и зависят от того, в составе какого комплекса движений они осуществляются.

Аналогичные различия были обнаружены при сравнении изменений скорости тока воздуха через нос на протяжении глухих и звонких согласных. Этот параметр анализировался, поскольку по его изменениям можно было судить о поведении нёбной занавески. Так как скорость тока воздуха через нос зависит не только от положения нёбной занавески, но и от целого ряда других факторов (силы выдоха, ширины голосовой щели, наличия или отсутствия преграды в полости рта и т. п.), при обработке осциллограмм принимались во внимание только те изменения регистрируемого сигнала, которые не могли быть отнесены за счет каких-либо сопутствующих явлений.

При произнесении взрывных согласных после ненанализированных гласных (в словах *тата*, *таба* и т. п.) на начальном участке смычки имело место временное увеличение скорости тока воздуха через нос (рис. 5.4). Это явление наблюдалось также и у других дикторов. Следует думать, что в начале согласного смыкание мягкого нёба с задней стенкой глотки не является полным. При этом условии увеличение сигнала после образования губной или нёбно-язычной смычки должно быть связано с повышением давления в полостях, расположенных ниже нёбной занавески, в ре-

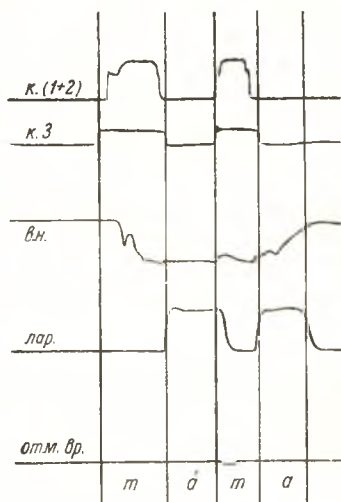


Рис. 5.4. Осциллограмма слова *тата*.

Объяснения в тексте.

зультате образования преграды в полости рта. Тогда последующее уменьшение скорости тока воздуха через нос должно быть вызвано дополнительным опусканием нёбной занавески. Исходя из этих соображений, была измерена длительность интервала от начала смычки до начала уменьшения скорости тока воздуха через нос при артикуляции каждого из четырех исследованных согласных. Если бы движения нёбной занавески были однотипными при произнесении как глухих, так и звонких согласных, наблюдае-

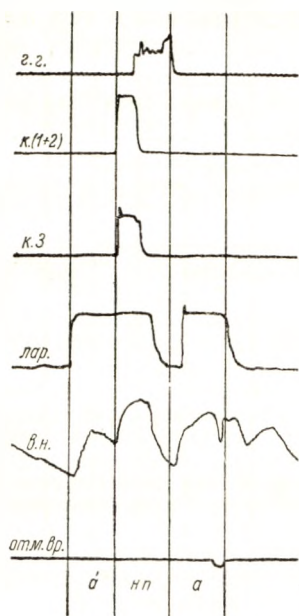


Рис. 5.5. Осциллограмма слова *анпа*.

мое уменьшение скорости тока воздуха через нос должно было бы происходить в обоих случаях в одно и то же время относительно начала предшествующего гласного. При этом длительность измеряемого интервала при артикуляции звонких согласных должна была бы уменьшаться вследствие запаздывания момента образования смычки.

Однако, как показали измерения, имеет место противоположное явление, а именно задержка начала смычки при артикуляции звонких согласных сопровождается запаздыванием подъема нёбной занавески (см. рис. 5.1).

Та же самая зависимость была обнаружена у второго диктора. Исследованные согласные произносились этим диктором в тех фонетических положениях, в которых они называются: после слога *на* и после назальных согласных *м* и *н*. Таким образом, мы имели возможность проследить изменения сигнала более подробно.

Как было обнаружено, при произнесении взрывных согласных в первом из исследованных фонетических положений во всех случаях имела место назализация начальных участков согласных. Аналогичным образом при произнесении губных согласных после *н* и язычных согласных после *м* уменьшение скорости тока воздуха через нос, которое следовало связать с подъемом нёбной занавески, в подавляющем большинстве случаев начиналось несколько позднее размыкания смычки первого согласного и, следовательно, значительно позднее начала смычки второго согласного (рис. 5.5).

Во всех исследованных ситуациях последующее уменьшение скорости тока воздуха происходило постепенно, так что сигнал достигал минимальных значений только к концу согласного. Так как вслед за этим обычно наблюдалось новое увеличение реги-

стрируемого сигнала, на протяжении согласного нельзя было выделить сколько-нибудь длительного отрезка времени, в течение которого отсутствовало бы движение воздуха через нос, можно было только определить момент, когда величина сигнала становится минимальной. Более того, этот минимум достигался при произнесении язычных согласных в ряде случаев, а при произнесении губных согласных в большинстве случаев только после размыкания соответствующих смычек.

Поскольку нас интересовали возможные различия в поведении нёбной занавески при артикуляции глухих и звонких согласных, во всех исследованных словах были измерены интервалы от начала первого гласного до начала уменьшения скорости тока воздуха через нос на протяжении взрывного согласного и от начала гласного до минимума регистрируемого сигнала. Последний интервал не был измерен при произнесении исследованных согласных после звука *м*.

Как оказалось, средняя длительность обоих измеренных интервалов при артикуляции звонких согласных больше, чем длительность тех же интервалов при артикуляции соответствующих глухих. Следует думать, что нёбная занавеска начинает движение и достигает максимального подъема в случае звонких согласных позже, чем в случае глухих согласных.

Таким образом, результаты анализа изменений скорости тока воздуха через нос оказались однозначными у обоих дикторов. И в том и другом случае изменения в движениях губ и языка при артикуляции звонких согласных сопровождались аналогичными изменениями в поведении нёбной занавески.

Далее, как было замечено при обработке осциллограмм, ток воздуха через нос в момент максимального подъема нёбной занавески в значительном числе случаев не прекращается полностью. Можно думать, что движения нёбной занавески не всегда оказываются достаточно сильными, так что полного смыкания нёбной занавески с задней стенкой глотки не происходит. Таким образом, можно было проверить, осуществляются ли движения нёбной занавески с одной и той же силой как при артикуляции глухих, так и при артикуляции звонких согласных. По литературным данным, величина внутриротового давления при артикуляции глухих согласных больше, чем при артикуляции звонких согласных (Fischer-Jorgensen, 1963). Если бы при этом сокращение нёбной занавески происходило в обоих случаях с одной и той же силой, скорость тока воздуха в момент максимального подъема нёбной занавески должна бы быть больше при артикуляции глухих согласных и меньше при артикуляции звонких согласных. Исходя из этих соображений, мы измерили минимальные значения скорости на протяжении всех четырех согласных при их произнесении как в интервокальном положении, так и после согласного *н*.

Как оказалось, при произнесении звука *t* в обоих исследованных ситуациях скорость тока воздуха в момент максимального подъема нёбной занавески близка к нулю, поступление воздуха в полость носа прекращается почти полностью. Следует думать, что смыкание нёбной занавески с задней стенкой глотки является в этом случае достаточно полным. При произнесении звука *д* минимальная скорость тока воздуха на протяжении смычки в обоих случаях заметно превышала нулевой уровень. Соотношение между изменениями скорости тока воздуха при произнесении

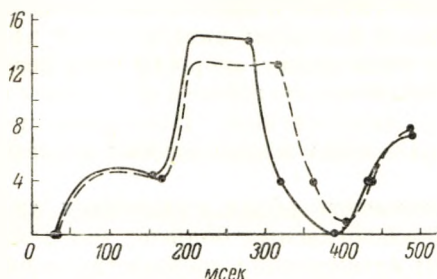


Рис. 5.6. Изменения скорости тока воздуха через нос на протяжении слов *анта* (сплошная линия) и *анда* (прерывистая линия).

По оси абсцисс — время, отсчитанное от начала первого гласного (в мсек.), по оси ординат — скорость тока воздуха (в условных единицах).

согласного *t* в слове *анта* и согласного *д* в слове *анда* иллюстрируется рис. 5.6. Для построения кривых, кроме средних скоростей в момент максимального подъема нёбной занавески, были подсчитаны средние значения регистрируемого сигнала на протяжении конечного сегмента согласного *н*, на протяжении конечного участка предшествующего гласного и на протяжении последующего гласного, а также длительность интервалов от всех этих точек до начала первого гласного. Аналогичное соотношение

сигналов было получено при произнесении *t* и *д* в интервокальном положении.

Таким образом, судя по этим данным, движения нёбной занавески при артикуляции *д* осуществляются с меньшей силой, так что полного прерывания потока воздуха не происходит.

При артикуляции *п* и *б* как в интервокальном положении, так и после *н* средние минимальные скорости тока воздуха через нос оказались практически одинаковыми и приблизительно равными тем, которые наблюдались при артикуляции *д*. Следовательно, результат в данном случае оказался менее четким. Можно думать, что смыкание нёбной занавески с задней стенкой глотки при артикуляции *п* является менее сильным, чем при артикуляции *t*, однако более сильным, чем при артикуляции *б*, иначе, как было сказано, величина регистрируемого сигнала была бы в случае *п* больше за счет большего внутриротового давления.

Таким образом, полученные данные свидетельствуют, что временные характеристики и, по-видимому, интенсивность разных движений при артикуляции звуков речи зависят от того, в составе какого комплекса движений они должны быть осуществлены. Следовательно, описание звука речи в виде набора

движений не является полным, если не указано, каким образом изменения в составе набора ведут к изменениям индивидуальных движений.

Далее, при произнесении одних и тех же звуков в разных контекстах были обнаружены существенные изменения в их длительности. Наиболее

четко это обнаружилось при сопоставлении одних и тех же звуков речи при их произнесении в начале слова и в других фонетических позициях. Как видно на рис. 5.7, полученном при усреднении 68 осциллограмм, длительность смычки начального *н* в слове *наппа* в три раза превышает длительность смычки второго *н*. Следовательно, говорить об инвариантности артикуляторных комплексов относительно контекста оказывается невозможным.

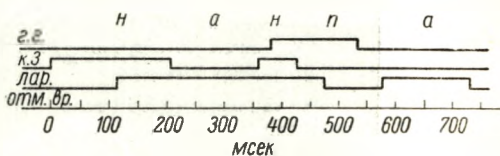


Рис. 5.7. Усредненная осциллограмма слова *наппа*. Отсчет времени (по оси абсцисс, в мсек.) от начала первого согласного.

Из всего сказанного становится очевидным, что поиски коррелятов дифференциальных признаков на уровне движений также необоснованы, как и на акустическом уровне. Для адекватного описания речевого процесса необходимо физиологическое исследование механизмов речеобразования. По-видимому, только в результате такого исследования могут быть найдены реальные принципы структурирования речевого потока.

2. Движения губ и нижней челюсти при артикуляции губных смычных согласных

Результаты, приведенные в предыдущем разделе данной главы, показали, что глухие смычные согласные характеризуются большей длительностью смычки, чем звонкие смычные согласные. Было высказано предположение, что это обусловлено большей силой сокращения мышц, обеспечивающих смычку при артикуляции глухого согласного. С физиологической точки зрения, такое предположение не является достаточно конкретным, так как смычка является результатом координированной деятельности целого комплекса мышц. Например, смыкание губ при *п*, *б*, *м* осуществляется как благодаря сокращению собственно губных мышц (*m. orbicularis oris*), так и благодаря сокращению мышц, поднимающих нижнюю челюсть. С другой стороны, длительность смычки должна определяться временными отношениями между сокращениями мышц-антагонистов, например мышц, поднимающих и опускающих нижнюю губу, и мышц, поднимающих и опускающих нижнюю челюсть.

Описываемые далее измерения были предприняты, чтобы в какой-то мере разобраться в этой заведомо сложной картине. В первой серии измерений применялась непрерывная фотоосцил-

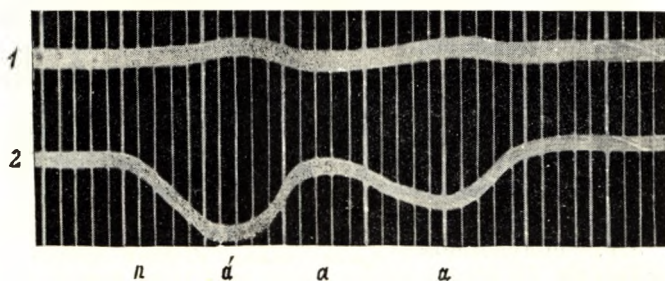


Рис. 5.8. Осциллограмма слова *papa*, произнесенного разговорным стилем.

1 — движения верхней губы; 2 — движения нижней губы. Вертикальные линии — отметка времени 20 мсек.

логографическая регистрация движений губ в вертикальной плоскости при произнесении слов *papa*, *баба*, *мама*. Методика регистрации описана в главе 2. В опытах участвовали 4 диктора, каждое слово повторялось по 4—8 раз в разговорном и полном стиле речи.

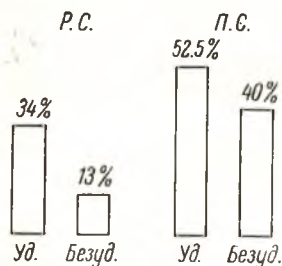


Рис. 5.9. Частота наличия стационарного участка в осциллограммах гласных в словах *papa*, *баба*, *мама*.

Р. С. — разговорный стиль, П. С. — полный стиль речи; Уд — ударный слог, Безуд. — безударный слог.

Полученные осциллограммы показали, что движение верхней губы ничтожно мало по сравнению с движением нижней (рис. 5.8). Поэтому казалось целесообразным на данном этапе работы ограничиться исследованием значительно более выраженного эффекта — перемещений нижней губы. Необходимо подчеркнуть, что движение нижней губы является сложным эффектом, складывающимся из движения нижней губы относительно нижней челюсти и перемещения нижней челюсти.

На осциллограмме, приведенной на рис. 5.8, трудно выделить стационарные участки. Чтобы определить, насколько типичным является наличие или отсутствие стационарных участков на протя-

жении произнесения гласных, все имевшиеся в распоряжении осциллограммы были оценены с этой точки зрения.

Считалось, что стационарный участок существует, если на кривой можно было выделить отрезок, длительностью не менее 20 мсек., на протяжении которого перемещение нижней губы не превышало 0.7 мм.

Полученные данные (рис. 5.9) показали, что стационарные участки наблюдаются практически только при полном стиле произнесения и преимущественно на ударных гласных. Это заставляет считать, что положение, к которому стремятся артикуляторные органы на протяжении гласного, в большинстве случаев не достигается.

Для удобства обработки осциллограмм каждая из них рассматривалась как последовательность отрезков. На рис. 5.10 показаны обозначения отрезков осциллограмм и их параметров, значения которых приводятся ниже.

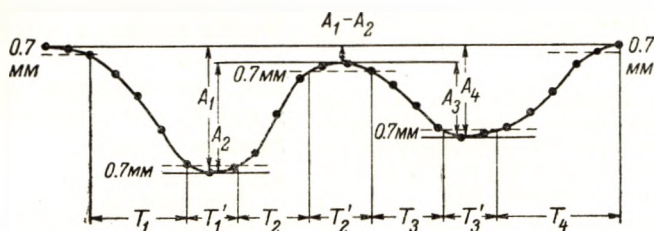


Рис. 5.10. Схема измерений осциллограммы движения нижней губы.

Объяснения в тексте.

1. Переход от согласного к гласному. Характеризуется временем перехода T_1 , величиной смещения A_1 и средней скоростью $\frac{A_1}{T_1}$.

2. Стационарный участок первого гласного. Характеризуется длительностью T_1' и величиной смещения A_1 . В тех случаях, когда T_1' было меньше 20 мсек., считалось, что стационарного участка нет.

3. Переход от первого гласного к согласному. Характеризуется временем перехода T_2 , величиной смещения A_2 и средней скоростью $\frac{A_2}{T_2}$.

4. Стационарный участок согласного. Характеризуется длительностью T_2' и величиной смещения $A_1 - A_2$.

5. Переход от второго согласного ко второму гласному. Характеризуется временем T_3 , величиной смещения A_3 и средней скоростью $\frac{A_3}{T_3}$.

6. Стационарный участок второго гласного. Характеризуется длительностью T_3' и величиной смещения A_4 .

7. Переход от второго гласного к окончанию слова. Характеризуется временем T_4 , величиной смещения A_4 и средней скоростью $\frac{A_4}{T_4}$.

Предварительная оценка всего материала в целом показала, что произнесение при полном стиле отличается от такового при разговорном большей длительностью фазы раскрытия губ, большей величиной максимального смещения губы на фазе раскрытия (в 1.5—4.0 раза) и большей скоростью движения как при поднимании, так и при опускании нижней губы.

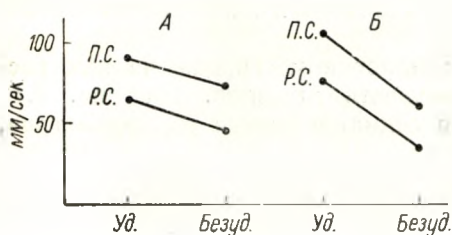


Рис. 5.11. Средние значения скорости (по оси ординат, в мм/сек.) перемещения нижней губы на переходах от согласного к гласному (А) и от гласного к согласному (Б).

Обозначения как на рис. 5.9.

баба диктором М. Л. разговорным (А) и полным (Б) стилем и кривые, отражающие значения скорости движений нижней губы за каждые 20 мсек. на протяжении всего слова. Отчетливо видно, что скорость при полном стиле произнесения больше, чем при разговорном.

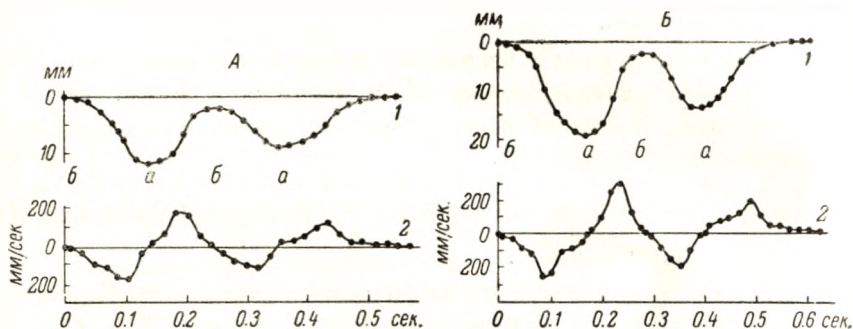


Рис. 5.12. Графики движений нижней губы (1) и их скорости (2) при разговорном (А) и полном (Б) стиле произнесения.

По оси абсцисс — время (в сек.); по оси ординат: для кривых 1 — движение (в мм), для кривых 2 — скорость (в мм/сек.).

Большая скорость движений при полном стиле, характеризующимся замедленным темпом речи, противоречит распространенному в фонетике представлению о том, что быстрый темп речи характеризуется также большими скоростями каждого отдельного движения (Stetson, Hudgins a. Moses, 1940). Этот факт противо-

речит также выводам, к которым пришли Кэлли, Антони и Алдалл (Kelly, Anthony a. Uldall, 1962) на основании исследования скорости переходов второй форманты гласного в сочетаниях губной согласный—гласный. Они нашли, что изменения темпа не влияют на скорость изменения частоты форманты. К сожалению, в сообщении не приводится достаточных количественных данных, которые позволили бы оценить справедливость этого заключения.

Исходя из физиологических данных, следует считать, что скорость, как и амплитуда движений, должна в первую очередь

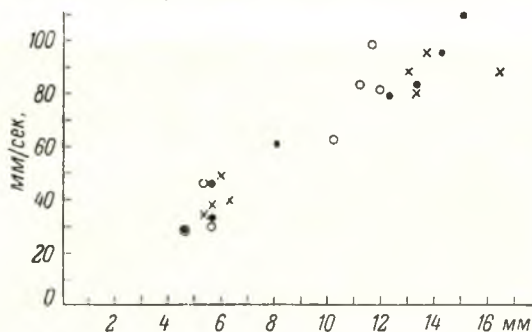


Рис. 5.13. Зависимость скорости движения нижней губы на отрезке перехода от первого гласного к согласному (по оси ординат, в мм/сек.) от величины максимальной амплитуды опускания нижней губы во время произнесения первого гласного (по оси абсцисс, в мм) при произнесении слов *papa* (точки), *мама* (крестики), *баба* (кружки).

определяться числом синхронно сокращающихся мышечных волокон, которое, в свою очередь, должно определяться числом возбужденных мотонейронов. Очевидно, что число возбужденных мотонейронов может меняться как в зависимости от характера импульсации в управляющей посылке, поступившей из высших отделов нервной системы (числа возбужденных волокон в эффекторном пути и частоты импульсации в каждом волокне), так и в зависимости от исходного состояния мотонейронов.

Разница в амплитуде и скорости движений при разговорном и полном стиле речи, вероятно, не может быть отнесена за счет различий в возбудимости мотонейронов. Однако можно предположить, что изменения возбудимости играют роль в стабилизации (уменьшении вариабельности) времени переходов от одного положения артикуляторного органа к другому его положению. Полученный материал показал, что амплитуда опускания нижней губы при произнесении первого слога (гласного) значительно варьировала по своей величине при многократных повторениях. Если бы скорость возвращения губы при переходе к следующему

слогу не зависела от амплитуды предыдущего опускания, то следовало бы ожидать также больших колебаний длительности этого перехода (возвращения). На самом деле время возвращения оказывалось довольно стандартным. Это заставляло предполагать, что скорость возвращения является функцией от положения, в котором находится орган в момент поступления следующей управляющей команды. Для проверки предположения были составлены графики типа рис. 5.13. Оказалось, что для всех дикторов существует отчетливая зависимость между амплитудой

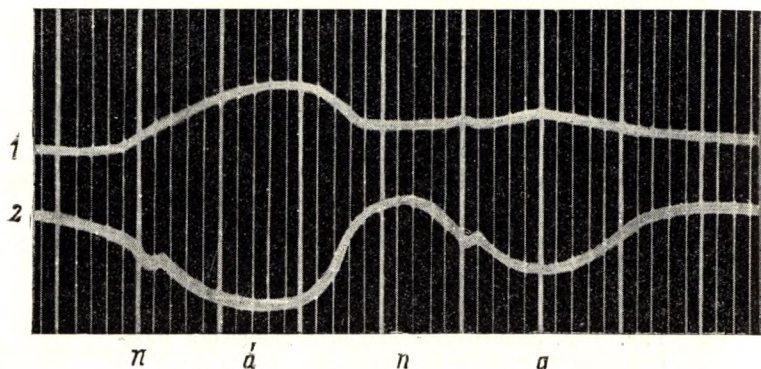


Рис. 5.14. Осциллограмма слова *nana*, произнесенного разговорным стилем.

1 — перемещения верхней губы; 2 — перемещения нижней губы; вертикальные линии — отметка времени 20 мсек.

опускания нижней губы на гласном предыдущего слога и скоростью подъема ее при переходе к следующему слогу, более выраженная при разговорном стиле произнесения. Очевидно, это явление может быть обеспечено механизмами регулирования, действующими на самых низких уровнях управления движениями. Опускание губы (нижней челюсти) должно приводить к рефлекторному повышению возбудимости центров антагонистических мышц, причем возбудимость должна повышаться тем больше, чем большим является опускание губы (нижней челюсти).

Перейдем к сравнению осциллограмм, соответствующих словам *nana*, *мама*, *баба*. В работе Фуджимура (Fujimura, 1961b), проведенной на одном дикторе, было показано, что опускание нижней губы после смычки согласного может иметь неравномерный характер, выражающийся в изломе кривой на начальном участке. Подобное явление наблюдалось нами систематически у одного диктора и было преимущественно выражено в случае артикуляции *п* (рис. 5.14). Фуджимура считает, что этот эффект вызывается толчком воздуха в момент взрыва согласного. Очевидно, что он не

имеет отношения к интересующим нас различиям в мышечной активности при артикуляции *п*, *б*, *м*.

