

354

Г. ФАНТ

АКУСТИЧЕСКАЯ ТЕОРИЯ РЕЧЕОБРАЗОВАНИЯ

ACOUSTIC THEORY OF SPEECH PRODUCTION

WITH CALCULATIONS BASED ON X-RAY STUDIES
OF RUSSIAN ARTICULATIONS

BY

GUNNAR FANT

Royal Institute of Technology Stockholm

MOUTON & CO. · 'S-GRAVENHAGE
1960

Г. ФАНТ

АКУСТИЧЕСКАЯ ТЕОРИЯ РЕЧЕОБРАЗОВАНИЯ

ПЕРЕВОД С АНГЛИЙСКОГО
Л. А. ВАРШАВСКОГО и В. И. МЕДВЕДЕВА

ПОД РЕДАКЦИЕЙ
В. С. ГРИГОРЬЕВА

ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА 1964

534
Ф 22
УДК 534.78

АННОТАЦИЯ

Монография известного шведского акустика Г. Фанта посвящена вопросам теории речеобразования, экспериментального изучения процессов речевой артикуляции, анализа звуков речи и их синтеза с помощью аналоговых вычислительных устройств. В книге устанавливаются связи между артикуляцией и акустическими характеристиками речи; рассмотрение вопросов проводится с физиологической, акустической и лингвистической точек зрения.

Книга может быть полезна для специалистов-акустиков, инженеров связи, физиологов, интересующихся вопросами речеобразования, и лингвистов.

ГЛАВНАЯ РЕДАКЦИЯ ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ

Г. Фант

Акустическая теория речеобразования

М., 1964 г., 284 стр. с илл.

Редактор В. А. Григорова.

Техн. редактор Л. В. Лихачева.

Корректор О. А. Сигал.

Сдано в набор 4/XI 1963 г. Подписано к печати 14/V 1964 г. Бумага 70×108/16. Физ. печ. л. 17,75+11 вкл. Условн. печ. л. 26,20. Уч.-изд. л. 24,53. Тираж 3 000 экз. Цена книги 1 р. 42 к. Заказ № 1853.

Издательство «Наука».

Главная редакция физико-математической литературы.
Москва, В-71, Ленинский проспект, 15.

Ленинградская типография № 2 имени Евгении Соколовой «Главполиграфпрома»
Государственного комитета Совета Министров СССР по печати. Измайловский проспект, 29.

ОГЛАВЛЕНИЕ

От редактора перевода	7
Предисловие автора	8
Введение. Теоретико-информационные аспекты анализа речи	11
§ 1. Канал речевой коммуникации	11
§ 2. Речевое сообщение. Избыточность	12
§ 3. Понятие о емкости канала применительно к спектрографическому анализу речи	14
§ 4. Структура сигнала и промежуточные коды в процессе речевой коммуникации	17
§ 5. Акустические корреляты элементов сообщения	22

ЧАСТЬ ПЕРВАЯ

АКУСТИЧЕСКАЯ ТЕОРИЯ РЕЧИ

Глава 1. Общая теория	27
§ 1. Речь как процесс фильтрации	27
§ 2. Сегментация	34
§ 3. F-картина	37
Глава 2. Теория цепей, эквивалентных речевому тракту	39
§ 1. Эквивалентные схемы акустических резонаторов и рупоров	39
§ 2. Методы численного расчета	47
§ 3. Общие уравнения преобразования для математического описания речевых сигналов	53
А. Функция передачи речевого тракта	53
Б. Излучение, источник и другие постоянно действующие факторы	55
В. Полное выражение преобразования Лапласа	57
Г. Обратное преобразование Лапласа	58
Глава 3. Аналитические связи в структуре речевых спектров	59
§ 1. Идеализированное спектральное описание сонорных звуков	59
§ 2. Связь между частотами формант и огибающей спектра	64
§ 3. Представление спектра согласных с помощью нулей и полюсов	71
Глава 4. F-картины сложных трубчатых резонаторов и рупоров	73
§ 1. Резонатор, состоящий из двух отрезков труб. Влияние огрубления	73
§ 2. Рупоры как резонаторы и связующие элементы системы	77
§ 3. Приближенное представление речевого тракта трехпараметрической моделью	80
А. Модели, состоящие только из цилиндрических отрезков труб	82
Б. Модели, в которых язычная область речевого тракта представлена рупором	88

ЧАСТЬ ВТОРАЯ

РАСЧЕТЫ, ОСНОВАННЫЕ НА РЕНТГЕНОГРАФИЧЕСКИХ ДАННЫХ

Глава 5. Методика рентгеновского исследования, испытуемое лицо и фонетический материал	99
Глава 6. Методы исследования и способы аппроксимации	100

Глава 7. Исследование гласных	109
§ 1. Расчетное определение частот формант и огибающих спектра	109
§ 2. Артикуляторные и акустические диаграммы гласных	113
§ 3. Соотношения между размерами резонаторов и формантными частотами	115
§ 4. Пространственное распределение звукового давления. Ширина полосы формант	128
А. Расчеты уровней формант	128
Б. Распределение звукового давления в речевом тракте	131
В. Зависимость ширины форманты от различных активных элементов в речевом тракте	135
Глава 8. Носовые звуки и назализация	139
§ 1. Физиологические данные	139
§ 2. Носовые звуки, образуемые при закрытом рте	143
§ 3. Назализация	147
Глава 9. Плавные	159
Глава 10. Щелевые, аффрикаты и взрывные	164
§ 1. Щелевые и аффрикаты	164
§ 2. Взрывные	179
§ 3. Идеализированные модели фрикативных и взрывных	184
§ 4. Заключение относительно характеристик источника для фрикативных и взрывных	195

ЧАСТЬ ТРЕТЬЯ

ВЫВОДЫ

Глава 11. Сегментация речи и определение ее структурных элементов	199
Глава 12. Связи между F -картиной и артикуляцией	201
Глава 13. Некоторые аспекты теории дифференциальных признаков	203
Глава 14. Замечания об акустической природе дифференциальных признаков	206

ПРИЛОЖЕНИЯ

I. Анализ речевой волны	217
§ 1. Измерение интенсивности	217
§ 2. Измерение спектра и формы волны	223
§ 3. Спектрограммы речевого материала, использованного для сравнения с результатами расчета согласных	230
II. Характеристики источника	231
§ 1. Голосовой источник	231
§ 2. Турбулентный и импульсный источники	239
III. Аналитическое исследование простых моделей резонаторов применительно к механизму речеобразования	247
§ 1. Одиночный резонатор Гельмгольца	247
А. Одно отверстие	247
Б. Два отверстия	249
§ 2. Сдвоенный резонатор Гельмгольца	250
§ 3. Труба постоянного сечения как акустический резонатор	256
§ 4. Системы из четырех отрезков труб. Уравнения преобразования для произвольного расположения источника	263
§ 5. Демпфирующее влияние потерь в последовательных и параллельных элементах резонатора, состоящего из двух отрезков труб	266
§ 6. Сводка расчетных формул для определения затухания в резонаторах из двух отрезков труб; приложение к модели речевого тракта	269
А. Резонатор Гельмгольца	270
Б. Стоячие волны в трубах	273
В. Эксперименты и расчеты, относящиеся к резонаторам с жесткими стенками, состоящим из одного и двух отрезков труб	275
Литература	278

ОТ РЕДАКТОРА ПЕРЕВОДА

Развитие автоматике, кибернетики и теории информации потребовало гораздо более глубокого изучения речевых сигналов, чем это было необходимо для решения сравнительно ограниченного круга чисто лингвистических вопросов. Если раньше традиционная фонетика основное внимание уделяла тому, как образуются и артикулируются звуки речи, то в настоящее время не меньший интерес представляют те особенности акустической картины речевого процесса, которые являются основой изучения восприятия речи и распознавания речевых образов. Поэтому естественно стремление установить связи между механизмом речеобразования и акустическими характеристиками речи, поскольку именно последние являются основным объектом восприятия.

Попытки установить эти связи, в особенности с появлением и развитием измерительной техники, позволяющей изучать достаточно детально акустическую картину речи, предпринимались неоднократно (работы Крендала, Дэнна, Чиба и Каджияма и др.); однако при этом допускался ряд упрощений, не всегда достаточно обоснованных. Книга Г. Фанта в значительной степени свободна от этого недостатка. Вопрос об акустической картине звуков речи и связи ее с характеристиками речевого аппарата рассматривается подробно и всесторонне. При этом наряду с полным решением, оказавшимся возможным при использовании электронной вычислительной техники, рассмотрен и ряд упрощенных моделей речевого тракта и определена степень применимости их для получения акустических характеристик звуков речи. Рассмотрена теория речеобразования, методы изучения речевых сигналов и расчета звукопроводов сложной конфигурации, а также ряд вопросов лингвистического истолкования акустической картины звуков речи.

Все основные расчеты проведены на материале русской речи; к русской речи относятся и основные экспериментальные результаты.

При подготовке перевода книги Г. Фанта было сочтено целесообразным дополнить ее вводной главой, заимствованной из первоначального гектографированного издания (1958 г.) и не включенной в основное издание. В ней автор рассматривает вопросы анализа речи и структуры речевого сигнала с точки зрения теории информации. Эта глава будет

представлять интерес для широкого круга читателей, не знакомых с этим аспектом теории речевых процессов.

Большая ценность книги Г. Фанта заключается в том, что помимо оригинального материала в ней приведено много данных, которые до сих пор были разбросаны в многочисленных журнальных публикациях. Приведенная в книге обширная библиография, относящаяся непосредственно к материалу книги и к ряду смежных вопросов, будет весьма полезна для читателя.

Обширный круг вопросов, рассмотренных в книге, делает ее в равной мере полезной для акустиков, инженеров связи, физиологов и лингвистов.

В. С. Григорьев

ПРЕДИСЛОВИЕ АВТОРА

Настоящая монография имеет целью уточнение и улучшение понимания связей между физиологическими данными, относящимися к речеобразованию, и физическими данными, характеризующими речевой сигнал. Часть этой работы имеет общий характер и включает сводку теоретических основ, используемых для описания речевых сигналов и для предсказания акустической картины этих сигналов на основании артикуляторных данных.

Часть I посвящена общей теории речеобразования и методам расчета, причем особое внимание обращено на акустику сложных резонансных систем; рассмотрение последних проводится с использованием эквивалентных схем. Эта часть дает теоретическую основу для исследований, изложенных в части II, а также для теории систем, состоящих из простых резонаторов, и теории источников звука, участвующих в образовании речи; оба эти вопроса рассмотрены в Приложениях.

В части II приведены результаты расчетов, основанных на рентгенографических данных; последние были получены для всех типичных русских гласных и согласных в 1951 г. А. С. Макмилланом и Г. Келеменом в Массачусетской больнице по глазным и ушным болезням при Гарвардской медицинской школе. Эта работа входила в план исследований, проводившихся проф. Р. Якобсоном из Гарвардского университета, проф. М. Халле из Массачусетского технологического института (MIT) и автором; рентгенографическое исследование артикуляции и основанные на нем расчеты являлись частью работы «Описание и анализ современной русской речи», проводившейся под руководством Р. Якобсона при содействии Отделения славянских языков и литературы Гарвардского университета. Эта работа субсидировалась фондом Рокфеллера.

Основной целью расчетов являлась попытка восстановления спектров звуков речи по физиологическим данным, относящимся к их образованию. Это потребовало сравнения спектров звуков в связной речи с соответствующими данными, полученными для протяжных форм артикуляции испытуемого лица во время рентгено съемки. Дополнительно сделаны некоторые критические замечания, относящиеся к традиционному описанию артикуляции соответствующих звуков.

Часть III представляет собой сводку соотношений между артикуляцией и акустической картиной с приложением к теории дифференциальных признаков. Первая глава этой части, являясь дополнением к началу части I, имеет вводный характер и более лингвистическую направленность.

Расчеты заняли длительное время, что связано отчасти с большой их сложностью. Вычисления производились на начальном этапе «вручную», и потому для их выполнения в практически приемлемые сроки пришлось прибегать к довольно грубым приближениям. На следующем этапе были использованы вычислительные машины, такие, как быстродействующая цифровая машина BESK в Стокгольме, позволившая выполнить более детальные расчеты в короткое время. Для гласных такие расчеты были выполнены в 1953—1954 гг.; расчеты для согласных были произведены в 1954—1957 гг. на аналоге речевого тракта в Стокгольмской лаборатории связи Отделения телефонии и телеграфии при Стокгольмском королевском технологическом институте (RIT); для носовых звуков расчеты несколько задержались из-за недостатка физиологических данных.

Проводившиеся независимо исследования в Массачузетском технологическом институте (MIT) и в RIT дали сходные результаты.

По окончании расчетной работы стало ясно, что рентгенографические данные и полученные по ним размеры речевого тракта не обладают той степенью точности и полноты, которая была бы желательна с точки зрения использования потенциальных возможностей современных вычислительных машин. Для детального изучения речи как артикуляторного события необходимо использовать скоростную кино-рентгено съемку и усовершенствовать методику измерения всех размеров речевого тракта.

Акустическую теорию гласных можно в настоящее время считать достаточно законченной. Что же касается согласных, то изучение даже их общих свойств еще далеко от завершения. В относящихся к ним расчетах были сделаны большие упрощения как исходных физиологических данных, так и теоретических предпосылок расчета. Несмотря на это, представляется, что результаты выполненных расчетов полезны с точки зрения критической оценки принятых приближений и дают представление о возможностях акустической теории речеобразования в ее современном состоянии.

Автор высоко ценит стимулирующие беседы с проф. Р. Якобсоном и проф. М. Халле, а также значительный труд, который они затратили на первоначальное изучение рентгенографического материала. Тесный контакт с руководимой профессором К. Н. Стивенсом группой по исследованию речи в Акустической лаборатории Массачусетского технологического института оказал существенное влияние на разработку методов перехода от артикуляторных данных к акустическим. Расчеты для носовых согласных стали возможными в результате определения размеров соответственных полостей, которое было выполнено д-ром Г. Бьюггреном из Саббатсбергской больницы в Стокгольме.

Автор также выражает большую признательность проф. Т. Лауренту из Королевского технологического института в Стокгольме за неизменную поддержку им настоящей работы и ценные советы.

Ценную редакторскую помощь по изданию оказал д-р В. Ясsem из Института техники польской Академии наук, Познань. Автор признателен также лектору Копенгагенского университета Эли Фишер-Йоргенсен, д-ру Г. Л. Фланагану из лаборатории фирмы Белл, проф. М. Халле и д-ру Г. М. Траби (RIT) за критические замечания по форме и содержанию предварительного гектографированного издания (1958). Автор приносит также благодарность М. Рихтеру, С. Фелицетти и другим сотрудникам лаборатории телефонии RIT за их работу по подготовке настоящей книги к выпуску в свет.

Проведение работы стало возможным благодаря финансовой поддержке ее фондом Рокфеллера, фондом Велленберга, Шведским государственным советом технических исследований, Шведским исследовательским институтом национальной обороны, а также Военно-воздушными силами США по договору AF 61 (514)-1084.

ВВЕДЕНИЕ

ТЕОРЕТИКО-ИНФОРМАЦИОННЫЕ АСПЕКТЫ АНАЛИЗА РЕЧИ

§ 1. Канал речевой коммуникации

Речевая коммуникация представляет собою процесс¹⁾ передачи информации от говорящего к слушающему. В описании этого процесса можно выделить ряд последовательных этапов:

- 1) Сообщение, подлежащее передаче.
- 2) Двигательная нервная активность, управляющая органами речи.
- 3) Положение и движения органов речи.
- 4) Акустические свойства источников голоса и системы резонаторов, образуемых полостями речевого аппарата.
- 5) Форма речевой волны, т. е. временная зависимость звукового давления в звуковом поле на некотором расстоянии от говорящего.
- 6) Механические процессы в улитке.
- 7) Нервная активность в проводящих путях слухового анализатора.
- 8) Сообщение, принятое слушающим.

Перечень этих этапов можно детализировать, разбив каждый из них на более мелкие. Например, на этапах 2 и 7 могло бы оказаться удобным для определенных целей рассматривать отдельно активность мозга и нервных проводящих путей. Слуховая и кинестетическая цепи обратной связи, позволяющие говорящему контролировать свою речь, здесь не рассматриваются; в этом нет необходимости при изучении структуры сигнала на каждом из указанных этапов. Некоторые этапы могут быть и укрупнены, если, например, объединить два или более из этапов 2—7.

Изучение речевой коммуникации имеет два аспекта: аналитический и статистический.

При аналитическом подходе должны быть определены последовательные этапы, дано общее описание сигнала на этих этапах и установлены причинные связи, действующие при преобразовании сигналов от этапа к этапу, начиная с речевого сообщения.

При втором подходе система и происходящие в ней процессы принимаются как нечто данное и проводится статистический анализ структуры сигнала, а также определяется вероятность появления различных сообщений; такой подход соответствует теоретико-информационному аспекту анализа речи. Он дает хорошую основу для оценки

¹⁾ Этот процесс был подробно рассмотрен Г. Е. Петерсоном [163, 165, 166]. (Список литературы см. в конце книги. — *Прим. ред.*)

эффективности различных методов описания речевой коммуникации, однако во многих случаях не может быть непосредственно использован за отсутствием необходимых количественных данных. Тем не менее ряд концепций теории информации может быть полезен при построении общей теории речевой коммуникации, и потому мы кратко рассмотрим их в настоящей работе.

§ 2. Речевое сообщение. Избыточность

Само сообщение составляет содержание начального и конечного этапа передачи информации¹⁾. Сообщение вводится в канал коммуникации на передающем его конце и сообщение же является конечным результатом передачи по каналу. Каждое сообщение представляет собой последовательный выбор одного символа из некоторого числа возможных. Количество символов конечно, и предполагается, что правила их использования согласованы между «приемником» и «передатчиком». Только в том случае, когда символы, представляющие собой элементы сообщения, определены, возможно применить теорию информации. Смысл сообщения или его значимость для данного слушателя, вообще говоря, из рассмотрения исключается. Правда, делались попытки расширить понятия теории информации таким образом, чтобы включить и семантическую сторону сообщений; см., например, [22].

В большинстве приложений теории информации к изучению языка, например в работе Шэннона [181], в качестве минимальных элементов использовались буквы. При изучении речевой коммуникации буквы заменяются фонемами, которые могут быть объединены в более крупные комплексы, а именно слоги, морфемы или слова, в зависимости от конкретных задач анализа. Однако во многих отношениях целесообразно выделять в фонеме более мелкие элементы — дифференциальные признаки²⁾. Вследствие повторяемости для разных фонем одних и тех же признаков число минимальных элементов оказывается уменьшенным. Подобное формальное описание речи должно быть, однако, дополнено признаками, определяющими относительную значимость и взаимосвязь слов во фразе, а также отношение к ним говорящего. Такие признаки экспрессивной стороны речи являются почти «табу» для теории информации; они могли бы быть охвачены этой теорией, только если бы могло быть определено конечное число возможных комбинаций подобных признаков, имеющих одинаковое значение как для говорящего, так и для слушающего.

На первом этапе процесса коммуникации мы имеем дело с источником информации. Основной оценкой его возможностей является число N различных используемых символов, без учета частоты их встречаемости. В этом случае количество информации на символ максимально и определяется как

$$H_m = \log_2 N \text{ дв. ед./символ.} \quad (B.1)$$

¹⁾ Основные понятия теории информации были введены Шэнноном и могут быть найдены у Шэннона и Уивера [182], Г. А. Миллера [152], Черри [22].

²⁾ Теория дифференциальных признаков, дающая описание сигналов на определенном этапе коммуникации, была развита Р. Якобсоном; см. [111, 112, 138, 114, 81, 115, 84]. Теоретико-информационный анализ речи на основе дифференциальных признаков как минимальных элементов был проведен Черри, Халле и Якобсоном [23].

Эта величина характеризует количество создаваемой источником информации в том частном случае, когда вероятности всех символов одинаковы и отсутствуют ограничения, связанные с порядком их следования. Однако в общем случае некоторые символы появляются чаще, чем другие, и вероятности появления их в контексте зависят от предшествующих и последующих символов. Вследствие этого среднее количество информации на символ H_a оказывается меньше H_m .

Одно из определений H_a , предложенное Шэнноном [182], выражается формулой

$$H_a = -\frac{1}{n} \sum p_{ni} \log_2 p_{ni} \text{ дв. ед./символ}, \quad (\text{B.2})$$

где p_{ni} — вероятности появления групп символов по n символов в группе. При этом предполагается, что n выбрано достаточно большим для того, чтобы можно было не считаться со взаимными влияниями одних групп на другие. Формулу (B.2) можно интерпретировать как среднее по тексту большого объема количество информации $\log_2(p_{ni}^{-1})$, приходящейся на один символ. Наименее часто встречающиеся символы несут, таким образом, наибольшую информацию.

После того как определено H_a , вводится определение избыточности R , выражающееся равенством

$$R = 1 - \frac{H_a}{H_m}. \quad (\text{B.3})$$

Если назвать H_a/H_m относительной информацией, то этим равенством избыточность определяется как 1 минус относительная информация.

Избыточность, таким образом, является мерой того, какая, в среднем, часть передаваемых источником символов не существенна для идентификации, или декодирования, сообщения. Например, если источник информации располагает тремя равновероятными символами A, B и C , но известно, что после появления B должно появиться C и, наоборот, что C может появиться только после появления B , то один из символов B или C является избыточным и потому может быть отброшен. В этом простом случае избыточность равна $1/3$.

Вероятность появления какой-либо буквы в письменном тексте в значительной мере зависит от предшествующей буквы и очень мало зависит от букв, удаленных от данной промежутком в $n = 10$ или более букв. Минимальная оценка Шэнноном [181] величины H_a близка к 1; это значит, что на каждый символ связного текста приходится количество информации, соответствующее одному двоичному выбору. При алфавите, содержащем двадцать семь букв (включая пробелы), максимальная информация была бы $H_m = \log 27 = 4,8$ дв. ед., так что избыточность составляет примерно 75%. Подобного рода расчеты для устной речи, если принять за символ фонему, дали бы несколько меньшую избыточность.

Информация, полученная каким-либо звеном коммуникационной цепи, не может быть, согласно определению, больше, чем переданная предшествующим звеном, но может быть меньше из-за шума или других помех в цепи. В частности, при восприятии сообщения человеком неизбежные отклонения в работе его механизма восприятия от идеального могут рассматриваться как шум. Передача сообщения при помощи кода, обладающего известной избыточностью, меньше подвержена вредному влиянию шума, но зато требует большего времени. Оба

эти противоречивых требования — краткости передачи, с одной стороны, и уверенности в отсутствии ошибок приема, с другой, — весьма существенны. Известно, например, что некоторая избыточность в формулировках в письме или при чтении лекции весьма желательна из педагогических соображений.

§ 3. Понятие о емкости канала применительно к спектрографическому анализу речи

Структура сигнала на каждом из этапов 2—7 может изучаться и оцениваться с трех точек зрения:

1) описания сигнала на данном этапе безотносительно к самому сообщению или к характеристикам сигналов на других этапах процесса коммуникации;

2) статистического и, если это возможно, аналитического изучения соотношений между характеристиками сигнала на каком-либо данном этапе и на остальных, безотносительно к сообщению;

3) изучения соотношений между элементами сообщения на этапе 1 и характеристиками сигнала на всех или некоторых из этапов 2—8.

Процесс описания сигналов (1) на каждом данном этапе может рассматриваться как сообщение, являющееся источником информации на этом этапе для внешнего наблюдателя. При этом важно оценить, в какой мере информационная емкость средств наблюдения согласована с информационной производительностью источника. Очевидно, что сигналы на этапе 5, имеющие форму звуковых волн, более удобны для наблюдения, чем на других этапах, так как вся структура сигнала здесь определяется одной временной функцией, будь то осциллограмма или спектральная картина, которые могут быть проанализированы с высокой степенью точности. При этом не теряются и коррелятивные связи, необходимые для исследований, проводимых в соответствии с задачами (2) и (3).

Значение тех ограничений точности, которые неизбежны при измерениях, относящихся к звуковой волне, может быть оценено при помощи формулы Шэннона для емкости канала. Передающая система с общей шириной полосы W гц при средней мощности сигнала S и средней мощности шума N может передавать информацию с предельной скоростью, которая определяется емкостью канала и равна

$$C = W \log_2 \left(1 + \frac{S}{N} \right) \text{ дв. ед./сек.} \quad (\text{B.4})$$

Хотя такая скорость и может быть достигнута, однако в реальных условиях, если необходимо устранить возможность ошибок, требующая для этого весьма сложная система кодирования привела бы к бесконечно большому запаздыванию. Приведенная формула может быть использована при рассмотрении как осциллографических, так и спектральных данных анализа речи.

Анализатор спектра, состоящий из полосных фильтров с шириной полосы пропускания B гц и охватывающий общий диапазон частот W гц, обладает емкостью, определяющейся формулой (B.4); эту емкость можно получить, просуммировав емкости W/B частотных каналов, примыкающих друг к другу. Каждый фильтр может опробоваться B раз в секунду. При большей частоте проб последние перестают быть достаточно независимыми, а при меньшей — теряется часть информации, содержа-

щейся в речевом сигнале. В анализаторах, подобных «Сонаграфу»¹⁾, расстояние между средними частотами соседних полос значительно меньше их ширины. Это создает графическую избыточность в получающейся картине, но не увеличивает информационной емкости.

В принципе каждая проба дает сведения об амплитуде и фазе, так же как каждая гармоника ряда Фурье. Однако в большинстве анализаторов фазовая информация не используется, поскольку она не существенна для фонетических целей; см. далее по этому поводу Приложение II, § 2.

Формула Шэннона (В.4), определяющая емкость канала, может быть интерпретирована следующим образом. Статистическая неопределенность амплитуды, обусловленная шумом и случайными ошибками измерения, кладет практический предел точности ее определения. Этот предел можно найти, введя квантованную шкалу с $(1 + S/N)^{1/2}$ возможными градациями амплитуды. Предполагая, что берется W проб, измеряемых по этой шкале, найдем, что количество различных комбинаций или, иначе говоря, равновероятных сигналов длительностью в 1 сек равно $(1 + S/N)^{W/2}$; логарифм этой величины по основанию 2 равен $\frac{1}{2} W \log_2(1 + S/N)$, что составляет половину той емкости канала, которая требовалась бы при передаче также и фазовой информации. Подобный прием квантования был впервые предложен Хартли [92] и развит далее Туллером [205]; см. также [136].

Амплитуды сигнала и шума обычно выражаются в виде отношений в логарифмической шкале, в которой за единицу принимается децибел (дб). Формула емкости канала может быть при этом представлена в следующем приближенном виде:

$$C = \frac{WD}{3} \text{ дв. ед./сек.} \quad (\text{В.5})$$

где D — отношение сигнал/шум,

$$D = 10 \log_{10} \frac{S}{N} \text{ дб.} \quad (\text{В.6})$$

Спектрограф с частотным диапазоном, охватывающим $W = 10\,000$ гц, и с точностью измерения амплитуды 1 дб при динамическом диапазоне 30 дб, т. е. при $D = 30$, имеет при измерении амплитуды информационную емкость, равную

$$C = \frac{1}{2} W \frac{D}{3} = 50\,000 \text{ дв. ед./сек.}$$

Полное использование всего этого количества данных означает описание сигнала при помощи W значений амплитуды в секунду, равномерно распределенных в плоскости время — частота. Каждому такому значению соответствует элементарная площадка в этой плоскости, определяющаяся полосой частот B гц и длительностью $1/B$ сек.

¹⁾ Ввиду отсутствия достаточно установившейся терминологии мы будем в дальнейшем называть звуковым спектрографом или просто спектрографом такого рода анализатор спектра, который дает спектральную картину в координатах частота — интенсивность — время. Саму спектральную картину при этом будем называть спектрограммой, а чтобы не смешивать ее со спектром в координатах частота — интенсивность, последний будет называться «спектральным разрезом». «Сонаграф» — фирменное наименование одного из типов спектрографа. (Прим. ред.)

Приведенная выше величина информационной емкости спектрографа, $S = 50\,000$ дв. ед./сек., характерна для всех систем передачи и записи речи, обладающих средним качеством. Она примерно в 1000 раз больше, чем поток информации в речевых сообщениях, считая, что скорость произнесения речи составляет около 10 фонем в секунду и что количество используемых при этом фонем около 32, т. е. 5 дв. ед./символ. Этот расчет показывает, насколько велико несоответствие или, применяя технический термин, «несогласованность» [44] между количеством информации, содержащимся в речи, и емкостью систем, обычно используемых для ее передачи.

Возникающие вследствие такой несогласованности проблемы являются общими как для экспериментальной фонетики, так и для техники связи, поскольку описание и кодирование сигналов представляют собой, по существу, вполне аналогичные операции с теоретико-информационной точки зрения; действительно, и в том и в другом случае целью является исчерпывающее описание речевого процесса. Наличие указанной выше большой несогласованности определяет заинтересованность инженеров связи в исследованиях речи, которые позволили бы осуществить разработку систем телефонной связи со значительно меньшей полосой частот, чем используемая обычно.

Фонетист, который поставил бы перед собой задачу исчерпывающего описания речи с использованием всей информационной емкости применяемого им спектрографа, столкнулся бы с астрономически большим числом возможных картин, подлежащих рассмотрению, и с невозможностью из-за этого решить поставленную задачу. Трудность заключается не в получении достаточного количества спектрографических данных, а в использовании этих данных для обеспечения максимально точного описания фонетических явлений. Использовать всю массу получаемых данных оказывается невозможным даже для самых детальных фонетических исследований. Возможность решения указанных задач определяется тем, что количество информации, содержащейся в акустической картине речи, не тождественно емкости каналов, используемых для ее передачи или исследования. В речевых сигналах имеют место упорядоченные соотношения как в частотной, так и во временной областях, обуславливающие избыточность в акустической картине. Наиболее отчетливо эта упорядоченность проявляется в тенденции к частотно-избирательной концентрации энергии, а также в квазистационарном характере спектральных картин. Возникающие вследствие этого связи будут более подробно рассмотрены далее. Наличие таких связей и приводит к тому, что действительная информация в акустической картине речи, определяемая безотносительно к коммуникативной функции речи, значительно меньше 50 000 дв. ед./сек., хотя и намного больше указанной выше цифры 50 дв. ед./сек. Точно эта величина не может быть измерена, но можно полагать, что высококачественную передачу речи можно осуществить в системах синтетической телефонии¹⁾, имеющих емкость 1000 дв. ед./сек.

Максимально возможное исключение избыточности в описании речевого сигнала необходимо не только для разработки подобных систем, но и для построения программ исследования речи для фонетических целей.

¹⁾ Формантные вокодеры описаны в работах [130, 1, 62, 64, 20, 168, 3]. Системы вокодеров достаточно установившегося типа описаны в работах [87, 36, 37, 206].

§ 4. Структура сигнала и промежуточные коды в процессе речевой коммуникации

Большая часть сведений о речи, которыми мы располагаем в настоящее время, охватывает вопросы речеобразования, т. е. этапы 3 и 4 указанного выше перечня. Разделение на этапы, содержащие только физиологические и только акустические данные, полезно с точки зрения более четкой классификации существующих сведений о речеобразовании.

Данные этапа 5 аналитически связаны с данными этапа 4 формулой, устанавливающей зависимость свойств речевой волны, например ее спектральной картины, от свойств источника и фильтра, входящих в состав речевого тракта; из этой формулы следует, что источник и фильтр однозначно определяют форму речевой волны. Разделение речеобразующего аппарата на источник и фильтр не является ни в коей мере новым в фонетике; достаточно напомнить, например, о принятом в фонетике различении звонких и глухих согласных по наличию голосового или шумового источника.

Частотно-избирательная передача через речевой тракт, обусловленная резонансными свойствами входящих в его состав полостей, определяет фильтровую функцию речевого аппарата. Эта функция, характеризующая структуру сигнала на этапе 4, может быть определена заданием размеров полостей артикуляторного аппарата, эквивалентной схемой системы этих полостей или коэффициентом передачи такой системы. Каждое из подобных представлений можно рассматривать как промежуточный этап в определении фильтровой функции.

Положения и движения артикуляторных органов на этапе 3 определяют размеры резонирующих полостей на этапе 4. Деятельность дыхательных органов и мышц гортани (этап 3) определяют свойства голосового источника при сонорных¹⁾ звуках на этапе 4, в основном независимо от артикуляции. В отличие от этого, для глухих согласных свойства источника определяются как дыхательными органами, так и положением органов, осуществляющих артикуляцию.

Классическое описание артикуляции звуков речи, т. е. описание положения артикуляторных органов при произнесении данного звука, относится, таким образом, к этапу 3. По данным такого описания могут быть сделаны некоторые выводы о конфигурации полостей речевого тракта, необходимые для этапа 4. Однако для полноты описания должны быть заданы еще характеристики источника, хотя некоторые заключения о характере источника могут быть получены, если известен способ артикуляции.

Средства, которыми мы в настоящее время располагаем для детального изучения речеобразования и исследования структуры сигнала, с теоретико-информационной точки зрения неудовлетворительны. Техника кинорентгена пригодна для физиологических исследований, но позволяет получить подробные данные только для медиального сечения речевого аппарата. Значительно труднее получить поперечные его размеры, хотя и это возможно, если применить специальный рентгенотехнический метод — томографию. Палатограммы могут дать сведения,

¹⁾ Вполне установившихся терминов, соответствующих английским «voiced — unvoiced», у нас нет. «Voiced» переведены как «сонорные»; к ним отнесены те звуки речи, которые образуются при участии голоса, независимо от того, гласные это или согласные. «Звонкий — глухой» оставлено как характеристика только шумных согласных. (Прим. перев.)

представляющие общий интерес. Однако мы не располагаем методами непрерывного наблюдения за источниками звука при речеобразовании. Поэтому часть информации, необходимой для декодирования связной речи по данным кинорентгена, оказывается упущенной.

Представление о характере источников легко получить по осциллографическим записям речи. Если на осциллограмме наблюдается регулярная, квазипериодическая структура с основной частотой, лежащей в пределах 60—600 *гц*, можно сказать почти уверенно, что звук сонорный. Можно получить и более точное количественное описание характеристик источника, если скомпенсировать влияние на форму речевой волны фильтрующих свойств, присущих полостям речевого аппарата. Работы в этом направлении пока немногочисленны, но теоретически такая компенсация вполне возможна, если учесть данные, соответствующие этапу 4, и их простую связь с формой речевой волны.

Достаточно подробные данные о передаче речевых сигналов на этапах 2 и 7 были бы весьма существенны, однако сомнительно, что удалось бы получить достаточно исчерпывающее представление о структуре сигнала на этих этапах. Прохождение сигнала по нервным путям связано с исключительно большим числом разветвлений и соединений, характеризующихся кажущимся, хотя и не функциональным, беспорядком.

Картина механических колебаний в улитке на этапе 6 дает представление о структуре сигнала в более доступной форме, однако по сравнению со спектрограммой, относящейся к этапу 5, она вряд ли является более подходящей формой описания речевого процесса. Действительно, механическое частотно-пространственное преобразование, осуществляемое основной мембраной улитки, составляет только один из этапов частотного анализа, осуществляемого органом слуха.

Если при изучении передачи речевых сигналов обратиться к процессам в нервной системе, то необходимо было бы отдельно рассматривать нервную деятельность в интервалах отсутствия и наличия звука. При этом не все данные, получаемые внешним наблюдателем, были бы обусловлены речевыми сигналами — часть их связана с иными процессами в живом организме. Подобным же образом обстоит дело и при рассмотрении структуры сигнала в форме речевой волны: не все обнаруживаемые в ней детали характерны для передаваемой речи. Часть этих деталей обязана своим происхождением шуму в помещении, часть их может быть обусловлена такими особенностями речеобразования у данного диктора, которые находятся вне его контроля. Так, тонкая структура картины, дающей интенсивность в функции времени и частоты для фрикативных согласных и других глухих звуков, представляется на широкополосной спектрограмме в виде вертикальной штриховки. Положения этих штрихов случайны и, как правило, не связаны с движением речевых органов. Они попросту отражают статистические свойства турбулентного шумового источника.

Таким образом, структура речевого сигнала, если рассматривать его в форме звуковой волны, представляется наиболее сложной; многие детали звуковой волны не зависят от намерений диктора, а часть их теряется при преобразовании речевого стимула в нервный процесс.

Независимое от других этапов изучение данных, наблюдаемых на каком-либо одном этапе, имеет известное значение для описания сигнала. Однако если при таком изучении имеется в виду основное назначение речи — служить средством общения, то будет очевидна необхо-

димось учитывать корреляцию между наблюдаемыми на разных этапах данными, с тем чтобы отбросить те из них, которые не обусловлены предыдущими или не обуславливают последующие. Правда, такой процесс «очистки» может оказаться трудно осуществимым с достаточной полнотой в силу большой сложности наблюдаемых данных и форм их кодирования. Особенно велики эти трудности в том случае, если выбранные при изучении переменные, определяющие процесс, не будут связаны наиболее непосредственно с физическими или физиологическими механизмами кодирования.

На первоначальных стадиях исследования, когда исследователь еще не ознакомился со всем ансамблем данных о речевых сигналах и их вероятностях, количество возможных выборов максимально, т. е. максимальна получаемая при каждом наблюдении информация. Лишенные взаимосвязи детальные данные, пока не открыт их код, имеют столь же высокую энтропию, как и шум. При этом следует считаться с опасностью того, что наблюдатель не сможет закодировать наблюдаемые им факты в удобном для осмысления виде в силу недостаточной емкости осуществляющего это кодирование канала. Однако, если код установлен, избыточность может оказаться полезной.

Как уже отмечалось выше, избыточность в самой акустической картине речи весьма значительна, не говоря даже о тех связях, которые обусловлены коммуникативной функцией речи. Ряд взаимосвязей, характерных для человеческой речи, отчетливо выявляется на спектрограммах. Сюда относятся, например: наличие более или менее острых резонансных пиков; относительно небольшие или сравнительно медленные изменения спектральной картины в пределах длительности одного звука речи; наличие во многих случаях отчетливых границ между звуками; тенденция к непрерывному во времени переходу некоторых резонансов одного звука в резонансы соседних; часто наблюдаемое чередование во времени более слабых и более интенсивных звуков, а также звуков с гармонической и шумовой структурой, обусловленное слоговым строением речи; малая интенсивность составляющих, расположенных в нижней части спектра для глухих звуков, и т. п.

Задание частоты, ширины и интенсивности резонансов речевого тракта, предложенное Джузом [121], представляет собой, видимо, один из наиболее эффективных методов сжатого описания для определенного ограниченного класса спектральных картин, а именно картин, относящихся к сонорным звукам. Такое описание, обоснованное теоретической работой Уэйбела [207], представляет собой во многих случаях практическую ценность, однако не устраняет полностью избыточности, так как не учитывает связей, обусловленных аналитическими соотношениями между выбранными переменными. Частично эти связи могут быть установлены на основании статистических данных. Однако более отчетливую их формулировку дает математический анализ [49, 52, 66] общих свойств фильтровой функции и источника на этапе 4, которыми и определяется наблюдающаяся на спектрограммах картина.

Описание конфигурации полостей речевого аппарата (этап 4), вследствие ограниченных пока возможностей измерения поперечных сечений в зависимости от расстояния по осевой линии между голосовой щелью и губами, ни в коем случае нельзя считать исчерпывающим; причиной этого являются чисто технические трудности.

Следует отметить, что такое описание все же сохраняет большую избыточность. Действительно, достаточно задать несколько точек ука-

занной выше зависимости, называемой далее функцией площади, чтобы, в силу непрерывности этой функции по длине речевого тракта, предсказать более или менее точно промежуточные точки. Помимо этого, ряд вариантов конфигурации полостей может быть исключен заранее вследствие их невозможности, если не говорить о таких особых случаях, как чревовещание.

Классическая фонетика обходится при описании артикуляции в основном тремя независимыми переменными: положением языка вдоль оси речевого тракта, высотой подъема языка и степенью огубления. Как показали Стивенс и Хауз [194, 195] и как видно из приведенных далее в настоящей работе соображений, возможно приближенное описание функции площади полостей речевого тракта при помощи трех параметров, подобных параметрам, используемым фонетистами для описания артикуляции гласных. При этом, вместо непрерывной кривой, для описания артикуляции достаточно трех чисел, позволяющих определить фильтрующие свойства полостей речевого аппарата через частоты формант.

Описание речевого тракта при помощи функции площади также обладает избыточностью в силу существования различных артикуляторных конфигураций, дающих примерно одни и те же частоты формант. В фонетике этот факт известен под названием компенсаторных форм артикуляции. Правда, имеющиеся данные показывают, что индивидуально каждый человек имеет достаточно стабильные артикуляторные навыки и что они в основном такие же, как и у других субъектов той же языковой группы. Это не значит, что компенсаторные формы артикуляции не существуют; однако между двумя естественными или синтезированными звуками, образованными нормальным и компенсаторным путем, подобие является только приблизительным: в некоторых случаях может быть обнаружено различие в ширине частотной полосы формант; кроме того, вряд ли может иметь место полное совпадение частот более чем двух формант. Поэтому есть основания утверждать, что, располагая необходимыми сведениями о корреляционных связях между звуковой волной речи и артикуляцией, можно сделать вполне определенное заключение об артикуляторной конфигурации по данным об акустической картине речи.

Аналитические выражения связей являются далеко не простыми. Классическое представление речевого тракта в виде двух связанных между собой простых резонаторов отражает в некоторой мере действительные соотношения только в ограниченном числе случаев, а интерпретация соотношений между полостями и частотами формант, основанная на подобной модели, была или слишком упрощенной, или неправильной. Вообще говоря, каждая полость или часть речевого тракта оказывает определенное влияние на частоты всех формант [24, 121, 39, 49, 53, 196, 8, 195]. Существенное значение для понимания корреляционных связей имеет анализ, выполненный Делаттром [30].

Соотношения между звуковой волной речи и восприятием, с одной стороны, и артикуляцией и звуковой волной, с другой, подобны в том отношении, что дают только ограниченную возможность определения картины сигнала на одном этапе по его картине на следующем.

Реакция, конечно, определяется стимулом, однако различные стимулы могут давать приблизительно одну и ту же реакцию. Так, например, слушатели, тренированные в различении передних гласных по степени открытости, реагируют при этом на изменения частоты первой

форманты. Однако умеренное повышение этой частоты может быть до известной степени компенсировано повышением частоты более высокой форманты [164]. Это явление можно рассматривать как пример слуховой компенсации, аналогичной компенсации артикуляторной.

Возможность подобных случаев компенсации обусловлена в основном избыточностью кодов, используемых для передачи данных от одного этапа к другому. Если бы набор символов, которыми располагает приемник, был увеличен так, чтобы их количество не уступало таковому у передатчика, компенсация всегда оказывалась бы неполной. Так, в приведенном выше примере поставленную перед слушателем задачу определения фонетического качества по признаку открытости — закрытости можно было бы дополнить задачей качественной оценки по признаку мужской голос — женский голос. Аналогия с артикуляторной компенсацией при этом была бы полной.

Таким образом, звуки речи можно оценивать по их качественным звуковым атрибутам, вне связи с функциями их в качестве фонем. Однако ряд имеющихся данных позволяет считать ухо особенно чувствительным к тем изменениям в звуковом качестве, которые выполняют различительную функцию в языке слушающего [132].

В связи с этим неоднократно поднимался вопрос о том, происходит ли декодирование принятого сообщения пофонемно или же по более крупным логическим единицам порядка одного и даже нескольких слов [146]. Однако вопрос этот не очень существен с точки зрения теории информации. Возможно, что слушающий и не осознает при восприятии фонему как самостоятельную единицу, за исключением тех случаев, когда он принимает участие в артикуляционных испытаниях или когда он находится в такой естественной ситуации, при которой от этого осознания зависит правильность принятого сообщения. Тем не менее различительная функция фонемы, или, более общо, дифференциального признака, все же очевидна и проявляется в изменении принятого сообщения при замене одного минимального символа другим.

Наличие избыточности, обусловленной связями между последовательными элементами сообщения, т. е. связями между предшествующими и последующими событиями, наряду с наличием эффективной речевой памяти дает слушающему возможность отвлекаться от декодирования структурного содержания сообщения и воспринимать, например, индивидуальные особенности голоса говорящего.

В лабораторных экспериментах, имеющих целью изучение разрешающей способности слуха¹⁾ для таких простых стимулов, как чистые тоны, время, затрачиваемое на принятие каждого решения, велико, а ансамбль атрибутов сравнительно невелик. Поэтому совершенно неправильно подсчитывать информационную емкость слуховой системы по психоакустическим данным, относящимся к таким простым стимулам, как это делал Якобсон [110]. Для синусоидальных волн средней частоты и интенсивности минимальное различимое изменение частоты составляет 0,3%, а амплитуды — 0,4 дБ. Если разделить область слуховых восприятий на элементарные площадки со сторонами, равными соответственно дифференциальным порогам по высоте и по интенсивности, то, используя данные Стивенса и Дэвиса [198], можно найти, что число таких площадок будет 350 000. Однако при таком количестве сигналов трудно

¹⁾ Введение в эту область исследований можно найти у Стивенса и Дэвиса [198], Ликлайдера [134]; по вопросам восприятия речи см. [136].

предположить, что абсолютная характеристика каждого из них хранится в памяти; поэтому условия распознавания сложных спектральных картин, подобных речи, нельзя установить на основании определения дифференциальных порогов для чистых тонов.

Было показано, что свойственная слуху человека информационная емкость лучше всего используется в том случае, когда воздействующие стимулы обладают множеством различных качественных признаков с несколькими, предпочтительно бинарными, ступенями для каждого из них; см. [170—172]. Речь, если иметь в виду акустические корреляты фонетических категорий, как раз и представляет собой «многомерный» стимул, и потому структура его оптимальна с точки зрения восприятия. Это дает основание полагать, что скорость декодирования случайно распределенных во времени фонем, составляющих самое большее 20—50 дв. ед./сек., характеризует максимальную скорость восприятия. В то же время данные по величинам пороговых изменений структуры стимулов, подобных звукам речи, полезны для установления верхних пределов точности описания речевых сигналов, необходимой при решении технических или фонетических задач.

Изучение с этой точки зрения требований, которые следует предъявлять к системам синтетической телефонии, дало следующую оценку достаточной точности: для частот формант 3%, для частоты основного тона голоса 1%, для суммарной интенсивности 1 дб, для интенсивности первой форманты 1 дб, второй форманты 3 дб и третьей 5 дб [63, 65].

Практическую пригодность тех или иных количественных шкал, используемых при описании речевых сигналов, целесообразно определять на материале синтезированной связной речи, позволяющей осуществлять желательные изменения ее структуры, но в то же время сохраняющей в достаточной мере естественность.

Полезные данные могут дать опыты с использованием избирательных частотно-временных искажений естественной речи. Такие опыты, вероятно, покажут, что отрезок синтезированной или каким-либо методом преобразованной связной речи может восприниматься как полностью идентичный с оригиналом, даже если ряд деталей такого отрезка будет отличаться от оригинала на величину, превышающую приведенные выше дифференциальные пороги для изолированных звуков. Непосредственный ответ на вопрос о значимости отдельных деталей в картине сигнала могут дать эксперименты, показывающие, какие детали и в какой мере могут быть устранены до появления ошибок в восприятии отдельных фонем [153].

Существенно также, наряду с определением допустимых изменений в картине сигнала, найти инвариантные признаки¹⁾, необходимые для сохранения правильного фонематического значения символов сообщения.

§ 5. Акустические корреляты элементов сообщения

Установление акустических коррелятов элементов сообщения представляет собой задачу в некоторых отношениях более простую, чем максимально подробное фонетическое описание этих элементов. Однако

¹⁾ Ценные результаты в этом направлении были получены экспериментально в опытах с синтезированными звуками, проводившихся в лабораториях Хаскинс [25, 34, 33, 133, 132, 159].

для уверенности в том, что при анализе речевых явлений не будут упущены существенные признаки, методы обработки данных должны быть одинаковыми при решении как первой, так и второй задачи. Недостаточно эффективный прибор или слишком грубые приближения могут дезориентировать исследователя.

Формальное представление речевого сообщения в виде последовательности фонемных символов иногда противопоставляется кажущейся непрерывности потока речи; см., например, [98]. Прежние наблюдения, производившиеся при помощи кимографов [148], а также более современные исследования работы речевого аппарата при помощи кинорентгенов, выполненные, например, в лабораториях Хаскинса [139], как будто подтверждают представление о непрерывности речевого процесса. Однако подобные наблюдения страдают неполнотой. Действительно, кимограф обладает довольно ограниченной информационной емкостью, будучи подобен низкочастотному фильтру; кинорентген дает сведения об одной только стороне речеобразования, а именно об артикуляции.

Большее количество информации содержится в спектрограммах, получаемых методами «видимой речи»; в этих спектрограммах выявляются как непрерывные, так и скачкообразные изменения спектрального состава речи. Это дает естественное членение акустической картины речи на отрезки, соответствующие звуку речи или части звука. Как указывалось в § 4 и более подробно рассмотрено в § 2 гл. 1, границы таких отрезков определяются либо изменением источника звука, воздействующего на речевой тракт, либо изменением фильтроприятия некоторых фонем существенную роль играет информация, содержащаяся в акустической структуре соседних отрезков, примером которых являются значения переходных признаков для восприятия взрывных и носовых согласных, не противоречит общему представлению о речи как о серии дискретных, последовательных во времени звуковых явлений [83].

Представление о фонеме как о пучке дифференциальных признаков находит подтверждение в тех фонетических выводах, которые могут быть получены на основании спектрограмм. Последние не только дают возможность расчленить речь, но позволяют идентифицировать отдельные отрезки ее как отвечающие носовым согласным, щелевым согласным, латеральным согласным, глухим и звонким согласным, слабым и сильным взрывным и т. д.; именно этим обусловлена возможность «чтения» по картинам «видимой речи». Однако сформулировать с наименьшей избыточностью качественные характеристики того или иного класса звуков или признаков не всегда просто. Многие из этих характеристик взаимосвязаны; если перечислять их все, описание признака становится непомерно длинным, а если попытаться определить сущность признака в краткой формулировке, можно потерять часть информации, необходимой инженеру для разработки систем, осуществляющих идентификацию фонем.

При фонетическом подходе к изучению речи найти приемлемый компромисс между точностью и простотой формулировки еще более трудно, чем при фонематическом. Это можно иллюстрировать следующим примером: обычно считается, что для акустического описания гласных достаточно указать частоты первых двух формант F_1 и F_2 ; но гласный [а], произнесенный мужским голосом, может иметь те же F_1 и F_2 , что и гласный [э], произнесенный девочкой [175]. Если эти два гласных

тем не менее воспринимаются как разные фонемы, то это, как правило, обусловлено информацией, которая содержится в более высоких формантах и в основной частоте голоса F_0 . Как раз эта информация и отбрасывается, если рассматриваются только F_1 и F_2 ; F_3 имеет значение для правильного восприятия только передних гласных [34], но она может быть использована и для нормализации всех гласных, произнесенных разными дикторами [162, 49, 51, 53].

Однако возможно допустить, что два гласных оказались бы идентичными по их физическим признакам даже при самом тщательном анализе и тем не менее воспринимались бы как разные фонемы¹⁾. В нормальной ситуации, если фиксирована «система координат» для данного диктора и данного контекста, основой различения [а] от [э] всегда остается более высокая первая форманта F_1 в соответствии с обычным кодом данного языка. Однако более высокая F_1 как раз представляет собой показатель компактности гласных; наличие же последней является условием, необходимым и достаточным для идентификации. Из этого примера видна важность относительных, а не абсолютных характеристик звука.

Однако как для технических приложений, так и для фонетического описания подобных относительных признаков недостаточно; желательно выявить и сопутствующие им физические характеристики, которые могли бы быть использованы для механического распознавания речи; естественно, эти характеристики должны быть функционально связаны с указанными выше признаками гласного, или, иными словами, должны отражать то обстоятельство, что при сравнении двух гласных фонем основой различения их является признак компактности. Учитывая все это, типичную разницу между фонемой [а] и фонемой [э] можно сформулировать как более высокое положение и F_1 и F_2 в [а].

Если бы разумный, снабженный надлежащим измерительным оборудованием наблюдатель с Марса должен был исследовать наш фонемный код, то вполне возможно, что он мог бы выявить большинство наших фонемных границ, а также смог бы в общих чертах классифицировать различные типы найденных отрезков звуковой картины. После этого он, вероятно, мог бы предположить, что люди Земли располагают ограниченным числом слов того или иного языка, и это позволило бы ему сделать вывод, что при такой классификации нецелесообразно учитывать все различия в мелких деталях подобных отрезков. Это позволило бы приблизительно найти действительное число дифференциальных признаков и фонем, отнеся ряд особенностей тех или иных отрезков к влиянию контекста или индивидуальности говорящего, при небольшом числе основных символов сообщения.

При всем том такой наблюдатель не достиг бы существенного успеха, если бы он не был способен связать свои наблюдения с лингвистической ситуацией, в которой находятся говорящий и слушающий, что необходимо для установления критериев сходства и различия между

¹⁾ Предложенная Джюзом [121] теория, согласно которой формантная картина оценивается с учетом средних величин и пределов изменений формантных частот у данного диктора, находит подтверждение в опытах с синтезированной речью, выполненных Ладефогедом и Бродбентом [125]. Эти опыты обнаружили наличие фонематических сдвигов в интерпретации речевых звуков в том случае, если в предшествующем контексте был произведен сдвиг формантных частот. Как подчеркивалось Якобсоном и Халле [115, 114], термин «фонематическое перекрытие» не имеет смысла, если учитывать относительный характер фонем и дифференциальных признаков.

двумя отрезками записи. Он мог бы найти такие критерии, если бы каждый из участвующих в опыте многократно повторил один и тот же, по его мнению, отрывок речи или если бы каждый из участвующих повторил отрывки, произнесенные другими. В результате этого он смог бы отделить контекстуальные и индивидуальные характеристики от основных, используя обычную технику структурной лингвистики, и таким путем расшифровать код¹⁾. По-видимому, успех в решении этой задачи в значительной мере зависел бы от его собственного языка, или, более общо, от основных для него способов коммуникации, например от того, осуществляет ли он ее акустическими средствами или какими-либо иными. В том случае, если бы коммуникация осуществлялась им при помощи речи, положение его немногим отличалось бы от положения земного исследователя, изучающего неизвестный язык.

Наряду с исследованиями, имеющими в виду детальное акустическое описание речи, а также ее собственно лингвистический анализ, возможен и чисто инженерный подход к изучению речи. Такой подход сводится в конечном счете к попыткам эмпирически уловить инвариантные элементы речи, используя те или иные математические операции над статистическими данными по структуре сигнала; при этом предполагается, что это позволит получить небольшое число существенных параметров, учитывающих особенности всех категорий речевых звуков. Следует отметить, что успехи в разработке систем компрессии речи, основанных на непрерывной передаче данных о трех формантах для гласных и согласных и дополнительных данных об основной характеристике источника, были до сего времени довольно ограниченными²⁾. Отчасти это обусловлено недостаточным использованием реально существующей прерывности спектральной структуры речи, а следовательно и различий между основными типами акустических картин. Некоторые из систем синтетической телефонии являются примером другой крайности, когда переоценивается значение фонематических границ и недооценивается роль переходных характеристик [89, 124]. Тем не менее несомненно, что в основу любой системы компрессированной речевой связи должны быть положены известные приближения; то же самое можно сказать и о системе фонетического описания. Но и в том, и в другом случае система оптимальных параметров может быть установлена только на основе детального изучения и правильной оценки значения различных физических характеристик речи.

Вопрос о том, как некоторое минимальное различие на этапе, соответствующем сообщению, реализуется на этапе акустической картины, является одной из основных проблем, рассмотренных в работе Якобсона, Фанта и Халле [114].

Настоящее исследование задумано как шаг к установлению аналитических связей между рчеобразованием и акустической картиной, имея в виду как вопросы, связанные с понятием дифференциальных признаков, так и вообще физические характеристики звука. Если такие связи будут установлены, станет значительно легче выявить правильное соответствие между накопленными до настоящего времени знаниями об артикуляторных признаках речи, частично унаследованными от класси-

¹⁾ См. далее [91, 81] и дискуссию о критериях сходства — различия, инициированную Фишер-Йоргенсен [57].

²⁾ См., например, [69]. В последние годы эти и подобные им системы были значительно улучшены; см., например, [130, 103].

ческой фонетики, и признаками, наблюдаемыми в акустической картине речи. Обратные заключения от акустической картины звука речи к его образованию могут оказаться особенно полезными в тех случаях, когда на основании данных анализа речевой волны желательно получить представление о речи как артикуляторном событии; одним из таких случаев является физиологическая интерпретация спектрограмм.

Сравнение корреляционными методами данных наблюдения на двух этапах может дать ценные результаты, как показано в работе Делаттра [30]. Однако дополнительно к этому аналитический подход, направленный на отыскание причинных связей, неизбежен при рационально поставленном исследовании; он помогает отделить в акустической картине избыточные признаки от тех, которые непосредственно связаны с артикуляторными характеристиками. Более глубокое понимание акустических особенностей речеобразования полезно также с точки зрения интерпретации спектральных данных с наименьшей избыточностью, так как естественные связи в структуре сигнала могут быть в этом случае определены гораздо полнее и точнее, чем на основе только эмпирических данных.

ЧАСТЬ ПЕРВАЯ

АКУСТИЧЕСКАЯ ТЕОРИЯ РЕЧИ

ГЛАВА I

ОБЩАЯ ТЕОРИЯ

§ 1. Речь как процесс фильтрации

Речевая волна представляет собой результат воздействия одного или нескольких источников звука на фильтрующую систему речевого тракта. Это простое правило, выраженное в терминах акустики и электротехники, означает, что речевая волна однозначно определяется характеристиками источников звука и фильтров. Это утверждение, очевидно, охватывает также основные фонетические концепции речеобразования.

Теория речеобразования, основанная на представлениях об источниках звука и фильтрах, характеризуется блок-схемой, приведенной на

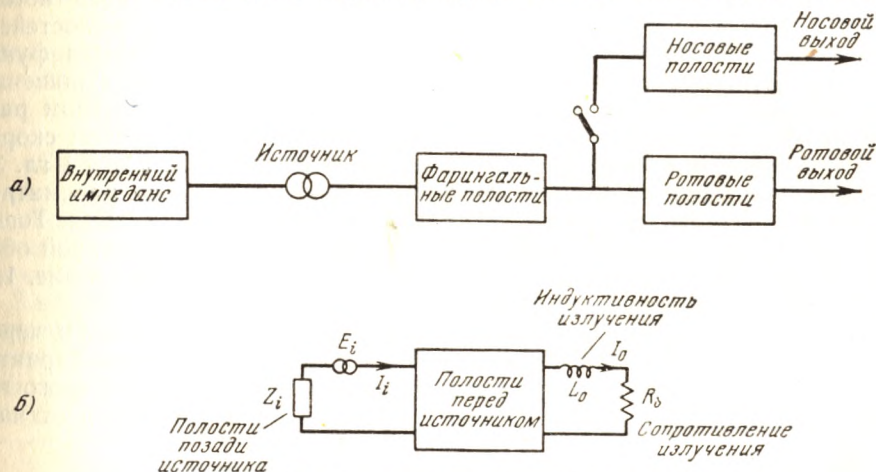


Рис. 1. Речеобразование как процесс фильтрации.

- а) Схематическое представление механизма образования звуков, создаваемых при участии голосового источника.
 б) Представление образования носовых звуков в виде четырехполюсника в общем случае, независимо от положения источника.

рис. 1. На ней показано несколько соединенных между собой фильтровых звеньев, каждое из которых представляет часть полостей речевого тракта. На рис. 1, а в качестве источника звука приняты голосовые связки. Носовая полость присоединена в точке схемы, соответствующей границе между фарингальной и ротовой областями тракта.

Подобные же блок-схемы, но только с двумя фильтровыми звеньями, представляющими соответственно переднюю и заднюю полости,

можно считать технической реализацией фонетических представлений. При этом сразу возникают трудности с определением истинных физиологических границ этих полостей. Можно считать область рта передней полостью, а область фаринкса — задней; можно разбить речевой тракт на две полости, исходя из их положения относительно источника звука; наконец, возможна еще одна разбивка (в случае гласных — дополнительная), при которой границей полостей считается область наибольшего сужения в речевом тракте. При фрикативных и взрывных звуках эта область расположена близко к источнику, но не обязательно совпадает с ним. Что касается гласных, то, в особенности для задних, место артикуляции, определяемое по наивысшей точке подъема языка, может оказаться далеко не совпадающим с местом наибольшего сужения, хотя с акустической точки зрения больше оснований именно его считать границей разделения всего тракта на заднюю и переднюю полости. В случае очень открытых гласных смысл подобного разделения для той части речевого тракта, которая лежит выше гортани, в значительной мере утрачивается; чем более открытым является гласный, тем менее обосновано рассмотрение отдельных частей речевого тракта как независимых резонаторов.

Схемы, подобные приведенной на рис. 1, б, используются при количественном рассмотрении речеобразования с использованием методов теории электрических цепей. Полости, расположенные перед ¹⁾ источником, представлены в этой схеме четырехполюсником, нагруженным импедансом излучения. Двухполюсник, расположенный позади источника, включает последовательно соединенные импеданс самого источника ²⁾ и импеданс всех задних по отношению к источнику полостей.

Полная фильтровая функция, в технике часто называемая функцией передачи, представляет собой частотную зависимость отношения двух величин: звукового давления в звуковом поле на известном расстоянии от рта говорящего и звукового давления или объемной скорости источника. При более детальном анализе (см. § 3 гл. 2 и § 1 гл. 3) частотно-зависимые свойства излучения ротового отверстия рассматриваются как самостоятельная составляющая фильтровой функции. Тогда сама система полостей речевого тракта характеризуется величиной объемной скорости на ее выходе, т. е. в проходе между губами; на рис. 1, б эта величина обозначена через I_0 .

Обозначив через S функцию, характеризующую источник, и через T — функцию, отображающую свойства фильтра, акустическую характеристику звука речи можно представить равенством: $P = ST$. Строго говоря, обе входящие сюда величины в общем случае зависят и от час-

¹⁾ Во избежание неясностей в дальнейшем следует отметить, что автор рассматривает положение различных элементов речевого аппарата, начиная от голосовой щели; таким образом, голосовая щель считается самым задним элементом, затем идут фаринкс, ротовая полость, ротовое отверстие. Однако при отсчете координаты по оси речевого тракта начало координат располагается, как правило (если особо не оговорено иное расположение начала), на выходе речевого тракта, как это часто делается при рассмотрении линий передачи. (Прим. ред.)

²⁾ Ток $i(t)$ ампер в электрической цепи соответствует объемной скорости $u(t)$ в $\text{см}^3/\text{сек}$ в акустической системе. Объемная скорость $u(t)$, равная $v(t)A$, представляет собой произведение скорости частиц $v(t)$ $\text{см}/\text{сек}$ и площади A см^2 поперечного сечения, перпендикулярного к направлению воздушного потока или к направлению колебаний. Напряжение $e(t)$ вольт соответствует давлению $p(t)$ $\text{дин}/\text{см}^2$. Отношение звукового давления при частоте f к объемной скорости при той же частоте представляет собой акустический импеданс $z = p(f)/u(f)$ для этой частоты.

тоты и от времени¹⁾. Имеется известная аналогия между фонетическим термином «фонация» и техническим «источник», так же как и между терминами «артикуляция» и «фильтр». Эта аналогия справедлива, если отличать фонацию от артикуляции, понимая под первой само создание звучания, тогда как артикуляцией формируется определенное фонетическое качество звука. Фильтровые свойства речевого тракта зависят от положения артикуляторов, что и позволяет считать понятия фильтра и артикуляции сопоставимыми, по крайней мере в том случае, если иметь в виду идеализированные модели речевого тракта. Это будет показано в гл. 4.

Говоря о речеобразовании, следует отметить, что источник S в формуле $P = ST$ представляет собой акустическое возмущение, наложенное на поток выдыхаемого воздуха; это возмущение вызывается либо препятствием в речевом тракте, обуславливающим наличие трения или внезапное открытие и закрытие прохода, либо, в случае сонорных звуков, квазипериодической модуляцией потока воздуха изменением ширины прохода между голосовыми связками. Таким образом, акустическому этапу фонации, к которому относится понятие источника, предшествует аэродинамический процесс выдоха (иногда вдоха) воздуха. Основными параметрами этого процесса являются объем выдыхаемого (вдыхаемого) воздуха в единицу времени и подсвязочное давление. Последнее не следует смешивать с акустическим понятием звукового давления, характеризующим источник.

Основным свойством голосового источника является периодичность создаваемого звука, которая определяется длительностью T_0 одного цикла работы голосовых связок; обратная ей величина представляет собой основную частоту голоса и равна

$$F_0 = \frac{1}{T_0}. \quad (1.2)$$

«Высота голоса» и «основная частота голоса» не являются синонимами, хотя и могут использоваться с одинаковым правом вследствие однозначного соответствия их друг другу. Строго говоря, высота есть ощущение, связанное с воздействием того или иного тона, а частота — физическое свойство звукового стимула.

Длительность цикла, которым определяется высота голоса, всегда несколько изменяется от периода к периоду. Частью эти изменения имеют систематический характер и связаны с интонационным рисунком речи, частью же представляют собой случайные или, точнее говоря, непреднамеренные колебания; однако эти колебания являются существен-

¹⁾ Частотную зависимость удобно выразить, используя преобразование Лапласа, как зависимость соответственных функций от комплексной частоты $s = \sigma + j\omega$:

$$P(s) = S(s) \cdot T(s), \quad (1.1)$$

т. е. преобразование Лапласа $P(s)$ от звукового давления в звуковом поле перед диктором является произведением соответственных преобразований для функции источника $S(s)$ и функции передачи $T(s)$.

В настоящей работе преобразование Лапласа широко используется; оно представляет собой удобный способ сжатой формулировки положений, относящихся к речеобразованию и структуре звуков речи. Преобразование Лапласа является более общим, чем преобразование Фурье, однако в ряде случаев последнее оказывается полезным как дополнительный способ трактовки акустических явлений. Так, в последующем весьма упрощенном описании образования сонорных звуков на рис. 2 использованы ряды Фурье, поскольку они широко применялись в фонетике со времен Гельмгольца. Зависимость от времени может быть отражена, если писать $P(s, t)$ вместо $P(s)$ и т. д.

ным признаком естественной человеческой речи. Только говорящая машина может обеспечить полное постоянство высоты звука, что обуславливает присущую такому постоянству монотонность.

Другой характеристикой голосового источника является огибающая спектра создаваемых им колебаний, т. е. зависимость $S(f)$ амплитуд составляющих спектра от их частоты. Огибающая определяется регистром голоса, основной частотой и громкостью речи, но отражает также и индивидуальные свойства голоса говорящего. В псевдонаучной литературе иногда ограничиваются только указанием числа обертонов в колебаниях, создаваемых голосовыми связками; однако этого недостаточно для правильного описания свойств голосового источника.

Одним из основных признаков классификации звуков речи в классической фонетике является характеристика источника. Термины «гармонический/негармонический спектр» или, во временном представлении, «периодический/непериодический звук» применяются соответственно к случаям наличия/отсутствия голосового источника. В апикальных или увулярных вибрантах также имеет место периодичность за счет периодических изменений фильтровой функции; в звонких вариантах основная энергия таких звуков обеспечивается голосовым источником. В глухих вариантах этих же звуков первичным является источник, расположенный в месте артикуляции, но частота колебаний не превосходит 30 гц, т. е. намного ниже возможных основных частот голосового источника. Для целей фонетической классификации термин «гармонический» следует применять только к звукам, образующимся с участием голосовых связок, исключая таким образом глухие вибрanty.

Приведенные термины «гармонический» или «периодический» не являются со строгой физической точки зрения достаточно правильными; всегда имеют место случайные изменения основной частоты колебаний голосовых связок, поэтому сонорные звуки можно рассматривать только как квазипериодические.

Термин «голос» используется нами и как характеристика категории источника, и как характеристика специфического вида звуковых колебаний. С точки зрения характера источника возможны следующие случаи:

- а) отсутствие источника (пауза);
- б) только голосовой источник;
- в) одновременно голосовой и шумовой источники;
- г) шумовой источник, один или несколько.

При разработке электрических синтезаторов речи «Водер» и «Вокодер» [36, 38], послуживших прототипами всех современных устройств для синтеза речи [37], были использованы двоичные категории источников: голос — отсутствие голоса и шум — отсутствие шума.

Термин «шумовой источник» относится к первичному акустическому возмущению в речевом тракте, возникающему при образовании шепотных, аспирированных, фрикативных и смычных звуков речи. Такой источник может быть непрерывным, как в случае звуков, которые могут произноситься протяжно, и импульсным, что имеет место для звуков, для которых малая длительность и определенная скорость нарастания и спадания являются критическими характеристиками [114]. В большинстве случаев шумовые источники по физическим условиям образования звука являются турбулентными; нетурбулентный шумовой источник образуется только тогда, когда происходит внезапный спад избыточного давления или резкое прерывание воздушной струи. Этот вид источника целесообразно называть «импульсным», как это принято в электронике.

Импульсный источник не идентичен с прерывистым турбулентным, хотя, как правило, оба они имеют место при образовании таких звуков, как глухие взрывные. При этом импульсный источник является результатом ударного воздействия на полости речевого тракта, получающегося в момент внезапного исчезновения избыточного давления при «взрыве», а турбулентный источник обусловлен случайными возмущениями воздушной струи, образующейся в момент «взрыва» и проходящей через сужения в речевом тракте.

Следует различать два вида турбулентного шума. Один из них — шум трения, или фрикативный шум, образующийся при наличии достаточно сильного сужения; при этом формирование спектра звука определяется в основном полостями и элементами речевого тракта, расположенными перед сужением, играющим главную роль в образовании звука, или в самом месте этого сужения. По длительности такой фрикативный интервал меньше у взрывных, чем у соответствующих фрикативных.

Другой вид турбулентного шума может быть назван «открытой аспирацией»; в фонетике он часто отмечается прибавкой индекса h к обозначению звука. Аспирация получается при более широком артикуляторном проходе, чем в случае фрикативных звуков. Более широкий проход и наличие в некоторых случаях более чем одного источника, например добавочного гортанного источника шума, усиливает роль формант, зависящих от всего речевого тракта в целом, а не только от его передних участков. В дальнейшем термин «аспирация» используется для обозначения специфического h -звука и не охватывает всей длительности того звука, который возникает при выдохе¹⁾.

Фрикация и аспирация, как они определены здесь, могут возникать как одновременно, так и друг за другом; возможны случаи, когда один из этих видов шума является основным. В шумовом интервале взрывного звука, во время которого происходит расширение артикуляторного прохода, аспирация должна следовать за фрикацией в том случае, если возникает и та и другая. Свойства голосового и шумового источников будут рассмотрены более подробно в Приложении I.

Аналитическое представление речеобразования путем разложения на две компоненты — источник и фильтр — можно продемонстрировать на простом примере, относящемся к сонорному звуку речи. Вследствие большого внутреннего сопротивления голосовой щели характеристикой источника можно считать заданный пульсирующий поток воздуха через голосовую щель. Этот поток, как функция времени, может быть представлен пилообразной кривой, приведенной на рис. 2 слева вверху; используя преобразование Фурье, можно получить спектр источника в виде ряда гармоник, как показано на рисунке слева внизу. Для получения амплитудного спектра звука после речевого тракта амплитуды каждой из гармоник спектра источника $|S(f)|$ умножаются на значение фильтровой функции $|T(f)|$ для частоты этой гармоники:

$$|P(f)| = |S(f)| \cdot |T(f)|. \quad (1.3)$$

Фаза каждой из гармоник после передачи по речевому тракту может быть получена как сумма фазы данной гармоники в спектре источника и фазы фильтровой функции для частоты этой гармоники, т. е.

$$\varphi_P(f) = \varphi_S(f) + \varphi_T(f). \quad (1.4)$$

¹⁾ Ср. аналогичную трактовку у Шаца [179] и Фишер-Йоргенсен [58].

Это — процесс синтеза, который может быть реализован во всех деталях говорящей машиной. Для фонетических исследований данные об огибающей амплитудного спектра $|P(f)|$ более удобны, чем данные о форме волны, представленной осциллографической записью звукового давления $p(t)$. Технической задачей спектрального анализа речи является получение огибающей функции $P(f)$ по временной зависимости $p(t)$ звука, воспринимаемого микрофоном. Методы, служащие для осуществления этой операции, рассмотрены в Приложении I. Фазовой функцией $\varphi_P(f)$, как правило, пренебрегают, так как она не добавляет существенной информации к амплитудным данным.

Резонансные пики, наблюдающиеся в спектральной картине звуков речи, называются формантами. На рис. 2 можно видеть, что каждому

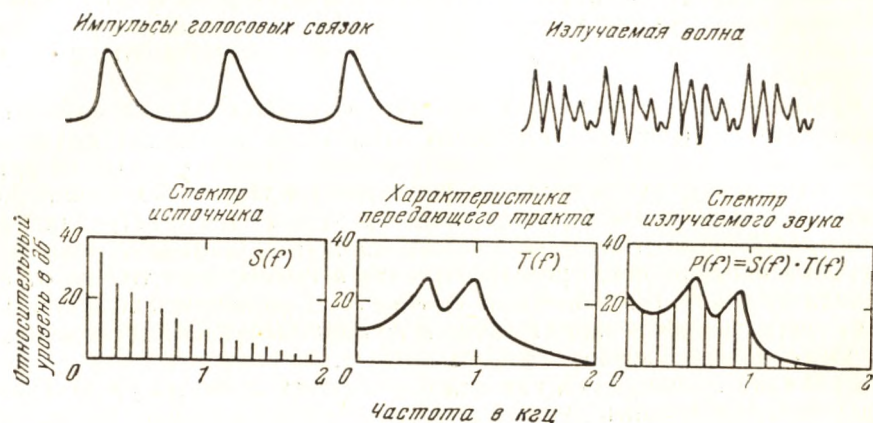


Рис. 2. Упрощенное представление образования спектра двухформантного сонорного звука как результата передачи спектра источника через фильтровую систему. Периодическая зависимость от времени изменений воздушного потока через голосовую щель преобразуется в спектр гармоник $S(f)$, который после умножения на характеристику фильтровой системы $T(f)$ дает спектр $P(f)$ произносимого гласного; этот же звук может быть представлен и как функция времени, которая показана наверху справа.

из резонансов соответствует область относительно больших значений функции передачи речевого тракта. Это избирательное свойство тракта, характеризующееся функцией $|T(f)|$, не зависит от свойств источника. Частотное положение максимумов $|T(f)|$, т. е. резонансные частоты, весьма близки к соответственным максимумам спектра $P(f)$ данного звука. Хотя, строго говоря, частоты резонансов и частоты этих максимумов не следует отождествлять — в большинстве случаев термины «резонансная частота» и «частота форманты» можно считать синонимами. Имея дело с сонорными звуками, для технических целей удобно считать формантные частоты атрибутами функции $T(f)$.

Основным принципом теории образования сонорных звуков является независимость, в первом приближении, фильтровой функции от свойств источника. В силу этого формантный пик только случайно может совпасть с частотой одной из гармоник источника. Частота форманты, таким образом, не зависит от расположения этих гармоник и может изменяться только в результате изменения артикуляции, при котором происходит изменение размеров различных участков речевого тракта и тем самым изменение фильтровой функции. Это означает в свою очередь, что по частотам формант можно судить о положении

артикуляторных органов; при этом, однако, надо помнить об ограничениях, обусловленных возможностью компенсаторных форм артикуляции.

Если формантные частоты неизменны, а основная частота голоса повышается на октаву, то в принципе удваивается только число импульсов в секунду, создаваемых голосовым аппаратом, удваивается расстояние между соседними гармониками, число же гармоник в заданном диапазоне уменьшается вдвое. Если некоторая форманта, например первая, до повышения основной частоты голоса была близка к 6-й гармонике, то после повышения основной частоты ближайшей к той же форманте будет 3-я гармоника. Поэтому не следует смешивать понятия частоты форманты и номера гармоники.

Если в речевом тракте имеется достаточно отчетливо выраженное сужение, образованное языком и разделяющее речевой тракт на переднюю и заднюю полости, то можно считать, что удовлетворяются необходимые условия для того, чтобы описывать речевой тракт как сдвоенный резонатор; условия эти, однако, не являются достаточными. Несмотря на это, подобная модель широко использовалась в литературе, причем предполагалось, что частота первой форманты определяется резонансом задней полости, а частота второй форманты — резонансом передней полости. Одна из задач настоящей книги — показать, что такая модель может иметь только ограниченное применение, и дать физиологически более обоснованную интерпретацию резонансов речевого тракта.

В гл. 4 показано, что каждая из формант определяется всеми участками речевого тракта, однако степень влияния этих участков на тот или иной резонанс различна и зависит от конкретной конфигурации речевого тракта. Модели этого тракта, составленные из отрезков труб или рупоров, являются более обоснованными и потому должны замещать упрощенные модели, в которых используются резонаторы Гельмгольца.

Интенсивность той или иной гармоники или группы гармоник в данной частотной области зависит как от спектра источника, так и от вида фильтровой функции. Как будет показано в гл. 3, уровень интенсивности группы спектральных составляющих в данной области частот понижается при уменьшении частоты низкочастотных формант или при удалении любой соседней форманты от рассматриваемой области.

Существенное влияние на спектр одного и того же звука оказывают изменения спектра источника, которые обусловлены индивидуальными особенностями голоса и голосовым усилием при произнесении звука. Уменьшение этого усилия, при неизменном частотном положении всех формант, приводит к понижению уровня всех гармоник, однако уровень более высоких гармоник снижается в большей мере, чем низких. Это обусловлено тем, что понижению интенсивности голоса обычно сопутствует увеличение крутизны спадающей огибающей спектра источника к высоким частотам.

Форманты как таковые обозначаются символами F_1 , F_2 , F_3 и т. д., в порядке их следования по шкале частот; для частот этих формант или соответствующих им резонансов речевого тракта приняты обозначения F_1 , F_2 , F_3 и т. д. В упрощенной модели, представленной на рис. 2, учитываются только две форманты. В действительных звуках всегда присутствуют и F_3 , и F_4 , хотя и с различной интенсивностью. Эти более высокие форманты особенно существенны в случае гласных переднего ряда.

Расстояние по частоте между формантами составляет для мужских голосов в среднем 1000 *гц*. Это статистическое среднее коррелировано со средней длиной речевого тракта. У женщин, для которых эта средняя длина меньше, расстояние между формантами больше и средние частоты формант более высоки. Аналогичные соотношения имеют место при сравнении детей со взрослыми. Более подробно эти соотношения рассмотрены в Приложении I.

По этой причине у двух дикторов, произносящих «один и тот же» гласный, частоты формант несколько разнятся в зависимости от конкретных размеров речевого тракта. Разброс частот формант может быть особенно велик, если рассматриваются все оттенки фонем, зависящие от контекста, и все возможные категории дикторов. Однако, как правило, у данного диктора в определенном контексте фонематическому различию между звуками соответствует вполне определенное различие в спектральной картине; это положение является основой теории дифференциальных признаков [114].

§ 2. Сегментация

Одним из основных вопросов при анализе речи является вопрос о степени членимости речевой волны. Обычный подход к решению этого вопроса сводится к тому, что за основу членения объективной картины речи принимаются лингвистические критерии, отражающиеся в фонематической транскрипции. Путем систематического сравнения акустических картин отдельных звуков в разных контекстах могут быть получены общие заключения об объективных признаках звука, типичных для данной фонемы. Важность этого рода исследований несомненна; однако, чтобы избежать неопределенности в разграничении следующих друг за другом речевых элементов, необходимо начать с сегментации и описания речевой волны на основе ее физической структуры; полученные в результате этого элементы явятся скорее фонетическими, а не фонематическими единицами. Широкие исследования именно в этом направлении необходимы для создания объективной основы для решения фонетических проблем. Книга «Видимая речь» Поттера, Коппа и Грина [174] содержит ценный материал по этому вопросу, однако необходимы еще более детальные и систематические исследования. Хотя эта работа и трудоемка, все же она представляется практически выполнимой ввиду ограниченности числа различных типов картин, подлежащих изучению.

В результате подобных исследований было бы получено описание акустической картины речи в виде последовательности во времени звуковых единиц, отчетливо отделенных друг от друга строго определенными границами. При установлении этих границ удобно исходить из представления о механизме образования звуков речи в виде источника звука и фильтра [49]. Так, граница между длительным носовым и следующим за ним длительным ротовым может быть определена по изменению системы фильтров речевого тракта, происходящему при раскрытии ротовой смычки; переход от звонкого длительного нёбного согласного к гласному, как можно видеть по изменению частотного положения формант, характеризуется более плавным изменением системы фильтров; граница между гласным и предшествующим ему интервалом аспирации, во время которого положение артикуляторных органов уже соответствует гласному, определяется в основном изменением источ-

ника; появление гласного после длительного зубного фрикативного согласного характеризуется относительно резким изменением как источника, так и свойств системы фильтров, поскольку эти свойства существенно зависят от местоположения в ней источника; на спектрограмме эти изменения обнаруживаются в виде смены случайной вертикальной штриховки в высокочастотной области, типичной для турбулентного звука, на формантные полосы с ясно очерченной структурой, расположенные преимущественно в низкочастотной области, что характерно для гласных; переход от паузы к последующему интервалу звучания определяется только источником при условии, что артикуляторные органы не успели, в отличие от случая взрывных, заметно изменить положение за время возбуждения источника.

В результате подобной чисто акустической сегментации может быть получено некоторое число минимальных звуковых единиц, имеющих размер, равный размеру звука речи или меньший; эти единицы могут быть классифицированы в соответствии с особенностями их образования. Число таких последовательных во времени звуковых единиц, как правило, больше числа символов фонетической или фонематической транскрипции. При желании согласовать эту транскрипцию со спектрографическими записями исследователь вынужден следовать некоторым условным правилам соотнесения акустических единиц с теми или другими графическими знаками. Это не приведет к недоразумениям в том случае, если все исследователи будут располагать подходящими средствами для обнаружения границ акустических отрезков речи и если правила соотнесения будут согласованы между всеми исследователями. Эти условия, однако, далеко не всегда выполняются; поэтому к имеющимся в литературе данным о длительности согласных следует относиться достаточно осторожно, в особенности в том случае, если эти данные основываются только на осциллографических или кимографических записях речи.

Хорошо известно, что идентификация слушающим данной фонемы часто зависит от признаков, содержащихся в нескольких последовательных во времени акустических отрезках, и притом не только в том отрезке или в тех отрезках, которые обычно относят к определенному символу традиционной фонетической транскрипции, но и в тех, которые принято приписывать предшествующей или последующей фонеме. Хорошо известным примером этого являются переходные смещения формант в пределах первой части звучания гласного, которые определяют идентификацию предшествующей согласной. Подобные «переходные» признаки имеют особо большое значение для различения носовых согласных, а также звонких взрывных, тогда как интенсивные фрикативные мало зависят от этих признаков.

Так, например, слово [tapa], содержащее четыре звука речи, может быть разбито на соответствующие этим звукам четыре акустические единицы; то же относится и к слову [paɪa]. Если магнитную пленку с записью этих слов разрезать в точках стыка таких единиц и затем склеить с заменой всех отрезков, содержащих звук [m], на отрезки, содержащие [n], то при прослушивании, как показали эксперименты, фонематический сдвиг не обнаруживается. Однако замена «гласного» одного из этих слов «гласным» другого приводит к такому сдвигу [143].

В качестве другого примера можно взять слова [pæt] и [kæt]. Здесь естественно принять за границу между согласным и гласным те моменты, когда появляются или исчезают голосовые колебания. Членение

в соответствии с этим признаком дает действительную границу между звуками; в этом можно убедиться, разрезав запись этих слов в точке появления голоса и начальное [p] заменив на начальное [k]. При прослушивании после такой замены имеет место явно выраженная тенденция слушающих идентифицировать каждое из полученных после склейки слов в соответствии с тем согласным сегментом, который в нем содержится теперь¹⁾. Это объясняется тем, что в момент появления голоса артикуляторные органы уже настолько приближены к положению, типичному для данного гласного, что основная часть переходного процесса оказывается законченной; первая, и наиболее существенная, часть перехода совершается за время аспираторного интервала [57, 58, 53]. Имеется и еще одна причина коммутативности шума согласного в этом примере, а именно его относительно большая интенсивность.

В случае взрывных согласных сегментация их на взрыв, фрикацию и аспирацию, в соответствии с § 1, не всегда может быть получена на основе спектрограмм. Отделить взрыв от фрикации часто бывает затруднительно по причине краткости первого. Длительность взрыва определяется переходными процессами в речевом тракте и по порядку величины равна обратной ширине полосы данной форманты; поэтому она пренебрежимо мала в высокочастотной области шума, например зубного взрывного звука. Интервал фрикации выявляется более отчетливо для нёбных и зубных по сравнению с губными, но переход к аспирации редко бывает резким. Различие между аспирацией и фрикацией проявляется в уменьшении интенсивности в области частот выше 4000 гц и в появлении формант, непрерывным образом переходящих в форманты последующего гласного. В интервалах, соответствующих аспирации, первая форманта чаще всего выражена весьма слабо, за исключением тех случаев, когда последующим звуком является какой-либо из задних гласных.

Акустическая сегментация слова [bæt] не может быть произведена, если пользоваться только признаком появления голоса, так как голосовой источник может функционировать на протяжении всего начального [b]. Наличие в интервале смычки «голосовой полосы»²⁾ в самой нижней части спектрограммы указывает на появление колебаний голосовых связок во время накопления избыточного давления в полостях рта перед размыканием смычки. Для русского языка наличие этой полосы является непременным атрибутом звонких взрывных. В случае шведского или английского языка голос может появляться только в момент самого взрыва или через очень малый промежуток времени после него. В этих случаях, при отсутствии «голосовой полосы», когда [p] и [b] различаются по признаку «сильные — слабые», звонкость [b] проявляется в том, что интервал времени от взрыва до появления колебаний голосовых связок больше для [p], чем для [b]. Кроме того, [b] не сопровождается аспирацией и обычно имеет очень слабо выраженный фрикативный интервал. Поэтому можно думать, что смещение формант начинается с появлением голосового источника в [b] и шумового источника в интервале аспирации в [p]. С артикуляторной точки зрения началом гласного в [pæt] является аспираторный интервал [p].

¹⁾ Эксперименты выполнялись Х. М. Траби.

²⁾ Этот термин («voice bar») введен в работе Поттера, Коппа и Грина «Видимая речь» для обозначения частотной области почернения на спектрограмме, соответствующей основной частоте голоса. (Прим. ред.)

Сходство процессов образования взрывных и носовых согласных становится ясным, если сравнить, например, начальное [b] со звонкой смычкой с начальным [m]. В интервале полной ротовой смычки, которая характерна для обоих звуков, передача звука во внешнюю среду в случае взрывного согласного происходит через колеблющиеся стенки ротовой полости, в случае же носового — через ноздри. Второй путь передачи звука является, очевидно, более эффективным. Однако перед размыканием губной смычки избыточное давление больше в случае взрывных. Это приводит к большей разнице интенсивностей для открытого и закрытого интервалов взрывных по сравнению с соответственной разницей для носовых.

§ 3. F-картина

Для того чтобы в акустической картине речи до конца отделить данные, относящиеся к источнику, от данных, относящихся к фильтровой функции, необходимо располагать детальным описанием речевого тракта диктора в процессе произнесения им речи. Такое описание позволяет найти свойства тракта, характеризующие его как систему передачи, т. е. определить фильтровую функцию. Характеристики источника для того или иного отрезка речи могут быть после этого получены вычитанием, в логарифмических масштабах, фильтровой функции из спектра рассматриваемого отрезка в соответствии с формулой (1.1). Такого рода анализ был выполнен для фрикативных и взрывных, и результаты его приведены в гл. 10.

Однако, имея в виду более общие фонетические выводы, оказывается возможным частично восстановить приближенные характеристики источника и фильтровую функцию, не обращаясь к детальным геометрическим данным о конфигурации речевого тракта. В известной мере такой подход сводится просто к физиологической интерпретации спектрограмм, основанной на общих представлениях о речеобразовании.

Тонкая структура спектра изучаемого звука или некоторой части его позволяет установить тип источника, т. е. определить, является ли он одиночным нестационарным или длящимся процессом; в последнем случае можно различить, имеет ли источник шумовой характер или тональный или обладает определенной комбинацией всех этих признаков. С другой стороны, тонкая структура может быть исключена из рассмотрения, если провести огибающую спектра, т. е. плавную кривую, охватывающую последовательные гармоники. Огибающая дает возможность выявить формантную структуру звука, или, иначе говоря, топологию формантных пиков.

При проведении огибающей следует по возможности сохранять основные провалы в спектре, соответствующие антирезонансам системы. Для решения определенной практической задачи огибающую можно сглаживать, в результате чего соседние формантные пики, не требующие по тем или иным соображениям изолированного рассмотрения, объединяются в одну форманту или, точнее, единую формантную группу. Такое сглаживание целесообразно, например, применить при выявлении основной форманты зубного взрывного или фрикативного звука, расположенной выше 4000 гц; в этом случае детальная формантная структура не представляет большого интереса. Аналогично при изучении восприятия может оказаться желательным рассматривать вторую и более высокие форманты переднего гласного как единую формантную группу.

Те форманты, частота которых определяется в основном ротовой частью речевого тракта, т. е. фарингальной, нёбной и губной областями, имеют наибольшее значение для анализа речи. Это определяется тремя обстоятельствами:

1) Возможностью получения фильтровой функции и тем самым общей формы спектра гласной по данным о частотах формант [52]. Используя это, можно осуществить разделение характеристик источника и фильтра, не прибегая к данным об артикуляции рассматриваемого звука.

2) Возможностью судить об артикуляции того или иного звука, если дана огибающая спектра и в особенности частоты тех формант, которые плавно переходят в форманты соседних гласных.

3) Значением переходных признаков для восприятия речи; эти признаки содержатся в изменениях формантных частот в отрезках акустической картины речи, примыкающих к согласному звуку.

Все это оправдывает целесообразность обобщения понятий «хаб» (hub) и «локуса» (locus)¹⁾ введением понятия *F-картины*.

F-картина в данный момент времени определяется совокупностью резонансных частот ротовой части речевого тракта, или тех резонансных частот, которые непрерывным образом переходят в ротовые резонансы соседних звуков. *F-картина* весьма близка к наблюдаемой совокупности частот формантных пиков на огибающей спектра таких, по преимуществу сонорных, звуков, для которых источник расположен в гортани. Поэтому желательно принять такой метод определения частот формант, при котором эти частоты полностью совпадали бы с резонансными частотами. Под *F-картиной* тогда будет подразумеваться совокупность частот F_1 , F_2 , F_3 , F_4 и т. д. Каждая из этих частот характеризует «локус» или «положение» на частотной шкале. F_2 -локус идентичен с «хаб» Поттера, Коппа и Грина [174], которые определяют его как «видимое или скрытое положение второй форманты на спектрограмме»²⁾.

Реализация *F-картины* в виде звука осуществляется только при наличии источника; таким источником не обязательно должны являться голосовые связки. Поскольку изменения положения артикуляторных органов имеют непрерывный характер и возможны только в ограниченных пределах, *F-картина* непрерывна; поэтому представляет определенный интерес и непрерывное наблюдение за нею в пределах достаточно большого речевого отрезка, включая даже интервалы отсутствия звука (паузы), где ее можно установить путем интерполяции.

Переходные процессы в сочетаниях согласный — гласный и гласный — согласный могут быть при этом представлены в виде последовательности *F-картин* во времени; такое представление дает в наиболее точном и компактном виде акустические корреляты артикуляции, т. е., точнее, конфигурации речевого тракта. Смещение формант в переход-

¹⁾ Соответствующие термины в нашей литературе отсутствуют. Термин «хаб» был введен в книге Поттера, Коппа и Грина [174], термин «локус» — в работах лабораторий Хаскинса. Содержание этих терминов разъясняется автором ниже. (Прим. ред.)

²⁾ Добавление к этому определению, сделанное Поттером, Коппом и Грином: «когда звук произносится изолированно», имело в виду возможность выдержки данной артикуляции. Здесь предлагается сохранить название «локус» для *F-картины* любого звука как характеристику предельного положения артикуляторных органов, например полной смычки согласного, а наименование «*F-положение*» использовать для характеристики произвольного отрезка связной речи. Стивенс и Хауз [195] пользуются названием «локус» как раз в этом смысле. Сотрудники группы из лаборатории Хаскинса (Делаттр и др. [33]) связывают понятие локуса в основном с эмпирически найденными ими правилами построения формантных смещений для их синтетических картин.

ных интервалах может быть как в сторону повышения их частоты, так и в сторону понижения. Принято также называть подобные смещения положительными или отрицательными, в зависимости от того, имеет ли согласный более низкий или, соответственно, более высокий F_n -локус, чем частота F_n последующего или предшествующего гласного [86].

Хорошо известно (см., например, [30, 195, 53]), что низкая F_1 является признаком артикуляторного затвора, очень низкая F_2 характерна для продвинутой назад артикуляции в комбинации с губной смычкой, очень высокая F_2 соответствует подъему языка к нёбу, а высокая F_3 является показателем передненёбной или зубной артикуляции. Зубным соответствует средне-высокое положение F_2 ; для губных F -картина изменяется с изменением положения языка. Результатом губной смычки является понижение всех локусов на ту или иную величину. Очень низкая F_3 является признаком ретрофлексных модификаций звуков. Соотношения между F -картиной и артикуляцией будут более подробно рассмотрены в § 3 гл. 4 и в части II.

При использовании спектрографических данных для фонетического описания звукового состава речи следует иметь в виду следующее общее положение: спектр гласных достаточно хорошо определяется F -картиной, однако для характеристики согласных эта картина должна быть дополнена огибающей спектра.

Как уже было отмечено выше, основной ход огибающей спектра может быть получен расчетным путем, если известны формантные частоты, т. е. F -картина. Хотя это и справедливо полностью для большинства звуков, в образовании которых участвует голосовой источник, однако к латеральным согласным и назализованным гласным это может быть отнесено только с некоторыми ограничениями. Форманты, которые видны в спектре назального звука¹⁾, только частично могут считаться входящими в состав F -картины, поскольку последняя охватывает только форманты, возникающие в ротовой части речевого аппарата.

В спектрах фрикативных согласных можно наблюдать второстепенные пики в их формантной области или изолированные слабо выраженные форманты. Хотя они и относятся в какой-то мере к F -картине, однако их недостаточно для того, чтобы восстановить весь спектр согласного.

ГЛАВА 2

ТЕОРИЯ ЦЕПЕЙ, ЭКВИВАЛЕНТНЫХ РЕЧЕВОМУ ТРАКТУ

§ 1. Эквивалентные схемы акустических резонаторов и рупоров

Математическое рассмотрение процесса речеобразования распадается на ряд последовательных операций. Первой из них является представление конфигурации речевого аппарата в виде функции площади, характеризующей изменение площади поперечного сечения речевого тракта вдоль его оси, начиная от голосовой щели и до излучающего отверстия рта. Второй операцией является приближенное

¹⁾ За отсутствием в русском языке фонетического термина, соответствующего термину «nasal pigtug», мы перевели его как «назальный звук». Это — звук, образующийся при опущенной нёбной занавеске во время ротовой смычки. (Прим. ред.)

представление этой функции системой по возможности малого количества отрезков, в пределах которых площади можно считать постоянными. Далее, для этой системы находится функция передачи, которая складывается с частотной характеристикой источника. Наконец, полученная таким путем суммарная частотная характеристика представляется в компактном виде как совокупность полюсов и нулей. При рассмотрении сонорных звуков наибольшее значение имеют частоты формант. Дополнительно, в целях дальнейшего уменьшения физической избыточности, можно проанализировать форму рассчитанного спектра подобно тому, как это было сделано в предыдущей главе.

Представление речевого тракта в виде эквивалентной схемы основывается на понятии акустического импеданса, определяемого как отношение звукового давления к объемной скорости. Его не следует смешивать с удельным импедансом, представляющим собой отношение давления к скорости частиц, а также с механическим импедансом, равным отношению силы к скорости частиц. Преимущество использования именно акустического импеданса определяется непрерывностью изменения как объемной скорости, так и звукового давления в пределах всей системы резонирующих полостей.

Теория простого резонатора Гельмгольца, как будет подробно показано в § 3 гл. 7, может быть полезна при рассмотрении характеристик передачи речевого тракта только частично, так как размеры используемых при этом резонаторов оказываются относительно большими. Имеющие более общее значение выводы можно получить, если всю систему резонирующих полостей разбить сечениями, нормальными к направлению распространения звуковых волн, на ряд отрезков, достаточно малых для того, чтобы считать площади поперечного сечения в их пределах постоянными. Эквивалентом такого отрезка с площадью поперечного сечения A и длиной l можно считать индуктивность $L = \frac{\rho l}{A}$ и следующую за ней в виде поперечной ветви емкость $C = \frac{IA}{\rho c^2}$;

последовательно с L включается сопротивление R , а параллельно с C — проводимость G . Подобная эквивалентная схема хорошо отображает свойства отрезка конечной длины в области частот, при которых длина волны остается больше $8l$.

В акустических системах, сверх того, надо считаться с наличием критической частоты, выше которой появляются колебания в радиальном направлении. Соответствующая этой частоте длина волны приблизительно в два раза больше наибольшего из поперечных размеров системы.

Вообще говоря, если принять, что потери, определяемые величинами R и $\frac{1}{G}$, относительно малы, резонансные частоты системы резонирующих полостей любой сложности определяются только элементами L и C ; элементы R и G , или, точнее говоря, отношения $\frac{R}{L}$ и $\frac{G}{C}$, определяют только ширину резонансных пиков.

Если длина отрезка не очень мала, более точное представление его в виде эквивалентной схемы можно получить, если разбить всю соответствующую его длине индуктивность на две равные половины и поместить их симметрично относительно емкости. Отрезок, имеющий площадь поперечного сечения A и длину l , представится в этом случае эквивалентной схемой в виде Т-образного звена, показанного в верхней

части рис. 3, в состав которого входят элементы

$$a = Z \operatorname{th} \frac{\Gamma}{2}, \quad b = \frac{Z}{\operatorname{sh} \Gamma}. \quad (2.1)$$

Характеристический импеданс Z и постоянная распространения Γ определяются распределенными параметрами L , C , R и G , рассчитанными на единицу длины, в соответствии с классическими выражениями

$$\left. \begin{aligned} Z &= \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \\ \Gamma &= l \sqrt{(R + j\omega L)(G + j\omega C)} = l(\alpha + j\beta) = l\gamma, \end{aligned} \right\} \quad (2.2)$$

где γ — комплексный коэффициент распространения, включающий коэффициент затухания α и фазовый коэффициент β .

В случае малости потерь

$$\left. \begin{aligned} Z &= Z_0 \left[1 - \frac{j\alpha_R}{\beta} + \frac{j\alpha_G}{\beta} \right], \\ \gamma &= \alpha_R + \alpha_G + j\beta, \end{aligned} \right\} \quad (2.3)$$

где

$$\left. \begin{aligned} Z_0 &= \sqrt{\frac{L}{C}}, \quad \alpha_R = \frac{R}{2Z_0}, \\ \beta &= \omega \sqrt{LC}, \quad \alpha_G = \frac{GZ_0}{2}. \end{aligned} \right\} \quad (2.4)$$

Поскольку индуктивность L и емкость C на единицу длины для акустической линии равны соответственно

$$L = \frac{\rho}{A}, \quad C = \frac{A}{\rho c^2}, \quad (2.5)$$

формулы (2.4) могут быть представлены в виде

$$\left. \begin{aligned} Z_0 &= \frac{\rho c}{A}, \quad \alpha_R = \frac{1}{c} \cdot \frac{R}{2L} = \frac{RA}{2\rho c}, \\ \beta &= \frac{\omega}{c}, \quad \alpha_G = \frac{1}{c} \cdot \frac{G}{2C} = \frac{G\rho c}{2A}. \end{aligned} \right\} \quad (2.6)$$

Если пренебречь потерями, то малый отрезок резонирующей системы полностью характеризуется его длиной и площадью поперечного сечения. Это приближенно справедливо для любой формы поперечного сечения, а не только для круговой.

Числовые расчеты, если только не пользоваться цифровыми или моделирующими электронными вычислительными машинами, оказываются довольно сложными и громоздкими. Поэтому целесообразно начать с некоторой простой аппроксимации речевого тракта, которая достаточна для целей настоящего исследования.

Один из возможных путей такого приближенного представления сводится к использованию рупора в качестве аналога значительного участка речевого тракта, на котором площадь поперечного сечения меняется непрерывно. Акустическая теория рупора в том виде, как она представлена у Морза [155], дает достаточно хорошую физическую основу для расчетов. Однако при рассмотрении рупора как части сложной

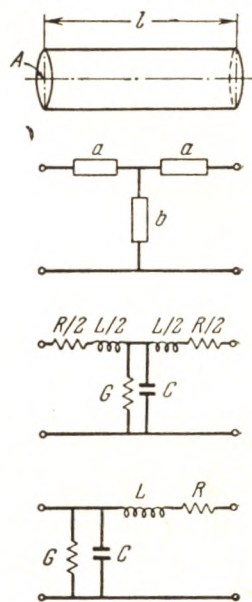


Рис. 3. Т-образная эквивалентная схема отрезка трубы и приближенное представление этой схемы, пригодное в области низких частот или при малой длине отрезка.

системы полостей необходимо, кроме волнового уравнения, описывающего работу рупора, располагать эквивалентной его схемой. Эквивалентные схемы для неоднородных линий с непрерывно изменяющимися параметрами, в точности соответствующих классу рупоров, рассмотренных Морзом, были получены Лоурентом [126, 129].

Общее выражение для площади поперечного сечения рупора $A(x)$, как функции расстояния x вдоль его оси, имеет вид

$$A(x) = \frac{A_0}{\text{ch}^2 \varepsilon} \text{ch}^2 \left(\frac{x}{h} + \varepsilon \right), \quad (2.7)$$

если площадь возрастает при возрастании x , и

$$A(x) = \frac{A_0 \text{ch}^2 \varepsilon}{\text{ch}^2 \left(\frac{x}{h} + \varepsilon \right)}, \quad (2.8)$$

если площадь убывает с возрастанием x .

Если $\varepsilon = 0$, получаем катеноидальный рупор; если $\varepsilon = \infty$, то рупор экспоненциальный; наконец, если $\varepsilon = \frac{x_0}{h} - j \frac{\pi}{2}$ и $h \rightarrow \infty$,

то рупор конический. Постоянные ε и h могут, разумеется, иметь значения и промежуточные между указанными, что соответствует рупорам различных промежуточных форм.

Пользуясь постоянными, входящими в предложенное Морзом уравнение рупора, элементы эквивалентных схем на рис. 4

можно, следуя Лоуренту, определить следующим образом:

Длина резонатора	l .
Номинальный характеристический импеданс . .	$Z_0 = \frac{\rho c}{A_0}$.
Номинальный коэффициент распространения . .	$\gamma_0 = \alpha + \frac{j\omega}{c}$.
Критическая частота рупора	$\omega_0 = \frac{c}{h}$.
Коэффициент распространения	$\gamma = \gamma_0 \left(1 - \frac{\omega_0^2}{\omega^2} \right)^{1/2} = \gamma_0 \tau$.
Постоянная распространения	$\Gamma = l\gamma$.
Характеристический импеданс	$\left\{ \begin{array}{ll} Z = \frac{Z_0}{\tau} & \text{при возрастающей} \\ & \text{площади,} \\ Z = Z_0 \tau & \text{при убывающей} \\ & \text{площади.} \end{array} \right.$
Продольный элемент эквивалентного Т-образного звена	$a = Z \text{th} \frac{\gamma l}{2}$.

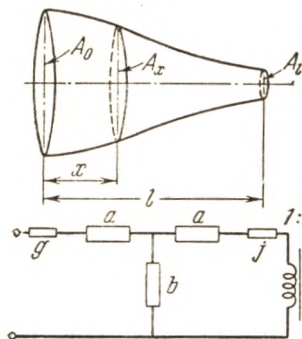
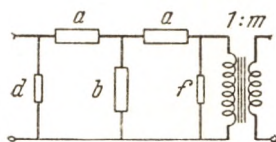
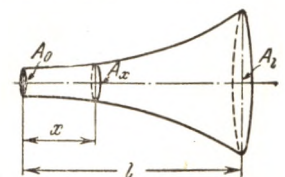


Рис. 4. Рупорные резонаторы и их эквивалентные схемы по Лоуренту [126, 129].

Поперечный элемент эквивалентного Т-образ-

ного звена $b = \frac{Z}{\text{sh } \gamma l}$.

Коэффициент трансформации импеданса $m^2 = \frac{A_0}{A_l}$.

Кроме указанных, в эквивалентную схему входят дополнительные последовательные и параллельные элементы, а именно:

$$\left. \begin{aligned} d &= Z_0 h \gamma_0 \text{cth } \frac{l}{h}, \\ f &= -Z_0 h \gamma_0 \text{cth } \left(\frac{l}{h} + \varepsilon \right) \end{aligned} \right\} \text{ в случае возрастающей площади,}$$

$$\left. \begin{aligned} g &= \frac{Z_0}{h \gamma_0} \text{cth } \frac{l}{h}, \\ j &= -\frac{Z_0}{h \gamma_0} \text{cth } \left(\frac{l}{h} + \varepsilon \right) \end{aligned} \right\} \text{ в случае убывающей площади.}$$

В частных случаях катеноидального, экспоненциального и конического рупоров приведенные выражения принимают вид:

	Возрастающая площадь		Убывающая площадь	
	d	f	g	j
Катеноидальный рупор . .	∞	$-Z_0 h \gamma_0 \text{cth } \frac{l}{h}$	0	$-\frac{Z_0}{h \gamma_0} \text{cth } \frac{l}{h}$
Экспоненциальный рупор .	$Z_0 h \gamma_0$	$-Z_0 h \gamma_0$	$\frac{Z_0}{h \gamma_0}$	$-\frac{Z_0}{h \gamma_0}$
Конический рупор	$Z_0 x_0 \gamma_0$	$-Z_0 (x_0 + l) \gamma_0$	$\frac{Z_0}{x_0 \gamma_0}$	$-\frac{Z_0}{(x_0 + l) \gamma_0}$

Катеноидальный рупор особенно полезен при моделировании речевого тракта, так как для него исчезает хотя бы один из добавочных элементов в эквивалентной схеме; кроме того, существенно то, что угол наклона касательной к образующей рупора равен нулю при $x=0$, что обеспечивает непрерывное сочленение рупора с другими элементами. Так, при образовании таких звуков, как [i] и [e], вся ротовая полость, включая небное сужение и передний и задний объемы, может быть представлена единой схемой, поскольку участки тракта по обе стороны от точки максимального сужения примыкают друг к другу полностью согласованно. То же относится и к тому случаю, когда ротовая полость сужается, т. е. представляет собой часть речевого тракта, функция площади которого имеет максимум. В задних гласных функция площади всего фаринкса имеет вид кривой, соответствующей расширению его поперечного сечения.

Изложенная теоретическая трактовка предполагает распространение волн только в одном измерении. При этом следует особо подчеркнуть, что отражение у выходного отверстия рта очень велико, так что упрощенная теория рупорного громкоговорителя, учитывающая только бегущие волны, в данном случае неприменима.

Наличие в эквивалентной схеме трансформатора указывает на первый взгляд на то, что мощность на выходе речевого тракта будет пропорциональна излучающей поверхности ротового отверстия. Однако на самом деле это правильно только для частот, значительно превышающих критическую частоту рупора, и, вообще говоря, только для участков спектра, расположенных между формантными пиками; более подробно этот вопрос освещен в § 2 гл. 4.

При частотах, значительно превышающих критическую, эквивалентная схема такая же, как для отрезка трубы с постоянной площадью поперечного сечения, равной площади горла рупора; единственным отличием является присутствие в случае рупора идеального трансформатора. При более низких частотах влияние трансформатора ослабляется наличием в схеме элементов d и f .

Распространение звуковых волн, даже в идеальных трубах с жесткими стенками, связано с потерями. В пограничном слое, прилегающем к внутренней поверхности трубы, возникают потери на трение. Их можно учесть, введя в схему последовательное сопротивление

$$R = \frac{S}{2A^2} (2\mu\omega\rho)^{1/2}, \quad (2.9)$$

где $\mu = 1,84 \cdot 10^{-4}$ э/см·сек — коэффициент вязкости. В случае трубы кругового сечения отношение площади A к периметру S равно половине радиуса r и сопротивление определяется выражением

$$R = \frac{1}{Ar} (2\mu\omega\rho)^{1/2}. \quad (2.10)$$

Обладающий вязкостью пограничный слой, кроме внесения потерь, уменьшает также эффективную величину площади, входящей в выражение для индуктивности L на единицу длины. С учетом этого обстоятельства имеем:

$$L = \left[\frac{\rho}{A} + \frac{S}{A} \left(\frac{\mu}{2\omega\rho} \right)^{1/2} \right]. \quad (2.11)$$

Однако в расчетах, относящихся к речевому тракту, поправочным членом в этом выражении можно пренебрегать.

Во многих руководствах по акустике, например [147, 202], потери на теплопроводность у стенок звукопровода учитываются как кажущееся увеличение вязкости. Однако такой способ учета этих потерь непригоден для построения эквивалентных схем. Действительно, потери на трение в эквивалентной схеме должны быть пропорциональны квадрату тока, а потери на теплопроводность — квадрату напряжения. Поэтому последние должны быть учтены путем введения в схему элемента с проводимостью G , определяемой формулой

$$G = S(\kappa - 1) \frac{\beta}{\rho c} \left(\frac{K_h}{2\omega C_p \rho} \right)^{1/2}, \quad (2.12)$$

где K_h — коэффициент теплопроводности воздуха, C_p — теплоемкость воздуха при постоянном давлении, C_v — то же при постоянном объеме, $\kappa = \frac{C_p}{C_v}$, $\beta = \frac{\omega}{c}$.

При обычных атмосферных условиях (1 атм, 20°С) найдем, подставляя соответствующие численные значения, следующие выражения для элементов схемы, учитывающих потери:

$$\left. \begin{aligned} R &= \frac{1,66 \cdot 10^{-3} f^{1/2}}{A \left(\frac{2A}{S} \right)}, & \alpha_R &= \frac{2,01 \cdot 10^{-5} f^{1/2}}{\frac{2A}{S}}, \\ G &= \frac{S \cdot 0,91 \cdot 10^{-5} f^{1/2}}{c}, & \alpha_G &= \frac{0,91 \cdot 10^{-5} f^{1/2}}{\frac{2A}{S}}. \end{aligned} \right\} \quad (2.13)$$

Эти выражения показывают, что пренебрегать при расчетах теплопроводностью нельзя, так как обусловленная ею часть коэффициента затухания составляет почти половину той части его, которая учитывает потери на трение. Однако в полостях речевого аппарата потери на теплопроводность, вероятно, малы по сравнению с потерями энергии, возникающими вследствие колебаний стенок полостей. Теоретически эти потери могут быть учтены введением в схему шунтирующего импеданса, который удобно представить в виде большой индуктивности и параллельного ей большого сопротивления. Обратная величина этого сопротивления будет определять проводимость потерь G_s . Согласно Ван-ден-Бергу [8], стенки фаринкса на интересующих нас частотах обладают положительным реактивным сопротивлением и обуславливают значительную часть затухания для первой форманты (см. также Приложение III, § 6, в котором приведены аналогичные расчеты Стивенса [192]). Влияние реактивной составляющей импеданса стенок на частоты резонансов несущественно, за исключением интервалов полной или почти полной смычки язычного или губного прохода (см. гл. 7, § 4, Б).

Независимо от физической природы явлений, вызывающих потери энергии у стенок звукопровода, можно видеть, что коэффициент затухания для отрезка звукопровода с постоянной вдоль его длины площадью поперечного сечения пропорционален периметру, какова бы ни была форма этого сечения. Это справедливо как для α_R , так и для α_G . Носовые полости характеризуются особенно большим коэффициентом

формы $\frac{S}{2\sqrt{A\pi}}$ в силу их довольно сложной конфигурации. Разветвление звукопровода в носовой полости на два канала увеличивает отношение общей поверхности его к объему. К такому же эффекту приводит и образование, при высоком подъеме языка, узкого прохода, поперечное сечение которого сильно отличается от круговой формы. Если для такого прохода принять эллиптическую форму с высотой, равной $1/9$ ширины, то коэффициент формы достигает величины 2. Коэффициент формы для кругового сечения равен, согласно определению, единице.

Расчеты показали, что скорости частиц в узких проходах могут оказаться достаточно большими для того, чтобы обусловить значительное нелинейное увеличение сопротивления трения при низких частотах. Расчеты, выполненные Ингардом [108], показали, что сопротивление в узком отверстии с резким переходом к прилежащим полостям определяется соотношением

$$R = \frac{1,66 \cdot 10^{-3} f^{1/2}}{Ar} \left[1 + 2r + 2r \cdot 0,7 \left(\frac{v}{100} \right)^{1,7} \right], \quad (2.14)$$

где r — радиус отверстия. Второй член в квадратных скобках — линейная часть концевой поправки, представляющая собой сумму поправок по ту и по другую сторону отверстия; третий член характеризует нелинейное возрастание концевой поправки. Этот член не зависит от длины отверстия, если длина отверстия велика по сравнению с колебательным смещением частиц v/ω .

Однако для расчетов, относящихся к речевому тракту, необходимость применять эту формулу ограничена, так как резкие изменения площади наблюдаются только у зубов и у верхних краев голосовых связок.

При образовании длительного сонорного или глухого звука к формантам или иным спектральным составляющим, характеризующим структуру сигнала, добавляется постоянная составляющая в виде постоянного потока воздуха. Как показано в § 1 Приложения II, при протекании потока воздуха через узкий проход в условиях существования турбулентности возникает сопротивление, зависящее от величины потока; это сопротивление определяется выражением

$$R_{\text{диф}} = 2R_{\text{пот}} = \frac{\rho v}{A}, \quad (2.15)$$

где A — площадь поперечного сечения, v — скорость частиц.

Нагрузкой фильтра, моделирующего речевой тракт, является импеданс излучения. Этот импеданс содержит активное сопротивление и последовательно включенное индуктивное сопротивление. Потери энергии в активном сопротивлении соответствуют излучаемой энергии, а индуктивное сопротивление соответствует эффективной массе соколеблющегося воздуха у ротового отверстия.

В приводимых ниже расчетах принята формула Морза [155] для излучающего поршня кругового сечения, расположенного на поверхности жесткой сферы радиуса $a = (A_0/\pi)^{1/2} = 9$ см. При частотах, для

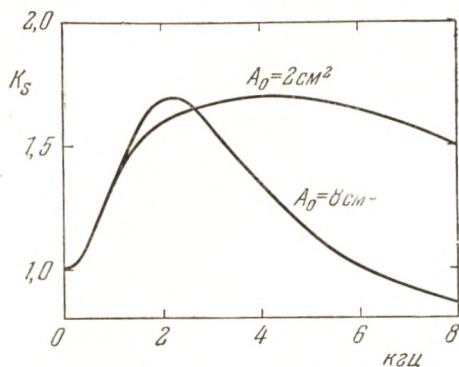


Рис. 5. Поправочный коэффициент к сопротивлению излучения $K_S = \frac{R_0}{\rho \omega^2 / 4\pi c}$ для частотного диапазона, наиболее существенного при расчете передачи по речевому тракту. При очень низких частотах влияние экрана пренебрежимо мало и сопротивление излучения равно половине его значения для случая бесконечного экрана. При более высоких частотах наличие сферического экрана с радиусом 9 см приводит к тому, что при площадях излучающей поверхности, указанных на кривых, значения сопротивления излучения приближаются к значению этого сопротивления для случая бесконечного экрана, но не достигают их.

которых $\frac{\omega a}{c} < 2$ или $f < 6000 \left(\frac{10}{A_0}\right)^{1/2}$, где A_0 — излучающая поверхность, т. е. во всей области частот, существенной для расчета спектров звуков речи, сопротивление излучения может быть выражено следующим образом:

$$R_0 = \frac{\rho \omega^2}{4\pi c} K_S(\omega). \quad (2.16)$$

Зависящий от частоты коэффициент $K_S(\omega)$ учитывает экранирующее действие головы (рис. 5). При очень низких частотах $K_S = 1$, как и следовало ожидать, так как при таких частотах источник можно рассматривать как точечный. При более высоких частотах K_S достигает максимального значения 1,7 в области 2000 гц, после чего начинает уменьшаться. В более ранних расчетных работах, которые были выполнены Чибя и Каджияма [24], Дэнном [39] и Ван-ден-Бергом [8], сопротивление излучения учитывалось на основе теории поршня, колеблющегося в бесконечном плоском экране; в этом случае значение K_S в формуле (2.16) равно 2.

Сопротивление излучения почти совершенно не зависит от площади излучающей поверхности до частоты порядка 3000 гц. При более высоких частотах большей площади соответствует меньшее сопротивление. Характер зависимости сопротивления излучения от частоты показан на рис. 6 пунктирной кривой. Пропорциональность ω^2 имеет место до

величин ω , которые соответствуют условию $\left[\omega \left(\frac{A_0}{\pi} \right)^{1/2} / c \right] = 1,85$; при сравнительно большой площади ротового отверстия, $A_0 = 8 \text{ см}^2$, вытекающая из этого условия частота равна 6500 гц. При более высоких частотах сопротивление излучения приближается по величине к характеристическому сопротивлению выходного участка речевого тракта $\rho c / A_0$.

Сплошные кривые на рис. 6 дают активное и реактивное сопротивление того звена в шведском электрическом аналоге речевого тракта LEA, которое учитывает излучение.

Как можно видеть из рисунка, соответствие с расчетными кривыми в пределах частот, интересующих нас при исследовании речи, хорошее.

Реактивное сопротивление излучения может быть для низких частот приближенно учтено концевой поправкой

$$l_0 = 0,8 \left(\frac{A_0}{\pi} \right)^{1/2}. \quad (2.17)$$

Если при переходе от звена к звену в сложной системе полостей имеет место резкое изменение площади от некоторой меньшей величины A_0 до большей A , то в месте перехода возникает некоторый эффект излучения. Соответствующая этому добавочному индуктивному сопротивлению внутренняя концевая поправка равна по Ингарду [108]

$$l_i = 0,48 A^{1/2} \left[1 - 1,25 \left(\frac{A_0}{A} \right)^{1/2} \right] \quad (2.18)$$

при условии $A_0 < 0,16 A$. Подобная поправка может иметь известное значение при расчете индуктивности прохода между зубами, индуктивности голосовой щели или входного импеданса гортанной трубки.

К точности расчета индуктивной составляющей импеданса излучения не следует предъявлять очень высокие требования, поскольку недостаточно определенным является действительное местоположение излучающей поверхности на выходе речевого тракта. Кроме того, поверхность эту вряд ли можно принимать просто за плоскость — форма ее, вероятно, в какой-то мере связана с кривизной губ и линии зубов и притом зависит не только от степени лабиализации, но и от частоты и формы передней полости. В расчетах, выполненных в настоящей работе, излучающая поверхность принята за плоскость, расположенную на расстоянии, не превышающем 0,5 см перед передними зубами; с этим допущением и производились расчеты для всех неогубленных гласных и согласных.

§ 2. Методы численного расчета

После того как речевой тракт представлен в виде эквивалентной цепной схемы, подобной приведенной на рис. 7, расчет может быть выполнен общепринятыми в теории цепей методами.

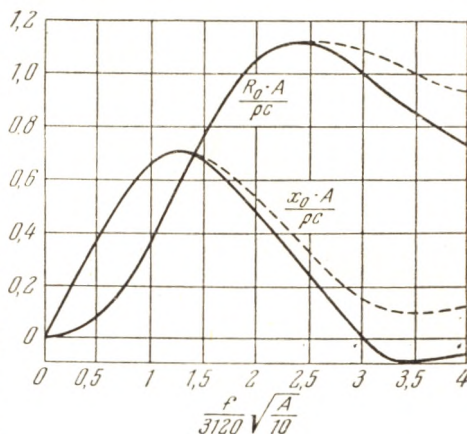


Рис. 6. Нормированное акустическое сопротивление $R_0 A / \rho c$ и нормированная акустическая индуктивность $x_0 A / \rho c$ в зависимости от нормированной частоты. Пунктирные кривые — нормированные активная и индуктивная составляющие импеданса излучения для случая малого кругового поршня, расположенного на поверхности сферы радиуса 9 см. Сплошные кривые — приближенные значения тех же величин, использованные при разработке LEA.

В случае сонорных звуков определению подлежит коэффициент передачи $\frac{U_0}{U_q}$, где U_0 — объемная скорость на выходе речевого тракта, а U_q — объемная скорость, создаваемая голосовым источником. Если источник звука расположен дальше по речевому тракту, то коэффициент передачи определяется как отношение U_0/E_v , где E_v — звуковое давление у источника. При этом источник E_v рассматривается как элемент,

включенный последовательно в продольное плечо схемы между v -м и $(v-1)$ -м звеном схемы¹⁾. Импеданс излучения обозначим через d_0 . Примем схему в виде цепочки симметричных Т-образных звеньев с импедансами a_n в продольных плечах и b_n в поперечных. Обозначив $a_n + b_n$ через d_n , можно получить следующие соотношения:

$$\frac{U_0}{U_q} = b_q \frac{\Delta_{0q}}{\Delta}, \quad (2.19)$$

$$\frac{U_0}{E_v} = \frac{\Delta_{0v}}{\Delta}, \quad (2.20)$$

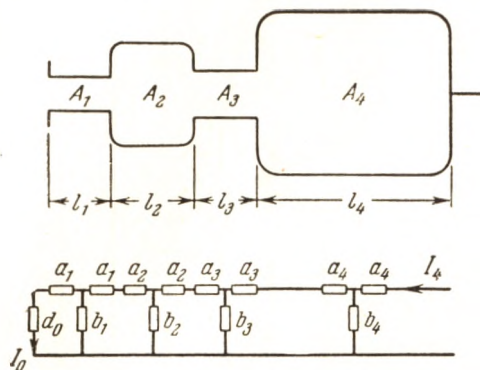


Рис. 7. Обобщенная эквивалентная схема четырехзвенного резонатора. Каждый из отрезков системы полостей представлен Т-образным звеном.

где Δ — главный определитель системы уравнений, относящихся к принятой схеме, а Δ_{0q} и Δ_{0v} — миноры.

Отсюда видно, что полюсы функции передачи речевого тракта определяются главным определителем, а возможные нули — минорами. Числитель $b_q \Delta_{0q}$ равен произведению импедансов всех поперечных элементов; отсюда видно, что в случае голосового источника коэффициент передачи не содержит нулей. Условие наличия нулей Δ_{0v} совпадает с условием бесконечного импеданса той части системы, которая расположена перед источником, считая по направлению от источника к голосовой щели.

Иллюстрация метода расчета приведена ниже на примере анализа четырехзвенной системы, представленной на рис. 7. Здесь, помимо голосового источника U_4 , включен источник E_3 , расположенный между звеньями 2 и 3.

Система уравнений для контуров 1, 2, 3 и 4 имеет вид:

$$(d_0 + d_1) U_0 - b_1 U_1 = 0; \quad (2.21a)$$

$$-b_1 U_0 + (d_1 + d_2) U_1 - b_2 U_2 = 0; \quad (2.21б)$$

$$-b_2 U_1 + (d_2 + d_3) U_2 - b_3 U_3 = E_3; \quad (2.21в)$$

$$-b_3 U_2 + (d_3 + d_4) U_3 = b_4 U_4, \quad (2.21г)$$

¹⁾ Именно такой способ включения E является существенным. Включение источника напряжения параллельно поперечному элементу цепной схемы невозможно, так как это означало бы короткое замыкание части системы, что нарушило бы ее структуру. Возможно параллельное включение источника тока; структура системы при этом сохранилась бы, но изменилась бы последовательность нулей по сравнению со случаем последовательного включения источника напряжения. В силу этого и такая возможность должна быть отброшена. Последовательное включение обеспечивает хорошее соответствие расчетных спектров с результатами измерений.

$$\Delta = \begin{vmatrix} d_0 + d_1 & -b_1 & — & — \\ -b_1 & d_1 + d_2 & -b_2 & — \\ — & -b_2 & d_2 + d_3 & -b_3 \\ — & — & -b_3 & d_3 + d_4 \end{vmatrix}, \quad (2.22)$$

$$\Delta_{0q} = \Delta_{04} = - \begin{vmatrix} -b_1 & — & — \\ d_1 + d_2 & -b_2 & — \\ -b_2 & d_2 + d_3 & -b_3 \end{vmatrix} = b_1 b_2 b_3, \quad (2.23)$$

$$\frac{U_4}{U_0} = \begin{vmatrix} (d_0 + d_1)/b_1 & -b_1/b_2 & — & — \\ -1 & (d_1 + d_2)b_2 & -b_2/b_3 & — \\ — & -1 & (d_2 + d_3)b_3 & -b_3/b_4 \\ — & — & -1 & (d_3 + d_4)/b_4 \end{vmatrix}, \quad (2.24)$$

$$\Delta_{03} = \begin{vmatrix} -b_1 & — & — \\ d_1 + d_2 & -b_2 & — \\ — & -b_3 & d_3 + d_4 \end{vmatrix} = b_1 b_2 (d_3 + d_4). \quad (2.25)$$

Нули определяются равенством $d_3 + d_4 = 0$.

Четырехзвенная схема была использована Дэнном [39] для расчета огибающих спектра гласных, причем были отброшены потери и было допущено, что звенья 1 и 3 являются просто индуктивностями. Расчеты четырехзвенной схемы с учетом потерь проводились автором [47] и Ван-ден-Бергом [8]. Представление звеньев 1 и 3 при всех этих расчетах в виде сосредоточенных элементов дает значительную ошибку для частот выше 3-й форманты. Полный анализ резонансной системы, состоящей из четырех отрезков труб, с использованием гиперболических функций, приведен в § 4 Приложения III.

Если при расчетах не учитываются потери, то полученные результаты должны быть дополнены полученными тем или иным путем данными о ширине формант. В случае пренебрежения потерями расчет по уравнениям (2.22)–(2.25) достаточно прост даже при большом числе звеньев. Присоединение каждого нового звена требует только двух дополнительных умножений и одного сложения. Если же основной задачей расчета является определение частот формант, столь же несложно, а может быть даже проще, вычислить импеданс, глядя со стороны голосовой щели; частоты формант после этого определяются обращением в бесконечность реактивной составляющей импеданса. Пренебрегая потерями, получим:

$$Z_{iq}(\omega) = jZ_q \operatorname{tg} \left[\frac{\omega l_q}{c} + \operatorname{arctg} \left[\frac{A_q}{A_{q-1}} \operatorname{tg} \left(\frac{\omega l_{q-1}}{c} + \dots \right. \right. \right. \\ \left. \left. \left. \dots + \frac{A_3}{A_2} \operatorname{tg} \left\{ \frac{\omega l_2}{c} + \operatorname{arctg} \left[\frac{A_2}{A_1} \operatorname{tg} \left(\frac{\omega l_1}{c} + \operatorname{arctg} \frac{Z_0}{Z_1} \right) \right] \right\} \dots \right] \right]. \quad (2.26)$$

При помощи номограммы, позволяющей определить $\operatorname{arctg}(m \operatorname{tg} \varphi)$, каждый рекуррентный шаг расчета сводится к двум операциям. Это — метод, который применялся для расчета резонансных мод колебаний в электрических объемных резонаторах сложной конфигурации.

Возможно использовать для расчета и другой метод — определение импеданса со стороны излучающего конца системы. Частоты формант

при этом определяются обращением этого импеданса, который обозначим через Z_{i0} , в нуль. Можно, наконец, разбив всю схему в какой-либо из узловых точек на две части, определить резонансные частоты из условия равенства нулю суммы реактивностей справа и слева от точки разбиения.

В общем случае, если учитываются и активные элементы схемы, коэффициент передачи является комплексной величиной:

$$\left. \begin{aligned} \frac{U_q}{U_0} &= \frac{1}{H(\omega)} = N(\omega) = N_b(\omega) + jN_a(\omega), \\ \varphi(\omega) &= \operatorname{arctg} \frac{N_a(\omega)}{N_b(\omega)}. \end{aligned} \right\} \quad (2.27)$$

Если выражение U_q/U_0 может быть получено в явной форме, то полюсы U_q/U_0 можно определить, подставив $j\omega$ вместо комплексной частоты $s = \sigma + j\omega$ и приравняв $N_b(s)$ нулю. Такой способ, однако, практически мало пригоден для схем с числом контуров, большим чем 2—3. При численных расчетах корни уравнения $N_b(s) = 0$ могут быть найдены приближенными методами. Совершенно очевидный метод заключается в вычислении абсолютных значений $\left| \frac{U_0}{U_q} \right|$ для до-

статочно большого числа значений независимой переменной ω ; построив по этим значениям кривую, легко определить по графику резонансные пики и ширину формант (по точкам на 3 дБ ниже пиковых уровней). Полученные таким образом значения резонансной частоты F_n и ширины резонансного пика B_n можно принять за приближенные значения круговой частоты ω_n и показателя затухания σ_n каждого из полюсов согласно соотношениям

$$\omega_n = 2\pi F_n, \quad \sigma_n = -\pi B_n. \quad (2.28)$$

Такого рода расчет эквивалентен измерению частоты формант и их ширины по амплитудным спектрам звуков речи. Однако при наличии даже малых потерь могут возникнуть значительные ошибки при определении параметров двух близко расположенных полюсов, соответствующих двойному пику. Два таких пика могут на графике слиться в один максимум, хотя два полюса никогда не могут совпасть, за исключением только предельного случая уменьшения площади сужения в речевом тракте до нуля.

Эффективный способ увеличения точности определения частот полюсов, предложенный Хаггинсом [104], состоит в определении частот, при которых вторая производная от фазы $\frac{d^2\varphi(\omega)}{d\omega^2}$ переходит через нуль.

Зависимость фазы от частоты для множителя, характеризующего один из резонансов и входящего в состав $H(j\omega)$,

$$H_1(j\omega) = \frac{\sigma_1^2 + \omega_1^2}{[j(\omega - \omega_1) + \sigma_1][j(\omega + \omega_1) + \sigma_1]}, \quad (2.29)$$

определяется равенством

$$\varphi_1(\omega) = \operatorname{arctg} \frac{\omega - \omega_1}{\sigma_1} + \operatorname{arctg} \frac{\omega + \omega_1}{\sigma_1}, \quad (2.30)$$

откуда

$$\frac{d\varphi_1(\omega)}{d\omega} = \frac{\sigma_1}{\sigma_1^2 + (\omega - \omega_1)^2} + \frac{\sigma_1}{\sigma_1^2 + (\omega + \omega_1)^2}, \quad (2.31)$$

$$\frac{d^2\varphi_1(\omega)}{d\omega^2} = 2\sigma_1 \left[\frac{\omega - \omega_1}{[\sigma_1^2 + (\omega - \omega_1)^2]^2} + \frac{\omega + \omega_1}{[\sigma_1^2 + (\omega + \omega_1)^2]^2} \right]. \quad (2.32)$$

Последнее выражение обращается в нуль при

$$\omega = \omega_1 \sqrt{\frac{2\omega_0}{\omega_1} - \frac{\omega_0^2}{\omega_1^2}}, \quad (2.33)$$

или

$$\omega_{1e} = \omega_1 \sqrt{2 \sqrt{1 + \frac{1}{4Q_1^2}} - 1 - \frac{1}{4Q_1^2}}, \quad (2.34)$$

где

$$\omega_0^2 = \omega_1^2 + \sigma_1^2 \quad \text{и} \quad Q_1 = \frac{\omega_1}{2\sigma_1}.$$

Значение ω_{1e} является хорошим приближением к действительной частоте полюса. Если $Q_1 > 1$, то декремент определяется соотношением

$$\left[\frac{d\varphi_1(\omega)}{d\omega} \right]_{\omega=\omega_{1e}} \approx \frac{1}{\sigma_1}. \quad (2.35)$$

Приведенный выше анализ может быть с успехом использован для определения частоты очень низко расположенной первой форманты, имеющей большую ширину; он применим и в случае двух близких по частоте более высоких формант, имеющих одинаковую ширину. В первом случае полюс, определяющий первую форманту, близок к своей сопряженной отрицательной паре.

Менее удобен способ определения полюсов путем вычисления частот, при которых фаза функции передачи равна $(2n-1)\frac{\pi}{2}$.

Условие, при котором два близких резонанса начинают сливаться в один максимум, состоит в том, что разность частот этих резонансов должна быть равна ширине полосы для каждого из них, взятого отдельно. Для случая очень низкочастотной первой форманты это означает, что частота F_1 должна быть равна половине ширины этой форманты. При таком условии влияние первой форманты на коэффициент передачи такое же, как и согласованного полувзвешенного низкочастотного фильтра-прототипа. Если Q первой форманты больше 1, частота резонансного пика ω_{pn} близка к круговой частоте ω_n полюса:

$$\omega_{p1} = \omega_1 \left(1 - \frac{1}{4Q^2} \right)^{1/2}, \quad (2.36a)$$

или

$$F_{p1} = F_1 \left(1 - \frac{1}{4Q^2} \right)^{1/2}. \quad (2.36b)$$

Эти соотношения показывают, что если принять частоту первой форманты равной 250 гц при ширине полосы 100 гц, то резонансный пик имеет место при частоте 245 гц. По приведенным выше соотношениям для частоты, при которой фаза сдвигается на 90° , найдем значение

255 гц , а частота, при которой обращается в нуль вторая производная, равна, как это следует из (2.32), 248,8 гц . Последнее значение ближе всего к истинной величине.

Эти расчеты, по существу, являются иллюстрацией известной теоремы, согласно которой фазовая характеристика минимально-фазовой цепи полностью определяется ее амплитудной характеристикой и наоборот. При анализе действительной речи высокоселективный фазовый анализ и методы, основанные на свойствах второй производной, находят лишь ограниченное применение, так как при использовании их трудно отличить небольшие нерегулярности или паразитные максимумы от действительных формант, характеризующих речевой тракт и входящих в состав F -картины. Это обстоятельство, впрочем, существенно только для устройств, предназначенных для автоматического «извлечения формант», поскольку в этих устройствах последовательным максимумам на спектре приписываются порядковые номера формант, независимо от происхождения этих максимумов, т. е. от того, являются ли они действительно формантами. Использование фазовой характеристики для определения полюсов возможно и при численных расчетах по заданной функции $\frac{U_0}{U_q} = H(j\omega)$, если заменить производные конечными разностями.

В настоящей работе для численных расчетов при определении полюсов использован другой метод. Он заключается в следующем: отбрасываясь от случая отсутствия потерь, т. е. полагая $N_a(\omega) = 0$, определяются частоты ω_{n1} , для которых $N_b(\omega) = 0$; показатель затухания σ_n и поправочный член $\Delta\omega_n$ к круговой частоте ω_n полюса s_n могут быть после этого получены путем использования линейной аппроксимации комплексной функции $N(j\omega)$ в окрестностях нуля:

$$s_n = j\omega_{n1} - \frac{N(\omega_{n1})}{N'(\omega_{n1})}, \quad (2.37)$$

$$N'(\omega_{n1}) = \left[\frac{dN(s)}{ds} \right]_{s=j\omega_{n1}} = \frac{1}{j} \left[\frac{dN_b(\omega)}{d\omega} + j \frac{dN_a(\omega)}{d\omega} \right]_{\omega=\omega_{n1}}, \quad (2.38)$$

$$\left. \begin{aligned} s_n &= \sigma_n + j(\omega_{n1} + \Delta\omega_n), \\ \sigma_n &= \frac{N_a N'_b}{N_a'^2 + N_b'^2}, \quad \Delta\omega_n = -\sigma_n \frac{N'_a}{N'_b}. \end{aligned} \right\} \quad (2.39)$$

В тех случаях, когда форманта характеризуется большим Q , поправочным членом $\Delta\omega_n$ можно пренебрегать.

Основная часть вычислительной работы, по крайней мере на первом этапе ее, падает на определение частот полюсов из равенства $N_b(\omega) = 0$. Имея дело со сложной схемой, желательно свести вычислительную работу к минимуму. Для этого полезно применить следующий прием: прежде всего намечаются предполагаемые более или менее вероятные значения резонансных частот; обозначим их через ω_{en} . Рассмотрим, например, определение $\omega_1 = 2\pi F_1$. По принятым значениям ω_{en} вычисляется $N_b(\omega_{e1})$ и на основании общих соотношений между частотами формант и огибающей спектра оценивается вероятная ошибка принятого значения ω_{e1} . С этой целью $N_b(\omega)$ представляется в виде произведения

$$N_b(\omega) = (1 - \omega^2/\omega_{e1}^2)(1 - \omega^2/\omega_{e2}^2)(1 - \omega^2/\omega_{e3}^2)(1 - \omega^2/\omega_{e4}^2) k_{r4}, \quad (2.40)$$

где k_{r4} — множитель, учитывающий влияние резонансов выше четвертого; определение этого множителя и его величина приведены в § 1 гл. 3. После дифференцирования получим:

$$N'_b(\omega) = \frac{2}{\omega_{e1}} \frac{N_b(\omega)}{(1 - \omega^2/\omega_{e1}^2)} \quad (\omega = \omega_{e1}). \quad (2.41)$$

Используя линейное приближение для выражения $N_b(\omega)$, полученного в результате решения контурных уравнений, при значении $\omega = \omega_{e1}$ и подстановке $\omega_1 = \omega_{e1} + \Delta\omega_1$, $\omega_2 = \omega_{e2} + \Delta\omega_2$, ..., где $\omega_1, \omega_2, \dots$ — истинные значения резонансных частот, можно найти следующую формулу для ожидаемого отклонения ω_{e1} от ω_1 :

$$\Delta\omega_1 = - \frac{N_b(\omega_{e1})}{N'_b(\omega_{e1})}, \quad (2.42)$$

причем $N'_b(\omega_{e1})$ находится из уравнения (2.41).