Для выявления этих различий было произведено сравнение значений длительности трех участков артикуляции *п*, *б*, *м*, а

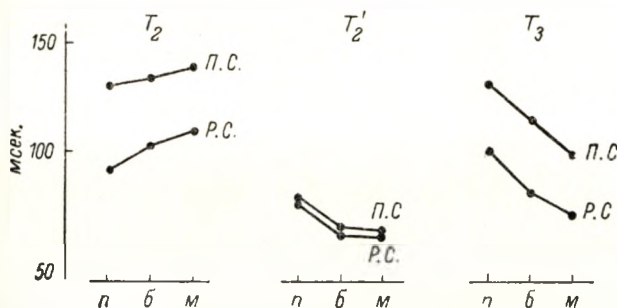


Рис. 5.15. Длительность различных участков интервокального согласного (по оси ординат, в мсек.) в словах *папа*, *баба*, *мама* при произнесении их полным и разговорным стилем.

Объяснения в тексте.

именно перехода от гласного к согласному (T_2), стационарного участка согласного (T_2') и перехода от согласного к гласному (T_3) (рис. 5.15). В соответствии с данными, описанными в предыдущем разделе главы, участок смычки согласного (стационарный участок T_2') оказался наиболее длинным в случае глухого согласного *л*. Однако длительность перехода от гласного к согласному (T_2) в случае *п* оказалась наименьшей. Укорочение длительности перехода при *п* обусловлено большей скоростью движения (рис. 5.16, А). Таким образом, можно утверждать, что увеличение длительности смычки глухого согласного должно быть отнесено в основном за счет большей силы (скорости) смыкания губ. Если принять за длительность согласного интервал времени от начала подъема нижней губы до начала обратного движения ($T_2 + T_2'$), то различия между согласными практически исчезают (рис. 5.16, Б).

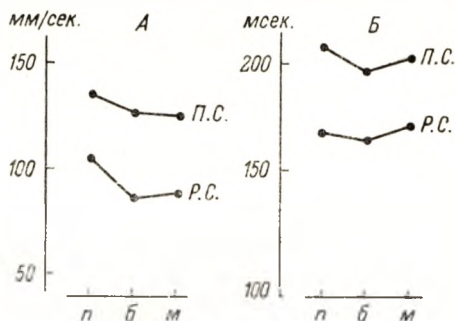


Рис. 5.16. Средняя скорость движения нижней губы (по оси ординат, в мм/сек.) при переходе от гласного к согласным *п*, *б*, *м* (А) и средняя длительность интервала (по оси ординат, в мсек.) от начала подъема нижней губы до начала опускания при артикуляции согласных *п*, *б*, *м* (Б).

ней челюсти. Исследование задержки момента максимального подъема нижней челюсти относительно начала звуко сочетания привело к парадоксальному заключению. Оказалось, что задержка минимальна в случае *м* и максимальна в случае *п*. Это заставляет считать, что скорость (а следовательно, и сила) движения челюсти является наименьшей в случае глухого согласного. В случае артикуляции губных согласных мышцы, поднимающие нижнюю челюсть, и мышцы, смыкающие губы, работают как синергисты. Обнаруженный факт позволяет предположить, что между этими группами существуют такие координационные отношения, которые обеспечивают возможность компенсации движения одной группы за счет другой.

Для проверки этого предположения были сняты осциллограммы движения губ и (для контроля) неподвижной нижней челюсти в условиях, когда диктор произносил слова *папа*, *баба*, *мама* с сомкнутыми зубами. При этом оказалось, что максимальное перемещение нижней губы относительно нижней челюсти в период произнесения гласных в два раза больше по сравнению с собственным перемещением нижней губы во время обычной речи (при движении нижней челюсти).

3. Временные отношения между смычкой и фонацией при артикуляции глухих смычных согласных

Анализ, проведенный в предыдущих главах, привел нас к простейшему артикуляторному комплексу типа СГ. Очевидно, физиологическое исследование такого рода элементов должно быть направлено уже на выяснение конкретных координационных отношений между движениями, параллельно участвующими в их артикуляции. Если слог СГ представляет собой организованную систему, то можно ожидать как простых, так и самых сложных и многообразных связей между образующими его движениями.

В настоящем разделе будут приведены данные, касающиеся временных отношений в такой паре явлений, как окончание смычки глухого согласного и начало фонации (переход от глухого согласного к гласному).

На осциллограмме слова *даты* (рис. 5.18) отмечен интервал τ между интересующими нас явлениями: прекращением смычки *к*. (*1+2*) и началом фонации (появление сигнала с ларингофона). С фонетической точки зрения, длительность этого интервала может складываться из длительности взрыва, длительности аффрицированной части в случае мягких согласных и длительности придыхания, которое Л. В. Щерба (1912) трактует как глухой участок гласного.

Исследование этого интервала τ представляет большой интерес, поскольку оно может помочь в решении вопроса о поведении голосовых связок при переходах от артикуляции согласного

к артикуляции гласного. А это, в свою очередь, должно помочь решению вопроса о некоторых механизмах управления артикуляцией.

Работе голосовых связок посвящен целый ряд исследований, но все они в основном рассматривают поведение связок только во время фонации гласного (Хохлов, 1960; Tarnóczy, 1951; van den Berg, 1956; van den Berg, Zantema, Doornenbal, 1957; Floyd. Negus a. Neil, 1957; Flanagan, 1958, 1959; Moore a. von Leden, 1958; Timcke, von Leden, Moore, 1958, 1959; Draper, Ladefoged, Whitteridge, 1959; Isshiki, 1961; Shimizu, 1961).

Нам не удалось найти работ, которые рассматривали бы поведение голосовых связок при артикуляции согласного в речевом потоке. Отсутствие такого рода работ объясняется большими методическими трудностями регистрации движений голосовых связок во время речи.

Из теоретических соображений очевидно, что для начала колебаний голосовых связок требуется выполнение следующих двух условий: голосовые связки должны быть сближены (сведены) и перепад давлений — надсвязочного и подсвязочного — должен достичь некоторого критического уровня. Можно представить себе следующие временные отношения в реализации этих условий:

1) голосовые связки достигают необходимого сближения, позволяющего начать голосообразование еще до окончания смычки согласного; в этом случае задержка начала фонации относительно конца смычки будет определяться лишь временем образования необходимого перепада давлений на связках, т. е. временем, в течение которого внутриротовое давление будет понижено до некоторого критического уровня;

2) голосовые связки достигают необходимого сближения лишь после прекращения смычки; тогда, если время сведения связок достаточно велико по сравнению со временем выравнивания внутриротового давления, длительность задержки начала фонации должна быть обусловлена временным интервалом сведения связок.

Если правильно первое предположение, т. е. что задержка начала фонации определяется временем выравнивания внутриро-

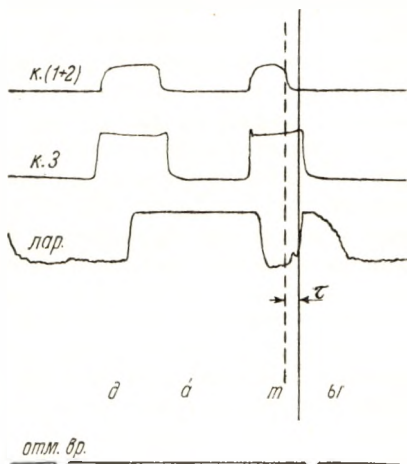


Рис. 5.18. Осциллограмма слова *даты*.

Объяснения в тексте.

что механизм управления работой связок в случае смычных и щелевых согласных одинаков.

Интересно рассмотреть распределение длительностей интервала между раскрытием смычки и началом голосообразования в случаях говорения с трубкой, когда наружное отверстие ее закрыто (рис. 5.20, 1) и когда открыто (рис. 5.20, 2). Можно видеть, что кривые особенно сильно различаются в своей начальной левой части. Правые же части кривых почти одинаковы. Область

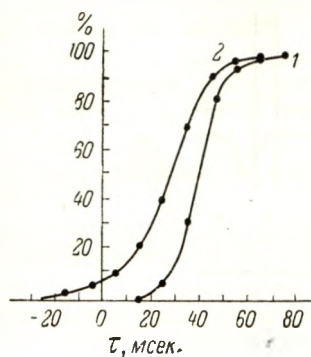


Рис. 5.20. Накопленная гистограмма длительностей интервала τ (в мсек.) при нормальном внутриротовом давлении (1) и при уменьшенном внутриротовом давлении (2).

значений, в которой лежит кривая 2 (от -25 мсек. до 65 мсек.), значительно превышает область значений, в которой лежит кривая 1 (от 15 мсек. до 75 мсек.).

Следующая задача состояла в исследовании зависимости длительности интервала τ между раскрытием смычки и началом фонации от фонетического его окружения (качества предшествующего согласного и последующего гласного) у разных дикторов.

Данные были получены на материале следующих слов и фраз: *даты, затычка, тычок, топить, тапочка, топятся, Тима, темень, темно, вытемнить, стукать, выстукать, у попа папа, а попа папа*. Они произносились в среднем темпе и повторялись от 50 до 70 раз. В эксперименте участвовало 4 диктора (2 женщины и 2 мужчин).

Полученные данные показали, что длительность исследуемого интервала меняется от диктора к диктору и зависит от твердости или мягкости согласного, характер же гласного на нее не влияет. Ниже приводятся средние значения интервала τ (в мсек.), полученные на материале двух дикторов. Разница между дикторами оказалась существенной ($P > 0.95$):

	Диктор Ю. К.	Диктор Л. Б.
<i>даты</i>	50	33
<i>затычка</i>	53	26
<i>тычок</i>	50	25
<i>тапочка</i>	36	23
<i>Тима</i>	111	83
<i>стукать</i>	35	21

Одной из возможных причин таких различий может быть разница в исходном альвеолярном давлении у разных дикторов (Malecot, 1955; Fischer-Jørgensen, 1963; Ladefoged a. McKinney, 1963). Так как во время смычки глухого согласного поток воз-

духа из легких полностью прекращается, то естественно считать, что внутриротовое давление за местом смычки соответствует альвеолярному давлению. Результаты измерений внутриротового давления у разных дикторов действительно показали большие различия. Кроме того, было обнаружено, что чем больше внутриротовое давление у данного диктора, тем меньше у него длительность интервала между раскрытием смычки и началом фонации (рис. 5.21). Эту же закономерность можно получить и на одном дикторе (рис. 5.22), заставляя говорить его с различной громкостью, поскольку изменения громкости связаны с изменениями как подсвязочного, так и внутриротового давления.

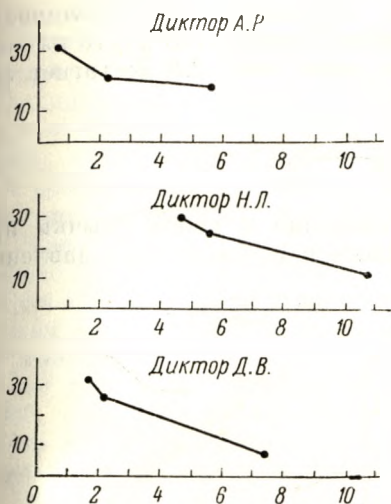


Рис. 5.21. Зависимость длительности интервала τ между концом смычки и началом фонации (по оси ординат, в мсек.) от величины внутриротового давления (по оси абсцисс, в см H_2O) при произнесении согласного n разными дикторами.

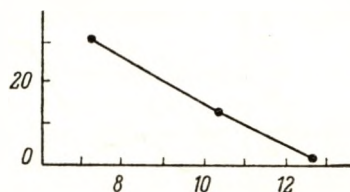


Рис. 5.22. Зависимость длительности интервала τ между концом смычки и началом фонации (по оси ординат, в мсек.) от величины внутриротового давления (по оси абсцисс, в см H_2O) при произнесении согласного n одним диктором с различной громкостью.

На первый взгляд может показаться, что эта зависимость противоречит тому, что мы наблюдали в случае говорения с открытой на конце трубкой, когда с уменьшением надсвязочного давления длительность интервала τ уменьшалась.

В действительности никакого противоречия здесь нет. Во время фазы взрыва происходит выравнивание внутриротового давления с атмосферным, ход которого в первом приближении может быть описан экспоненциальной зависимостью

$$P_{p.} = P_{p. н.} e^{-\alpha t},$$

где $P_{p.}$ — текущее значение внутриротового давления, $P_{p. н.}$ — внутриротовое давление в момент начала размыкания смычки,

α — постоянный коэффициент, t — время, отсчитываемое от момента начала размыкания смычки.

Отсюда переменный перепад давлений над и под связками (ΔP) может быть определен, как

$$\Delta P = P_{\text{пс}} - P_{\text{р. п.}} = P_{\text{пс}} - P_{\text{р. п.}} e^{-\alpha t},$$

где $P_{\text{пс}}$ — подсвязочное давление. Так как фонация начинается при определенной величине перепада давлений ($\Delta P_{\text{ф.}}$), то из этого выражения можно найти время τ , за которое ΔP достигнет значения $\Delta P_{\text{ф.}}$:

$$\tau = \frac{1}{\alpha} \ln \frac{P_{\text{р. п.}}}{P_{\text{пс}} - \Delta P_{\text{ф.}}}$$

При нормальных условиях говорения к концу смычки происходит выравнивание надсвязочного и подсвязочного давлений,

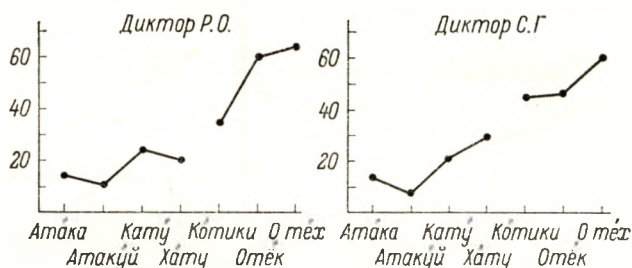


Рис. 5.23. Длительность интервала τ между концом смычки и началом фонации (по оси ординат, в мсек.) при произнесении t и t' в составе разных слов.

т. е. $P_{\text{р. п.}} = P_{\text{пс}}$. При этом условии оказывается, что чем больше будет величина этих давлений, тем ближе к единице будет выражение, стоящее под знаком логарифма, и тем меньше будет интервал τ , т. е. тем быстрее начнется фонация.

В случае же говорения с трубкой, вставленной в рот, выравнивания внутриротового и подсвязочного давлений не происходит, и если говорить с постоянной громкостью (что и соблюдалось в соответствующих экспериментах), то $P_{\text{пс}}$ можно считать величиной постоянной. Из последнего условия следует, что в этом случае с уменьшением $P_{\text{р. п.}}$ значение τ будет также уменьшаться. Таким образом, искусственное понижение внутриротового давления во время фазы смычки глухого согласного приводит к укорочению задержки начала фонации.

Как уже было сказано, длительность задержки начала фонации существенно различна для группы твердых и группы мягких согласных. Это подтверждают приводимые ниже средние значе-

ния интервала τ (в мсек.), полученные при стократных произнесениях одним диктором различных слов с мягким и твердым m .

После m	После m'
24.0	52.7
27.5	56.6
28.3	60.7
28.3	83.3

Для проверки этого факта были проведены дополнительные измерения. С помощью электронного измерителя статистического распределения длительности интервала (см. главу 2) были определены длительности интервала τ в словах: *атака, атакуй, коту, хату, котики, отек, о тех*. Каждое слово повторялось 100 раз. В опыте участвовало 2 диктора. На рис. 5.23 приведены средние длительности интервала τ для разных слов.

Увеличение длительности интервала τ при мягких согласных можно объяснить следующим образом. Известно, что «палатализация влечет за собой существенные изменения в основной артикуляции согласного. Это имеет место особенно при палатализации переднеязычных смычных, приобретающих при этом нередко более или менее выраженный аффрикатный характер» (Зиндер, 1960, стр. 134). Палатализованные смычные согласные (прежде всего m' , k' , n') характеризуются тем, что их последнюю фазу составляет не взрыв, а щелевой элемент.

В случаях аффрикатизации щелевая фаза, заменяющая фазу взрыва, для m' иногда достигает 42% всей длительности согласного. Поскольку в мягких согласных последнюю фазу составляет не взрыв, а щелевой элемент, сопротивление выходного отверстия рта в этих случаях возрастает по сравнению с сопротивлением выходного отверстия при твердых согласных. За счет этого меняются условия выравнивания внутриротового и атмосферного давлений, что должно приводить к увеличению задержки начала фонации гласного.

Таким образом, подводя итог, можно полагать, что голосовые связки сближаются еще до раскрытия смычки и длительность задержки фонации определяется только временем образования нужного перепада надсвязочного и подсвязочного давлений.

В настоящее время не существует сколько-нибудь определенной теории восприятия речи, хотя попытки рассмотреть эту проблему неоднократно предпринимались в течение последних нескольких лет (Чистович, 1961, 1962; Licklider, 1952; Fry a. Denes, 1958; Ladefoged, 1958; Stevens, 1960; Liberman, Cooper, Harris a. MacNeilage, 1962, и др.). В результате этих попыток отчетливо выявились основные вопросы, без решения которых принципиально невозможно построить теорию восприятия речи.

Первым и кардинальным теоретическим вопросом является вопрос о единицах восприятия — единицах решений и о процедурах принятия решений, используемых человеком. Как вполне резонно утверждает Миллер (Miller, 1962), наибольшая надежность распознавания речи обеспечивалась бы в том случае, если бы единицы восприятия были максимально крупными — соответствовали, например, предложениям или синтагмам. При этом было бы нецелесообразным принимать текущие решения о более мелких элементах речи, так как принятие решений означает сокращение информации и может быть сопряжено с ошибками. Однако для того чтобы предложения могли быть единицами восприятия, необходимо, чтобы весь возможный набор предложений был заранее известен человеку. Как уже говорилось ранее (см. главу 1), это допущение не является реальным. Необходимо принять, что распознавание предложения происходит не путем идентификации его с образом одного из множества предложений, хранимых в памяти, но путем лингвистического анализа. Для того чтобы анализ стал возможным, необходимо, однако, перейти от динамической спектрограммы (физической картины) к смыслоразличительным элементам. Иначе говоря, это означает, что должно быть текущее принятие решений о сравнительно мелких смыслоразличительных единицах.

Кроме того, для того чтобы пользоваться предложением в качестве первичной единицы распознавания, необходимо иметь возможность запомнить весь его звуковой образ на короткое время. Для запоминания всей информации, содержащейся в динамической спектрограмме предложения, требуется колоссальная оперативная память. Как показывает сам Миллер (Miller, 1956), объем оперативной памяти человека очень ограничен. Это подтверждается результатами опытов по восприятию звуковых последовательностей (Чистович, Клаас, Алекин, 1961), которые показывают, что для запоминания последовательности сигналов человеку необходимо текущим образом принять решения о каждом из них.

Итак, вопреки точке зрения Миллера приходится считать, что человек текущим образом перекодирует воспринимаемый им сигнал в последовательность символов, соответствующих каким-то смыслоразличительным единицам. Так как разборчивость предложения значительно выше разборчивости изолированных зву-

ков и слогов, очевидно, что решения о мелких смысловоразличительных единицах не должны быть категоричными и окончательными. Иначе говоря, приходится предположить, что хотя человек и принимает текущие решения о мелких смысловоразличительных элементах, он делает это каким-то весьма удачным способом, допускающим снятие существовавшей неопределенности на следующих этапах процесса распознавания предложения.

Наиболее распространена точка зрения, что первичные текущие решения, принимаемые человеком, соответствуют фонемам. Однако существуют также определенные доводы в пользу того, что до принятия решения о фонеме человек опознает характеристики артикуляции слышимого сигнала (Чистович, 1960а, б, 1961, 1962; Liberman, 1957; Liberman, Cooper, Harris a. MacNeilage, 1962). Высказывались также предположения о том, что первичные единицы восприятия соответствуют слогам (Cooper, Delattre, Liberman, Borst a. Gerstman, 1952).

Мы постараемся дальше показать, что противоречия между этими гипотезами могут быть устранены, если допустить, что процедура принятия фонемных решений у человека не совпадает с той, которую пытались использовать в ряде технических устройств для автоматического распознавания фонем. Чтобы разобраться в этом вопросе, полезно сначала рассмотреть основные допущения, используемые при автоматическом распознавании, и выяснить, в какой мере они согласуются с существующими экспериментальными данными о распознавании фонем человеком.

1. Состояние вопроса о восприятии речи

Наиболее распространенная гипотеза пофонемного распознавания основывается на предположении о том, что акустический речевой поток непосредственным образом отражает фонемную последовательность; он состоит из последовательности звуков речи, каждый из которых в результате соответствующей обработки может быть отождествлен с определенной фонемой. Хотя известно, что характеристики звуков речи в потоке зависят от окружения, предполагается, что все звуки речи, относимые лингвистами к одной фонеме, обладают некоторыми общими инвариантными признаками, по которым они отличаются от звуков, относимых к другим фонемам. При этом сама фонема рассматривается как определенная комбинация (набор) инвариантных акустических признаков.

С точки зрения этой гипотезы задача распознавания может считаться решенной, если будет правильно выбран принцип классификации фонем по инвариантным акустическим признакам и будут найдены методы автоматического обнаружения этих акустических признаков. Наиболее серьезным примером попытки классификации фонем на акустическом уровне является система

На основании этих данных сотрудники Хаскинских лабораторий (Cooper, Delattre, Liberman, Borst a. Gerstman, 1952; Liberman, Ingeman, Lisker, Delattre a. Cooper, 1959) предполагают, что слог является той минимальной единицей, на уровне которой возможен переход от акустического сигнала к смысло-различительным элементам языка. Отсутствие однозначных соответствий между звуками речи и фонемами объясняется ими с точки зрения моторной теории восприятия речи (Liberman, 1957; Liberman, Cooper, Harris a. MacNeilage, 1962).

Согласно этой теории, в основе фонемной классификации лежат не акустические, а артикуляторные признаки речевых элементов. Различные акустические сигналы, возникающие при произнесении одного и того же согласного в разных слогах, относятся человеком к одной и той же фонеме не потому, что они обладают некоторыми общими для всех них акустическими признаками, но потому, что при производстве этих сигналов используются одни и те же комплексы артикуляторных движений. Если бы человек не знал, что эти сигналы производятся одним и тем же способом, он не имел бы оснований объединить их в одну группу.

Рассматривая образование системы фонемной классификации в процессе становления речи у ребенка, моторная теория подчеркивает роль явлений имитации воспринимаемых звуков. В процессе имитации создаются условия, благоприятные для образования условнорефлекторных связей между группами звуковых сигналов и комплексами артикуляторных движений (более подробное обсуждение этих вопросов см.: Чистович, 1961, 1962; Liberman, Cooper, Harris a. MacNeilage, 1962). Предполагается, что эти условнорефлекторные связи играют существенную роль также и в процессе распознавания речи взрослым человеком.