Этот способ был использован при проведении первой серии расчетов на вычислительной машине BARK¹⁾ в Стокгольме. Работа машины программировалась так, чтобы после нахождения ошибки и автоматического ввода нового исправленного значения вновь определить ошибку, и т. д. Сходимость этого процесса оказалась очень хорошей; для вычисления частот каждого из полюсов с ошибкой не более 1% достаточно было 2—5 циклов последовательного приближения.

Этот же метод, дополненный интерполяцией, можно рекомендовать для определения частот резонансов полной эквивалентной схемы и в тех случаях, когда нет возможности использовать для расчета быстродействующие вычислительные устройства. Для десятизвенной эквивалентной схемы речевого тракта расчет при этом не более сложен, чем расчет первых десяти гармоник по осциллограмме.

Быть может, фонетисты в настоящее время еще мало склонны пользоваться численными расчетами; однако тот, кто пожелал бы проверить математическим путем физиологические данные, полученные при помощи рентгена, при невозможности использовать электронные счетные устройства непрерывного действия может использовать описанный здесь метод вычисления.

§ 3. Общие уравнения преобразования для математического описания речевых сигналов

А. Функция передачи речевого тракта

Функция передачи речевого тракта для неназализованных звуков речи определяется отношением объемной скорости через ротовое отверстие к объемной скорости через голосовую щель, которое можно представить в виде

$$\frac{U_0}{U_q} = H(s) = \frac{K_G}{\prod_1^{\infty} \left(1 - \frac{s}{\hat{s}_n}\right) \left(1 - \frac{s}{\hat{s}_n^*}\right)} = \frac{K_G K_{rg}(s)}{\prod_{n=1}^g \left(1 - \frac{s}{\hat{s}_n}\right) \left(1 - \frac{s}{\hat{s}_n^*}\right)}, \quad (2.43)$$

где $\hat{s}, \hat{s}_n^*$ — сопряженные пары полюсов. Постоянная K_G учитывает потери, обусловленные колебаниями стенок ротовой полости; потери эти

¹⁾ Двоичная цифровая вычислительная машина Шведского совета по вычислительной технике в Стокгольме.

незначительны и потому можно принять $K_G = 1$. В уравнении (2.43) бесконечное произведение, соответствующее идеальному случаю распространения звуковой волны в одном измерении, заменено приближенным выражением, включающим только первые g полюсов, с поправочным множителем $K_{rg}(s)$, учитывающим влияние более высоких полюсов с частотами выше частоты полюса с порядковым номером g^1). При нулевой частоте $H(s)$ приближается к 1, как это и должно быть из условия непрерывности объемной скорости.

Если нужно учесть связь с подгортанной или носовой полостями или если источник расположен не в голосовой щели, а в более близких к ротовому отверстию участках речевого тракта, то в приведенное выражение войдет дополнительная «функция нулей» $H_z(s)$:

$$H(s) = H_p(s) H_z(s). \quad (2.44)$$

Здесь функция полюсов $H_p(s)$ тождественна с $H(s)$ в формуле (2.43), а функция $H_z(s)$ может быть записана в виде произведения, включающего масштабный множитель K_z :

$$H_z(s) = K_z \cdot s \prod_1^{\infty} \left(1 - \frac{s}{s_n}\right) \left(1 - \frac{s}{s_n^*}\right). \quad (2.45)$$

В случае носовых гласных звук распространяется частично по носовому тракту, причем часть эта (если пренебречь активными элементами тракта) равна $\frac{L_M}{L_N + L_M}$, где L_M — эффективное значение индуктивности всего ротового, а L_N — носового тракта, начиная от язычка:

$$L_M = \rho \int_{\text{язычок}}^{\text{губы}} \frac{dx}{A(x)}. \quad (2.46)$$

Если учесть активные элементы соответствующих трактов R_M и R_N , включенные последовательно с L_M и L_N , то масштабные множители для ротового и носового трактов определяются следующим образом:

$$\left. \begin{aligned} K_{zM} &= \frac{L_N}{L_N + L_M} \cdot \frac{s - s_N}{s - s_{NM}}, \\ K_{zN} &= \frac{L_M}{L_N + L_M} \cdot \frac{s - s_M}{s - s_{NM}} = 1 - K_{zM}, \end{aligned} \right\} \quad (2.47)$$

где

$$s_N = -\frac{R_N}{L_N}, \quad s_M = -\frac{R_M}{L_M}, \quad s_{NM} = -\frac{R_N + R_M}{L_N + L_M}. \quad (2.48)$$

Если $\frac{R_N}{L_N} > \frac{R_M}{L_M}$, то для очень низких частот имеет место некоторое преобладание звукопередачи через рот; однако этот эффект полностью компенсируется, если отдачи звука через носовое и ротовое отверстия складываются, так как $K_{zM} + K_{zN} = 1$. Конечный импеданс стенок речевого тракта оказывает шунтирующее действие, аналогичное действию носовой полости; вообще уравнения преобразования в том виде, как они приведены для учета влияния носового тракта, могут

¹⁾ Определение множителя $K_{rg}(s)$ дано Фантом [55]. Этот множитель играет существенную роль при синтезе речи, см. § 1 гл. 3.

быть отнесены и к случаю любой другой системы, шунтирующей ротовые полости. Аналогичные рассуждения можно использовать для определения множителя K_z , учитывающего влияние полостей, расположенных позади источника. При частотах, значительно более низких, чем сопряженные нули и полюсы всей системы, объемная скорость, развиваемая источником $E_v(s)$, определяется равенством

$$\lim_{s \rightarrow 0} U_v(s) = E_v(s) s C_b,$$

где

$$C_b = \frac{1}{\rho c^2} \int_{\text{источник}}^{\text{выход тракта}} A(x) dx. \quad (2.49)$$

Таким образом, C_b представляет собой емкость, эквивалентную полному объему всех полостей, расположенных до источника. В случае глухих звуков голосовые связки нельзя принимать за границу речевого тракта; трахею и легкие также следует считать входящими в его состав, хотя влияние их и сказывается в основном на низких частотах.

Поскольку объемная скорость, создаваемая источником, является также и объемной скоростью на выходе, $K_z = C_b$ и

$$H_z(s) = C_b \cdot s \prod_1^{\infty} \left(1 - \frac{s}{s_n}\right) \left(1 - \frac{s}{s_n^*}\right). \quad (2.50)$$

Как видно из этого выражения, функция $H_z(s)$ имеет единственный нуль в начале координат. Если источник расположен в сильно суженном проходе, влияние этого нуля и первой пары сопряженных полюсов функции $H_p(s)$ на достаточно удаленных частотах эквивалентно одному низкочастотному полюсу на вещественной оси, соответствующему внутреннему импедансу источника, в основном состоящему из последовательных индуктивности и активного сопротивления прохода.

Если все сопряженные нули $H_z(s)$ эффективно нейтрализуются полюсами $H_p(s)$, то задние полости могут быть вовсе исключены из рассмотрения, и вместо источника, характеризующегося заданным звуковым давлением, можно рассматривать источник заданной объемной скорости, которая определяется соотношением

$$U_v = \frac{E_v(s)}{R_v + sL_v}, \quad (2.51)$$

где $R_v + sL_v = Z_v(s)$ — импеданс суженного прохода.

Переход от источника заданной э. д. с. к источнику заданного тока (пользуясь здесь электротехническими терминами) полезен не только при рассмотрении звуков, создаваемых голосовой щелью, но и при расчете характеристик передачи передней (по отношению к источнику) полости речевого тракта в случае взрывных и фрикативных, если при этом соблюдается условие слабой связи между передними и задними полостями.

Б. Излучение, источник и другие постоянно действующие факторы

Переход от объемной скорости через ротовое отверстие $U_0(\omega)$ к звуковому давлению $P_1(\omega)$ в звуковом поле на расстоянии l от говорящего может быть выполнен приближенно с помощью следующего

энергетического соотношения:

$$U_0^2 R_0 = \frac{P_1}{\rho c} 4\pi l^2. \quad (2.52)$$

Предполагая, что излучатель находится в сферическом экране, имеем:

$$R_0 = \frac{\rho \omega^2}{4\pi c} K_S(\omega), \quad (2.53)$$

откуда получим:

$$\frac{P_1}{U_0} = \frac{\rho \omega}{4\pi l} \sqrt{K_S(\omega)}. \quad (2.54)$$

Однако принятое здесь предположение о независимости излучения от направления справедливо только для низких частот. Учет направленности излучения, а также того обстоятельства, что R_0 растет быстрее, чем ω^2 , может быть выполнен на основании данных Морза [155], относящихся к малому поршню кругового сечения на поверхности сферы:

$$\frac{P}{U_0} = \frac{\rho \omega}{4\pi l} K_T(\omega), \quad (2.55)$$

где $K_T(\omega)$ — полный поправочный множитель.

Частотная зависимость $20 \log_{10} K_T(\omega)$ для сферы радиусом 9 см приведена на рис. 8. Максимальное значение вносимой

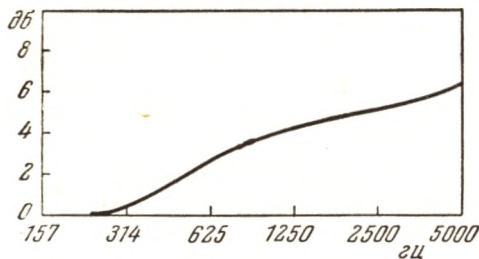


Рис. 8. Частотная зависимость осевой интенсивности звука, излучаемого малым участком поверхности сферы радиуса 9 см, учитывающая влияние направленности и отклонения частотной зависимости сопротивления излучения от пропорциональности ω^2 .

этим множителем поправки составляет 7 дБ на высоких частотах и 5 дБ на 2000 Гц. Приведенная кривая является, однако, только одной из постоянно присутствующих частотных зависимостей, которые нужно учитывать наряду с огибающей спектра источника и характеристикой излучения.

К таким постоянно действующим частотно-зависимым факторам, увеличивающим затухание на частотах, превышающих 5000 Гц, относится влияние морганиевых желудочков, образующих с ложными головчатыми связками добавочный резонатор [12], шунтирующее действие грушевидных пазух, расположенных с боковых сторон гортани, и влияние различных поперечных мод колебаний, при которых перемещение частиц происходит перпендикулярно к направлению воздушного потока. Отклонение сопротивления излучения, стремящегося при очень высоких частотах к значению $\rho c/A_0$, от приведенной выше зависимости от ω также оказывает некоторое, хотя и весьма незначительное, влияние. К другим второстепенным постоянно действующим факторам относится экранирующее действие головы и туловища, которое, однако, еще недостаточно хорошо изучено.

Характеристики голосового источника вместе с остальными постоянными зависимостями, снижающими отдачу на высоких частотах,

могут быть приближенно учтены четырьмя полюсами на отрицательной вещественной оси:

$$U_q = \frac{U_{q0}}{\prod_{r=1}^4 \left(1 - \frac{s}{s_r}\right)}. \quad (2.56)$$

Первые два из этих полюсов, характеризующих голосовой источник, равны приближенно $s_{r1} \approx s_{r2} = 2\pi \cdot 100$ гц; однако имеют место индивидуальные различия и зависимость от голосового усилия. Третий и четвертый полюсы приблизительно равны $s_{r3} = 2\pi \cdot 2000$ и $s_{r4} = 2\pi \cdot 4000$. Последние два полюса существенны при синтезе речи с помощью аналога речевого тракта в виде линии передачи; для формантного синтезатора они не имеют значения, так как такой синтезатор рассчитан на воспроизведение спектра только до 3500 гц.

В. Полное выражение преобразования Лапласа

Формулы (2.43), (2.55), (2.56) позволяют получить следующее полное выражение для идеального сонорного, неназализованного звука на расстоянии l см от губ:

$$P_d(s) = \frac{\rho s}{4\pi l} K_T(s) \frac{U_{q0}}{\prod_{r=1}^4 \left(1 - \frac{s}{s_r}\right)} \frac{K_{rg}(s)}{\prod_{n=1}^g \left(1 - \frac{s}{\hat{s}_n}\right) \left(1 - \frac{s}{\hat{s}_n^*}\right)}, \quad (2.57)$$

в котором учтено влияние направленности, характеристик излучения, характеристик голосового источника и другие факторы, обуславливающие подавление высоких частот. Каждая из g первых формант представлена парой сопряженных полюсов, влияние же более высоких формант учитывается множителем $K_{rg}(s)$. Выражение (2.57) представляет собой формальное описание процесса звукообразования в операторной форме и обосновывает целесообразность введения фильтровой функции, определяемой ниже формулой (3.11).

Излучение звука через ротовое отверстие для назализованных гласных определяется выражением

$$P_l(s) = \frac{\rho s}{4\pi l} K_T(s) \frac{U_{q0}}{\prod_{r=1}^4 \left(1 - \frac{s}{s_r}\right)} \frac{s - s_N}{s - s_{NM}} \frac{L_N}{L_N + L_M} \times \\ \times \frac{\prod_{b=1}^f \left(1 - \frac{s}{s_b}\right) \left(1 - \frac{s}{s_b^*}\right)}{\prod_{b=1}^f \left(1 - \frac{s}{\hat{s}_b}\right) \left(1 - \frac{s}{\hat{s}_b^*}\right)} \frac{K_{rg}(s)}{\prod_{n=1}^g \left(1 - \frac{s}{\hat{s}_n}\right) \left(1 - \frac{s}{\hat{s}_n^*}\right)}, \quad (2.58)$$

в котором влияние назализации представлено отдельным множителем.

Неназализованный звук, создаваемый источником $E(s)$ при конечной величине связи между частями схемы, расположенными перед источником и позади него, характеризуется выражением

$$P_l(s) = \frac{\rho s}{4\pi l} K_T(s) E(s) s C_b \frac{\prod_{n=1}^h \left(1 - \frac{s}{s_n}\right) \left(1 - \frac{s}{s_n^*}\right)}{\prod_{n=1}^g \left(1 - \frac{s}{s_n}\right) \left(1 - \frac{s}{s_n^*}\right)} \frac{K_{rg}(s)}{K_{rh}(s)}. \quad (2.59)$$

Множители $K_{rg}(s)$ и $K_{rh}(s)$ могут быть, очевидно, опущены в том случае, если в этих выражениях будет сохранено бесконечное число сопряженных полюсов и нулей.

В случае слабой связи передней и задней частей тракта те из g полюсов, которые зависят в основном от переднего резонатора, могут быть выделены в отдельный множитель. Следует отметить, что если источник расположен в очень узкой щели, самой низкой паре полюсов, характеризующих первую форманту, может соответствовать аperiodический процесс. При этом s_n и s_n^* отрицательны, вещественны и не равны друг другу.

Г. Обратное преобразование Лапласа

Обратное преобразование Лапласа $P(s)$ дает звуковое давление как функцию времени. Это преобразование для выражения (2.57), распространенное на случай стационарного периодического процесса, имеет вид¹⁾

$$P(t) = \sum_{m=0}^v \left\{ \sum_1^4 A_r e^{s_r(t-mT_0)} + \right. \\ \left. + (-1)^n \sum_{n=1}^g A_n e^{s_n(t-mT_0)} \cos[\omega_n(t-mT_0) + \varphi_n] \right\} \quad (2.60)$$

и представляет собой суперпозицию конечного числа затухающих колебаний и неperiodических экспоненциальных процессов, возбуждающихся при каждом из m последовательных импульсов голосового источника, начиная с $m=0$ до $m=v$, т. е. до последнего импульса в данном отрезке речи. Наличие множителя $(-1)^n$ показывает, что в электрическом синтезирующем устройстве с параллельно включенными контурами должна обеспечиваться противоположность фаз соседних контуров²⁾. Постоянные A_r , A_n и φ_n , очевидно, полностью определяются заданием системы полюсов и постоянно наличествующих частотных факторов, входящих в выражение для $P(s)$ ³⁾. Начальные амплитуды тесно связаны с амплитудами формант, как показано Фантом [55]⁴⁾.

Как правило, период основного тона $T_0 = \frac{1}{F_0}$, а также начальные амплитуды A_r и A_n изменяются от периода к периоду. Кроме подобных

¹⁾ Более детальное изложение см. в [55].

²⁾ Это было показано также Уэйбелом [207].

³⁾ См. также [66]. Соответствующие зависимости рассмотрены в § 3 гл. 7.

⁴⁾ В этой работе приведено также прямое и обратное преобразования для стационарных периодических звуков.

временных изменений, обусловленных нестабильностью источника, могут иметь место также постепенные изменения во времени частот полюсов s_r и $s_n = \sigma_n \pm j\omega_n$. Исходя из этого, можно определить смещение формант в пределах периода основного тона голоса.

В общем случае частоту и ширину полосы формант можно представить как функции времени в виде

$$F_n(t) = \frac{\omega_n(t)}{2\pi}, \quad B_n(t) = -\frac{\sigma_n(t)}{\pi}. \quad (2.61)$$

Такие изменения можно видеть на широкополосных спектрограммах, полученных на «Сонаграфе» для низкого мужского голоса при переходе, например, от губного звонкого взрывного к гласному. При узкополосном анализе такие быстрые изменения не могут быть обнаружены в силу большого времени усреднения фильтра. Изменения ширины полосы формант или, точнее говоря, величины σ_n в пределах одного периода основной частоты голоса можно видеть на осциллографических записях. Заметные изменения B_1 во времени могут иметь место вследствие зависимости величины сопротивлений голосовой щели и других очень узких проходов в речевом тракте от потока воздуха, а также вследствие наложения эффектов, связанных с нарушениями непрерывности потока через голосовую щель за время периода основной частоты голоса. Возможно, что пикфактор в пределах периода имеет более непосредственное значение для слухового восприятия, чем ширина полосы или постоянная затухания.

ГЛАВА 3

АНАЛИТИЧЕСКИЕ СВЯЗИ В СТРУКТУРЕ РЕЧЕВЫХ СПЕКТРОВ

§ 1. Идеализированное спектральное описание сонорных звуков

Относительная значимость формант сонорных звуков уменьшается с повышением их номера после F_2 . Форманты F_1 и F_2 являются основными характеристиками гласных¹⁾. F_3 и F_4 заметно влияют на фонетическое качество передних гласных; для задних гласных они менее важны. Наряду с фонетической информацией F_3 и F_4 содержат, так же как и F_0 , информацию об индивидуальных особенностях голоса диктора.

Из формулы преобразования Лапласа, приведенной в § 3 гл. 2, следует, что идеальный неназализованный гласный полностью и однозначно определяется характеристикой источника и данными о частоте и ширине полосы формант. Уровни формант полностью определяются этими данными; поэтому сведения об уровнях избыточны [52].

Количество переменных, необходимых для описания акустической картины речи, сводится, таким образом, вместо параметров 20—40 гармоник, характеризующих гармонический состав этой картины, к сведениям о четырех формантах и о характеристиках источника, как это вытекает из теории преобразования Лапласа. Сведения о ширине полосы формант до известной степени предопределяются значениями их частот, а характеристики источника являются более или менее

¹⁾ См. § 2 Приложения 1, где даны определения, пределы изменений и описана техника измерений.

постоянными для каждого диктора и зависят в основном от голосового усилия.

Таким образом, в качестве независимых переменных, определяющих сонорный звук, остаются частоты формант, т. е. F -картина (см. также § 3 гл. I). Большое значение формант для описания речи известно из практики анализа и синтеза звуков речи. Интересно, однако, отметить, что это значение подтверждается математическим анализом. Взаимозависимость между частотами формант и формой огибающей спектра имеет существенное теоретическое и практическое значение при решении вопросов анализа и синтеза речи. Конечно, фактически ни один из гласных не является идеальным; на спектральном разрезе можно наблюдать второстепенные небольшие пики, провалы и даже дополнительные «форманты». Они обусловлены особенностями реального спектра голосового источника, а также наличием связи между речевым трактом и подгортанными полостями. Одной из причин указанных искажений идеальной картины спектра является назальность; аналогичные явления могут явиться следствием взаимной модуляции между составляющими спектра в аппаратуре для записи и анализа речи.

Если желательно сохранить при описании все доступные наблюдения детали, следует использовать сведения о всех гармониках, несмотря на то, что им присуща значительная избыточность. Но и в этом случае полезно знать связь между частотами формант, т. е. F -картиной, и формой спектра, для того чтобы решить, что является формантой и что надо считать просто искажением.

Зависимость спектральной картины сонорных звуков от характеристик источника и фильтра можно представить, если отбросить фазовую информацию, в следующей форме:

$$|P(f)| = |U(f)| |H(f)| |R(f)|. \quad (3.1)$$

Здесь $|U(f)|$ представляет собой частотную характеристику источника¹⁾, $|H(f)|$ — частотную характеристику тракта и $|R(f)|$ — характеристику излучения, которая связывает объемную скорость через ротовое отверстие с давлением в звуковом поле. Произведение $|H(f)| |R(f)|$ представляет собой полную фильтровую функцию $T(f)$, см. формулу (1.3).

Целесообразно, однако, преобразовать выражение (3.1) таким образом, чтобы выделить и объединить все функции частоты, которые не изменяются от одного сонорного звука к другому. Выделенные спектр источника и характеристика излучения в совокупности образуют спектр, огибающая которого падает приблизительно на 6 дБ на октаву. $|R(f)|$ приблизительно пропорционально частоте f ; примем, что $|U(f)|$ приблизительно пропорционально $1/f^2$ на частотах выше 100 гц. Тогда можно написать:

$$|U(f)| |R(f)| = P_k \frac{(f/100)}{1 + (f/100)^2}, \quad (3.2)$$

где P_k — постоянный множитель, определяющийся данным значением уровня звукового давления.

¹⁾ Вместо символа $S(f)$, принятого для обозначения источника в обобщенном уравнении (1.3), здесь предпочтительнее использовать символ $U(f)$, обозначающий спектр объемной скорости источника.

Функция передачи речевого тракта $|H(f)|$, согласно формуле (2.43), может быть в свою очередь разложена на два основных множителя; из них один является поправочным множителем, который может быть отнесен к числу факторов, не изменяющихся от звука к звуку, а другой соответствует первым четырем формантам, характеризующим звук. Таким образом,

$$|H(f)| = k_{r4}(f) |H_1(f)| |H_2(f)| |H_3(f)| |H_4(f)|; \quad (3.3)$$

здесь $|H_1(f)|$ представляет собой резонансную частотную характеристику, соответствующую первой форманте, $|H_2(f)|$ — второй форманте, и т. д. Поправочный множитель $k_{r4}(f)$ учитывает влияние всех формант, расположенных выше четвертой, на ход частотной характеристики в нижележащей области частот. Конечно, можно ограничиться заданием в явной форме только трех первых формант; в этом случае влияние всех остальных формант будет учитываться другим поправочным множителем, который обозначим через $k_{r3}(f)$.

О значении множителей $k_{r4}(f)$ и $k_{r3}(f)$, которыми ни в коем случае не следует пренебрегать, можно судить по рис. 9. На этом рисунке приведена функция передачи трубы длиной 17,5 см, один конец которой открыт, а на другом конце находится источник с бесконечно большим сопротивлением. Такая труба представляет собой резонатор, воспроизводящий в идеализированном

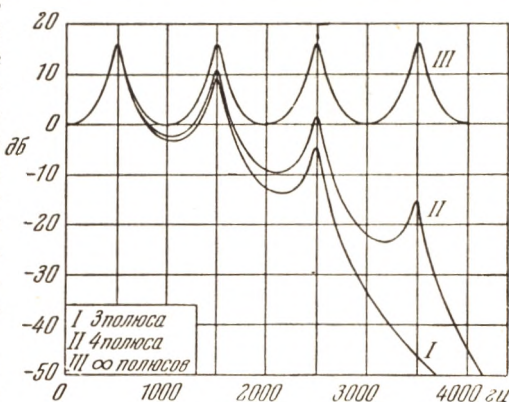


Рис. 9. Частотные характеристики отношения объемной скорости на выходе к объемной скорости на входе для идеализированного резонатора с формантными частотами, равными нечетным кратным от 500 гц, и с шириной полосы формант 100 гц. Кривая III — точное решение в соответствии с теорией линии передачи, короткозамкнутой на удаленном конце и разомкнутой на входе, т. е. в предположении наличия источника тока на входе. Кривая II — приближенное решение, учитывающее только четыре форманты, т. е. функция передачи системы с четырьмя полюсами. Кривая I — приближение, учитывающее только три полюса.

виде артикуляцию нейтрального открытого гласного. Частоты формант при такой артикуляции равны нечетным кратным от 500 гц, т. е. 500, 1500, 2500, 3500, ... гц; спектральный уровень во всех впадинах расположенных между формантными пиками, один и тот же. Потери приняты не зависящими от частоты, так что все форманты имеют одинаковую ширину полосы, которая принята равной 100 гц; уровни всех формантных пиков также одинаковы.

Если такую функцию передачи попытаться воспроизвести при помощи некоторого числа резонансных контуров с надлежащим образом выбранными частотами резонансов, то окажется, что спектральные уровни такой передающей системы быстро падают на высоких частотах, но хорошо согласуются с ходом частотной характеристики трубы на низких частотах.

Разность между ординатами кривой III на рис. 9, представляющей в логарифмическом масштабе $|H(f)|$, и ординатами кривой II, соответствующей в таком же масштабе произведению $|H_1(f)H_2(f)H_3(f)H_4(f)|$, дает множитель $k_{r4}(f)$. Этот множитель можно получить в аналитической

форме¹⁾), используя разложение в ряд и суммируя доли, вносимые 5-м и всеми остальными более высокими (до ∞) полюсами:

$$20 \log_{10} k_{r4} = 0,54x^2 + 0,00143x^4 \text{ (дб)}. \quad (3.4a)$$

Аналогично

$$20 \log_{10} k_{r3} = 0,72x^2 + 0,0033x^4 \text{ (дб)}, \quad (3.4б)$$

$$20 \log_{10} k_{r2} = 1,06x^2 + 0,0102x^4 \text{ (дб)}; \quad (3.4в)$$

здесь $x = \frac{f}{f_1}$ и $f_1 = \frac{c}{4l_{\text{полн}}}$; $l_{\text{полн}}$ — общая длина речевого тракта.

Постоянно присутствующие частотные зависимости характеризуются кривыми, приведенными на рис. 10. Кривая I на этом рисунке представляет частотную зависимость

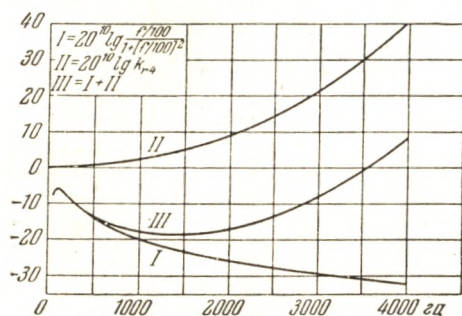


Рис. 10. Постоянно присутствующие частотные зависимости, определяющие спектр гласных, а именно: огибающая идеализированного спектра голосового источника, частотная зависимость излучения и влияние всех полюсов выше четвертого. I — сумма повышающейся на 6 дБ/окт характеристики излучения и падающей после 100 гц на 12 дБ/окт огибающей спектра идеализированного голосового источника; II — характеристика влияния всех полюсов, лежащих выше четвертого, на спектральные уровни при частотах ниже 4000 гц; III — общая частотная поправка к сумме резонансных кривых, соответствующих первым четырем формантам, которую необходимо вводить при расчете огибающей спектра гласного по формантным данным или при расчете этого спектра на четырехрезонансном аналоге речевого тракта в случае питания последнего генератором с равномерным спектром.

произведения $|U(f)R(f)|$ согласно формуле (3.2), кривая II — частотную зависимость множителя k_{r4} . Кривая III = I + II дает частотную характеристику всех постоянно присутствующих зависимостей, которая должна быть обеспечена в синтезирующем устройстве, состоящем из четырех формантных контуров, соединенных каскадно. Практическая необходимость учета этих факторов была продемонстрирована при помощи шведской «говорящей машины» OVE [50, 53].

С математической точки зрения понятию форманты, характеризующей качественную сторону спектральной картины, эквивалентно понятие полюса функции комплексной частоты. Частотные характеристики звука, согласно § 3 гл. 2, могут быть, исходя из этого, получены, если задан ряд точек, а именно полюсы и нули, в плоскости комплексных частот; эти точки показаны на рис. 11.

На этом рисунке обе координаты точек в плоскости умножены на $\frac{1}{2\pi}$ и оси координат повернуты на 90° по сравнению с обычным их расположением, для того чтобы ось, по которой откладываются значения действительных частот, была горизонтальна.

Каждая форманта на рисунке представлена парой сопряженных полюсов, соответствующих частоте форманты F_n , взятой со знаком плюс и минус. Ординаты на рисунке, т. е. величины $\frac{\sigma_n}{2\pi}$, дают половину ширины полосы форманты, т. е. $\frac{B_n}{2}$. Частотная характеристика голосового источника определяется полюсами, расположенными на оси «ширины полос» (ось ординат). В случае функции, характеризующей источник согласно равенству (3.2), ей соответствуют два полюса

¹⁾ Подробное см. [55].

(точнее — двукратный полюс) при частоте 100 гц. На рис. 11 полюсы источника находятся при 100, 2000, 4000 и 5000 гц. Такое расположение полюсов дает приближенно частотную характеристику источника, принятую Стивенсом и др. [196]¹⁾. Характеристика излучения определяется нулем в начале координат; формантные полюсы расположены при частотах, являющихся нечетными кратными от 500 гц, причем предполагается, что ширина полосы форманты увеличивается с частотой, как это обычно и имеет место в действительности.

После того как источник, излучение и передача по речевому тракту охарактеризованы точками в плоскости комплексных частот, любая

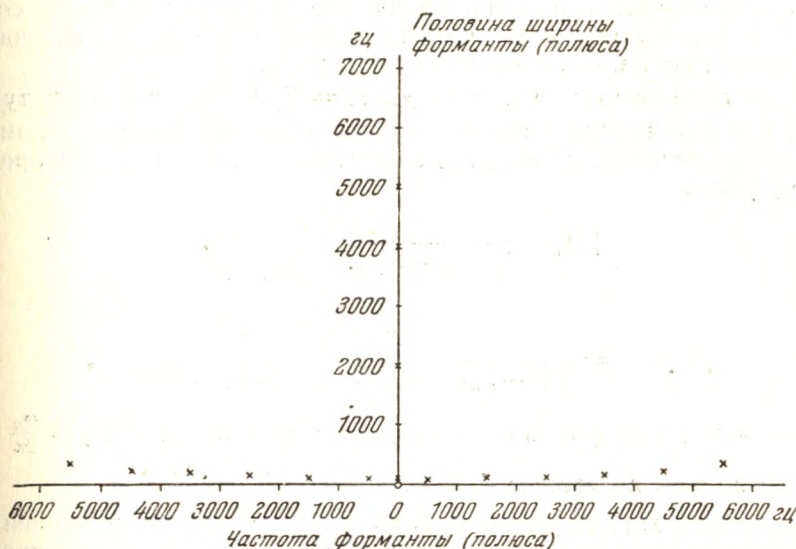


Рис. 11. Представление основных параметров, определяющих гласный, в плоскости комплексных частот. $\times$ — полюс, $\circ$ — нуль. Сопряженные полюсы характеризуют форманты идеализированного нейтрального гласного. Нуль при нулевой частоте является аппроксимацией характеристик излучения. Полюсы на отрицательной вещественной оси (оси «ширины формант») дают приближенную частотную характеристику источника, использованную в одной из ранних работ.

из функций $U(f)$, $H(f)$, $R(f)$, так же как и их произведение в формуле (3.1), легко могут быть получены при помощи следующих графических приемов.

Проводятся векторы, соединяющие все полюсы с точкой $(f, 0)$ на оси частот, а также векторы, соединяющие полюсы с началом координат; вычисляется произведение длин всех векторов, проведенных к точке $(f, 0)$, а также векторов, проведенных к началу координат; если нули в плоскости комплексных частот отсутствуют, то отношение второго произведения к первому дает функцию, характеризующую систему, или функцию системы; если нули имеются, функция, определяемая полюсами, делится на функцию, определяемую нулями, которая получается способом, совершенно аналогичным только что рассмотренному. Если нуль находится в начале координат, то соответствующий ему множитель получается несколько иным способом, а именно берется только расстояние от точки $(f, 0)$ до начала координат.

¹⁾ Как показали опыты синтеза на OVE, эта характеристика не даёт преимуществ по сравнению с характеристикой, определяющейся двукратным полюсом при 100 гц.

Идеальные спектры гласных характеризуются отсутствием нулей, кроме нуля в начале координат, который вносится характеристикой излучения. Как было отмечено раньше, это приводит к возрастанию амплитуд огибающей спектра на $+6$ дБ на октаву; при частотах выше 100 гц это возрастание компенсируется полюсом на оси «ширины полос», характеризующим источник. Влияние каждого из остальных полюсов, характеризующих систему, приводит к падению огибающей спектра на более высоких частотах, равному -6 дБ на октаву.

Полюсы и нули, соответствующие конечным действительным частотам, т. е. точкам, не лежащим на оси «ширины полос», представлены всегда сопряженными парами. Таким образом, каждой точке, соответствующей полюсу или нулю при положительной частоте, отвечает сопряженная точка при отрицательной частоте.

Пара сопряженных полюсов $s_n, s_n^* = F_n \pm j\omega_n$ определяет ту долю $H_n(f)$, которую вносит в результирующую кривую форманта номера n . Абсолютная величина этой доли, в соответствии с векторным представлением, равна

$$|H_n(f)| = \frac{s_n s_n^*}{|(s - s_n)(s - s_n^*)|_{s=j\omega}}, \quad (3.5a)$$

или

$$|H_n(f)| = \frac{F_n^2 + (B_n/2)^2}{\sqrt{(f - F_n)^2 + (B_n/2)^2} \sqrt{(f + F_n)^2 + (B_n/2)^2}}, \quad (3.5b)$$

где, как и раньше, форманта характеризуется частотой $F_n = \frac{\omega_n}{2\pi}$ и шириной ее полосы по уровню -3 дБ, равной $B_n = \frac{\sigma_n}{\pi}$.

Частотная характеристика $H_n(f)$ тождественна с функцией передачи напряжений в резонансном контуре RLC , определенной как отношение напряжения на конденсаторе к действующей в контуре э. д. с.

§ 2. Связь между частотами формант и огибающей спектра

Форма кривой $|H_1(f)|$ для первой форманты с частотой 250 гц и шириной полосы 100 гц показана на рис. 12. На низких частотах $|H_n(f)|$ стремится к 1, т. е. к 0 дБ, а при частоте форманты значение $|H_n(f)|$ весьма близко к

$$Q = \frac{F_n}{B_n}. \quad (3.6)$$

Как правило, Q для формант достаточно велико; это позволяет представить уравнение (3.5) в упрощенном виде:

$$|H_n(f)| = \frac{1}{\sqrt{(1 - x^2)^2 + x^2/Q^2}}, \quad (3.7)$$

где $x = \frac{f}{F_n}$ — относительная частота¹⁾. В силу того, что Q велико, та

¹⁾ Более точно, $x = \frac{f}{\sqrt{F_n^2 + (B_n/2)^2}}$ и $Q = \frac{\sqrt{F_n^2 + (B_n/2)^2}}{B_n}$.

часть резонансной кривой, которая соответствует значениям $x > \sqrt{2}$, может быть приближенно представлена в виде

$$|H_n(f)| \approx \frac{1}{x^2 - 1} \quad \text{при } x > \sqrt{2}, \quad (3.8)$$

и на более высоких частотах

$$|H_n(f)| \approx \frac{1}{x^2} \quad \text{при } x > 2,$$

т. е.

$$|H_n(f)| \approx \frac{F_n^2}{f^2} \quad \text{при } f > 2F_n. \quad (3.9)$$

Таким образом, при частоте $f = \sqrt{2} F_n$ величина $|H_n(f)|$, пройдя максимум, вновь падает до 1, т. е. до 0 дб; при еще более высоких частотах скорость спада составляет 12 дб/окт.

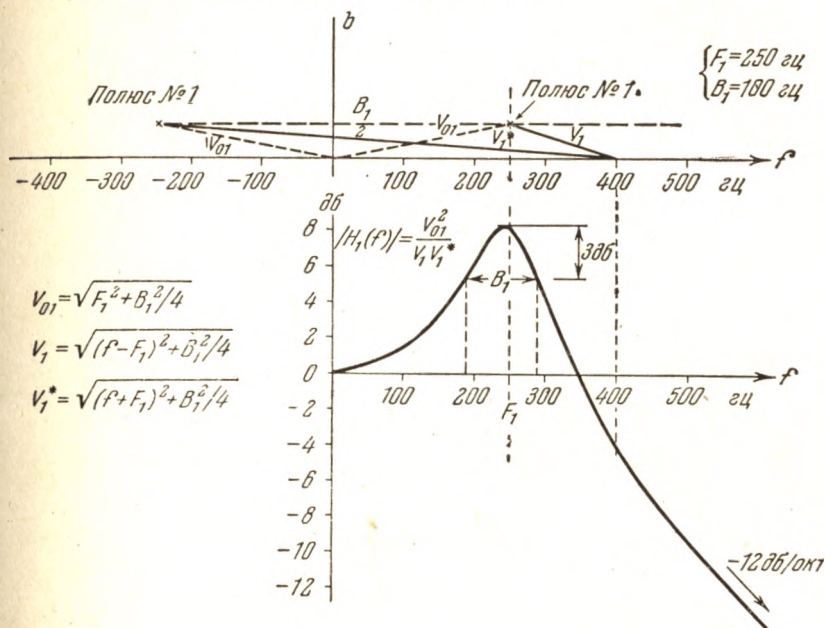


Рис. 12. Получение резонансной кривой, соответствующей форманте, по частоте и ширине последней, т. е. по двум величинам, характеризующим полюс функции передачи. Частные значения этих величин, принятые в настоящем примере, а именно частота форманты 250 гц и ширина 100 гц, можно считать типичными для первой форманты переднего гласного высокого подъема. Построение начинается с того, что на график наносятся два сопряженных полюса, т. е. точки $B_1/2, \pm F_1$. Резонансная кривая определяется как $20 \log_{10} (V_{01}^2 / V_1 V_1^*)$, где V_{01} представляет собой длину вектора, соединяющего начало координат с полюсами, а V_1 и V_1^* — длины векторов, проведенных от точки f на оси частот соответственно до положительного и отрицательного полюсов.

Поскольку $|H_n(f)|$ зависит не только от f , но и от F_n , то сдвиг F_n вниз на октаву, означающий возрастание x в два раза, приводит к уменьшению уровня огибающей спектра на всех частотах, отвечающих условию $x > 2$, на 12 дб; на частотах, близких к первоначальной частоте форманты, но превышающих последнюю, уменьшение значительно больше. Это — основное правило, определяющее взаимосвязь между частотами формант и формой огибающей спектра.

Построение спектральной картины идеального гласного при помощи частотно зависимых величин $|H_1(f)|$, $|H_2(f)|$, $|H_3(f)|$ и $|H_4(f)|$, описываемых равенствами (3.7), и неизменного для всех гласных множителя $k_{r4}(f) |U(f)| |R(f)|$, определяемого равенствами (3.2) и (3.4а), показано на рис. 13. Этот рисунок соответствует F -картине для гласного,

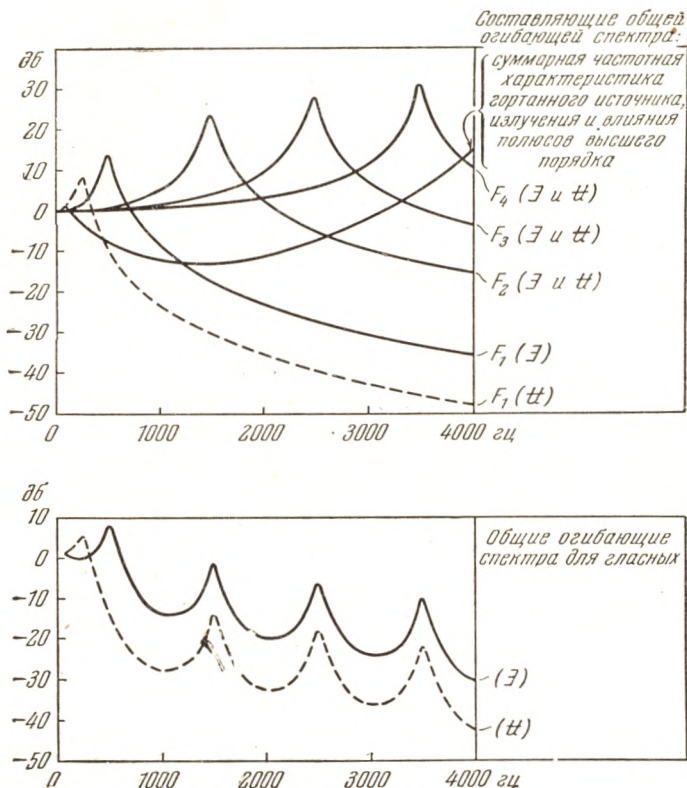


Рис. 13. Влияние сдвига частоты первой форманты на уровни огибающей спектра, показанное на примере двух гласных. Огибающие спектра этих гласных получены путем суммирования резонансных кривых. [э] представляет собой идеализированный нейтральный гласный, соответствующий представлению речевого тракта в виде открытой трубы; частоты формант такого гласного 500, 1500, 2500, ... гц. Сдвиг одной только первой форманты с 500 до 250 гц дает новый звук, близкий к [а]. Огибающая спектра этого звука выше 500 гц по форме практически совпадает с огибающей первого звука, но уровни на 12 дб ниже.

частоты формант которого равны нечетным кратным $F=500$ гц; здесь показано также изменение, происходящее при сдвиге F_1 с 500 гц до 250 гц, при сохранении прежнего частотного положения остальных формант. Отдельные сомножители, входящие в выражения (3.1) и (3.3), выражены в дб и произведения заменены суммами логарифмов, т. е. осуществлено суммирование спектральных уровней, выраженных в дб:

$$20 \log_{10} |P(f)| = 20 \log_{10} k_{r4}(f) |U(f) R(f)| + 20 \log_{10} |H_1(f)| + \\ + 20 \log_{10} |H_2(f)| + 20 \log_{10} |H_3(f)| + 20 \log_{10} |H_4(f)|, \quad (3.10)$$

или

$$20 \log_{10} |P(f)| = 20 \log_{10} k_{r4}(f) P_k \frac{f/100}{1 + (f/100)^2} - \sum_{n=1}^4 10 \log_{10} \left[\left(1 - \frac{f^2}{F_n^2} \right)^2 + \frac{f^2}{F_n^2 Q_n^2} \right] \quad (\text{дб}). \quad (3.11)$$

Ширина полосы для всех формант принята равной 100 гц.

Внизу на рис. 13 представлена полученная расчетным путем огибающая спектра идеальной нейтральной гласной, обозначенной символом [э], которая характеризуется частотами формант $F_1 = 500$ гц, $F_2 = 1500$ гц, $F_3 = 2500$ гц и $F_4 = 3500$ гц, а также огибающая спектра, получающаяся при сдвиге F_1 до 250 гц. Последний спектр определяет звук, фонетическое качество которого близко к [у], как в норвежском «hus». Обращает на себя внимание разница на 12 дб в уровнях огибающей для [у] по сравнению с [э] при частотах, значительно превышающих F_1 ; суммарная интенсивность при сдвиге F_1 вниз также оказывается меньшей.

Метод расчета, использованный при построении огибающей на рис. 13, был в дальнейшем усовершенствован с целью выполнения расчета спектров сонорных звуков по заданным частотам формант. При этом было признано целесообразным не исходить из полученных в результате эксперимента сведений о частотах формант, а провести математический синтез огибающих при систематическом изменении этих частот. Ступени изменения составляли 250 гц и 125 гц для F_1 , 250 гц и 500 гц для F_2 и 500 гц для F_3 ; частота F_4 была принята постоянной и равной 3500 гц. Фонетические символы, которыми отмечены полученные расчетным путем огибающие на рис. 14, представляют собой только указание на ближайший звук из алфавита Международной фонетической ассоциации (МФА), так что приведенные на рисунке спектры не претендуют на то, чтобы представлять собой физическую интерпретацию стандартных символов МФА.

Существенно, однако, то, что самая форма полученных спектров обнаруживает большое сходство со спектрами таких естественных звуков, частоты формант которых близки к принятым при расчете; далее, при анализе естественных звуков было установлено, что теоретические выводы правильно отражают те изменения уровней формант, которые обусловлены изменениями F -картины этих звуков [52].

Используя взаимосвязь между частотами формант и их относительными уровнями, можно решать два типа задач. Примером первого типа, математически более простого, является рис. 13. Он отвечает на вопрос: как изменяется уровень огибающей на некоторой заданной частоте в случае сдвига по частоте одной или нескольких формант. Решение такой задачи получается весьма просто, если использовать элементарную резонансную кривую или, точнее, фильтровую функцию, соответствующую форманте и описываемую равенством (3.7) или приближенными равенствами (3.8) и (3.9).

Задачи второго типа, более сложные, преследуют цель определения уровня формантного пика при изменении частоты форманты. Сложность этого типа задач связана с тем, что приходится учитывать расстояния сдвигаемой форманты по частоте от всех остальных. В таких случаях решение легче всего получить графически, для чего форманты представляются точками в комплексной плоскости частот, и от этих

точек проводятся векторы до точки, соответствующей интересующей нас частоте; такое построение было рассмотрено выше и вытекает из основной формулы (3.5) для элементарной фильтровой функции. При этом необходимо учитывать все постоянно присутствующие зависимости — спектр голосового источника, влияние излучения и влияние всех более высоких, по отдельности не рассматриваемых полюсов. Суммарное влияние этих зависимостей иллюстрируется рис. 13.

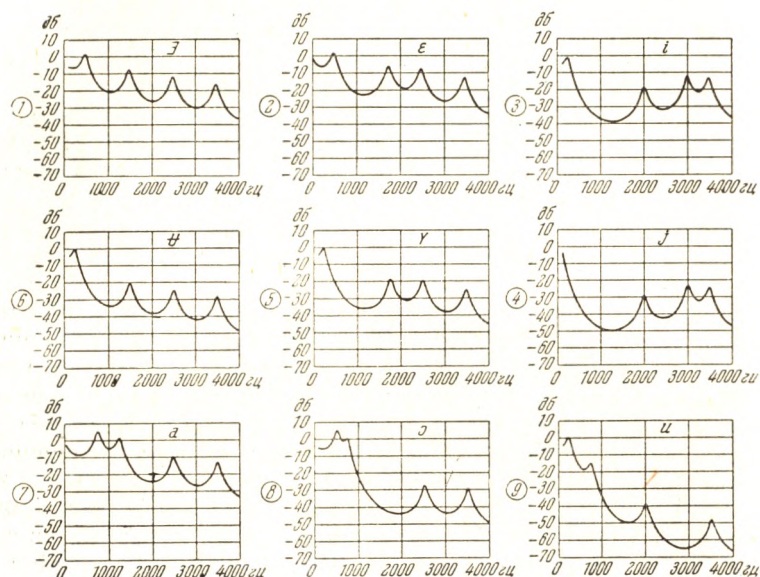


Рис. 14. Огибающие спектров гласных, полученные расчетным путем по элементарным резонансным кривым, соответствующим каждой из формант. При расчете принят источник со спаданием огибающей спектра 12 дБ/окт. Ширина формант принята равной 100 Гц. Для каждой из огибающих показан наиболее подходящий фонетический символ из алфавита МФА; однако приведенные спектры не предлагаются считать нормой фонетического качества. Эти спектры приведены лишь как иллюстрация возможности предсказания формы спектра и спектральных уровней по частотам формант.

На основании такого метода анализа можно сделать ряд общих выводов:

1) Сдвиг форманты F_n по частоте в основном влияет на уровни огибающей на частотах, лежащих выше F_n . На этих частотах уровни увеличиваются на 12 дБ при каждом смещении форманты F_n вверх на одну октаву [см. уравнение (3.8) и рис. 13 и 14]. Влияние сдвига форманты можно сформулировать в виде «правила низкочастотного фильтра»: каждая форманта действует как низкочастотный фильтр с граничной частотой $\sqrt{2} F_n$, где F_n — частота форманты. Влияние сдвига F_1 в сторону более высоких частот можно видеть, сопоставляя огибающие для [э], [у], [е], [ү] и [и], [і].

2) Если две сравнительно близкие форманты сближаются, то уровни обоих пиков, соответствующих этим формантам, увеличиваются на 6 дБ, а уровень впадины между ними — на 12 дБ при каждом уменьшении расстояния между формантами вдвое. Если две форманты сходятся очень близко, остается только один видимый максимум, как показано в § 2 гл. 4. Результат возрастания F_1 при одновременном уменьше-

нии F_2 можно видеть, если сравнить кривую для [а] с одной из кривых для [ε] или [Э].

3) Изменения интенсивности первой форманты в результате сдвига ее частоты F_1 не очень велики, если только F_1 не подходит очень близко к F_2 . Увеличение L_1 на 6 дб/окт, которое могло бы иметь место вследствие увеличения Q_1 при возрастании F_1 и неизменном B_1 , почти полностью компенсируется ходом суммарной характеристики источника и излучения, отличающейся падением на 6 дб/окт. Влияние более высоких полюсов несколько уменьшает падение этой суммарной характеристики на частотах выше 500 гц; поэтому в этой области частот имеет место некоторое увеличение L_1 при возрастании F_1 .

4) Интегральная интенсивность синтетических звуков, показанных на рис. 14, определяется главным образом F_1 ; первая форманта во всех случаях дает больший вклад в интегральную интенсивность, чем любая другая. С другой стороны, с точки зрения восприятия громкости значимость второй форманты больше, однако и в этом случае доминирующей остается первая, если только уровень громкости звука не очень мал. Располагая звуки в порядке возрастания интегральной интенсивности, можно получить ряд [а], [э], [ε], [Э], [i], [u], [ч], [y], [j], который характерен и для естественных звуков.

Артикуляторное сужение вызывает сдвиг F_1 вниз, чему соответствует уменьшение интенсивности, особенно существенное для более высоких формант. Кривая 4 на рис. 14 отмечена символом согласного [j] по той причине, что она отличается от кривой для [i] в основном более низкой F_1 ; наличие шумовых составляющих не обязательно для идентификации фонемы [j]. Так, слог [ji] оказалось возможным синтезировать на установке ОВЕ одним только подбором частот формант при соотношении интенсивностей, определяемом приведенными выше аналитическими зависимостями. Этого достаточно и для слогового деления в фразах вроде «How are you», «I love you» [50, 53].

Рассмотренные выше общие взаимосвязи между частотами формант и их уровнями полезны при анализе основных спектральных особенностей и дифференциальных признаков звуков с четкой формантной структурой, а именно «компактности» и противопоставленной ей «диффузности», а также «низкотональности» и противопоставленной ей «высокотональности», в соответствии с определением этих терминов, данным Якобсоном и др. [114] и позже Якобсоном и Халле [115]. Такой анализ был проведен в одной из прежних работ [52].

Суждение о компактности или низкотональности относительно, т. е. возможно только в результате сравнения двух противопоставленных по этому признаку звуков. В идеале, однако, для признаков звуков следует выявлять атрибуты, поддающиеся точному безотносительному количественному определению; иначе говоря, необходимо найти параметр, являющийся как бы общим знаменателем для всех рассматриваемых минимальных пар. Это желательно, но не всегда достаточно, например, если ставится задача детального фонетического описания.

Компактность гласных, определяемую как степень концентрации энергии спектра в области 1000 гц, не легко выразить количественно, несмотря на кажущуюся простоту этого определения. Можно было бы принять в качестве одной из границ компактности гласный [а], у которого этот признак наиболее сильно выражен. Аналогично гласный [i], как наиболее высокотональный, можно было бы считать границей признака высокотональности.

Однако выбор таких физических отправных точек приводит к неизбежной зависимости между этими двумя признаками в том смысле, что изменение параметра, характеризующего один из признаков, связано с изменением другого признака. Максимально высокотональный звук оказывается и максимально диффузным, т. е. минимально компактным.

С лингвистической точки зрения такая зависимость несущественна. Действительно, если высокотональные фонемы (передние гласные) отделены от низкотональных (задних гласных), нет смысла использовать дальше противопоставление по признаку высокотональный — низкотональный. Признак же компактности, характеризующий степень открытости, может быть использован для дальнейшего различения звуков в пределах передних и задних гласных.

Однако с инженерной точки зрения желательно выбрать физические критерии таким образом, чтобы они удовлетворяли условию ортогональности, или, иными словами, найти независимые спектральные атрибуты и параметры (см. также часть III).

Одним из примеров использования ортогональной системы параметров является выбор сдвига F_1 в сторону возрастания в качестве критерия увеличения компактности и сдвига F_2 в сторону убывания в качестве критерия увеличения низкотональности.

Для перехода от сдвига частот формант к изменениям формы спектра можно воспользоваться приведенным выше правилом (1) в тех случаях, когда нас интересуют только наиболее существенные особенности спектра. Из этого правила следует, что компактность, если определять ее только по значению F_1 , проявляется в основном в увеличении уровня интенсивности на всех участках спектра. Если использовать еще и правило (2), то можно видеть, что повышение уровня интенсивности в области частот между F_1 и F_2 особенно ясно выражено при сближении F_2 и F_1 .

Основные особенности спектра, связанные с признаком низкотональности, оцениваемой по значению F_2 , также могут быть получены на основании правила (1). Сдвиг F_2 вверх повышает уровень интенсивности всех вышележащих формант, а сдвиг F_2 вниз дает обратный эффект. При этом повышение уровня первой форманты, в соответствии с правилом (2), заметно только в том случае, когда F_2 при сдвиге вниз подходит близко к F_1 . Таким образом, в спектральной картине низковоисотональность проявляется как сдвиг основной энергии спектра вниз — вверх по частоте. Следует, однако, отметить, что этот сдвиг обусловлен не только изменением частоты второй форманты, но и изменениями уровня других формант [52].

Другой возможной системой ортогональных параметров являются величины $(F_1 + F_2)$ и $(F_2 - F_1)$, которые получаются поворотом осей координат (F_1, F_2) на 45° . Величина $(F_2 - F_1)$, т. е. расстояние по частоте между второй и первой формантами сонорного звука, является мерой рассредоточения энергии по спектру. Величина $(F_1 + F_2)/2$ характеризует приближенно «центр тяжести» спектра.

Предельным случаем значений параметра $(F_2 - F_1)$, соответствующим малому рассредоточению, является, согласно правилу (2), одноформантный звук, а предельным случаем малых значений $(F_1 + F_2)/2$ является, согласно правилу (1), звук с сильно выраженной низкотональностью или, пользуясь специфическим термином, с весьма низким центром тяжести.

Как показано в части III, такая система параметров имеет некоторые преимущества по сравнению с системой параметров F_1 и F_2 ; ее, однако, не следует рассматривать как непосредственное количественное выражение признаков компактности и низкотональности. Артикуляторным коррелятом высокого/низкого ($F_2 - F_1$) является сдвиг языка назад/вперед, причем одновременное приближение его к нёбу дает максимальные значения ($F_2 - F_1$). Основным артикуляторным коррелятом низкого ($F_1 + F_2$) является огубление, однако известное влияние на значение этого параметра оказывает сдвиг языка назад.

§ 3. Представление спектра согласных с помощью нулей и полюсов

Как видно из уравнений (2.58) и (2.59), наличие связи с носовой полостью или же с подгортанной областью, включающей трахею и легкие, характеризуется дополнительным множителем, который учитывает появляющиеся вследствие такой связи добавочные нули и полюсы. Относящаяся к надгортанной части речевого тракта функция $H(f)$ в формуле (3.1) должна быть, таким образом, умножена на функцию $N(f)$, отражающую влияние шунтирующих полостей.

В случае связи с носовой полостью сопряженные полюсы этой функции характеризуют носовые форманты, однако каждая такая форманта сопровождается антирезонансом, т. е. парой сопряженных нулей функции $N(f)$. Влияние на частотную характеристику одного из полюсов и одного из нулей определяется произведением двух функций: $|N_{pn}(f)|$, соответствующей полюсу, и $|N_{zn}(f)|$, соответствующей нулю. Принимая потери малыми, так же как это было сделано в связи с уравнением (3.7), найдем:

$$|N_n(f)| = |N_{pn}(f)| |N_{zn}(f)| = \left[\frac{(1 - x_{zn}^2)^2 + x_{zn}^2/Q_{zn}^2}{(1 - x_{pn}^2)^2 + x_{pn}^2/Q_{pn}^2} \right]^{1/2}, \quad (3.12)$$

где

$$x_{pn} = \frac{f}{F_{pn}}, \quad x_{zn} = \frac{f}{F_{zn}}, \quad Q_{pn} = \frac{F_{pn}}{B_{pn}}, \quad Q_{zn} = \frac{F_{zn}}{B_{zn}},$$

причем величины F_{pn} и B_{pn} характеризуют полюсы, а F_{zn} и B_{zn} — нули, определяющиеся носовой полостью.

Если частота и ширина полосы какого-либо полюса и нуля совпадают, то произведение $|N_{pn}(f)| |N_{zn}(f)|$ равно 1, т. е. происходит полная взаимная компенсация этого полюса и нуля. Если различие полюса и нуля по частоте и ширине полосы невелико, то в результате их суммарного действия сохраняется в общем плавная кривая, если не считать небольшого местного максимума или минимума или и максимума, и минимума вблизи каждой из критических частот. Пример такого случая представлен на рис. 15, где показаны значения функции $N_n(f)$, содержащей одинаковые по ширине полосы, но немного разнящиеся по частоте полюс и нуль. После пика и впадины функция на высоких частотах асимптотически приближается к некоторой предельной величине, которая определяется, согласно формуле (3.12), выражением

$$20 \log_{10} |N_{pn}(f)| |N_{zn}(f)| \approx 20 \log_{10} \left(\frac{F_{pn}}{F_{zn}} \right)^2 \text{ дБ при } f \gg F_{pn}. \quad (3.13)$$

Общая форма спектра гласной не может быть существенно искажена

подобного рода множителем. Однако по мере увеличения связи с системой шунтирующих полостей полюсы и нули все более расходятся по частоте, что приводит к более существенному влиянию их на форму спектра. Вдобавок при увеличении связи несколько сдвигаются по частоте и некоторые полюсы функции $H(f)$, соответствующей системе ротовых резонаторов. Эти особенности рассмотрены более подробно в § 3 гл. 8 в связи с расчетами для назализованных гласных.

Статистически среднее расстояние между формантами системы полостей, общая длина которой равна l , определяется величиной $c/2l$, где c — скорость распространения звука. Поэтому, если принять длину речевого тракта равной 17,5 см,

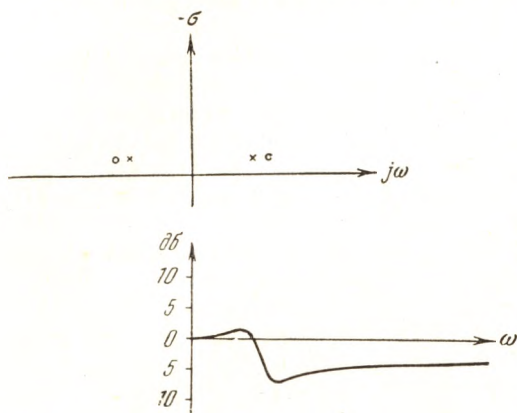


Рис. 15. Нуль и полюс в плоскости комплексных частот и соответствующая такой паре частотная характеристика. При сближении нуля и полюса они могут полностью скомпенсировать друг друга.

то среднее расстояние между формантами, обусловленными ротовой полостью и, следовательно, входящими в состав F -картины, будет равно приблизительно 1000 гц. Аналогично среднее расстояние между носовыми формантами равно $c/2l_n$, где l_n — общая длина носовой полости от язычка до выходного отверстия ноздрей. Точно так же среднее расстояние между нулями должно быть равно $c/2l_n$, где l_n — полная осевая длина шунтирующей системы. Полюсы и нули, отвечающие этой системе, следуют вдоль шкалы частот, чередуясь друг с другом.

Соображения, подобные только что приведенным, справедливы и при учете влияния полостей, расположенных перед источником и позади него при образовании фрикативных и взрывных звуков. Среднее расстояние между нулями в этих случаях равно $c/2l_b$, где l_b — длина всей системы полостей от места расположения источника до голосовой щели.

Если передние и задние полости очень четко разграничены, т. е. связь между ними мала, каждый из нулей задних полостей почти совпадает с соответственным полюсом; это значительно уменьшает уровень форманты, определяемой таким полюсом, т. е. уровень одной из формант задней полости. Поэтому при слабой связи фильтровая функция речевого тракта с достаточно хорошим приближением определяется только формантами системы передних полостей. В состав этой системы обычно входит узкий проход в месте артикуляторного сужения. Степень связи между передними и задними полостями частотно зависима, причем эта связь постепенно возрастает с возрастанием частоты. Тем не менее общая форма фильтровой функции речевого тракта может по-прежнему определяться в основном резонансами передних полостей, форманты же задних полостей, или, точнее говоря, форманты речевого тракта в целом, часто выявляются только в виде тонкой структуры спектра на фоне широких формантных областей, обусловленных передними полостями. В подобных случаях плотность распределения фор-

мант по шкале частот, т. е. среднее расстояние между формантами, может явиться критерием, позволяющим отнести ту или иную форманту к системе полостей всего речевого тракта в целом или только к передней его части. Примеры явлений такого рода приведены в § 3 гл. 10.

ГЛАВА 4

F-КАРТИНЫ СЛОЖНЫХ ТРУБЧАТЫХ РЕЗОНАТОРОВ И РУПОРОВ

§ 1. Резонатор, состоящий из двух отрезков труб. Влияние огубления

Графический метод определения резонансных частот системы, состоящей из двух отрезков труб, иллюстрируется на рис. 16. Искомые частоты находятся из того условия, что на границе переднего отрезка 1

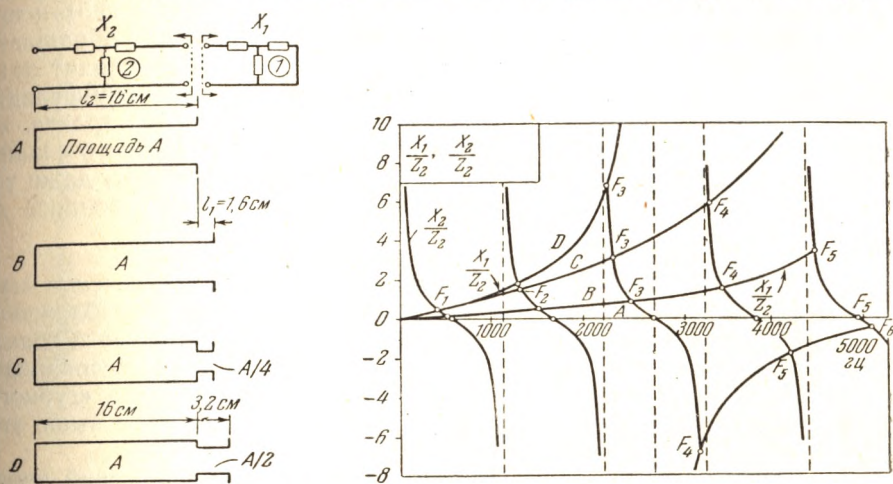


Рис. 16. Эффект удлинения передней части или уменьшения ее площади для резонатора в виде одиночного отрезка трубы. Частоты формант определяются условием $X_1 = -X_2$ (здесь $X_1 = Z_1 \tan \frac{2\pi f l_1}{c}$, $X_2 = -Z_2 \cotg \frac{2\pi f l_2}{c}$), т. е. точками пересечения кривых, представляющих собой частотные зависимости индуктивного сопротивления для основной и «губной» частей трубы. Большее выпячивание или округление губ приводит к понижению частот всех формант, лежащих ниже некоторой граничной частоты; последняя определяется четвертьволновым резонансом «губной» части. Это правило справедливо при любой конфигурации основного резонатора.

и заднего 2 сумма реактивностей, получающихся, если смотреть вправо и влево от этой границы, должна быть равна нулю или

$$Z_1 \tan \frac{\omega l_1}{c} = Z_2 \cotg \frac{\omega l_2}{c}. \quad (4.1)$$

Частоты резонансов определяются точками пересечения кривых, соответствующих правой и левой частям равенства. Если источник расположен между обоими отрезками, в функции передачи появляются нули, частоты которых определяются обращением в бесконечность импеданса заднего отрезка.

Если передний отрезок отсутствует, то резонансы определяются точками пересечения кривой, представляющей реактивное сопротивление заднего отрезка, с осью частот; однако при этом нужно всегда

учитывать концевую поправку. Наличие обусловленного этой поправкой индуктивного элемента, а также короткого, не очень узкого переднего отрезка приводит к убыванию резонансных частот по линейному закону

с коэффициентом $\frac{l_2 + l_{1e} \frac{A_2}{A_1}}{l_{2e}}$; эффективные длины $l_{1e} = l_1 + l_{1t}$ и $l_{2e} = l_2 + l_{2t}$ включают обычные концевые поправки. Импеданс переднего отрезка обращается в бесконечность при частоте, для которой

$$\operatorname{ctg} \frac{\omega l_1}{c} = \frac{\omega l_{1t}}{c}, \quad (4.2)$$

или приближенно

$$f = \frac{c}{4l_{1e}}.$$

Независимо от фактической конфигурации полостей речевого тракта, его входное реактивное сопротивление (глядя со стороны губ) непрерывно растет с частотой, за исключением тех точек, где оно обращается в бесконечность; в этих точках сопротивление скачком переходит от положительных к отрицательным значениям. Обратная по знаку функция, очевидно, непрерывно падает; поэтому удлинение речевого тракта, что эквивалентно присоединению добавочного отрезка у «губного» конца, неизбежно приводит к сдвигу вниз всех резонансных частот, даже тех, которые расположены выше частоты четвертьволнового резонанса добавленного отрезка, равной $f = \frac{c}{4l_{1e}}$.

С другой стороны, уменьшение площади поперечного сечения «губного» отрезка приводит, при прочих равных условиях, к приближению частот резонанса к частотам бесконечного импеданса той части тракта, которая расположена позади губ. Это означает понижение резонансных частот, лежащих ниже частоты четвертьволнового резонанса «губного» отрезка, и некоторое возрастание тех, которые расположены выше этой частоты, определяемой равенством (4.2).

Подобные явления можно рассматривать как акустический коррелят фонетического понятия «огубления», которое имеет в виду вытягивание губ, или уменьшение площади прохода между ними, или то и другое вместе. Количественной мерой огубления можно, очевидно, считать отношение эффективной длины l_{1e} к площади поперечного сечения A_1 «губного» отрезка системы¹⁾. В частотном диапазоне в пределах первых четырех формант результат огубления, независимо от того, как оно осуществляется, один и тот же: понижение всех резонансных частот. При этом те из формант, частоты которых зависят главным образом от полости, расположенной непосредственно позади губ, изменяются значительно больше, чем те, которые отвечают стоячим волнам в фарингальной полости.

Присоединение дополнительного отрезка к концу системы, расположенному в глубине фарингальной полости, приводит к понижению всех тех формант, частоты которых ниже критической частоты $c/4l_g$, где l_g — длина дополнительного отрезка. При этом резонансы, связанные с фарингальной полостью, изменяются сильнее, чем другие. Физиологически подобное изменение системы осуществляется опусканием всей гортани в целом.

¹⁾ Стивенс и Хауз [194, 195] используют обратное отношение A/l .