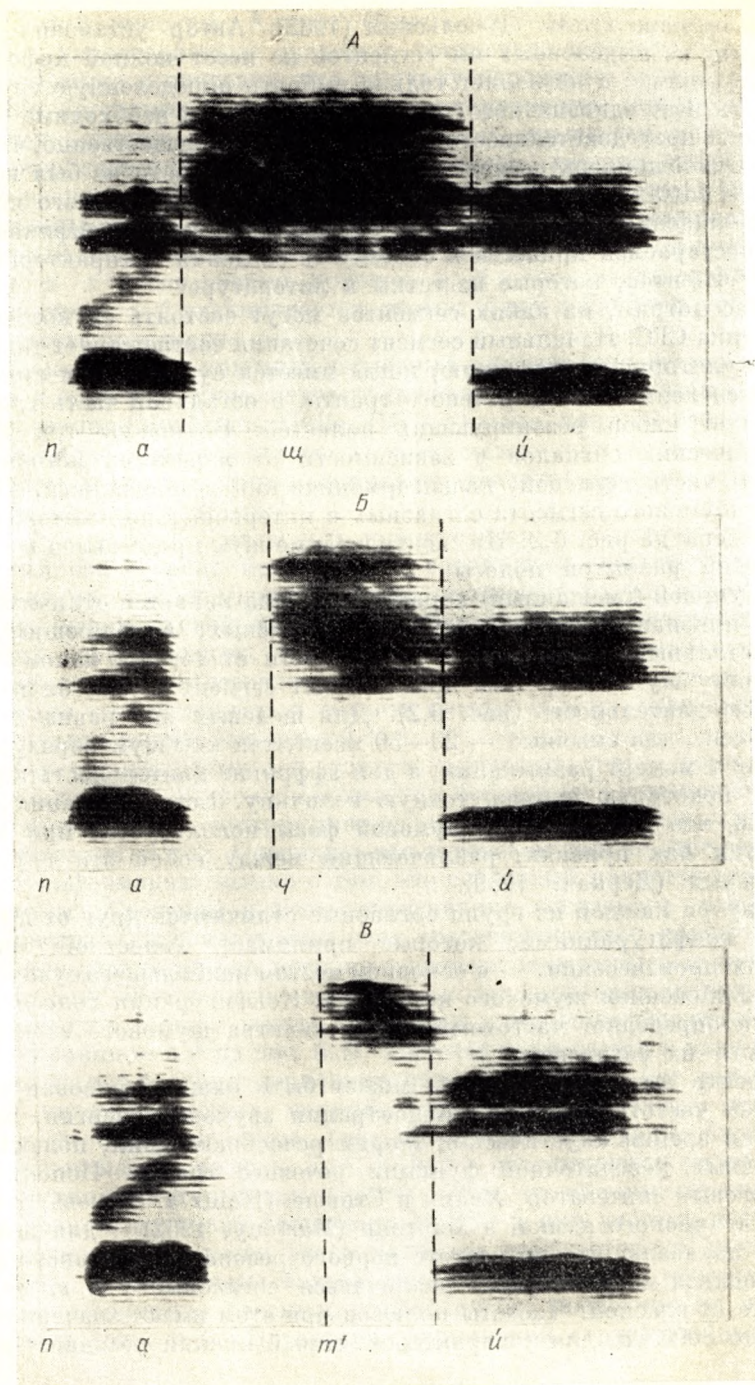
Из моторной теории восприятия следует весьма существенный практический вывод: поиски инвариантных признаков фонем на акустическом уровне, на которые уже было затрачено очень много труда и времени, вряд ли приведут к положительным результатам.

В соответствии с особенностями артикуляции слог в акустическом выражении представляет собой последовательность сигналов, каждый из которых соответствует определенному изменению в состоянии речевого тракта и, таким образом, несет информацию о тех или иных признаках фонем, описывающих этот слог.

Первая систематическая попытка рассмотрения слога как последовательности сигналов (сегментов), экспериментального выделения этих сегментов и изучения особенностей их восприятия

Рис. 6.2. Динамические спектрограммы звукосочетов *пацц* (А), *пачч* (В) и *патт* (В).

Пунктиром отмечены границы шумовых сегментов согласных *ц*, *ч* и *т*.



была сделана Н. И. Дукельским (1962). Автор установил, что каждый из выделенных им сегментов не несет полной информации о фонеме, а позволяет только выбрать определенную группу фонем. Для однозначного опознавания фонемы необходимо воспринять последовательность сегментов, причем существенно, чтобы временной порядок сегментов в последовательности не был искажен. К сожалению, Н. И. Дукельским не было дано четкого принципа членения на сегменты, и не осуществлена их классификация. Мы постараемся привести в систему те сведения о характеристиках сегментов, которые известны в литературе.

Рассмотрим, из каких сегментов могут состоять звуко сочетания типа СГС. Начальный сегмент сочетания соответствует такому артикуляторному состоянию, когда имеется сужение или смычка в определенном месте речевого тракта, а остальная часть тракта образует набор резонирующих полостей. Возможны три типа акустических сигналов в зависимости от характера источника звука: чисто шумовой, квазигармонический и смешанный. Примеры шумового сегмента согласных в интервокальном положении приведены на рис. 6.2. На спектрограмме шум представлен в виде широкой размытой полосы.

Шумовой (смешанный) характер сигнала является отличительным признаком группы шумных согласных, выделяющим их из остальных согласных. В зависимости от того, к какому согласному шумной группы принадлежит сегмент шума, он имеет разную длительность (рис. 6.2). Для щелевых она равна 100—200 мсек., для смычных — 20—50 мсек., так как шум образуется только в момент размыкания, а для аффрикат длительность шума имеет некоторую промежуточную величину. Есть основания полагать, что длительность шумовой фазы используется при восприятии как признак, различающий между собой эти группы согласных (Деркач, 1959).

Внутри каждой из групп согласные отличаются друг от друга теми конфигурациями, которые принимает голосовой тракт при их произнесении, — в основном местом наибольшего сужения, т. е. положением шумового источника. Конфигурация голосового тракта определяет частотный состав спектра шумового сегмента каждого из согласных.

Спектр шумового сегмента может быть охарактеризован или нижней частотой области концентрации звуковой энергии, или, с точки зрения акустической теории речеобразования, полюсами и нулями передаточной функции речевого тракта. Используя аналоговый синтезатор, Хейнц и Стивенс (Heinz a. Stevens, 1961) для английского языка, а Мэртони (Mártony, 1962) — для шведского показали, что по частоте первого резонанса и первого антирезонанса могут хорошо различаться согласные *ʃ* и *s*, независимо от гласной. Частоты полюсов при этом имеют значения от 2500 до 8000 гц, для *ʃ* характерен самый низкий резонанс, для

s — самый высокий. Харрис (Harris, 1958) использовала в качестве характеристики щелевых согласных нижнюю границу полосы шума и показала, что по этому параметру хорошо различаются согласные *s* и *ʃ* при их изолированном предъявлении. Но как в одних, так и в других работах отмечается, что не все щелевые согласные, взятые вне контекста, хорошо различимы: *f* и *θ* распознаются только при их предъявлении в составе слогов. Из данных Н. И. Дукельского (1962) о восприятии изолированной шумовой части начальных щелевых русского языка следует, что *s*, *ш* и *х* распознаются хорошо, а *ʃ* — неудовлетворительно. Разборчивость *ʃ* увеличивается, если к шуму добавляется некоторая часть гласного.

Спектры шума смычных согласных оказываются весьма близкими к спектрам шума щелевых согласных с аналогичным местом образования (Hughes, a. Halle, 1956). Однако, как уже говорилось выше, связь спектральных характеристик шумового сегмента смычных согласных с их восприятием значительно более сложна, чем в случае щелевых.

Если щелевой согласный произносится с участием голоса, первый его сегмент представляет собой звук, спектр которого состоит из двух компонент: шума, имеющего те же характеристики, что и глухой вариант, и низкочастотных квазигармонических составляющих, создаваемых голосовыми связками. Эти составляющие представлены на спектрограмме в виде темной, довольно четкой полосы в области низких частот — так называемая полоса голоса (см. рис. 6.1).

Для русских смычных звонких согласных полоса голоса является первым по времени сегментом согласного. Предъявленный изолированно он воспринимается как неопределенное гудение (Дукельский, 1962). Очевидно, что в речевом потоке этот сегмент несет информацию только о том, что согласный является звонким смычным. Сведений о месте образования согласного в нем не содержится. Для получения этих сведений человек должен воспринять следующие по времени сегменты: шум взрыва и переход к гласному. Шум взрыва звонких смычных согласных значительно слабее шума взрыва глухих согласных и может совсем отсутствовать (сравни *т* и *д* на рис. 6.1). Переход выражается в виде изменения частот формант начального участка гласного (см. рис. 6.1).

К числу согласных, первый сегмент которых также практически не содержит шума, относятся звуки и группы сонантов: в русском языке это *м*, *н*, *л*, *р*, *ј*. *М* и *н* объединяются по способу артикуляции, который заключается в том, что оба они образуются при полном смыкании голосового тракта и при участии носового резонатора. Следствием этого является присутствие в звуковом сигнале, кроме полосы голоса, двух-трех формант, которые похожи на форманты гласных, однако слабее последних по интенсивности (рис. 6.3). Благодаря формантной структуре носовые

звуки по своим акустическим свойствам ближе к гласным, чем к согласным.

По акустической теории речеобразования спектр носовых звуков (в отличие от гласных) характеризуется не только полюсами передаточной функции голосового тракта, но и нулями (как правило, важен первый антирезонанс). Синтезируя назальные согласные с помощью аналогового синтезатора, Наката (Nakata, 1959) нашел, что английские *m*, *n* и *ŋ* наилучшим образом воспринимаются, если частота второй форманты находится в следующих областях: 800—1300 гц для *m*, 1500—2000 гц для *n* и 2000—2500 гц для *ŋ*. Положение форманты внутри указанных областей зависит от характера следующего гласного. Добавление перехода к стационарному участку назального согласного увеличивает надежность его распознавания.

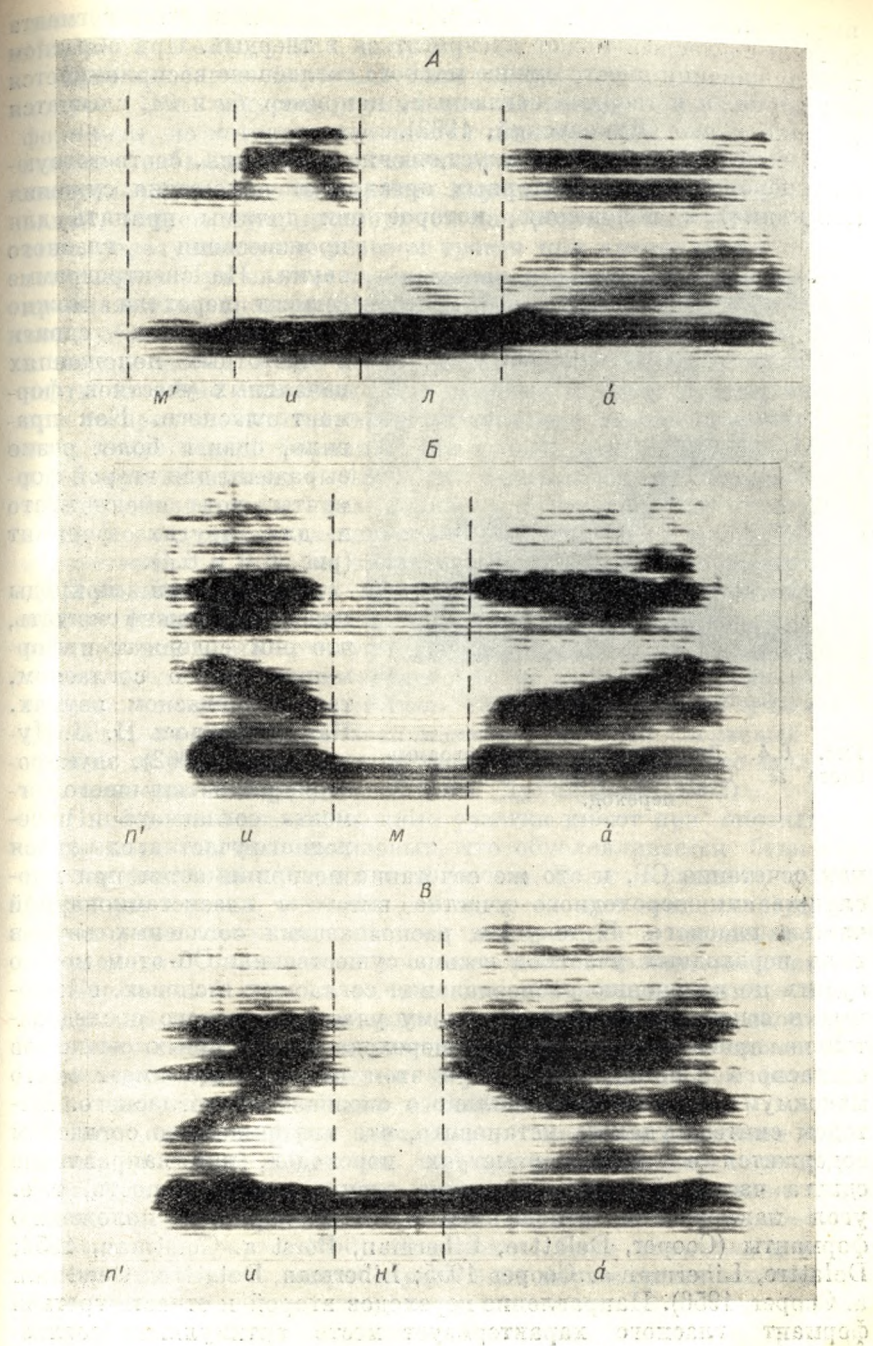
Существуют также данные о том, что спектральные характеристики сегмента, соответствующего смычке носовых естественных звуков, не содержат достаточно информации для различения этих звуков между собой (Malécot, 1956). Значительная часть полезной информации заключена в переходе от согласного к гласному.

Так же как и для носовых согласных, формантная структура первого сегмента, соответствующего фазе смычки, характерна для плавных *l* и *r*. Артикуляторная и акустическая картина этих звуков довольно сложна и мало изучена. Относительно согласного *l* можно только сказать, что для его опознавания недостаточно той информации, которая содержится в данном сегменте: предъявленный изолированно он воспринимается как гласный *y* (Halle, 1959).

Согласные русского языка могут различаться также по признаку палатализации. Конфигурация голосового тракта при артикуляции мягких согласных отличается от конфигурации его при артикуляции соответствующих твердых дополнительным подъемом средней части спинки языка. Признак мягкости может быть выражен в акустическом сигнале двумя способами. Во-первых, информация о мягкости содержится в спектральной характеристике целевого сегмента, как это имеет место, например, для согласных *с*, *х*, *ш* и *л*, т. е. для большинства целевых и плавных. Во многих случаях первая часть мягкого согласного может совпадать по спектральным характеристикам с соответствующим твердым согласным. Следовательно, шум в целевых мягких согласных фактически разделяется на два сегмента, из которых мягким является только конечный. Во-вторых, для слогов с мягкими согласными характерно присутствие особого сегмента — *и*-образного перехода, на основании которого человек, очевидно, извлекает

Рис. 6.3. Спектрограммы звуко сочетаний *милă* (А), *нимă* (Б) и *пинă* (В).

Пунктиром отмечены границы смычки сонорных согласных *м*, *н* и плавного *л*.



информацию о мягкости (рис. 6.4). При удалении этого сегмента мягкий согласный может превратиться в твердый. При обратном прослушивании вместо одного мягкого согласного воспринимается два звука: *и* и твердый согласный, например *ба* и *да*, слышатся как *аиб* и *айд* (Дукельский, 1962).

Рассмотрим подробнее акустические изменения, соответствующие переходу артикуляторных органов от положения сужения (смыкания) к положению, которое они должны принять для

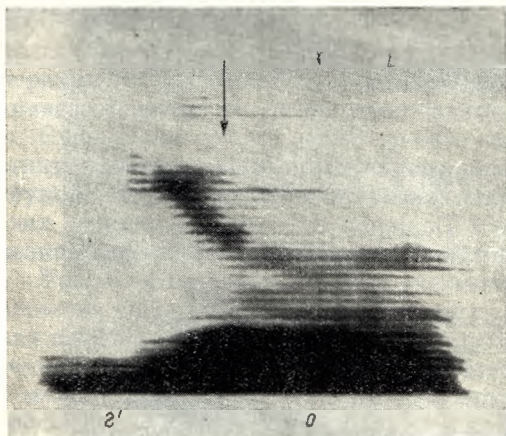


Рис. 6.4. Динамическая спектрограмма слога *гё*. Стрелкой указан *и*-образный переход.

произнесения гласного звука. На спектрограмме сегмент перехода можно наблюдать как сдвиги в частотных положениях начальных участков формант гласного. Как правило, сдвиги более резко выражены для второй форманты, хотя имеют место и для других формант (рис. 6.3 и 6.4).

Исходя из природы переходов, можно ожидать, что они содержат информацию как о согласном, так и о гласном звуках. Как отмечалось Н. И. Дукельским (1962), звук, состоящий из шумового сегмента согласного и переходного участка, слышится

как сочетание СГ, и это же сочетание воспринимается при прослушивании переходного участка вместе с квазистационарной частью гласного. В текущем распознавании согласных звуков роль переходных участков весьма существенна. Об этом можно судить по изменению распознавания согласных в слогах, в которых к выделенному стационарному участку гласного последовательно прибавляются отрезки перехода и остальных сегментов согласного (Ohman, 1962). При этом на переходе имеет место максимум увеличения правильного опознавания согласного. Методом синтеза удалось установить, что информация о согласном содержится в таких параметрах переходов, как направление сдвига частот формант, величина сдвига и его скорость, т. е. угол наклона по отношению к стационарному положению форманты (Cooper, Delattre, Liberman, Borst a. Gerstman, 1952; Delattre, Liberman a. Cooper 1955; Liberman, Delattre, Gerstman a. Cooper, 1956). Направление переходов второй и отчасти третьей формант гласного характеризует место артикуляции соглас-

ного (Cooper, Delattre, Liberman, Borst a. Gerstman 1952; Halle, Hughes a. Radley, 1957, и др.).

Для описания перехода было введено специальное понятие локуса. Под локусом понимается начальное значение частоты форманты, из которого начинается переход. Локус или непосредственно наблюдается на спектрограмме, или находится путем экстраполяции. С точки зрения акустической теории речеобразования, локусы интерпретируются как резонансные частоты полостей, образующихся в речевом тракте при артикуляции данного согласного. При слабом возбуждении резонаторов эти частоты не проявляются в спектре (Stevens a. House, 1956, 1961). Для согласных, образованных путем сужения голосового тракта в одном и том же месте, локус имеет постоянное значение в сочетаниях почти со всеми гласными. Это правило хорошо подтверждается при восприятии синтезированных губных и переднеязычных гласных: локус второй форманты для губных имеет значение 720 гц, для переднеязычных — 1800 гц. Однако локус заднеязычных звуков обнаруживает два варианта в зависимости от огубленности последующего гласного: в сочетаниях с *i, e, ε, a* — 2400 гц, в сочетаниях с *u, o* он значительно ниже и более расплывчат.

Добавление сегмента перехода к предшествующим ему сегментам фазы смычки (щели) существенно улучшает возможность опознавания места артикуляции согласного. Если для прослушивания предъявляется сочетание СГ, из которого удален первый сегмент согласного и оставлен только переходный участок и гласный, то слушатель воспринимает смычный согласный независимо от того, был ли этот переход ранее связан со щелевым, смычным или сонорным согласным (Дукельский, 1962).

Наибольший вес переходные участки имеют при опознавании начальных смычных согласных, что обуславливается бедностью акустического выражения этих звуков. Если в сигнале отсутствует шум взрыва, что часто наблюдается при звонких взрывных и сонорных согласных, то вся информация о признаке места образования оказывается сосредоточенной только на переходе.

Переходный сегмент начального смычного согласного может содержать также информацию о его глухости или звонкости: при отсутствии звонкого участка смычки звонкость может быть опознана по длительности и форме перехода первой форманты (Liberman, Harris, Hoffman a. Griffith, 1957).

Еще одно качество переходов — скорость движения в частотно-временной плоскости, или длительность перехода второй форманты, — может служить признаком для различения групп смычных, полугласных и гласных звуков (Liberman, Delattre, Gerstman a. Cooper, 1956). Эти данные были получены для синтетических звуков; вопрос о роли этих параметров в естественной речи нигде не обсуждался, за исключением спорной во многих отношениях работы Лехисте и Петерсона (Lehiste a. Peterson,

1961). Если отмеченное свойство переходов используется и при восприятии естественной речи, то этим, очевидно, можно объяснить причину плохого распознавания полугласных, предъявляемых изолированно от контекста.

Как уже упоминалось, наряду с признаками согласного переходный участок содержит также сведения о гласном, однако они весьма неопределенны, т. е. характерный тембр гласного отражен мало. В настоящее время нет четких данных о том, какой сигнал можно принять в качестве границы окончания переходного сегмента. В случаях синтезированных сигналов границей считается начало стационарной части гласного, однако на спектрограммах реальных звуков, как правило, трудно определить, где начинается стационарный участок гласного. Такое положение в значительной мере связано с неопределенностью в отношении признаков восприятия гласных звуков. Основным качеством гласных, позволяющим различать их друг от друга, является тембр, который создается типичной формантной структурой спектра. Согласно акустической теории речеобразования, формантная частота определяется как характерная частота резонансной полости голосового тракта и соответствует термину «полус» в передаточной функции. Спектр каждого гласного достаточно полно описывается частотами трех или четырех первых формант, соотношением их интенсивностей и частотой основного тона, которые определенным образом связаны с акустическими характеристиками артикуляторного тракта (Stevens a. House, 1955; Fant, 1960). Обычно для характеристики гласных достаточно указать значения частот двух первых формант (Potter a. Steinberg, 1950; Peterson, 1952). Гласные, синтезированные на основании этих значений, хорошо определяются при восприятии (Miller, 1953).

Основная трудность, касающаяся механизмов распознавания гласных, состоит в том, что в естественной речи частоты формант являются непрерывными функциями времени, так как конфигурация речевого тракта непрерывно меняется. Поскольку вопрос о способах оценки непрерывно изменяющихся сигналов до сих пор совершенно не ясен, мы не будем больше останавливаться на гласных и перейдем к дальнейшему описанию состава сочетания СГС. Следующий сегмент соответствует переходу от гласного к согласному, в спектрограмме акустического сигнала он выражается быстрыми частотными сдвигами формант гласного. Характер начальных и конечных переходов несколько различен (Halle, Hughes a. Radley, 1957), и локусы формант не совпадают (Malécot, 1958). Исследования восприятия конечных переходов (Дукельский, 1962), показали, что основная информация, которую они содержат, касается места образования согласного.

Принимая во внимание, что для восприятия речевого элемента большое значение имеет временная последовательность его составляющих, естественно ожидать, что роль начальных и конеч-

ных переходов в текущем распознавании будет различна. Если начальные переходы в большинстве случаев делают лишь более определенным восприятие признака, содержащегося в предыдущем сегменте, то конечные переходы несут в себе первые сведения о согласном, которые могут быть дополнены или исправлены на основании анализа следующих сегментов (исключение представляет случай импловзивных согласных).

Недостаток сведений о конечных переходах вынудил нас поставить некоторые эксперименты для исследования значения

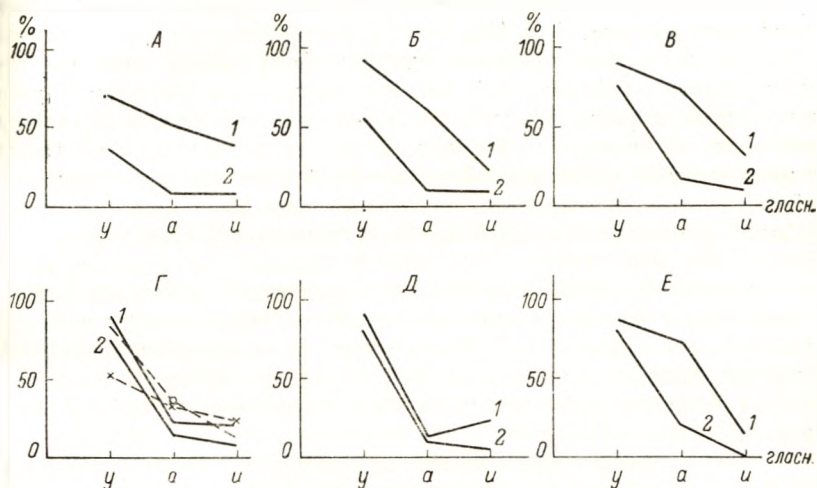


Рис. 6.5. Результаты распознавания места артикуляции смычных согласных по конечным переходам гласных.