Для определения частот первых четырех резонансов системы, состоящей из двух отрезков труб, если заданы их длины и площади поперечного сечения, можно использовать номограмму, приведенную на

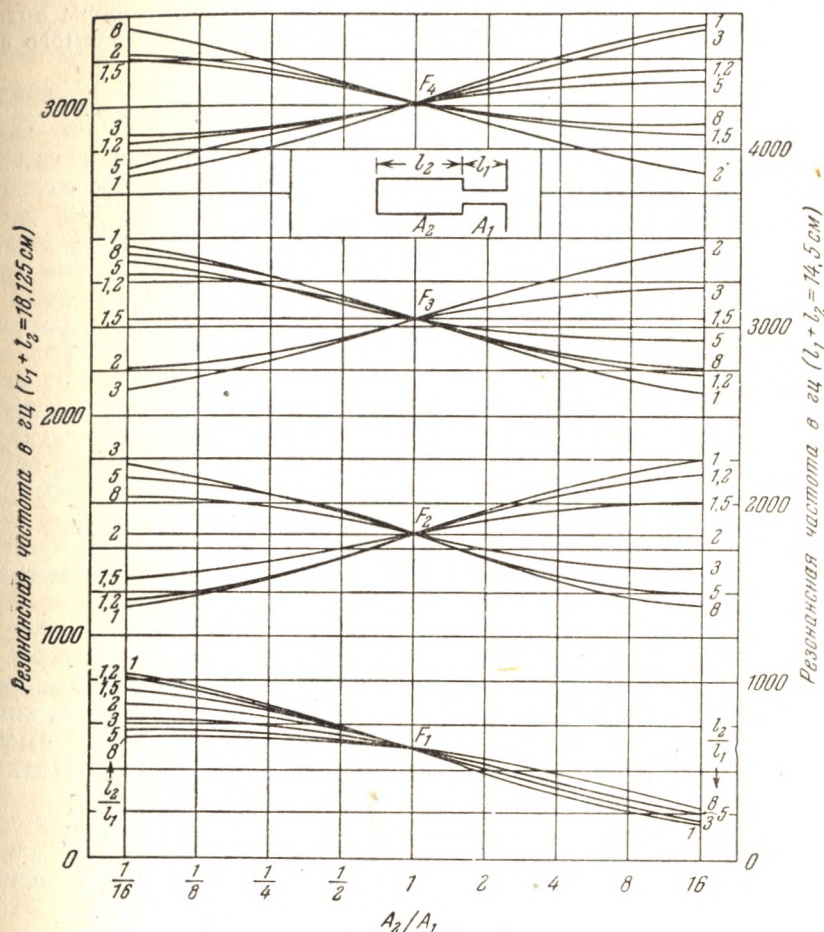


Рис. 17. Номограмма для расчета резонатора, состоящего из двух отрезков труб. Резонансные частоты могут быть определены по заданному отношению площади поперечного сечения заднего отрезка A_2 к площади поперечного сечения переднего A_1 и по отношению их длин l_2/l_1 . Для отсчета ординат даны две шкалы, соответствующие двум значениям общей длины — 14,5 см и 18,125 см. Для других значений этой длины резонансные частоты могут быть определены по правилу обратной пропорциональности. Влияние реактивной составляющей импеданса излучения не учитывается; его можно приближенно учесть, считая поправку $0,8 \sqrt{A_1/\pi}$ включенной в длину l_1 ; здесь A_1 — площадь выходного отверстия переднего отрезка.

рис. 17. Длина l_1 при этом должна включать как внутреннюю, так и внешнюю концевые поправки для переднего отрезка.

Номограмма охватывает диапазон изменений отношения l_2/l_1 от 1 до 8 и отношения A_2/A_1 от 1/16 до 16. Из условий резонанса

$$\frac{A_2}{A_1} \operatorname{tg} \frac{\omega l_1}{c} \operatorname{tg} \frac{\omega l_2}{c} = 1 \quad (4.3)$$

можно видеть, что l_1 и l_2 можно поменять местами; таким образом, данное отношение l_1/l_2 дает точно такие же резонансные частоты, как и

равное ему обратное отношение l_2/l_1 . Поэтому, если $\frac{l_2}{l_1} < 1$, можно использовать номограмму, взяв величину, обратную этому отношению. Это обстоятельство показывает возможность компенсаторных форм артикуляции. Для идеализированной системы резонаторов описываемого вида

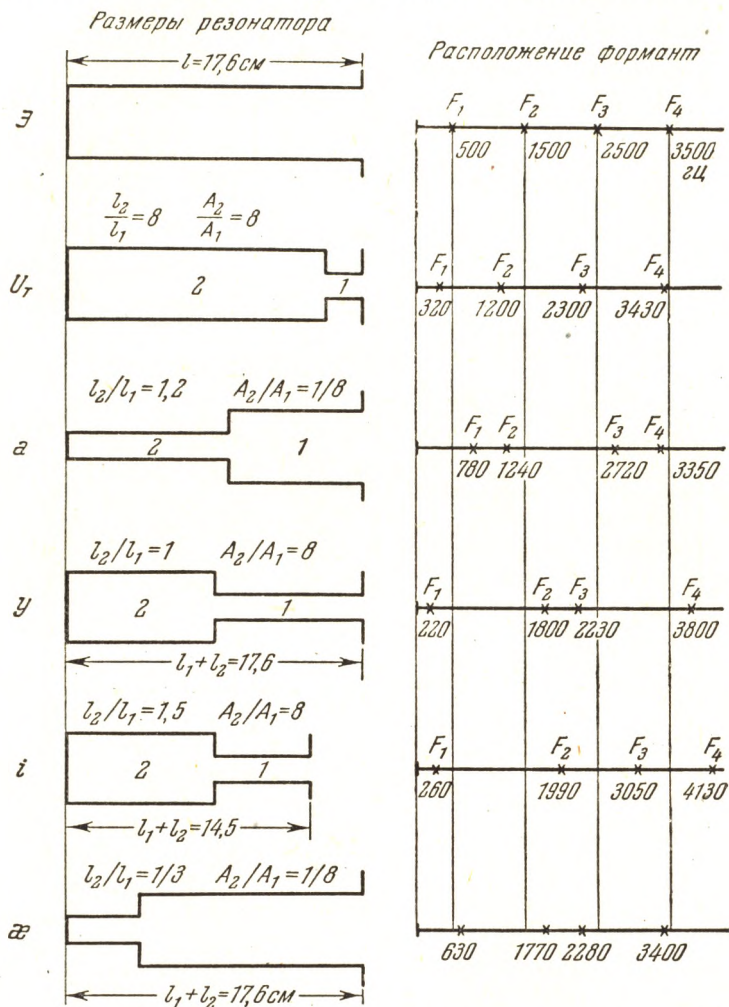


Рис. 18. Резонаторы из двух отрезков труб, для которых расположение формант подобно расположению их для некоторых гласных. Резонатор в виде одиночного отрезка трубы, открытого с одного конца, с частотами формант, являющимися нечетными кратными 500 гц, представляет собой эталон «нейтрального» звука. Синтетический гласный с таким расположением формант обозначен символом [э]. Все остальные резонаторы и соответствующие им расположения частот формант, кроме второго сверху, являются хорошим артикуляторным и акустическим приближением для гласных, указанных на рисунке.

компенсация может быть полной, если не учитывать влияния импеданса излучения. Таким образом, сужение переднего отрезка системы дает тот же эффект, что и расширение заднего, обладающего той же длиной. С другой стороны, влияние расширения переднего отрезка точно такое

же, как влияние сужения заднего той же длины, при условии, конечно, что отношение площадей остается неизменным.

Пользуясь номограммой, легко видеть, что уменьшение отношения A_2/A_1 , т. е. отношения площади поперечного сечения переднего и заднего отрезков, всегда сопровождается повышением частоты первой форманты. Если передний и задний отрезки не различаются существенно по длине, а именно, если $2 > \frac{l_2}{l_1} > \frac{1}{2}$, изменение A_2/A_1 в сторону уменьшения сопровождается понижением частоты второго резонанса. Нечто подобное имеет место в ряду гласных [i], [e], [ε], [æ], [a], где увеличение компактности сопровождается увеличением низкочастотности, как показано в § 2 гл. 3. Отношение F_2/F_1 возрастает с увеличением A_2/A_1 быстрее всего в том случае, когда $l_1 = l_2$. Если при этом $A_2 > A_1$, то имеет место наибольшее сближение F_2 и F_3 .

Следует заметить, что традиционное фонетическое представление об открытости или закрытости звука исходит из положения точки наивысшего подъема языка в ротовой полости. Гласный, подобный [a], считается обычно передним открытым. Однако действительно существенной особенностью артикуляции [a] является относительно суженный фаринкс, как будет показано в § 3 гл. 7, имеющем более фонетический уклон.

На рис. 18 показано несколько систем, составленных из двух отрезков труб, и приведены характерные для этих систем частотные положения резонансов. Фонетические символы, указанные на рисунке, проставлены исходя из сходства в расположении резонансов этих систем с расположением формант соответственных звуков; можно, однако, заметить и известное сходство в конфигурации этих систем со структурой полосей речевого аппарата, соответствующей тому или иному звуку.

Модель из двух отрезков труб, состоящая из узкой задней трубки и широкого переднего отрезка, представительна для звука [a] в такой же или даже в большей степени, чем сдвоенный резонатор Гельмгольца, который был рассмотрен в предыдущем разделе. Следует отметить резкую противоположность в конфигурациях системы для [a], с одной стороны, и для [y] или [i], с другой; следствием этого различия является и противоположность в расположении резонансных частот или, иначе говоря, в характере F-картины. Уменьшение длины переднего отрезка в конфигурации, соответствующей звуку [y], обеспечивает сдвиг вверх F_2 и F_3 , необходимый для получения [i]. Увеличение площади A_1 переднего отрезка в системе, соответствующей [i], привело бы к повышению F_1 , т. е. к приближению звука к [e]. В модели для [æ] может показаться странным то, что суженным является задний отрезок. Тем не менее, как показывают рентгенограммы, это действительно типично для очень открытых передних гласных. Сужение прохода у передней части неба является для этого звука до известной степени иллюзией.

Следует отметить возможность «артикуляторной компенсации», т. е. возможность получения той же формантной картины при помощи трубы постоянного сечения с расширением на ее переднем конце.

Может встречаться и комбинация этих двух способов артикуляции.

§ 2. Рупоры как резонаторы и связующие элементы системы

Артикуляторные условия, приводящие к повышению частоты первой форманты при почти неизменном положении остальных, могут быть воспроизведены при помощи модели, в которой весь речевой тракт

представлен в виде рупора, как показано на рис. 19. Труба постоянного сечения, имеющая длину 17,6 см и принятая за модель для некоторого нейтрального гласного, дает формантную картину, в которой частоты формант равны нечетным кратным 500 гц. Рупор такой же длины, но с плавно изменяющейся площадью сечения дает почти такую же картину для формант выше первой; однако частота первой форманты заметно возрастает в том случае, если площадь поперечного сечения воз-

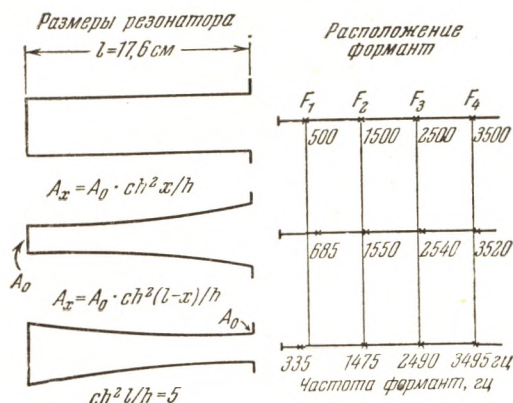


Рис. 19. Простейшие артикуляторные модели, дающие значительное изменение F_1 при незначительном изменении частот остальных формант. Приведенные данные получены расчетом, основанным на теории рупорных резонаторов. Увеличение площади поперечного сечения резонатора от горла к устью дает весьма высокую F_1 , а уменьшение этой площади — низкую F_1 . Расположение более высоких формант не отличается заметно от расположения резонансов нейтрального звука, что объясняется плавным изменением площади поперечного сечения резонатора, исключающим возможность отражения.

растет с расстоянием, и уменьшается, если она убывает. Данные, указанные на рис. 19, были получены расчетным путем в предположении рупора катеноидального типа с использованием параметров эквивалентной схемы, приведенных в § 1 гл. 2.

Пример использования рупора в качестве элемента модели речевого тракта приведен на рис. 20, иллюстрирующем способ графического определения резонансных частот на модели для гласного [i]. Подобная модель дает возможность отобразить в единой эквивалентной схеме всю ротовую часть речевого тракта, включающую нёбное сужение, переднюю полость и плавный переход к фарингальной полости. Резонансные частоты определяются точками пересечения

двух кривых — падающей и поднимающейся, причем последняя характеризует реактивное сопротивление передней части системы, если смотреть со стороны вершины фарингальной полости. Это дает значение $F_1 = 250$ гц, $F_2 = 2000$ гц и $F_3 = 3000$ гц, что соответствует спектрографическим данным и данным, полученным для модели рис. 18.

Однако по сравнению с последней моделирование ротовой полости при помощи рупора представляется более соответствующим действительной конфигурации речевого тракта. С тем, чтобы еще более приблизиться к действительности, можно дополнить систему отрезком трубы, отображающим гортань; с таким добавлением получается модель б на рис. 20. В спектре звука [i], получающемся при помощи такой модели, вторая и четвертая форманты при сохранении той же общей длины речевого тракта имеют более высокие частоты. Дальнейшее обсуждение влияния гортани и связанных с нею полостей дано в гл. 6.

Обратное по сравнению с [a] соотношение передней и задней частей системы для [i] очень просто учитывается в модели речевого тракта, использующей рупор. Достаточно в модели рис. 20 поменять местами глосовую щель и ротовое отверстие так, чтобы $x = 0$ соответствовало закрытому, а $x = 16,5$ см открытому концу, чтобы получить хорошую модель артикуляции звука [a], более естественную по конфигурации, чем система из двух отрезков труб на рис. 18. Для большинства звуков

речи такие участки речевого тракта, как образуемый языком суженный проход и прилежащие к нему плавные переходы к передней и задней полости, могут быть с очень хорошим приближением отображены при помощи катеноидального рупора. Подобное отображение составляет основной элемент одной из трехпараметрических моделей речевого тракта, описываемых в § 3 гл. 4.

Учитывая изложенное в § 1 гл. 1, § 1 гл. 2, § 1 гл. 3 и в Приложении II, следует отметить, что привлечение теории рупоров к рассмотрению речеобразования не может дать большей информации о спектрах звука, чем та, которая содержится в «функции системы» речевого тракта при задании всех ее полюсов и нулей. Конфигурация речевого тракта

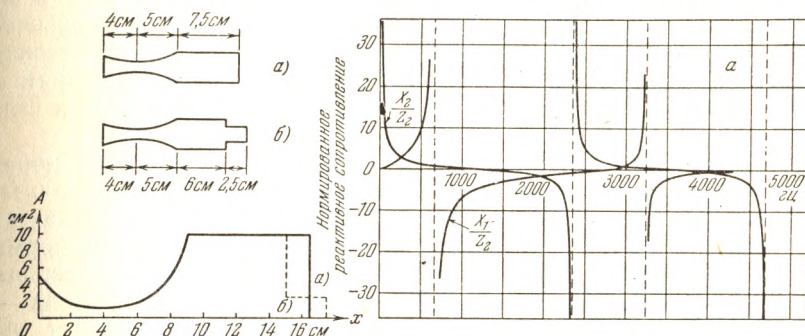


Рис. 20. Простые модели резонаторов для гласного [i], использующие катеноидальный рупор для отображения передней, средней и задней частей ротовой полости и отрезок цилиндрической трубы для отображения фарингальной части речевого тракта. В варианте (б) добавлена гортанная трубка. Метод расчета формантных частот для более простой модели (а) показан на графике реактивных сопротивлений, где X_1 — реактивность ротовой полости, на которую мы смотрим со стороны фарингальной, в сечении, отделяющем обе эти полости одну от другой; X_2 — реактивность фарингальной полости относительно того же сечения. Форманты определяются частотами, для которых $X_1 + X_2 = 0$.

сама по себе мало что говорит, например, о мощности, отдаваемой речевым трактом; вообще говоря, проще связывать мощность с частотой первой форманты, а не с формой полостей речевого аппарата. Так, широкий проход между губами еще не является показателем высокого уровня звука: положение языка, оказывая определенное влияние на частоту первой форманты, может существенно повлиять и на отдачу мощности.

Как было показано в предшествующих разделах, излучение звука не возрастает при увеличении излучающей поверхности, если частоты формант сохраняются неизменными. Поэтому невозможно увеличить акустическую мощность путем изменения одного только артикуляторного фактора (фильтровой функции), не изменяя при этом фонетического качества гласного, определяющегося частотами формант.

Функция площади рупора, входящего в состав резонатора, может быть с любой степенью точности аппроксимирована ступенчатой кривой; поэтому весь рупорный резонатор может быть представлен в виде ряда коротких трубчатых отрезков. Акустический эффект изменений системы таких простых резонаторов, заменяющих рупор, показан на рис. 21. Все резонаторы имеют длину 4 см, что по порядку величины соответствует длине передней полости палатализованных и ретрофлексных щелевых. Расчеты выполнялись на электрическом аналоге речевого тракта LEA (см. гл. 6). На входе аналога был включен источник с большим внутренним активным сопротивлением, дающий

напряжение с равномерным спектром. Поэтому до частоты, равной примерно половине частоты первой форманты, огибающая спектра на выходе имеет подъем 6 дБ на октаву в силу прямой пропорциональности излучения частоте ω .

Первым из примеров, представленных на рис. 21, является модель резонатора в виде одного отрезка трубы с основным резонансом при частоте $\frac{c}{4l_e} = 1750$ и с более высокими резонансами при частотах, представляющих собой примерно нечетные кратные 1750, т. е. 5250 и 8750 гц. Шкала частот для спектра этого резонатора, как и для всех других, полученных при помощи LEA, — логарифмическая.

По сравнению с этим «нейтральным» эталоном модель 2 и еще в большей мере модель 6 отличаются типичным для рупора повышением частоты первой форманты, большим демпфированием всех резонансов, т. е. увеличением ширины полосы формант, а также повышением уровня спектра в интервалах между формантами, что приводит к сглаживанию формантных пиков.

При увеличении площади устья рупора с увеличением кривизны, образующей уровни формантных пиков, расположенных значительно выше критической частоты рупора, зависят от двух действующих в противоположных направлениях факторов: один из них — трансформаторное действие рупора, увеличивающее объемную скорость на выходе пропорционально квадратному корню из площади устья; второй — увеличение затухания, обусловленного излучением, что снижает уровни пиков пропорционально площади устья и квадрату частоты этих пиков. Суммарный эффект увеличения площади устья, если принять сопротивление излучения за основной диссипативный элемент и считать его малым и постоянным по величине, сводится к тому, что уровни пиков уменьшаются, а уровни впадин между ними увеличиваются пропорционально квадратному корню из площади устья.

Как можно видеть из приведенных на рисунке примеров, при сходящейся форме рупора острота формантных пиков возрастает. Такой же результат дает добавление к трубе или расширяющемуся рупору отрезка трубы, соответствующего огублению; это можно видеть на спектрах моделей 4, 7 и 8, на которых видно и одновременное понижение частоты первой форманты.

Форма модели 5 сходна с конфигурацией передней полости речевого аппарата при произнесении согласного [к]. Здесь, наряду с основной формантой, имеющей частоту 1500 гц, появляется еще одна форманта с частотой 6500 гц и группа формант, расположенных около частоты 10 000 гц. При этом, правда, надо учитывать, что уровни высокочастотных формант, как и для всех моделей рис. 21, преувеличены по сравнению с таковыми для действительной речи в силу специфических характеристик источника, выбранного для проведения расчета на аналоге (см. гл. 10).

§ 3. Приближенное представление речевого тракта трехпараметрической моделью

Резонатор, состоящий из двух отрезков труб, непригоден для отображения конфигурации речевого тракта, если образуемое языком сужение явно выражено. В этом случае модель должна состоять по меньшей мере из трех или даже четырех элементов (если учитывать еще и влияние огубления). Независимо от формы, цилиндрической или

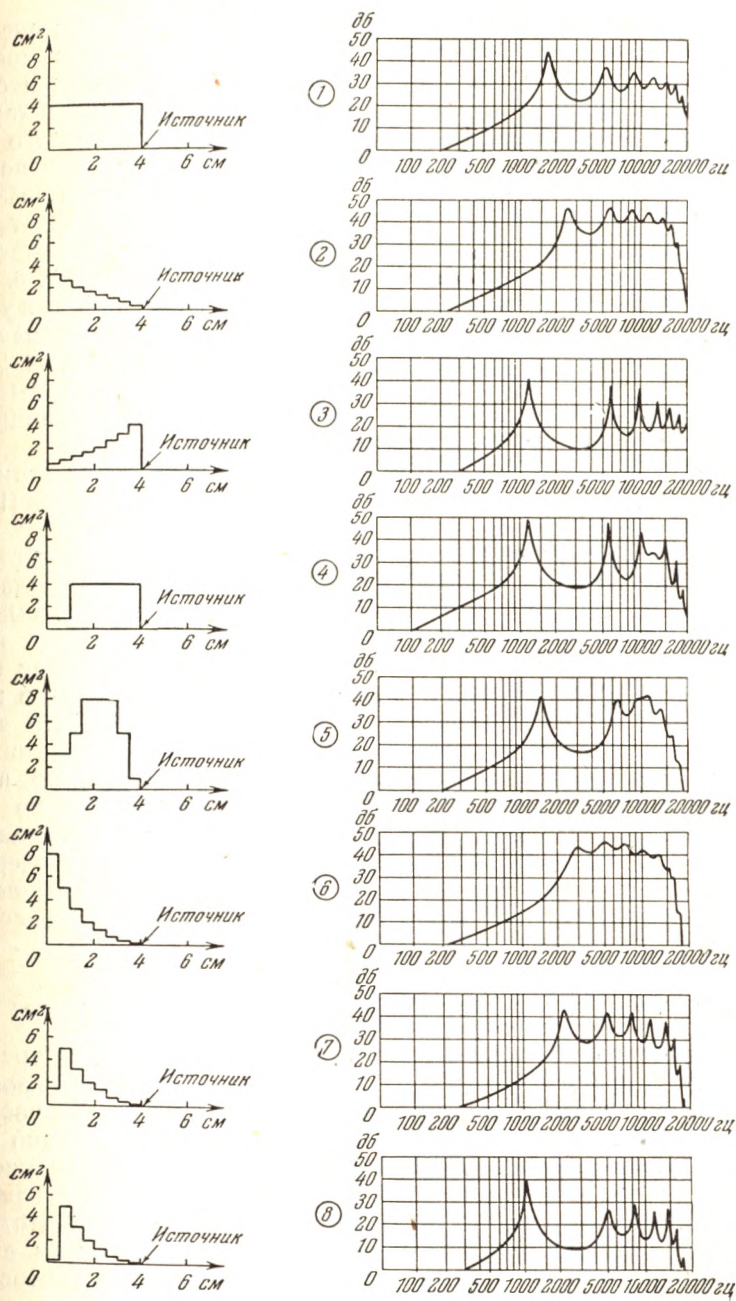


Рис. 21. Частотные характеристики коэффициента передачи некоторых простых резонаторов, имеющих длину 4 см каждый. Коэффициент передачи определяется отношением давления в излучаемой волне к объемной скорости источника с равномерным спектром и бесконечно большим внутренним сопротивлением. Характеристики рассчитаны при помощи ЛЕА. Начало абсцисс в диаграммах, представляющих функцию площади, соответствует излучающей поверхности в отверстии губ.

рупорообразной, отдельных участков системы, можно свести число артикуляторных параметров к трем. В фонетическом плане они могут быть соотнесены с местом артикуляции, степенью открытости и степенью огубления. При этом принимается, что место артикуляции характеризуется эффективным центром основного сужения, образуемого языком, т. е. местом максимального сужения в системе полостей речевого аппарата, образующимся приближением языка к твердому или мягкому нёбу или к стенке фаринкса. Степень открытости определяется минимальной площадью поперечного сечения прохода, образуемого сужением. Степень огубления характеризуется отношением l/A , т. е. отношением длины губного прохода к его площади. В случае неогубленных звуков, которые образуются с не очень большим удалением друг от друга верхних и нижних резцов, при определении отношения l/A необходимо учитывать и проход между зубами. В качестве параметра, характеризующего огубление, можно использовать вместо отношения l/A обратный ему индекс проводимости A/l .

Модель речевого тракта может быть полностью задана этими тремя переменными¹⁾. Такой подход и использовали Стивенс и Хауз [194, 195], выполнившие обширные исследования с помощью модели, у которой функция площади поперечного сечения в месте сужения, образуемого языком, и в прилежащих к нему частях передней и задней полостей характеризовалась параболическим законом. Длина всей системы и наибольшая площадь поперечного сечения передней и задней полостей были приняты постоянными. Аналогичная модель, основанная на представлении функции площади гиперболой, была исследована в процессе настоящей работы. Более простая модель, в которой место язычного сужения воспроизводилось при помощи цилиндрического отрезка, также была исследована довольно подробно, причем оказалось, что, несмотря на менее естественную форму, такая модель ведет себя, по крайней мере для первых трех резонансов, подобно модели с рупорообразной формой полости. В то же время, благодаря более простым очертаниям, на упрощенной модели легче проанализировать зависимость резонансов от размеров полостей; поэтому в первую очередь и будет рассмотрена именно эта модель.

А. Модели, состоящие только из цилиндрических отрезков труб

На рис. 22 показано несколько моделей, состоящих из трех отрезков, и соответственные F -картины, т. е. совокупности их резонансных частот, или, на техническом языке, спектры их собственных частот. Эти модели приводятся здесь с целью показать влияние изменений положения места сужения, образуемого языком, и общей длины системы и в особенности продемонстрировать связь частоты основного резонанса полости, ближайшей к ротовому отверстию, с любой из частот первых четырех формант. Первая сверху модель является простым резонатором, состоящим из одного отрезка трубы; частоты резонансов в этом

¹⁾ Такое представление подобно классическому артикуляторному описанию звуков по положению наивысшей точки языка, однако не идентично ему. Так, велярное или задненёбное положение этой точки может сочетаться с местом максимального сужения в полости фаринкса для гласных [а] и [ɑ]. Согласно классической терминологии [а] является открытым по отношению к конфигурации полости рта; однако в плане понятия о площади эффективного сужения этот звук может иметь такую же степень открытости, как и [i]; см. также § 2 гл. 7.

случае равны нечетному кратному 500 гц. Основной резонанс такой модели соответствует стоячей волне на частоте, при которой длина трубы равна $1/4$ длины волны. Четвертьволновый резонанс определяет также частоту F_2 для второй модели, частоту F_3 для третьей и частоту F_4 для четвертой.

Близость F_2 и F_3 в модели 3 обусловлена совпадением в несвязанной системе четвертьволнового резонанса передней полости на частоте $c/4l_1$ и полуволнового резонанса задней полости на частоте $c/2l_2$ при $l_2 = 2l_1$. Если место наибольшего подъема языка переместить несколько вперед или укоротить переднюю полость, как это сделано на модели 4,

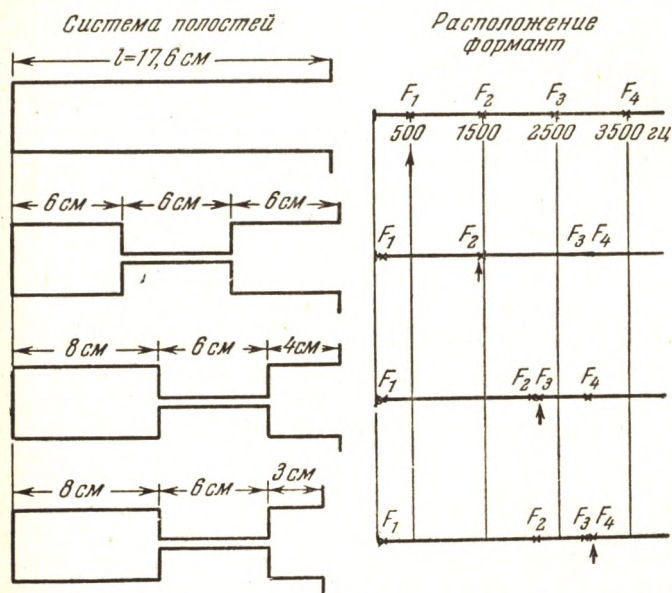
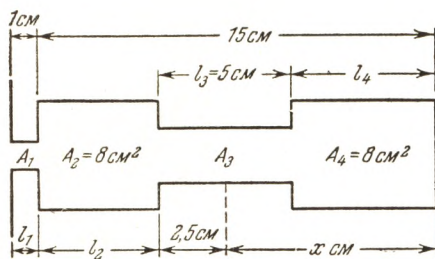


Рис. 22. Модели резонатора из трех отрезков труб и соответственное расположение формантных частот. Размеры в каждом случае выбраны так, чтобы частота основного резонанса переднего отрезка совпадала соответственно с 1-й, 2-й, 3-й и 4-й формантой. Стрелки на F-картинах указывают местоположение основного резонанса передней полости. Эти идеализированные модели отображают некоторые основные черты артикуляции велярных и палатальных согласных.

то сближаются частоты F_3 и F_4 . В этом случае F_2 почти целиком определяется полуволновым резонансом задней полости, а F_3 обусловлена полуволновым резонансом сужения, т. е. отрезка трубы, соединяющего обе полости. Модели 2, 3, 4 можно использовать при рассмотрении образования согласных [к] или [г], стоящих перед [а], [æ] и [i] соответственно. Дополнительное огубление в модели 2 вызывает понижение частоты «ротовой» форманты, как это имеет место в слове [ки].

Были проведены детальные исследования модели подобного типа, состоящей из трех и четырех отрезков труб. В отличие от моделей рис. 22, за длину переднего отрезка была принята его физическая длина, т. е. не учитывалась концевая поправка. Для случая модели из трех отрезков, соответствующей отсутствию огубления, была принята общая длина 15 см. Дополнительный отрезок, имитирующий проход между губами, всюду имел длину 1 см. Площадь поперечного сечения отрезков, расположенных по обеим сторонам отрезка, имитирующего язычное

сужение, в этой модели была равна 8 см^2 , как это видно из рис. 23. Длина отрезка, соответствующего язычному сужению, оставалась постоянной и равной 5 см во всех тех случаях, когда сохранялись «полости» по обеим сторонам его, но его место и площадь поперечного сечения изменялись. Смещения этого отрезка по направлению к переднему и заднему концам модели доводились до получения конфигураций, при которых длина l_2 или l_4 становилась равной нулю; в этих случаях



	Кривая	A_1	l_1	Общая длина
$A_3 = 0,65$	1	8	0	15
	2	4	1	16
	3	2	1	16
	4	0,65	1	16
	5	0,16	1	16
$A_3 = 2,6$	1	8	0	15
	2	4	1	16
	3	2	1	16
	4	0,65	1	16
	5	0,16	1	16

Рис. 23. Трехпараметрическая модель речевого тракта, состоящая из четырех отрезков труб, которые соответствуют губному проходу, передней полости, язычному проходу и задней полости. Полная длина без губного прохода 15 см ; площадь поперечного сечения двух основных полостей сохраняется постоянной и равной 8 см^2 . Длина язычного прохода 5 см , за исключением тех случаев, когда он занимает крайнее переднее или заднее положение, т. е. когда исчезает либо передняя, либо задняя полость. Тремя параметрами являются: 1) местоположение язычного прохода, 2) площадь его поперечного сечения, 3) отношение длины губного прохода к его площади. Таблица значений этих параметров относится к номограммам рис. 24.

($2,5 < x < 12,5$) длина l_3 прохода уменьшалась настолько, чтобы обеспечить постоянство общей длины системы. При крайнем переднем или крайнем заднем положении сужения резонатор обращается в простую трубу длиной 15 см , к которой присоединяется отрезок с изменявшейся площадью поперечного сечения, характеризующий губной проход.

Результаты расчетов, выполненных на электрическом аналоге LEA, показаны на рис. 24 и 25. Для кривых 2, 3, 4 и 6 площади поперечного сечения отрезка, характеризующего влияние губ, равны соответственно 4, 2, 0,65 и $0,16 \text{ см}^2$. Кривая 1 относится к случаю, когда отрезок, имитирующий губы, исключен. Рис. 24 и 25 относятся соответственно к случаям площади поперечного сечения язычного сужения в $0,65$ и $2,6 \text{ см}^2$.

Как видно из кривых, влияние изменений места сужения качественно одно и то же при различных площадях поперечного сечения этого сужения. Различие заключается в большем отклонении от F -картины для нейтрального случая простой трубы при уменьшении площади поперечного сечения язычного сужения.

Важнейшие видимые изменения, вызываемые перемещениями язычного сужения от наиболее заднего положения к переднему, выражаются в переходе от высокой F_1 и низкой F_2 , как в задних гласных, к F_2 , близкой к F_1 , как в передних гласных высокого подъема. Место артикуляции, обуславливающее минимальную F_2 , находится в $1-2 \text{ см}$ впереди от координаты, при которой максимально значение F_1 . Подобно этому максимум F_2 имеет место при координате приблизительно на $0,5 \text{ см}$ впереди от координаты наименьшей F_3 . Связанная с этим близость F_2 и F_3 встречается при средненёжном положении языка.

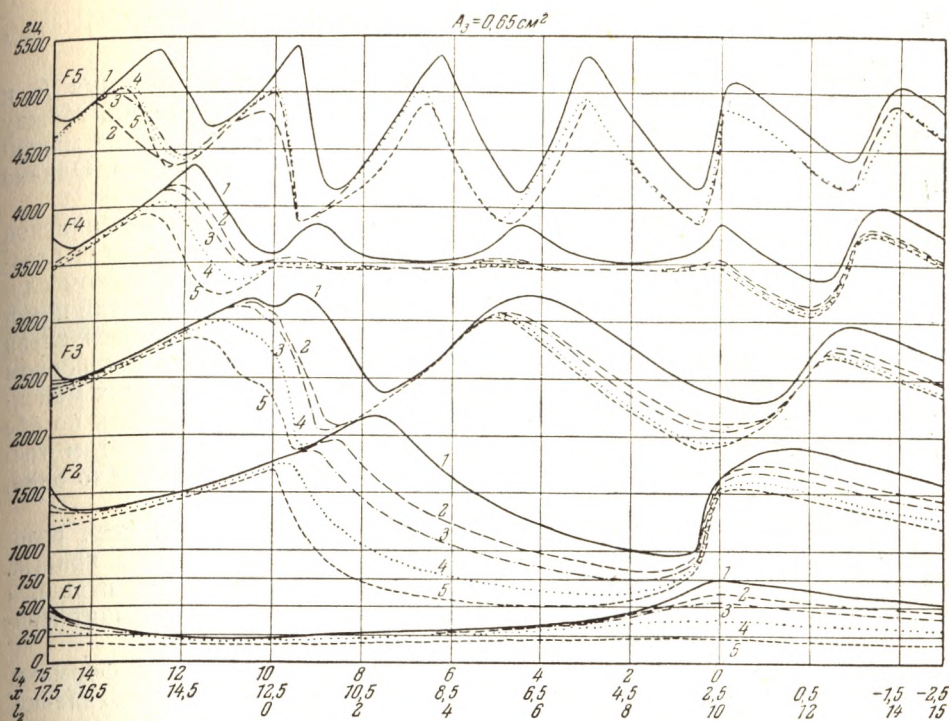


Рис. 24. Номограммы, дающие связь между частотами F_1, F_2, F_3, F_4 и F_5 и положением язычного прохода в трехпараметрической модели. Площадь поперечного сечения язычного прохода $A_3 = 0,65 \text{ см}^2$; l_1 — длина заднего отрезка, x — координата сужения, l_2 — длина переднего отрезка.

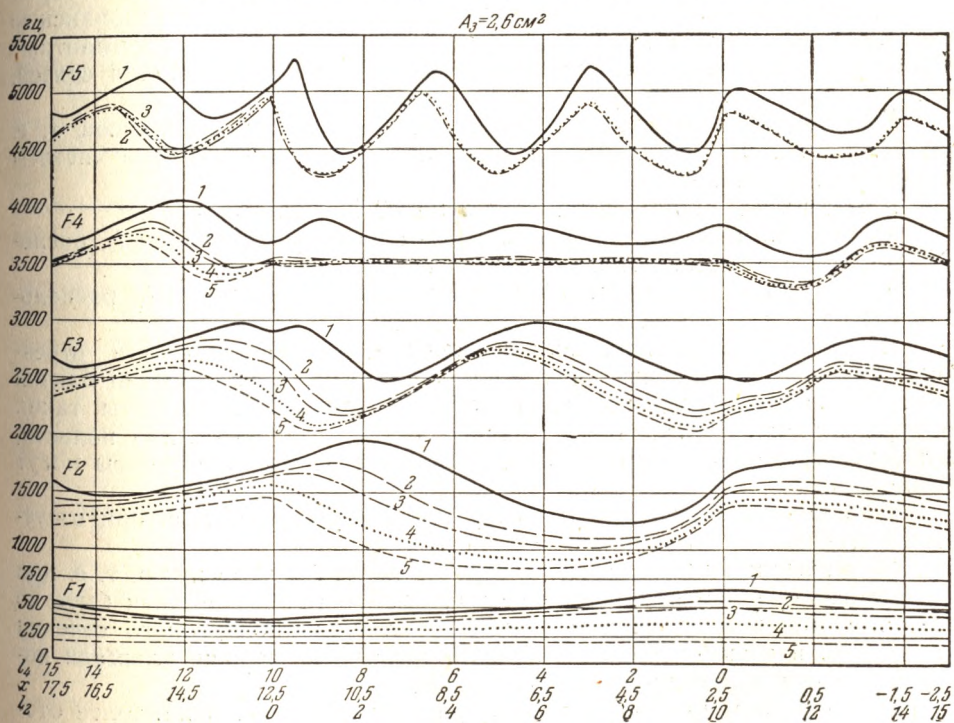


Рис. 25. То же, что на рис. 24, для $A_3 = 2,6 \text{ см}^2$.

Далее, после минимума F_2 наблюдается еще один максимум F_2 и минимум F_3 , что соответствует гортанному месту артикуляции. Между этими двумя областями сближения F_2 и F_3 расположен один из максимумов F_3 , и еще один находится перед средненёбной областью. Однако пределы изменения F_3 при этом не очень велики.

Если продвижение языка вперед ведет к повышению резонансной частоты, то можно сделать вывод, что резонанс этот определяется главным образом той полостью, длина которой уменьшается, в данном случае передней полостью. Аналогично можно считать, что те участки нотограммы, где резонансные частоты поднимаются, если двигаться слева направо, т. е. если язык оттягивается назад, определяются в основном задней полостью. Отсюда следует, что частоты формант в точках максимумов и минимумов в равной мере зависят от обеих полостей, а по обе стороны от положения язычного прохода, соответствующего этим точкам, на частоту формант оказывает преимущественное влияние то одна, то другая из полостей.

Увеличение степени огубления приводит к понижению всех резонансных частот. Это понижение наибольшее для тех участков кривых, на которых частота резонанса определяется по преимуществу передней полостью. Огубление также сдвигает области сближения формант к более переднему положению язычного прохода.

Если язык передвигается назад от области сближения формант, частота более высокой форманты изменяется в том же направлении, в котором изменялась частота более низкой форманты до точки ее максимума. Общей причиной этой кажущейся непрерывности является изменение несвязанной резонансной частоты той полости, длина которой уменьшается при перемещении языка. Две несвязанные резонансные частоты совпадают в центре области сближения, но две формантные частоты, т. е. резонансные частоты всей системы при наличии связей, могут совпадать только при полной смычке.

Для суждения о происхождении первых четырех резонансов, т. е. F_1 , F_2 , F_3 , F_4 , представляют интерес следующие резонансные частоты несвязанной системы:

I. Резонансы Гельмгольца (формула (3.4)), определяемые:

а) передним резонатором, состоящим из полости длиной l_2 , расположенной перед язычным проходом, и губного отверстия;

б) задним резонатором, образуемым полостью длиной l_4 , расположенной позади от язычного прохода, и язычным сужением.

II. Резонансы стоячих волн в каждой из полостей или отрезков, концы которых нагружены по-разному, а именно: один близок к условиям открытого конца, а другой — закрытого. Частоты таких четвертьволновых резонансов равны $nc/4l$, где l — длина полости, а n соответственно равно 1, 3, 5 и т. д. Подобные резонансы могут возникать:

а) в передней полости при отсутствии огубления, т. е. при отсутствии отрезка, имитирующего губной проход;

б) в язычном проходе в том случае, когда l_2 равно нулю, т. е. при предельно переднем положении языка. Губной проход должен быть по площади много меньше, чем язычный; площадь поперечного сечения последнего в свою очередь должна быть заметно меньше площади поперечного сечения задней полости;

в) в язычном проходе, когда l_4 равно нулю, т. е. в наиболее отодвинутом назад положении языка.

III. Резонансы стоячих волн в той полости или отрезке трубы, оба конца которого нагружены одинаково, т. е. оба или почти открыты, или почти закрыты. Частоты этих полуволновых резонансов равны $nc/2l$, где $n = 1, 2, 3$ и т. д. Подобные резонансы могут возникать:

- а) в передней полости при значительном огублении,
- б) в задней полости,
- в) в язычном проходе, соединяющем две большие полости.

IV. Резонансы промежуточного между этими тремя главными типами характера; они встречаются, когда один из концов резонирующей системы полуоткрыт;

а) основной резонанс передней или задней полости в случае, промежуточном между Ia и IIa;

б) следующий за ним более высокий резонанс, т. е. промежуточный между вторым резонансом типа II с частотой, равной $3c/4l$, и первым резонансом типа III с частотой, равной $c/2l$. Аналогично, следующий резонанс будет на частоте, лежащей между частотами $5c/4l$ и $2c/2l$, но относительно ближе к последней. Положение дальнейших резонансов ясно из графического построения, приведенного на рис. 16.

Детальное представление о связи резонансов с той или иной полостью может быть получено, если проследить за ходом кривых на рис. 24 справа налево. Данные на рис. 25 менее показательны, так как благодаря широкому язычному проходу в этом случае частоты формант в меньшей мере зависят только от одного какого-либо отрезка системы. При координатах сужения позади от координаты максимума F_1 характер кривой 1 из группы кривых, относящихся к F_1 на рис. 24, определяется четвертьволновым резонансом передней полости. По мере увеличения огубления, чему соответствуют кривые 1, 2, 3, 4, 5, резонанс передней полости изменяется по характеру от типа IIa к типу IVa и, наконец, к Ia,

Если язычное сужение передвигается вперед после прохождения максимальной F_1 , то F_1 все в большей мере начинает определяться резонансом задней полости. Область зависимости от одной из полостей для этой форманты больше, чем для любых других формант, и охватывает большие пределы изменений координаты сужения, по крайней мере при наличии огубления.

Частота F_2 впереди от координаты минимума F_2 и до координаты максимума F_2 определяется по преимуществу основным резонансом передней полости. При дальнейшем продвижении языка вперед этот резонанс определяет F_3 и даже F_4 , как уже показано на рис. 22.

Интересно проследить за полуволновым резонансом задней полости, т. е. резонансом типа IIIб. При смещении язычного сужения назад этот резонанс после прохождения максимума F_2 все в меньшей степени определяет F_2 и все в большей мере соответствует F_3 . После прохождения язычным сужением положения, соответствующего минимуму F_2 , основной резонанс задней полости типа Ib или IVa начинает определять не F_1 , а F_2 .

Участок кривой F_3 между средним ее максимумом и последующей областью сближения F_2 и F_3 связан с $3/4$ -волновым резонансом передней полости при отсутствии огубления и приближается к полуволновому по мере увеличения огубления, как показано выше в пункте IVб. При координате сужения в пределах 0—1,5 см резонансная частота F_3 соответствует четвертьволновому резонансу сужения, как об этом сказано в пункте IIв.

При крайних заднем или переднем положениях сужения вся модель представляет простую трубу с добавлением или без добавления отрезка, имитирующего огубление. В последнем случае резонансы имеют место при нечетных кратных 540 гц, т. е. частоты, при которой длина в 15 см с добавлением концевой поправки $0,8\left(\frac{8}{\pi}\right)^{1/2}$ см составляет четверть длины волны. Эффект максимального огубления заключается в понижении частоты одного из этих резонансов на величину, соответствующую четвертьволновому резонансу при физической длине 15 см, т. е. 590 гц.

Частота F_4 в основном определяется полуволновым резонансом язычного сужения и, таким образом, в значительной мере независима от артикуляторных переменных, за исключением ограниченной области переднего положения сужения, где F_4 связана с передней полостью. F_5 связана с различными резонансами стоячих волн более высокого порядка. Данные, относящиеся к F_5 , не имеют существенного фонетического значения и являются в большей степени свойством данной модели, а не действительной речи. Это же частично относится и к F_4 . В естественной речи на F_4 может в большей или меньшей степени оказывать влияние четвертьволновый резонанс гортанной трубки; последняя, однако, не необходима для возникновения четвертой форманты (см. § 2 гл. 7).

Если бы модель речевого тракта вела себя полностью как двойной резонатор Гельмгольца, то наблюдались бы только два резонанса и только одна область сближения резонансных частот, в пределах которой расположен максимум более низкочастотного и минимум более высокочастотного резонанса. Местоположение сужения, при котором становятся равными резонансные частоты несвязанных передней и задней полостей, близко к центру области наибольшего сближения формант.

Б. Модели, в которых язычная область речевого тракта представлена рупором

Представленная на рис. 26 трехпараметрическая модель речевого тракта является более подходящей основой для определения частот формант по артикуляторным данным, чем модель рис. 23, так как конфигурация ее ближе к конфигурации естественного речевого тракта. В такой модели области сужения представлены катеноидальным рупором. Для того чтобы проводить расчеты этой модели с помощью электрического аналога LEA, необходимо аппроксимировать функцию площади рупора ступенчатой кривой, как это показано на рисунке. Удобной основой для численных расчетов является полная эквивалентная схема, описанная в § 1 гл. 2.

Не считая огубления, все размеры модели однозначно определяются координатой центра рупорообразной части $x_{\min}$ и площадью поперечного сечения $A_{\min}$ в этой точке, поскольку площадь поперечного сечения рупора возрастает симметрично по обоим сторонам от координаты центра по закону:

$$A_x = A_{\min} \operatorname{ch}^2(x - x_{\min})/h, \quad (4.4)$$

где A_x — значение функции площади для координаты x , h — постоянная, характеризующая раствор рупора и обусловленная тем, что пло-

щадь должна достигать предельной величины 8 см^2 на расстоянии $\pm 4,75 \text{ см}$ от координаты центра и далее оставаться постоянной в области передней и задней полостей:

$$\frac{h}{4,75} = \text{arch} \left(\frac{8}{A_{\min}} \right)^{1/2}. \quad (4.5)$$

Гортанная трубка и шунтирующие полости, соответствующие грушевидным пазухам (sinus piriformis), сохраняют фиксированные размеры и учитываются во всех вычислениях как добавление к задней полости. Расстояние по оси от основания гортанной полости до конца модели у губ принято равным $16,5 \text{ см}$ для случая неогубленных звуков

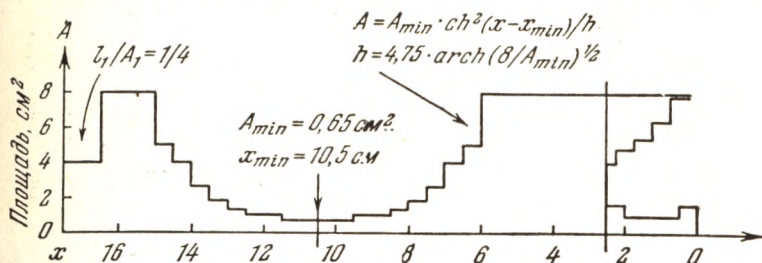


Рис. 26. Трехпараметрическая модель речевого тракта, использующая рупорообразное отображение язычного прохода. Рупор аппроксимирован ступенчатой функцией, соответствующей возможностям расчета при помощи электрического аналога LEA. Модель содержит дополнительные полости постоянных размеров, отвечающие гортанной трубке и грушевидным пазухам. Координата x — расстояние сужения от голосовых связок в сантиметрах.

и $17,5 \text{ см}$ при огублении, когда вводится дополнительный отрезок, соответствующий губному проходу длиной $l_1 = 1 \text{ см}$. Эффективную длину области язычного сужения можно определить как длину цилиндрической трубы с площадью сечения $A_{\min}$, имеющей такой же импеданс на низких частотах, как и рупор. Из формулы (4.4) следует, что эта длина l_e равна

$$l_e = A_{\min} \int_{x_{\min}-4,75}^{x_{\min}+4,75} \frac{dx}{A(x)} = 2h \text{th} \left(\frac{4,75}{h} \right). \quad (4.6)$$

В нижеследующей таблице приведены значения эффективной длины как функции минимальной площади поперечного сечения $A_{\min}$.

Влияние изменения «места артикуляции» $x_{\min}$ при фиксированной небольшой «открытости»

$A_{\min} = 0,65 \text{ см}^2$ и при пяти различных степенях огубления показано на рис. 27, который можно сравнить с рис. 24. Аналогичные графики рис. 28, построенные для случая $A_{\min} = 2,6 \text{ см}^2$, можно сравнить с рис. 25. Сравнение показывает, что топология номограмм этих двух моделей, определяющая координаты областей сближения формант, одинакова. Они также хорошо согласуются с данными, приведенными Стивенсом и Хаузом [194, 195]. Такое совпадение показывает, что зависимость формант от полостей не изменяется коренным образом, если часть речевого тракта, соответствующая язычному сужению, заменена рупором вместо цилиндрической трубы.

$A_{\min}, \text{см}^2$	0,16	0,32	0,64	2,0	4,0	8,0
$l_e, \text{см}$	3,6	4,1	4,7	6,3	7,6	9,5

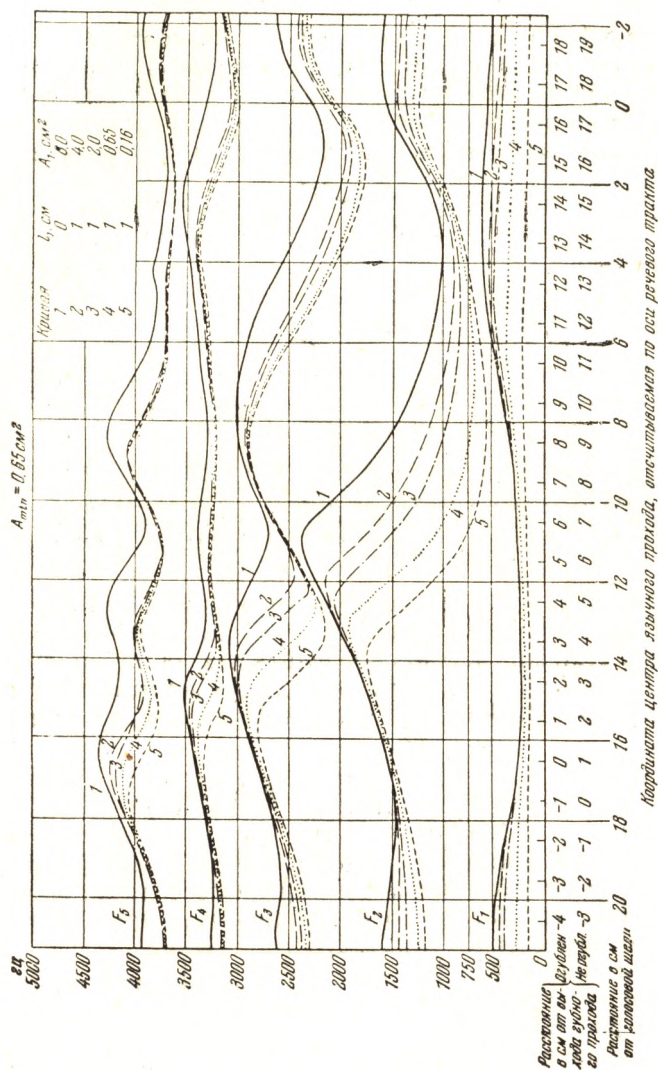


Рис. 27. Номограммы для определения F_1 , F_2 , F_3 , F_4 и F_5 для трехпараметрической модели рис. 26 в зависимости от положения язычного прохода. Минимальная площадь поперечного сечения сужения, представляющего язычный проход, $A_{\min} = 0,65 \text{ см}^2$. Пять степеней огубления: кривая 1 соответствует нулевой длине губного прохода, кривые 2—5 соответствуют данным таблицы.

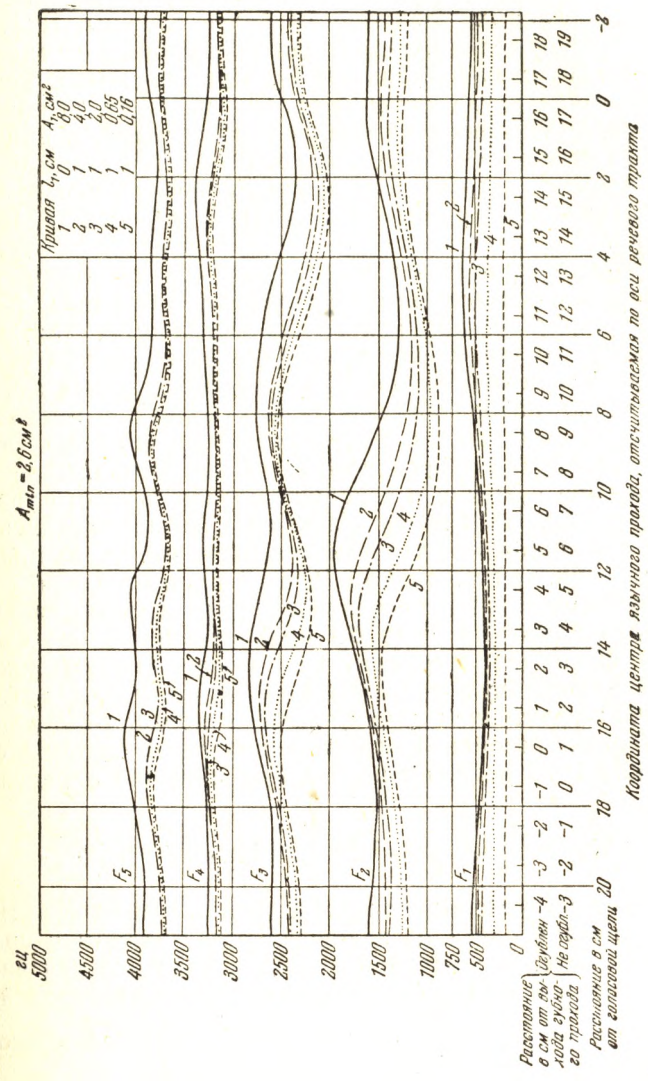


Рис. 28. То же, что на рис. 27, для $A_{\text{min}} = 2,6 \text{ см}^2$.

Эффект изменения площади поперечного сечения в области язычного сужения в зависимости от его положения показан на рис. 29, построенном для случая минимального огубления.

Увеличение площади сужения ведет к повышению F_1 , если сужение расположено в передней половине модели. При положении, характерном для задних гласных, зависимости более сложны. В этом случае имеется оптимальная величина площади поперечного сечения сужения, которой соответствует максимальная F_1 . При максимальном значении площади $A_{\min} = 8 \text{ см}^2$, когда сужение вообще исчезает, F_1 приближается к величине, соответствующей нейтральному гласному, а при очень малой площади F_1 становится равной нулю. В области номограммы, соответствующей задним гласным, изменения F_1 меньше, чем изменения F_2 . Основное влияние «централизации» задней гласной сказывается в сдвиге F_2 вверх.

Общее правило состоит в том, что уменьшение контрастов в площадях поперечного сечения речевого тракта сближает F -картину с F -картиной, характерной для нейтрального гласного. Так, можно видеть, что F_2 растет с увеличением площади сужения при условии, что центр сужения расположен ближе к голосовой щели, чем к губам. Если место артикуляции находится в передней половине модели, частота второй форманты стремится к нейтральному положению, уменьшаясь при увеличении площади поперечного сечения прохода.

При положении сужения в середине модели речевого тракта изменения $A_{\min}$ не сказываются на F_2 . На первый взгляд казалось бы более естественным предположить, что такая независимость F_2 появится при расстоянии сужения от губ, равном $1/3$ всей длины модели, где резонатор в виде трубы постоянного сечения имеет узел скорости для F_2 . Однако такое предположение было бы правильным, только если бы сужение можно было рассматривать как весьма тонкую диафрагму с отверстием. В наших двух моделях, а также в моделях Стивенса и Хауза [194, 195] эффективная длина язычного прохода была порядка $1/3$ всей длины. Условие резонанса для цилиндрической системы, состоящей из трех отрезков труб равной длины l , приближенно определяется выражением

$$\operatorname{ctg} \varphi = \frac{A_3}{A_2} \operatorname{tg} \left(\varphi + \operatorname{arctg} \frac{A_2}{A_1} \operatorname{tg} \varphi \right), \quad \varphi = \frac{\omega l}{c}, \quad (4.7)$$

которое вытекает из условия равенства нулю суммы реактивных сопротивлений справа и слева от границы между задним и средним отрезками. При частоте, при которой $\varphi = 90^\circ$, правая и левая части уравнения обращаются в нуль независимо от отношения площадей. Частота этого резонанса $F_2 = \frac{c}{4l} = \frac{3c}{4l_{\text{полн}}}$, где $l_{\text{полн}} = 3l$ является полной длиной модели, совпадает с «нейтральным» положением F_2 .

Тот же результат получается при таком положении сужения, при котором модель голосового тракта приближается к системе из двух отрезков труб, в которой длина заднего или переднего отрезка равна $1/3$ полной эффективной длины всей системы. Независимо от отношений площадей F_2 сохраняет в этом случае то же значение, что и для резонатора из одной трубы, соответствующего нейтральной гласной. Это вытекает из формулы для резонатора из двух отрезков труб:

$$\operatorname{ctg} \varphi_2 = \frac{A_2}{A_1} \operatorname{tg} \varphi_1, \quad (4.8)$$

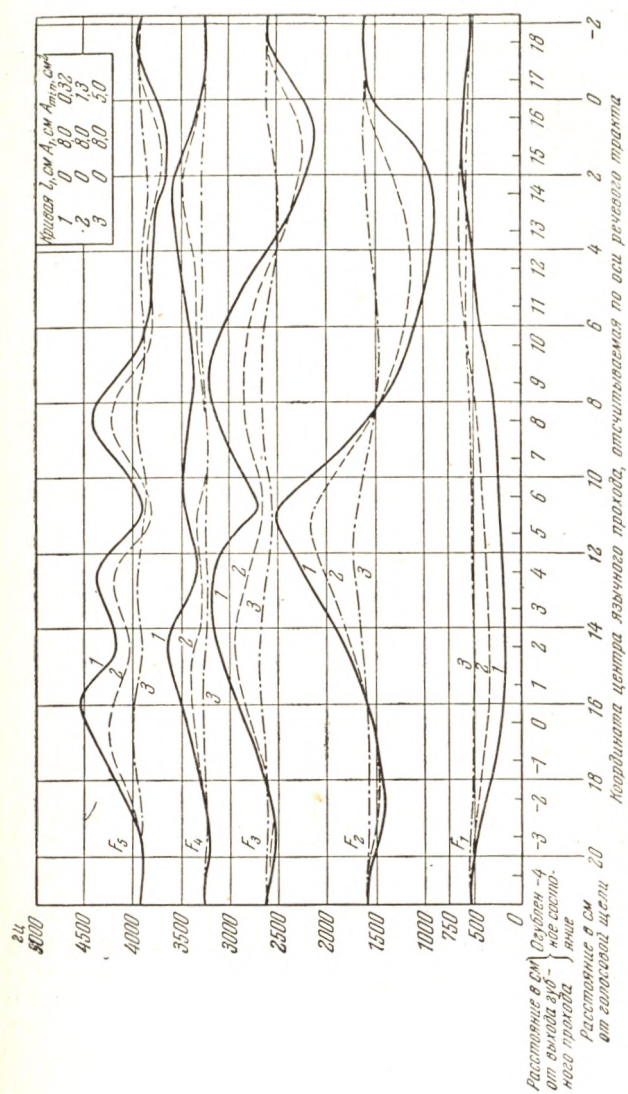


Рис. 29. То же, что на рис. 27, но губной проход не учитывается. Три значения минимальной площади язычного прохода.

где $\varphi_1 = 2\varphi_2 = \pi$ или $\varphi_2 = 2\varphi_1 = \pi$. Этим объясняется, почему кривые, характеризующие изменения F_2 при разных значениях $A_{\min}$, сходятся при положении места сужения сразу же за губами. Эффективная длина переднего отрезка включает концевую поправку и половину длины участка язычного прохода. Постоянство F_2 при близости сужения к гортанному концу системы не поддается столь простому объяснению ввиду наличия там добавочных полостей.

Некоторые существенные особенности взаимосвязи между изменениями артикуляции и изменениями частот формант легко уяснить, если обратиться к картине стоячих волн в модели речевого тракта, состоящей

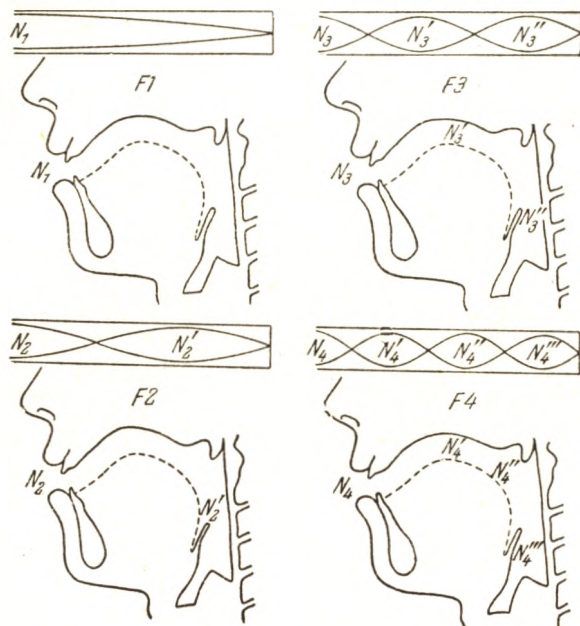


Рис. 30. Картины стоячих волн объемной скорости для каждого из первых четырех резонансов речевого тракта, представленного в виде трубы, согласно Чибя и Каджияма [24].

из одного отрезка трубы. Как указывает Чибя [24], «если в трубе образуется сужение, то частота резонанса трубы понижается или повышается в зависимости от того, находится ли это сужение вблизи от точки максимальной объемной скорости (N) или максимального избыточного давления (P)». Распределение амплитуд объемной скорости (тока) в модели нейтрального гласного показано на рис. 30 для первых четырех резонансов. Пространственное распределение давлений обратно распределению объемной скорости, так что минимуму объемной скорости соответствует максимум давления. Распределение давления в моделях для различных гласных рассмотрены в гл. 7, § 4, В.

На частоте первого резонанса максимум объемной скорости находится у губ, причем объемная скорость имеет значительную величину во всей передней части ротовой полости. Отсюда следует, что сужение прохода между губами или в пределах ротовой полости понижает частоту первой форманты. Максимум объемной скорости именно в губном проходе характерен не только для первого, но и для всех других резо-

нансов, так что при уменьшении площади губного прохода снижаются частоты всех формант.

При втором резонансе существует дополнительный минимум объемной скорости, т. е. максимум давления, на расстоянии $\frac{2}{3}$ всей длины речевого тракта, считая от голосовой щели. Этой же координате соответствует такое местоположение язычного прохода в двух рассмотренных выше трехпараметрических моделях, при котором F_2 достигает максимального значения. Третьему резонансу трубы, являющейся моделью нейтрального гласного, соответствуют по длине трубы два минимума объемной скорости — на расстоянии $\frac{4}{5}$ и $\frac{2}{5}$ от голосовой щели; в полном соответствии с этим находятся и наблюдаемые на номограммах рис. 24, 25 и 27—29 особенности. Минимум частоты второго резонанса, как это видно из номограмм, совпадает с тем, что можно ожидать на основании картины стоячих волн в трубе; совпадение это, однако, имеет место только для умеренных степеней сужения язычного прохода. При значительном уменьшении площади поперечного сечения язычного прохода координата минимума F_2 сдвигается назад, к голосовой щели. Это обусловлено комбинированным влиянием изменений объема наряду с изменениями площади сужения.

Правила, связывающие изменения формантных частот с местными сужениями или расширениями трубы, легко получить, рассматривая импедансы. Уменьшение площади поперечного сечения в месте максимума объемной скорости равносильно последовательному включению сосредоточенной индуктивности, поскольку емкостью суженного участка можно пренебречь в силу минимума давления на этом участке. Если уменьшение площади поперечного сечения трубы происходит на участке, близком к координате минимума объемной скорости, т. е. максимума давления, то, наоборот, можно не учитывать распределенную индуктивность в пределах этого участка и считаться только с его емкостью. Однако увеличение сосредоточенной индуктивности приводит к понижению частоты резонанса, а уменьшение емкости — к ее повышению. Это следует непосредственно из того общего положения теории цепей, что реактивное сопротивление (в нашем случае представляющее сумму реактивных сопротивлений перед рассматриваемым сечением и позади него) должно быть равно нулю на частоте резонанса и возрастать с возрастанием частоты.

Таким образом, влияние местного сужения или расширения в речевом тракте может быть оценено заранее, если известна картина стоячих волн в системе до того, как произошло артикуляторное изменение, вызывающее подобное сужение или расширение.

Следует, однако, отметить, во-первых, что использование одного отрезка трубы в качестве исходной системы при такой оценке допустимо только при условии отсутствия огубления и, во-вторых, что положение максимумов и минимумов до известной степени зависит от степени сужения.

Для резонансов выше третьего эта теория оправдывается в меньшей степени ввиду довольно значительных по длине размеров язычного прохода, образующего сужение.

Если в исходной нейтральной модели речевого тракта добавляется огубление, то и исходная картина стоячих волн должна быть соответственным образом изменена. Уменьшение выходного отверстия в трубе, соответствующей исходной модели, приводит к смещению минимума объемной скорости на частоте второго резонанса от координаты,

Таблица 1

Влияние исключения из системы гортанных полостей и уменьшения общей длины рупорообразной части системы (рис. 26).
 x — расстояние центра сужения от заднего конца модели (в см), d — расстояние центра от переднего конца системы (в см).

А. Численные значения соответствуют кривой I на рис. 27; отрезок, соответствующий губному проходу, отсутствует; $A_{\text{min}} = 0,65 \text{ см}^2$; полная длина модели 16,5 см; учтена гортанная трубка и грушевидные па-
 зухи

В. Конфигурация резонатора такая же, как в А, грушевидные пазухи исключены

С. То же, что и В, но без учета гортанной трубки. Полная длина 14 см
 D. То же, что и С, но общая длина модели уменьшена до 11 см без уменьшения рупорообразной части системы

d	А				В				С				D			
	x	F_1	F_2	F_3	x	F_1	F_2	F_3	x	F_1	F_2	F_3	x	F_1	F_2	F_3
—4	20,5	530	1590	2620	20,5	560	1695	2740	18	580	1740	2930	15	725	2185	3690
—3	19,5	500	1520	2565	19,5	530	1630	2685	17	540	1680	2885	14	670	2100	3670
—2	18,5	435	1465	2600	18,5	460	1570	2715	16	480	1625	2930	13	580	2085	3810
—1	17,5	340	1460	2680	17,5	385	1580	2780	15	300	1635	3060	12	480	2155	4040
0	16,5	300	1525	2795	16,5	330	1640	2900	14	340	1710	3240	11	395	2325	4280
1	15,5	260	1620	2915	15,5	285	1750	3005	13	300	1840	3440	10	350	2570	4310
2	14,5	245	1745	3015	14,5	265	1900	3080	12	280	2015	3550	9	340	2875	4140
3	13,5	250	1910	3070	13,5	260	2080	3120	11	275	2220	3449	8	345	3075	4085
4	12,5	260	2100	3005	12,5	275	2285	3080	10	285	2445	3465	7	380	3035	4260
5	11,5	280	2310	2815	11,5	300	2470	2935	9	305	2610	3200	6	435	2860	4200
6	10,5	295	2370	2705	10,5	315	2460	2855	8	330	2470	3290	5	545	2540	3830
7	9,5	320	1970	2860	9,5	350	2050	2925	7	370	2050	3285	4	750	2110	3720
8	8,5	355	1640	2975	8,5	400	1710	3020	6	430	1710	3255	3	1030	1815	3760
9	7,5	410	1410	3015	7,5	475	1460	3055	5	525	1460	3215	2	1195	1860	3805
10	6,5	480	1245	2990	6,5	600	1295	3030	4	710	1320	3200	1	1135	2215	3735
11	5,5	555	1125	2895	5,5	715	1200	2945	3	880	1350	3090	0	1025	2560	3620
12	4,5	600	1050	2710	4,5	770	1180	2780	2	880	1660	2895	—1	930	2650	3720
13	3,5	625	1015	2520	3,5	760	1235	2575	1	825	2015	2780	—2	850	2540	3980
14	2,5	635	1055	2345	2,5	720	1375	2410	0	790	2160	2940	—3	790	2380	3945
15	1,5	625	1205	2215	1,5	680	1595	2305	—1	705	2085	3180	—4	745	2260	3800
16	0,5	600	1440	2180	0,5	640	1765	2360	—2	660	1930	3240	—5	735	2200	3715

равной $2/3$ полной длины речевого тракта (считая от голосовой щели в направлении к губному проходу). Это обстоятельство отражается в сдвиге вперед максимума F_2 на номограммах трехпараметрической модели речевого тракта по мере увеличения степени огубления, как можно видеть на рис. 24, 25 и 27—29.