А, Б, В, Г, Д, Е — данные разных дикторов: 1 — для сочетаний, в которые входили переднеязычные согласные; 2 — для сочетаний с губными согласными. По оси ординат — процент распознавания согласных как переднеязычных.

конечных переходных участков в текущем распознавании согласных.

В опытах слушателям были предъявлены звуки, полученные из слогов ГС путем удаления участка, соответствующего смычке согласного. Удаление смычного участка осуществлялось с помощью электронного ключа (Вайтулевич, Кожевников, Лебедев и Сороко, 1962), управляемого сигналами с датчиков, регистрирующих губную или переднеязычную смычки. Ключ, открытый при произнесении гласного, запирался импульсом, соответствующим началу смычки согласного, и оставался закрытым до окончания произнесения звуко сочетания. По оставшемуся гласному и переходу слушатели должны были определить, к какой фонетической группе — переднеязычных или губных — принадлежит конечный согласный.

Результаты опытов (рис. 6.5) показали, что решение о месте образования согласного, принимаемое человеком, сильно зависит от характера предшествующего гласного: после низкоформантных гласных переходы опознаются преимущественно как переднеязычные согласные, а после высокоформантных — как губные. Картина такова, как если бы оценка признака места образования определялась направлением перехода второй форманты. Аналогичные данные об идентификации согласных по конечным переходам были получены Н. И. Дукельским (1962). Построенные на основании этих данных кривые изображены пунктиром на рис. 6.5 (Г).

После сегмента перехода следует фаза щели или смычки. Акустические свойства этой части сигнала для сонантов и звонких согласных аналогичны тем, которые проявляются в начальных сегментах. В случае же глухих, щелевых и смычных согласных на этом участке наблюдается по крайней мере два сегмента: звонкий и глухой. Это явление известно в фонетике как озвончение глухих согласных в интервокальном положении (Бондарко, Вербицкая, Зиндер, 1960). Величина озвонченного участка может составлять 60% общей длительности согласного, и на этом участке глухие звуки ничем не отличаются от соответствующих звонких. Длительность глухого интервала имеет фонематическое значение. Если она оказывается меньше некоторой критической величины, то глухой согласный вообще не воспринимается (Bastian, Delattre a. Liberman, 1959).

Остается рассмотреть свойства сегмента, соответствующего взрывному шуму смычных согласных, и его роль в формировании образа этих звуков в восприятии.

Для текущего распознавания большое значение имеет тот факт, что шум в акустическом сигнале следует после перехода и участка смычки, так что на основании оценки спектра шума должно быть принято окончательное решение о месте образования согласного. По данным Малеко (Malécot, 1958), действительно, различие конечных смычных по признаку места образования на основании спектра шума имеет весьма определенный характер. Необходимо отметить также, что в число существенных параметров данного сегмента должна входить его длительность, так как большой отрезок шума после смычки имеет место при наличии сочетания смычного согласного с щелевым и в случае аффрикат.

2. Фонема как первичная единица решений

Фактический материал, приведенный в предыдущем разделе, убедительно показывает, что человек, принимая решение о фонеме, использует информацию, распределенную на участке речевого потока, соответствующем по меньшей мере слогу. Это является достаточно целесообразным, так как таким образом

достигается полное использование информации о фонеме, содержащейся в акустических характеристиках речевого сигнала. Следует ли из этого, что единицей решения является слог, т. е. что первичные решения, которые принимает человек, — это решения о слогах? Иными словами, следует ли допустить, что в памяти человека имеются образы всех возможных слогов и решение о слоге эквивалентно выбору одного из множества слоговых образов, а запоминание слога соответствует запоминанию символа, обозначающего выбранный образ?

Ряд соображений говорит, что такой путь распознавания является мало рациональным. Во-первых, слогов оказывается слишком много. Согласно Сивертсен (Sivertsen, 1961), набор слогов в английской речи состоит из 4400 или, с учетом просодических характеристик, из 30 800 единиц. Во-вторых, количество информации, требующееся для достаточно подробного описания акустической картины слога, слишком велико. Следовательно послоговое распознавание требует большой памяти, как долговременной, так и оперативной. В-третьих, исследования артикуляции показывают, что инвариантными в отношении окружения могут быть только слоги типа СГ, ССГ, СССГ, так как именно они являются артикуляторными комплексами, произносимыми последовательно.

Естественно, что именно такие слоги целесообразно было бы использовать в качестве элементарных отрезков речевого потока, на основании которых принимаются первичные решения. Однако такие слоги не соответствуют морфемам (слог может быть образован на стыке двух слов, например, *урок окончен*), и, следовательно символы, обозначающие эти слоги, не могут быть прямо использованы в качестве смысловоразличительных единиц на следующих этапах распознавания. Для того чтобы последовательность слоговых символов стала доступной для ее обработки на уровне распознавания и анализа предложения, необходимо было бы предварительно превратить каждый слоговой символ в последовательность соответствующих ему фонемных символов.

Таким образом, система распознавания, использующая слоги в качестве первичных единиц решений, кажется слишком громоздкой и неэкономичной. Поэтому интересно рассмотреть другие возможные пути решения проблемы, при которых единицы решений соответствовали бы фонемам, но участок речевого потока, на основании которого принимались бы фонемные решения, соответствовал бы слогу. Такие пути являются теоретически возможными, если отказаться от предвзятой точки зрения о том, что процесс распознавания одной фонемы не может осуществляться параллельно с процессом распознавания другой фонемы.

Допустим, что в памяти распознающего устройства имеется набор образов простейших акустических явлений (сегментов). Это соответствует тому, что в нервной системе образованы спе-

циальные схемы (блоки) для обнаружения таких явлений, как например шум с максимумом энергии в определенном участке спектра, толчок (взрыв), пауза, формантный переход с определенными свойствами и т. д. При восприятии речевого сигнала указанные схемы вырабатывают символы, обозначающие соответствующие акустические явления как таковые. Эти символы, в свою очередь, поступают на вход дешифратора фонем, на выходе которого возникает фонемный символ при условии, если было образовано определенное логическое высказывание из символов, соответствующих акустическим явлениям. Иначе говоря, фонема является логической функцией от ряда акустических явлений. При этом допустимо, что одной и той же фонеме соответствует несколько различных логических функций, и что одно и то же акустическое явление может входить в качестве аргумента в логические функции, соответствующие различным фонемам. Например, экспериментальные фонетические данные заставляют предполагать, что *и*-образный переход в слове *тя* должен входить как аргумент в логические функции, соответствующие как фонеме *т'*, так и фонеме *а* (Дукельский, 1962).

Итак, считая доказанным, что человек, принимая первичные решения, основывается на отрезках речевого потока, соответствующих слогу, можно сформулировать две гипотезы относительно единиц первичных решений. Одна из них состоит в том, что первичные решения соответствуют слогам. Другая гипотеза состоит в том, что первичные решения соответствуют фонемам, причем фонема является логической функцией от последовательности акустических явлений (сегментов) на протяжении слога.

Пытаясь найти эксперимент, результаты которого позволили бы предпочесть одну из этих гипотез, мы воспользовались следующими соображениями. Если слог является единицей решения, это должно означать, что и гласный, и согласный этого слога опознаются одновременно. Тогда, предъявляя испытуемому открытые слоги типа *СГ* и прося его в одном случае записывать гласный, а в другом согласный, мы вправе ожидать, что задержки записи от начала слога в обоих случаях будут одними и теми же.

Если единица решения соответствует фонеме, следует ожидать, что задержка записи согласного может быть меньше задержки записи гласного (конечный участок гласного уже не должен содержать информации о согласном).

Эксперимент проводился следующим образом: диктор, с соответствующими датчиками для регистрации артикуляторных параметров, читал таблицы слогов; четверо испытуемых, находящихся рядом с диктором, но не видящих его, записывали согласный или гласный карандашом на металлических пластинах. Контакт карандаша с пластиной замыкал электрическую цепь и приводил к появлению сигнала, который записывался на ленте чернило-

пишущего осциллографа параллельно с артикуляторными сигналами от диктора.

В опытах с записью согласных использовалась таблица слогов, начинающихся с $t, t', \partial, \partial', л, л', н, н', з, з', с, с'$ и кончающихся на гласные y, o, a . Испытуемые пользовались русским алфавитом и записывали твердый и мягкий согласный одной и той же буквой. Каждый из 4 испытуемых записал в общей сложности по 720 согласных. В опытах с записью гласных использовалась таблица слогов, начинающихся с согласных $t, t', \partial, \partial', л, л', с, с'$ и кончающихся на гласные y, o, a, u . Пользуясь русским алфавитом, испытуемые записывали гласные, следующие за твердыми согласными, буквами y, o, a . Гласные, следующие за мягкими согласными, записывались буквами $ю, ё, я, и$. Каждый из испытуемых записал в общей сложности по 840 гласных.

Задержки записи гласных и согласных были подсчитаны относительно одного и того же явления в артикуляции слога, а именно момента размыкания смычки (цели) согласного (размыкание контакта $k.3$ на искусственном нёбе). Предварительные подсчеты показали, что задержки записи твердых и мягких согласных практически не отличаются друг от друга, разница между задержками c и $з$, а также $л$ и $н$ не является статистически достоверной. Это позволило объединить все согласные в четыре группы (t, t'), (∂, ∂'), ($л, л', н, н'$), ($с, с', з, з'$) и подсчитать для этих групп средние значения величины задержки записи согласного (рис. 6.6, 1).

Обработка результатов опытов с записью гласных не позволила обнаружить какой-либо зависимости величины задержки от качества гласного, однако выяснилось, что величина задержки записи гласного отчетливо зависит от качества предшествующего согласного. Поэтому все записи гласных были разбиты на восемь групп, соответственно характеру предшествующего согласного. Средние значения величин задержек записи гласных, полученные после такой разбивки материала, приведены на рис. 6.6, 2 и 3. Из рис. 6.6 следует, что человек распознает согласный раньше, чем гласный. Очевидно, что это противоречит гипотезе о слоговом

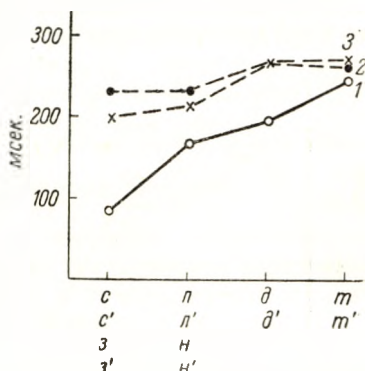


Рис. 6.6. Задержки буквенной записи согласных в зависимости от их качества (1) и задержки буквенной записи гласных в зависимости от качества предшествующих согласных (2 — твердый согласный, 3 — мягкий согласный).

По оси абсцисс — согласные тех групп слогов, для которых вычислялись средние величины задержки, по оси ординат — величина задержки (в мсек.).

распознавании и позволяет предполагать, что первичные решения, принимаемые человеком, — это решения о фонемах.

Очень интересным является факт зависимости задержки записи гласного от качества предшествующего согласного. Нет оснований ожидать каких-либо различий в физической картине гласных, следующих за разными переднеязычными согласными, которые могли бы быть ответственны за эту зависимость. Приходится предположить, что причину обнаруженной зависимости следует искать не непосредственно в физической картине гласных, но в организации процедуры принятия фонемных решений в мозгу человека.

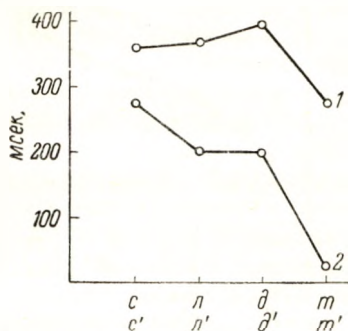


Рис. 6.7. Задержки буквенной записи согласных (в мсек.) относительно начала исходного согласного (1) и длительности исходных согласных (2).

Рис. 6.6 наводит на мысль, что решение о гласном не может быть принято до того, как будет принято решение о согласном. Если решение о согласном по каким-то причинам задерживается, то решение о гласном также задерживается. Однако для того, чтобы рассматривать этот вопрос, нужно сначала выяснить, чем определяется разница в задержках для различных согласных.

На рис. 6.6 приведены задержки записи согласного относительно момента размыкания контакта $k.3$ у диктора, что в акустическом сигнале должно соответствовать началу перехода от согласного к гласному и появлению фонации при m, m' .

Возникает предположение, что полученная при таком подсчете зависимость задержки от качества согласного обусловлена тем, что на самом деле постоянна задержка от начала согласного до буквенной записи. Так как при регистрации артикуляторных сигналов диктора записывались также сигналы с микрофона и ларингофона, не представляло труда вычислить задержки записи согласного от начала согласного (рис. 6.7, 1). Оказалось, что и в этом случае задержки зависят от качества согласного. На рис. 6.7 приведена также кривая (2), показывающая среднюю длительность согласных. Можно видеть, что согласные одинаковой длительности ($л$ и $д$) характеризуются разными задержками, наименьшей оказывается задержка для $т$, имеющего наименьшую длительность.

Полученные факты относительно задержки записи согласного могут быть в самом общем виде описаны следующим образом. Во-первых, запись производится всегда уже после того, как согласный окончится, т. е. человек сможет услышать не только стационарную часть согласного, но и переход к следующему гласному. Во-вторых, задержка записи относительно конца согласного

тем больше, чем меньшую информацию о согласном можно получить на основании его стационарного участка.

Как было показано в предыдущем разделе главы, стационарный участок щелевых согласных содержит почти всю информацию, необходимую для выбора фонемы, стационарный участок сонорных согласных содержит значительно меньше информации, стационарный участок звонких смычных согласных почти не содержит информации, а во время смычки глухих смычных согласных в начале слова вообще не образуется звука, т. е. информация полностью отсутствует. Однако ссылки на информацию еще не дают возможности понять механизмы явления. Некоторые представления о возможных механизмах задержки фонемного распознавания согласного будут даны в следующем разделе.

3. Процедура принятия фонемных решений

Приняв, что единицы решений у человека соответствуют фонемам, естественно пойти дальше и постараться выяснить, каким образом осуществляется переход к фонеме от последовательности акустических явлений (сегментов).

До сих пор мы довольно неопределенно говорили о том, что фонема должна быть логической функцией от ряда акустических явлений. При попытке представить себе техническое устройство, которое могло бы обеспечить вычисление подходящих логических функций, сразу же возникают определенные трудности. Во-первых, акустические явления поступают последовательно во времени; следовательно, должно быть обеспечено запоминание предыдущих акустических явлений до поступления нового явления. Во-вторых, совершенно очевидно, что последовательность поступления акустических явлений оказывается весьма важной. Это доказывается резкими нарушениями распознавания фонем при так называемом инверсальном прослушивании (см.: Дукельский, 1962). Например, при воспроизведении в обратном направлении магнитной записи сочетания *атя* оно воспринимается как последовательность, состоящая из четырех фонем — *айфьта* или *айсьта*.

Логично рассматривать человека как конечный автомат. Конечным автоматом называется система с m -входными символами (состояниями входов) $a_1, \dots, a_m$, n -внутренними состояниями, $q_1, \dots, q_n$ и k -выходными символами $y_1, \dots, y_k$. Изменение состояний системы происходит в дискретные моменты времени $1, 2, \dots, t, \dots$; при этом внутреннее состояние и выход в момент t зависят от входного символа в момент $t-1$ и внутреннего состояния в момент $t-1$, т. е.

$$q(t) = \delta[a(t-1), q(t-1)] \text{ и } y(t) = \lambda[a(t-1), q(t-1)],$$

где δ — функция переходов, λ — функция выходов. Возможны другие варианты временной зависимости: например, выход мо-

жет зависеть от входного символа в тот же момент времени, и т. д. Так как новый входной символ может переводить автомат в новое внутреннее состояние, то тем самым обеспечивается запоминание этого символа. Так как выходной символ определяется не только поступившим в данный момент входным символом, но и существующим внутренним состоянием, то тем самым он может зависеть от последовательности поступавших ранее входных символов (предыдущей истории автомата).

Если принять, что входными символами автомата являются акустические явления, а выходными символами являются фонемы, то задача описания человека как распознающего фонемы автомата сводится к определению набора возможных внутренних состояний $q_1, \dots, q_n$, определению функций переходов δ и функций выходов λ . К сожалению, изменения внутреннего состояния системы, как правило, остаются скрытыми от непосредственного наблюдения.

Уже несколько лет назад у одного из нас возникла идея о том, нельзя ли найти такие внешние реакции человека, которые в какой-то мере отражали бы не образование выходных фонемных символов, но изменения во внутреннем состоянии системы. Казалось вероятным, что если такие реакции принципиально могут быть обнаружены, то их следует искать в речевом аппарате. Во-первых, в теоретических работах, касающихся возможных моделей восприятия (Маккей, 1956), высказывались соображения о том, что двигательные отделы нервной системы должны принимать участие в процессе обработки информации о внешнем сигнале. Во-вторых, имелись отдельные факты, говорящие не только о том, что в процессе восприятия речи наблюдается активность речевых мышц (Соколов, 1960), но и что эта активность почему-то важна для восприятия. Так, Лурия (Лурия, 1947) показал, что понимание речи афатиком существенно ухудшается, если тот во время слушания речи зажимает язык между зубами. Наконец, уже упоминавшаяся выше моторная теория восприятия речи (Liberman, 1957; Liberman, Cooper, Harris a. MacNeilage, 1962) предполагает, что процесс фонемного распознавания включает в качестве промежуточного звена построение артикуляторного образа слышимого элемента речи.

Если процесс восприятия речи сопровождается скрытой активностью речеобразующих органов, то, очевидно, не так уж трудно найти пути для того, чтобы сделать эту активность явной. Наиболее простой путь состоит в том, чтобы попросить испытуемого перейти от внутреннего слежения за воспринимаемым сообщением к его повторению вслух.

Явление текущей имитации слышимой речи было впервые описано Черри (Cherry, 1953), затем это явление использовали в своих работах и другие исследователи (Saslow, 1958; Kupfmüller, 1961).

После того как нам удалось экспериментально воспроизвести эффект текущей имитации (Чистович, Алякринский, Абульян, 1960), был проведен ряд работ (Чистович, 1960б, 1961, 1962; Чистович и Клаас, 1962), направленных на изучение характеристик этих реакций и выяснение их сходства или различия с реакциями, в основе которых несомненно должен лежать выбор определенной фонемы (запись буквы, название фонемы, соответствующей услышанному звуку). Было обнаружено, что временные задержки (скрытые периоды) имитации существенно короче задержек буквенной записи; выяснилось также, что уменьшение полезной информации в сигнале не изменяет задержек имитации, существенно увеличивая задержки буквенной записи и названия фонем. Наконец, что особенно интересно, было обнаружено, что человек начинает имитировать слышимый согласный (*б, г, д*) в то время, когда он еще принципиально не может знать, какая именно фонема произносится диктором (Чистович, 1962).

Все эти факты заставили считать, что текущая имитация отражает такие явления в нервной системе, которые происходят до того, как фонемный символ может быть выработан, и что, вероятно, эти явления можно соотнести с изменениями внутреннего состояния автомата, вырабатывающего фонемные символы.

Если состояние речевого тракта при имитации отражает внутреннее состояние распознающей системы, чрезвычайно интересно исследовать, как будет изменяться артикуляторное состояние на протяжении имитации согласного. Как уже говорилось в предыдущем разделе данной главы, согласные особенно интересны тем, что полезная фонемная информация содержится в последовательности различных акустических явлений (сегментов).

Первая серия экспериментов была проведена на материале звуко сочетаний типа ГСГ, в которых использовались согласные *п, т, б, д, м, н, ф, с, в, з* и гласный *а*. Диктор, с соответствующими датчиками для регистрации артикуляторных параметров, читал таблицу (200 звуко сочетаний) перед микрофоном. Находящийся в другой комнате испытуемый, также с датчиками, слышал звуко сочетания через телефоны и текущим образом их имитировал. Артикуляторные сигналы как от диктора, так и от испытуемого параллельно записывались на ленте чернилопишущего осциллографа. В опытах участвовал один диктор и, поочередно, два испытуемых.

Пример записи сочетания *ана* приведен на рис. 6.8. Можно видеть, что имитация начинается значительно раньше, чем окончится согласный диктора, т. е. заведомо до того, как вся информация о согласном будет получена. На рис. 6.9 приведены величины задержек повторения согласных. Можно видеть, что задержки весьма малы и практически не зависят от качества согласного.

Во всех без исключения случаях испытуемые начинали имитировать согласный раньше окончания исходного согласного. Так как испытуемый образовывал смычку (щель), в ряде случаев еще не имея достаточных сведений о том, что эта смычка должна быть образована, естественно было ожидать большого количества ошибок. Действительно, оказалось, что только в 77% случаев повторение начиналось с правильной смычки. В остальных случаях или дело начиналось с неправильной смычки, к которой

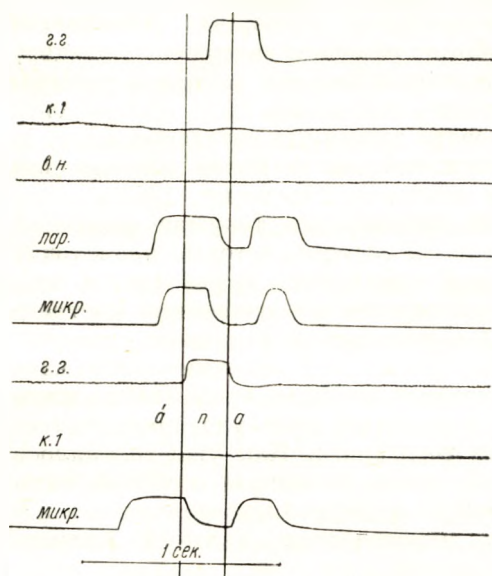


Рис. 6.8. Осциллограмма сочетания *апа* при быстром повторении.

Три нижние кривые соответствуют артикуляторным параметрам диктора, верхние — испытуемого.

затем присоединялась правильная, или сразу же возникали две смычки, затем неправильная смычка исчезала. Например, при повторении *б* в сочетании *аба* испытуемый сначала образовывал и язычную и губную смычку, затем язычная смычка прекращалась (рис. 6.10). Естественно, что в ряде случаев наблюдалась только одна неправильная смычка, которая так и не исправлялась, однако эти случаи мало интересны, так как они могли быть обусловлены ошибками слухового анализа. На рис. 6.11, *м*, приведено количество случаев, когда неправильная смычка (или одновременно две смычки) наблюдалась в начале повторения (одинарная штриховка) и когда это сохранялось до конца повторения согласного (двойная штриховка). Из диаграммы следует, что в большинстве случаев происходит исправление. Принципиально то же самое происходило и с другими признаками артикуляции согласного: повторение звонкого согласного могло начаться с прекращения голоса, однако затем голос появлялся, повторение щелевого согласного могло начаться с полной смычки, однако затем возникла щель (рис. 6.11, *зв* и *щ*).

В результате такого текущего исправления конечный участок артикуляторной картины повторения согласного значительно больше соответствовал исходному согласному, чем начальный участок. Для одного испытуемого полное совпадение характеристики конечного участка повторения с характеристикой, требуе-

мой исходным согласным, наблюдалось в 83% случаев, для другого — в 75% случаев. Эти величины меньше тех, которые следовало бы ожидать для обычным образом определенной разборчивости согласных, однако необходимо учесть некоторые погрешности методики регистрации, приводящие к ошибкам уже при оценке данных экспериментатором. Например, если возникающий при повторении *в* высокочастотный шум был недостаточно сильным, мы не могли его обнаружить и считали, что испытуемый произнес *б*. Отсутствие сигнала *в.н.* при повторении *м* заставляло нас считать этот звук звуком *б*, хотя отсутствие сигнала могло быть обусловлено недостаточной чувствительностью датчика (в опытах использовался рупорный датчик, от которого впоследствии отказались). Итак, полученные данные показывают, что текущая имитация согласного начинается с создания в значительной мере случайного артикуляторного состояния; по мере поступления звуковой информации это состояние изменяется, так что в конце концов начинает соответствовать исходному согласному.

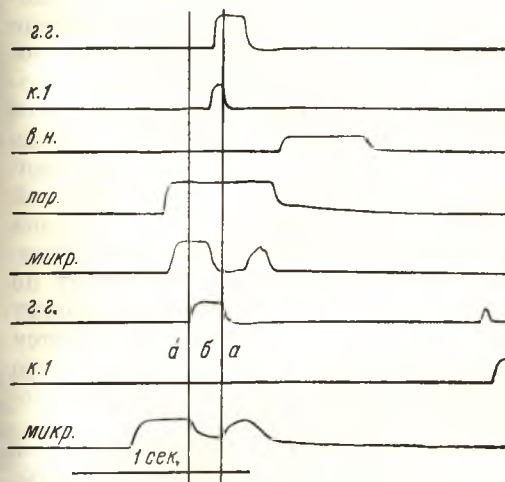


Рис. 6.10. Осциллограмма сочетаний *аба* при быстром повторении.

Три нижние кривые соответствуют артикуляторным параметрам диктора, верхние — испытуемого.

если он есть) несет двоякую информацию: с одной стороны, о месте образования и глухости или звонкости согласного, и, с другой — о том, что согласный кончен, начинается гласный.

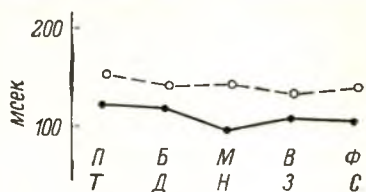


Рис. 6.9. Задержки образования смычки (щели) согласного (в мсек.) у двух испытуемых от начала смычки (щели) согласного у диктора.

Точки соответствуют медианам распределений задержек для каждой пары согласных.

Очень интересен вопрос, используются ли в процессе изменения состояния данные об уже достигнутом состоянии артикуляции. Для его выяснения полезно рассмотреть задержки окончания повторения согласного (рис. 6.12) от конца исходного согласного. Как известно, окончание исходного согласного (переход и взрыв,

Из рис. 6.12, А следует, что если согласный с самого начала повторялся правильно (верхняя гистограмма), то испытуемый его заканчивал и переходил к гласному с задержками всего около 100 мсек. Если же согласный был начат неправильно (нижняя гистограмма), то переход к гласному задерживался. На рис. 6.12, Б неправильно начатые согласные разделены на три более мелкие категории: 1) исправленные до окончания исходного согласного (верхняя гистограмма), 2) исправленные после окончания исходного согласного (средняя гистограмма), 3) вообще не исправленные согласные (нижняя гистограмма). Можно видеть, что большие задержки окончания согласного приходится на вторую группу. Эти данные несомненно показывают, что переход к новому состоянию артикуляции зависит как от внешнего акустического сигнала, так и от имевшегося в данный момент артикуляторного состояния.

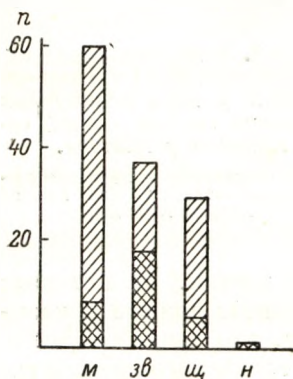


Рис. 6.11. Количество ошибок имитации согласных, заключающихся в неправильном воспроизведении места образования согласного (м), звонкости или глухости (зв), щелевого или смычного характера согласного (щ), назальности или неназальности (н).

Объяснения в тексте.

Следующая группа экспериментов была проведена на материале изолированных слогов типа СГ. Начальные опыты были проведены с теми же согласными, которые использовались в сочетаниях ГСГ. Было обнаружено, что повторение б, д, м, н в слогах СГ очень часто начинается с образования одновременно как губной, так и язычной смычки. Затем неправильная смычка прекращается, и конечный участок имитации согласного уже соответствует по артикуляторному состоянию исходному согласному. Это явление представляется

достаточно понятным, так как начальный участок согласного или вообще не содержит информации о месте образования согласного (б, д) или этой информации очень мало (м, н).

Так как аналогичная картина уже наблюдалась нами в предыдущих опытах, было решено этим эффектом больше не заниматься и обратить основное внимание на воспроизведение при имитации таких признаков согласных, как твердость или мягкость, и смычный или щелевой характер. Для эксперимента были выбраны переднеязычные согласные т, т', д, д', с, с', з, з', л, л', н, н', в сочетании с гласными у, о, а. В опытах участвовали те же двое испытуемых, каждый повторил по 360 слогов.

Результаты показали, что первым проявлением имитации согласного является образование контакта языка с нёбом. Задержки начала образования контактов относительно начала

исходного согласного оказались значительно большими, чем в опытах с сочетаниями ГСГ. У одного испытуемого они составили в среднем 160 мсек., у другого — 200 мсек. Было обнаружено также, что величина задержки зависит от качества согласного. У обоих испытуемых задержка росла в следующем порядке: (m, m') , (c, c') , (l, l') , (n, n') , $(з, з')$, (∂, ∂') . Проще всего предположить, что причина этой зависимости кроется в разной интенсивности сигнала.

Изменение артикуляторного состояния на протяжении повторения согласного отчетливо проявлялось в изменении конфигу-

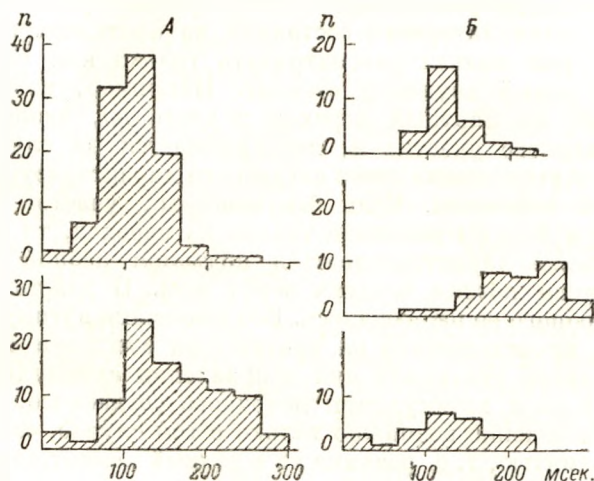


Рис. 6. 12. Гистограммы задержек размыкания смычки (щели) согласного у испытуемого относительно размыкания смычки (щели) согласного у диктора.

Объяснения в тексте.

рации контактов языка с нёбом. Так, имитация $з, з'$ в 40 случаях из 120 начиналась с образования контакта в точке $к.(1+2)$, который нормально характеризует только смычные согласные. В 31 случае из 40 этот контакт в дальнейшем разомкнулся, появился поток воздуха через рот и шум в полосе выше 4 кгц. Так как в случае $с, с'$ такой эффект почти не наблюдался, естественно считать, что он обусловлен весьма малой интенсивностью шума на начальном участке $з, з'$ и действительным сходством этого участка сигнала с начальным участком звонкого смычного согласного.

Другое интересное явление состояло в том, что 36 из 180 повторений мягких согласных $\partial', л', н'$ были начаты как твердые согласные (отсутствовал контакт в точке $к.11$). В 25 случаях

из этих 36 произошло изменение артикуляторного состояния к концу имитации согласного — контакт в точке *к.11* был образован. Полученный факт вполне согласуется с фонетическими данными о том, что начальный участок мягкого согласного в русском языке воспринимается как твердый согласный: основная информация о мягкости заключена в *и*-образном переходе от согласного к гласному (Дукельский, 1962).

Итак, артикуляторное состояние, создаваемое в начале имитации согласного, вполне определенно отражает свойства начального участка исходного согласного и, следовательно, может рассматриваться как реакция на этот начальный участок. Однако изменение артикуляторного состояния на протяжении имитации согласного уже нельзя рассматривать только как реакцию на некоторый новый звуковой сегмент. Например, известно, что сам по себе формантный переход в слове *на*, изолированный от стационарного участка *н*, воспринимается на слух как согласный *д*. Артикуляция этого согласного характеризуется подъемом нёбной занавески. Если бы конечный участок имитации согласного *н* был бы реакцией только на переход, то он должен был бы также характеризоваться подъемом нёбной занавески (прекращением потока воздуха через нос). В действительности ничего подобного не наблюдалось. Все изменения артикуляторного состояния, происходившие на протяжении имитации согласного, имели характер уточнения или исправления артикуляции, приближавших ее к артикуляции исходного согласного. Очевидно, что такое поведение возможно только в том случае, если система обладает памятью, т. е. реакция ее в данный момент определяется не только действующим сейчас сигналом, но и предыдущими сигналами.

Очевидно, что информация о предыдущем сигнале (сегменте) содержится в артикуляторном состоянии, вызванном этим предыдущим сегментом и существующем в тот момент времени, когда воспринимается следующий сегмент. Поэтому проще всего предположить, что изменение в артикуляторном состоянии определяется не только внешним сигналом (сегментом воспринимаемого звука), но и тем артикуляторным состоянием, которое к этому времени достигнуто. Обозначив начальное состояние артикуляции при имитации согласного как q_1 , а следующее состояние как q_2 , мы можем записать, что $q_1 = f(a_1)$, $q_2 = f(a_2, q_1)$, где a_1, a_2 — последовательные сегменты слышимого согласного.

Роль предыдущего состояния артикуляции особенно отчетливо выявилась при исследовании факторов, определяющих момент перехода от согласного к гласному. Очевидно, что внешним сигналом, вызывающим переход от согласного к гласному, должно быть окончание согласного диктором. В таком случае естественно ожидать, во-первых, что временные задержки окончания согласного должны быть одними и теми же при повторении

разных согласных. Во-вторых, следует ожидать, что должна быть высокая корреляция между длительностью согласного у диктора (длительностью смычки) и интервалом времени от начала согласного у диктора до размыкания смычки у испытуемого; зависимость между этими двумя величинами должна выражаться прямой линией с наклоном в 45° (насколько вырастет длительность исходного согласного, настолько, в среднем, должна возрасти и задержка окончания имитации согласного от начала исходного согласного).

Результаты обработки материала показали, что первое предположение не оправдывается. Как следует из рис. 6.13, задержки окончания согласного сильно варьируют в зависимости от качества согласного. Не оправдалось также и второе предположение (рис. 6.14). Можно видеть, что только при больших длительностях исходного согласного наклон кривой близок к 45° . Тем не менее зависимость отчетливо существует, коэффициент корреляции существенен (0.51 и 0.71).

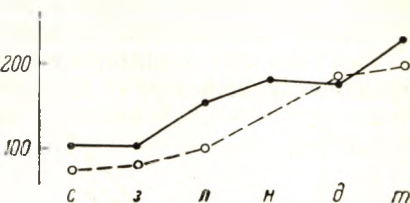


Рис. 6.13. Задержки окончания согласного (в мсек.) у двух испытуемых относительно окончания согласного у диктора для разных согласных.

Точки соответствуют медианам распределения задержек.

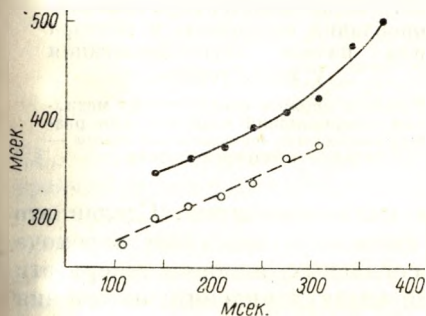


Рис. 6.14. Средняя величина задержек окончания согласного у двух испытуемых относительно начала исходного согласного (по оси ординат, в мсек.) в зависимости от длительности исходного согласного (по оси абсцисс, в мсек.).

начало фонации несколько запаздывало относительно начала смычки, можно было допустить, что необходимым признаком артикуляторного состояния является наличие фонации.

Итак, мы могли сформулировать предположение, что сигналом для размыкания смычки (перехода к гласному) является

сложное событие, состоящее из окончания исходного согласного и начала фонации у самого испытуемого. Результаты проверки этого предположения приведены на рис. 6.15. Можно видеть, что теперь наклон кривых действительно равен 45° . Корреляция между величинами интервалов оказалась весьма высокой. Коэффициент корреляции составил 0.78 и 0.82.

Получив эти данные, казалось возможным принять, что для шумных согласных c , c' сигналом, равноценным началу фонации

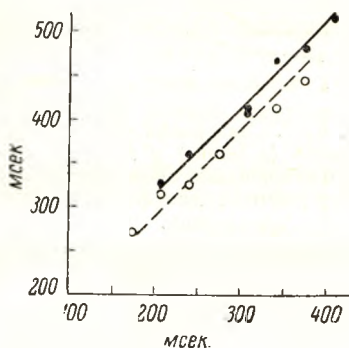


Рис. 6.15. Зависимость между интервалом от начала исходного согласного до конца согласного у двух испытуемых (по оси ординат, в мсек.) и интервалом от начала исходного согласного до последнего по времени из двух явлений: окончания исходного согласного и начала фонации у испытуемого (по оси абсцисс, в мсек.).

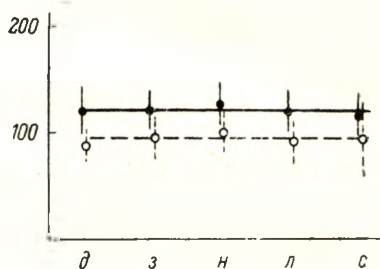


Рис. 6.16. Задержки окончания согласного (в мсек.) у двух испытуемых относительно последнего по времени из двух явлений: окончания согласного у диктора или начала звукообразования у испытуемого.

Точки и кружки соответствуют медианам распределений задержек для разных согласных, вертикальные линии — зоны неопределенности.

при звонких согласных, является появление шума. Сделав это предположение, мы могли снова вычислить задержки перехода к гласному при имитации разных согласных, измеряя теперь эти задержки от последнего по времени из двух явлений: окончания исходного согласного и появления шума или голоса у испытуемого. Оказалось, что величины задержек перехода к гласному, определенные таким образом, практически совпадают при разных согласных (рис. 6.16). Итак, мы можем утверждать, что переход к новому артикуляторному состоянию при имитации определяется как внешним сигналом, так и предыдущим артикуляторным состоянием.

Как уже говорилось, исследование текущего повторения было предпринято в предположении, что изменения артикуляторного состояния при имитации отражают изменения во внутреннем состоянии автомата, преобразующего акустические сигналы в дискретные фонемные символы.

Оставаясь в пределах этого предположения, поставим вопрос о том, в какой момент времени должен образовываться фонемный символ. Ответ на него достаточно очевиден. Фонемный символ целесообразно выдавать, когда система переходит из состояния согласного в состояние гласного. Иначе говоря, появление фонемного символа на выходе автомата должно совпадать по времени с окончанием согласного при имитации.

В предыдущем разделе настоящей главы были приведены результаты исследования задержек буквенной записи согласных. Так как в опытах с имитацией использовались те же слоги типа

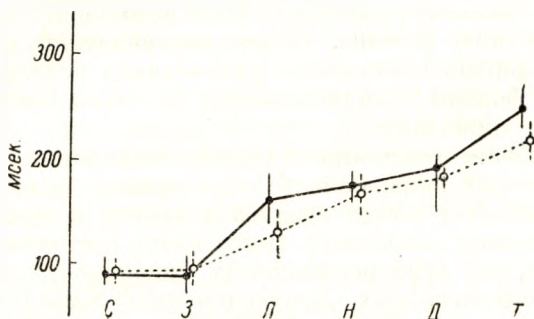


Рис. 6.17. Задержки окончания согласного, создаваемого при имитации (прерывистая линия), и задержки буквенной записи согласных (сплошная линия) от окончания исходного согласного.

Точки — средние значения задержек по группе испытуемых, вертикальные линии соединяют крайние значения индивидуальных средних по отдельным испытуемым.

СГ, что и в опытах с буквенной записью, мы могли произвести сравнение задержек буквенной записи с задержками окончания имитации согласного. Совпадение данных оказалось вполне удовлетворительным (рис. 6.17). Нам кажется, что полученный результат представляет собой весьма веский довод в пользу того, что изменения артикуляции при имитации действительно отражают изменения во внутреннем состоянии автомата, вырабатывающего фонемные символы.

Попробуем рассмотреть теперь вопрос, какой набор символов необходим, чтобы описать состояния артикуляции при имитации (внутренние состояния автомата). Не имея пока экспериментальных данных о том, осуществляются ли во время имитации согласного также некоторые движения, характеризующие следующий гласный (например, движения огубленности), мы будем говорить только о компонентах артикуляции согласного.

Очевидно, что конечный участок имитации согласного, как правило, может быть исчерпывающе описан в терминах фонем. Определенные затруднения возникают, когда артикуляторное

состояние характеризуется наличием двух смычек (щелей), временно образованных в различных участках речевого тракта. Оно может быть описано или одновременным указанием двух фонемных символов, или указанием признаков группы возможных фонем.

Как известно, для описания фонемы не обязательно присваивать каждой из них отдельный специальный символ; вполне достаточно перечислить значения дифференциальных признаков.¹ В фонологической литературе фонема рассматривается как многомерный вектор, компоненты которого соответствуют дифференциальным признакам. Указание значений всех компонент эквивалентно полному описанию фонемы, отсутствие указаний относительно одной какой-нибудь компоненты равносильно указанию группы, включающей фонемы, различающиеся по неуказанному дифференциальному признаку.

Если внутреннее состояние автомата описывается в терминах дифференциальных признаков, то образование согласной фонемы на выходе автомата может сводиться просто к копированию и передаче на выход последнего внутреннего состояния или, вернее, значений тех дифференциальных признаков, которые относятся к системе согласных. Допущение такого механизма приводит к некоторым весьма интересным следствиям. Важнейшее из них состоит в том, что процесс образования фонемы не включает операции выбора, так как экспериментально было показано, что реакции имитации не являются реакциями выбора (Чистович, 1960б; Чистович, Клаас, 1962; Saslow, 1958). Тем самым снимаются основные возражения, выдвигавшиеся против гипотезы текущего фонемного распознавания (см.: Алекин, Клаас, Чистович, 1962; Miller, 1962).

Эти возражения заключались, во-первых, в том, что реакции выбора осуществляются у человека с очень большими задержками, которые растут с увеличением неопределенности сигнала. Следовательно, если бы решение о фонеме производилось путем выбора одной из возможных фонем, то человек просто не успевал бы принимать фонемные решения при текущем восприятии речи.

Следующее возражение состояло в том, что информация о фонеме, заключенная на участке речевого потока, соответствующем звуку речи или слогу, во многих случаях недостаточна для однозначного выбора фонемы. Так как оставшаяся неопределенность может быть ликвидирована при распознавании слов и предложе-

¹ Дифференциальные признаки довольно просто соотносятся с основными различительными особенностями артикуляции согласных (глухость—звонкость, смычка—щель и т. д.), однако в принципе можно допустить, что связь между артикуляторными особенностями и дифференциальными признаками является весьма сложной. Система дифференциальных признаков может рассматриваться просто как система промежуточного кода, используемого при синтезе и распознавании речевых комплексов.

ний, очевидно, что нерационально принимать текущие фонемные решения. При этом неизбежно должны возникнуть ошибки, которых можно было бы избежать, не производя выбора на фонемном уровне.

Если выработка фонемы состоит просто в передаче на выход сведений о дифференциальных признаках, накопленных в процессе последовательных изменений внутреннего состояния автомата, то в случае недостаточной внешней информации фонемный сигнал на выходе автомата также не будет полным. Он будет содержать сведения только о тех дифференциальных признаках, которые были определены, и, следовательно, будет эквивалентен указанию не одной фонемы, но некоторой группы фонем. Это хорошо согласуется с экспериментальными данными относительно распределения ошибок при восприятии согласных в шуме и при частотных ограничениях (Чистович, 1955; Miller a. Nicely, 1955). Было показано, что, полностью не распознав согласный, человек тем не менее устанавливает и помнит определенные признаки этого согласного, позволяющие отнести его в ту или другую группу фонем.

4. Роль ритмической структуры сообщения в распознавании смысловых единиц

Описанные ранее данные привели к представлению, что результатом текущего фонемного преобразования речевого потока является превращение его в последовательность групп символов. Каждая группа соответствует отдельной фонеме, а каждый символ соответствует отдельному дифференциальному признаку. Если речь не вполне разборчива, то группа символов не является полной и, соответственно, не содержит исчерпывающей информации о фонеме, указывая лишь группу возможных фонем. Такая последовательность групп символов могла бы рассматриваться как образ сообщения, используемый на последующих этапах распознавания и анализа предложения. Однако очевидно, что сообщение не может быть исчерпывающе описано последовательностью фонем. Естественно ожидать, что в образе сообщения должны также содержаться сведения о ритмической структуре (распределении ударений по слогам) и других просодических характеристиках сообщения.

Целью описываемых ниже экспериментов было выяснение вопроса о том, в какой мере ритмическая структура сообщения используется при распознавании крупных смысловых единиц — слов и предложений. С этой точки зрения анализировались ошибки, допущенные при распознавании фраз в условиях сильного искажения речи. Ожидалось, что если ритмика сообщения является его важным полезным признаком, используемым как для членения слышимой речи на слова, так и для выбора (расознавания) слов, то фразы, ошибочно принятые вместо переданных, должны совпа-

дать с переданными по числу слов и по ритмике чередования ударенных слогов с безударными.

Работа преследовала также и другую цель; хотелось получить некоторые предварительные представления о возможных процедурах принятия решений о смысловых единицах. Вопрос о процедурах тесно связан с вопросом об объеме оперативной памяти, в которой может быть записан образ сообщения. Рассмотрим это на простом примере. Предположим, что в образе, возникшем в результате предварительного фонемного преобразования, содержится полная информация о гласных и о распределении ударений в исходной фразе, но не содержится информации о месте образования согласных. В таком случае образ предложения *Ваня писал письмо* можно считать эквивалентным следующей записи:

			п'	с		п'	с'		
в	н'	т'	ш		т'	щ	м	о	
а	а	и	а	л	и				
з	м'	к'	х		к'	х'	н		
			ф			ф'			

Если объем оперативной памяти не ограничен, то наилучшим способом распознавания будет запоминание сначала всего образа сообщения, а затем просмотр всех возможных фраз (осмысленных и бессмысленных), которые могут соответствовать этому образу, и выбор методами анализа предложений тех фраз, которые являются осмысленными. В данном случае возможны только две осмысленные фразы, которые будут соответствовать образу: *Ваня писал письмо* и *Ваня писал письмо*. Так как первая из них встречается чаще, то, очевидно, именно ее следует предпочесть (значение частоты встречаемости для выбора слов было показано экспериментально; обзор литературных данных и обсуждение эффекта даются в работе Савина (Savin, 1963)). Если объем оперативной памяти не ограничен, то вероятность правильного опознавания, естественно, должна возрастать с увеличением длины фразы за счет наличия грамматических и вероятностных связей между словами. Экспериментально установлено, что увеличение длины слов приводит к увеличению их разборчивости (Savin, 1963). Хорошо известно также, что фразовая разборчивость выше словесной (Поварков и Глекин, 1956; Покровский, 1956), однако, есть ли предел для роста разборчивости с увеличением длины сообщения, не установлено. Если объем памяти не ограничен, то предела быть не должно: чем длиннее фраза, тем лучше должна быть разборчивость.

Иная процедура распознавания будет рациональной в том случае, если объем памяти ограничен, так что в ней может одновременно находиться образ только некоторого отрезка предложения. Очевидно, что в этом случае необходимо текущим образом принимать решения о словах или словосочетаниях и освобождать

память для записи в нее образов следующих слов. Чем меньше информации содержится в образе слова, тем большим должен быть набор слов (словосочетаний), который должен быть просмотрен, и тем, естественно, больше должно быть время, затрачиваемое на поиск подходящего слова. В этом случае следует ожидать, что должна существовать оптимальная длительность фразы, при которой разборчивость будет максимальной. Эта длительность должна соответствовать объему оперативной памяти. При больших длительностях фразы, в условиях сильных искажений, должны наблюдаться пропуски начальных или срединных частей фразы, связанные с нехваткой времени на их текущий просмотр и установление смысла. С целью проверки такого предположения в наших опытах варьировалась длина переданных фраз.