Интересно выяснить, насколько значения частот формант рупорной модели зависят от фиксированных полостей гортани, т. е. от гортанной трубки и грушевидных пазух, а также насколько изменение отношения длины сужения ко всей длине модели могло бы изменить получаемые результаты.

В табл. 1 даны частоты первых трех формант для четырех конфигураций модели.

Исключение шунтирующих полостей у голосовой щели приводит к увеличению максимальных значений F_1 и F_2 , причем F_1 возрастает наиболее заметно. Данные, приведенные в рубриках *C* и *D* таблицы, характерны для женского и детского речевого аппарата, характеризующегося небольшой длиной. У женщин гортанная трубка и глотка значительно короче, чем у мужчин. Согласно данным Чибя и Каджияма [24], в среднем гортань мужчин на 25% длиннее гортани женщин, различие же в длине ротовой полости составляет лишь 10%.

Стивенс и Хауз [194, 195], обозначаемые в последующих ссылках начальными буквами СХ, составили более детальные номограммы, связывающие параметры трехпараметрической модели с частотами первых трех формант. Их модель имеет меньшую общую длину в случае максимальной делабиализации и относительно большую длину язычного прохода, чем принято в нашей модели на рис. 26—29. Пределы изменений положения сужения в моделях СХ составляют 4—13 см от голосовой щели, что соответствует области 3—12 см для нашей модели из цилиндрических отрезков и 4,5—13,5 см для рупорообразной модели. Если площадь сужения в нашем случае выразить через эквивалентный радиус, то оказывается, что он больше, чем радиус модели СХ в конфигурациях, характерных для передних гласных, предполагая, что F_1 в обоих случаях одинакова. Значения $\frac{A}{l}$ для ротового отверстия у СХ соответствуют непосредственно нашим значениям A_1 , так как в нашей модели l_1 постоянно и равно 1 см для кривых 2, 3, 4 и 5.

Любая из моделей, рассмотренных выше, может быть использована для изучения взаимоотношений между артикуляцией и формантной картиной¹⁾. Используя нашу рупорообразную модель (рис. 26) и номограмму, относящуюся к площади поперечного сечения язычного прохода в $0,65 \text{ см}^2$ (рис. 27), возможно образовать приемлемые варианты

¹⁾ Сравнивая свойства этих моделей речевого тракта с пределами изменения F-картины естественной человеческой речи, необходимо отметить, что в нашей модели F_1 не достигает достаточно высокого максимального значения, а в американской модели затруднительно получить достаточно низкие значения F_2 . Эти различия частично обусловлены слишком большим шунтирующим влиянием полости, представляющей грушевидные пазухи в нашей модели, и слишком длинным язычным проходом в модели СХ. Более естественным было бы уменьшать объем грушевидных пазух в нашей модели по мере того, как место язычного сужения приближается к гортанной области. Данные столбцов *A* и *B* табл. 1 дают представление о том максимальном эффекте, который можно было бы при этом получить.

В противоположность нашей модели длина рупорообразной части в модели СХ зависит от площади поперечного сечения в точке максимального сужения. Зависимость эта, однако, сохраняется только для $A_{\min} > 0,5 \text{ см}^2$. Для радиусов $< 0,4 \text{ см}$ в модель СХ вводится дополнительное местное сужение в центре рупора.

гласных [a] и [i] при положении сужения соответственно на расстоянии 4 и 12 см от голосовой щели и при отсутствии огубления. Гласный [o] можно получить, руководствуясь кривой 3, при положении центра сужения на расстоянии 6 см от голосовой щели, а гласный [u] (кривая 5) — при положении сужения на расстоянии 7—10 см впереди от голосовой щели и площади губного прохода 0,16 см². Наконец, гласная [i] может быть получена в соответствии с кривой 3 при положении сужения на расстоянии 11 см от связок (см. § 2 гл. 7).

Артикуляция зубных согласных может быть воспроизведена в нашей модели при координатах сужения 14—16 см, для которых F_2 изменяется в пределах 1800—1600 гц, а F_3 — в пределах 3050—2900 гц. Как было отмечено Стивенсом и Хаузом, огубление в стадии смычки не называется на F_2 и F_3 . Однако в случае неполной смычки, если допустить, что $A_{\min}$ соизмеримо с площадью губного прохода, может наблюдаться уменьшение F_3 .

Увулярным, велярным и нёбным согласным соответствуют пределы координаты сужения 8—13 см. Увулярному месту артикуляции на расстоянии 8 см от голосовой щели соответствует средний максимум F_3 . В естественной речи этот максимум нейтрализован большей осевой длиной пути распространения звука от связок до губ, за счет обхода высокого заднего подъема в области увулярного сужения. При сопоставлении данных, полученных на модели, с естественной речью необходимо всегда помнить о подобных изменениях длины. Поэтому в случае увулярной артикуляции для всех частот следует вводить поправку порядка 10%. Эта поправка постепенно исчезает по мере продвижения места артикуляции вперед.

Рассматривая F -картины нёбных согласных, интересно отметить, что в то время как при средненёбном положении языка, т. е. при $x = 11$ см, F_2 и F_3 при отсутствии огубления сближаются на частоте приблизительно 2500 гц, огубление такой же степени, как и для гласной [u], т. е. $\frac{A}{l} = 0,16$, ведет к сдвигу F_2 вниз до 750 гц. Изучение рентгенограмм показывает, что различные степени коартикуляции [k] и [g] с различными гласными обуславливаются по преимуществу площадью губного прохода, а не изменением объема передней полости. Уменьшение степени сближения F_2 и F_3 при увеличении нёбно-язычного прохода, имеющем место при переходном процессе после палатализованного звука, видно из рис. 29.

F -картина губных может быть экстраполирована по конфигурации любого гласного путем добавления достаточно узкого губного прохода. До тех пор, пока площадь поперечного сечения этого прохода значительно меньше, чем площадь язычного прохода, это не приводит к заметному дополнительному понижению формант даже при полностью сомкнутых губах.

Следует далее заметить, что ретрофлексная артикуляция не может быть отображена рупорообразной моделью, если не добавить еще одно местное сужение, расположенное в ротовой области. Добавочная фарингализация может, однако, быть воспроизведена подходящим смещением назад рупорообразной части системы, чем создается вторичное место артикуляции.

ЧАСТЬ ВТОРАЯ

РАСЧЕТЫ, ОСНОВАННЫЕ НА РЕНТГЕНОГРАФИЧЕСКИХ ДАННЫХ

ГЛАВА 5

МЕТОДИКА РЕНТГЕНОВСКОГО ИССЛЕДОВАНИЯ, ИСПЫТУЕМОЕ ЛИЦО И ФОНЕТИЧЕСКИЙ МАТЕРИАЛ

Испытуемый — мужчина 38 лет, родившийся в Москве, актер по профессии. При записи русских фонем испытуемому давалась инструкция такого рода: «Сказать [s], как в слове [sat]». В случае длительных и гласных испытуемый сохранял первоначальное положение речевых органов в течение всего периода съемки рентгенограммы. В случае взрывных снимались две рентгенограммы — одна для периода смычки, вторая для положения артикуляторных органов после раствора.

Техника рентгенографии описана А. С. Макмилланом и Г. Келемемом следующим образом¹⁾:

«Очертания языка, свода верхней челюсти и стенок глотки делались более четкими с помощью смеси бария с водой, в которую был добавлен слизистый отвар из акации. Благодаря этому бариевая смесь была более плотной при том же количестве воды. Прилипание ее к слизистой оболочке рта было очень хорошим. Бариевая смесь наносилась шпателем, и испытуемый размазывал ее языком по поверхности рта. Для получения хорошего снимка необходимо, чтобы голова испытуемого находилась в естественном разговорном положении и чтобы это положение сохранялось за все время съемки. В то же время важно, чтобы голова не удерживалась какими-либо скрепками или другими приспособлениями в напряженном состоянии. В поднятом состоянии голова опиралась затылочной областью на подголовник. Нижний угол кассеты располагался на уровне плеч (против дельтовидной мышцы) в 22 см от средней сагитальной линии лица. Чтобы предупредить искажения, обусловленные таким значительным расстоянием кассеты от испытуемого, рентгеновская трубка была удалена на расстояние до 2 м; при этом условии можно пренебрегать искажениями. Центральный луч был направлен через венец первого верхнего коренного зуба.

Поддерживая постоянство этого направления, можно было производить съемку в одних и тех же условиях день ото дня. Для большей гарантии точности позиции головы положение ее фиксировалось деревянным штيفом, на который ориентировался кончик носа. Использовалась вращающаяся рентгеновская трубка, время экспозиций составляло $1/30$ сек при режиме работы 75 кв и 480 ма. Съемки производились на мелкозернистую пленку фирмы Дюпон, проявление велось по стандартной вакуумной методике. Контуры речевого тракта наносились на кальку при помощи просмотрового устройства с сильным источником

¹⁾ A. S. MacMillan, G. Kelemen, Radiography of the Supraglottic Speech Organ, A.M.A. Archives of Otolaryngology 55 (1952), p. 681—682.

света. Внутренняя поверхность верхних резцов, свод верхней челюсти, положения языка, его спинки и основания и положения мягкого нёба зарисовывались вместе с контуром задней стенки фаринкса, надгортаника и черпаловидных хрящей».

Для получения дополнительных данных были сделаны гипсовые слепки ротовой полости испытуемого. Слепки разрезались в нескольких



Рис. 31. Фронтальный снимок испытуемого во время рентгено съемки.

местах и по разрезам определялись точные размеры. Форма губ определялась по фотоснимкам лица испытуемого; см. рис. 31, который был сделан одновременно с рентгено-съемкой.

Произносимые звуки записывались на магнитофонной ленте. К сожалению, акустические свойства рентгеновского кабинета были очень плохими, так что магнитофонные записи могли быть использованы только для сравнения положения формант гласных. Дополнительные записи речи испытуемого¹⁾, сделанные в заглушенной камере, использовались для сравнения данных, по-

лученных в естественных согласных, с вычисленными. Спектрограммы этого речевого материала показаны в Приложениях.

За исключением фонемы [i], снимки делались только одного аллофона данной фонемы. В случае же [i] были изучены как [i], так и [i̥]. Рентгенограмма для последнего варианта фонемы приведена на рис. 32.

ГЛАВА 6

МЕТОДЫ ИССЛЕДОВАНИЯ И СПОСОБЫ АППРОКСИМАЦИИ

Рентгеновское исследование было предпринято для получения основных физиологических данных, необходимых для изучения связи между артикуляцией и акустической картиной, в первую очередь с точки зрения дифференциальных признаков. Теперь, 6 лет спустя, чувствуется, что объем физиологических исследований был непропорционально мал по сравнению с большим количеством вычислительных работ, предпринятых для выполнения намеченной задачи. По мере увеличения объема исследования стало ясно, что было бы желательно иметь более детальные артикуляционные данные, особенно в виде палатограмм, а также использовать не одного, а нескольких дикторов. Однако недавно Конечной и Завадовским [123] было опубликовано обширное исследование русской речи. Оно содержит данные рентгеновского исследования и палатограммы для всех русских фонем, полученные для нескольких дикто-

¹⁾ Запись речи производилась при среднем уровне —10 VU, для того чтобы нелинейные искажения были наименьшими. Подобные же меры предосторожности были приняты при спектрографическом анализе, что необходимо для получения чистых, отчетливых спектрограмм.

ров. Эта работа¹⁾ является ценным дополнением к представленным здесь данным, при получении которых преследовались по преимуществу аналитические цели.

Некоторые из размеров речевого тракта невозможно измерить в условиях естественной речи. Так, при оценке истинной формы и боковых размеров полости фаринкса при заднеязычной артикуляции приходится в известной мере руководствоваться догадками.



Рис. 32. Рентгенограмма при произнесении гласного [ɜ].

Единственное, кроме настоящего, исследование функций площади речевого тракта принадлежит Чибя и Каджияма [24], и потому естественно было привлекать их данные в сомнительных случаях. Так, считаясь с этими данными, было принято, что уменьшение наблюдаемого расстояния между задней стенкой фаринкса и языком, например в задних гласных [ɑ] и [o], сопровождается одновременным уменьшением боковых размеров; таким образом, уменьшение площади идет быстрее, нежели в прямой пропорции к видимому на снимках уменьшению

¹⁾ Другие рентгеновские исследования, имеющие общий интерес, представлены в работах [173, 178, 161, 187, 24, 71, 72]. Дальнейшую библиографию см. в [141].

линейных размеров. Это, конечно, относится и к ротовой полости в тех случаях, когда язык приближается к твердому нёбу.

Другими источниками неопределенности при построении функции площади являются незнание истинного размера воздушного прохода в месте артикуляции фрикативных и аффрикат и недостаточно отчетливое представление об интервале раствора взрывных, т. е. их фрикативной и аспирированной фаз. Величина объема воздуха по обоим сторонам и под кончиком языка также весьма гадательна. Выполненные за последнее время рентгенографические исследования шведской речи¹⁾ показывают, что эта величина подвержена значительным индивидуальным отклонениям. Оценка боковых путей для [I] является чистой догадкой.

Важные данные о размерах носового прохода были получены путем прямых измерений на слепках, снятых на трупах²⁾. Конечно, эти величины не могут быть приняты в качестве нормы, так как известно, что носовая полость значительно варьирует по ширине от одного лица к другому.

Трудно установить, имеет ли место полное закрытие носового прохода или же остается небольшой проход в носовую полость, даже если очертания языка и мягкого нёба видны отчетливо, что не всегда бывает. По-видимому, небольшая степень назальности часто встречается при артикуляции открытых гласных, хотя она и не оказывает заметного влияния на их качество. Для удобства было принято проводить центральную линию в очертании стенок полостей на рентгеновских рисунках в виде сплошной кривой, а представляющие интерес боковые контуры и тыльный контур гортанной трубки обозначать штриховой кривой.

Эффективная центровая линия, характеризующая распространение акустических волн вдоль речевого тракта, может оказаться недостаточно определенной в случае острых изгибов в тракте и сложной формы поперечного сечения.

Как правило, ось x проводилась через центры тяжести поперечных сечений последовательных участков речевого тракта. Начало координат помещалось в той плоскости у губ, где возникает излучение звука; ось x простиралась вплоть до дна гортанной полости. Трудно определить точно, где собственно возникает излучение, поскольку фронт волны у ротового отверстия не является плоским; звуковые волны здесь расходятся и в боковых направлениях. Губная смычка начинается со смыкания боковых частей ротового отверстия, что приводит к некоторому смещению начала координат вперед, даже если губы при этом не вытягиваются. Для нелабиализованных звуков было принято, что начало координат находится не далее чем на расстоянии 0,5 см от передней поверхности зубов.

Возможные ошибки, связанные с неподдающимися точному определению размерами, в общем не являются критическими для акустической интерпретации полученных данных. Размеры по длине являются более важными, чем поперечные, так как большинство формант связано с возникновением стоячих волн. Наличие неопределенностей в определении размеров речевого тракта не должно поэтому удерживать исследователя от максимального использования произведенных измерений. Ошибки, вызванные отклонением в отдельных местах от истин-

¹⁾ Проведенные д-ром Г. Эдхольмом, Королевская больница, Стокгольм.

²⁾ Данные получены д-ром Гуннаром Бьогренгом, Саббатсбергская больница, Стокгольм.

ных размеров, частично компенсируются избыточностью, определяющей условием плавного хода получаемой функции между более точными значениями ее в ряде других точек речевого тракта.

Источники, участвующие в создании звуков речи, трудно поддаются непосредственному изучению прямыми методами. Наиболее подходящим путем для получения распределения энергии источника по частоте и внутреннего импеданса его является анализ физиологических данных, данных о воздушном потоке и спектрограмм исследуемых звуков речи. Использование результатов такого анализа сводится к вычитанию из спектрограмм расчетных значений фильтровой функции; остаток и будет представлять собой характеристику источника. Как указывалось в гл. 2, весьма вероятно, что действие голосового источника, так же как и источника фриктивных звуков, связано с наличием нелинейных импедансов. Однако теоретические основы процесса генерации как голосовых, так и шумовых звуков недостаточно ясны из-за отсутствия исчерпывающих данных о действительных условиях протекания воздуха через сужения речевого тракта.

При расчетах согласных мы комбинировали различные варианты огибающей спектра источника и импеданса звукообразующей щели с тем, чтобы оценить приемлемость тех или иных исходных допущений. Следует учитывать, что на форму спектра фриктивных согласных влияют как наклон огибающей спектра источника, так и его импеданс. Это в известной мере ограничивает возможность однозначного определения истинного наклона огибающей спектра источника путем сопоставления результатов измерений и расчетных данных; подробнее этот вопрос рассмотрен в § 4 гл. 10.

Большинство расчетов спектра сонорных было вначале выполнено на основе наклона огибающей спектра источника, принятого Стивенсом и др. [196] для аналога речевого тракта MIT. Однако сравнение полученных спектров со спектрами естественных звуков показало, что уровни составляющих на низких частотах преуменьшены. Было найдено, что лучшее совпадение дает спектр с огибающей, имеющей падение 12 дБ на октаву. Поэтому все вычисления были выполнены еще раз с использованием такого наклона огибающей. Та и другая огибающие показаны на рис. 33. В обоих случаях к спектру самого источника добавлен подъем 6 дБ/окт для учета частотной характеристики излучения.

Расчеты для русских гласных осуществлялись численными методами, описанными в § 2 гл. 2, а также с помощью быстродействующей вычислительной машины BESK¹⁾. При графическом представлении спектральных данных этой серии вычислений принята была линейная

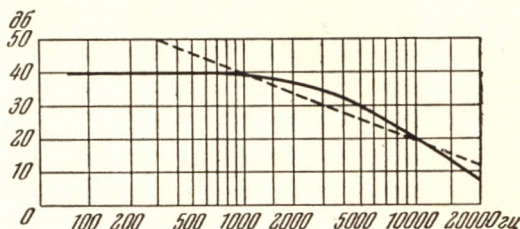


Рис. 33. Огибающие спектра голосового источника с добавлением подъема 6 дБ/окт. Сплошная кривая относится к источнику, принятому на одной из первых стадий расчета; соответствующее этой кривой расположение комплексных частот показано на рис. 11. Пунктирная кривая относится к стандартному голосовому источнику, огибающая спектра которого имеет наклон —12 дБ/окт.

¹⁾ Бинарная электронная вычислительная машина Шведского совета по вычислительной технике. Ранее для вычислений использовалась более простая вычислительная машина BARK.

шкала частот. Все остальные вычисления были выполнены на электрическом аналоге LEA¹⁾; автоматический пишущий индикатор уровня на выходе аналога давал сразу огибающую спектра. Логарифмическая частотная шкала измерителя уровня была сохранена при окончательной обработке данных. Идеалом являлось бы использование некоторого

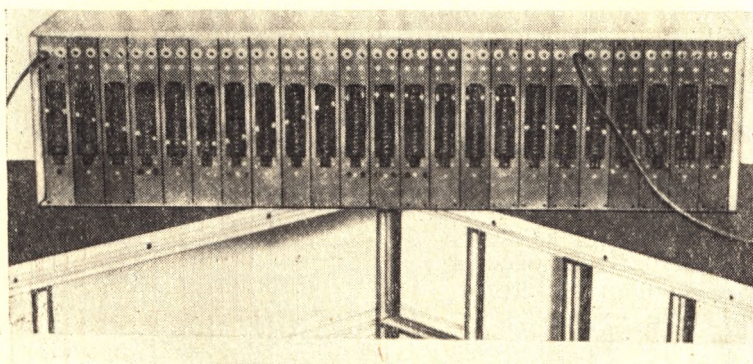


Рис. 34. Электрический аналог (синтезатор речи) LEA, использованный для выполнения большей части приведенных в настоящей работе расчетов. На LEA по положению кнопок отдельных фильтровых звеньев можно непосредственно видеть установленную кривую функции площади (от голосовой щели до губ).

приближения к шкале мелов, т. е. линейной шкалы до 1000 гц и логарифмической для более высоких частот. Логарифмическая шкала удобна для представления характеристик фрикативных и взрывных, но приводит к излишнему подчеркиванию первой форманты гласных и звон-

Таблица 2
Стандартные значения площадей LEA

№	Площадь, см ²	№	Площадь, см ²
1	16,0	9	2,6
2	13,0	10	2,0
3	10,5	11	1,6
4	8,0	12	1,3
5	6,4	13	1,0
6	5,2	14	0,65
7	4,0	15	0,32
8	3,2	16	0,16

ких согласных. Это необходимо помнить при оценке спектральных кривых (см. также § 2 гл. 2). Аналог LEA²⁾ (рис. 34) содержит 45 фильтровых звеньев, каждое из которых представляет участок речевого тракта с длиной по осевой линии 0,5 см и с изменяемой площадью поперечного сечения. Эти площади могут изменяться ступенями от 0,16 см² до 16 см² согласно табл. 2.

Каждое из 45 звеньев включает поперечную емкость C , продольную индуктивность L и сопротивление потерь R_s . Параллельно каждому второму емкостному элементу C включена проводимость $G = \frac{1}{R_p}$. Сопротивления потерь R_s и R_p осуществлены в виде потенциометров. Значения L и C могут изменяться ступенями при помощи ручек, выве-

1) В лаборатории передачи речи Отдела телефонии и телеграфии Королевского технологического института в Стокгольме.

2) Сконструированный В. Эльясоном и С. Вадфорсом.

денных на переднюю панель; эти ручки перемещаются перпендикулярно к горизонтальной «оси распространения» колебаний через горизонтальный ряд последовательных звеньев, как это видно на рис. 34. Положение ручек воспроизводит, таким образом, общий характер функции площади в системе. Такой зрительный контроль оказывается ценным, когда нужно имитировать те или иные конфигурации речевого тракта.

Элементы L и C каждого из звеньев длиной в 0,5 см обычно изменяются одновременно, так чтобы сохранялось постоянство их произведения

$$(LC)^{1/2} = \frac{0,5}{c}, \quad (6.1)$$

где c — скорость звука. При 35°С $c = 35\,300$ см/сек. Граничная частота фильтра равна

$$F_c = \frac{1}{2\pi(LC)^{1/2}} = \frac{c}{\pi} = 11\,250 \text{ гц}. \quad (6.2)$$

Предусмотрена, однако, и возможность изменять значения L и C независимо друг от друга, для того чтобы то или иное звено могло имитировать секцию с длиной большей или меньшей, чем 0,5 см. Отношение L и C определяет площадь поперечного сечения A в аналоге согласно соотношению

$$Z_e = \frac{\rho c}{A} k_e = \left(\frac{L}{C}\right)^{1/2}, \quad (6.3)$$

где Z_e — характеристический импеданс звена в электрических омах, $k_e = 112,5$ — множитель для перевода акустических омов в электрические. Так, если L и C удвоить, то площадь аналога A не изменится, но длина участка увеличится с 0,5 до 1 см. Это приведет также к тому, что граничная частота f понизится до 5625 гц.

Удлинение участков указанным способом применялось при образовании носовых согласных и назализованных гласных. В этом случае 12 звеньев LEA используются для имитации носовой полости. При этом на ротовой и носовой выходы LEA включаются, как нагрузка, отдельные импедансы излучения. Суммирующая схема позволяет объединять оба эти выхода.

Стандартные звенья LEA осуществлены в виде вставных блоков. Соединение между любыми двумя последовательными звеньями может быть разорвано с помощью трехполюсной вилки. Две смежные продольные ветви могут быть, таким образом, по желанию разъединены и в этом месте может быть введен дополнительный источник в случае имитации турбулентного звука. Для присоединения такого источника должен быть использован тщательно отсимметрированный трансформатор.

Сопrotивление источника сонорных звуков было принято равным 200 акустическим омам, т. е. 22 500 электрическим омам. Согласно анализу, проведенному в § 4 гл. 7, затухание, вводимое этим сопротивлением, невелико и довольно хорошо соответствует затуханию при естественной речи. Потери в схемных элементах LEA в общем малы по сравнению с потерями, обусловленными сопротивлением источника и сопротивлением излучения. Ширина формант в спектрах, вычисленных с помощью LEA, в ряде случаев меньше, чем в спектрах естественной речи, особенно для первой форманты. Импеданс излучения в LEA дает демпфирующий эффект, очень близкий к значениям, показанным на рис. 6.

Необходимо упомянуть еще о некоторых других систематических отклонениях от действительности, принятых при вычислениях. Влияние грушевидных пазух, расположенных в глубине полости глотки по обеим сторонам гортанной трубки, не учитывалось, так как их шунтирующее действие не очень легко ввести в вычисления. На рис. 35 показано влияние, которое оказывает исключение этих впадин на огibaющую

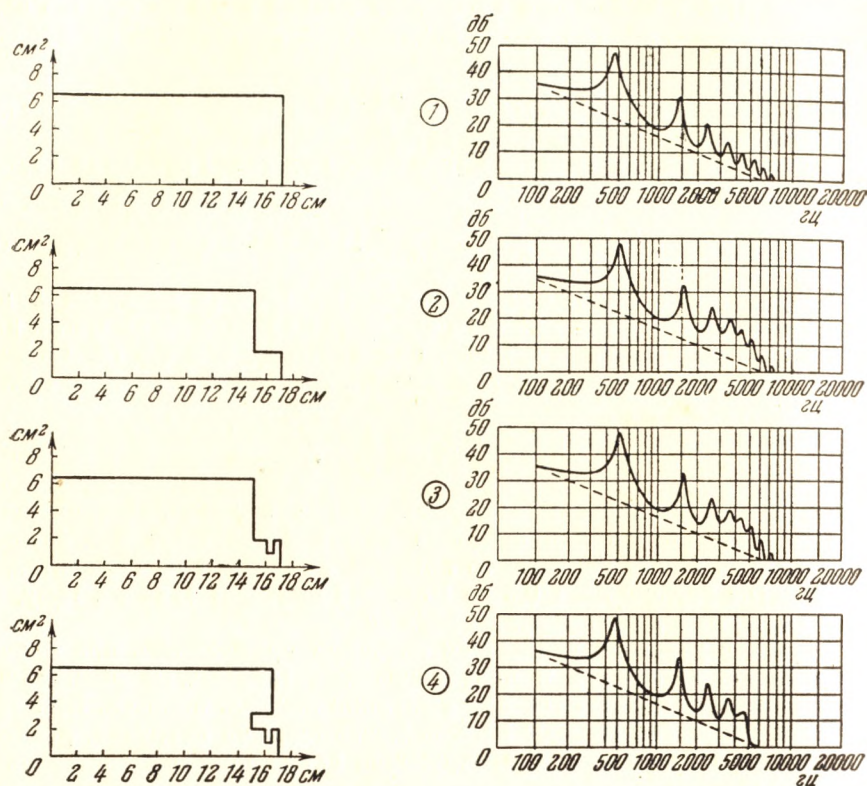


Рис. 35. Функции площади и соответственные огibaющие спектра гласных для резонатора из отрезков цилиндрических труб при различных модификациях такого резонатора в области, отображающей гортань.

- 1) Одиночный отрезок трубы с площадью поперечного сечения $6,5 \text{ см}^2$ и физической длиной 17 см (эффективная длина с учетом концевой поправки на излучение равна $18,2 \text{ см}$).
- 2) Площадь на протяжении первых двух сантиметров от конца трубы уменьшена до 2 см^2 , что соответствует гортанной трубке.
- 3) То же, что (1), но с добавлением ложных голосовых связок.
- 4) То же, что (3), но с добавлением полости, шунтирующей гортанную трубку и отображающей грушевидные пазухи.

спектра. Здесь также показано влияние на спектр гортанной трубки и морганиевых желудочков. Остальная часть тракта здесь представлена простой открытой на конце трубой, что соответствует конфигурации для нейтральной гласной. Можно видеть, что резонатор из одной трубы дает падение огibaющей спектра порядка -6 дБ на октаву, обусловленное выбранным наклоном огibaющей спектра источника -12 дБ/окт. Частоты формант определяются нечетными кратными частоты 475 гц , соответствующей четвертьволновому резонансу $c/4l_e$ трубы.

При уменьшении площади поперечного сечения у дальнего конца трубы образуется нагрузочная трубка длиной 2 см и площадью 2 см^2 ,

имитирующая гортанную трубку, что приводит к увеличению плотности полюсов в области частот 3000—5000 гц благодаря появлению дополнительного полюса и небольшого смещения соседних полюсов. Поэтому

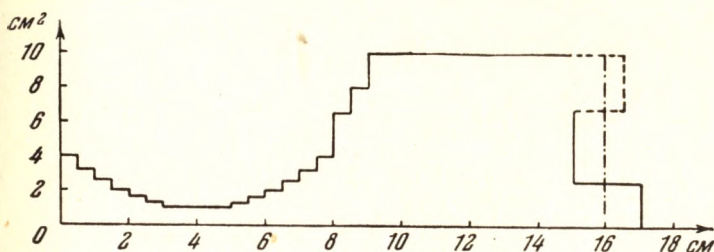
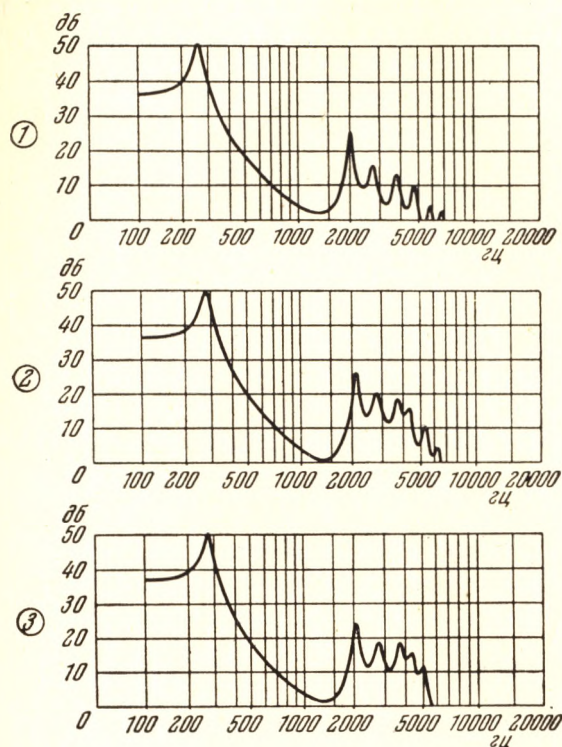


Рис. 36. Огибающие спектра гласного [и], полученные на модели с рупоро-образной частью, отображающей ротовую полость.
 1) Фаринкс и гортань представлены отрезком трубы постоянного сечения; общая длина модели 16 см.
 2) То же, что (1), но с добавлением отрезка трубы длиной 2 см, представляющего гортань, и с укорочением отрезка, соответствующего фаринксу; общая длина 17 см.
 3) То же, что (2), но с добавлением шунта, отображающего грушевидные пазухи.

есть основания ассоциировать F_4 с резонансом гортанной трубки, как это предлагали Совиярви [187, 188] и Чибя и Каджияма [24] (см. § 3 гл. 7). Уровни в этой частотной области несколько повышаются, но зато выше 5000 гц падение огибающей становится круче. Введение

ложных голосовых связок, как показано на кривой 3, дает лишь небольшой добавочный эффект, сходный с влиянием низкочастотного фильтра. Наконец, из кривой 4 видно, что шунтирующее влияние грушевидных пазух резко увеличивает крутизну спада после 5000 гц, что вызвано появлением нуля, т. е. антирезонанса, чуть выше этой

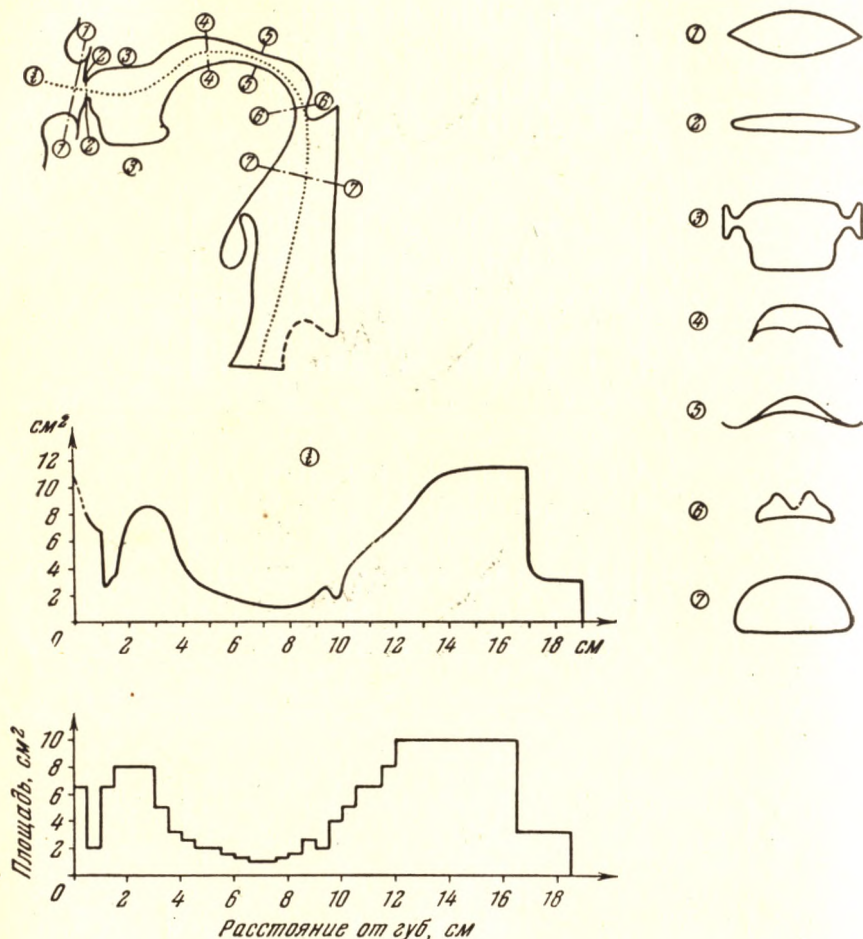


Рис. 37. Способ получения ступенчатого приближения к функции площади на основании рентгенограмм. Определяются площади для ряда поперечных сечений речевого тракта, нормальных к осевой линии, на протяжении от голосовой щели до губ, после чего строится плавная кривая, представляющая собою функцию площади. Эта кривая разбивается на ряд ступеней с длиной по оси абсцисс 0,5 см, как это требуется для численных расчетов или для расчета при помощи ЛЭА.

частоты. Можно видеть далее, что небольшое повышение частоты всех формант, лежащих ниже 3500 гц, вызываемое наличием гортанной трубки, компенсируется шунтирующим действием грушевидных пазух.

Таким образом, влияние системы нагрузочных полостей гортани сказывается главным образом в явно выраженном подавлении высоких частот, начиная с граничной частоты 5000 гц. Как указывает

Ван-ден-Берг [12], подобные эффекты¹⁾ можно учесть, отнеся их к характеристикам источника.

На рис. 36 показано ступенчатое приближение к модели звука [i] (рис. 20), образованной из рупорообразной части, имитирующей ротовую полость, и цилиндра, воспроизводящего фарингальную полость. Из рис. 36 можно видеть, что ступенчатое приближение дает формантные частоты (кривая 1), очень близкие к полученным ранее на основании рупорообразной модели, и что добавление гортанной трубки (кривая 2) и грушевидных пазух (кривая 3) дает такой же эффект подавления высоких частот, как и в случае простой трубы на рис. 35. В модели рис. 36 объем задней полости увеличен так, чтобы включать объем гортанных полостей, и полная длина модели равна 16 см; к этому случаю и относится кривая 1 на этом рисунке.

Из этих данных можно видеть, что теоретически пренебрежение грушевидными пазухами, допущенное во всех дальнейших вычислениях в настоящей работе, приводит к несколько преувеличенной оценке частот тех формант, которые зависят в основном от фарингальной полости. Этот эффект имеет, вероятно, наибольшую величину в случае задних гласных (см. табл. 1, стр. 96)²⁾. Кроме этого эффекта, указанное пренебрежение приводит к тому, что подавление области спектра выше 5000 *гц* оказывается менее выраженным.

Процесс получения по рентгенограмме функции площади, необходимой для вычислений, иллюстрируется рис. 37 для случая гласного [i]; рентгенограмма этого гласного представлена на рис. 32.

Первым шагом является по возможности близкая к действительности оценка формы и площади поперечных сечений речевого тракта для ряда точек вдоль осевой линии. При этом могут оказаться необходимыми некоторые уточнения самой осевой линии. Определение площади производится на основе всех данных, которые могут быть для этого использованы, включая зубные слепки и фронтальные рентгенограммы фаринкса. Подобные измерения дают отправные точки для построения функции площадей. Промежуточные значения после этого находятся по медиальной рентгенограмме с учетом условий непрерывности функции. Наконец, полученная функция преобразуется в ступенчатую кривую, необходимую для проведения численных расчетов или расчетов с помощью LEA.

ГЛАВА 7

ИССЛЕДОВАНИЕ ГЛАСНЫХ

§ 1. Расчетное определение частот формант и огибающих спектра

Контуры речевого тракта, полученные на основе рентгеновских снимков (рис. 38), и соответствующие им функции площади (рис. 39) для гласных [a], [o], [u], [i], [i], [e] дают исходные физиологические

¹⁾ Фильтровый эффект должен быть в основном приписан комбинации гортанной трубки и грушевидных пазух, а не морганиевым желудочкам, как это предполагает Ван-ден-Берг.

²⁾ Интересно отметить, что расчетные данные для гласных лучше совпадают с естественными в том случае, когда не учитывается влияние грушевидных пазух. Возможно, это объясняется компенсаторным увеличением F_1 , вызванным небольшим шунтирующим действием носовой полости, голосовой щели или фарингальных полостей, хотя здесь могут сказаться и другие источники систематических ошибок. Указанная тенденция, уже отмечавшаяся в § 2 гл. 4, подтверждается результатами проведенных за последнее время расчетов для шведских гласных.

данные для расчетов. Прежде чем перейти к рассмотрению особенностей образования этих гласных, приведем вкратце результаты расчетов.

Сопоставление частот формант протяженных звуков, произносимых испытуемым во время съемки рентгенограмм, с частотами формант, вычисленными на основании функций площади с помощью BESK, дано

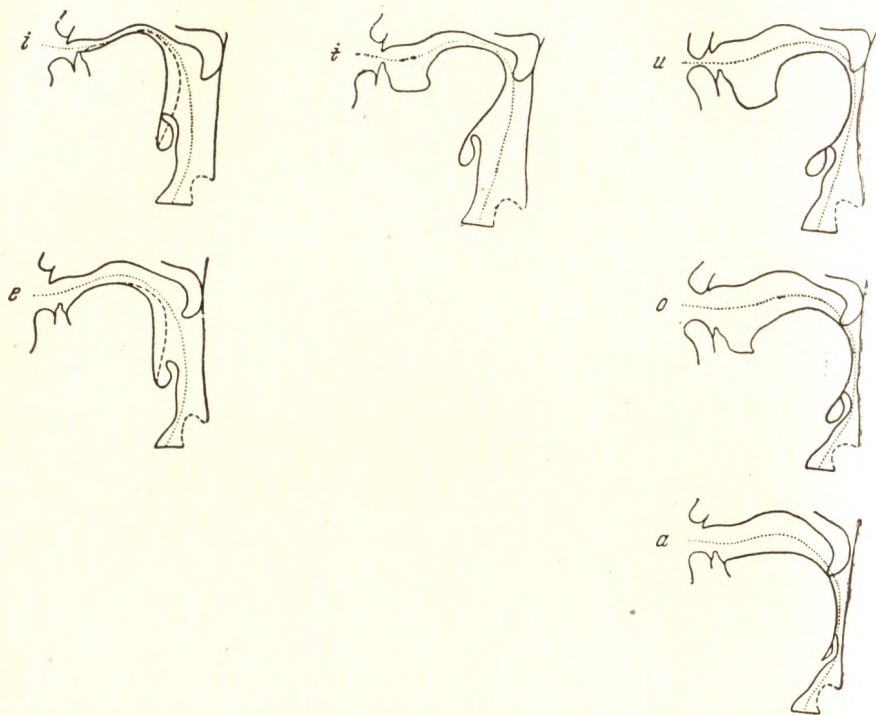


Рис. 38. Контурсы речевого тракта в медиальном сечении по данным рентгенограмм для русских гласных [а], [о], [у], [й], [и], [е].

на рис. 40. Совпадение оказалось даже более полным, чем ожидалось; отметим, например, почти полное совпадение F_1 и F_2 для [i] и [й]. Среднее отклонение расчетных данных от действительных было около 5% для F_2 и F_3 и 10% для F_1 . Максимальное отклонение F_1 наблюдалось для гласного [а]; оно составляло 84 гц. Для F_2 и F_3 наибольшее абсолютное различие между действительными и вычисленными данными имеет место для звука [е], где расчетные значения на 200 гц выше действительных.

Расчеты были повторены и на LEA для выяснения точности результатов, обеспечиваемой этим устройством, и установления возможности применения его в дальнейшей работе. Средние квадратичные отклонения данных, полученных на BESK и на LEA, составляют 3% для F_1 , 1,3% для F_2 и 1,6% для F_3 . По сравнению с BESK, LEA дает в среднем отклонение на +0,3% для F_1 , -0,8% для F_2 , -1,8% для F_3 . Это объясняется частично различными видами аппроксимации, так как при использовании LEA данные для функции площади квантованы шестнадцатью уровнями, как показано в табл. 3. Точность вычислений на LEA вполне достаточна для практических целей.

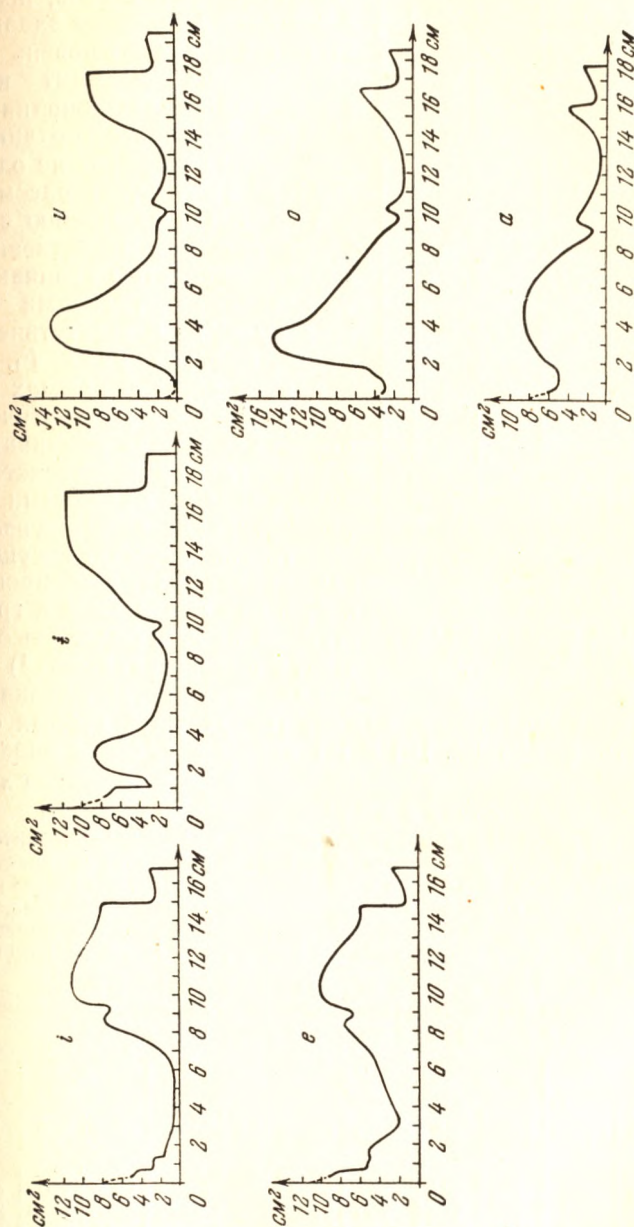


Рис. 39. Функции площади, полученные по данным рис. 38.

Огибающие спектра, полученные с помощью BESK, показаны на рис. 41. Они имеют в общем такой же вид, что и огибающие спектров

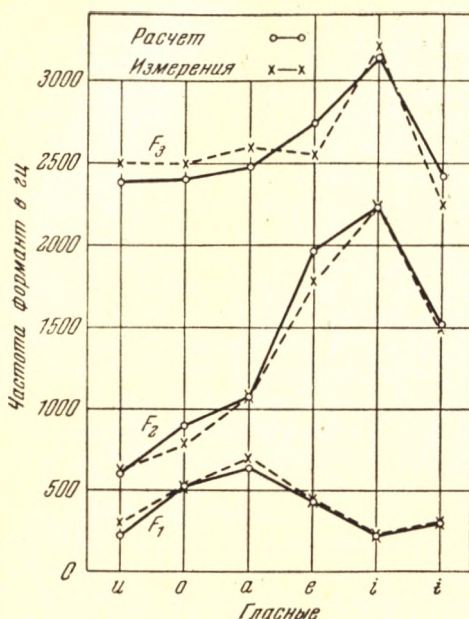


Рис. 40. Частоты формант шести русских гласных, полученные расчетом и измерением. Для расчета использовалась быстродействующая вычислительная машина BESK.

гипотетических звуков, полученных в § 3 гл. 3 только по заданным частотам формант. Уровень F_3 и более высоких формант несколько выше благодаря дополнительному влиянию резонанса гортанной трубки. Ширина формант не одинакова; причина этого будет рассмотрена в § 3 гл. 7. По сравнению с огибающими спектров естественной речи расчетные спектры отличаются преувеличенными уровнями для частот выше F_3 и отсутствием «голосовой полосы» ниже F_1 , которая видна на спектрограммах для данного испытуемого. Акустические соотношения между различными русскими гласными, представленные диаграммами, подобными рис. 40, рассматривались Джоунзом [119].

Третья форманта существенна как основа для нормирования F_1 и F_2 и как дополнение к группе, состоящей из F_2 и более высоких формант у высокотоновых¹⁾ гласных. Однако для установления акустических соотношений достаточно, вообще говоря, двух первых формант.

Низкотоновость фонем [u], [o] и [a] связана с более низким F_2 по

Таблица 3

Сравнение расчетных и измеренных частот формант

BESK: Численные расчеты на быстродействующей цифровой вычислительной машине. Речевой тракт разбит на 20 последовательных однородных участков, обладающих каждый определенной площадью поперечного сечения и определенной длиной.

LEA: Каждое последовательное звено, соответствующее отрезку длиной 0,5 см по оси речевого тракта, состояло из индуктивности, включенной в продольное плечо звена, и емкости в поперечном плече, которые выбирались так, чтобы при квантовании функции площади соседние сечения отличались множителем $\sqrt[3]{2}$.

Измеренные значения частот формант получены путем спектрографического анализа звука, произносимого испытуемым во время рентгено съемки.

Гласные	F_1			F_2			F_3			F_4		F_5	
	BESK	LEA	Испыт.	BESK	LEA	Испыт.	BESK	LEA	Испыт.	BESK	LEA	BESK	LEA
[u]	231	240	300	615	610	625	2375	2370	2500	3320	3400	4000	3950
[o]	510	500	535	900	860	780	2400	2320	2500	3220	3500	3920	3800
[a]	616	630	700	1072	1072	1080	2470	2400	2600	3410	3550	3820	4000
[e]	432	420	440	1959	1960	1800	2722	2750	2550	3500	3410	4400	4150
[i]	222	230	240	2244	2220	2250	3140	2970	3200	3700	3570	4655	4400
[i]	296	285	300	1577	1480	1480	2413	2320	2230	3450	3200	4150	4200

¹⁾ В русской оригинальной и переводной литературе сейчас по преимуществу приняты используемые здесь наименования дифференциальных признаков. (Прим. перев.)

сравнению с другими гласными. Различие между [e] и [i] проще всего сформулировать, используя в качестве критерия форманту F_1 , которая в первом гласном выше, чем во втором.

Этот критерий компактности может быть также использован и для ряда [a], [o], [u]. Продвинутая вперед артикуляция [u], [o] и [a] между

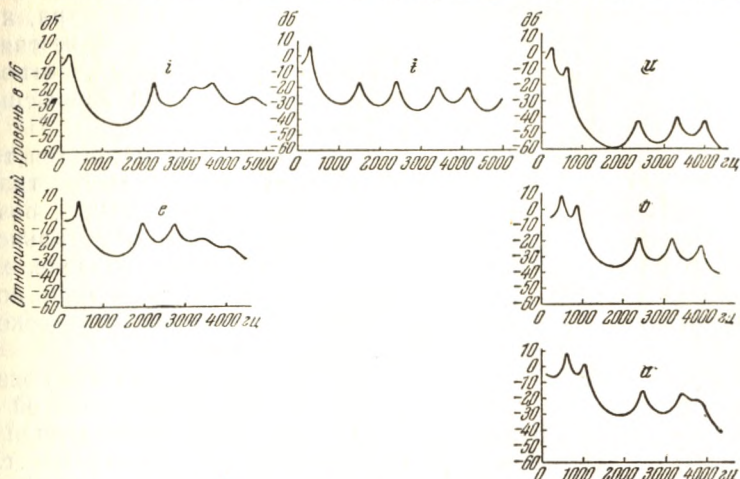


Рис. 41. Огибающие спектров шести русских гласных, полученные расчетом при помощи вычислительной машины BESK. Спектральные уровни на частотах выше 3000 гц сильно преувеличены, что указывает на более быстрое, чем 12 об/окт, спадание огибающей спектра голосового источника на высоких частотах.

двумя дизными¹⁾ (палатализованными) согласными ведет к появлению не рассматриваемых здесь аллофонов [j̥], [ø] и [æ]²⁾, акустически отличающихся более высокой F_2 . Равным образом палатализация аллофона [i], когда этот гласный находится рядом с дизным согласным, и аллофона [e], когда этот гласный стоит перед дизным согласным, характеризуется более высокими F_2 и F_3 и более низкой F_1 , чем соответственно [i] и [e].

§ 2. Артикуляторные и акустические диаграммы гласных

Имеются два существенно различных метода артикуляторного описания гласных. Одним из них является классическое описание по положению наивысшей точки языка по горизонтали и вертикали. Другой заключается в более или менее точном описании размеров воздушных полостей, расположенных друг за другом вдоль речевого тракта от голосовой щели до губ. Наиболее полной формой такого описания является график функции площади. Второй метод дает численные величины, необходимые для определения математическим путем резонансных частот речевого тракта, тогда как первый представляет собой попросту подходящую условную основу для фонетической классификации. Положение наивысшей точки языка хорошо коррелирует с соответствующими акустическими данными, но не характеризует однозначно размеры резонатора.

¹⁾ См. предыдущее примечание. (Прим. перев.)

²⁾ См., например, [18].

Одним из приближенных методов представления геометрии речевого тракта является трехпараметрическая модель, введенная Стивенсом и Хаузом и примененная здесь в несколько измененном виде (§ 3 гл. 4). В такой модели место и площадь поперечного сечения язычного прохода вместе с показателем огубления однозначно определяют все размеры системы; однако первые два параметра, как будет

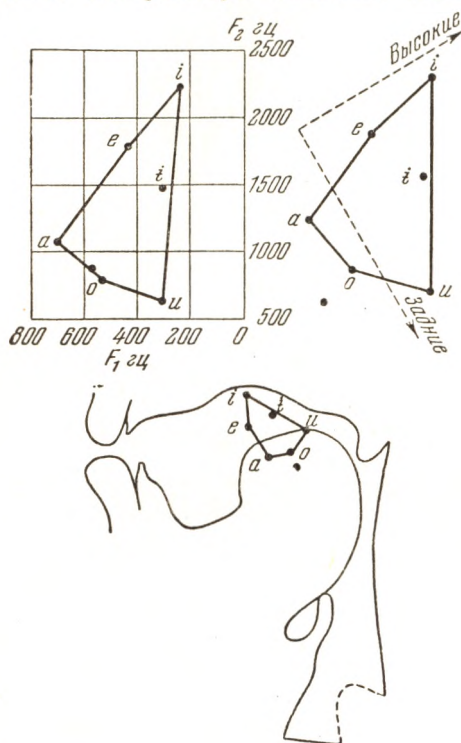


Рис. 42. Акустическая и артикуляторная характеристики гласных в двух измерениях. Диаграмма частот формант в координатах F_1 , F_2 подобна диаграмме, дающей положение наивысшей точки языка для отдельных гласных. Контур речевого тракта приведен для гласного [u].

показано в § 3 гл. 7, не связаны какими-либо простыми соотношениями с перемещениями наивысшей точки языка. Широко применяется также интерпретация речевого тракта как сдвоенного резонатора; поэтому такое представление детально рассмотрено в данном разделе, хотя применимость этой модели по сравнению с трехпараметрической более ограничена. Настоящий раздел посвящен сопоставлению классической диаграммы гласных, отражающей положение языка, и диаграмм гласных в формантной плоскости F_1 , F_2 . Причинные связи между размерами резонатора и частотами формант будут рассмотрены в следующем разделе. На контурах речевого тракта, полученных по рентгенограммам (рис. 38), положение языка можно определить как наивысшую точку поверхности языка в средней (сагиттальной) плоскости. Эти точки были определены для каждого гласного и затем нанесены на контур, относящийся к гласному [u], как это показано на рис. 42. Полученная таким образом фигура отражает традиционные отношения между наиболее задними [u] и [o], причем [u] соответствует более высокому положению наивысшей точки языка, чем у []. Таким же образом из гласных [i] и [e], являющихся наиболее передними, [i] характеризуется более высоким положением наивысшей точки, чем [e]. Для гласного [ɛ] характерно положение этой точки между таковыми для [u] и [i], а для [a] — ниже и более вперед, чем для [o]. На рис. 42 представлена та же диаграмма положения верхней точки языка после ее поворота на 60° по часовой стрелке и линейного увеличения ее размеров. Такое преобразование приводит к поразительному сходству артикуляторной диаграммы с диаграммой, представляющей зависимость F_2 от F_1 , если значение F_1 откладывать влево так, чтобы образующаяся фигура была расположена во «втором квадранте». Это сходство обеих диаграмм оказывается даже большим, чем можно было ожидать на основании аналогичных исследований Джюза [121]. Следует отметить, что смещение языка, ведущее к сдвигу вниз по вертикали точки, при-

нятой за показатель его положения, связано не только с возрастанием F_1 , но одновременно и с некоторым уменьшением F_2 . Аналогично перемещение верхней точки языка вперед по горизонтали ведет к повышению F_1 наряду с более заметным повышением F_2 . Эти вторичные эффекты — уменьшение F_2 при понижении языка и возрастание F_1 при продвижении его вперед — несколько усиливаются, если диктор откидывает голову назад. У нашего диктора была склонность к такому положению головы, что надо принимать во внимание, сравнивая наши результаты с данными других авторов, например Д. Джуонза (1934 г.) [118].

Упомянутые выше вторичные эффекты, связанные со сдвигом языка, уменьшаются, если высоту последнего измерять не по вертикали, а по нормали к своду заднего отдела полости рта в области мягкого нёба и передней части нёбной занавески. Подъем языка, или, более точно, сдвиг выбранной представляющей точки перпендикулярно к этой поверхности обуславливает уменьшение F_1 ¹⁾ при почти постоянном F_2 , а сдвиг этой точки параллельно поверхности ведет к возрастанию F_2 при относительно постоянном F_1 .

Степень такой корреляции, конечно, зависит, помимо перемещения языка, от степени огубления и других изменений, влияющих на конфигурацию воздушных камер, например от изменений положения гортани или мягкого нёба и от понижения челюсти. На рис. 42 показана еще одна дополнительная точка положения языка для гласного [o]. Этот звук был образован с более низким положением языка, бóльшим открытием челюстей и бóльшим огублением, чем при обычном [o], также показанном на рисунке. Различия между этими вариантами на акустической диаграмме F_1F_2 меньше, чем на артикуляторной диаграмме, и идут в ином направлении.

§ 3. Соотношения между размерами резонаторов и формантными частотами

В теории речеобразования часто встречаются три неправильных положения, которые уже упоминались в этой работе. Одно из них относится к участию в речеобразовании фаринкса. В классических артикуляторных схемах, например Флетчера [70] и Д. Джуонза [118], показывается неизменная фарингальная полость, относительно которой строятся различные профили верхней части языка. Это приводит ко второму неправильному положению, согласно которому наивысшая точка языка совпадает с местом минимальной площади поперечного сечения в речевом тракте, отделяющим переднюю часть его от задней.

Имеющиеся в фонетической литературе рентгенографические данные не дают оснований считать эти гипотезы правильными. Большие изменения фарингальной полости и образование в ней области максимального сужения, например при задних гласных, были отмечены многими исследователями — Расселом [178], Хефнером [94], Дэнном [39], Мальмбергом [144].

Третье ошибочное положение является следствием чрезмерно упрощенного рассмотрения акустических свойств сложной системы резонаторов. Популярное представление о связи F_1 с задней полостью и F_2 с передней лишь чисто внешне соответствует факту корреляции между

¹⁾ Что согласуется с более ранними данными Делаттра [30].

уменьшением ротовой полости и увеличением фарингальной в ряду [a], [æ], [e] и [i], что связано с уменьшением F_1 и увеличением F_2 . Несмотря на фундаментальный вклад в понимание происхождения формант гласных, сделанный Дэнном [39], неполная и частично ошибочная упрощенная теория сдвоенного резонатора до сих пор доминирует в фонетической литературе.

История развития акустической теории резонансов гласных представляется, конечно, интересной, но подробный обзор имел бы сомнительную ценность из-за несовершенства инструментальных методов и теоретического анализа в ранних работах.

Наши данные сходятся во многом с выводами Дэнна [39] о природе образования трех первых формант; однако настоящее исследование характеризуется более широким охватом вопроса и использует более полную модель речевого тракта. Возникновение некоторых резонансов, например F_3 , вследствие образования стоячих волн, по-видимому, не было достаточно ясно понято некоторыми фонетистами, цитировавшими Дэнна. Новая книга Хэла [80] содержит большое число ссылок на старую европейскую литературу, дополняющих исторические сведения, приведенные у Дэнна [39] и Чибя и Каджияма [24]. Предлагаемая Хэла физиолого-акустическая интерпретация как более высоких, чем второй, так и дополнительных резонансов не-правильна и отражает тенденцию теоретиков старой школы связывать каждую форманту с отдельной полостью или с отдельным источником звука.

Интересно, однако, что Хэла приходит к выводу, что при сравнении [u] и [o] имеет место обратная классическим представлениям связь F_1 с задней полостью и F_2 с передней полостью, что в известном смысле, если сравнивать задние и передние гласные, соответствует нашим расчетам.

Ярким примером интуитивного понимания связи между формантами и полостями является система физиологической классификации, предложенная Совиярви [187, 188] (частично использующая и встречающиеся в более ранней литературе предположения). Эта система учитывает наличие большого числа формант и соответствующих им субполостей или областей в пределах речевого тракта (7 изменяющихся и 11 постоянных). В некоторых своих частях эта система является следствием неполноты используемой этим автором акустической теории; однако некоторые положения, например относящиеся к резонансу гортанной трубки, являются правильными и находят подтверждение в нашей работе; другие же положения, например такие, как о резонансах носовой полости и резонансах трахеи, правильны частично или требуют более детального рассмотрения.

Поразительно простая теория взаимоотношений между F_1 , F_2 и F_3 была предложена Ганешсундарамом [76]. Его теория «каскадной модуляции» предполагает, что $F_3 - F_2$ всегда равно $2F_1$, так что F_3 является верхней, а F_2 — нижней боковой полостью, расположенными на расстоянии $\pm F_1$ от резонанса ротовой полости F_2 , подавленного по амплитуде.

Отвлекаясь от не соответствующих действительности физических предпосылок этой теории, противоречащих принципу линейности системы речевого тракта, надо признать, что основные наблюдения автора приблизительно соответствуют данным для некоторого ограниченного числа гласных. Как видно из описанных в гл. 4 моделей (из двух отрезков труб и трехпараметрической), имеется явная тенденция к изменениям F_1 в направлении, противоположном изменению F_2 , и к изменениям F_2 , противоположным движению F_3 . Это в особенности справедливо для конфигурации, представленной двумя отрезками труб, при одинаковой длине этих отрезков; здесь правило, предложенное Ганешсундарамом, выдерживается точно. Частично приложимость его к речи может быть объяснена тенденцией языка занимать положение между палатальной и фарингальной точками максимального сужения и тенденцией сужения занимать половину длины речевого тракта. Однако имеются и заметные отклонения от правила $F_3 - F_2 = 2F_1$ (или такого же правила с заменой F_2 на F_1 и наоборот, как это предлагается для задних гласных), такие, например, как очень большое расстояние между F_3 и F_2 при очень низкой F_3 для [i] и, наоборот, высокая F_1 при близости F_3 и F_2 , как в [æ] и в ретрофлексных вариантах гласных.

Целью настоящего раздела является дополнение приведенной в гл. 4 и Приложении III общей теории моделей речевого тракта акустическими и физиологическими данными, полученными для шести русских гласных, исследованных в настоящей работе. Среди других вопросов,

представляющих существенный интерес, рассмотрен и вопрос о пригодности сдвоенного резонатора Гельмгольца в качестве модели речевого тракта и применимости трехпараметрических номограмм (гл. 4, § 3, Б) к описанию исследованных гласных. Детально рассмотрено также количественное выражение относительной роли той или иной части речевого тракта как фактора, определяющего каждую из формант.

В разделе Б табл. 4 сведены артикуляторные данные, характеризующие функции площади речевого тракта для различных гласных. Хорошее соответствие частот формант, полученных расчетом и измеренных по магнитофонным записям данного испытуемого (рис. 39), является гарантией достаточной правильности (в основном) функций площади, хотя и не исключена возможность некоторых отклонений в деталях, что отмечалось выше, в гл. 6.

Одной из наиболее бросающихся в глаза особенностей функций площади является обратная конфигурация для [а] по сравнению с [i], что подтверждает положения гл. 4.

В ряду [а], [о], [u], [i], [i], [e] место артикуляции, определяющееся эффективным центром язычного сужения, движется вверх и вперед от точки, расположенной на 4 см выше голосовой щели, к точке, находящейся в 4 см сзади от излучающей поверхности ротового отверстия. Длина системы резонаторов наибольшая для звука [u] и равна в этом случае 19,5 см вследствие большого пути воздушного потока, огибающего высоко поднятую и отодвинутую назад спинку языка. Площадь поперечного сечения A_{2mln} в истинном месте артикуляции изменяется от звука к звуку (в порядке их следования в табл. 4) относительно мало, за исключением перехода от [i] к [e]. В среднем A_{2mln} равна примерно 1 см².

Если учитывать только положение наивысшей точки языка, то можно говорить о возрастании степени открытости ротовой полости в ряду [u], [о], [а]. Если же рассматривать основное сужение в системе, то в этом ряду имеет место приблизительно постоянная степень открытости, а для данного диктора — даже некоторая тенденция к увеличению закрытости.

Между высотой подъема языка и положением мягкого нёба существует отчетливо выраженная положительная корреляция, отвечающая неполному закрытию прохода в носовые полости при очень открытых гласных, наблюдаемому вполне определенно в случае [а], но сомнительному для [e] и [о]. Влияние различной степени назализации на характеристики спектра гласных будет разобрано в § 1 гл. 8.

Систематизация размеров резонаторов может быть осуществлена, как было отмечено в § 2 гл. 7, несколькими различными методами.

Один из вариантов такой систематизации заключается в указании размеров ротовой и фарингальной полостей, исходя из определенного физиологического их разграничения.

Это означает, что граница между двумя полостями будет установлена независимо от артикуляции, например в определенной фиксированной плоскости в области язычка. Тогда задние гласные [u], [о], [а] будут характеризоваться большим объемом ротовой полости, чем передние гласные [i], [i], [e].

Если же в качестве границы между передними и задними полостями принимать положение истинного места артикуляции, то можно видеть, что отношение объемов передней и задней полостей может быть

Таблица 4

А. Площадь поперечного сечения А в зависимости от расстояния x , отсчитываемого от переднего конца речевого тракта. Данные для шести гласных квантованы ступенями в $2^{1/3}$ в соответствии с техническими характеристиками электрического аналога LEA

	[a]	[o]	[u]	[i]	[l]	[e]
$x, \text{см}$	$A, \text{см}^2$	$A, \text{см}^2$	$A, \text{см}^2$	$A, \text{см}^2$	$A, \text{см}^2$	$A, \text{см}^2$
0,0	5,0	3,2	0,65	6,5	4,0	8,0
0,5	5,0	3,2	0,65	6,5	4,0	8,0
1,0	5,0	3,2	0,32	2,0	3,2	5,0
1,5	5,0	3,2	0,32	6,5	1,6	5,0
2,0	6,5	6,5	2,0	8,0	1,3	4,0
2,5	8,0	13,0	5,0	8,0	1,0	2,6
3,0	8,0	13,0	10,5	8,0	0,65	2,0
3,5	8,0	16,0	13,0	5,0	0,65	2,6
4,0	8,0	13,0	13,0	3,2	0,65	2,6
4,5	8,0	10,5	13,0	2,6	0,65	3,2
5,0	8,0	10,5	13,0	2,0	0,65	4,0
5,5	8,0	8,0	10,5	2,0	0,65	4,0
6,0	8,0	8,0	8,0	1,6	0,65	4,0
6,5	6,5	6,5	6,5	1,3	1,3	5,0
7,0	5,0	6,5	5,0	1,0	2,6	5,0
7,5	4,0	5,0	3,2	1,0	4,0	6,5
8,0	3,2	5,0	2,6	1,3	6,5	8,0
8,5	1,6	4,0	2,0	1,6	8,0	6,5
9,0	2,6	3,2	2,0	2,6	8,0	8,0
9,5	2,6	2,0	2,0	2,0	10,5	10,5
10,0	2,0	1,6	1,6	4,0	10,5	10,5
10,5	1,6	2,6	1,3	5,0	10,5	10,5
11,0	1,3	1,3	2,0	6,5	10,5	10,5
11,5	1,0	0,65	1,6	6,5	10,5	10,5
12,0	0,65	0,65	1,0	8,0	10,5	8,0
12,5	0,65	1,0	1,0	10,5	10,5	8,0
13,0	0,65	1,0	1,0	10,5	10,5	6,5
13,5	1,0	1,3	1,3	10,5	10,5	6,5
14,0	1,6	1,6	1,6	10,5	8,0	6,5
14,5	2,6	2,0	3,2	10,5	8,0	6,5
15,0	4,0	3,2	5,0	13,0	2,0	1,3
15,5	1,0	4,0	8,0	13,0	2,0	1,6
16,0	1,3	5,0	8,0	10,5	2,6	2,0
16,5	1,6	5,0	10,5	10,5	3,2	2,6
17,0	2,6	1,3	10,5	—	—	—
17,5	—	1,3	10,5	3,2	—	—
18,0	—	1,6	2,0	3,2	—	—
18,5	—	2,6	2,0	3,2	—	—
19,0	—	—	2,6	3,2	—	—
19,5	—	—	2,6	—	—	—

Б. Параметры резонатора, полученные по функциям площади, приведенным в разделе А. Гласные следуют в порядке места расположения язычного сужения

		[a]	[o]	[u]	[ɨ]	[i]	[e]
Полная длина, см		17,0	18,5	19,5	18,5	16,5	16,5
Расстояние центра сужения от язычного сужения, см	от переднего конца	13,0	12,0	11,0	7,5	4,0	4,0
	от заднего конца	4,0	6,5	8,5	11,0	12,5	12,5
Минимальная площадь язычного сужения $A_2 \text{ min}$, см ²		0,7	0,8	1,0	1,1	0,5	2,2
Объем передней полости V_1 , см ³		59,0	70,0	54,0	27,0	6,0	11,0
Объем задней полости V_2 , см ³		8,4	13,0	31,0	72,0	73,0	76,0
Отношение длины горла резонатора к его площади, 1/см	l_1/A_1	0,75	0,9	5,3	0,7	0,5	0,6
	l_2/A_2	5,2	4,2	4,2	3,5	7,5	1,4
Отношение объемов V_1/V_2		7,0	5,4	1,7	0,38	0,08	0,14
Параметр $R = \frac{V_1 l_1 A_2}{V_2 l_2 A_1}$		1,0	1,2	2,2	0,076	0,053	0,06

использовано для отграничения [u], [o], [a] от остальных гласных, а также в качестве характеристики соотношений между [o], с одной стороны, и [u] и [a], с другой. Среди передних гласных отношение объемов отличает [e] от [i]; к сожалению, у нас не было «твердого» варианта [e], который можно было бы сравнить с [ɨ] — «твердым» вариантом [i]. Данные показывают для [ɨ] большую величину отношения объемов, чем для [e], представленного «мягким», т. е. палатализованным, вариантом. Если же использовать в качестве критерия отношение объема рта к объему фарингальной полости, то можно обнаружить, что как [i], так и [ɨ] образуются при меньшем отношении объемов, чем [e].

Здесь возникает существенный вопрос о том, возможно ли определение координат центра артикуляторного сужения с большей точностью. Рассматривая функцию площади, приведенную на рис. 39, можно видеть, что координата центра определяется довольно легко. Для правильного определения этой точки и связанного с ним $A_{2\text{min}}$ надо принять во внимание всю форму образуемого языком горба и решить, как расположить рупорообразное соединительное звено модели так, чтобы обеспечить наилучшее соответствие с исходной функцией площади. При асимметричных сужениях получаемая при этом координата не совпадает в точности с координатой абсолютного минимума функции площади. Не следует также думать, что $A_{2\text{min}}$ будет точно совпадать со значением минимума площади, получаемым непосредственным измерением.

С другой стороны, для открытых гласных типа [æ], [œ] место артикуляции становится неопределенным и потому не может быть использовано как граница, необходимая для определения соотношения

полостей. В таких случаях нет физических предпосылок для использования сдвоенного резонатора Гельмгольца в качестве математической модели речевого тракта.

Задание же места сужения и его площади поперечного сечения в соответствии с трехпараметрическим представлением позволяет с достаточной точностью определить F -картину, так как ошибка при определении местоположения центра сужения быстро уменьшается по мере приближения конфигурации речевого тракта к трубе постоянного сечения, для которой формантная картина определяется только полной длиной и степенью огубления.

Следует отметить, что при представлении речевого тракта в виде сдвоенного резонатора Гельмгольца внутреннее сужение принимается целиком за горло заднего резонатора. Поэтому граница между двумя резонаторами не совпадает с центром сужения. Данные о размерах горла представлены в табл. 4 отношением l/A . Обратное отношение A/l представляет собой индекс проводимости горла резонатора. Отношение l/A , умноженное на плотность воздуха ρ , определяет акустическую индуктивность горла; в нее частично входят и индуктивности двух прилегающих полостей. Величина ее может быть получена как сумма индуктивностей для последовательных отрезков длиной 0,5 см:

$$L = \frac{\rho}{2} \left(\frac{1}{A_a} + \frac{1}{A_b} + \dots + \frac{1}{A_q} \right), \quad (7.1)$$

что можно рассматривать как аппроксимацию интеграла

$$L = \rho \int_{x_a}^{x_b} \frac{dx}{A(x)}. \quad (7.2)$$

Если различие в площадях сужения и полостей не очень велико или если площадь изменяется плавно, возникает неопределенность в выборе границ интегрирования. В таких случаях нами было принято правило: доводить интегрирование почти до центра примыкающей полости. Такое правило основано на низкочастотном приближении входного импеданса цилиндрической полости, закрытой на другом ее конце. Добавочная масса эквивалентна $1/3$ массы, заключенной во всей трубе. Следует отметить, что величина l/A как мера огубления должна определяться по-разному для случая сдвоенного резонатора и для случая трехпараметрической модели. В последней добавочная индуктивность, соответствующая передней части ротовой полости, исключается и учитывается только масса в проходе между губами с соответственными концевыми поправками.

С целью попытаться последовательно применить теорию сдвоенного резонатора Гельмгольца функция площади для [i] и [o] будет ниже насильственно интерпретирована в свете этой теории; для этого придется в известной мере отступить от физических данных. Отступление это довольно значительно и в большой мере произвольно, так как для этих звуков все язычное сужение вместе с ротовой полостью действуют как индуктивность, которая с основной полостью резонирует на частоте F_1 . Ротовая полость и язычное сужение ведут себя при этом как отрезок трубы, в которой возникают резонансы стоячих волн, более или менее связанные с подобными же резонансами в задней полости.

Для перехода от размеров системы резонаторов к резонансным частотам с помощью формул для сдвоенного резонатора необходимо прежде всего найти резонансные частоты отдельно для переднего и заднего резонаторов. Затем истинные значения F_1 и F_2 определяются по формуле (III.10) или по номограмме рис. 102. Как было установлено ранее, две «несвязанные» резонансные частоты должны находиться между F_1 и F_2 , причем геометрическое среднее их равно геометрическому среднему из F_1 и F_2 .

Результаты расчета сведены в табл. 5, которая показывает, что теория сдвоенного резонатора явно непригодна для определения F_2 звуков [e] и [i]. Вычисленные значения F_2 этих звуков в большей мере подходят для F_3 . В то же время первая форманта этих двух гласных, так же как и две первые форманты гласных [o], [a], [u], [i], определяется с приемлемой точностью, хотя и с тенденцией к преувеличению получающихся значений частот. Эти отклонения от действительных величин зависят, конечно, от принятых условно пределов интегрирования при вычислении элементов индуктивности.

Таблица 5

Результаты расчетов по формулам для сдвоенного резонатора Гельмгольца и сравнение их с данными, полученными на электрическом аналоге LEA

Гласная	Формантные частоты, гц				
	электрический аналог			сдвоенный резонатор	
	F_1	F_2	F_3	F_1	F_2
[a]	630	1070	2400	705	1015
[o]	500	860	2320	535	905
[u]	240	610	2370	235	640
[i]	285	1480	2320	322	1470
[e]	230	2220	2970	232	3350
[e]	420	1960	2750	460	3150

Одной из возможностей охарактеризовать взаимосвязь между частотами формант и резонаторами в свете теории сдвоенного резонатора является определение параметра $R = \frac{V_1 l_1 A_2}{V_2 l_2 A_1}$; его значения приведены в табл. 4, Б.

Согласно выводам, приведенным в § 2 Приложения III, значение R , меньшее единицы, означает, что F_1 в большей мере связано с задним резонатором, а F_2 — с передним. Обратное имеет место для значений R , больших единицы. Следует помнить, что влияние отдельного резонатора складывается из влияний его горла и полости. Было показано, что при значениях R , близких к 1, горло переднего резонатора влияет на F_1 несколько больше, чем на F_2 , тогда как передняя полость более тесно связана с F_2 . Аналогично, горло заднего резонатора, т. е. соединительная секция между двумя резонаторами, влияет больше на F_2 , чем на F_1 , в то время как объем задней полости влияет на F_1 несколько больше, чем на F_2 .

В табл. 6 сведены результаты вычисления, на основе теории сдвоенного резонатора Гельмгольца, коэффициентов, характеризующих зависимость частот резонансов от отдельных элементов. Для сравнения,

Таблица 6

Коэффициенты, характеризующие связь между изменениями резонансных частот и элементов системы, полученные на основе теории сдвоенного резонатора (CP) и измеренные на электрическом аналоге (Ан). Все коэффициенты отрицательны, за исключением отмеченных звездочкой

Гласный	Ан CP	$\frac{\Delta F_1 C_1}{F_1 \Delta C_1}$	$\frac{\Delta F_2 C_1}{F_2 \Delta C_1}$	$\frac{\Delta F_1 C_2}{F_1 \Delta C_2}$	$\frac{\Delta F_2 C_2}{F_2 \Delta C_2}$	$\frac{\Delta F_1 L_1}{F_1 \Delta L_1}$	$\frac{\Delta F_2 L_1}{F_2 \Delta L_1}$	$\frac{\Delta F_1 L_2}{F_1 \Delta L_2}$	$\frac{\Delta F_2 L_2}{F_2 \Delta L_2}$	$\frac{\Delta F_1 C_1}{F_1 \Delta C_1}$	$\frac{\Delta F_2 C_2}{F_2 \Delta C_2}$	$\frac{\Delta F_1 L_1}{F_1 \Delta L_1}$	$\frac{\Delta F_2 L_2}{F_2 \Delta L_2}$
[a]	Ан	0,33	0,50	0,33	0,21	0,18	0,21	0,14	0,23	0,35	0,04	0,06	0,11
	CP	0,22	0,28	0,28	0,22	0,30	0,20	0,20	0,30	—	—	—	—
[o]	Ан	0,16	0,47	0,32	0,16	0,21	0,22	0,19	0,19	0,39	0,03	0,04	0,02
	CP	0,22	0,28	0,28	0,22	0,33	0,17	0,17	0,33	—	—	—	—
[u]	Ан	0,18	0,34	0,26	0,24	0,43	0,09	0,06	0,30	0,15	0,05	0,01	0,13
	CP	0,24	0,26	0,26	0,24	0,43	0,07	0,07	0,43	—	—	—	—
[i]	Ан	0,05	0,60	0,51	0,08	0,07	0,41	0,30	0,06	0,06	0,72	0,02	0,04
	CP	0,05	0,45	0,45	0,05	0,11	0,39	0,39	0,11	—	—	—	—
[i]	Ан	0,06	0,01	0,53	0,71	0,04	0,04	0,45	0,00	0,27	0,08	0,33	0,08*
	CP	0,00	—	0,50	—	0,03	—	0,47	—	—	—	—	—
[e]	Ан	0,09	0,07	0,55	0,55	0,14	0,06	0,30	0,03*	0,14	0,42	0,11	0,08*
	CP	—	—	—	—	—	—	—	—	—	—	—	—

приведены истинные значения, найденные при помощи электрического аналога.

Приведенные в таблице коэффициенты определяются как увеличение в процентах частоты данной форманты, обусловленное уменьшением на 1% определенного элемента резонатора. Так, например, коэффициент $\frac{\Delta F_1}{F_1} \cdot \frac{C_1}{\Delta C_1}$ показывает степень зависимости частоты первой форманты от относительного изменения объема передней полости, а коэффициент $\frac{\Delta F_1}{F_1} \cdot \frac{L_1}{\Delta L_1}$ — степень зависимости частоты той же форманты от изменения размеров горла переднего резонатора¹⁾. Согласно теории сдвоенного резонатора эти коэффициенты изменяются в пределах от 0 до 0,5. Первое предельное значение соответствует отсутствию зависимости, второе — максимальной степени зависимости данной форманты от изменения параметров данного элемента резонатора. Другие правила, относящиеся к идеальным условиям, состоят в том, что сумма коэффициентов, характеризующих влияние какого-либо элемента на две форманты, является постоянной, равной 0,5, и что коэффициент, характеризующий связь, например передней полости с F_1 , равен коэффициенту, связывающему заднюю камеру с F_2 , а коэффициент, определяющий связь F_2 с передней полостью, равен соответствующему коэффициенту, определяющему связь F_1 с задней. Между коэффициентами, определяющими влияние горла каждого из резонаторов на те же две форманты, имеют место подобные же симметричные соотношения.

Практическим способом введения относительных изменений в электрическом аналоге являлось укорочение полостей путем исключения одного звена в месте, соответствующем максимальной площади поперечного сечения. Изменения индуктивности осуществлялись сужением звена, расположенного в центре горла резонатора, т. е. у губ или у язычного сужения.

¹⁾ Здесь C_1 — емкость, эквивалентная объему полости резонатора, L_1 — индуктивность, эквивалентная массе, заключенной в его горле. (Прим. ред.)

Из табл. 6 видно достаточное соответствие расчетных и измеренных значений для [u] и [i], что позволяет считать теорию сдвоенного резонатора приемлемой для этих звуков. Однако для [o] и [a] имеется явное расхождение, проявляющееся, например, в большей, чем по расчету, измеренной величине влияния укорочения передней полости на F_2 . Это расхождение определяется тем, что в теории сдвоенного резонатора не учитываются волновые явления: максимальное значение рассматриваемого коэффициента возрастает до единицы, если учитывать эти явления в полной мере. Передняя полость в случае [a] действует скорее как простая труба, чем как резонатор Гельмгольца, и укорочение этого звена эквивалентно уменьшению как емкости, так и индуктивности.

Вследствие неправильной формы речевых полостей мало целесообразно связывать коэффициенты, характеризующие резонансы стоячих волн, с объемом этих полостей. Более показательными являются размеры полостей по их длине, если не считаться со значительной долей произвольности при установлении границ полости, имеющей плавно изменяющиеся очертания. Поэтому было признано желательным установить относительный сдвиг формантных частот без нормирования относительно размеров резонаторов.

В табл. 7 приведены данные, характеризующие эффект удаления звена единичной длины в одной из двух основных полостей. Из этих данных можно сделать следующие выводы.

Таблица 7

Относительные приращения (в процентах) формантных частот при удалении отрезка длиной 0,5 см в точках, соответствующих местам максимальной и минимальной площади поперечного сечения речевого тракта, полученные при помощи LEA

Гласный	Передняя полость					Задняя полость				
	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$
[a]	2,2	3,4	2,4	4,2	1,2	5,0	3,1	0,8	0,9	6,5
[o]	1,6	4,9	4,1	0,7	2,2	5,0	2,5	0,4	2,8	1,7
[u]	2,1	3,9	1,7	0,6	4,2	4,2	3,8	0,7	3,0	0,8
[i]	0,7	9,0	0,9	13,7	1,4	3,6	0,5	5,1	1,8	1,8
[e]	1,3	0,2	6,1	1,3	7,4	3,5	4,7	0,5	3,0	0,2
[e]	2,2	1,6	3,3	6,6	2,6	3,0	3,4	2,2	1,5	3,2

Гласный	Губы					Язычный проход					Гортанная трубка				
	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$	$\frac{\Delta F_1}{F_1}$	$\frac{\Delta F_2}{F_2}$	$\frac{\Delta F_3}{F_3}$	$\frac{\Delta F_4}{F_4}$	$\frac{\Delta F_5}{F_5}$
[a]	2,5	3,7	1,3	3,0	0,6	3,2	3,7	2,5	1,1	1,9	2,2	1,4	0,6	0,6	14,5
[o]	3,6	5,7	1,7	0,0	0,4	3,6	4,5	2,1	1,6	1,8	1,8	4,3	0,4	7,0	9,2
[u]	6,7	2,1	0,0	0,3	0,0	0,5	3,5	4,5	0,7	3,7	0,5	0,8	0,2	8,0	11,8
[i]	1,0	0,4	0,4	0,6	0,5	5,8	1,3	1,5	2,0	2,2	1,7	0,0	2,5	2,8	5,0
[e]	0,9	0,2	5,6	1,1	6,6	5,2	0,4	5,2	0,3	4,0	0,9	2,9	0,8	11,5	1,3
[e]	1,4	1,0	2,4	5,3	4,3	6,0	1,6	1,8	2,2	2,6	0,5	1,8	4,2	3,9	8,2

Первая форманта. Частота первой форманты F_1 в общем больше зависит от объема задней полости, чем от объема других полостей. Исключением является звук [а], где на F_1 в такой же мере влияет изменение объема передней полости, что и равное ему по относительной величине изменение объема задней полости. Однако, поскольку задняя полость для [а] много короче передней, процентное увеличение F_1 при укорочении задней полости удалением элемента единичной длины больше, чем изменение F_1 , вызванное удалением такого же элемента в середине передней полости.

F_1 в гласных [е], [i] и [i̯] почти полностью определяется объемом задней полости и наиболее узким участком в ротовой полости. В гласной [u] губной участок оказывает большее влияние на F_1 , чем язычное сужение. Влияние задней полости на F_1 для [u] несколько больше, чем передней.

Вторая форманта. Лишь для гласного [i̯] ротовая полость с примыкающими к ней проходами является основным элементом, определяющим F_2 . В случае [i] F_2 явно соответствует полуволновому резонансу задней полости. Для второй форманты [е] имеет место подобная же, но менее выраженная тенденция к большей зависимости от задней, чем от передней полости. Вторая форманта задних гласных [u], [o] и [а] несколько больше зависит от передней полости, чем от задней. Если изменять объем полостей одинаково в процентном отношении к исходному их объему, эта зависимость выявляется весьма отчетливо. Однако, если изменять объем путем уменьшения длины на одну и ту же абсолютную величину, зависимость F_2 для [u] и [а] от той и другой полостей становится одинаковой. В случае [u] F_2 значительно сильнее зависит от относительных размеров язычного прохода, чем от размеров губного участка. Эти две части всей системы оказывают примерно одинаковое влияние на F_2 как для [а], так и для [o]. Губной участок практически совсем не оказывает влияния на F_2 для звука [i̯] и только в небольшой степени влияет на F_2 для [е].

Третья форманта. Согласно приведенным в таблице данным, F_3 для [u], [o] и [а] зависит главным образом от участка, расположенного перед язычным сужением. Это справедливо также и для [i̯]. В случае [i̯] основным определяющим фактором для F_3 является задняя полость. Для [е] влияние на эту форманту более равномерно распределено между всеми участками речевого тракта. Исходя из физических граничных условий для отдельных полостей, можно считать, что частота F_3 для [i̯] в основном определяется четвертьволновым резонансом по обоим сторонам язычного сужения, но можно интерпретировать ее и как полуволновый резонанс язычного сужения и передней полости. F_3 для [i̯] есть не что иное, как полуволновый резонанс задней полости, а F_3 звука [u] представляет собой полуволновый резонанс передней полости. Для [а] и [o] F_3 связано с 3/4-волновым резонансом в системе полостей, расположенных перед язычным сужением.

Четвертая и пятая форманты. Для звуков [o], [u] и [i̯] имеет место отчетливое влияние полости гортани на F_4 ; система полостей, расположенных перед язычным проходом, оказывает заметное влияние на F_4 в случае звуков [а], [i̯] и [е], в то время как задние полости оказывают большее влияние на эту форманту в случае [u], [o] и [i̯]. В значительной мере F_5 зависит также от гортанной трубки, кроме звука [i̯], для которого определяющей является передняя полость. Тесная связь F_5 с задней полостью существует для [а] и с передней полостью — для [u].

Интересно исследовать влияние изменений объема, вызванных увеличением площади поперечного сечения в одном звене системы в тех или иных местах модели. Такие измерения были произведены в местах максимальной площади каждой из двух основных полостей (табл. 8).

Таблица 8

Процентное уменьшение формантных частот при увеличении на 1% объема передней или задней полости путем увеличения площади единичного звена, расположенного в месте максимальной площади. Отрицательные значения отмечены звездочкой

Гласный	$\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_2}{F_2} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_1}{F_1} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_2}{F_2} \frac{C_2}{\Delta C_2}$	$\frac{\Delta F_1}{F_1} \frac{C_1}{\Delta C_1}$	$\frac{\Delta F_3}{F_3} \frac{C_2}{\Delta C_2}$
[a]	0,07	0,19	0,23	0,11	0,18	0
[o]	0,05	0,33	0,37	0,22	0,25	0
[u]	0,08	0,28	0,20	0,15	0,06	0
[i]	0,02*	0,39	0,49	0,02*	0,04	0,45
[i]	0,01*	0	0,53	0,39*	0,04	0
[e]	0,08*	0,01*	0,42	0,26*	0,01	0,23

Эти данные должны были бы совпадать с приведенными в табл. 6, если бы вызванное изменением площади увеличение объема давало тот же эффект, что и вызванное изменением длины. Это, как можно видеть, не имеет места в силу конечной скорости распространения звука. По сравнению с табл. 6 местное увеличение площади поперечного сечения задней полости для [u] оказывает меньшее влияние на F_1 . Относительно мало влияние и местных изменений площади передней полости на F_1 для звуков [o] и [a].

Значительное изменение F_1 для [a] при укорочении передней полости (табл. 7) обусловлено тем, что удаленное для укорочения звено играло по преимуществу роль индуктивности, а не емкости. Полезно сравнить этот эффект с характером изменений резонансной частоты отрезка трубы, закрытой на одном конце. Небольшие изменения площади в середине такой трубы не влияют на резонансную частоту. Увеличение площади сечения вблизи закрытого конца вызывает максимальное понижение резонансной частоты, а такое же изменение вблизи открытого конца приводит к максимальному ее повышению (рис. 19). Можно также показать, что местное удвоение площади поперечного сечения на участке несколько впереди от середины трубы не влияет на F_1 .

Приведенные выше исследования относительных изменений позволяют сделать общий вывод о том, что хотя теория сдвоенного резонатора Гельмгольца позволяет достаточно удовлетворительно предсказать значения F_1 и F_2 в гласных [u], [o], [a] и [i] и значения F_1 в [i] и [e], представление полостей в виде сосредоточенных емкостных элементов ведет к недооценке связи F_2 с передней полостью и, до известной степени, связи F_1 с задней полостью. Однако в звуках [u], [o] и [a] существенно влияние на F_1 и F_2 обеих полостей, а в звуке [e] влияние этих полостей на F_2 и F_3 даже почти одинаково. Все проведенные здесь исследования относительных изменений указывают на отчетливую зависимость F_2 для [i] от задней полости и на связь F_3 этого гласного с передней полостью; для [i] обнаруживается отчетливая зависимость F_3

от задней полости. Степень открытия губ особенно сильно влияет на F_1 для [u], F_2 для [i] и F_3 для [i] и оказывает приблизительно одинаковое влияние на F_1 и F_2 в звуках [o] и [a]. Основное язычное сужение, т. е. горло задней полости, влияет на F_1 передних гласных [i], [i] и [e] в большей мере, чем на другие их форманты. В случаях [u] и [a] степень сужения язычного прохода более критична для F_2 , чем для F_1 , а в случае [o] она оказывает одинаковое влияние как на F_1 , так и на F_2 .

Теория сдвоенного резонатора должна использоваться с большой осторожностью. Следует заметить, что ни эта, ни какая-либо другая теория не может дать формул расчета частоты первых двух формант по известным размерам резонатора, удовлетворяющих одновременно требованиям простоты и универсальности применения. Что касается теории сдвоенного резонатора, то основным препятствием для ее использования является ошибочность даваемых ею значений F_2 для высоких передних гласных.

Некоторых ограничений в использовании теории двойного резонатора можно избежать, если придать понятию «форманта» несколько иное содержание, рассматривая ее с точки зрения критериев восприятия. Как известно из опытов по синтезированию звуков речи [32, 34], почти все гласные могут быть успешно синтезированы на основе двух формант. В задних гласных и других звуках, где F_2 ближе к F_1 , чем к F_3 , синтез можно осуществить, используя только F_1 и F_2 , со значениями, соответствующими естественным звукам. С другой стороны, в типичных передних гласных, где F_2 ближе к F_3 , верхняя форманта синтезированных звуков должна быть расположена ближе к F_3 , а в случае очень «высокого» [i] — даже выше F_3 . Учитывая полученные выше результаты анализа, можно сказать, что ротовая полость является основным фактором, определяющим F_2 или «среднее по восприятию» из F_2 и всех более высоких формант (ср., например, связь F_3 в звуке [i] с передней полостью [53]).

Эти факты, подтверждающие теорию, согласно которой два резонанса являются основным показателем окраски гласных, не могут, конечно, полностью реабилитировать представление о том, что низкая форманта определяется только задним резонатором, а высокая — только передним. Низкая форманта всегда зависит от всего речевого тракта в целом; это верно и для второй форманты задних гласных. В этой связи следует отметить, что гласные, где F_2 очень близка к F_1 , могут быть синтезированы с помощью одной форманты, средней с точки зрения восприятия и лежащей приблизительно посередине между F_1 и F_2 для [a], но ближе к F_1 для [u]. Частота этой одной форманты близка к собственной частоте несвязанного резонанса переднего резонатора.

Если принять, что все артикуляторные характеристики звука сводятся к четырем элементам, рассматриваемым в теории сдвоенного резонатора Гельмгольца (которая, как это было показано выше, далека от совершенства), то удобно использовать следующие приближенные формулы для вычисления обеих резонансных частот такой системы:

$$\left. \begin{aligned} F_{\text{низ}} &= \frac{c}{2\pi} \left[V_2 \left(\frac{l_1}{A_1} + \frac{l_2}{A_2} \right) \right]^{-1/2}, \\ F_{\text{выс}} &= \frac{c}{2\pi} \left[\left(\frac{A_1}{l_1} + \frac{A_2}{l_2} \right) \frac{1}{V_1} \right]^{1/2}. \end{aligned} \right\} \quad (7.3)$$

Эти формулы дают достаточно хорошее приближение при условии

$$\frac{V_1}{V_2} \ll 1 + \frac{l_2 A_1}{l_1 A_2}, \quad (7.4)$$

которое, однако, не соблюдается в случае задних гласных. Поэтому для задних гласных получается на 10—20% более низкое значение $F_{\text{выс.}}$. В общем все же формулы (7.3) и (7.4) могут быть использованы для оценки относительных изменений, если помнить при этом, что в задних гласных как передний, так и задний объемы оказывают существенное влияние на оба резонанса. При этом имеют место следующие взаимозависимости, которые были использованы Якобсоном и др. [114].

Частота первой форманты возрастает с уменьшением объема задней полости и с увеличением индекса проводимости A/l любого суженного прохода. Частота верхней форманты возрастает с уменьшением объема передней полости и с увеличением индекса проводимости первого или второго горла, связанного с этой полостью, т. е. при делабиализации звука или при расширении язычного прохода, являющегося горлом заднего резонатора. Если проводимость одного горла не равна проводимости другого, горло, характеризующее большим значением индекса проводимости, будет определять в основном верхнюю форманту, а положение низкой форманты будет зависеть от горла с меньшей проводимостью, т. е. с большей величиной l/A . Если проводимость одного горла резонатора равна проводимости другого, то низкая форманта будет больше зависеть от горла, расположенного дальше от голосовых связок, т. е. от губного участка.

Однако два объема и два горла могут лишь в редких случаях изменяться независимо. Так, при артикуляции передних гласных понижение языка вызовет не только увеличение проводимости внутреннего горла резонатора, но и увеличение объема передней полости, причем это увеличение может быть еще усилено большей степенью растройства челюстей. Оба эти сопутствующие друг другу, но действующие в противоположных направлениях изменения элементов резонатора могут привести к их взаимной частичной компенсации, обуславливая тем самым тенденцию к сохранению постоянства верхней форманты. При средненёбной точке артикуляции в результате соответственных артикуляторных движений F_2 уменьшается, а F_1 возрастает. В задних гласных, типа [а], сужение прохода между спинкой языка и стенкой фаринкса обычно сопровождается и уменьшением объема задней полости. В результате этого происходит частичная компенсация сдвига нижней форманты. Однако верхняя форманта будет при этом уменьшаться по частоте, так как объем передней полости остается приблизительно постоянным.

Все ограничения и трудности, связанные с применением аналитических формул, полученных на основе теории сдвоенного резонатора, исчезают, если влияние артикуляции на F -картину рассматривается на основе модели, учитывающей корреляцию между размерами элементов системы; такого типа модель описана Стивенсом и Хаузом [194, 195] и дополнительно развита в § 3 гл. 4 настоящей работы. При использовании подобной модели необходимо задать место артикуляции, степень открытия прохода в этом месте, а также длину и площадь поперечного сечения губного участка. Номограммы, связывающие эти данные с резонансными частотами речевого тракта, должны заменить или по крайней мере дополнить существующие в фонетических руководствах аналитические формулы резонаторов.

Мы установили, что частотное положение первых трех формант проанализированных здесь шести русских гласных может быть найдено с точностью до 5—10% с помощью данных о размерах речевого тракта, приведенных в табл. 4, Б. Эти данные использованы в номограммах рис. 27—29.

§ 4. Пространственное распределение звукового давления. Ширина полосы формант

А. Расчеты уровней формант

Следующие разделы посвящены изучению распределения звукового давления в полостях речевого тракта и в излучаемой речевой волне, а также влияния различных диссипативных элементов на демпфирование формант гласных, т. е. на ширину их полосы.

Конфигурации речевого тракта для шести русских гласных, проанализированных в настоящей работе, были использованы для численных расчетов и для расчетов с помощью аналога в соответствии с методами, рассмотренными в § 2 гл. 4 и в гл. 6.

Электрический аналог LEA оказался особенно полезным для этой цели, так как распределение напряжения вдоль звеньев аналога отражает распределение давлений внутри речевого тракта. Этот аналог позволяет также легко изучать влияние добавления или удаления определенных диссипативных элементов в том или ином месте модели. Метод численных расчетов, однако, более удобен в том случае, когда необходимо учитывать распределенные диссипативные элементы в соответствии с известным выражением для показателя затухания речевого тракта. Эти вычисления легко выполнимы, если располагать цифровой вычислительной машиной. В § 1 гл. 7 было показано, что формантные частоты, вычисленные с помощью машины BESK, вполне сравнимы с результатами, полученными на аналоге LEA, вплоть до третьей форманты. Таким образом, эти два метода до известной степени взаимозаменяемы.

В табл. 9 приведены данные о ширине полосы формант и уровнях звукового давления формант, полученные путем численных расчетов и использованные при построении приведенных ранее на рис. 41 огибающих спектра. Диссипативными элементами, использованными в этих расчетах, являлись сопротивление излучения и распределенные потери на различных участках резонатора. Эти потери определяются коэффициентом затухания

$$\alpha = 0,007 \left(\frac{\pi}{A} \right)^{1/2} \cdot \left(\frac{1}{2} + \frac{f}{4000} \right). \quad (7.5)$$

Эта формула построена на основании опытных данных так, чтобы обеспечить ширину полосы формант в тех же пределах, что и в естественных звуках¹⁾. Зависимость от площади поперечного сечения здесь та же, что и в случае классического резонатора с потерями на стенках, но частотная зависимость принята линейной вместо степенной зависимости с показателем 1/2, получаемой из теории систем с жесткими стенками. Постоянный и пропорциональный частоте члены в приведенной

¹⁾ Формула (7.5) введена здесь лишь для расчета уровней формант. Стивенс [192] для определения ширины полосы формант успешно использовал данные об импедансе тканей человеческого тела (см. Приложение III, § 6, Б).

Таблица 9

Ширина полосы формант в гц и уровень звукового давления формант в дб относительно 0,0002 дин/см². Расчетные данные

Гласные	Ширина полосы формант, гц					Уровни формант								
						на расстоянии 5 см перед губами			у губ			у голосовой щели		
	B_1	B_2	B_3	B_4	B_5	L_1	L_2	L_3	L_1	L_2	L_3	L_1	L_2	L_3
[а]	57	72	130	175	200	91	85	70	111	104	87	135	121	97
[о]	54	65	100	135	155	90	86	66	112	107	86	134	121	95
[у]	69	50	110	115	110	84	73	37	114	101	68	137	121	86
[и]	43	125	77	134	140	88	67	69	108	82	85	136	86	104
[і]	60	75	240	230	330	84	68	66	106	86	83	136	108	88
[е]	39	95	170	325	310	90	77	70	108	93	92	132	108	105

линейной зависимости становятся равны друг другу при частоте 2000 гц. Принятая зависимость является практически компромиссом между требованиями учета диссипативных элементов в речевом тракте, зависящих от частоты с отрицательным показателем степени и обусловленных потерями на колебания стенок, и учета нелинейного затухания в сужениях.

Голосовым источником при расчетах является источник постоянной объемной скорости, равной 150 см³/сек; частотные свойства источника определялись двойным полюсом при частоте 100 гц на отрицательной вещественной оси частот. Таким образом, огибающая спектра источника выражалась формулой

$$|U_q(f)| = \frac{150}{1 + f^2/100^2} \text{ см}^3/\text{сек}, \quad (7.6)$$

со скоростью спадания 12 дб на октаву на частотах выше граничной частоты 100 гц. Такой источник был принят как стандарт во всех расчетах сонорных звуков, приведенных в настоящей работе¹⁾.

Ширина полосы формант была численно определена с помощью BESK на основании выражения (2.39). Относительные уровни различных формант излучаемого звука также вычислялись на BESK с учетом множителя $K_T(\omega)$ (рис. 8), характеризующего возрастание давления быстрее, чем прямо пропорционально ω .

Отношение звукового давления у губ к звуковому давлению в точке на расстоянии 5 см впереди от губ было найдено путем использования формулы (2.55), связывающей звуковое давление в излучаемой волне с объемной скоростью в ротовом отверстии. Это отношение равно

$$\frac{P_{\text{губ}}}{P_l} = \frac{0,8 (A/\pi)^{1/2} 4\pi l}{AK_T(\omega)}. \quad (7.7)$$

Давление у губ определяется по объемной скорости и индуктивной составляющей импеданса излучения, равной $L_0 = 0,8 \left(\frac{A}{\pi} \right)^{1/2} \frac{\rho}{A}$.

Функция передачи, связывающая звуковое давление у губ со звуковым давлением у входа речевого тракта, определялась как отношение соответственных напряжений на LEA.

¹⁾ См. также формулу (3.2), где принято во внимание и сопротивление излучения.

Наконец, абсолютные уровни звукового давления определялись по объемной скорости источника звука на входе системы, т. е. в голосовой щели, при частоте первой форманты $[a]$, определенной по уравнению (7.6). Величина объемной скорости, умноженная на соответственный входной импеданс, измеренный на LEA, и дает искомое значение звукового давления. Значение входного импеданса должно быть исправлено в соответствии с разницей в ширине формант, полученной в результате вычислений на LEA и BESK.

Вычисленные уровни давления достаточно показательны, так как выбранная для расчетов функция, характеризующая источник, дала в результате нормальные значения уровней формант соответственных звуков. Абсолютные значения всех уровней давления получены, исходя из величины постоянной составляющей потока, равной $150 \text{ см}^3/\text{сек}$. Это значение близко к среднему значению расхода воздуха при фонации [24].

Единственные экспериментальные данные, с которыми можно сравнить приводимые здесь расчеты, относятся к шведским гласным [45]. В табл. 10 данные об уровнях формант для естественных шведских

Таблица 10

Гласный		L_1/F_1	L_2/F_2	L_3/F_3
[a]	R	91/616	85/1070	70/2375
	S	91/680	86/1070	64/2520
[e]	R	90/432	77/1960	70/2720
	S	91/440	76/1795	73/2385
[i]	R	84/222	68/2240	66/3140
	S	88/255	64/2065	67/2960
[u]	R	84/231	73/615	37/2375
	S	89/310	76/730	47/2230

звуков обозначены символом S, а вычисленные для русских — символом R. Приведенные величины относятся к звуковому давлению на расстоянии 5 см от губ.

Часть различий в уровнях формант можно объяснить различием частот формант. Так, бо́льшая на 10 дБ величина уровня L_3 в естественном шведском [u] частично может быть обусловлена остаточным множителем $40 \log_{10} \left(\frac{310}{231} \cdot \frac{730}{615} \right) = 8 \text{ дБ}$, который вытекает из уравнения (3.9).

Измеренные уровни L_1 для «открытых» и «закрытых» гласных, как видно из таблицы, отличаются относительно немного, хотя эти звуки разнятся соответственно более высоким и более низким положением F_1 . Это можно объяснить тем, что средний спектр диктора падает быстрее, чем функция, принятая в расчетах, по крайней мере в области частот от 200 до 1000 гц. Имеются также данные, указывающие на то, что скорость спадания спектра голосового источника человека на частотах выше 2000 гц больше, чем 12 дБ на октаву.

Следует, кроме того, учитывать, что амплитуда звукового давления форманты обратно пропорциональна ширине ее полосы, так что удвоение ширины уменьшает уровень форманты на 6 дБ.

Достоверных сведений о ширине полосы каждой из формант в естественной речи не существует, но можно, по-видимому, считать, что данные о ширине полосы формант, полученные расчетным путем, отклоняются в общем от типичных естественных данных не более чем на 50%. Систематическим фактором, имеющим второстепенное значение, является различие между уровнем интенсивности формант и уровнем форманты, определенным по огибающей спектра, который и был предметом расчета. При условии, что ширина форманты меньше, чем частота основного тона F_0 , уровень интенсивности ее не более чем на 1 дБ превышает уровень, найденный по огибающей.

Рассматривая уровни формант и спектры голосового источника, необходимо иметь в виду, что средний наклон спектра изменяется при форсировании речи; в случае большого голосового усилия уровень основной составляющей голоса относительно мал, а уровни высокочастотных формант велики по сравнению с первой формантой, которая в нормальных условиях в основном определяет интенсивность сонорных звуков (см. § 1 Приложения II).

Б. Распределение звукового давления в речевом тракте

Пространственное распределение звукового давления в речевом тракте от губ до голосовой щели показано на рис. 43, на котором приведены отдельные кривые для первых трех формант каждой из шести гласных. Эти кривые дают абсолютные величины звукового давления в линейной шкале без учета фаз; на оси ординат отмечены уровни, выраженные в дБ.

На частоте первой форманты звуковое давление имеет минимальное значение у губ и возрастает до максимума у голосовой щели. В случае такого типичного переднего гласного, как [i], где речевой тракт для первой форманты ведет себя как простой резонатор, звуковое давление возрастает почти строго линейно в ротовой части, представляющей собою горло резонатора, и остается постоянным в пределах фарингальной полости. Такое распределение как раз соответствует представлению горла в виде индуктивности, а основной полости — в виде емкости в эквивалентной схеме. Степень повышения давления вдоль горла будет, очевидно, определяться отношением индуктивного сопротивления горла к индуктивному сопротивлению излучения, т. е. в случае горла постоянного поперечного сечения — отношением его физической длины к концевой поправке.

С другой стороны, кривая распределения L_1 для [u] характеризуется резким подъемом давления позади губ, причем звуковое давление в передней полости лишь немного меньше, чем в задней. Это объясняется тем, что в этом случае доминирующую роль играет индуктивное сопротивление губного прохода.

На частоте второй форманты кривые давления должны иметь дополнительный минимум, а на частоте третьей форманты — два минимума между передним и задним концами системы. Минимум L_2 обычно имеет место в верхней половине фарингальной полости и мало меняется по своему местоположению. Отсюда может быть сделан вывод о том, что ларингофон не следует помещать в слишком высокой части горла, а предпочтительней располагать его на уровне гортанной полости [8]. В ряде случаев минимумы звукового давления сильно задемпфированы; это справедливо для третьей форманты [i] и второй форманты [i].

В резонаторе, имеющем вид отрезка трубы длиной l , отношение величины давления $P(x, \omega)$ в точке, расположенной в x см от

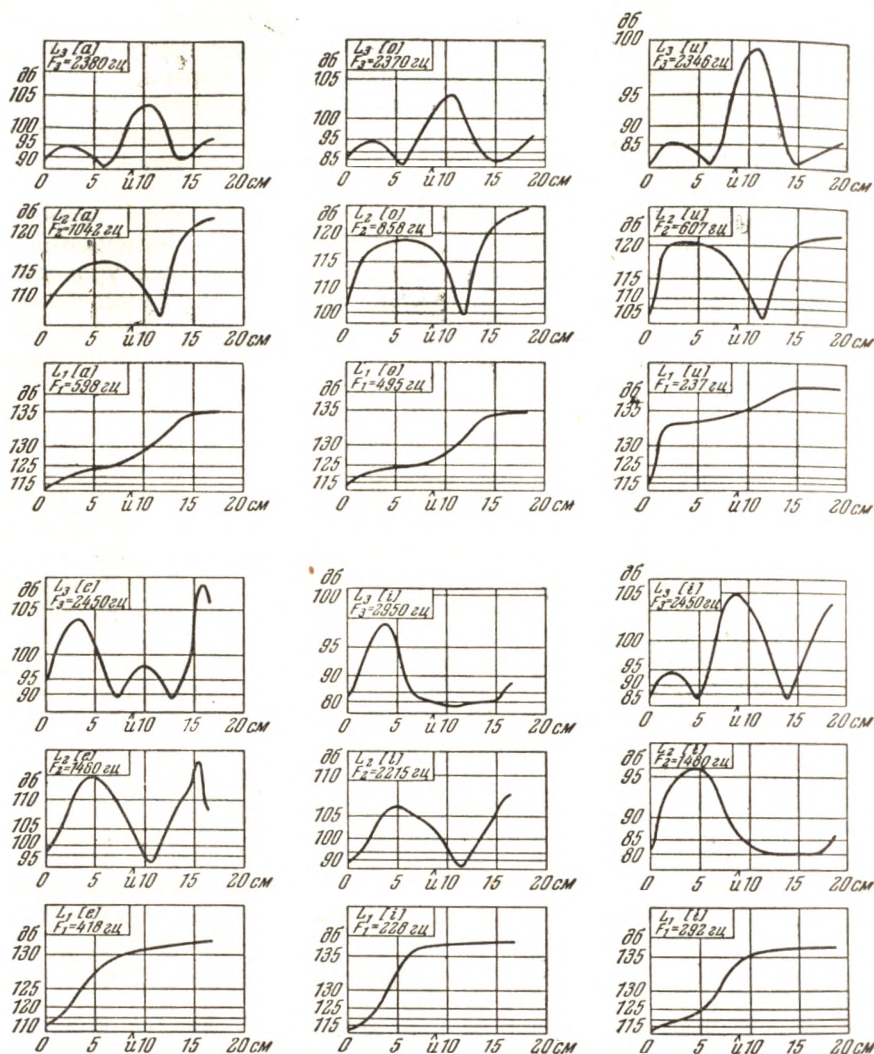


Рис. 43. Распределение звукового давления вдоль речевого тракта, полученное расчетом для каждой из первых трех формант русских гласных. Соответственные функции площади приведены на рис. 39. Начало координат — у губ.

переднего, открытого конца, к объемной скорости $U_q(\omega)$ источника, расположенного на противоположном конце трубы, равно

$$\frac{P(x, \omega)}{U_q(\omega)} = Z \frac{\operatorname{sh}\left(j \frac{\omega}{c} + \alpha\right) x}{\operatorname{sh}\left(j \frac{\omega}{c} + \alpha\right) l}. \quad (7.8)$$

Аналогично коэффициент передачи объемных скоростей равен

$$\frac{U(x, \omega)}{U_q(\omega)} = \frac{\operatorname{ch} \left(j \frac{\omega}{c} + \alpha \right) x}{\operatorname{ch} \left(j \frac{\omega}{c} + \alpha \right) l}. \quad (7.9)$$

Резонансными являются те частоты, при которых $\cos \frac{\omega l}{c} = 0$. При частоте второй форманты, определяющейся условием $\frac{\omega_2 l}{c} = \frac{3\pi}{2}$, давление будет минимальным, когда $\sin \frac{\omega_2 l}{c} = 0$, т. е. при $x = \frac{2}{3} l$, что и характерно для диаграмм изменения F_2 , приведенных на рис. 43. Отношение максимального давления к минимальному приблизительно равно $1/\alpha x$, причем максимум имеет место при $x = \frac{l}{3}$. Объемная скорость сдвинута по отношению к давлению на угол 90° во времени и пространстве, причем максимальные величины объемной скорости совпадают с минимальными величинами давления и наоборот, как это следует из синусоидального и косинусоидального вида соответственных функций.

Эти соотношения хорошо известны из элементарной теории стоячих волн в трубах и линиях передач. Общим является положение, согласно которому, независимо от формы полостей речевого тракта, нули давления находятся там, где реактивное сопротивление, рассматриваемое по направлению к губам, равно нулю. Если в то же время имеют место условия резонанса, реактивное сопротивление, рассматриваемое по направлению к голосовой щели, должно быть также равно нулю. При наличии сильного демпфирования единственным видимым признаком пространственного положения нуля является фаза давления, которая близка к нечетному кратному 90° и быстрее всего изменяется в области узлов давления. Кривые давления и объемной скорости, полученные Чибя и Каджияма [24], относятся к случаю отсутствия потерь, но в общих чертах подтверждают наши данные.

Расположенный ближе к переднему концу максимум давления для частоты второй форманты находится в небной полости. Это было предсказано Ван-ден-Бергом [11], который высказал предположение, что вибрации мягкого неба с частотой второй форманты вызывают ощущение, которое играет в известной мере роль сигнала обратной связи, позволяющего певцу придавать то или иное качество своему голосу. Однако по величинам давления, приведенным на рис. 43, можно с большей вероятностью считать, что действительно эффективным стимулом скорее является первая форманта, которая в этой области имеет уровень на 15 дБ более высокий и лежит в таком диапазоне частот, где тактильная чувствительность наибольшая [177]. Абсолютный уровень L_1 у мягкого неба достигает 136 дБ для [i] при среднем голосовом усилии и по крайней мере на 10 дБ больше при пении.

Известный интерес представляет вопрос о том, в какой мере наши кривые распределения давления соответствуют результатам измерения. Результаты измерения уровня звукового давления внутри ротовой полости и вне ее, выполненного Бекеши [5] с помощью зонда, и уровни звукового давления для первой форманты, усредненные по гласным [o], [a] и [e] по нашим расчетным данным, приведены на рис. 44. Все данные приведены к одному уровню по величинам звукового давления у края

губ. Видно, что расхождение уровней для остальных точек не превышает 2 дб.

В связи с тем, что в фарингальной полости уровни звукового давления велики, может возникнуть вопрос о том, в какой степени возможна передача первой форманты непосредственно через стенки полостей речевого тракта. Простую оценку такой возможности можно получить следующим образом. Допустим, что колебания стенок полости управляются только массой, которую можно считать распределенной по площади 50 см² при толщине 1 см и плотности 1 г/см³. Индуктивное сопротивление,

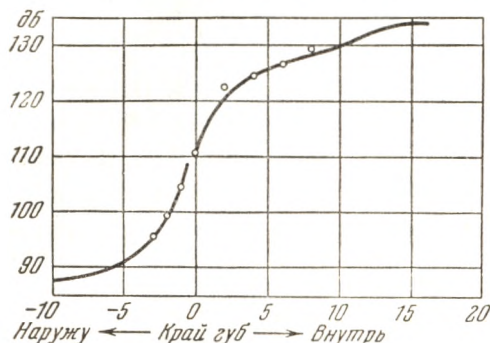


Рис. 44. Расчетные уровни звукового давления для первой форманты внутри и вне рта, усредненные по гласным [о], [а], [е], и сравнение их с данными измерения суммарного уровня звукового давления, выполненного Бекеша [5]. За исходный уровень для последних данных принят такой уровень, при котором в обоих случаях обеспечивается одинаковое звуковое давление у губ. Пространственная координата — расстояние от губ (в см).

соответствующее этой массе, равно $20 \cdot 10^{-3}$ акустических ом и только в четыре раза больше индуктивности узкого прохода в ротовой полости, обеспечивающей резонанс с фарингальной полостью объемом 80 см³ на частоте 300 гц; это соответствует, например, F_1 для гласного [i]. Таким образом, шунтирующее действие колеблющихся стенок полости вызовет повышение частоты F_1 в $(1 + 0,25)^{1/2}$ раз, т. е. приблизительно на 10%. Эта поправка является достаточно большой и ее следует принимать в расчет при вычислениях. Объемная скорость потока, создаваемого стенками полости, будет также равна $1/4$ объемной скорости в

ротовом отверстии. Излучение через стенки горла создаст, таким образом, звуковое давление на 12 дб меньше, чем нормальное излучение через рот на достаточном расстоянии от диктора.

Это, конечно, только грубая оценка, однако она находит некоторое подтверждение в данных измерений звукового давления во время звонкой смычки, предшествующей взрывным звукам [b], [d], [g], причем в этих условиях звук передается только через стенки¹⁾.

Принятое выше значение индуктивного сопротивления стенки соответствовало бы частоте первой форманты порядка 150 гц, что является наименьшим возможным значением F_1 при любых условиях. Измерения, выполненные во время звонкой смычки, действительно дают величину F_1 такого порядка, и уровень на 12 дб ниже уровня L_1 для [i].

Пространственное распределение звукового давления нельзя использовать непосредственно для выводов о зависимости формант от полостей речевого тракта. Необходимо сначала нормировать это распределение относительно площадей поперечного сечения; кроме того, немаловажную роль играет объемная скорость. Реактивная энергия за один период в сечении речевого тракта с площадью $A(x)$ является суммой потенциальной энергии, пропорциональной $A(x) \cdot P^2(x)$, и кине-

¹⁾ Следует также иметь в виду наличие известного низкочастотного излучения стенками грудной клетки, которое рассматривал Ван-ден-Берг и др. [15]. Этот звук возникает за счет проходящих в подгортанную полость звуковых колебаний, образующихся в воздухе. Измерение амплитуды колебаний в разных частях человеческого тела было выполнено Бекеша [5].

тической энергии, пропорциональной $U^2(x)/A(x)$. Сумма этих двух величин, как функция координаты места этого сечения, могла бы явиться подходящей мерой зависимости данной форманты от различных частей системы. Представляющие непосредственный интерес для теории речеобразования выводы из приведенного рассмотрения состоят в том, что эффект шунтирования речевого тракта какой-либо дополнительной ветвью имеет наибольшее значение в тех местах, где звуковое давление наивысшее, и что влияние дополнительных последовательных элементов наибольшее в тех местах, где велика объемная скорость. Активная часть импеданса, введенного дополнительно в систему, увеличивает демпфирование форманты, а реактивная часть приводит к частотному сдвигу форманты.

В. Зависимость ширины форманты от различных активных элементов в речевом тракте

Демпфирующее влияние различных активных элементов в модели речевого тракта будет прежде всего показано на примере гласного [i]. Для расчетов, результаты которых приведены на рис. 45, была принята показанная на рис. 36 идеализированная конфигурация, а именно рупообразная форма ротовой полости и постоянная величина площади поперечного сечения фаринкса. Кривые 1—3 на этих рисунках соответствуют нормальным условиям демпфирования в LEA, т. е. нормальному демпфирующему действию излучения, сопротивлению голосовой щели, равному 5рс , и минимальным возможным потерям в катушках и конденсаторах аналога. Кривая 1 относится к фаринксу без учета гортанной трубки. При получении кривой 2 учтена гортанная полость, а кривой 3 — грушевидные пазухи, шунтирующие выход гортанной полости.

Шунтирующая ветвь была удалена в экспериментах, результаты которых отражены на кривых 4—8. При получении кривой 4 сопротивление излучения было уменьшено вдвое. Это дает заметное увеличение уровня и соответствующее уменьшение ширины полосы третьей и более высоких формант. Кривая 5 была получена при том же сопротивлении излучения, что и кривая 4, но при площади поперечного сечения язычного прохода, уменьшенной настолько, чтобы F_1 сдвинулась вниз до 180 гц , т. е. до положения, более соответствующего согласному [j].

Понижение F_1 привело к обычному эффекту понижения уровня более высоких формант. Вследствие того, что F_3 возросло по частоте и приблизилось к F_4 , как это видно из кривой 5, происходит общее относительное увеличение уровня этих двух формант, из которых F_3 соответствует основному резонансу полости рта, а F_4 — основному резонансу полости гортани. При нормальной величине сопротивления излучения эти два резонанса сливались бы в общий пик, что часто встречается в случае [j], где третья форманта включает и F_3 , и F_4 . Условия для кривой 6 отличаются от условий для кривой 2 лишь тем, что сопротивление голосовой щели было уменьшено до $2,5\text{рс}$, т. е. его шунтирующее действие усилено в 2 раза.

Как видно из кривых распределения давления (рис. 43), давление в том конце речевого тракта, где расположена голосовая щель, весьма велико для всех формант всех гласных. Поэтому сопротивление голосовой щели влияет на все форманты, как это и видно из кривой 6.

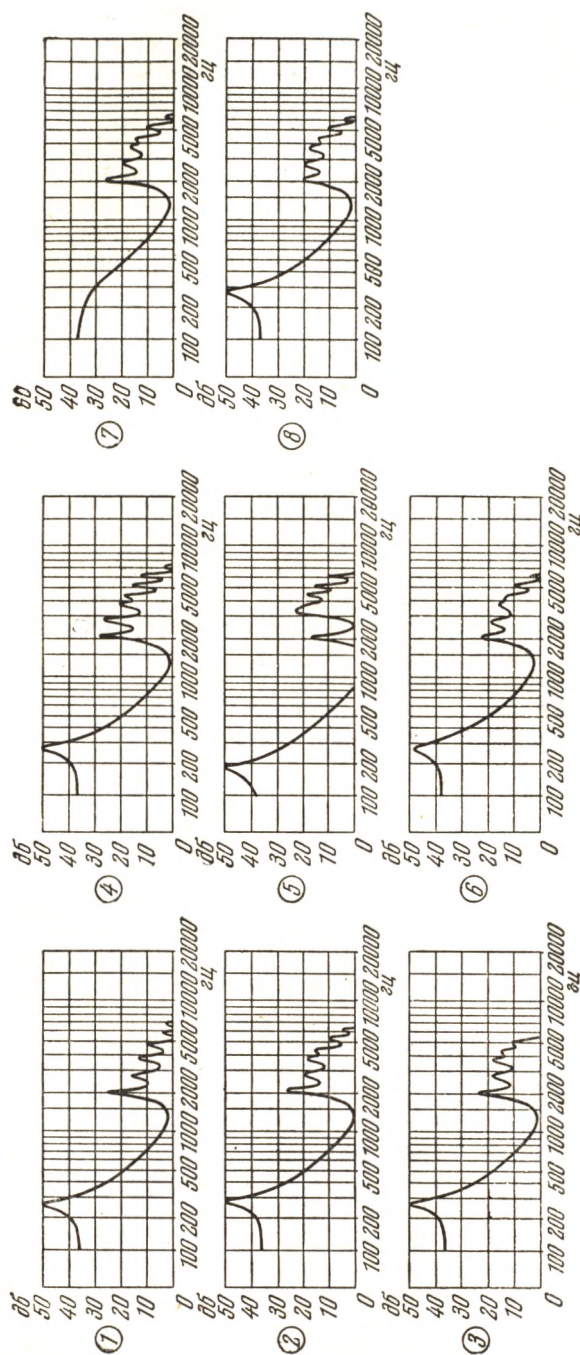


Рис. 45. Влияние изменения факторов, определяющих затухание формант в модели гласного [i] с рупорообразным элементом по рис. 36.
1) — 3) То же, что на рис. 36.

4) Конфигурация гортани такая же, как в (2), но сопротивление излучения уменьшено вдвое против нормальной его величины.

5) То же, что (4), но площадь сужения, образующего язычный проход, уменьшена вдвое. Виден двойной пик $F_3 + F_4$.

6) То же, что (2), но сопротивление голосовой щели $2,5 \text{ рс}$ вместо 5 рс .

7) То же, что (2), но последовательно с язычным проходом включено сопротивление $0,25 \text{ рс}$.

8) То же, что (2), но сопротивление включено последовательно в середине фаринкса.

Кривая 7 иллюстрирует влияние включенного последовательно с язычным проходом сопротивления в 0,25рс. Эта кривая отличается от кривой 2 главным образом большим демпфированием первой форманты. Этого и следовало ожидать ввиду большой объемной скорости в горле резонатора для F_1 . Почти стопроцентная обусловленность второй форманты полуволновым резонансом задней полости проявляется в виде отсутствия сколько-нибудь заметного демпфирования этой форманты. Как вторая, так и третья форманты характеризуются максимумом давления в области язычно-нёбного сужения и потому малой объемной скоростью в сужении. В то же время последовательное сопротивление, расположенное в середине фарингальной полости, оказывает существенное влияние на уровень и ширину полосы второй форманты, как это видно из кривой 8.

Было предпринято исследование влияния сопротивления излучения, сопротивления голосовой щели, а также потерь на вязкость и теплопроводность, на относительное изменение ширины полосы каждой из пяти первых формант в гласных. Влияние излучения и сопротивления голосовой щели определялось на LEA, а демпфирующее влияние стенок полостей определялось путем ряда численных расчетов, выполнявшихся с помощью вычислительной машины BESK.

Полученные в результате измерений на аналоге LEA данные (табл. 11), показывающие влияние излучения и голосовой щели на ширину трех первых формант, хорошо согласуются с расчетами на BESK, но для четвертой и пятой формант были основания считать, что данные, полученные с помощью LEA, более точны.

Таблица 11

Доля, вносимая в ширину формант (гц): А) сопротивлением излучения; Б) параллельным сопротивлением голосовой щели в 5 рс; В) потерями на трение и теплопроводность, в предположении идеально твердых стенок; соответственный коэффициент затухания $\alpha = 2,92 (\pi/A)^{1/2} \cdot 10^{-5} \cdot f^{1/2} S_A$ непер/см, где $S_A = 2$ — коэффициент, учитывающий форму и характер поверхности стенок

Гласные	А					Б					В			А + Б + В		
	B_1	B_2	B_3	B_4	B_5	B_1	B_2	B_3	B_4	B_5	B_1	B_2	B_3	B_1	B_2	B_3
[a]	4	13	35	110	29	84	70	24	52	194	17	20	33	107	103	92
[o]	3	13	14	1	14	52	32	12	230	94	16	19	31	71	64	57
[u]	0	0	1	0	0	18	16	12	270	48	15	16	40	33	32	53
[i]	0	50	8	23	24	18	4	44	48	88	11	16	24	29	70	76
[ɪ]	0	2	190	43	400	16	88	44	172	76	14	22	36	30	112	256
[e]	3	28	85	240	255	20	50	140	76	112	11	19	27	34	107	252

При слабом демпфировании, т. е. малой ширине полосы формант, частота формантного пика совпадает с той частотой, при которой входной импеданс в эквивалентной схеме модели речевого тракта является чисто активным. Увеличение ширины полосы формант, вызываемое активным сопротивлением источника, прямо пропорционально соответственной проводимости. Входная проводимость при частоте форманты пропорциональна ширине ее полосы и определяется диссипативными элементами, за исключением сопротивления источника. Данные о ширине полосы формант, приведенные в табл. 9, могут быть использованы в силу этого для расчета активных входных сопротивлений речевого

тракта. Вычисленные значения находятся в пределах 1,6—7,3 ρc для первой форманты, 0,14—5,8 ρc для второй и 0,45—4,2 ρc для третьей. Таким образом, эти сопротивления по порядку величины близки к ρc , как это было ранее установлено Ван-ден-Бергом [8] для модели [i].

Демпфирующее влияние сопротивления голосовой щели определяется не только его абсолютной величиной, но зависит и от характеристического сопротивления $(L/C)^{1/2}$ резонатора, которое изменяется с конфигурацией речевого тракта и типом резонанса. Так, первая форманта для [a] зависит от сопротивления голосовой щели значительно больше, чем первые форманты [u], [i], [i] или [e]. Этого и следовало ожидать, учитывая суженную фарингальную область при произнесении [a] чему соответствует высокое характеристическое сопротивление (см. формулы (2.6) и (III.107)).

Входной импеданс речевого тракта, наблюдаемого со стороны губ, в области частот данной форманты аналогичен импедансу последовательного резонансного контура. Ввиду того, что сопротивление излучения пропорционально ω^2 , его демпфирующим влиянием на первую форманту можно пренебречь, что видно из аналитических формул и расчетов, приведенных в § 6 Приложения III. Демпфирующее влияние как сопротивления излучения, так и сопротивления голосовой щели увеличивается по мере увеличения зависимости частоты рассматриваемой форманты от передней и соответственно задней полости; это видно из сравнения данных приведенной выше таблицы с данными, рассмотренными в § 2 гл. 7. Так, данные о ширине полосы формант отражают одинаковую важность передней и задней полостей для первых двух формант задних гласных [u], [o], [a], а также вполне определенную связь второй форманты с задней полостью и третьей форманты с передней полостью для [i] и второй форманты с передней полостью для [i].

Вследствие трансформаторного эффекта в полостях речевого аппарата и сравнительного постоянства конфигурации в области гортанной трубки зависимость демпфирования от изменений сопротивления голосовой щели не так велика, как от изменений сопротивления излучения.

Для звука [u] сопротивление излучения, как видно, не является существенным фактором, оказывающим демпфирующее влияние на любую из его формант; это и следовало предполагать, имея в виду малое раскрытие губ при образовании этого звука и появление в силу этого индуктивности последовательно с активным сопротивлением излучения. Также относительно невелико и влияние сопротивления голосовой щели, при величине его $5\rho c$, на демпфирование первых трех формант этого звука.

Последний столбец включен в табл. II для того, чтобы показать, какая ширина полосы формант получается, если просуммировать влияние всех тех диссипативных элементов, которые поддаются теоретическому расчету. Необходимо, однако, помнить, что индуктивность, включенная последовательно с сопротивлением голосовой щели, должна уменьшать его демпфирующее влияние на третью и более высокие форманты и на вторую форманту в случае высокого ее частотного положения.

Потери в стенках полостей на теплопроводность и трение у стенок полостей, определенные по формуле для системы с идеально жесткими стенками, были введены в вычисления с добавочным коэффициентом $S_A = 2$, учитывающим форму и характер поверхности стенок полостей; см. § 1 гл. 2 и § 6 Приложения III. Большая часть этого коэффициента

учитывает обычно встречающиеся отклонения формы поперечных сечений речевого тракта от круговой; свойства поверхности стенок полостей речевого аппарата, которыми определяется трение, неизвестны, причем трудно отделить потери на трение от потерь, связанных с колебаниями этих стенок. Известно только, что первые доминируют в узких проходах, играющих роль горла резонатора Гельмгольца.

Ширина полосы формант, вычисленная как сумма $A + B + V$ (табл. 11), менее показательна, чем данные табл. 9, которые были получены без учета сопротивления голосовой щели, на основании эмпирической формулы (7.5) для постоянной затухания полостей речевого аппарата. Основной недостаток данных табл. 11 заключается в том, что первые две форманты $[a]$ оказываются задемпфированными слишком сильно, а первые форманты других гласных, кроме $[o]$, — недостаточно задемпфированными. Предварительные исследования показателя затухания для резонансов речевого тракта, имеющих место при закрытой голосовой щели, дают довольно обоснованные значения ширины полос около 50 $гц$ вплоть до частоты 2000 $гц$, с быстрым возрастанием для более высоких частот. Резонансы в области частот 300—500 $гц$ могут иметь даже меньшую ширину, достигающую до 25 $гц$. Если сложить значения V_1 , получающиеся из столбца V , со значениями из столбца A , то получающиеся величины не согласуются с измеренными. Представляется вероятным, что колебания стенок полостей¹⁾ несколько увеличивают V_1 , как предположил Ван-ден-Берг. Демпфирующее влияние голосовой щели при фонации отвечает, по-видимому, сопротивлению ее порядка 5—15 $рс$. Это сопротивление изменяется в течение цикла фонации.

Величина постоянного потока воздуха для гласного $[u]$ удовлетворяет условиям возникновения турбулентности в губном проходе, что, согласно формуле (III.104), дает прирост V_1 , достигающий до 100 $гц$, если считать $u = 150 \text{ см}^3/\text{сек}$, $A_{\min} = 0,2 \text{ см}^2$ и коэффициент относительного изменения F_1 в зависимости от губного прохода для $[u]$ равным 0,86 согласно табл. 6. Зависящий от скорости постоянного потока воздуха прирост ширины полосы форманты имеет порядок 15 $гц$ только для $[i]$; число Рейнольдса Re , равное 1000, может оказаться ниже критического значения, необходимого для возникновения турбулентности.

ГЛАВА 8

НОСОВЫЕ ЗВУКИ И НАЗАЛИЗАЦИЯ

§ 1. Физиологические данные

Теория назализации, т. е. взаимодействия полостей носа и рта при формировании звуков, образующихся без ротовой смычки, тесно связана с теорией собственно носовых согласных, т. е. звуков, образуемых при открытых носовых полостях и полностью сомкнутом ротовом проходе. С акустической точки зрения, независимо от фонемной классифи-

¹⁾ Последние исследования Хауза и Стивенса [102] показывают, что потери в речевом тракте, обусловленные колебаниями стенок, определяют основную часть демпфирования первой форманты (см. Приложение III, § 6, Б).

кации, представляют интерес следующие физиологические противопоставления, характеризующие этот класс звуков. Одно из них относится к функции мягкого нёба как органа, обуславливающего передачу звука или только через рот при закрытом проходе в носовую полость, как при ротовых звуках, или только через нос, как при велярных носовых согласных, или одновременно через носовой и ротовой проходы. В послед-

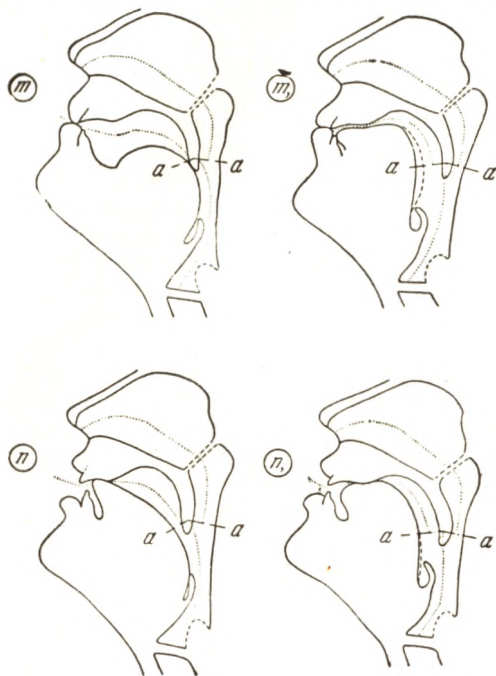


Рис. 46. Контуры речевого тракта по рентгенограммам для носовых согласных: непалатализованных [m] и [n] и палатализованных [m_п] и [n_п]. Сечения а-а в области язычка показывают принятую границу между ротовой, носовой и фарингальной полостями, т. е. место, в котором эти части речевого тракта соединяются между собой.

нем, промежуточном, положении мягкого нёба открытие или закрытие рта приводит к различению назализованных гласных от назального звука. Другое противопоставление относится к наличию или отсутствию препятствия внутри системы носовых полостей. Закрытый носовой проход, например, в результате распухания мягких тканей при насморке, ведет к явно выраженному ослаблению назального звука и изменению качества назализованных гласных. Так называемые «носовые гласные», например в системе гласных французского языка, образуются, наряду с элементом назализации, и за счет специфического положения язычка.

Назализация, как правило, встречается в гласных тогда, когда они расположены вблизи или, лучше, между двумя носовыми согласными. В ударных слогах ассимилированная назальность может сочетаться с укорочением или

даже отсутствием интервала неопределенного назального звука в носовых согласных; в этом случае как раз назализация гласного становится главным признаком, говорящим о наличии в слоге носовой фонемы. Вероятно, небольшая степень назализации свойственна всем очень открытым гласным. Хорошо известна тенденция к опусканию мягкого нёба в таких открытых гласных, как [a] и [e]; это можно видеть и на рис. 38.

Носовой тракт речевого аппарата менее доступен для изучения, чем ротовой. Анатомическая изменчивость ширины носовых проходов и различное количество слизи, наполняющей полости и сужения, проявляются в изменчивости деталей спектрографических картин для разных дикторов или для одного и того же диктора в разных случаях.

Измерение размеров носовых полостей связано с весьма грубыми приближениями. Для нашего диктора общая длина носовых путей, измеренная на основании рентгенограмм по кратчайшему расстоянию от язычка до отверстия ноздрей, оказалась равной 12,5 см, согласно рис. 46, показывающему сагиттальный разрез головы при непалатализованных [m] и [n] и палатализованных [m_п] и [n_п]. Функция площади,

отражающая изменение площади поперечного сечения на различных участках носового тракта, была рассчитана на основе выполненных Г. Бьюггеном¹⁾ слепков носового тракта свежего трупа. Эти данные были приведены к длине носового пути, установленной для нашего диктора.

Левый и правый носовые каналы идут приблизительно параллельно на протяжении 8 см от отверстия ноздрей и затем сливаются, образуя носоглотку. Каждая из передних половин носовых каналов содержит

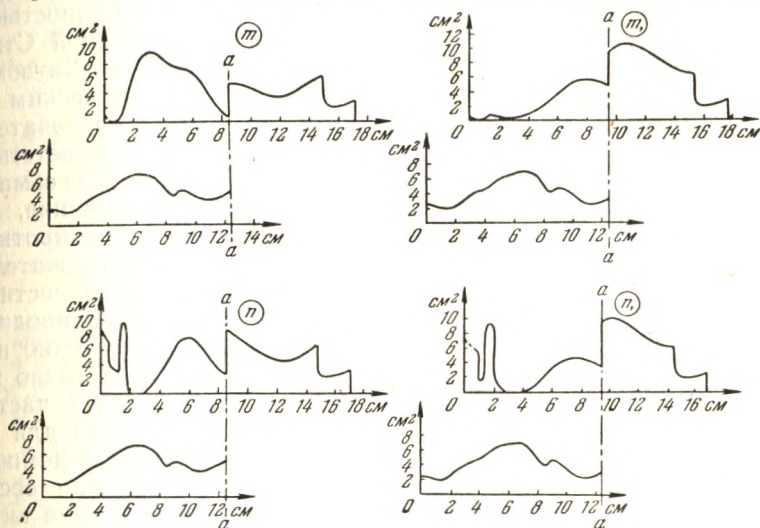


Рис. 47. Функции площади для носовых согласных [m], [n], [m,] и [n,]. Ротовая и носовая части речевого тракта показаны раздельно. Место соединения между ними отмечено линиями а—а.

нижнюю, среднюю и верхнюю ветви, полностью сообщающиеся между собой по всем поперечным сечениям. Они, по-видимому, связаны слишком тесно для того, чтобы рассматривать их как независимые резонаторы. Если правая и левая части полностью симметричны, они будут функционировать акустически как единая система полостей. Именно такая идеальная конфигурация была положена в основу расчетов; поэтому следует ожидать дополнительного размытия спектральной картины вследствие возможной асимметрии, которая вызывает различия в частотах формант, определяющихся правым и левым носовыми проходами и смешивающихся определенным образом при излучении через нос. Причины более сильного демпфирования резонанса в носовом проходе по сравнению с ротовой частью речевого тракта понятны и заключаются в большей величине отношения периметра к площади поперечного сечения на всем пути носового прохода, кроме носоглотки. Коэффициент формы, входящий в уравнение (III.103), здесь достигает значения, равного 3. Затухание увеличивается также благодаря волосяному покрову ноздрей.