Искажение речи достигалось путем частотных ограничений. Запись фраз на магнитную ленту производилась через один из фильтров анализатора СЗЧ с полосой пропускания от 906 до 1141 гц и затуханием сигнала на частотах вне полосы пропускания не менее чем на 28 дб.

Как следует из данных акустической теории речи (Fant, 1960) и анализа ошибок в восприятии отдельных звуков при частотных ограничениях (Варшавский и Литвак, 1955; Чистович, 1955; Miller a. Nicely, 1955), примененные нами искажения должны были прежде всего повести к резкому ухудшению возможностей распознавания места образования согласных и различения гласных друг от друга.

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 слов, в большинстве случаев два или три; учитывалось количество полных (знаменательных) слов, имеющих словесное ударение и выражающих отдельные понятия (Грамматика русского языка, 1952). Как правило, переданные словосочетания не содержали служебных слов; если они имелись, то при обработке данных они присчитывались как безударные слоги к последующим полным словам. Фразы были составлены так, что являлись одной синтагмой и произносились слитно (без пауз между словами). После каждой фразы следовала пауза длительностью 20 сек.

Таблицы фраз воспроизводились магнитофоном МАГ-59 и прослушивались группой из нескольких человек, которым предлагалось по возможности полно, без пропусков, записать услышанное; разъяснялось, что записывать следует также и при отсутствии уверенности в том, что фраза понята правильно. Фактически это не всегда удавалось, неразобранные сообщения испытуемые полностью или частично прочеркивали или же оставляли пропуски в записи. Всего было передано и прослушано 1039 фраз (2617 слов), из них коротких, содержащих от трех до семи сло-

гов, — 704 фразы (1592 слова) и более длинных — 335 фраз (1025 слов).

Чтобы сравнивать именно звучание, произношение фраз, мы при обработке материала избавлялись от правописания, изображая каждое переданное сообщение и все варианты записи его испытываемыми фонетическими знаками. Превращение неударного *е*, следующего за мягким согласным, в *и* считалось правильным приемом, так как это соответствует современным произносительным нормам русского языка. Несоблюдение мягкости или твердости согласных не считалось за ошибку.

Суммарная оценка полученных записей показала, что фразовая разборчивость по всему переданному материалу равнялась 30%. Ошибки заключались, во-первых, в замене слов и целых фраз другими словами и фразами, которые могли отличаться от переданных как незначительным количеством звуков, так и большинством звуков. Во-вторых, часто наблюдались пропуски (прочерки) слов и фраз. Длинные фразы пропускались (полностью или частично) чаще, чем короткие. Третий вид ошибок — превращение сказанного слова в группу слогов, утративших словесное ударение, наблюдался редко (3% переданных слов). При этом слоги обычно все же использовались для формирования осмысленных единиц: полностью, или при частичном пропуске, присоединялись к соседним словам. Наблюдались случаи, когда из двух полных слов образовывалось одно, по числу слогов соответствующее двум переданным. Обратное явление — образование двух слов из одного переданного и, следовательно, возникновение лишнего ударения, наблюдалось всего один раз.

Для оценки воспроизведения ритмики переданного сообщения из материала были отобраны все записи фраз, совершенно не соответствующие по смыслу сказанному, например:

Передано

*пошлй в лёс
балёт на льдѹ
мировбй рекбрд
болънобй въздорвел
зелёнобй крокодйл
черёшни поспѣли
живѹ воспомина́ннйми
напра́шивае́шься на комплиме́нт
уве́ренный и споко́бный гблос
са́мая вы́сокая оце́нка*

Записано

*почти нѣт
полѣт на луноѹ
недово́льный ко́т
багрѣм вътяну́ли
наве́рно приходи́л
лежи в постѣли
жела́ет понима́ния
вопрѣс на комби́нат
доро́га из друго́го по́ля
ста́рая фотоплѣнка*

Для всех таких фраз было учтено соответствие с переданными по числу полных слов (что соответствует числу словесных ударений) и по числу слогов во фразе. Кроме того, рассматривалось, на какие слоги записанных неправильных слов приходились ударения, т. е. соответствовало ли положение их положению ударений в сказанной фразе. В большинстве случаев, когда число

записанных слогов равнялось числу переданных, оценка не вызывала сомнений. При неодинаковом числе слогов приходилось использовать в качестве ориентиров правильно распознанные слоги и звуки и рассматривать, можно ли было при данном числе слогов расположить ударения более удачно. Например, сказано *приле́жный ученик*, а записано *желёзный кйт*. Совершенно очевидно, что пропущено два слога в середине фразы, а положение второго словесного ударения — на последнем слоге фразы — правильное, так же как и первого. В случае, когда было записано *бродяга Ма́ша* вместо *трудо́ная зада́ча* — первое ударение неправильное, а второе, на предпоследнем слоге фразы, правильное; пропущен один слог. При записи *трудо́ная Лю́ся* вместо *прыга́й выше́*, очевидно, что окончание *ай* превращено в *ая* (*аја*) благодаря чему появился второй звук *а* (лишний слог) в середине фразы, а оба ударения, на первом и на предпоследнем слогах фразы, соответствуют переданным (приняты правильно); ритмика фразы частично нарушена за счет лишнего слога.

Те же соображения использовались при оценке правильности распределения слогов между словами. В приведенных выше примерах считалось, что границы между словами найдены правильно. В записи *два́ бу́льдо́га* вместо *заши́ть на́до* сохранено правильное количество слогов во фразе, и правильно второе ударение, но слоги неправильно разделены по словам и неправильно первое ударение.

Весь полученный материал обнаруживал существенное отличие разборчивости фраз длиннее семи слогов от разборчивости более коротких фраз. Поэтому результаты обработки записей, совершенно неправильных по смыслу, были разбиты на соответствующие две группы. В группе коротких фраз материала значительно больше, так как коротких сообщений вообще было больше, и они, кроме того, чаще записывались без пропуска слов. Вместе с тем совсем не оказалось записей фраз длиннее десяти слогов без полных пропусков слов и одновременно целиком неправильных по смыслу.

Анализ материала показал, что записи, неправильные по смыслу, плохо соответствуют звуковому составу прослушанных сообщений. Они содержат всего немногим более 50% правильно принятых гласных, менее 40% правильных согласных и совсем уже мало соответствуют переданным фразам по составу слогов. Более детальное рассмотрение записей позволяет считать, что замены согласных не являются случайными. Так, вместо произнесенного *фсе́фпрóшлaм* один из слушавших записал фразу, которая звучит *фхóтплáтный*. Из восьми согласных фонем правильно воспроизведены только *ф* и *п*, однако все сонанты заменены также сонантами, а глухие согласные — другими глухими.

Эти неправильные по звуковому составу фразы тем не менее достаточно хорошо соответствовали переданным по числу слов

и слогов. Из 79 коротких фраз 68 совпали с переданными по числу слов, а 57 — по числу слогов. В длинных фразах число слов также в большинстве случаев сохранялось неизменным (16 случаев из 22), однако записанные слова часто оказывались короче переданных. В результате только 6 из 22 длинных фраз совпали с переданными по числу слогов.

Наибольшее соответствие с переданными сообщениями обнаружилось по словесным ударениям. В коротких и в длинных фразах соответственно 90% и 84% переданных словесных ударений были воспроизведены на правильном месте. Так как записанные фразы достаточно точно отражали ритмику чередования ударных слогов с безударными произнесенной фразы, очевидно, что выбор фраз был явно ограничен ритмикой переданного сообщения.

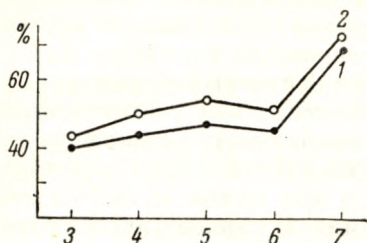


Рис. 6.18. Разборчивость слов в зависимости от порядка следования их во фразах из двух слов.

1 — соответствует первым словам, 2 — вторым. По оси абсцисс — число слогов в фразе, по оси ординат — доля правильно принятых слов (в %).

Другим существенным ограничением, вводимым человеком, являлось то, что предложения или словосочетания допускались только осмысленные (грамматически правильные и с удовлетворительно соответствующими друг другу лексемами).

Анализ ошибок позволяет утверждать, что при выборе фраз использовалась информация о рит-

мике сообщения и о признаках отдельных звуков речи. Очевидно, что те же сведения использовались и при осуществлении правильного выбора фраз.

Воспроизведение ритмики сообщения, так же как и отсутствие в записях бессмысленных фраз, говорит о том, что решения принимались сразу о больших, осмысленных отрезках речи. В литературе высказывались соображения в пользу того, что единицей решения является целая фраза (Miller, 1962). Однако осмысленная грамматически правильная фраза может быть получена и в результате такой процедуры принятия решений, когда единицами решений являются слова (словоформы). Миллер, Хейзе, и Лихтен (Miller, Heise a. Lichten, 1951) показали, что слова в составе фраз распознаются лучше, чем предъявляемые в случайном порядке, и объяснили эффект повышения разборчивости тем, что контекст фразы сокращает число возможных слов, среди которых слушающий должен осуществить выбор. В самое последнее время этот эффект детально исследовали Стоу, Харрис и Хэмптон (Stowe, Harris, a. Hampton, 1963). Иначе говоря, можно представить себе, что выбор первого слова ограничивает словарь, из которого выбирается второе слово, и так же обстоит дело со всеми

последующими выборами. При этом может происходить увеличение вероятностей правильного приема в направлении от первого слова к последнему. Если же решение принимается сразу о целой фразе, положение слова во фразе не должно влиять на его разборчивость.

Сопоставление разборчивости слова, первого во фразе, с разборчивостью второго, проведенное на материале 570 фраз, состоявших из двух слов, показало, что в действительности место, занимаемое словом во фразе, отчетливо влияло на его разборчивость (рис. 6.18). Во всех пяти группах фраз различной длины последние слова распознавались лучше, чем первые. Оба слова оказывались пропущенными в среднем одинаково часто, возрастание процента правильно принятых слов происходило за счет уменьшения ошибок (рис. 6.19, А).

Рассмотрение 418 фраз из трех слов различной длины (от пяти до двенадцати слогов) обнаружило, что среднее во фразе слово распознавалось хуже, чем первое и последнее (рис. 6.19, Б). Этого следовало ожидать, если единицами распознавания являлись слова. Миллер (Miller, 1962) наблюдал, что распознавание слов на фоне шума улучшается с удлинением пауз между словами. В нашем случае фразы произносились слитно, по несколько слов в секунду, и большие задержки решений об искаженных словах должны были препятствовать преимущественно текущему распознаванию слов, средних во фразе, которые начинали звучать в то время, когда еще не было принято решение о первом слове, и которые все еще обрабатывались, когда непосредственно за ними поступало последнее слово.

Итак, полученные данные заставляют предположить, что человек принимает решение о словах, не дожидаясь окончания всей фразы. Такое поведение является целесообразным, если объем оперативной памяти ограничен, и она не может вместить образа всей фразы в целом. Как уже говорилось, в случае не ограниченной памяти следовало ожидать, что разборчивость фраз должна увеличиваться с увеличением длины фразы.

С целью проследить, как изменялась разборчивость фраз в зависимости от их длины, весь полученный материал был разбит на группы по числу слогов, составлявших переданную фразу

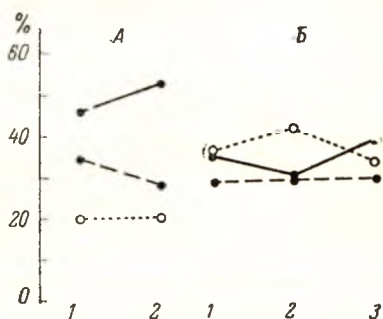


Рис. 6.19. Характеристика приема слов в зависимости от их места во фразах, состоящих из двух слов (А) и из трех слов (Б).

По оси абсцисс — порядковый номер слов во фразе, по оси ординат — доля (в %) правильно принятых (сплошная линия), ошибочно записанных (пунктирная линия), и пропущенных (пунктир) слов.

(при прослушивании фразы различных размеров следовали в случайном порядке). Оказалось, что некоторое возрастание разборчивости с удлинением сообщения действительно имело место, но только в пределах до семи слогов (рис. 6.20, А). Дальнейшее увеличение сообщения вызывало падение разборчивости. Повышению фразовой разборчивости соответствовало несомненное и систематическое снижение процента неправильно записанных слов (рис. 6.20, Б). Прослушав фразы из трех слогов, испытуемые одинаково часто записывали правильные слова и ошибочные, тогда как при прослушивании фраз из семи слогов правильных

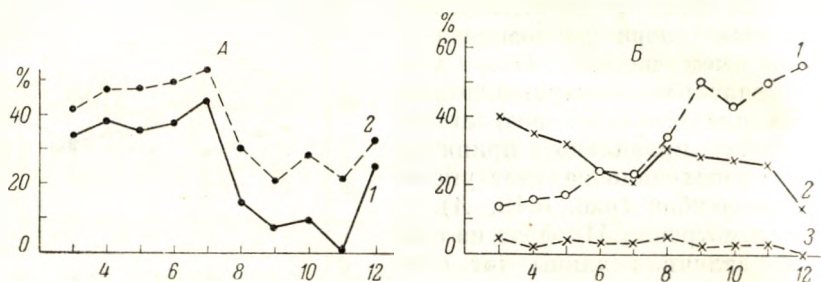


Рис. 6.20. Характеристика приема фраз и слов в зависимости от числа слогов в сообщении (по оси абсцисс).

А — разборчивость (в %) фраз (1) и слов (2). Б — доля (в %) пропущенных слов (1), ошибочно записанных слов (2), и слов, утративших словесное ударение (3).

записей слов производилось вдвое больше, чем ошибочных. В результате разборчивость фраз возрастала, несмотря на то, что процент пропущенных слов возрастал с удлинением сообщения.

Низкая фразовая разборчивость сообщений длиннее семи слогов связана в основном с пропуском слов (46 вместо 19%), а также и целых фраз (29 вместо 12%). Правильные и ошибочные слова записывались одинаково часто.

Резкое снижение разборчивости фраз, превысивших размеры семи слогов, заставляет считать, что примерно семью слогами исчерпывается емкость памяти на образ речевого сообщения. Это хорошо согласуется с известными данными о том, что человек может запомнить и воспроизвести серию бессмысленных слогов, также не превышающую в среднем семи единиц (Miller, 1956).

Нет оснований предполагать, что решения о словах задерживались до исчерпания возможностей запоминания образа сообщения. Очевидно, слова распознавались текущим образом, по мере их поступления, в коротких сообщениях так же, как и в длинных. Об этом говорит то, что зависимость разборчивости слова от его места в сообщении отчетливо видна на материале коротких фраз (см. рис. 6.18). Однако не все слова узнавались по мере их поступления. Можно думать, что по окончании сооб-

щения, не превысившего емкости памяти, оказывалось возможным использовать сохранившиеся в памяти признаки сообщения для отставленного решения о слове, не узнанном текущим образом. Если же фраза была длинна, то образ слова, т. е. сведения о его размерах, положении в нем ударения и т. д., утрачивался; тогда решение о неузнанной части фразы, например о сказуемом, могло приниматься только по догадке — на основании одних лингвистических вероятностей, без ограничения признаками слова и, следовательно, с большею вероятностью ошибки.

Результаты настоящего исследования показывают, что отрезки слышимой речи, о которых принимаются решения, во-первых, превышают размеры слога (поскольку ошибочные записи доказывают, что выбор был ограничен ритмикой чередования ударенных слогов с безударными), и, во-вторых, не выходят за пределы порядка семи слогов. Совершенно очевидно, что такими отрезками речи являются слова. Представление о слове как единице решения (восприятия) получает распространение в современной лингвистике. Данные в пользу пословного восприятия речи приводятся в обзоре А. А. Леонтьева (1961).

Процедуру распознавания слышимой речи человеком можно представить примерно следующим образом. Поступающие друг за другом звуки речи записываются в памяти как наборы характеристик их по признакам; гласные записываются с марками, обозначающими степень их ударности. После восприятия ударного слога намечается условная граница слова, и ищется в словаре подходящее слово.

Если решение принимается, тем самым отмечаются границы отрезка, включенного в слово, и сокращается словарь последующих выборов.

Словесные ударения позволяют узнать число полных слов во фразе, и в то же время положение ударения является важным дифференциальным признаком слова. Таким образом, отрезки сообщения, более крупные, чем слоги, приобретают новый полезный акустический параметр — ритмику сообщения. Большая разборчивость фраз и слов по сравнению с разборчивостью бессмысленных слогов несомненно в значительной мере зависит от того, что, распознавая осмысленное сообщение, человек использует дополнительный признак, не разрушаемый частотными искажениями и шумами, — ритмику чередования ударных слогов с безударными.

Г Л А В А 7

О ВОЗМОЖНЫХ ПРОЦЕДУРАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

Описанные в предыдущих главах результаты исследования организации речевого процесса и процесса восприятия речи позволяют высказать некоторые соображения о существующих процедурах автоматического распознавания речи и о возможных путях совершенствования таких процедур.

Весь материал, касающийся вопросов организации артикуляторных движений, с несомненностью показывает, что речь не может рассматриваться как простая последовательность звуков речи, каждый из которых обладает своими собственными характеристиками, независимыми от характеристик соседних элементов. Это заставляет считать, что процедуры фонемного автоматического распознавания (Мясников, 1943; Dreyfus-Graf, 1950, 1962; Smith, 1951; Wiren a. Stubbs, 1956; Fry a. Denes, 1958; Steinbuch, 1958; Grabbe, 1960; Sakai, Doshita, Nagata, Sekimoto, 1962, и др.) вряд ли можно считать перспективными. Как было показано в главе, посвященной организации слога, артикуляторные органы изменяют свое положение в соответствии с последующим гласным уже с самого начала произнесения первого согласного в слоге. Два согласных в слоговой группе типа ССГ не являются последовательными в том смысле, что следующий согласный артикулируется только после того, как был произнесен первый согласный. Не только при разговорном, но и при полном стиле произнесения второй согласный, как правило, начинается на протяжении артикуляции первого согласного. Если дело идет о сочетании двух смычных неносовых согласных с разным местом образования, то акустическая картина, соответствующая сочетанию двух согласных, находящихся между гласными (ГССГ), должна иметь следующий вид: переход от гласного к первому согласному должен быть типичным для первого согласного; затем

должна следовать одна удлиненная (глухая или звонкая) смычка; затем должен быть взрыв и переход ко второму гласному, характерные для второго согласного. Естественно, что для того чтобы такая акустическая картина могла быть предусмотрена в системе для автоматического распознавания речи, целесообразно, чтобы единицы распознавания в этой системе соответствовали слогам (слоговым группам), но не фонемам.

Результаты, полученные в настоящей работе, заставляют считать, что слог как артикуляторный комплекс обладает тем обязательным признаком, что он заканчивается гласным. Это означает, что принципиально возможно ввести формальную процедуру членения речевого потока на слоговые группы, основывающуюся на наблюдении гласного. Одна из возможных процедур может заключаться в следующем: с момента начала гласного (довольно легко определяемого по акустическим признакам) производится слежение за изменением формант (и огибающей сигнала). Пока изменения происходят в одном направлении, слог считается продолжающимся. Как только направление движения формант изменяется, определяется начало нового слога, обработка данных о первом слоге заканчивается, полученные результаты передаются в устройство памяти или на выход системы.

Конечный согласный (группа согласных) в конце синтагмы может в таком случае рассматриваться как редуцированный слог (слоговая группа). Выделение такого особого элемента не должно представлять больших трудностей, так как между синтагмами существуют паузы.

Попытки использовать для автоматического распознавания отрезки речевого сигнала, приближающиеся по длительности к слогам, были предприняты Олсоном и Беларом (Olson a. Belar, 1956, 1961). Возражения против использования слогов в качестве единиц распознавания могут быть двух типов. С одной стороны, число слогов, возможных в языке, значительно превышает число возможных фонем. Это как будто бы должно приводить к тому, что устройство, распознающее слоги, должно быть более сложным, чем устройство для фонемного распознавания. С другой стороны, для целого ряда целей (устный ввод данных в вычислительные машины, машинный перевод устной речи, кодирование команд и т. д.) кажется более рациональным использование в качестве единицы распознавания такого крупного элемента речи, как слово (обсуждение этого вопроса см.: Сапожков, 1963). Миллер (Miller, 1962), рассматривая вопрос о возможных единицах распознавания речи, приходит к выводу, что наибольшая надежность может быть обеспечена, если единицы распознавания будут соответствовать словам или фразам.

Рассмотрим сначала второе возражение. Как показывают наши данные, человек, принимая окончательное решение о слышимом сообщении, в очень большой мере использует сведения

о ритмической структуре этого сообщения. Однако ритмическая структура, будучи признаком синтагмы (слова), является общей для большего числа синтагм (слов) и в очень малой степени зависит от конкретной фонетической характеристики слогов, входящих в эту синтагму (слово). Поэтому, чтобы записать ритмическую структуру синтагмы, нет необходимости перечислять фонетические характеристики каждого конкретного слога. Достаточно указать число слогов, пронумеровать их и относительно каждого указать некоторый набор данных, касающихся длительности, интенсивности и, вероятно, частоты основного тона. Более того, ритмическую структуру принципиально нельзя описать, предварительно не расчленив синтагму на слоги и не заменив конкретные акустические изображения слогов на абстрактные сигналы, обозначающие эти слоги.

Как показали результаты нашей работы, ритмическая структура синтагмы является инвариантной (независимой от темпа речи) только на уровне слогов и более крупных элементов речи. Это означает, что процесс нормирования синтагмы по времени, необходимый для того, чтобы исключить влияние темпа речи, может применяться только к последовательностям сигналов, обозначающих слоговые команды (элементы), но не к оригинальной речи. Изменения темпа в разной мере влияют на длительность гласных и согласных, они по-разному сказываются на стационарных и переходных участках. Поэтому нормирование по времени принципиально не может быть осуществлено простым сжатием или растягиванием оригинального сигнала (процедура, применявшаяся Denes a. Mathews, 1960).

Таким образом, можно утверждать, что для опознания слова (синтагмы) целесообразно располагать не его непосредственным акустическим изображением (например, динамической спектрограммой), но уже извлеченными из этого изображения сведениями о его ритмической структуре и конкретных фонетических характеристиках входящих в слово слогов. Процедура распознавания слов, основывающаяся на сравнении оригинальной спектрограммы слова с эталонными спектрами (спектрально-временными распределениями) целых слов (Petrick a. Willett, 1960; Denes a. Mathews, 1960, и др.), должна приводить к худшим результатам, не говоря уже о том, что в этом случае резко возрастает объем памяти распознающего устройства.

Рассмотрим теперь вопрос, следует ли из принятия слогов в качестве единиц распознавания обязательное увеличение памяти распознающего устройства по сравнению со случаем фонемного распознавания. Очевидно, что ответ зависит от того, как будет организована процедура распознавания слогов.

За последние годы основной метод, принятый при распознавании речи, состоит в сравнении спектрально-временных распределений анализируемых звуков с эталонами — типовыми рас-

пределениями, заложенными в памяти устройства (Forgie a. Forgie, 1959; Foulkes, 1963; Welch a. Wimpres, 1961). Во всех этих методах за основу взята теория статистического опознавания образов. Отклонение реального принимаемого сигнала от эталона рассматривается как результат действия помехи; задача заключается в том, чтобы найти эталон, наименьшим образом отличающийся от принимаемого сигнала (обсуждение этого метода на примере распознавания русских гласных см.: Чистович, 1957, 1958). Если пользоваться таким методом, то принятие слогов в качестве единиц распознавания должно привести к существенному усложнению распознающего устройства. Число эталонов должно быть в этом случае, естественно, равно числу возможных слогов.