Все расчеты были выполнены на электрическом аналоге LEA, описанном в гл. 6. Ряд фильтровых звеньев, соответствующих каждому отрезку тракта с эффективной длиной в 1 см и различной площадью

¹⁾ Личное сообщение.

поперечного сечения, соединялся каскадно и создавал носовую ветвь электрического аналога. В систему было введено нагрузочное звено, воспроизводившее эффект излучения и смещения звука, излучаемого через нос и рот. Носовой тракт был соединен с ротовой системой в плоскости сечения $a-a$ на рис. 46 и на соответственных графиках функции площади (рис. 47). Такая система полностью идентична использованной Стивенсом и Хаузом [195] и Хаузом [99] в опытах с электрическим аналогом MIT. Эти исследователи вводили в цепь, соответствующую носовому проходу, весьма значительное демпфирование, распределенное в виде сопротивлений, включенных последовательно с катушками индуктивности и конденсаторами; это приводит к тому, что потери частично не зависят от частоты, частично пропорциональны квадрату частоты. В нашем исследовании для получения необходимого демпфирования был использован сосредоточенный элемент в виде активного сопротивления в 1 ps , шунтирующего носовой тракт на полпути по длине прохода. Такое простое решение было выбрано на основании эмпирических данных о большем затухании второго носового резонанса, расположенного в области 1000 гц , по сравнению с затуханием первого и более высоких резонансов. При выполнении расчетов был использован стандартный голосовой источник со скоростью спадания огибающей спектра в 12 db на октаву.

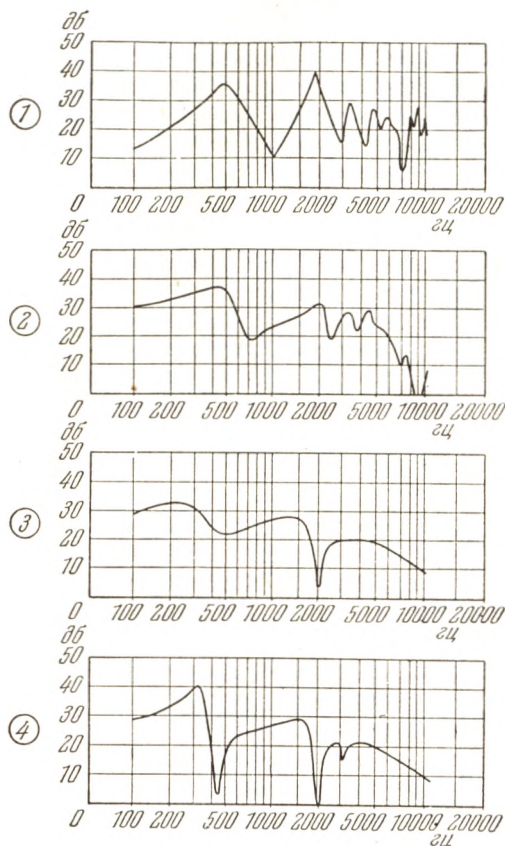


Рис. 48. Уровень напряжения (в db) на входных зажимах электрической цепи, эквивалентной носовому тракту, питаемой от источника переменной частоты через сопротивление 5 ps . Эти кривые дают частотную зависимость входного импеданса носового тракта.

- 1) Площадь поперечного сечения входа в носовой тракт $2,6 \text{ см}^2$. Функция площади соответствует [ш.], см. рис. 47; демпфирование определяется сопротивлением 1 ps , шунтирующим носовую полость.
- 2) Площадь входного сечения уменьшена до $0,65 \text{ см}^2$.
- 3) Площадь входного сечения уменьшена до $0,16 \text{ см}^2$, а площадь поперечного сечения наиболее узкой части прохода в ноздрях уменьшена до $0,32 \text{ см}^2$ на протяжении 2 см .
- 4) То же, что (3), но сопротивление, шунтирующее носовой тракт, устранено.

жения на входе носового тракта при питании его от генератора звуковых частот через сопротивление 5 ps . Эти данные представляют особый интерес с точки зрения вопроса о назализации.

Кривая 1 относится к тому случаю, когда звено, связывающее носовой проход с нёбно-глоточным, имело площадь поперечного сечения $2,6 \text{ см}^2$ и длину $1,5 \text{ см}$. Нули импеданса имеют место при частотах $0, 1000, 2800, 4200 \text{ гц}$. Между ними находятся максимумы. Как было

показано в § 4 гл. 7, ожидаемое среднее расстояние между двумя последовательными максимумами или минимумами должно быть $\frac{c}{2l_{\text{полн}}} = \frac{35\,300}{25} = 1\,400$ гц. Кривая 2 относится к случаю, когда площадь поперечного сечения связующего звена уменьшена до $0,65 \text{ см}^2$; это ведет к понижению частот нулей импеданса на несколько сотен герц. Кривая 3 характеризует влияние на входной импеданс уменьшения площади поперечного сечения связующего звена до $0,16 \text{ м}^2$, а общей площади выходных отверстий ноздрей до $0,32 \text{ см}^2$ на протяжении 2 см. Кривая 4 показывает, что удаление демпфирующего сопротивления ведет к обострению максимума на частоте 300 гц и первого минимума на частоте 430 гц.

§ 2. Носовые звуки, образуемые при закрытом рте

Рассмотрение носовых согласных проще, чем назализованных гласных, так как здесь необходимо принять во внимание только излучение через носовые отверстия во время собственно назального звука. Спектры носовых согласных были вычислены на основе конфигураций

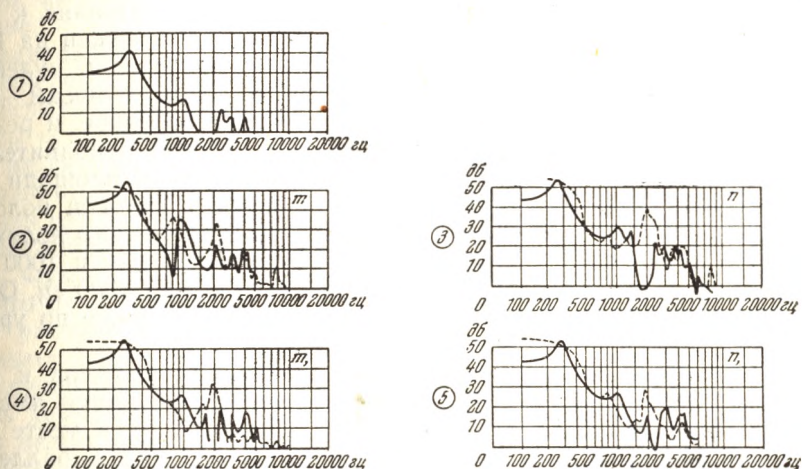


Рис. 49. Расчетные (сплошные кривые) и измеренные (пунктирные кривые) спектры «назального звука».

- 1) Шунтирующая ротовая полость в модели звука [п.] полностью отключена.
- 2) — 5) Звуки [m], [n], [p], [q].

резонаторов, приведенных на рис. 47, и соответствующих данных спектрографического анализа речи диктора. Эти спектры показаны на рис. 49. Кривая 1 отражает эффект исключения из всей системы ротовой полости, шунтирующей носовой тракт при произнесении [п.]; при этом фаринкс и носовой тракт образуют простую систему каскадно соединенных звеньев без ответвлений, которая, по-видимому, должна иметь место при образовании велярных носовых звуков. Частота первой форманты равна 300 гц и соответствует основному резонансу фарингальной полости в совокупности с носовой системой, которая при этом играет роль горла резонатора. Вторая форманта рассчитанного гипотетического велярного носового звука имеет частоту 1000 гц, очень близкую

к частоте первого минимума импеданса носового тракта, который виден на кривой *I* рис. 48. Частота третьей форманты 2200 гц, очевидно, связана в основном с полуволновым резонансом фарингальной полости, а частота четвертой форманты 2900 гц отражает второй минимум импеданса части носовой системы, расположенной выше языка. Форма спектра, представленная кривой *I*, подобна спектру звука [u +]; различие заключается лишь в большем демпфировании второй и более высоких формант.

Влияние полости рта как боковой камеры, шунтирующей передачу звука по носоглоточной системе, вызывает сдвиги резонансных частот и появление антирезонансов. Последние наблюдаются на частотах, при которых импеданс ротового шунта равен нулю, так что весь ток, создаваемый гортанным источником, проходит через него. На диаграмме 2 рис. 49, представляющей собой вычисленные и измеренные спектры, этот эффект ясно виден. Ротовая полость образует резонатор Гельмгольца, закрытый у губного конца и соединенный с фарингсом довольно узким проходом по обеим сторонам язычка, центр которого прикасается к спинке языка.

Очевидный результат наличия связи с полостью рта заключается в увеличении полного объема, что сдвигает основной резонанс к более низкой частоте, 250 гц. Антирезонанс при 800 гц расположен на 150 гц ниже второго пика расчетного спектра. Анализ полюсов и нулей системы показывает, что этот пик соответствует двум полюсам — резонансу носовой полости при 1000 гц, уже упомянутому выше, и резонансу при 900 гц, обусловленному полостью рта. Путем дополнительных опытов на аналоге было установлено, что увеличение площади поперечного сечения прохода, осуществляющего связь с ротовой полостью, как в случае неполного опускания мягкого нёба или низкого положения языка, приводит к сдвигу антирезонанса ротовой полости к 1000 гц и, как следствие, к нейтрализации носовой форманты при 1000 гц. Остающийся при этом пик в области 1100 гц был на 15 db ниже по уровню, чем двойной пик на диаграмме *I*.

Приближенный способ определения положения полюсов и нулей на шкале частот состоит в следующем. Сперва вычисляется импеданс в узловой точке носоглоточной системы у язычка, т. е. в месте связи двух систем полостей. Этот импеданс представляет собой параллельное соединение импеданса той части носового тракта, которая расположена выше язычка, и импеданса расположенной ниже фарингальной полости. Частоты максимального импеданса такой цепи должны совпадать с резонансными частотами всей носоглоточной системы без ротового шунта; к такому именно случаю относится кривая *I*. Затем строится частотная зависимость выходного импеданса шунтирующей ротовой полости, взятого с обратным знаком, причем элементы активного сопротивления не учитываются. Полюсы полной системы находятся, как это было установлено раньше, как точки пересечения этой кривой с полученной ранее кривой реактивного сопротивления носоглоточной системы ¹⁾.

¹⁾ Рассмотрение упрощается, если пользоваться проводимостями. Обозначая через $Y_n = \frac{1}{Z_n}$ проводимость носового тракта, через Y_p — проводимость фаринкса и через Y_m — проводимость ротовой полости, получим условие резонанса для всей связанной системы в виде $Y_n + Y_p + Y_m = 0$. Частоты несвязанных резонансов фарингальной системы должны удовлетворять условию $Y_n + Y_p = 0$.

Поскольку входной импеданс цепи, составленной только из реактивных сопротивлений, должен постоянно возрастать с повышением частоты, не считая скачков от $+\infty$ до $-\infty$ в точках максимумов, и, кроме того, ход реактивного сопротивления шунтирующей системы должен начинаться от $-\infty$ при закрытом губном проходе, можно прийти к заключению, что частота нуля реактивности ротового шунта предшествует частоте бесконечного реактивного сопротивления, причем эти частоты тесно сближаются при уменьшении площади связывающего прохода.

Остальная часть речевого тракта при частоте, на которой импеданс шунта равен 0, ведет себя как емкость, т. е. как отрицательное реактивное сопротивление. Поэтому можно полагать, что частота полюса, т. е. резонансная частота всей системы, будет расположена между точками нуля и бесконечности импеданса шунта, т. е. в данном частном случае на 100 гц выше нуля, расположенного при 800 гц. Эти графические приемы в основном те же, что и представленные на рис. 16; см. также рис. 15, показывающий влияние пары полюс—ноль на огибающую спектра. Следующая пара полюс—ноль шунтирующей системы находится приблизительно при 3500 гц. Она имеет меньшее значение.

Правильность расчета спектра [m] подтверждается близким соответствием его реальным спектрам. Так как при измерении спектра были использованы фильтры с полосой пропускания 150 гц, то пики и провалы огибающей спектра оказываются расширенными. Пунктирная линия на графике 2 рис. 49 показывает, что первая пара полюс—ноль естественного [m] лежит на 100—250 гц ниже, чем в расчетном спектре, а именно при частоте 550 гц для нуля и 800 гц для полюса.

Эти данные хорошо согласуются с теми, которые были получены Отиаи, Фукумура и Накатани [158]. Из рис. 4 этой работы можно видеть, что нуль находится на частоте 550 гц и полюс на частоте 900 гц. Наша расчетная кривая спектра [m] в основном согласуется также с данными Хауза [99], если не считать несколько меньшего демпфирования первой и более высоких носовых формант. В естественной речи степень демпфирования, по-видимому, имеет величину, промежуточную между величинами, принятыми в американских и шведских расчетах. У нашего испытуемого имелась явно выраженная носовая форманта с частотой 2000 гц, наблюдаемая у некоторых дикторов и обусловленная, вероятно, большой близостью и, следовательно, взаимным усилением первого резонанса стоячей волны в фаринксе и первого резонанса стоячей волны в полости носа, т. е. формант с частотами 2200 и 2800 гц на графике 1 рис. 49.

Этот пик на частоте 2000 гц является основным фактором, обуславливающим несоответствие расчетных и измеренных спектров для остальных носовых звуков. Постоянная форманта носовой полости, находящаяся около 1000 гц, повторяется в расчетных спектрах остальных носовых звуков (графики 3—5). Она видна на частоте 800 гц в спектрах [n] и [п.], а также, хотя и менее отчетливо, в спектре [m.]. Резонанс ротовой полости в расчетных спектрах находится на частоте 1400 гц, но в измеренных спектрах речи нашего испытуемого он имеет частоту 1200 гц. Эта разница в 200 гц сохраняется весьма устойчиво и может быть объяснена тем, что спектры рассчитывались для случая напряженной артикуляции при протяжном произношении, а спектрограммы естественных звуков были получены при анализе связной речи,

состоящей из слогов типа согласный — гласный. Соответственные широкополосные спектрограммы этой речи приведены в Приложениях.

Основной антирезонанс ротовой полости в расчетном спектре [п] имеет частоту 1800 *гц*, т. е. лежит выше ротового резонанса, в то время как в спектре [т] антирезонанс лежит ниже, чем резонанс ротовой полости. Эта разница является следствием меньшего объема ротовой полости при произнесении [п].

Изменяющаяся по своему частотному положению форманта находится на частоте 900 *гц* в расчетном спектре [т], но на частоте 1400 *гц* в спектре [п] и появляется при 1700 *гц* в расчетном спектре [т.] и 2000 *гц* для [п.]. Она идентична второй форманте назализованных гласных, которые образуются при раскрытии ротовой смычки. В случае [т.] и [п.] она уже не может быть отнесена к ротовым формантам из-за большой площади прохода, связывающего носовую и ротовую полости.

Расположение F_2 в естественных [т.] и [п.] не очень хорошо выявляется на спектрограммах и спектральных разрезах согласного. Исходя из непрерывности перехода к форманте последующей гласной, частоту F_2 можно принять равной 2000 *гц* для обоих звуков. Более низкое положение F_2 для [т.] по сравнению с [п.] вызвано заметно выраженным понижением гортани, как это видно из рентгенограммы на рис. 46. Типичным для всех палатализованных звуков является положение языка, характерное для артикуляции [i]. Это обуславливает возникновение второй форманты с частотой 2000 *гц*, что связано с возникновением стоячей волны в фаринксе.

Частота первого нуля в [т.] и [п.] зависит только от конфигурации ротовой полости, в соответствии с общими правилами, относящимися к случаю наличия шунтирующих ветвей. Возникновение этого нуля можно упрощенно объяснить, если рассматривать ротовую полость как трубу с эффективной длиной в 4 см, закрытую у дальнего конца. Частота нуля импеданса в этом случае равна $\frac{c}{4l} = \frac{33500}{4 \cdot 4} = 2100$ *гц*, т. е.

имеет величину, близкую к полученным при более точном расчете значениям 1800 *гц* для [т.] и 2200 *гц* для [п.], как это видно на кривых 4 и 5. Этот антирезонанс находится, таким образом, в области пика на частоте 2000 *гц*, имеющегося в расчетных спектрах [т.] и [п.], и вызывает избирательное понижение уровня. Второй антирезонанс ротового шунта, если основываться на идеализированном представлении ротовой полости для [т.] и [п.] в виде трубы, получается на частоте $3c/4l = 6300$ *гц*. Расчет при помощи аналога LEA дает соответственно 5600 и 6400 *гц*. Для [т] второй нуль находится на 3500 *гц* и для [п] — на 5600 *гц*.

Суммируя изложенное выше, можно сказать, что носовые звуки содержат весьма устойчивые форманты, обусловленные носовым трактом и фаринксом. Они находятся приблизительно на частотах 250, 1000, 2000, 3000 и 4000 *гц*, причем самая низкая из них имеет наибольшую интенсивность. Форманта на 1000 *гц* не всегда может быть обнаружена на спектрограммах ввиду ее малой интенсивности и подъема на высоких частотах, обычно применяющегося в спектрографах. Форманта на частоте 2000 *гц*, а также более высокие форманты могут оказаться сдвинутыми только в случае сжатия фаринкса, что имеет место при сочетании носовых с некоторыми задними гласными. Имеются также форманты, зависящие от ротовых полостей, но они сильно ослаблены ввиду их большой близости к нулям. Спектр назального звука содер-

жит редуцированную вторую форманту гласных с частотой F_2 , непрерывно переходящую в F_2 прилежащего интервала открытия ротового прохода. Прекращение ротовой смычки ведет к радикальному сдвигу нулей и связанному с этим внезапному возрастанию интенсивности «ротовых» формант, но на положение F_2 открытие ротового прохода оказывает лишь малое влияние. Форманты более высокого порядкового номера, например F_3 , ведут себя подобным же образом.

Форманты назального звука обозначаются в дальнейшем как N_1 , N_2 , N_3 и т. д., в том порядке, в каком они встречаются в спектре. Таким образом, N_1 соответствует F_1 , N_3 соответствует F_2 и в некоторых случаях N_4 соответствует F_3 . Под соответствием здесь подразумевается непрерывность перехода, а не идентичность, так как, например, форманты, обозначаемые как N_1 или F_1 , могут в одинаковой мере зависеть от носовых и ротовых полостей.

Некоторые авторы считают [25, 143], что изменения положения формант в интервалах ротового звука, следующего за назальным звуком, являются наиболее важным для слуха признаком, определяющим различие отдельных носовых фонем. Такое предположение подкрепляется фактом сильного понижения уровня F -формант в интервале назального звука. Однако слышимые остатки F -картины во время произнесения носовых звуков имеют все же некоторое значение, как это можно показать, если при пропевании назального звука передвигать язык из положения, характерного для [i], в положение для [u] [183].

§ 3. Назализация

В последнее десятилетие теория назализации привлекает большое внимание. Кроме лингвистического и фонетического интереса, изучение назализации имеет существенное значение для инженеров связи, занимающихся вопросами синтетической телефонии. Искажение спектральных характеристик гласных вследствие появления дополнительных пиков может внести ошибки в определение порядкового номера формант, если только следящая за положением формант система не будет пренебрегать накладывающейся на гласный структурой, зависящей от назализации.

Как было показано в § 1 гл. 3, используя сведения об избыточности спектральных картин, возможно создать такие инструкции для анализа, вне зависимости от того, производится этот анализ машиной или человеком, чтобы обеспечить правильную идентификацию.

Спектрографические признаки назализации исследованы достаточно подробно. Джуз [121] указывает на добавочную форманту, лежащую выше F_1 при частоте 1000 $гц$, и на тенденцию к появлению дополнительных формант между всеми основными формантами гласных. Автор также указывает на присутствие антирезонанса с частотой 900 $гц$; правда, он не всегда встречается, в силу чего этот признак не считается существенным.

Смит [184] дает следующий перечень признаков «открытой назальности»:

- 1) нет существенных изменений в основном тоне;
- 2) F_1 значительно ослаблена;
- 3) появляется форманта с частотой 1000 $гц$;
- 4) F_2 часто ослаблена и расположена несколько выше по частоте;
- 5) иногда появляется форманта с частотой 2000 $гц$;

6) F_3 ослабляется и частота ее понижается;
7) F_4 усиливается;
8) резонансы выше F_4 обнаруживают тенденцию к ослаблению.
Делаттр [31] дает следующие признаки, полученные в результате обширных исследований, проведенных методом синтеза звуков речи по спектрографическим картинам:

- 1) F_1 ослабляется (первичный признак);
- 2) появляется носовая форманта на частоте 250 $гц$ (вторичный признак);
- 3) имеется, за исключением случая открытых гласных, несущественная форманта с частотой 2000 $гц$;
- 4) F_2 не изменяется;
- 5) F_3 понижается (несущественный признак);
- 6) F_4 смещается вниз по шкале частот (несущественный признак).

Хауз и Стивенс [100] синтезировали назализованные гласные с помощью устройства, аналогичного использованному в настоящей работе, и исследовали восприятие синтезированных звуков. Выводы из этих экспериментов в основном следующие:

- 1) F_1 ослабляется с одновременным возрастанием ее ширины и частотного положения;
- 2) суммарный уровень гласного уменьшается;
- 3) имеются различные вторичные признаки, например появление антирезонанса на частотах 900—1800 $гц$, исчезновение F_3 , нарушения регулярности в области высоких формант и иногда появление добавочных спектральных пиков.

Хаттори, Ямамото и Фуджимура [93] считают, что основными характеристиками назализации являются:

- 1) усиление интенсивности спектра в области около 250 $гц$;
- 2) избирательное уменьшение интенсивности спектра в области около 500 $гц$ (этот антирезонанс приписывается шунтирующему влиянию системы носовых полостей, являющейся ответвлением речевого тракта);
- 3) слабые, рассеянные спектральные составляющие (обусловленные, по мнению авторов, передачей звука через ноздри).

Эти выводы были сделаны на основании спектрограмм, полученных при различных экспериментальных условиях.

Изучение признаков назализации представляет значительные трудности, особенно если требуются детальные данные, так как акустические характеристики назализованных звуков зависят от диктора, от звука, на который накладывается назализация, а также от степени и типа связи носовых полостей с ротовыми. Из приведенного здесь краткого обзора более ранних исследований видно, что наиболее существенным признаком назализации является явное уменьшение интенсивности первой форманты. Из обнаруженного некоторыми исследователями наличия антирезонанса и увеличения ширины полосы форманты аналитически следует и уменьшение интенсивности форманты. С точки зрения теории цепей антирезонанс, обусловленный шунтирующим ответвлением, всегда сопровождается резонансом; поэтому удобно описывать спектральные характеристики назализации, задавая частоты для каждой из пар полюсов и нулей, как это было сделано в § 3 гл. 3 (см. рис. 15 и формулу (3.12)).

Полюс и нуль каждой такой пары четко разделяются в том случае, когда существует сильная связь шунтирующей ветви с основной систе-

мой. В противном случае, при слабой связи, нуль близок к полюсу, и минимум огибающей спектра, обусловленный нулем, не будет выражен достаточно отчетливо. В предельном случае полного закрытия прохода полюс и нуль совпадают, так что получается полная взаимная компенсация их влияния на форму спектра. Весьма вероятно, что первый нуль, обусловленный шунтирующей ветвью, в случае назализации близок к полюсу, соответствующему в спектре назализованных гласных форманте, обозначаемой через F_1 ; следствием этого будет максимальное ослабление уровня интенсивности первой форманты.

Эти общие положения и уравнения, приведенные в § 3 гл. 2, применимы при рассмотрении звука, излучаемого только ртом или только через носовые отверстия. Приведем более полную трактовку назализации с учетом явлений, возникающих при сложении этих двух компонентов. На основе формул (2.43) и (2.48) результат сложения звуков, излучаемых ротовым и носовым отверстиями, при условии достаточного удаления приемника звука (микрофона) от диктора, может быть выражен следующим образом:

$$P_M(s) + P_N(s) = U_q(s) H(s) R(s) [k_{zM}(s) H_{zM}(s) + k_{zN}(s) H_{zN}(s)], \quad (8.1)$$

где $P_M(s)$ и $P_N(s)$ — лапласовы преобразования для ротовой и носовой составляющих назализованного звука. Обе составляющие содержат множитель $U_q(s)$, характеризующий источник, множитель $H(s)$, соответствующий полюсу, и характеристику излучения $R(s)$.

То обстоятельство, что полюс характеризуется в суммарном спектре одним только общим множителем, обусловлено известным положением, согласно которому собственная частота для данной моды колебаний одинакова в любой ветви сложной цепи. Функция нулей ротовой составляющей

$$H_{zM}(s) = \prod_{i=1}^{\infty} [1 + (s + \bar{\sigma}_{Mi})^2 / \bar{\omega}_{Mi}^2] \quad (8.2)$$

и функция нулей носовой составляющей

$$H_{zN}(s) = \prod_{j=1}^{\infty} [1 + (s + \bar{\sigma}_{Nj})^2 / \bar{\omega}_{Nj}^2] \quad (8.3)$$

содержат только комплексные сопряженные нули. Частота i -го нуля для звука, излучаемого ртом, будет таким образом равна

$$\frac{1}{2\pi} (\omega_{Mi}^2 - \sigma_{Mi}^2)^{1/2}.$$

Несмотря на довольно значительное демпфирование в носовой полости, все же соответствующие нулю частота и ширина полосы могут быть определены с использованием приближения, соответствующего большому Q :

$$F_{zM_i} = \frac{\bar{\omega}_{Mi}}{2\pi}; \quad B_{zM_i} = \frac{\bar{\sigma}_{Mi}}{\pi}. \quad (8.4)$$

Приравнявая выражение (8.1) нулю, найдем уравнение, определяющее положение нулей при суммарном излучении ртом и носом:

$$k_{zM}(s) H_{zM}(s) + k_{zN}(s) H_{zN}(s) = 0. \quad (8.5)$$

Коэффициенты $k_{zM}(s)$ и $k_{zN}(s)$, как это следует из формулы (2.47), определяют соотношение между передачей звука по носовой и ротовой

ветвям речевого тракта при низких частотах, т. е. частотах, лежащих значительно ниже первого сопряженного нуля. Каждый из этих коэффициентов содержит нуль и общий для обоих полюсов на отрицательной вещественной оси в плоскости комплексных частот. Эти полюсы и нули определяются активными и индуктивными элементами двух ветвей системы (см. формулу (2.48)). В суммарном звуке оба нуля складываются и образуют новый нуль при частоте, равной частоте общего полюса; это следует из соотношения $k_{zN}(s) + k_{zM}(s) = 1$. Поэтому эти весовые функции могут быть заменены частотно-независимыми множителями

$$k_M = \frac{L_N}{L_M + L_N}, \quad k_N = \frac{L_M}{L_M + L_N}. \quad (8.6)$$

Уравнение (8.5) можно решить обычным способом, подставив $s = \sigma + j\omega$ и приравнявая вещественную и мнимую составляющие нулю.

Если рассматривается только первый нуль каждого из звуков, излучаемых через рот и через нос, можно показать, что сумма содержит только один нуль, частота и ширина полосы которого равны соответственно

$$\left. \begin{aligned} F_{z1} &= F_{zM1} \cdot k_M^{-1/2} (1+a)^{-1/2}, \\ B_{z1} &= B_{zM1} \left[\frac{1 + a B_{zN1}/B_{zM1}}{1+a} \right]^{-1/2}, \\ \text{где} \quad a &= \frac{k_N F_{zM1}^2}{F_{zN1}^2 k_M}. \end{aligned} \right\} \quad (8.7)$$

Если нуль для звука, излучаемого через нос, расположен более высоко, чем нуль для звука, излучаемого ртом, что обычно наблюдается в назализованных звуках, можно ожидать, что в суммарном звуке нуль будет иметь промежуточную ширину и частоту, но будет ближе к частоте нуля для ротовой ветви, особенно в том случае, когда связь с носовой полостью мала. Весовые коэффициенты k_M и $k_N = 1 - k_M$ могут включать также множители, учитывающие возможное различие в ослаблении звука при передаче его к микрофону от рта и от носа. Можно поэтому ожидать, что при перемещении микрофона в звуковом поле от губ к носу первый нуль будет постоянно смещаться от частоты F_{zM1} к частоте F_{zN1} . Так как связь между носовой и ротовой полостями через излучение не отсутствует полностью, точное совпадение с одним из этих идеализированных значений не достигается. Если k_N велико вследствие сужения ротового прохода, возникает некоторое превышение отдачи через ротовое отверстие на низких частотах, обусловленное большим значением R/L для носового пути. Это приводит, как это было показано раньше, к появлению некомпенсированного полюса $k_{zM}(s)$. Величина этого превышения составляет примерно 1—2 дБ при 100 гц.

Влияние второго нуля каждой из систем — носовой и ротовой — на первый нуль суммарной отдачи может быть легко учтено при численных расчетах. Этот эффект в общем мал, так как влияние нуля на ту часть спектра, которая расположена ниже его частоты, невелико.

Метод приближенного определения полюсов и нулей назализованных гласных, если даны формантные частоты для неназализованного варианта этих гласных, а также полюсы и нули входного импеданса носового тракта, не отличается от метода, рассмотренного раньше для

звуков, при образовании которых проход в носовой тракт открыт, а ротовое отверстие закрыто (см. § 1 гл. 8). Сначала вычисляется импеданс со стороны язычка всей ротовой системы, включая полость рта и фаринкса, причем потери не учитываются, т. е. входной импеданс принимается чисто реактивным. Точки, в которых этот импеданс обращается в бесконечность, совпадают с частотами полюсов идеальной неназализованной системы и поэтому близки к частотам формант.

Значение реактивного импеданса должно возрастать от нуля при нулевой частоте, при которой импеданс по направлению к губам обращается в нуль, и проходить через нуль каждый раз, когда импеданс по направлению ко рту и к фаринксу обращается в нуль. Если кривую входного реактивного сопротивления носового тракта нанести с обратным знаком ординат на график кривой импеданса ротовой системы, то полюсы совокупности обоих этих трактов могут быть найдены по точкам пересечения обеих кривых. Графические построения такого рода позволили установить ряд общих правил:

1) Нулями F_{zNi} в спектре звука на выходе носового тракта являются исключительно точки, в которых импеданс ротовых полостей обращается в нуль, если смотреть со стороны язычка; нулями F_{zM_i} в спектре звука на выходе ротового тракта являются исключительно точки, в которых обращается в нуль входной импеданс носового тракта.

Первый нуль на выходе носового тракта в общем случае расположен между F_2 и F_3 неназализованной системы, но в [u] может быть ниже F_2 , а в [i] — выше F_3 .

2) Связь с носовой полостью вносит столько же дополнительных формант, сколько нулей содержит входной импеданс носового тракта.

3) Первый по порядку следования полюс F_{p1} , который не обязательно соответствует F_1 , всегда расположен ниже первого нуля и лежит между частотой F_1 , имеющей место при отсутствии связи с носовым трактом, и частотой первого максимума импеданса носового тракта.

4) При малых величинах связи с носовой системой первый нуль на выходе ротового тракта F_{zM1} расположен чуть выше первого полюса F_{p1} всей системы в целом. По мере увеличения связи с носовой системой F_{zM1} передвигается вверх по шкале частот быстрее, чем F_{p1} . Это тем более справедливо для нуля в спектре общей отдачи F_{z1} в силу относительного увеличения отдачи через нос и более высокого частотного положения первого нуля звука, излучаемого носовым трактом. При условии, что носовой проход относительно широк и свободен от сужений, первый нуль в спектре звука, излучаемого носовым трактом, может достигать частоты 1800 гц, т. е. оказаться даже выше F_2 . Противоположным предельным случаем является полное перекрытие носового прохода в том или ином месте, в особенности у выхода его, в комбинации с малой площадью связи у мягкого неба. Это ведет к снижению частоты первого нуля до частот ниже 500 гц.

5. При наличии сужения на выходе носового тракта происходит понижение по частоте всех полюсов и нулей входного импеданса носовой ветви. Однако если площадь прохода в носовую полость у язычка велика, то нули при этом снижаются незначительно.

Сужение в носовом тракте посередине между язычком и ноздрями ведет к значительному снижению по частоте второго полюса и нуля носового импеданса, помимо небольшого сдвига вниз первой пары полюс — нуль (ср. графики для трехпараметрической модели в § 3 гл. 4).

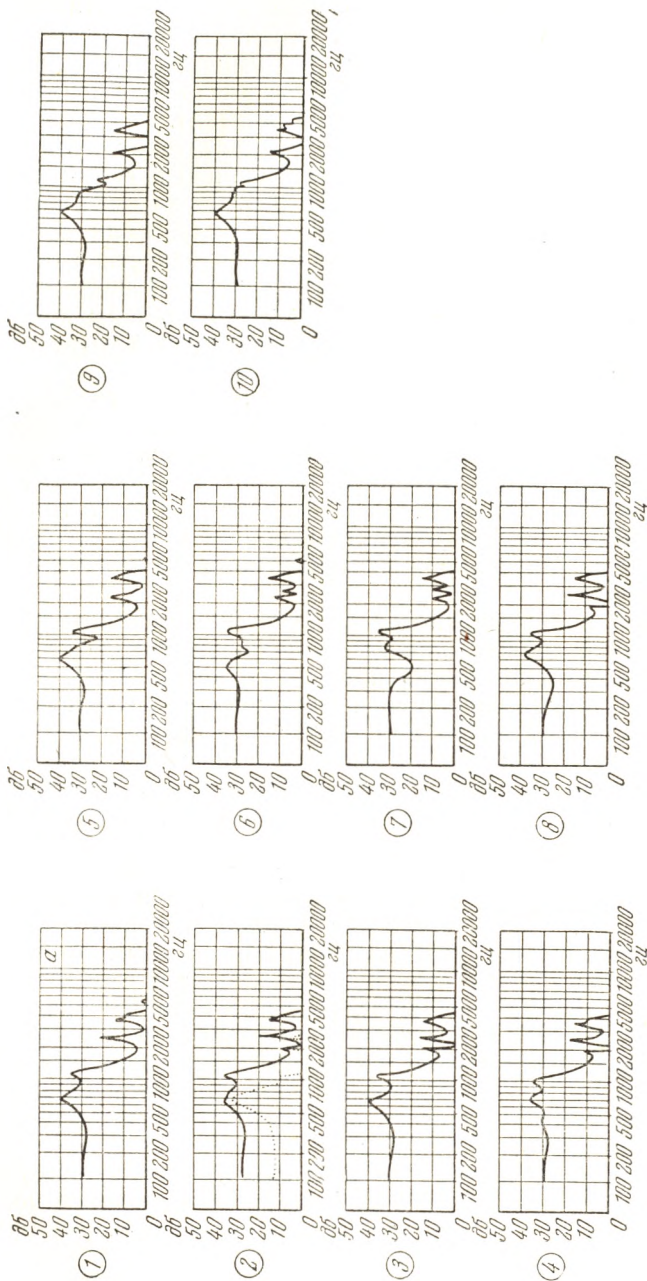


Рис. 50. Влияние связи с носовой полостью на расчетные огибающие спектра звука α . Сопротивление голосовой щели 5 гс .

1. Связь с носовой полостью отсутствует.
2. Излучение через рот (сплошная кривая) и через нос (пунктирная кривая) в том случае, когда площадь носоглоточного прохода равна $0,16 \text{ см}^2$.
3. Суммарное излучение через рот и через нос при условиях (2).
4. То же, что (3), но область наибольшего сужения на выходе носового тракта уменьшена до $0,32 \text{ см}^2$.
5. То же, что (3), но площадь носоглоточного прохода увеличена до $0,65 \text{ см}^2$.
6. То же, что (3), но при уменьшенной до величины, указанной в (4), площади выхода носового тракта.
7. То же, что (6), но при полном закрытии ноздрей.
8. То же, что (7), но площадь носоглоточного прохода уменьшена до $0,16 \text{ см}^2$.
9. Нормальная площадь выхода носового тракта; площадь носоглоточного прохода $2,6 \text{ см}^2$.
10. То же, что (9), но с уменьшением площади входа в ротовую полость до $0,65 \text{ см}^2$, получающимся при приближении мягкого неба к задней части языка.

На основе этой теории, при использовании кривых, полученных с помощью электрического аналога LEA, были определены приводимые ниже данные для нескольких назализованных звуков. Спектры, представленные на рис. 50, относятся к различным степеням и типам назализации при конфигурации речевого тракта, соответствующей гласному [a]; функция площади для этого гласного приведена в таблице в § 3 гл. 7. Кривая 1 представляет спектр гласного без присоединения звена, имитирующего носовой тракт. Первые четыре форманты имеют соответственно частоты 600, 1070, 2400 и 3550 *гц*. Кривая 2 иллюстрирует влияние весьма слабой связи с носовым трактом (при площади поперечного сечения велярно-фарингального прохода в 0,16 *см*²) на спектры звуков, излучаемых носом и ртом, а кривая 3 представляет собой спектр суммарного звука. Наиболее ярко выраженным эффектом на последнем спектре является отчетливое расщепление третьей форманты и уменьшение ее уровня, или, точнее говоря, появление носовой форманты с частотой 2000 *гц* и антирезонанса между этой формантой и F_3 . Этот эффект обусловлен второй парой полюс — нуль входного импеданса носового тракта. Влияние первой пары не так явно вследствие расположения ее между F_1 и F_2 и относительно большого демпфирования ее сопротивлением голосовой щели, приводящего к большему, чем в нормальной речи, значениям $B1$ и $B2$ (см. § 3 гл. 7).

Частота первой носовой форманты F_{p2} равна 660 *гц*, а первый нуль для ротового звука имеет частоту $F_{zm1} = 700$ *гц*. Первый нуль F_{zn1} для носового звука расположен при 1200 *гц*, т. е. чуть выше $F_2 = F_{p3}$. В спектре носового звука доминирует двойной пик, соответствующий F_{p1} и F_{p2} , а F_2 почти полностью компенсируется F_{zn1} .

Различие между спектрами ротового и носового звука на низких частотах определяется степенью связи с носовым трактом, величина которой характеризуется отношением индуктивностей и была принята равной $\frac{L_N}{L_M} = 6,4$, чему соответствует коэффициент шунтирования

$k_M = 0,87$, а также возрастанием на 5% частоты нуля F_{z1} в результирующем звуке по сравнению с F_{zm1} . При такой слабой связи ни F_1 , ни F_2 не изменяются существенно от добавления носовой полости. Интересно, что уменьшение уровня первого формантного пика, видимое на спектре ротового звука, компенсируется звуком, излучаемым через нос, так что в результирующем звуке уменьшение этого пика составляет всего лишь 1 *дб*. Антирезонансный эффект первого нуля F_{z1} ослаблен и проявляется в уменьшении на 1,5 *дб* уровня провала между $F1$ и $F2$.

Кривая 4 рис. 50 показывает влияние уменьшения площади поперечного сечения носового прохода в месте, удаленном на 1—4 *см* от отверстий ноздрей. Остальные условия оставались такими же, как и для кривой 3. При этом выявляется носовая форманта F_{p1} , расположенная ниже F_1 и имеющая частоту 450 *гц*. Кроме того, наблюдается также уменьшение уровня первой форманты на 3,5 *дб* по сравнению с таковым на кривой 1, соответствующей неназализованному гласному. Раздвоение третьей форманты еще сохраняется, но ее низкочастотная часть F_{p3} слабее, чем на кривой 3.

Кривые 5, 6 и 7 относятся к случаю, когда площадь поперечного сечения у носоглоточного прохода равна 0,65 *см*², причем кривая 6 построена для такого же суженного выхода из носовой полости, что и кривая 4. Первый нуль F_{z1} сдвигается до 700 *гц* на кривой 6 до 900 *гц* на соответствующей более широким ноздрям кривой 5. Полное закрытие

ноздрей сдвигает нуль вниз до 500 гц и первый полюс носового тракта до 300 гц. Наблюдается также подъем F_1 до 780 гц и превышение F_4 над F_3 и F_{p2} .

Кривая 8 показывает, как влияет на спектр уменьшение площади носоглоточного прохода до $0,16 \text{ см}^2$ при закрытых ноздрях. Первый нуль снижается до 350 гц, но выражен слабее, чем на кривой 7; уровень первой форманты уменьшается лишь на 1,5 дб по сравнению с уменьшением уровня на 8 дб на кривой 7, соответствующей площади $0,65 \text{ см}^2$. Помимо слабой носовой форманты, всегда находимой ниже F_3 при

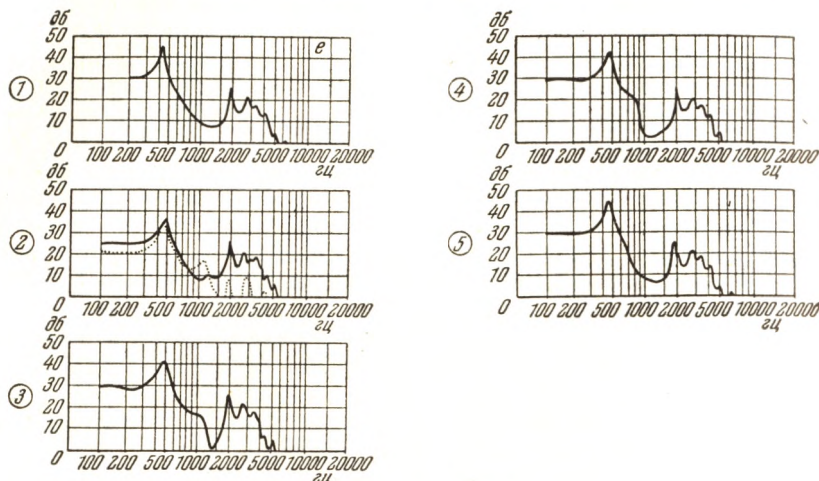


Рис. 51. Влияние связи с носовой полостью на огибающую спектра гласного [е].

- 1) Связь отсутствует.
- 2) Площадь носоглоточного прохода $2,6 \text{ см}^2$; сплошная кривая относится к выходу ротового, пунктирная — носового тракта.
- 3) Сумма излучения через рот и через нос.
- 4) То же, что (3), но после уменьшения площади поперечного сечения носоглоточного прохода до $0,65 \text{ см}^2$.
- 5) Площадь поперечного сечения носоглоточного прохода уменьшена до $0,16 \text{ см}^2$.

наличии некоторой степени назализации, изменения спектра незначительны.

Кривая 9 относится к предельному случаю очень открытого носового тракта, т. е. к случаю большой площади сечения как входа в носовую полость, так и выхода из нее. Первый нуль результирующего звука находится чуть ниже F_2 , что приводит к значительному уменьшению уровня второй форманты. Этот эффект выражен еще сильнее на кривой 10, которая была получена при дополнительном уменьшении площади поперечного сечения на входе в ротовую полость, что соответствует весьма низкому положению небной занавески, приближенной к языку.

Влияние назализации на гласный [е] показано на рис. 51. Расчеты были произведены только для случая более широкой передней части носового тракта. На графике 1 представлен спектр неназализованного гласного; здесь $F_1 = 420 \text{ гц}$, $F_2 = 1960 \text{ гц}$, $F_3 = 2760$ и $F_4 = 3410 \text{ гц}$. На графике 2 приведены раздельно спектры носового и ротового звуков при площади небно-глоточного прохода $2,6 \text{ см}^2$. Сумма этих двух составляющих показана на графике 3. Здесь виден сдвиг частоты первой форманты до 450 гц и увеличение ширины ее полосы. Уровень формант-

ного пика снижается на 7 дБ в звуке, излучаемом через рот, и лишь на 4 дБ в суммарном звуке. Вторая и более высокие форманты не изменяются существенно, однако появляется отчетливая дополнительная форманта на частоте 1100 гц, которую следует относить к $Fp2$. Она находится чуть выше первого нуля в звуке, излучаемом через рот; в суммарном звуке этот нуль повышается до 1350 гц.

Была проведена численная проверка величины сдвига этого нуля в носовом звуке, которая заключалась в следующем: сначала был вычислен коэффициент шунтирования $k_M = \frac{L_N}{L_M + L_N} = 0.62$. Он находится

из отношения между интенсивностью звуков на выходе ротового и носового трактов на низких частотах; согласно графику 2, это отношение равно 1,8 или 5 дБ. Полученное по графику 2 отношение было проверено также интегрированием функций площади для ротового и носового трактов согласно формуле (7.1). Далее решали уравнение (8.5), пренебрегая шириной полосы нулей и замечая, что первый нуль спектра на выходе ротового тракта имеет частоту 1050 гц, второй — частоту 2900 гц, а первый нуль на выходе носового тракта соответствует частоте 2650 гц.

Расчет дал значение частоты для первого нуля суммарного звука, равное 1350 гц, что совпадает с данными, полученными на аналоге. Уменьшение площади носоглоточного прохода до 0,65 см² понижает первый нуль до 1000 гц, как видно из графика 4, а дальнейшее уменьшение этой площади до 0,16 см² сдвигает нуль к 800 гц (см. график 5). Соответствующая носовая форманта лежит чуть ниже этого нуля и проявляется только в изменении наклона огибающей спектра выше F_1 . Этот эффект часто наблюдается на спектрограммах.

На рис. 52 приведены спектры назализованных [i] и [u]. Графики 2 и 3 относятся к большей площади сечения элемента связи — 2,6 см². Можно видеть, что интенсивность звука, излучаемого через нос, на 8,5 дБ больше излучаемого ртом. В общем влияние назализации здесь такое же, как и для звука [e], за тем исключением, что первый нуль здесь достигает частоты 1800 гц, что вызывает явное ослабление второй и более высоких формант. Поскольку в этом случае доминирует носовая составляющая звука, представляет интерес выяснить, насколько этот спектр отличается от спектра, получающегося при полной смычке в месте нёбного сужения. Как видно из сравнения кривых графиков 3 и 4, такая смычка вносит только небольшие дополнительные изменения; это позволяет считать, что кривая графика 3 характеризует собой уже высокую степень назализации. Ввиду узости прохода между языком и нёбом в случае гласного [i] и наличия поэтому большой индуктивности L_M в ротовом тракте, становится ясным, что передача через носовой тракт может быть весьма заметной даже при малой величине связи между ротовым и носовым трактами.

Кривая графика 5 относится к случаю площади сечения элемента связи, равной 0,32 см², а кривая 6 — к случаю площади сечения 0,16 см². Даже в последнем случае частота первой форманты достаточно высока для того, чтобы качество гласной соответствовало в большей мере [I], а не [i]. В силу указанных в § 2 гл. 3 связей между частотами формант, их уровнями и уровнями всех других точек на огибающей спектра, при сдвиге F_1 с 220 до 315 гц должно иметь место повышение уровня высоких формант на $40 \log_{10} \left(\frac{315}{220} \right) = 6$ дБ. Причина того, что такое повышение

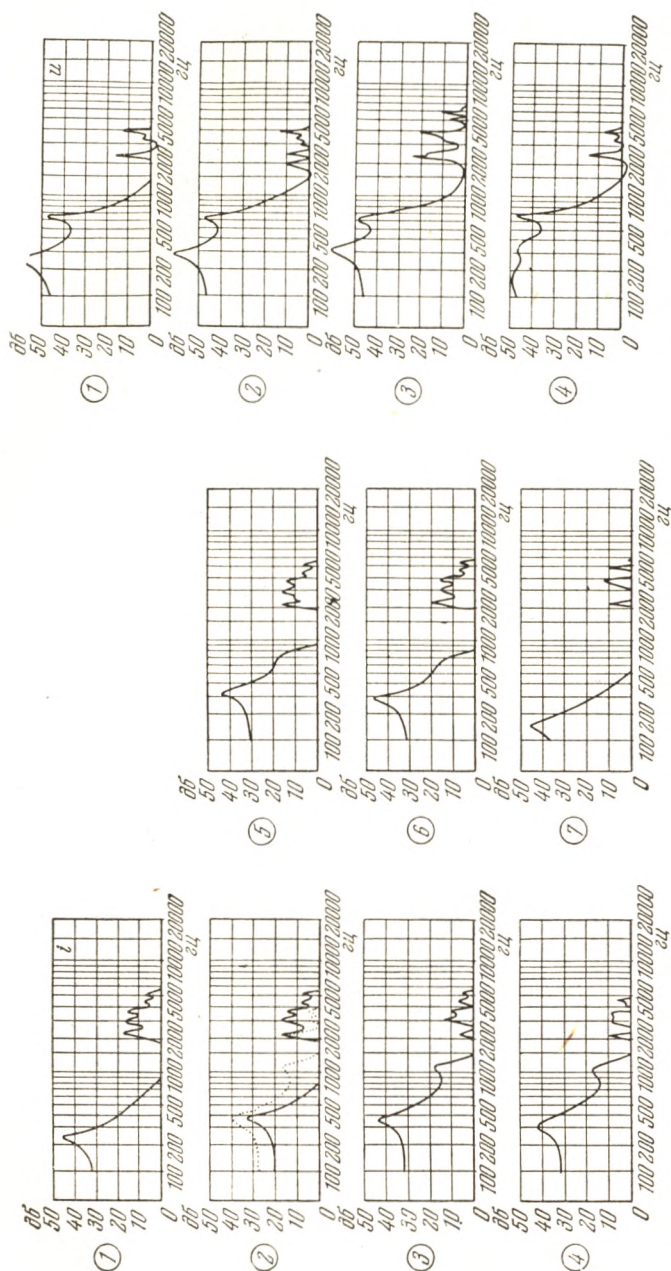


Рис. 52. Влияние связи с носовой полостью на отгибание спектра гласных [I] (графики 1—4 на рисунке справа).

1) Гласный [I], связь с носовой полостью отсутствует.

2) Площадь носоглоточного прохода $2,6 \text{ см}^2$ (сплошная кривая — выход ротового, пунктирная — носового тракта).

3) Сумма излучения через рот и через нос.

4) Нёбно-язычный проход сужен до $1/4$ его нормальной величины, а в середине полностью закрыт.

5) То же, что (3), но при уменьшении площади носоглоточного прохода до $0,32 \text{ см}^2$.

6) То же, что (3), но после уменьшения площади носоглоточного прохода до $0,16 \text{ см}^2$.

7) То же, что (1), но при уменьшенной до $1/4$ нормальной величины площади поперечного сечения нёбного прохода.

Справа:

1) Гласный [u], связь с носовой полостью отсутствует.

2) Сумма отдачи на выходе ротового и носового трактов при площади носоглоточного прохода $0,16 \text{ см}^2$.

3) То же, что (2), но при площади носоглоточного прохода $0,65 \text{ см}^2$.

4) Площадь носоглоточного прохода $0,32 \text{ см}^2$, ноздри полностью закрыты.

уровня не наблюдается (ср., например, кривые графиков 1 и 5 рис. 52), заключается в компенсации этого повышения в основном первым полюсом и нулем носового тракта, которые вызывают снижение уровня высоких частот, имеющее характер, показанный на рис. 15. Так как величина этого снижения равна $40 \log_{10} \frac{F_{z1}}{F_{p1}}$, можно сделать вывод, что отношение нуля и полюса для носового тракта приблизительно равно отношению значения F_1 после назализации к значению F_1 до назализации.

Кривые графиков 1—4 справа на рис. 52 относятся к влиянию назализации на гласный [u]. Как видно из кривой графика 2, даже очень слабая связь с носовой полостью вызывает заметное демпфирование всех формант, кроме первой. Если площадь сечения элемента связи с носовой полостью равна $0,65 \text{ см}^2$, в назализованном [u] наблюдается сдвиг F_1 до 300 гц. При этом нули и полюсы так комбинируются в области между F_3 и F_5 , что происходит усиление F_3 и F_5 . Наконец, как видно из кривой графика 4, полное закрытие носового выхода вызывает почти полное сглаживание первой форманты из-за сдвига первого нуля с 900 гц на кривой графика 3 к частоте, близкой к F_1 .

После этих наблюдений, сделанных на идеальном, но неодушевленном «субъекте» LEA, было необходимо выяснить, как проявляется назализация в реальной речи. На рис. 53 представлены спектральные разрезы и осциллограммы для стационарной части гласного [a], следующего за начальными согласными [n], [n:], [z] и [z,]; участки, для которых были получены эти данные, отмечены на спектрограммах, приведенных в Приложениях. По записям интенсивности было установлено, что интегральный уровень гласного, следующего за носовым согласным, в среднем на 2 дб ниже, чем у той же гласной после других согласных. На спектральных разрезах, приведенных на рис. 53, виден антирезонанс, имеющий неизменную частоту 300 гц. Этот антирезонанс определяется индивидуальным качеством голоса, присутствующим данному диктору, и, по-видимому, не связан с назализацией. Другой индивидуальной особенностью спектрограмм для этого диктора является то, что в общем малый уровень третьей форманты значительно возрастает в гласных с ассимилированной назальностью.

Включение носовых полостей относительно мало сказывается на частотном положении двух первых формант. Внутри каждой пары звуков F_1 понижается на 50 гц в гласной, следующей за носовой согласной. При переходе от [(z)a] к [(n)a] наблюдается уменьшение уровня первой форманты L_1 на 2 дб и второй форманты L_2 на 1 дб, тогда как уровень третьей форманты L_3 увеличивается на 5 дб. В [(n,)a] по сравнению с [(z,)a] L_1 уменьшается на 5 дб, L_2 — на 1 дб, но L_3 увеличивается на 11 дб. Второй нуль на частоте 1800 гц отчетливо выражен как для [(n)a], так и для [(n,)a]. В [(n)a] первый нуль расположен на 800 гц, а в [(n,)a] он, по-видимому, на 50 гц ближе к F_1 вследствие значительного уменьшения L_1 . Связанная с этим нулем носовая форманта имеет частоту 1000 гц.

Такое распределение полюсов и нулей между F_1 и F_2 подобно распределению на кривой 6 рис. 50. Эти данные сходятся со спектрограммами рис. 1 из статьи Делаттра [31], где виден отчетливый антирезонанс на частоте 900 гц. Однако к выводам Делаттра можно добавить, что ослабление первой форманты в значительной мере вызвано этим нулем и что наиболее низкая форманта на 250 гц, обозначенная Делаттром

как FN_1 , мало отличается в назализованных и неназализованных частях как его спектрограмм, так и наших, приведенных на рис. 53. Автоматический регулятор уровня в «Сонаграфе» повышает уровень воспроизведения при ослаблении первой форманты, что ведет к увеличению уровня наиболее низкой форманты на спектрограмме.

Как можно было видеть из рис. 50, носовые проходы должны быть очень узкими для того, чтобы первая «носовая» форманта, накладывающаяся на спектр [а], имела частоту 250 гц. Такие условия, по-видимому, имели место при образовании назализованного [а] на рис. 1 в работе Хаттори и др. [93]. Спектр назализованного [а] при закрытии ноздрей, приведенный этими японскими исследователями, характеризуется тем же признаком в диапазоне частот ниже второй форманты, как и спектр синтезированного нами звука, приведенный на рис. 50, кривая 7.

Третий признак назальности, указываемый Делаттром и ранее упомянутый Смитом [184], т. е. форманту с частотой 2000 гц, можно видеть в большинстве спектров синтезированных звуков рис. 50; он часто представляется в виде расщепленной третьей форманты. Две форманты этой группы обусловлены вторым резонансом носового тракта и полуволновым резонансом либо ротовой полости, либо всей фарингальной области рта и носа. Если мягкое небо опущено очень низко, полная длина фаринкса может стать основным фактором, определяющим нижнюю из двух формант; но если связь с носовой полостью очень мала, эта форманта может определяться в основном носовой полостью. Опыты с синтезом показали, что влияние связи с носовыми полостями на расположение пиков и провалов в области F_3 — F_5 очень изменчиво и не подчиняется каким-либо простым правилам, так как эффект изменения степени связи зависит от индивидуальной конфигурации носового тракта.

В практике анализа речи принято иллюстрировать акустические особенности звуков спектрами. Однако осциллограммы, записанные при большой скорости движения бумаги, иногда дают более полную и даже более ясную картину структуры сигнала. Осциллограммы на рис. 53 были получены на чернильном пишущем осциллографе «Мингограф», который обеспечивает хорошую запись в диапазоне частот от 0 до 800 гц. Этот диапазон достаточен для того, чтобы видеть на осциллограмме форму волны F_1 и дополнительные форманты, а также пульсации давления, создаваемые голосовыми связками.

Наиболее отчетливо акустические особенности назализации выявляются при сопоставлении [п, а] и [з, а]. Колебания, характеризующие форманту, значительно более демпфированы в [а] после [п,], чем в [а] после [п]. Если отбросить обусловленную формантами волнистость, то остаются пульсации давления, создаваемые голосовыми связками; они более отчетливо видны в [п, а], но заметны и в [п, а]. Уменьшение интенсивности первой форманты при наличии антирезонанса является процессом, идентичным тому, который используется для выделения формы волны голосового источника при помощи «обратных фильтров» (см. § 1 Приложения II).

Выводы Делаттра [31] и Хауза и Стивенса [100] о том, что уменьшение интенсивности первой форманты является основным воспринимаемым признаком назализации, были подтверждены опытами с синтезирующим устройством OVE. Уменьшение L_1 при увеличении B_1 приводило к заметной назальности звука. Опыты с электрическим аналогом LEA показали, что добавление носовых полостей также ведет

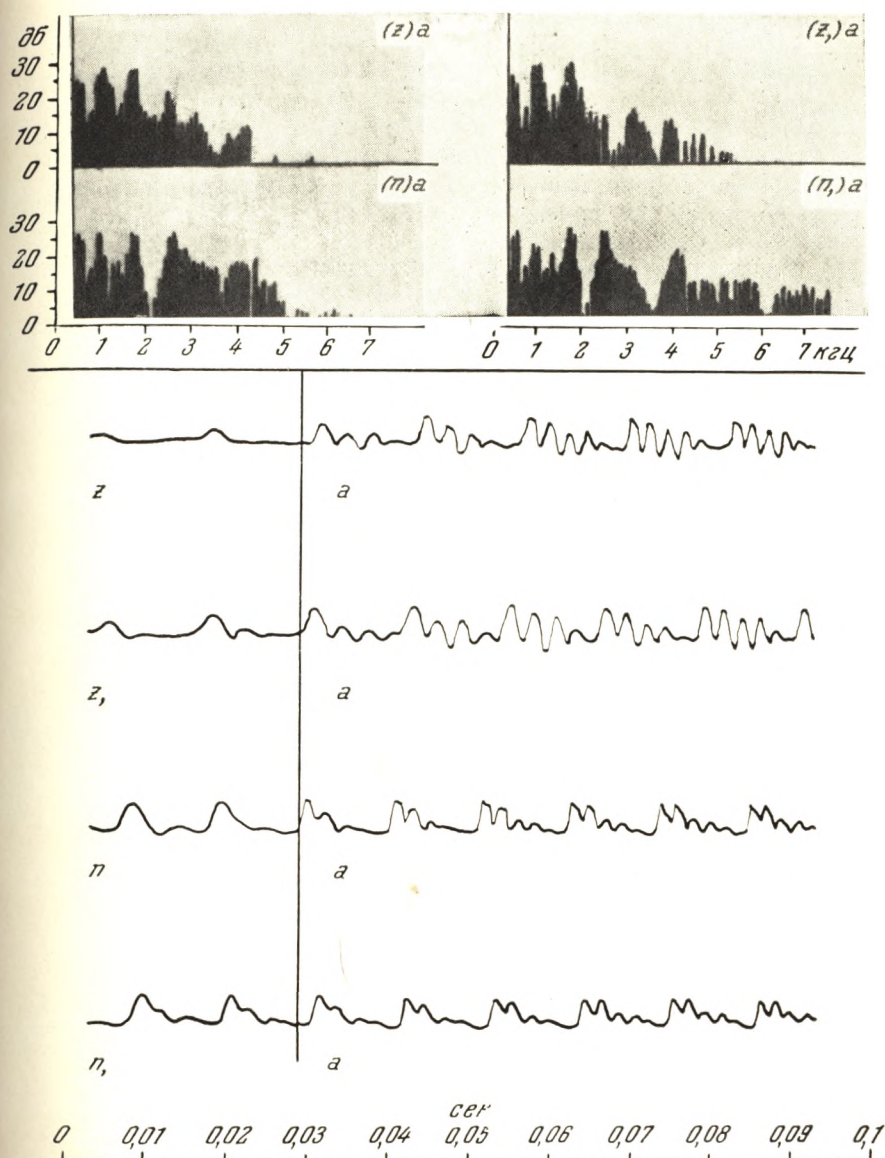


Рис. 53. Спектральные разрезы, полученные при помощи «Сонаграфа», и осциллограммы, записанные «Миньографом», слогов [za], [z, a], [pa], [p, a]. Спектральные разрезы относятся к 3–5 периодам после открытия ротовой смычки. Частотный диапазон при снятии осциллограмм ограничен частотой 1000 гц. Подавление F_1 на осциллограмме типично для назализованного гласного.

к изменению качества звука, характерному для сдвига частоты F_1 вверх. Это изменение связано с действительным повышением F_1 или, точнее, с повышением средней частоты комбинации из F_1 и первого носового резонанса. Так как для нашего испытуемого характерны довольно открытые носовые проходы и отверстия ноздрей, то и первый нуль носового резонанса и связанная с ним носовая форманта расположены, в общем, выше F_1 . Один из теоретических выводов, вытекающих из этого факта, заключается в том, что назализация может компенсировать слишком нейтральное положение языка при артикуляции гласного [а]. Таким образом, назализацией возможно скомпенсировать расширение фаринкса или увеличение объема, лежащего позади места максимального сужения. Без этой компенсации качество гласной приближается к звуку «шва»¹⁾. Построение спектров звуков речи на основании рентгенограммы и синтеза на LEA подтверждают наличие такой возможности²⁾. Расчетное F_1 задних гласных всегда получается слишком низким; это можно видеть, например, для [а] на рис. 40. С другой стороны, теоретически возможен и противоположный эффект в том случае, если носовые проходы узки. Следует также учитывать возможность небольшого положительного сдвига F_1 , обусловленного вибрациями стенок фаринкса, как это показано в § 3 гл. 7; однако этот сдвиг составляет всего лишь около 3% при частоте $F_1 = 600$ гц. Теоретически подобный эффект может дать и шунтирующее влияние индуктивного сопротивления голосовой щели, однако оценить его порядок затруднительно ввиду отсутствия физиологических данных и достаточно достоверных сведений о характеристиках потока воздуха.

На субъективную оценку степени назализации влияет величина и характер изменений основного тона в синтезированных звуках. Ван-ден-Берг и Фант путем прослушивания установили, что появление вибрато может нейтрализовать назальное качество звуков, синтезированных с помощью LEA при наличии в нем цепи, имитирующей носовой тракт.

ГЛАВА 9

ПЛАВНЫЕ

Акустическая теория образования плавных [l] и [r] весьма проста в том случае, если можно пренебречь турбулентным шумом, возникающим в суженном проходе. Как показывают спектрограммы и опыты по синтезу, такое упрощение вполне оправдано. Акустическая теория звонкого [r] не отличается от акустической теории гласных, так как при его образовании отсутствуют шунтирующие боковые полости, обуславливающие появление антирезонансов. Теория латеральных осложняется необходимостью введения функции нулей, подобно тому как это имело место при носовых звуках, хотя здесь она имеет меньшее значение.

На рис. 54 и 55 показаны контуры речевого тракта, построенные по рентгенограммам, и функции площади для палатализованного и непалатализованного [l] и [r]. Место артикуляции продвинуто дальше вперед для [l], чем для [r], и для [l.] — дальше, чем для [r.].

¹⁾ Транскрипция древнееврейского слова, означающего «ничто». Этот термин употребляется для обозначения гласного неопределенного тембра, называемого также нейтральным гласным (Ж. Марузо, Словарь лингвистических терминов, ИЛ, 1960, стр. 338). (Прим. перев.)

²⁾ Неопубликованные данные.

О различии между соответственными парами можно говорить как о различии по точке контакта кончика языка с зубами или альвеолами. Типичные артикуляторные отличия в положении языка для палатализованных и непалатализованных

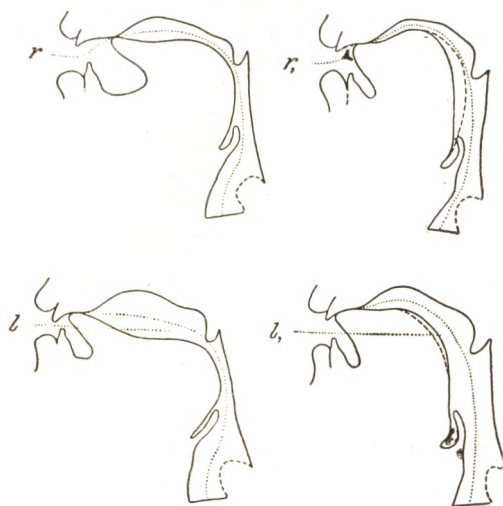


Рис. 54. Контуры речевого тракта по рентгенограммам.

звуков хорошо видны при сравнении [г.] с [г] и [l.] с [l]. Они имеют тот же характер, что и для [i] и [u]. Несмотря на фиксированное положение кончика языка, корень языка сохраняет достаточную свободу для того, чтобы приблизиться к верхней части фаринкса при произнесении [l] и [г]. В результате такого приближения полость, расположенная позади места артикуляции, приобретает конфигурацию, подобную двойному резонатору Гельмгольца. Поэтому в случае непалатализованных звуков образуется вторичное место артикуляции у язычка или несколько ниже его. С другой стороны, положение языка для палатализованных звуков обуславливает конфигурацию задней

полости, характеризующуюся свободным фарингосом с большой площадью поперечного сечения, постепенно сужающимся по направлению к области артикуляторной преграды в ротовой полости. При образовании латеральных эта преграда доходит до полного смыкания, а для

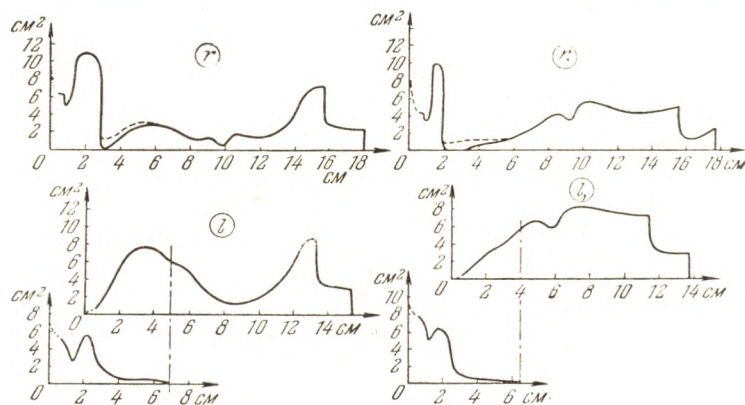


Рис. 55. Функции площади непалатализованных [l] и [g] и палатализованных [l.] и [g.].

дрожащих [г] и [г.] в произношении нашего испытуемого периодически меняется с частотой дрожания.

Латеральный проход, соединяющий внутренние полости при [l] и [l.] с полостями, находящимися спереди языка, можно считать начинающимся вблизи от задних коренных зубов. Поэтому можно лишь предположительно судить о площади поперечного сечения, характеризующей

проход, который может быть как односторонним, так и двухсторонним. Однако для выполнения расчетов существенна только длина этого прохода.

Полученные путем расчета спектры показаны на рис. 56 и 57, на которых сплошные и пунктирные линии относятся к спектрам, вычисленным с помощью электрического аналога LEA. Спектры, изображен-

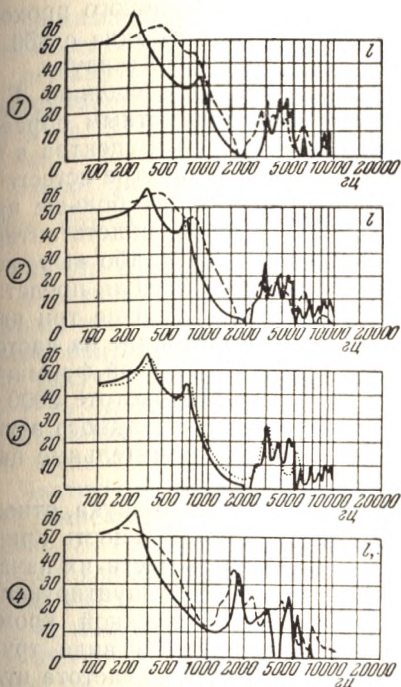


Рис. 56.