Однако к настоящему времени накоплен ряд данных, которые заставляют предполагать, что в случае речи такая процедура распознавания может быть далеко не оптимальной. Если бы речь представляла собой последовательность стационарных участков и все трудности распознавания сводились бы к тому, что спектральные характеристики этих стационарных участков варьируют от диктора к диктору, то данный метод, очевидно, был бы наилучшим. Однако в речи, особенно произносимой в быстром темпе, вообще может не обнаруживаться стационарных участков. Кроме того, темп речи может меняться в очень широких пределах. В силу недостаточной синхронности движений различных органов, участвующих в артикуляции, сильно флюктуируют по длительности также и отдельные сегменты одного и того же звука при его повторных произнесениях. Все это приводит к тому, что успешность использования рассматриваемой процедуры распознавания становится критическим образом зависимой от успешности процедуры предварительного членения речи на элементы. К сожалению, ни один из описанных до сих пор методов членения речи не является достаточно надежным.

Как показывают данные о восприятии речи человеком (глава 6), задача может решаться принципиально иным способом. Выделение слога как элементарной единицы речевого потока, на основании которой принимаются решения, еще не означает, что единицы решений должны соответствовать слогам. Вполне допустимы такие процедуры, при которых единицы решений соответствуют фонемам, однако для принятия этих решений используется информация, распределенная на участке слога. При этом минимальными носителями полезной информации в слоге могут считаться акустические явления, происходящие при переходе речевого тракта из одного состояния в другое. Так как эти вопросы довольно подробно обсуждались в предыдущей главе, мы не будем к ним возвращаться и сошлемся только на некоторые интересные с этой точки зрения факты, полученные при исследовании артикуляции.

Исследование артикуляторных движений, сделанное в настоящей работе, показывает, что инвариантным является только набор движений, которые производятся на протяжении слога. Временные отношения между отдельными движениями отнюдь не являются постоянными. Осуществление требуемого набора движений означает, что обязательно произойдут также и определенные изменения в акустическом сигнале. Например, поднятие кончика языка к верхним зубам при произнесении *m* обязательно приведет к изменению частоты второй форманты предшествующего гласного, размыкание голосовых связок обязательно приведет к прекращению акустического сигнала и т. д.

Но если дело обстоит таким образом, то определение слога может быть сведено к перечислению ряда акустических явлений, обнаруженных на протяжении произнесения этого слога. Так, как каждое акустическое явление в речи соответствует определенному изменению состояния артикуляторных органов, то такое перечисление акустических явлений будет эквивалентно перечислению артикуляторных изменений, осуществленных на протяжении этого слога.

Если акустические явления выделены, переход от них к дифференциальным признакам и фонемам уже не должен представлять особых трудностей.

Вопрос о том, какие именно акустические явления должны оцениваться на протяжении слога, стал в настоящее время достаточно понятным благодаря созданию акустической теории речеобразования и многочисленным исследованиям восприятия как синтетической речи, так и сегментов естественной речи. Практически совершенно неясным остается, однако, вопрос о конкретных способах автоматического выделения и измерения этих явлений. Нам представляется, что исследование этого вопроса является одной из важнейших задач работы по автоматическому распознаванию речи. Можно думать, что весьма существенные данные в этом направлении могли бы быть получены с помощью методов «анализа через синтез», разработанных в Массачусетском технологическом институте (Stevens, 1960; Bell, Fujisaki, Heinz, Stevens, House, 1961).

Очевидно, однако, что полностью задача автоматического распознавания речи не сможет быть решена до тех пор, пока не будут разработаны методы машинного анализа предложения. Вероятно, решение этой задачи применительно к письменным текстам еще не будет означать ее решения для случая устной речи. Специфическими особенностями символической записи устного предложения должно быть отсутствие интервалов между словами и неполное распознавание отдельных фонем.

ЛИТЕРАТУРА

- А в а п е с о в Р. И. 1956. Фонетика современного русского литературного языка. М.
- А л е к с е е в М. А. 1963. Об основных закономерностях условнорефлекторной деятельности мозга в связи с теорией поиска. В сб.: Физиологические механизмы координации движений человека, М.: 3—10.
- А л ё к и н Р. О., Ю. А. К л а а с, Л. А. Ч и с т о в и ч. 1962. Временные задержки буквенной записи человеком-наблюдателем. Акустич. журн., 8: 26—33.
- А н д р е е в Н. Д. и Л. Р. З и н д е р. 1963. О понятиях речевого акта, речи, речевой вероятности и языка. ВЯ, 3: 15—21.
- А р т ё м о в В. А. 1956. Экспериментальная фонетика. М.
- Б а с с и н Ф. В. и Э. С. Б е й н. 1957. О применении электромиографической методики в исследовании речи. Материалы совещания по психологии (1—6 июля 1955). М.: 315—327.
- Б е р н ш т е й н Н. А. 1947. О построении движений. М.
- Б о г о м о л о в В. Н. 1961. Устройства с датчиками Холла и датчиками магнитосопротивления. М.—Л.
- Б о г о р о д и ц к и й В. А. 1909. Опыт физиологии общерусского произношения в связи с экспериментально-фонетическими данными. Казань.
- Б о г о р о д и ц к и й В. А. 1930. Фонетика русского языка в свете экспериментальных данных. Казань.
- Б о н г а р д М. М. 1961. Моделирование процесса узнавания на цифровой счетной машине. Биофиз., 6: 129—141.
- Б р а в е р м а н Э. М. 1962. Опыты по обучению машины распознаванию зрительных образов. Автомат. и телемех., 23: 349—364.
- Б р о в ч е н к о Т. А. 1954. Сравнительный анализ гласных украинского и английского языков. Уч. зап. МГПИИЯ, 8: 24—55.
- Б р о к О. 1910. Очерк физиологии славянской речи. Энциклопедия славянской филологии, 5, 2. СПб.
- В а й т у л е в и ч С. Ф., В. А. К о ж е в н и к о в, А. П. Л е б е д е в, В. И. С о р о к о. 1962. Электронный ключ с малым уровнем помех. Приборы и техн. экспер., 5: 185—186.
- В а р ш а в с к и й Л. А. и И. М. Л и т в а к. 1955. Исследование формантного состава и некоторых других физических характеристик звуков русской речи. В книге: Проблемы физиологической акустики, М.—Л., 3: 5—17.
- В и н е р Н. 1958. Кибернетика. М.
- В о р о н и н Л. Г. 1952. Анализ и синтез сложных раздражителей у высших животных. Л.
- В у д в о р т с Р. 1950. Экспериментальная психология. М.
- Г в о з д е в А. Н. 1949. О фонологических средствах русского языка. М.—Л.

- Гельфанд И. М., В. С. Гурфинкель, М. Л. Цетлин. 1962. О тактиках управления сложными системами в связи с физиологией. В сб.: Биологические аспекты кибернетики, М.: 66—73.
- Глезер В. Д. 1954. Зрачкороваспирительная реакция. Физиол. журн. СССР, 38 : 604—611.
- Грамматика русского языка. 1952. Под ред. Виноградова В. В., Е. С. Истриной, С. Г. Бархударова. Т. I. Фонетика и морфология. М.
- Деркач М. Ф. 1958. Статистика восприятия глухих взрывных и щелевых согласных в зависимости от их длительности. В сб.: Вопросы статистики речи, Л.: 40—45.
- Дукельский Н. И. 1962. Принципы сегментации речевого потока. М.—Л.
- Жгенти С. М. 1960. Некоторые новшества в Научно-исследовательской лаборатории экспериментальной фонетики Тбилисского государственного университета. Тр. Тбилисск. гос. унив., 93 : 399—412.
- Живкин Н. И. 1958. Механизмы речи. М.
- Зиндер Л. Р. 1960. Общая фонетика. Л.
- Златоустова Л. В. 1954. Длительность гласных и согласных звуков русского языка. Уч. зап. Казанск. гос. унив. им. В. И. Ульянова-Ленина, 114, 6 : 99—123.
- Златоустова Л. В. 1962. Фонетическая структура слова в потоке речи. Казань.
- Иванов В. В. и С. К. Шаумян. 1961. Лингвистические проблемы кибернетики и структурная лингвистика. В сб.: Кибернетику на службу коммунизму, 1, М.—Л.: 218—234.
- Качурин Л. Г. 1962. Электрические измерения аэрофизических величин. Л.
- Ковалевский В. А. 1962. Корреляционный метод распознавания изображений. Журн. вычислит. матем. и матем. физики, 2 : 684—691.
- Кожевников В. А. и В. С. Шупляков. 1962. О приемах одновременной регистрации артикуляционных и акустических параметров речи. Вopr. психол., 6 : 128—134.
- Кратин Ю. Г. 1955. К методике записи колебаний электрических потенциалов речевой мускулатуры. Журн. высш. первн. деят., 5 : 591—594.
- Крид Р., Д. Денни-Броун, И. Икклс, Е. Лидделл и Ч. Шеррингтон. 1935. Рефлекторная деятельность спинного мозга. М.—Л.
- Криксов В. Г. 1958. Реостатно-емкостные генераторы синусоидальных колебаний. Киев.
- Кузьмин Ю. И. 1963. Динамическая палатография. Вopr. психол., 1 : 137—141.
- Кучстман К. И. и Л. А. Орбели. 1924. О последствиях деафферентации задней конечности у собак. Изв. Ленингр. научн. ин-ст. им. П. Ф. Лесгафта, 9, 2 : 187—195.
- Левацов М. М., В. А. Цукерман. 1961. Фотографический метод регистрации и временной развертки нистагма и произвольных движений глаза. Журн. ушных, носовых и горловых болезней, 4 : 21—24.
- Левидов И. И. 1939. Певческий голос в здоровом и больном состоянии. М.—Л.
- Леонтьев А. А. 1961. Психолингвистика и проблема функциональных единиц речи. В сб.: Вопросы теории языка в современной зарубежной лингвистике, М.: 163—190.
- Ликлайдер. 1960. Симбиоз человека и машины. Зарубежная радиоэлектроника, 9 : 84—96.
- Лурия А. Р. 1947. Травматическая афазия. М.
- Лурия А. Р. 1962. Высшие корковые функции человека и их нарушения при локальных поражениях мозга. М.

- Маккей Д. М. 1956. Проблема образования понятий автоматами. В сб.: Автоматы, М.: 307—325.
- Матусевич М. И. 1948. Введение в общую фонетику. Л.
- Москаленко Ю. Е. и А. И. Науменко. 1956. К теории методики электроплетизмографии. Физиол. журн. СССР, 42: 312—316.
- Мясников Л. Л. 1943. Объективное распознавание звуков речи. Журн. техн. физ., 13, 3: 109—115.
- Нечаев Г. К. и Н. П. Удалов. 1961. Реле и датчики с полупроводниковыми термосопротивлениями. М.—Л.
- Орбели Л. А. 1938. Лекции по физиологии нервной системы. М.—Л.
- Орбели Л. А. 1949. Вопросы высшей нервной деятельности. М.—Л.
- Орлов В. В. 1961. Плетизмография. М.—Л.
- Павлов И. П. 1951а. Проба физиологического понимания симптомологии истерии. Полн. собр. соч., 3, 2, М.—Л.: 195—218.
- Павлов И. П. 1951б. Физиология высшей нервной деятельности. Полн. собр. соч., 3, 2, М.—Л.: 219—234.
- Парtridge Г. 1961. Электронные измерительные приборы. М.—Л.
- Петрович Н. Т., А. В. Козырев. 1954. Генерирование и преобразование электрических импульсов. М.
- Поварков А. И., Г. В. Глекин. 1956. Зависимость между различными видами артикуляции русской речи. В сб.: Восприятие звуковых сигналов в различных акустических условиях. М.: 74—82.
- Покровский Н. Б. 1956. Разборчивость речи в различных акустических условиях и основы методики ее расчета. В сб.: Восприятие звуковых сигналов в различных акустических условиях. М.: 57—68.
- Рач Т. К. 1960. Двигательные системы. В кн.: Экспериментальная психология. I. М., 216—283.
- Сапожков М. А. 1963. Речевой сигнал в кибернетике и связи. М.
- Сергиевский М. В. и Ю. Н. Иванов. 1961. Краткий обзор исследований по физиологии дыхания за последние 10 лет. Тр. Куйбышевск. мед. инст., 18.
- Скалозуб Л. Г. 1963. Палатограммы и рентгенограммы согласных фонем русского литературного языка. Киев.
- Соколов А. Н. 1960. Динамика и функции внутренней речи (скрытой артикуляции) в процессе мышления. (По данным электрографии речевой мускулатуры). Изв. АПН РСФСР, 113: 149—182.
- Томашевский Б. 1929. О стихе. Л.
- Туричин А. М. 1951. Электрические измерения неэлектрических величин. М.—Л.
- Фонарев А. М. 1956. Прибор для электрической регистрации микродвижений языка при скрытой артикуляции. Изв. АПН РСФСР, 81: 99—101.
- Харкевич А. А. 1959. Опознавание образов. Радиотехника, 14, 5: 12—22.
- Хессе Г. 1962. Использование датчика сопротивления для физиологических целей. В сб.: Электроника в медицине, Рига: 60—64.
- Хохлов А. В. 1960. Исследования элементов механизма фонации методом эндоларингографии. Автореф. дисс. Л.
- Хомский Н. 1962. Синтаксические структуры. В сб.: Новое в лингвистике, 2, М.: 412—527.
- Цемель Г. И. 1961. Автоматическое (объективное) распознавание звуков речи. Зарубежная радиоэлектроника, 4: 52—73.
- Чаргейшвили А. К., Т. Л. Тохадзе и В. В. Польшин. 1959. Электромиографическое исследование речи как способ изучения функционального состояния слухового анализатора. Вестн. оториноларинг., 3: 9—13.
- Чернышев В. 1915. Законы и правила русского произношения. Пгр.
- Чистович Л. А. 1955. Влияние частотных ограничений на разборчивость русских согласных звуков. В сб.: Телефонная акустика, 1—2, Л.: 35—42.

- Чистович Л. А. 1957. О методах обработки спектрограмм гласных звуков. *Биофиз.*, 2 : 441—451.
- Чистович Л. А. 1958. Применение статистических методов к определению фонетической принадлежности индивидуального гласного звука. В сб.: *Вопросы статистики речи*, Л. : 26—35.
- Чистович Л. А. 1959. Различение интервала времени между двумя короткими звуковыми импульсами. *Акустич. журн.*, 5 : 480—484.
- Чистович Л. А. 1960а. Восприятие звуковой последовательности. *Биофиз.*, 5 : 671—676.
- Чистович Л. А. 1960б. Классификация звуков речи при их быстром повторении. *Акустич. журн.*, 6 : 392—398.
- Чистович Л. А. 1961. Текущее распознавание речи человеком, I. В сб.: *Машинный перевод и прикладная лингвистика*, 6, М. : 39—79.
- Чистович Л. А. 1962. Текущее распознавание речи человеком, II. В сб.: *Машинный перевод и прикладная лингвистика*, 7, М. : 3—44.
- Чистович Л. А., В. В. Алякринский, В. А. Абульян. 1960. Временные задержки при повторении слышимой речи. *Вопр. психол.*, 1 : 114—120.
- Чистович Л. А., Л. В. Бондарко. 1963. Об управлении артикуляционными органами в процессе речи. (Анализ нарушений речи при задержке обратного акустического сигнала). Конференция по структурной лингвистике, посвященная базисным проблемам фонологии. Тез. докл., М. : 103—105.
- Чистович Л. А. и Ю. А. Клаас. 1962. К анализу скрытого периода «произвольной» реакции на звуковой сигнал. *Физиол. журн. СССР*, 48 : 899—906.
- Чистович Л. А., Ю. А. Клаас и Р. О. Алёкин. 1961. О значении имитации для распознавания звуковых последовательностей. *Вопр. психол.*, 5 : 173—182.
- Чистович Л. А., Ю. А. Клаас, Ю. И. Кузьмин. 1962. Текущее распознавание звука речи. (Опыты с быстрым повторением и буквенной записью согласных). *Вопр. психол.*, 6 : 26—29.
- Шеррингтон Ч. С. Интегративная деятельность нервной системы. Приложение в кн.: Крид Р., Д. Денни-Броун, И. Икклс, Е. Лидделл и Ч. Шеррингтон. 1935. Рефлекторная деятельность спинного мозга. М.—Л. : 163—257.
- Шефтель И. Т. 1956. Полупроводниковые термочувствительные сопровствления. Л.
- Щерба Л. В. 1912. Русские гласные в качественном и количественном отношении. СПб.
- Щерба Л. В. 1915. О разных стилях произношения и об идеальном фонетическом составе слов. *Зап. неофилолог. общ. при имп. Петрогр. унив.*, 8 : 339—347.
- Щерба Л. В. 1955. Фонетика французского языка. М.
- Abercrombie D. 1957. Direct palatography. *Ztschr. f. Phonetik*, 10, 1 : 21—25.
- Arnold G. E. 1957. Morphology and physiology of the speech organs. *Manual of phonetics*. Edited by L. Kaiser, Amsterdam : 31—64.
- Atkinson C. J. 1953. Adaptation to delayed side-tone. *Journ. Speech and Hearing Disorders*, 18 : 386—391.
- Atkinson C. J. 1954. Some effects on intelligibility as the sidetone level and the amount of side-tone delay are changed. *Proc. Iowa Acad. Sci.*, 61 : 334—340.
- Bastian J., P. Delattre and A. M. Liberman. 1959. Silent interval as a cue for the distinction between stops and semivowels in medial position. *Journ. Acoust. Soc. Am.*, 31 : 1568.
- Beaumont J. T. and B. M. Foss. 1957. Individual differences in reacting to delayed auditory feed-back. *Brit. Journ. Psychol.*, 48, 2 : 85—89.

- Bell C. G., H. Fujisaki, J. M. Heinz, K. N. Stevens and A. S. House. 1961. Reduction of speech spectra by analysis-by-synthesis techniques. *Journ. Acoust. Soc. Am.*, 33 : 1725—1736.
- Bellarmino L. 1885. Anwendung der graphischen Methode bei Untersuchung der Pupillenbewegungen. *Photocoreograph. Arch. Ges. Physiol.*, 37 : 107—122.
- Berg van den Jw. 1956. Direct and indirect determination of the mean subglottic pressure. *Folia Phoniatrica*, 8 : 1—24.
- Berg van den Jw., J. T. Zantema and P. Doornenbal. 1957. On the air resistance and the Bernoulli effect of the human larynx. *Journ. Acoust. Soc. Am.*, 29 : 626—631.
- Bergeijk van W. A. and E. E. David. 1959. Delayed hand-writing. *Percept. and Motor Skills*, 9 : 347—357.
- Black J. W. 1951. The effect of delayed side-tone upon vocal rate and intensity. *Journ. Speech and Hearing Disorders*, 16 : 56—60.
- Blankinship W. A. 1962. Method of time normalization for speech. *Journ. Acoust. Soc. Am.*, 34 : 242—243.
- Butler R. A. and F. T. Galloway. 1957. Factorial analysis of the delayed speech feedback phenomenon. *Journ. Acoust. Soc. Am.*, 29 : 632—635.
- Chang H. T., T. C. Ruch and A. A. Ward. 1947. Topographical representation of muscles in motor cortex of monkeys. *Journ. Neurophysiol.*, 10 : 39—56.
- Chase R. A. 1958. Effect of delayed auditory feedback on the repetition of speech sounds. *Journ. Speech and Hearing Disorders*, 23 : 583—590.
- Chase R. A., S. Sutton, D. First. 1959. Bibliography: delayed auditory feedback. *Journ. Speech and Hearing Res.*, 2 : 193—200.
- Chase R. A., G. Guilfoyle. 1962. Effect of simultaneous delayed and undelayed auditory feedback on speech. *Journ. Speech and Hearing Res.*, 5 : 144—151.
- Chase R. A., S. Harvey, S. Standfast, J. Rapin and S. Sutton. 1961. Studies on sensory feed-back. I. Effect of delayed auditory feed-back on speech and keytapping. *Quart. Journ. Exper. Psychol.*, 13 : 141—152.
- Cherry E. C. 1953. Some experiments on the recognition of speech with one and with two ears. *Journ. Acoust. Soc. Am.*, 25 : 975—979.
- Chocholle R. 1942a. Variation des temps de réaction auditifs avec l'intensité. *C. R. Soc. Biol.*, 136, 7—8 : 334—335.
- Chocholle R. 1942b. Variation des temps de réaction auditifs avec la fréquence. *C. R. Soc. Biol.*, 136, 9—10 : 372—373.
- Cooper F. S., P. C. Delattre, A. M. Liberman, J. M. Borst and L. J. Gerstman. 1952. Some experiments on the perception of synthetic speech sounds. *Journ. Acoust. Soc. Am.*, 24 : 597—606.
- Crabbe H. J. F. 1960. Recognition and control of speech. *Automation Progress*, 5, 1 : 15—17.
- Creelman C. D. 1962. Human discrimination of auditory duration. *Journ. Acoust. Soc. Am.*, 34 : 582—593.
- Delattre P. 1962. Some factors of vowel duration and their crosslinguistic validity. *Journ. Acoust. Soc. Am.*, 34 : 1141—1143.
- Delattre P. C., A. M. Liberman and F. S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journ. Acoust. Soc. Am.*, 27 : 769—773.
- Denes P. 1955. Effect of duration on the perception of voicing. *Journ. Acoust. Soc. Am.*, 27 : 761—764.
- Denes P. and M. V. Matthews. 1960. Spoken digit recognition using time-frequency pattern matching. *Journ. Acoust. Soc. Am.*, 32 : 1450—1455.
- Dieth E. 1950. *Vademekum der Phonetik*. Bern.

- Draper M. H., P. Ladefoged and D. Whitteridge. 1959. Respiratory muscles in speech. *Journ. Speech and Hearing Res.*, 2: 16-27.
- Dreyfus-Graf J. 1950. Sonograph and sound mechanics. *Journ. Acoust. Soc. Am.*, 22: 731-739.
- Dreyfus-Graf J. 1962. The phonetograph and the quantization of the sound waves. *Speech Communication Seminar*. Stockholm.
- Eccles J. C. 1939. The spinal cord and reflex action. *Ann. Rev. Physiol.*, 1: 363-384.
- Faaborg-Andersen K. 1957. Electromyographic investigation of intrinsic laryngeal muscles in humans. *Acta Physiol. Scandinav.*, Suppl. 140.
- Fairbanks G. 1954. A theory of the speech mechanism as a servosystem. *Journ. Speech and Hearing Disorders*, 19: 133.
- Fairbanks G. 1955. Selective vocal effects of delayed auditory feedback. *Journ. Speech and Hearing Disorders*, 20: 333-345.
- Fairbanks G., N. Guttman. 1958. Effects of delayed auditory feedback upon articulation. *Journ. Speech and Hearing Res.*, 1, 1: 12-22.
- Fant G. 1960. Acoustic theory of speech production with calculations based on x-ray studies of Russian articulations. The Hague.
- Fischer-Jørgensen E. 1963. Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck. *Phonet.*, 16, 1-3: 19-36.
- Flanagan J. L. 1958. Some properties of the glottal sound source. *Journ. Speech and Hearing Res.*, 1: 99-116.
- Flanagan J. L. 1959. Estimates of intraglottal pressure during phonation. *Journ. Speech and Hearing Res.*, 2: 168-172.
- Floyd W. F., V. E. Negus and E. Neil. 1957. Observations on the mechanism of phonation. *Acta Otolaryngologica*, 48, 1-2: 16-25.
- Fónagy I. and K. Magdics. 1960. Speed of utterance in phrases of different lengths. *Language and Speech*, 3: 179-192.
- Forgie J. W. and C. D. Forgie. 1959. Results obtained from a vowel recognition computer program. *Journ. Acoust. Soc. Am.*, 31: 1480-1489.
- Foulkes J. D. 1961. Computer identification of vowel types. *Journ. Acoust. Soc. Am.*, 33: 7-11.
- Fry D. B. and P. Denes. 1958. An analog of the speech recognition process. *Sympos. on Mashinization of Thought Process*, 1: 376-386.
- Fujimura O. 1961a. Effects of vowel context on the articulation of stop consonants. *Programm of 61. Meeting of Acoust. Soc. of Amer.*: 19.
- Fujimura O. 1961b. Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. *Journ. Speech and Hearing Res.*, 4: 233-247.
- Halle M. 1959. The sound pattern of Russian. 's-Gravenhage.
- Halle M., G. W. Hughes and I. P. A. Radley. 1957. Acoustic properties of stop consonants. *Journ. Acoust. Soc. Am.*, 29: 107-116.
- Harris C. M. 1953. A study of the building blocks in speech. *Journ. Acoust. Soc. Am.*, 25: 962-969.
- Harris K. S. 1958. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1: 1-7.
- Harris K. S., M. M. Schvey and G. F. Lysaught. 1962. Component gestures in the production of oral and nasal labial stops. *Journ. Acoust. Soc. Am.*, 34: 743.
- Heffner R. M. S. 1940. A note on vowel length in American speech. *Language*, 16, 1: 33-47.
- Heinz J. M. and K. N. Stevens. 1961. On the properties of voiceless fricative consonants. *Journ. Acoust. Soc. Am.*, 33: 589-596.

- Henry F. M. 1948. Discrimination of the duration of a sound. *Journ. Exper. Psychol.*, 38 : 734—743.
- Highlema n W. H. 1962. Linear decision functions with application to pattern recognition. *Proc. IRE*, 50 : 1567.
- Hildebrandt B. 1961. Die arithmetische Bestimmung der durativen Funktion. *Ztschr. f. Phonetik*, 14 : 328—336.
- Hoshiko M. S. 1960. Sequence of action of breathing muscles during speech. *Journ. Speech and Hearing Res.*, 3 : 291—297.
- Horridge G. A. 1961. The organisation of the primitive central nervous system as suggested by examples of inhibition and the structure of neuropile. «Nervous inhibition», ed. by E. Florey : 395—410.
- House A. S. and G. Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journ. Acoust. Soc. Am.*, 25 : 105—113.
- Hudgins C. V. and R. H. Stetson. 1937. Relative speed of articulatory movements. *Arch. néerl. phonét. expér.*, 13 : 85—95.
- Hughes G. W. and M. Halle. 1956. Spectral properties of fricative consonants. *Journ. Acoust. Soc. Am.*, 28 : 303—312.
- Hunt C. C. 1955. Temporal fluctuation in excitability of spinal motoneurons and its influence on monosynaptic reflex response. *Journ. Gener. Physiol.*, 38 : 801—811.
- Hunt C. C. and E. R. Perl. 1960. Spinal reflex mechanisms concerned with skeletal muscle. *Physiol. Rev.*, 40 : 538—579.
- Isshiki N. 1961. Voice and subglottic pressure. *Studia Phonologica*, 1 : 86—94.
- Jakobson R., G. L. M. Fant and M. Halle. 1952. Preliminaries to speech analysis. The distinctive features and their correlates. *M. I. T. Acoust. Lab. Techn. Rep.*, 13.
- Jakobson R. and M. Halle. 1956. *Fundamentals of language*. 's-Gravenhage.
- Jespersen O. 1904. *Phonetische Grundfragen*. Leipzig.
- Jespersen O. 1926. *Lehrbuch der Phonetik*. Leipzig.
- Jones D. M. A. 1909. *The pronunciation of English*. Cambridge.
- Kaiser L. 1929. A few remarks concerning occlusives in Russian. *Arch. néerl. phonét. expér.*, 4 : 77—80.
- Kaiser L. 1934. Some properties of speech muscles and the influence thereof on language. *Arch. néerl. phonét. expér.*, 10 : 121—133.
- Kaiser L. 1939. Biological and statistical research concerning the speech of 216 Dutch students. *Arch. néerl. phonét. expér.*, 15 : 1—76.
- Kaiser L. (Editor). 1957. *Manual of phonetics*. Amsterdam.
- Kaplan H. M. 1960. *Anatomy and physiology of speech*. N. Y.
- Katsuki Y. 1950. The function of phonatory muscles. *Jap. Journ. Physiol.*, 1 : 29—36.
- Kelly J., J. K. Anthony and E. Uldall. 1962. Tempo and transitions. *Speech Communication Seminar*. Stockholm.
- Küpfmüller K. 1961. Regelungsvorgänge beim Sprechen und Hören. *Proc. 3rd Internat. Congr. on Acoust.*, 1 : 170—187.
- Ladefoged P. 1957. Use of palatography. *Journ. Speech and Hearing Disorders*, 22 : 764—774.
- Ladefoged P. 1958. The perception of speech. *Sympos. on Machinization of Thought Processes*, 1 : 399—408.
- Ladefoged P. and N. P. McKinney. 1963. Loudness, sound pressure and subglottal pressure in speech. *Journ. Acoust. Soc. Am.*, 35 : 454—460.
- Lee B. S. 1950. Effects of delayed speech feedback. *Journ. Acoust. Soc. Am.*, 22 : 824—826.
- Lee B. S. 1951. Artificial stutter. *Journ. Speech a. Hearing Disorders*, 16 : 53.

- Lehiste I. and G. E. Peterson. 1961. Transitions, glides and diphthongs. *Journ. Acoust. Soc. Am.*, 33 : 268—277.
- Liberman A. M. 1957. Some results of research on speech perception. *Journ. Acoust. Soc. Am.*, 29 : 117—123.
- Liberman A. M., F. S. Cooper, K. S. Harris and P. F. MacNeilage. 1962. A motor theory of speech perception. *Speech Communication Seminar*. Stockholm.
- Liberman A. M., P. Delattre and F. S. Cooper. 1952. The role of selected stimulus — variables in the perception of the unvoiced stop consonants. *Amer. Journ. Psychol.*, 65 : 497—516.
- Liberman A. M., P. Delattre, L. J. Gerstman and F. S. Cooper. 1956. Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journ. Exper. Psychol.*, 52 : 127—137.
- Liberman A. M., K. S. Harris, H. S. Hoffman and B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journ. Exper. Psychol.*, 54 : 358—368.
- Liberman A. M., F. Ingeman, L. Lisker, P. Delattre and F. S. Cooper. 1959. Minimal rules for synthesizing speech. *Journ. Acoust. Soc. Am.*, 31 : 1490—1499.
- Licklider J. C. R. 1952. On the process of speech perception. *Journ. Acoust. Soc. Am.*, 24 : 590—594.
- Lisker L. 1957. Closure duration and the intervocalic voiced—voiceless distinction in English. *Language*, 33, 1 : 42—49.
- Ludwig H. and J. E. Hind. 1955. Electronic switch eliminates transients. *Electronics*, 28, 10 : 163—165.
- Lysaught H., R. J. Rosov and K. S. Harris. 1961. Electromyography as a speech research technique with an application to labial stops. *Journ. Acoust. Soc. Am.*, 33 : 842.
- Lullies H. 1953. *Physiologie der Stimme und Sprache*. In *Lehrbuch der Physiologie*.
- MacNeilage P. F. 1963. Electromyographic and acoustical study of the production of certain final clusters. *Journ. Acoust. Soc. Am.*, 35 : 461—463.
- Malécot A. 1955. An experimental study of force of articulation. *Studia Linguistica*, 9, 1 : 35—44.
- Malécot A. 1956. Acoustic cues for nasal consonants. *Language*, 32 : 274—284.
- Malécot A. 1958. The role of releases in the identification of released final stops. *Language*, 34 : 370—380.
- Marill T. 1961. Automatic recognition of speech. *IRE Trans. on Human Factors in Electronics*, HFE-2, 1 : 34—38.
- Mártony J. 1962. Some experiments on perceptual cues for Swedish fricatives. *Speech Communication Seminar*. Stockholm.
- Menzerath P. und A. de Lacerda. 1933. *Koarticulation. Steuerung und Lautabgrenzung*. Berlin.
- Miller G. A. 1956. The magical number seven, plus or minus two: some limits in our capacity for processing information. *Psychol. Rev.*, 63 : 81—97.
- Miller G. A. 1962. Decision units in the perception of speech. *IRE Trans. on Information Theory*, 8 : 81—83.
- Miller G. A., E. Galanter and K. H. Pribram. 1960. *Plans and the structure of behavior*. N. Y.
- Miller G. A., G. A. Heise and W. Lichten. 1951. The intelligibility of speech as a function of the context of the test materials. *Journ. Exper. Psychol.*, 41 : 329—335.
- Miller G. A. and P. E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journ. Acoust. Soc. Am.*, 27 : 338—352.

- Miller R. L. 1953. Auditory tests with synthetic vowels. *Journ. Acoust. Soc. Am.*, 25 : 114—121.
- Moll K. L. 1960. Cinefluorographic techniques in speech research. *Journ. Speech and Hearing Res.*, 3 : 227—241.
- Moore P. and H. von Leden. 1958. Dynamic variations of the vibratory pattern in the normal larynx. *Folia Phoniatrica*, 10 : 205—238.
- Nakata K. 1959. Synthesis and perception of nasal consonants. *Journ. Acoust. Soc. Am.*, 31 : 661—666.
- Öhman S. E. G. 1962. Perception of gated Swedish consonants in vc and cv syllables. *Speech Communication Seminar*. Stockholm.
- Olson H. F. and H. Belar. 1956. Phonetic typewriter. *Journ. Acoust. Soc. Am.*, 28 : 1072—1081.
- Olson H. F. and H. Belar. 1961. Phonetic typewriter III. *Journ. Acoust. Soc. Am.*, 33 : 1610—1615.
- Passy P. 1887. *Les sons du français*. Paris.
- Peters R. W. 1954. The effect of changes in side tone delay and level upon rate of oral reading of normal speakers. *Journ. Speech and Hearing Disorders*, 19 : 483—490.
- Peterson G. E. 1952. The information-bearing elements of speech. *Journ. Acoust. Soc. Am.*, 24 : 629—637.
- Petrick S. R. and H. M. Willett. 1960. Digital automatic word recognition procedure. *Journ. Acoust. Soc. Am.*, 32 : 1517.
- Pollack I. and L. Ficks. 1954. Information of elementary multidimensional auditory displays. *Journ. Acoust. Soc. Am.*, 26 : 155—158.
- Potter R. K., G. A. Kopp and H. C. Green. 1947. *Visible speech*. N. Y.
- Potter R. K. and J. C. Steinberg. 1950. Toward the specification of speech. *Journ. Acoust. Soc. Am.*, 22 : 807—820.
- Raab D. H. 1962. Effect of stimulation-duration on auditory reaction-time. *Amer. Journ. Psychol.*, 75 : 298—301.
- Rall W. and C. C. Hunt. 1956. Analysis of reflex variability in terms of partially correlated excitability fluctuation in a population of motoneurons. *Journ. gen. Physiol.*, 39 : 397—422.
- Rao P. and N. A. Nayak. 1956. Auditory reaction time as a function of frequency of the signal tone. *Current Sci.*, 25, 8 : 255.
- Rousselot P. J. 1901—1908. *Principles de phonétique expérimentale*. 1, 2, Paris.
- Ruhm H. B., W. A. Cooper. 1962. Low sensation level effects of pure-tone delayed auditory feedback. *Journ. Speech and Hearing Res.*, 5 : 185—193.
- Sakai T., S. Doshita, K. Nagata and T. Sekimoto. 1962. *Phonetic typewriter*. *Speech Communication Seminar*. Stockholm.
- Saslow M. G. 1958. Reaction time to Consonant-vowel Syllables in ensembles of various sizes. *Rep. Res. Lab. Electronics. MIT.* July : 143—144.
- Savin H. B. 1963. Word-frequency effect and errors in the perception of speech. *Journ. Acoust. Soc. Am.*, 35 : 200—206.
- Schatz C. D. 1954. The role of context in the perception of stops. *Language*, 30 : 47—56.
- Sebestyen E. 1961. Recognition of membership of classes. *IRE Transactions on Information Theory*, 7 : 44—50.
- Sebestyen E. 1962. Pattern recognition by an adaptive process of sample set construction. *IRE Transactions on Information Theory*, 8 : 82—91.
- Shimizu K. 1961. Experimental studies on movements of the vocal cords during phonation by high voltage radiograph motion pictures. *Studia phonologica*, 1 : 111—116.
- Sivertsen E. 1961. Segment inventories for speech synthesis. *Language and Speech*, 4 : 27—89.

- Small A. M. and T. R. Campbell. 1962. Temporal differential sensitivity for auditory stimuli. *Amer. Journ. Psychol.*, 75 : 401-411.
- Smith C. P. 1951. A phoneme detector. *Journ. Acoust. Soc. Am.*, 23 : 446-451.
- Spilka B. 1954. Some vocal effects of different reading passages and time delays in speech feedback. *Journ. Speech and Hearing Disorders*, 19 : 37-47.
- Spuehler H. E. 1962. Delayed sidetone and auditory flutter. *Journ. Speech and Hearing Res.*, 5 : 124-132.
- Steinbuch K. 1958. Automatische Spracherkennung. *NTZ-Nachr. tech. Ztschr.*, 11 : 446-454.
- Stetson R. H. 1951. *Motor phonetics*. Amsterdam.
- Stetson R. H., C. V. Hudgins and E. R. Moses. 1940. Palatograms change with rates of articulation. *Arch. néerl. phonét. expér.*, 16 : 52-62.
- Stevens K. N. 1960. Toward a model for speech recognition. *Journ. Acoust. Soc. Am.*, 32 : 47-55.
- Stevens K. N. and A. S. House. 1955. Development of a quantitative description of vowel articulation. *Journ. Acoust. Soc. Am.*, 27 : 484-493.
- Stevens K. N. and A. S. House. 1956. Studies of formant transitions using a vocal tract analog. *Journ. Acoust. Soc. Am.*, 28 : 578-585.
- Stevens K. N., A. S. House. 1961. An acoustical theory of vowel production and some of its implications. *Journ. Speech and Hearing Res.*, 4 : 303-320.
- Stevens K. N., A. S. House. 1963. Perturbation of vowel articulations by consonantal context: an acoustical study. *Journ. Speech and Hearing Res.*, 6 : 111-128.
- Stott L. H. 1935. Time-order errors in the discrimination of short tonal durations. *Journ. Exper. Psychol.*, 18 : 741-766.
- Stowe A. N., W. P. Harris and D. B. Hampton. 1963. Signal and context components of word-recognition behavior. *Journ. Acoust. Soc. Am.*, 35 : 639-644.
- Sumby W. H., D. Chambliss and I. Pollack. 1958. Information transmission with elementary auditory displays. *Journ. Acoust. Soc. Am.*, 30 : 425-429.
- Tarnóczy T. H. 1951. The opening time and opening-quotient of the vocal cords during phonation. *Journ. Acoust. Soc. Am.*, 23 : 42-44.
- Timcke R., H. von Leden and P. Moore. 1958. Laryngeal vibrations: measurements of the glottic wave. Part I. The normal vibratory cycle. *Arch. of Otolaryngology*, 68 : 1-19.
- Timcke R., H. von Leden and P. Moore. 1959. Laryngeal vibrations: measurements of the glottic wave. Part II. Physiologic variations. *Arch. of Otolaryngology*, 69 : 438-444.
- Truby H. M. 1959. Acoustico-cineradiographic analysis considerations with especial reference to certain consonantal complexes. *Acta Radiologica*. Suppl. 182.
- Welch P. and R. S. Wimpless. 1961. Two multivariate statistical computer programs and their application to the vowel recognition problem. *Journ. Acoust. Soc. Am.*, 33 : 426-434.
- Wiren J. and H. Stubbbs. 1956. Electronic binary selection system for phoneme classification. *Journ. Acoust. Soc. Am.*, 28 : 1082-1091.
- Witting C. 1953. New techniques of palatography. *Studia Linguistica*, 7, 1 : 54-68.
- Yngve V. H. 1960. A model and an hypothesis for language structure. *Proc. Amer. Philosoph. Soc.*, 104 : 444-466.
- Zimmerman S. and S. M. Sapon. 1958. Note on vowel duration seen cross-linguistically. *Journ. Acoust. Soc. Am.*, 30 : 152-153.

СОКРАЩЕНИЯ

для обозначения регистрируемых
артикуляторных и акустических явлений

<i>в.н.</i>	— ток воздуха из носовых отверстий, регистрируемый при помощи термоанемометра.
<i>в.н. руп.</i>	— ток воздуха из носовых отверстий, регистрируемый при помощи рупорного датчика.
<i>в.р.</i>	— ток воздуха из ротового отверстия, регистрируемый при помощи термоанемометра.
<i>в.р. руп.</i>	— ток воздуха из ротового отверстия, регистрируемый при помощи рупорного датчика.
<i>г.г.</i>	— губная смычка, регистрируемая при помощи контактного датчика.
<i>давл.</i>	— давление в полости рта.
<i>дых.</i>	— дыхательные движения грудной клетки.
<i>к.1, к.2, к.3. . . к.19</i>	— касание языком соответствующих точек (см. рис. 2.15) датчика «искусственное нёбо» (эти же сокращения используются для обозначения расположенных в этих точках электродов).
<i>к.(1+2)</i>	— касание языком нёба по крайней мере в одной из точек <i>к.1</i> или <i>к.2</i>
<i>лар.</i>	— огибающая сигнала ларингофонного тракта.
<i>микро.</i>	— огибающая сигнала микрофонного тракта.
<i>н. чел.</i>	— движения нижней челюсти.
<i>огубл.</i>	— огубленность, т. е. «выпячивание» губ, характерное для губных гласных.
<i>отм. вр.</i>	— отметка времени 1 сек.

В качестве транскрипционных знаков при обозначении звуков русской речи используются буквы русского алфавита. Средняяязычный сонант обозначается знаком *ȷ*. Мягкость обозначается косым штрихом, расположенным сверху и справа от основного символа (например, *т'*).

О Г Л А В Л Е Н И Е

	Стр.
Г л а в а 1. Физиология речеобразования и вопросы автоматического распознавания и синтеза речи	3
Г л а в а 2. Методы исследования деятельности речеобразующего аппарата	10
1. Датчики потоков воздуха из ротового и носовых отверстий	13
2. Датчик внутриротового давления	18
3. Датчики для регистрации дыхательных движений	20
4. Методы кинематографической и фотоосциллографической регистрации артикуляторных движений губ	21
5. Электрические методы регистрации артикуляторных движений губ	27
6. Динамическая палатография	31
7. Микрофонный и ларингофонный тракты	44
8. Электронные устройства для обработки сигналов о деятельности речеобразующего аппарата	45
9. Электронные измерители времени, автоматизирующие обработку данных	51
10. Комплексная регистрация артикуляторных и акустических параметров речи	57
11. Основные возможности использования комплексной регистрации артикуляторных и акустических параметров речи	63
Г л а в а 3. Организация временной программы синтагмы	78
1. Определение понятия артикуляторной программы. Синтагма как единица артикуляции	78
2. Влияние темпа речи на длительность различных элементов синтагмы	85
3. Флюктуация длительности слогов и более крупных отрезков синтагмы	97
4. Влияние фонетического состава на временные характеристики слова (синтагмы)	109
Г л а в а 4. Организация слога	122
1. Слог типа СГ	122
2. Сложные слоговые конструкции. Исследование движений огубленности	126
3. Сложные слоговые конструкции. Временные отношения внутри комплекса согласных	132
4. Нарушения артикуляции в условиях задержанной обратной слуховой связи	140

Глава 5. Взаимоотношения между элементами артикуляторных комплексов	157
1. Элементы артикуляции и дифференциальные признаки	157
2. Движения губ и нижней челюсти при артикуляции губных смычных согласных	165
3. Временные отношения между смычкой и фонацией при артикуляции глухих смычных согласных	174
Глава 6. Исследования процесса восприятия речи	182
1. Состояние вопроса о восприятии речи	185
2. Фонема как первичная единица решений	198
3. Процедура приятия фонемных решений	203
4. Роль ритмической структуры сообщения в распознавании смысловых единиц	215
Глава 7. О возможных процедурах автоматического распознавания речи	224
Литература	229
Сокращения	239

РЕЧЬ. АРТИКУЛЯЦИЯ И ВОСПРИЯТИЕ

*Утверждено к печати
Институтом физиологии им. И. П. Павлова
Академии наук СССР*

Редактор издательства Г. С. Катинас
Художник Д. А. Андреев
Технический редактор М. Е. Зендель
Корректор Н. П. Яковлева

Сдано в набор 12/IX 1964 г. Подписано к печати 15/XII
1964 г. РИСО АН СССР № 2-92В. Формат бумаги 60×
×90^{1/16}. Бум. л. 7^{3/8}. Печ. л. 15^{1/4}=15^{1/4} усл. печ. л.
Уч.-изд. л. 15.93. Изд. № 2284. Тип. вак. № 944. М-25196.

Тираж 2500. ТП 1964 г. № 942. Цена 1 р. 22 к.

Ленинградское отделение издательства «Наука»
Ленинград, В-164, Менделеевская лин., д. 1

1-я тип. издательства «Наука», Ленинград, В-34, 9 лин., д. 12

ИЗДАТЕЛЬСТВО «НАУКА»

ГОТОВИТ К ПЕЧАТИ НОВЫЕ КНИГИ:

Бионика. (Научный совет по комплексной проблеме «Кибернетика»). 43 л. Цена 3 р. 30 к.

Первая отечественная книга, освещающая различные проблемы новой науки — бионики. Рассматриваются проблемы опознавания образов и построения самообучающихся машин, моделирование нервных клеток, проблемы аэро- и гидродинамики биологических систем, биологические и медицинские аспекты бионики.

Бжалава И. Т. Психология и кибернетика. 18 л. Цена 1 р. 50 к.

Книга посвящена интереснейшей проблеме связи психологии и кибернетики. Изучаются принципы, на основе которых следует проводить аналогию между поведением организма и машины, дается детальное описание фиксированной установки на деятельность, раз-
бираются модели восприятия, памяти, мышления и т. д.

Биологическая электроника и кибернетика. (Научный совет по комплексной проблеме «Кибернетика»). 25 л. Цена 1 р. 90 к.

Биологическое значение и функциональная детерминация миграционного поведения животных. 6 л. Цена 42 к.

Достижения современной физиологии нервной и мышечной системы. 12 л. Цена 84 к.

Проблемы современной нейрофизиологии. 12 л. Цена 84 к.

Сложные формы поведения. 20 л. Цена 1 р. 50 к.

Вопросы физиологии интероцепции, вып. 2. Труды Института физиологии им. И. П. Павлова, т. XI. 40 л. Цена 3 р.

Л. К. Чередниченко. Физиологическая калориметрия. 9 л. Цена 60 к.

Ладыгина-Котс Н. Н. Предпосылки человеческого мышления. 20 л. Цена 1 р. 40 к.

Книга посвящена вопросу о предыстории возникновения человеческого сознания. Это богато иллюстрированное экспериментальное исследование.

ИМЕЕТСЯ В ПРОДАЖЕ КНИГА

От простого к сложному. Элементы развития высшей нервной деятельности ребенка. (Экспериментальные исследования). 1964. 224 стр. Цена 1 р.

В книге излагаются результаты исследований становления различных сторон нервной деятельности детей в возрасте от нескольких часов жизни до нескольких лет.

ЗАКАЗЫ НА КНИГИ ИЗДАТЕЛЬСТВА «НАУКА» ПРОСИМ НАПРАВЛЯТЬ ПО АДРЕСУ:

Москва, К-12, Б. Черкасский пер., 2/10, Контора «Академкнига»,
отдел «Книга — почтой»; Ленинград, Д-120, Литейный пр., 57,
Ленотделение «Академкнига», отдел «Книга — почтой»

**ЗАКАЗЫ НА КНИГИ ВЫПОЛНЯЮТСЯ НАЛОЖЕННЫМ
ПЛАТЕЖОМ**