- 1) Спектр непалатализованного [л], полученный расчетом по функции площади, приведенной на рис. 55 (сплошная кривая), и измеренный в связной речи (штриховая кривая).
- 2) То же, что (1), но при расчете площадь наименьшего поперечного сечения бокового прохода была увеличена вдвое, а носоглоточного — уменьшена во столько же раз.
- 3) Сплошная кривая — то же, что (2). Пунктирная кривая иллюстрирует эффект устранения апикальной преграды при закрытии бокового шунтирующего прохода.
- 4) Расчетный и измеренный спектры палатализованного [л].

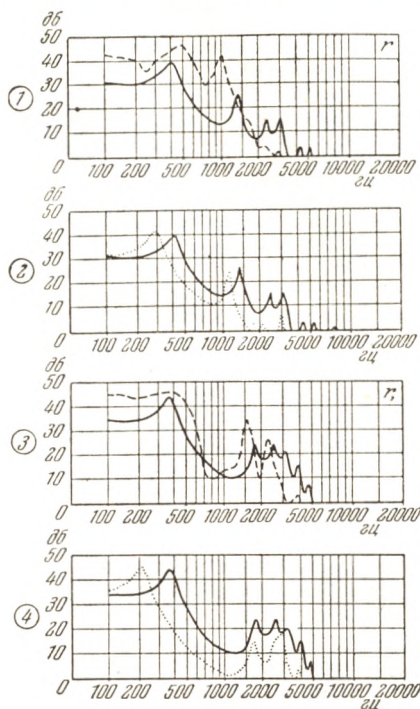


Рис. 57.

- 1) Спектр непалатализованного [г] в фазе максимального раствора, рассчитанный по функции площади, приведенной на рис. 55 (сплошная кривая), и полученный измерением (штриховая кривая).
- 2) Расчетные спектры [г] в условии (1) и в фазе наименьшего раствора (пунктирная кривая).
- 3) Расчетный и измеренный спектры палатализованного [г].
- 4) Расчетные спектры [г] в фазах наибольшего и наименьшего раствора.

ные штриховой линией, соответствуют отрезкам записи звуков, полученным по данным для соответствующих отрезков односложных слов, произнесенных нашим испытуемым. Эти отрезки были подвергнуты детальному анализу в координатах время — частота — интенсивность, как это показано в Приложениях (рис. 82 и 83).

Кривая 1 на рис. 56 позволяет сравнить спектр естественного [л] с вычисленным на основе данных о функции площади, приведенных на рис. 55. Из сравнения можно видеть, что оба спектра имеют много общего: характерно, например, низкое положение F_2 и обусловленное этим значительное расстояние F_2 от частот более высоких формант; это

соответствует таким вторичным характеристикам артикуляции данного звука, как наличие дополнительного сужения в системе полостей, расположенных позади от первичного места артикуляции. Более высокое положение $F_1 = 350$ гц в спектре естественного звука по сравнению с $F_1 = 220$ гц в расчетном позволяет считать, что ширина бокового прохода в расчете была преуменьшена. Поэтому были проведены новые расчеты, в которых площадь минимального сечения латерального прохода была увеличена с $0,3$ см² до $0,65$ см², а площадь фарингального прохода уменьшена с $1,3$ см² до $0,65$ см². Последнее сдвигает F_2 вниз с 850 гц до 620 гц, что несколько ниже, чем в спектре естественного звука.

Увеличение площади бокового прохода ведет к смещению F_1 с 220 гц до 290 гц. Фарингальное сужение не влияет заметным образом на спектральную картину выше F_2 . Форма верхней части спектра в основном одинакова как в расчетных спектрах, так и в спектре естественного звука; однако в расчетном спектре значения частот больше примерно на 5—10%. В спектре естественного звука можно видеть антирезонанс на частоте 1800 гц, сопровождающийся весьма слабо выраженной формантой у 2000 гц. Следующий пик на частоте 2700 гц представляет собой четвертую форманту, после которой имеются еще три пика до частоты 5000 гц. В расчетном спектре виден антирезонанс на частоте 2100 гц и плато на частоте 2300 гц, соответствующее третьей форманте. Четвертая форманта расчетного спектра находится на частоте 2900 гц и соответствует пику на частоте 2700 гц в спектре естественного звука; выше этого пика в расчетном спектре видны еще дополнительные пики до частоты 5000 гц.

Сходство расчетного спектра и спектра естественного звука, относящееся в особенности к антирезонансу, показывает правильность принятой в расчете длины боковых проходов, в том числе и места их начала в задней части рта. Антирезонанс, очевидно, обусловлен шунтирующим влиянием ротовой полости, находящейся позади подвижной кромки языка. Эту полость можно приближенно представить в виде трубы с закрытым дальним концом. Частота антирезонанса, т. е. частота нуля, равна в этом случае $\frac{c}{4l_s}$, где l_s — длина трубы, c — скорость звука. Поэтому частоте нуля 2000 гц соответствует эффективная длина шунтирующей системы $l_s = 4,4$ см.

На графике 3 рис. 56 приведен расчетный спектр, иллюстрирующий эффект снятия апикальной преграды и замыкания боковых проходов. Для сравнения на этом же графике приведена расчетная кривая графика 2. Как видно из пунктирной кривой, антирезонанс в этих условиях исчезает, а слабая третья форманта на частоте 2250 гц, характерная для латеральных естественных звуков, сдвигается по направлению к четвертой форманте при 3000 гц. Пятая форманта остается без изменений, а шестая и седьмая, сдвинутые вверх по частоте на 10%, ослаблены. Третья форманта зависит от передней ротовой полости, включая и боковые проходы, лежащие впереди смычки. Четвертая форманта, на частотах 2900 гц и 3000 гц соответственно, определяется в основном полуволновым резонансом полости между язычком и точкой апикальной смычки. Пятая, шестая и седьмая форманты определяются резонансом гортанной трубки и образованием различных стоячих волн в системе двойного резонатора, расположенного над гортанью.

Такое скопление формант, не меньше пяти, в частотной области, лежащей между 2250 гц и 5000 гц, отстоящих друг от друга на 700 гц,

часто встречается на спектрограммах и является отчасти причиной «звонкости» стационарного участка латерального звука. Тесное расположение формант несомненно связано с общей длиной системы, которая, как это видно из функции площади, составляет, включая ротовой шунт, $l_{\text{полн}} = 22$ см. При такой длине расстояние между формантами должно быть равным $\frac{c}{2l_{\text{полн}}} = 800$ *гц*.

Помимо подавления близкой к нему третьей форманты, антирезонанс обуславливает некоторый подъем спектра, начинающийся с частоты, лежащей на 40% выше антирезонанса, как это видно из кривых спектрального уровня для единичного нуля или полюса, приведенных в § 2 гл. 3. Третью форманту и антирезонанс можно рассматривать как пару полюс — нуль, которые взаимно нейтрализуются и поэтому не влияют на форму спектра, не считая небольшого постоянного повышения интенсивности, равного $40 \log_{10} \frac{F_3}{F_{21}} = 1,6$ *дб*, на частотах выше F_3 , как это следует из формулы (3.13).

В результате нейтрализации третьей форманты нулем, третьей по порядку становится F_4 вместо F_3 , четвертой F_5 вместо F_4 и т. д. С помощью синтеза было установлено, что очень хороший звук [I] может быть получен без введения нуля и что сдвиг F_3 к частоте 2900 *гц*, близкой к F_4 , увеличивает естественность звучания. Идентификация звука обуславливается резким сдвигом F_1 вверх в момент перехода латерального звука в соседний гласный.

Частоты F_1 и F_2 в [I] примерно в равной степени зависят от полостей, расположенных спереди и сзади от фарингального сужения. Зависимость F_1 от суженного латерального прохода и F_2 от фарингального сужения является вполне определенной и аналогична той, которая приведена в § 2 гл. 7 для [u]. Для палатализованного [I.] зависимость F_1 и F_2 от полостей такая же, как для гласного [i]. Отсюда следует, что F_2 определяется полуволновым резонансом в той области комбинированной системы рот — фаринкс, которая расположена позади места образования артикуляторной преграды. Так как это место для [I.] более продвинуто вперед, чем для [i], то длина полости больше и F_2 поэтому расположена ниже.

Полученные в результате расчетов частоты формант для [I.]: $F_1 = 210$ *гц*, $F_2 = 1700$ *гц*, $F_3 = 2500$ *гц* и $F_4 = 3050$ *гц* хорошо совпадают с данными, полученными для естественных звуков, а именно: $F_1 = 230$ *гц*, $F_2 = 1600$ *гц*, $F_3 = 2300$ *гц* и $F_4 = 3100$ *гц*. Третья форманта связана с передней частью ротовой полости, как в [I], а четвертая форманта соответствует цельноволновому резонансу в задней части ротовой полости. Как обычно, F_1 соответствует основной моде колебаний всей системы в целом. По расчетным данным нуль расположен на частоте 3600 *гц*. Его более высокое по сравнению с [I] положение зависит в основном от рупорообразной формы шунтирующей полости. В спектрах естественных звуков этот нуль выражен менее четко, но, по-видимому, он совпадает с минимумом на частоте 2600 *гц*. Очевидно, и здесь в расчете длина шунтирующей полости оказалась преуменьшенной.

Образование [r] и [r.] во многих отношениях сходно с образованием [I] и [I.]. Сдвинутое более назад место артикуляции при [r] уменьшает наибольшее возможное значение площади поперечного сечения системы внутренних полостей. Вторичное место артикуляции, которое разделяет заднюю полость при [r] на две части, находится в области языка, т. е.

выше, чем при [l]. По этой причине F_2 в большей мере зависит от передней из двух внутренних полостей и имеет соответственно более высокую частоту. F_3 по-прежнему обусловлена ротовой полостью, расположенной у передней части языка.

На спектре естественного [г] F_1 имеет довольно высокую частоту (500 гц), как показано на штриховой кривой графика 1 рис. 57. Это объясняется тем, что для анализа был взят временной участок звука, соответствующий наибольшему открытию в интервале дрожания. В спектре естественного звука отчетливо видно, что F_1 лежит выше, а F_2 ниже, чем в расчетном спектре, и что выше F_2 огибающая спектра падает значительно быстрее. Частично это вызвано более высоким положением F_2 в расчетном спектре. Влияние уменьшения площади поперечного сечения апикального прохода при переходе от открытого к закрытому интервалу показано на графике 2. Помимо уменьшения уровня спектра выше F_1 , которое можно было ожидать ввиду сдвига F_1 вниз, наблюдается отчетливое смещение частоты F_3 вниз к области около 2000 гц. Это вызвано исчезновением индуктивности апикального прохода, шунтирующего переднюю ротовую полость в открытой фазе [г].

Совпадение расчетных значений F_1 , F_2 и F_3 в [г.] с результатами измерений для естественного звука очень хорошее, как это видно из графика 3 рис. 57.

Рассчитанные значения $F_2=1700$ гц и $F_3=2400$ гц на 200 гц выше, чем в спектре естественного звука. В последнем отсутствуют четвертая и более высокие форманты, имеющиеся в расчетной кривой. Связь трех первых формант спектра [г.] с конфигурацией речевого тракта в основном такая же, как для [l]. Так, F_2 является полуволновым резонансом внутренних полостей, а F_3 в наибольшей степени зависит от передней части ротовой полости.

Влияние уменьшения площади поперечного сечения апикального прохода показывает график 4. Из него можно видеть, что это уменьшение вызывает очень небольшое снижение F_2 и сдвиг F_4 по частоте вниз по направлению к F_3 . Уменьшение уровня спектра выше F_1 , связанное со сдвигом вниз при закрытии апикального прохода, является, очевидно, главной характеристикой интервала минимальной интенсивности в пределах периода дрожания. Как видно из спектрограммы рис. 83, период колебаний при дрожании у нашего диктора равен 40 мсек, т. е. частота дрожания равна 25 гц. В пределах более открытого интервала видны два периода основного тона $F_0=100$ гц; следы еще двух периодов этого тона видны в закрытом интервале дрожания.

ГЛАВА 10

ЩЕЛЕВЫЕ, АФФРИКАТЫ И ВЗРЫВНЫЕ

§ 1. Щелевые и аффрикаты

Предшествующие разделы были посвящены изучению звуков, в образовании которых участвует голосовой источник. Шепотные плавные и гласные теоретически отличаются от своих голосовых вариантов главным образом случайным характером тонкой структуры и более сильно выраженной высокочастотной частью спектра ввиду шумового характера источника, расположенного в голосовой щели. Однако в образовании

таких «вокалических»¹⁾ звуков может участвовать также шум, возникающий в сужении, расположенном выше голосовой щели. Такой шум не является обязательным атрибутом этих звуков, но может значительно повлиять на их спектр, как, например, в [i], [l] или [r], находящихся под ударением. Начальная часть плавного, следующего за глухим согласным, может быть в результате ассимиляции с последним лишена голосовой компоненты и таким образом становится аспирированным²⁾ или скорее щелевым звуком вследствие возникновения шума в узком проходе над голосовой щелью; этот вопрос рассматривается в § 1 гл. 1 и в § 2 Приложения II.

Акустически общим признаком всех звуков, образованных при участии системы резонаторов заданной конфигурации, является ряд определенных формантных частот речевого тракта, т. е. *F*-картина. Различие в местоположении источника и в огибающей его спектра сказывается лишь на относительных уровнях интенсивности формант. Следует ожидать, что полости, расположенные впереди источника в непосредственной близости от него, оказывают наибольшее влияние на спектральную картину любого звука.

Необходимо при этом иметь в виду одно усложняющее обстоятельство, состоящее в том, что голосовая щель может быть раскрыта и может оказывать заметное влияние на частоты формант и их затухание. Если площадь голосовой щели велика, первая форманта повышается по частоте и сильно демпфируется. Это общее положение относится ко всем звукам, которые образуются полностью или частично за счет источника, расположенного выше голосовой щели, т. е. к взрывным, аффрикатам и щелевым.

С акустической точки зрения, если рассматривать только стационарные спектры звуков, нет больших различий между оглушенным фриктивным [r] и фриктивным согласным [ʃ] с одной и той же артикуляцией, или между соответствующими им сонорными [r] и [ʒ]. Различия между ними заключаются главным образом в наличии или отсутствии модуляции интенсивности, которая для [r] обуславливается одним или несколькими колебаниями кончика языка.

Не следует переоценивать роль шумовых компонент звонких щелевых, аффрикат и взрывных. Так, второстепенное значение шума при восприятии фонемы [j] явилось основанием для классификации ее как глайда [81, 82]. Для идентификации фонемы [v] также, по-видимому, решающим моментом является голосовой источник, как было показано в опытах по синтезу звуков на шведской синтезирующей машине OVEII. Удаление шумовой компоненты сказывается больше на натуральности, чем на разборчивости этого звука. Однако из структурно-теоретических соображений все же следует считать [f] и [v] минимальной парой.

Рентгенограммы, полученные у нашего диктора при произнесении [v], [v.], [z], [z.] и [ʒ], являются очень схожими с соответственными картинами для глухих [f], [f.], [s], [s.] и [ʃ], что позволяет принимать за основу

¹⁾ Этот термин, соответствующий дифференциальному признаку *vocalic*, использован, например, в переводе «Основ фонологии» Н. С. Трубецкого, ИЛ, 1960, стр. 106. (Прим. перев.)

²⁾ Термин «аспирация» здесь противопоставляется «фрикативности» в первую очередь ввиду наличия более сильной связи между различными частями речевого тракта, но кроме того и в силу участия в создании звука голосовой щели, либо самой по себе, либо в комбинации с другими, одновременно возникающими источниками. Короче, аспирация является эквивалентом звука [h].

при расчетах одни и те же значения функции площади речевого тракта как для звонких, так и глухих согласных. Дополнительно к глухим согласным, указанным выше, на рис. 58 приведены рентгенографические контуры для велярного фрикативного [x], глайда [j], нёбной аффрикаты [č] и палатализованного варианта [š], который Халле [81, 82] обозначает как [šč], с тем чтобы подчеркнуть, что палатализация обусловлена последующей фонемой [č], представляющей в ряде диалектов удлинение палатализованного фрикативного. Рентгенографический контур зубной аффрикаты [c] не совсем ясен и поэтому не показан, но на основании

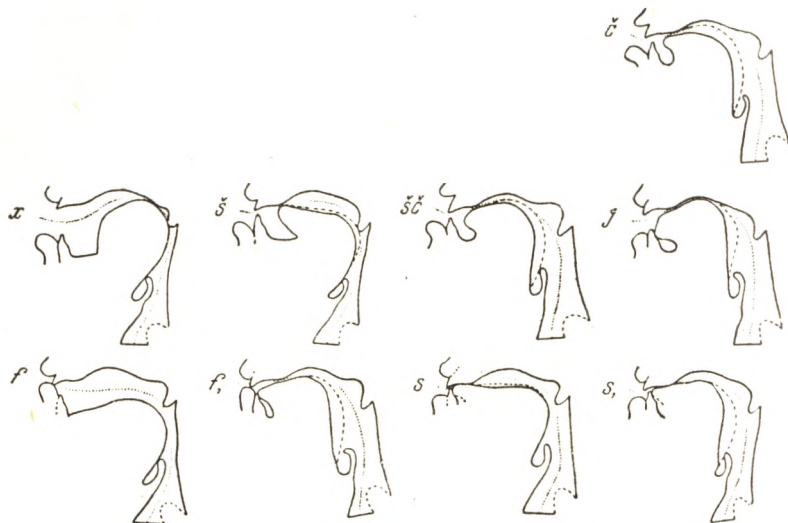


Рис. 58. Контурсы речевого тракта, построенные по рентгенограммам для аффрикаты [č] и фрикативных [x], [š], [šč], [j], [f], [f'], [s] и [s'].

акустических данных не приходится, видимо, сомневаться в том, что ее артикуляция очень близка к [s].

Функции площади для рассматриваемых звуков показаны на рис. 59. Следует отметить, что размеры полости впереди места артикуляции меньше для губно-зубных [f] и [f'], чем для зубных [s] и [s'], и что для тех и других эта передняя полость меньше, чем для всех других звуков. При определении на этом рисунке размеров по длине необходимо учесть, что излучающая поверхность расположена на 0,5 см позади от начала, ввиду чего начальная часть каждой из кривых изображена пунктиром.

Действительную ширину очень узкого прохода в месте или, скорее, в области артикуляции нельзя точно определить по сагиттальным рентгеновским снимкам, так как поперечные размеры, как правило, недоступны для измерения. Дополнительная палатография и томография речевого аппарата диктора была бы полезна для определения этих размеров. За отсутствием таких данных нами были частично использованы сведения, приводимые в литературе. Минимальная возможная площадь поперечного сечения, которую можно было воспроизвести на электрическом аналоге LEA, равнялась $0,16 \text{ см}^2$; она была принята как стандартное значение для очень узких проходов при фрикативных, аффрикатах и взрывных. Некоторая неточность в определении этих размеров не

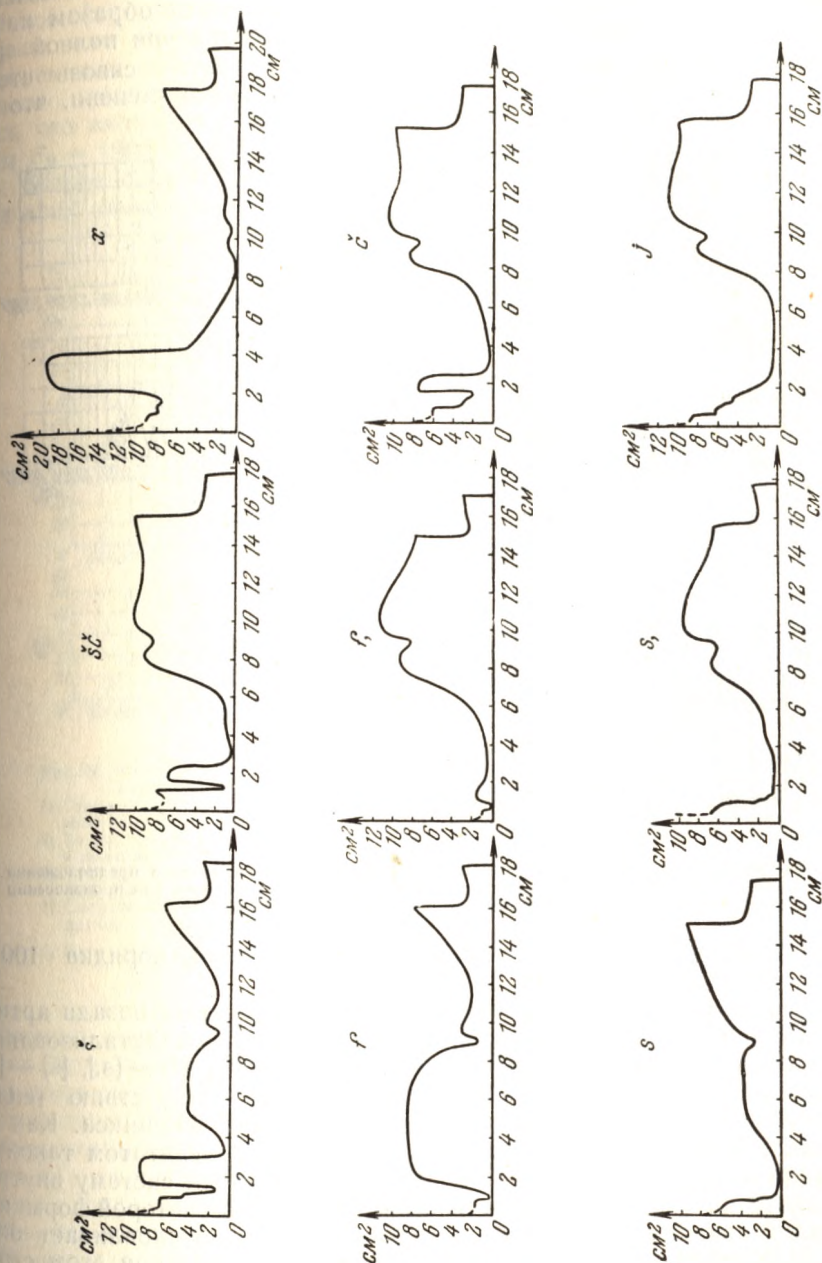


Рис. 59. Функции площади для звуков, указанных на рис. 58.

оказывает серьезное влияние на расчеты, так как размер по длине для узких проходов более критичен, чем площадь поперечного сечения. Ошибка в определении ширины сказывается главным образом на частоте F_1 основного резонанса системы. Однако даже при полной артикуляторной смычке имеет место прохождение звука сквозь стенки резонирующих полостей речевого аппарата в такой степени, что это

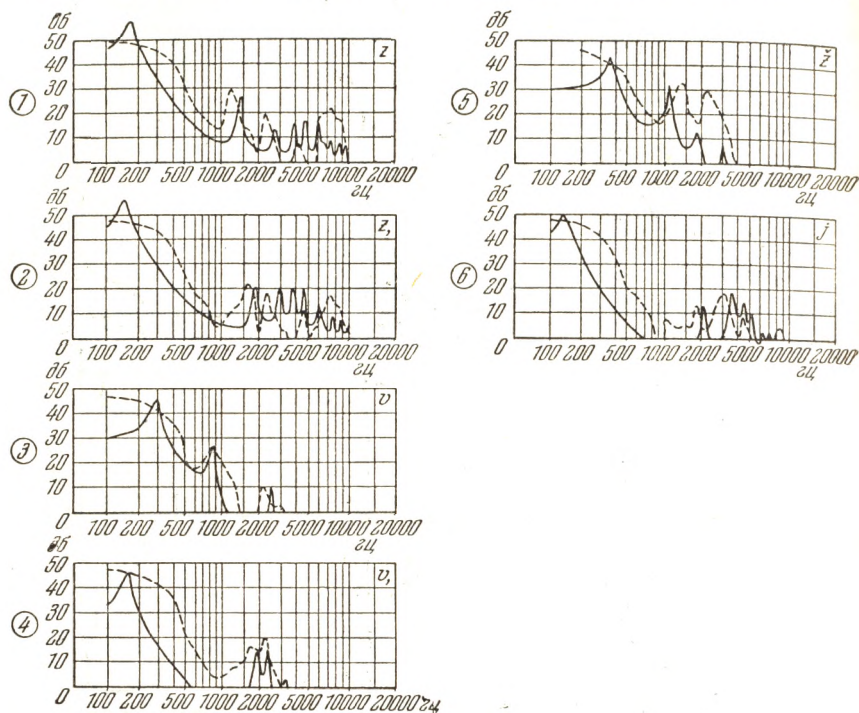


Рис. 60. Спектры $[z]$, $[z.]$, $[v]$, $[v.]$, $[z̃]$ и $[j]$, полученные расчетным путем в предположении стандартного голосового источника (сплошные кривые) и измеренные при произнесении в связной речи (штриховые кривые).

ограничивает минимальное значение F_1 величиной порядка 100 гц (см. § 3 гл. 7).

Различие в конфигурации полостей, расположенных позади артикуляторного сужения, для твердых и мягких, т. е. непалатализованных и палатализованных согласных, таких как $[f] - [f.]$, $[s] - [s.]$, $[ʃ] - [ʃ.]$, сводится попросту к наличию или соответственно отсутствию тенденции к сужению в области язычка и верхней части фаринкса. Как это было выше отмечено при рассмотрении плавных, результатом такой веяризации или, скорее, фарингализации, разделяющей систему внутренней полости на две части, является снижение частоты второй форманты.

Палатализация, в противоположность этому, обуславливает оптимальные условия для получения высокого F_2 , так как при этом структура речевого тракта приближается к единой, не разделенной на части полости с плавным сужением по направлению к закрытому концу в ротовой полости.

Спектры составляющих, обусловленных голосовым источником, в звуках $[z]$, $[z.]$, $[v]$, $[v.]$, $[z̃]$ и $[j]$, полученные расчетным путем с помощью

аналога LEA, и соответственные данные измерений для этих звуков, произнесенных нашим диктором в слогах, состоящих из согласного и гласного [a], приведены на рис. 60. Если не считать тенденции к повышению всех формантных частот на 10—20% в расчетных спектрах по сравнению с измеренными, качественно они весьма сходны. Как и следовало ожидать, влияние палатализации наибольшее для губно-зубных, что видно из сравнения расчетного $F_2 = 850$ гц для [v] с расчетным $F_2 = 1900$ гц для [v,].

Общее соответствие между расчетными и измеренными данными для этих звуков говорит о том, что роль шумовой компоненты в есте-

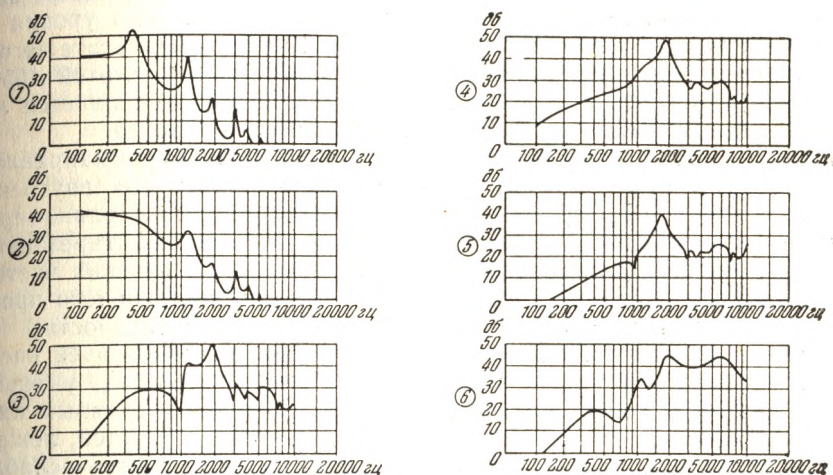


Рис. 61. Влияние местоположения источника и изменений величины демпфирующих элементов при конфигурации речевого тракта, соответствующей звуку [ʃ].

- 1) Стандартный голосовой источник с наклоном огибающей — 12 дБ/окт, сопротивление голосовой щели 5рс.
- 2) То же, что (1), но с добавлением последовательного сопротивления 0,25рс, включенного в апикальной части язычного прохода.
- 3) Источник с равномерным спектром расположен в язычном проходе, остальное как в (2).
- 4) Сопротивление голосовой щели уменьшено до 0,33рс, остальное как в (3).
- 5) Сопротивление голосовой щели вновь увеличено до 5рс, а сопротивление язычного прохода увеличено до 2рс.
- 6) Источник у зубов, последовательно с ним сопротивление $\frac{3}{40}$ рс, остальное как в (3).

ственных звуках, в том числе и для [j], не очень велика. Однако в спектрах естественных [z] и [z,] видна область высокочастотных формант, простирающаяся от 5 до 10 кгц, которая, видимо, обусловлена зубным источником шума. Из приведенных в Приложении спектрограмм слогов, произнесенных нашим диктором, можно видеть, что третья форманта для [z] имеет хаотическую тонкую структуру, указывающую на наличие шума. Учитывая еще и более высокое F_2 в этом примере, можно заключить, что этот звук при измерении спектра произносился с более передней артикуляцией, чем во время съемки рентгенограмм.

Неестественно малая ширина полосы некоторых формант в расчетных данных вызвана тем, что при расчете не было учтено демпфирование, вносимое сужением. Сопротивление голосовой щели принималось равным 5рс, как и для гласных. Следует также иметь в виду, что при спектральном анализе естественной речи использовался фильтр

с шириной полосы пропускания 150 гц, что дает эффект кажущегося расширения форманты на величину того же порядка.

Влияние изменений в величине активных сопротивлений, в расположении источника и в его характеристиках показано на рис. 61 для неизменной конфигурации речевого тракта, соответствующей рентгенографическим данным для звука [š]. Кривая 1, идентичная кривой для [ž] на рис. 60, соответствует стандартному голосовому источнику с наклоном огибающей спектра —12 дб/окт и сопротивлению голосовой щели 5рс. Формантная картина здесь сходна с таковой для [г], например, по положению F_3 , типичному для звуков с ретрофлексной артикуляцией. Как видно из кривой 2, введение сопротивления в 0,25рс последовательно с апикальным сужением ведет к уменьшению уровня первой форманты, имеющей частоту 350 гц, на 13 дб, в результате чего она демпфируется почти полностью. Вторая форманта с частотой 1050 гц ослабляется на 8 дб, а третья, с частотой 1900 гц, уменьшается лишь на 4 дб. Четвертая и пятая форманты почти не изменяются.

Кривая 3 показывает влияние смещения источника по направлению к апикальному сужению при сохранении величины его внутреннего сопротивления и при изменении спектра источника на равномерный. Влияние смещения источника сказывается в появлении антирезонансов, т. е. нулей на 0, 950, 3000, 3900, 4900 гц и на более высоких частотах. Это — частоты, при которых импеданс со стороны апикального прохода по направлению к голосовой щели обращается в бесконечность.

Первый нуль обуславливает сильное ослабление уровней спектра ниже 200 гц. Второй нуль, расположенный чуть ниже F_2 , уменьшает уровень второй форманты настолько, что F_3 становится главным пиком. Этого следовало ожидать ввиду связи F_3 с системой полостей, лежащих впереди источника. Более высокие форманты также заметно ослаблены благодаря наличию соответственных нулей.

Кривая 4 иллюстрирует эффект демпфирования при уменьшении сопротивления голосовой щели до 0,33рс. Эта величина характеризует активное входное сопротивление трахеи, согласованно нагруженной в месте ее перехода в бронхиальную систему и легкие, если считать площадь поперечного сечения трахеи равной 3 см² и голосовую щель широко открытой. При таком большом демпфировании последние следы F_1 и F_2 пропадают, а нуль, расположенный чуть ниже F_2 , становится едва различимым. По всей вероятности, такое демпфирование превышает действительную величину в естественной речи, особенно для частот выше 1000 гц. Как следует из кривой 5, подобное же влияние на спектр оказывает десятикратное увеличение сопротивления сужения при сохранении сопротивления голосовой щели равным 5рс.

Наконец, кривая 6 показывает, что смещение источника к зубам при сохранении всех условий, принятых для кривой 3, за исключением последовательно включенного сопротивления зубного прохода, равного 0,067рс, ведет к появлению нулей между первыми тремя полюсами. Соответствующие этим нулям и полюсам антирезонансы и форманты заметно демпфируются сопротивлением источника. Уровни спектральных составляющих, лежащих выше главного пика, расположенного на частоте 1900 гц, более или менее одинаковы до пика на частоте 6000 гц. Этот пик является основным резонансом полости, находящейся перед зубами, т. е. образованной межгубным проходом.

Повышение уровней в области спектра, лежащей выше 2000 гц, которое наблюдается при очень продвинутом вперед расположении

источника. обуславливается появлением комплексно сопряженных нулей на низких частотах, одного на 700 гц и второго на 1300 гц, вместо одного нуля на 950 гц, наблюдаемого при расположении источника в месте язычного сужения. Используя эквивалентную электрическую схему, можно показать, что перенос источника от заднего к переднему сужению, ограничивающему изолированную полость, ведет к небольшому сдвигу частот нулей и к введению дополнительного нуля на частоте резонанса, определяющегося объемом полости и индуктивностью сужения, расположенного позади полости. В примере, рассмотренном выше, этот нуль лежит на частоте 700 гц. Он вводит в фильтровую функцию дополнительный множитель $1 - \frac{f^2 - jfB_z}{F_z^2}$, где F_z — частота нуля, B_z — ширина его полосы (см. формулу (3.12)).

На частоте f , превышающей F_z на 40%, этот множитель становится почти равным единице, а на высоких частотах дает подъем огибающей спектра с наклоном, приближающимся к +12 дб/окт. Таким образом, влияние этого нуля противоположно влиянию единичного полюса, рассмотренного в § 3 гл. 3.

Из графика 2 рис. 62 можно видеть, что кривая, полученная для естественного произнесенного диктором звука [š] (штриховая кривая), ближе по общему ходу к расчетной кривой, полученной при расположении сужения в области языка (сплошная линия), чем к кривой, соответствующей расположению источника в зубном проходе (пунктирная линия). В обоих случаях применялся источник с равномерным спектром. Огибающая мгновенного спектра для естественных звуков, взятая из произнесенных испытуемым односложных слов (спектрограммы их приведены в Приложениях), в области частот 300—9000 гц отличается на величину до 12 дб от расчетной для случая расположения источника в сужении; на частотах ниже 800 гц огибающая спектра естественного звука ниже расчетной на постоянную величину 8—10 дб.

В области низких частот лучшее соответствие спектров получается при расположении источника в зубном проходе; однако для того, чтобы имело место лучшее соответствие для частот выше главного пика, огибающая спектра источника должна была бы падать на 12 дб/окт.

Нет сомнения, что главный пик на 1650 гц, наблюдаемый в естественном звуке, представляет собой F_3 и соответствует пику в области 1900 гц на расчетном спектре. Полученная при расчете частота $F_2 = 1050$ гц достаточно близка к $F_2 = 1150$ гц в естественном звуке. Очень быстрое падение уровней спектра от 1050 гц до 900 гц наблюдается на обеих кривых. Возможно, что в образовании естественного [š] участвовало два источника, из которых один расположен в зубном проходе, а другой, со сравнительно малыми уровнями спектра на частотах ниже 1000 гц, — в язычном. Имеющийся в естественном звуке нуль на частоте 750 гц, возможно, обусловлен наличием значительной связи полости рта с трахеей.

Интересно отметить, что спектр, рассчитанный нами для [š], и спектр [š] в слове «shack», полученный Юзом и Халле [105], в значительной мере сходны.

Расчет спектра согласного [x] является, пожалуй, самой успешной из всех наших попыток восстановления спектральных характеристик звуков в связной речи на основании полученных для того же лица рентгенограмм изолированных, протяжно произносимых звуков. Кроме небольшого провала на 750 гц и связанного с ним небольшого плато на

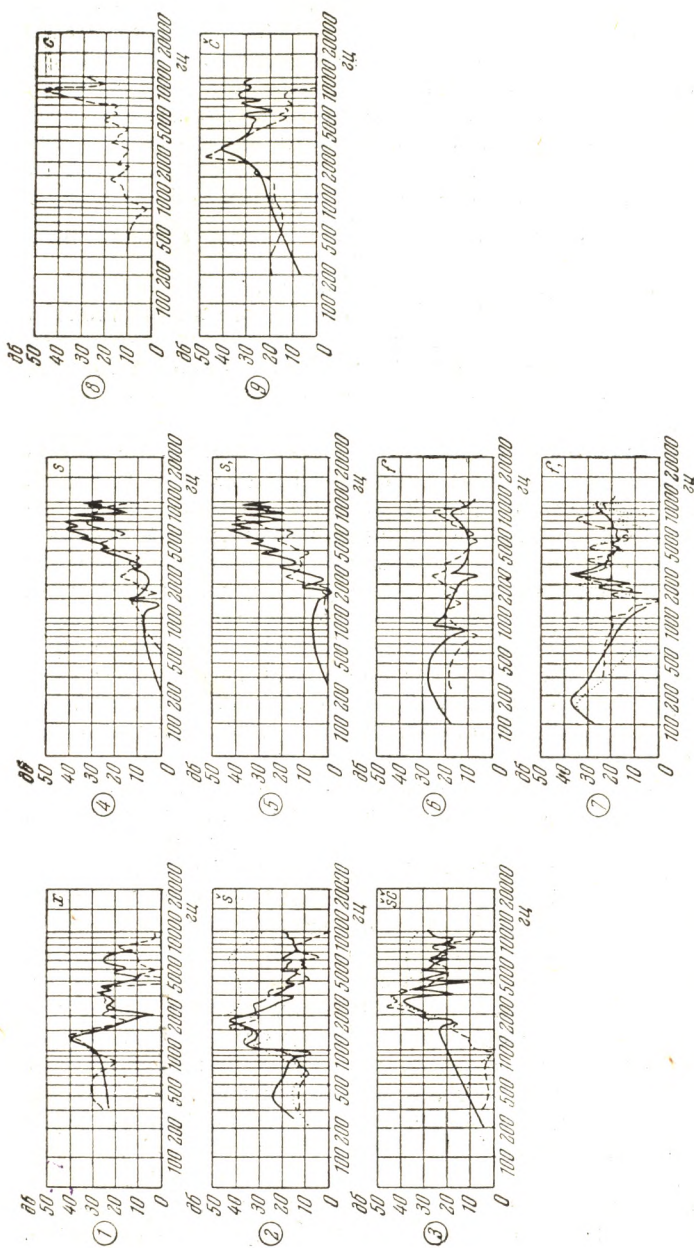


Рис. 62. Спектры фрикативных и аффрикат, расчетные (сплошные и пунктирные кривые) и измеренные (штриховые кривые), при произнесении диктором в связной речи. Сопротивление толстовой щели при расчетах принято равным 5 π с. Источник — в середине артикуляторного сужения, если особо не оговорено иное его местоположение.

1) [x]: сплошная кривая: источник с равномерным спектром и внутренним сопротивлением 0,25 π с в язычном проходе.
2) [ʃ]: то же, что (2), но сопротивление источника 2,5 π с.

3) [ʂ]: источник с равномерным спектром и внутренним сопротивлением 0,25 π с у зубов, сопротивление язычного прохода 2 π с.

4) [s]: источник с равномерным спектром и внутренним сопротивлением 0,25 π с и сопротивлением 0,25 π с.

5) [z]: условие те же, что в случае (4).
6) [ʒ]: губно-зубной источник с наклонно огибающей спектра — 6 дБ/окт и сопротивлением 0,25 π с; пунктирная кривая: губно-зубной источник с наклонно огибающей спектра — 6 дБ/окт и сопротивлением 0,25 π с; сплошная кривая: губно-зубной источник с наклонно огибающей спектра — 6 дБ/окт, сопротивление губно-зубного прохода такое же, как в предыдущем случае.

8) Измеренный спектр зубной аффрикаты [tʃ].

9) [tʃ]: источник с равномерным спектром и сопротивлением 2,5 π с в язычном проходе.

450 гц, что может быть объяснено наличием конечной связи с трахеей, приводящей к повышению F_1 и появлению нуля, для этого звука имеет место большое сходство в расположении пиков и провалов на расчетном и действительном спектрах. Основной пик на частоте 1300 гц соответствует резонансу полости, лежащей впереди от места артикуляторного сужения, и представляет собою F_2 . Нули, имеющие в расчетных спектрах частоты 2000 гц и 4200 гц, соответствуют нулям 1800 гц и 4050 гц в естественном звуке. Полученное сходство подтверждает правильность принятого при расчете предположения о том, что источник, расположенный в области сужения, является основным.

При расчете был принят шумовой спектр источника со спаданием уровня огибающей, равным 6 дб/окт. Увеличение наклона огибающей еще на 6 дб/окт для частот выше 4000 гц дало бы, вероятно, еще лучшее соответствие кривых. Сопротивление источника полностью демпфирует F_1 и компенсирует недостаточность затухания, вносимого стандартным в наших расчетах сопротивлением голосовой щели, которое для данного случая, пожалуй, слишком велико и не обеспечивает достаточной активной нагрузки тракта.

Спектрограмма слова [ха], приведенная в Приложении, заслуживает некоторых замечаний. На ней видны вертикальные штрихи, отчетливо ограниченные во времени, с интервалами 30—50 мсек; они могут быть приписаны либо колебаниям ложных голосовых связок, либо вибрациям мягкого нёба в месте сужения. Последнее объяснение кажется более вероятным ввиду высокого положения F_1 . Механизм таких колебаний можно представить себе как нерегулярно повторяющиеся полные смычки и следующие за ними прорывы воздуха с появлением восстанавливающей силы, обусловленной эффектом Бернулли. В результате возникает нерегулярная модуляция выдыхаемого воздуха и, как следствие, колебательная сила, действующая на заднюю часть мягкого нёба и язычок; в предельном случае этот процесс может привести к возникновению увулярной трели.

С другой стороны, такие же или подобные им явления кажущегося «квантования» звуковой энергии часто наблюдаются в шумовом интервале велярных, палатальных¹⁾ и даже губных звуков. Эти кванты энергии, по-видимому, обусловлены в большей мере толчками прорывающихся порций воздуха, а не эффектами, связанными с турбулентностью. Огибающая спектра источника со спаданием 6 дб/окт, принятая в расчете, как раз соответствует спектру ступенчатой функции, характеризующей внезапное включение потока воздуха. Записи изменения интенсивности во времени для слога [ха] показывают, что такими «квантами» модулированы главным образом составляющие спектра, лежащие ниже 3000 гц. В интервалах между «квантами» имеет место шум меньшей интенсивности, обусловленный источником с менее крутым спаданием огибающей спектра. Теоретически не исключена возможность, что появление подобных звуковых квант вызывается модуляцией потока воздуха колебаниями ложных голосовых связок. Однако до сих пор все эти явления еще не имеют удовлетворительного объяснения.

В отношении расчетных кривых для звука [šč] (см. график 3 рис. 62) остаются в силе те замечания, которые были сделаны при рассмотрении звука [š]; здесь более вероятным, чем для [š], является участие зубного

¹⁾ Двойной взрыв у [k] наблюдала Фишер-Йоргенсен [57].

источника. В пользу этого говорит наличие нуля у 1000 гц, хорошее совпадение кривых ниже основного пика на 2700 гц и возможность использования при расчете меньшего наклона огибающей спектра для согласования расчетного спектра со спектром естественного звука в области частот выше основного пика. Некоторые остающиеся после этого различия в уровнях высоких частот могут быть отнесены на счет более низкого частотного положения главного пика в естественном звуке; обоснованность такого предположения подкрепляется тем, что в спектре естественного звука F_3 имеет частоту 2400 гц и F_4 — частоту 3000 гц по сравнению с $F_3 = 2700$ гц и $F_4 = 3300$ гц в расчетном спектре.

Основное акустическое различие между [šć] и [š] заключается в более высоком частотном положении основного пика в спектре [šć], что свидетельствует о малых размерах передней полости. Это относится и к случаю сравнения [š] и [x]. В спектре [x] доминирующей является F_2 , тогда как в [š] F_3 несколько более выражена, чем F_2 , а в [šć] основной формантой является F_3 . Уровень следующей за ней форманты F_4 меньше уровня F_3 всего на 3 дб.

В спектре аффрикаты [č], снятом для середины интервала фрикативного шума, также доминирует F_3 ; и здесь следующей по интенсивности является F_4 , имеющая уровень на 10 дб более низкий, чем F_3 ; это видно яснее на спектрограмме. Поэтому для [č] зубной источник менее вероятен. Остатки F_2 в шумовом спектре [č] и [šć] не являются существенными ввиду их низкого уровня; однако более высокая частота F_2 по сравнению с F_2 для [š] и [x] проявляется в более отчетливо выраженном положительном переходе, видимом на спектрограмме.

При расчетах для непалатализованного [f] был использован губно-зубной источник со спаданием огибающей в 6 дб/окт. Как расчетный спектр, так и спектр естественного звука в общем сглаженные, однако и в том и в другом имеется пик при 8500 гц. Остаточная формантная структура в спектре естественного звука усложнена добавочной формантой на 1500 гц, которая, возможно, обусловлена сильной связью с трахеальной системой. Ноль в спектре естественного звука на частоте 750 гц соответствует, вероятно, первому сопряженному нулю на 800 гц в расчетном спектре.

Использованный для сравнения спектр естественного звука соответствует спектральному разрезу II спектрограммы, приведенной в Приложении. Следует отметить, что F_2 в спектре разреза III, соответствующего моменту непосредственно перед началом гласного, расположена выше, чем начальная точка перехода F_2 в пределах гласного. Это представляет собой типичный эффект прекращения связи с трахеей в начале фонации и согласуется с тем, что в звуке [v] нашего испытуемого F_2 на 200 гц ниже, чем в [f]. Те форманты, которые в существенной мере определяются фарингальной полостью, всегда несколько выше по частоте при открытой голосовой щели. Здесь огубление в комбинации с фарингализацией создает условия, обеспечивающие необходимую зависимость F_2 от задней полости (см. § 3 гл. 4 и § 2 гл. 7).

На графике 7 рис. 62 приведены: расчетная кривая для [f], полученная при губно-зубном источнике, со спаданием огибающей спектра в 6 дб/окт; спектр естественного звука; расчетная кривая, полученная при расположении источника в гортани, причем спектр источника был принят таким же, как для первой кривой. Спектр естественного звука и расчетный спектр для губно-зубного источника одинаковы по своему общему характеру. Уровень основного пика на частоте 2400 гц, соот-

ветствующего F_3 , на 14 дБ выше уровня F_2 в естественном спектре и на 13 дБ в расчетном. Между частотами F_1 и F_2 находится хорошо выраженный антирезонанс, обусловленный первым сопряженным нулем с частотой 1500 гц в расчетном спектре и 1300 гц в спектре естественного звука. Влияние следующего нуля, лежащего чуть выше F_2 , невелико; в силу близости нуля к F_2 он компенсируется полюсом F_2 .

Еще более определенно нейтрализация проявляется в расчетном спектре для тех полюсов и нулей, которые лежат выше F_3 . Однако в спектре естественного звука отчетливо видны и F_4 и F_5 , что указывает на наличие добавочного источника, который расположен так, что уменьшает взаимное перекрытие полюсов и нулей этих формант. Вторичный источник в области небного сужения может возникать в сильно палатализованных губных, если губы раскрываются скорее, чем понижается язык. Как видно из спектрограммы, приведенной здесь, спектральный разрез естественного звука [f.] относится к моменту, когда начинается аспирация, накладывающаяся на фриктивный шум; это типично для конечного участка глухого согласного в начальной позиции. Наличие дополнительного источника в голосовой щели весьма вероятно ввиду хорошего совпадения F_3 , F_4 и F_5 естественного звука с расчетными данными, полученными для случая расположения источника в голосовой щели. Основная форманта фриктивного интервала для [f.] лежит на частоте 7500 гц и связана с небольшой полостью впереди верхних передних зубов, в комбинации с импедансом излучения. В расчетных спектрах эта форманта лежит на частоте 9000 гц. Пик на 8500 гц в спектре [f.] имеет то же происхождение.

На графиках 4 и 5 представлены огибающие спектров соответственно для [s] и [s.]. При их определении был использован источник с равномерным спектром, расположенный при координате, соответствующей краю верхних передних зубов. Как для расчетных кривых, так и для спектра естественного звука типичным является их сходство с характеристикой высокочастотного фильтра с граничной частотой 2500 гц и крутизной среза 15—25 дБ/окт. Крутизна эта больше для расчетных кривых, чем для спектров естественных звуков. Максимум спектра естественных звуков лежит в области 6000—9000 гц, а в расчетных спектрах примерно на 20% ниже.

Как будет показано более подробно дальше, типичная форма спектра [s] определяется полуволновым резонансом в проходе, образованном языком и альвеолярными выступами. Этот проход можно рассматривать как трубку с эффективной длиной $l = 2,5$ см. Поэтому основные форманты [s] должны иметь частоту полуволнового резонанса этой трубки $\frac{c}{2l} = 7000$ гц, далее частоту $\frac{2c}{2l} = 14\,000$ гц, соответствующую следующей моде колебаний, и дополнительно частоту резонанса губной полости, расположенной перед зубами. Характеристика типа высокочастотного фильтра подчеркивается еще нулем, находящимся на частоте 3000 гц и отвечающим четвертьволновому резонансу сужения, расположенного позади источника.

Детали формантной структуры естественного звука, налагающиеся на основную огибающую спектра, не очень хорошо воспроизводятся расчетным спектром, за исключением более высокого положения F_2 и F_3 в палатализованном [s.]. Значительно ослабленная F_2 и сильно выраженная F_3 в спектре естественного [s.] заставляет предполагать наличие источника, расположенного у входа в альвеолярный канал со-

стороны твердого нёба. Существенных различий между спектрами [s] и [s₁] по общей форме нет.

Основной пик в спектре аффрикаты [c] также расположен в области 7000—8000 гц, но он более узок, чем в [s]; это говорит о более простой

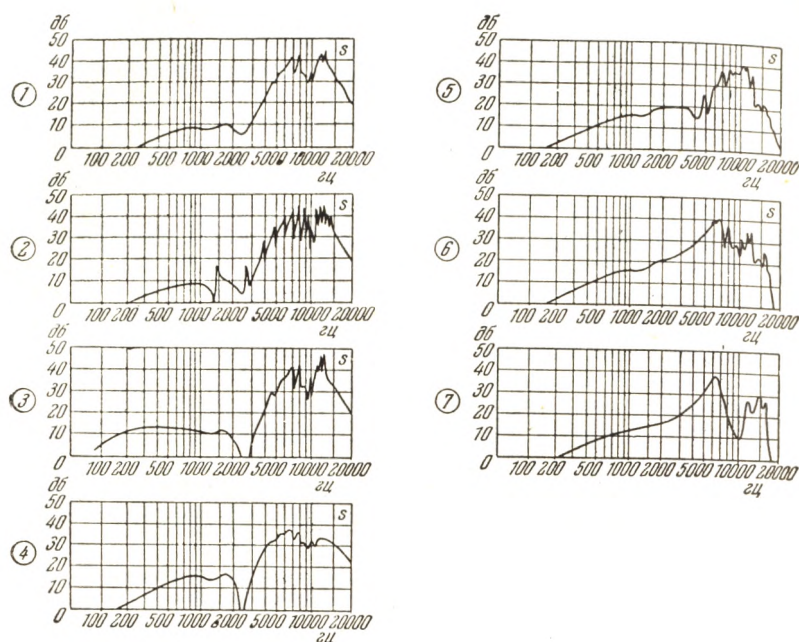


Рис. 63. Влияние на расчетный спектр [s] демпфирования голосовой щели, величины сопротивления источника и полостей позади источника. Спектр источника равномерный. При расчете, производившемся на электрическом аналоге LEA, были установлены следующие значения площади поперечного сечения последовательных звеньев, считая от излучающей поверхности у губ:

№ звена	1	2	3	4	5	6
Площадь в см ²	6,5	0,65	0,16	0,16	0,16	0,32

Эти значения были приняты как приближение к виду функции площади на переднем ее участке; остальные звенья аналога были установлены в соответствии с видом функции площади для [s], приведенной на рис. 59.

- 1) Источник с сопротивлением 0,025рс расположен перед звеном 2, т. е. в месте, соответствующем зубам: сопротивление голосовой щели 0,33рс; между звеньями 4 и 5 введено сопротивление 2рс; эти условия были приняты для расчета [s], соответствующего графику 4 рис. 62.
- 2) Сопротивление голосовой щели устранено, в остальном как в (1).
- 3) Сопротивление голосовой щели, как в (1); исключено последовательное сопротивление 2рс.
- 4) Сопротивление источника у зубов увеличено до 2рс, в остальном как в (3).
- 5) Источник перед звеном 3, в остальном как в (4).
- 6) Источник перед звеном 5, в остальном как в (4).
- 7) Позади звена 6 — короткое замыкание; в остальном как в (6).

структуре системы, типа простого резонатора. Расчеты для [c] не производились.

При расчете спектров [s] были выполнены некоторые систематические вариации условий синтеза. Конфигурация полости на протяжении

первых 3 см системы, отображаемая в параметрах первых шести звеньев аналога LEA, была следующей:

№ звена	1	2	3	4	5	6
Площадь, см ² . . .	6,5	0,65	0,16	0,16	0,16	0,32

Размеры остальной части речевого тракта были взяты в соответствии с данными рентгенограммы.

Как видно из рис. 63, налагающаяся формантная структура может быть подавлена, если включить в сужение большое последовательное сопротивление или сделать сопротивление голосовой щели малым. Результатом введения большого последовательного сопротивления в области зубов является уменьшение разницы в уровнях частей спектра, лежащих выше и ниже граничной частоты 2500 гц.

Такое же выравнивающее влияние, но в более сильной степени, оказывает перемещение источника от зубов к внутреннему входу в язычный проход. Промежуточное положение источника¹⁾ приводит к менее быстрому, чем это обычно имеет место, подъему огибающей спектра [s] в области от 3000 до 6000 гц, как это можно видеть, сравнивая график 5 с графиками 4 и 6.

Если нагрузку язычного прохода полостями, расположенными позади него, заменить коротким замыканием, больших изменений спектра не происходит, если не считать появления нуля на частоте 10 000 гц и полного исчезновения формантной структуры огибающей спектра.

Влияние некоторых систематических изменений конфигурации системы резонаторов, участвующих в образовании [s], показано на рис. 64. Общая длина модели равнялась 3 см, источник располагался на расстоянии $x = 0,5$ см от излучающего конца. Сопротивление источника было равно 0,025 рс. Задний конец системы был коротко замкнут.

Из сравнения графиков 1 и 2 видно, что увеличение площади поперечного сечения сужения с 0,16 до 0,32 см² ведет к сдвигу первого нуля с 2500 до 3000 гц, а один из полюсов в области 4000—8000 гц смещается вниз. Расширение передней части прохода вызывает дальнейший сдвиг вниз одного из этих полюсов и первого нуля.

Графики 4, 5 и 6 показывают, что сужение первого звена длиной 0,5 см, входящего в систему губного резонатора, не вызывает заметных сдвигов нуля, но существенно понижает частоту полюса, отвечающего полости, расположенной спереди от сужения. Интересно, что уменьшение степени открытия губ с 6,5 см² до 0,16 см² сдвигает этот полюс с 8000 гц не более чем до 3500 гц. Основные спектральные характеристики при огублении сохраняются, если не считать повышения частоты резонанса губно-зубной полости. Однако расчетным данным для частот выше 10 000 гц нельзя придавать большое значение ввиду ограниченной приложимости на этих частотах одномерной трактовки распространения звука в речевом тракте.

Палатограммы звука [s] ([35], стр. 187; [123]) показывают большую изменчивость ширины и формы канала между языком и зубами у различных лиц и в разных языках. Наряду с конфигурацией речевого тракта при произнесении [s], приближающейся к отрезку трубы (такая

¹⁾ Согласно Майер-Эпплеру [150], эффективное положение источника должно быть у места наибольшего сужения и не обязательно у края верхних резцов. Это утверждение не опровергается настоящими исследованиями, но все же положение источника у края зубов представляется более вероятным.

конфигурация и была принята в наших расчетах), можно встретить формы артикуляции, при которых речевой тракт правильнее рассматривать как трубу с сужением в середине¹⁾. Влияние сужения в центральной

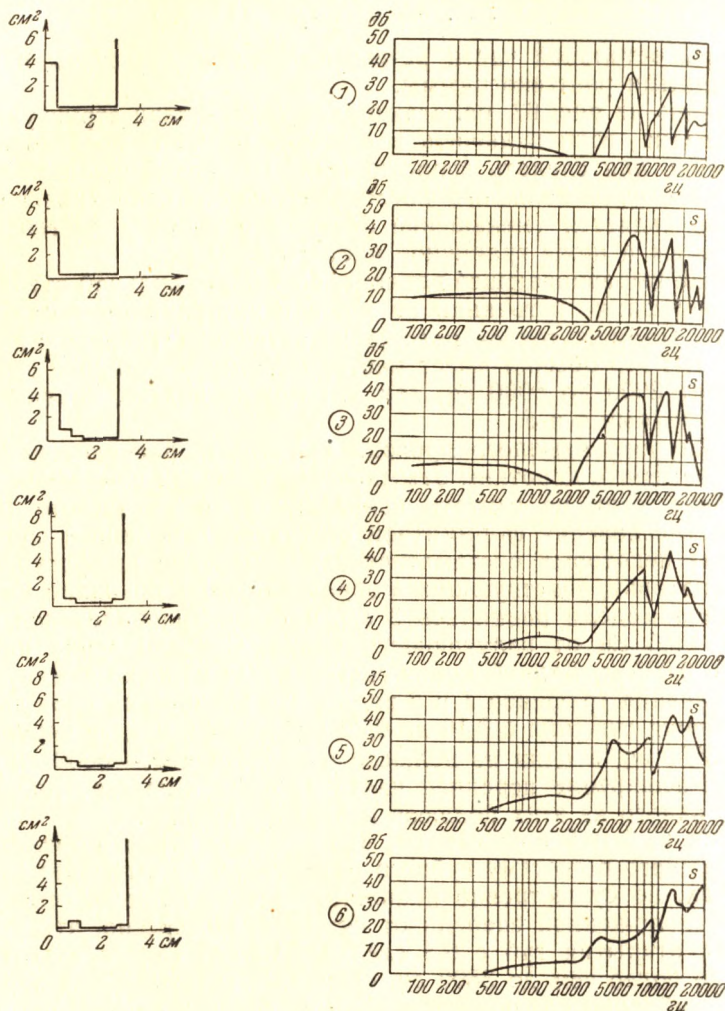


Рис. 64. Сильно упрощенные модели передней полости, соответствующей артикуляции [s], и получающиеся на таких моделях спектры. Влияние задних полостей представлено коротким замыканием в месте, соответствующем первым 3 см модели. Минимальная площадь поперечного сечения для модели (2) составляет 0,32 см², для остальных — 0,16 см². Источник с равномерным спектром расположен в точке с координатой $x=0,5$ см, отсчитываемой от переднего конца модели; сопротивление источника 0,025 рс.

части трубы длиной l , открытой с обоих концов, не очень велико. Основной резонанс трубы без сужения, имеющий место на частоте $f = \frac{c}{2l}$, сдвигается за счет сужения незначительно, а следующий ре-

¹⁾ Согласно Свенд Смит (личное сообщение), место максимального сужения обычно находится у «Papilla incisiva palati».

зонанс этой трубы на частоте $f = \frac{2c}{2l}$ сдвигается вниз вплотную к $\frac{c}{2l}$.

Если основное сужение имеет длину порядка 1,5 см, оно вызывает приблизительно такие же изменения, как показанные на графиках 4 или 5 рис. 64.

§ 2. Взрывные

Для спектрального анализа щелевых и аффрикат использовались спектральные разрезы, относящиеся к интервалу максимальной интенсивности, который наблюдается в середине или ближе к концу звука. Спектральные разрезы взрывных звуков были взяты также для интервала максимальной интенсивности, который, вообще говоря, приходится на начало глухого участка звука, отождествляемого со взрывом. У палатализованных зубных имеется сильно выраженная тенденция к аффрикации, которая ускоряет наступление момента максимальной интенсивности.

Участки, для которых определялись спектральные разрезы, можно классифицировать согласно терминологии, введенной в § 1 гл. 1, следующим образом: 0 — звонкая смычка, I — начальный переходный процесс при взрыве, II — интервалы фрикативного звука, III — интервалы аспирации или аспирации, смешанной с фрикацией.

Все расчеты и опытные данные (см. Приложения), приводимые здесь, относятся либо к интервалу I или II, либо к их комбинации в том случае, когда звуки очень коротки или когда разделить участки I и II вообще невозможно.

Расчеты были выполнены только для глухих взрывных. Сравнивая контуры речевого тракта для этих звуков, приведенные на рис. 65, а также соответственные кривые для функции площади (рис. 66), с данными рис. 58 и 59, можно видеть большое сходство между [k] и [x], [k.] и [j], [p] и [f], [p] и [f.], [t] и [s] и [t.] и [s.].

Все рентгенограммы для рассматриваемых звуков были сняты в момент полной смычки перед взрывом, но при проведении расчетов было принято, что в этот момент сохраняется узкий проход с очень малой площадью поперечного сечения 0,16 см². Выбор того или иного значения этой площади не очень критичен для общей формы получаемых спектров.

Для расчетов взрывных был принят источник со спаданием огибающей спектра в 6 дб/окт и внутренним сопротивлением 0,25 рс за исключением случая звука [t.], который рассматривался как щелевой, поэтому

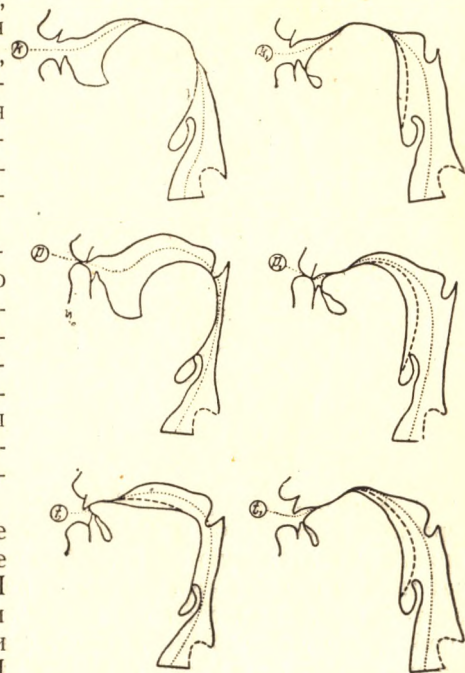


Рис. 65. Контуры речевого тракта, построенные по рентгенограммам для [k], [p], [t], [k.], [p.], [t.]

в этом случае был использован источник с равномерным спектром и внутренним сопротивлением 2рс . За недостатком данных об импедансе голосовой щели не делалось попыток имитировать условия открытой голосовой щели, и для всех расчетов было принято стандартное значение ее сопротивления, равное 5рс , соответствующее ее полузакрытому состоянию при фонации. Такое сопротивление не обеспечивает достаточного демпфирования низкочастотных формант и в большей мере соответствует условиям для звонких взрывных. С другой стороны, активная согласованная нагрузка гортанной трубки, равная $0,33\text{рс}$, приводит

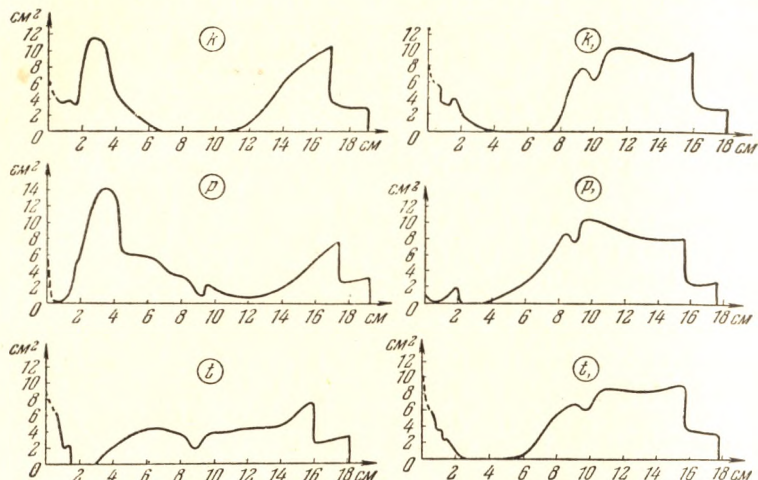


Рис. 66. Функции площади для [k], [p], [t] и [k:], [p:], [t:].

к слишком большому демпфированию всей области частот выше первой форманты; вероятно, в действительности импеданс голосовой щели содержит значительную реактивную составляющую. До тех пор пока просвет между голосовыми связками не достигает максимума, импеданс голосовой щели является по преимуществу индуктивным; это приводит к небольшому повышению частот формант, как это было показано выше для щелевых. Для учета влияния активных элементов, включенных последовательно с индуктивностью голосовой щели, они должны быть, как показано в § 3 гл. 7, преобразованы в параллельное сопротивление.

Предполагаемое большое сопротивление источника для случаев щелевых, аффрикат и аффрицированных взрывных представляет собой, по-видимому, наибольшее возможное значение; это следует из формул (II.10) и (II.14), если принять большую скорость $v = 6000 \text{ см/сек}$ потока воздуха, протекающего через узкий проход с площадью поперечного сечения $A = 0,08 \text{ см}^2$. Для взрывных можно ожидать большой изменчивости сопротивления источника, так как в широких пределах изменяются размеры сужения и условия протекания воздуха. Значения, принятые для расчета, дают вполне приемлемые спектры, причем значения эти не очень критичны для формы спектров; дальнейшее рассмотрение этого вопроса см. в § 4 гл. 10.

Расчеты спектров [p] и [p:] были выполнены для источника, расположенного на расстоянии $0,5 \text{ см}$ от переднего конца тракта, смоделированного на аналоге. Для [t] и [t:] источник располагался при координ-

нате $x = 1$ см, а для [к] и [к.] — в середине язычного сужения. Небольшое различие в принятых для расчета положениях источника для губных и зубных не имеет существенного значения.

Расчетные и естественные спектры для глухих взрывных приведены на рис. 67, где представлены также спектры некоторых звонких взрывных. Общее соответствие основной формы расчетных и естественных спектров и средних спектральных уровней на низких, средних и высоких частотах позволяет считать, что характер огибающей спектра источника был выбран правильно.

Совпадение расчетного и естественного спектров [к] на графике 1 настолько хорошее, что каждый пик на спектре естественного звука может быть отождествлен с соответствующим пиком на расчетной кривой. В области частот, лежащих ниже нуля, расположенного на частоте 5000 гц, расчетная кривая дает частоты максимумов, превышающие соответственные частоты естественного спектра примерно на 20%. Это значит, что естественный звук был артикулирован с несколько более продвинутым вперед положением языка. Что касается расчетной кривой, то главный пик на частоте 1200 гц является основной модой ротовой полости и идентичен F_2 .

Следующая, более высокая мода колебаний в ротовой полости соответствует в F -картине форманте F_3 . При частоте 3100 гц на расчетной кривой находится нуль, соответствующий первому после $f = 0$ значению частоты, при котором импеданс по направлению от источника к голосовой щели обращается в бесконечность. Этот нуль, так же как и нуль на частоте 4900 гц, уменьшает уровни четвертой и пятой формант, обусловленных фарингальной полостью и гортанной трубкой. Более высокая формантная область, между 7000 и 9000 гц, связана с третьим и четвертым резонансами ротовой полости. В естественном звуке [к] первый сопряженный нуль находится на частоте, близкой к F_3 . Поэтому второй резонанс ротовой полости представлен формантой F_4 .

Основной пик палатализованного [к.] охватывает F_3 и F_4 . Вторая форманта нейтрализована нулем, что не очень ясно видно в расчетном спектре. Ближайший более высокий нуль имеет частоту 4900 гц.

За исключением несколько более отчетливо заметных следов остаточной формантной структуры в естественных спектрах [р] и [р.] они по распределению энергии в спектре в основном сходны с расчетными. Лежащий на частоте 1300 гц нуль в спектре [р.] представляет собой хорошо выраженный антирезонанс, разделяющий спектр звука на нижнюю и верхнюю части, подобно тому как это имеет место для звука [f]. Одной из основных причин слабой выраженности формантной структуры в расчетном спектре [р.] является весьма малая величина площади поперечного сечения, принятая при расчете для вторичного сужения у твердого нёба; вследствие этого полюсы и нули полости фаринкса сближаются друг с другом. Аналогично причиной отсутствия формантной структуры в области частот выше F_2 в расчетном спектре [р] может быть слишком малая площадь сужения или слишком большое сопротивление источника.

Двойной пик $F_2 + F_3$ палатализованных губных часто наблюдается в виде кажущейся концентрации энергии в центральной части спектрограммы, в результате чего лабиализованный интервал имеет характер нёбного звука. Примеры такого рода приводились Халле, Хьюзом и Редли [86]. Однако отличить спектры губных можно по нулю, расположенному чуть ниже F_2 , и по более сильно выраженной области низких

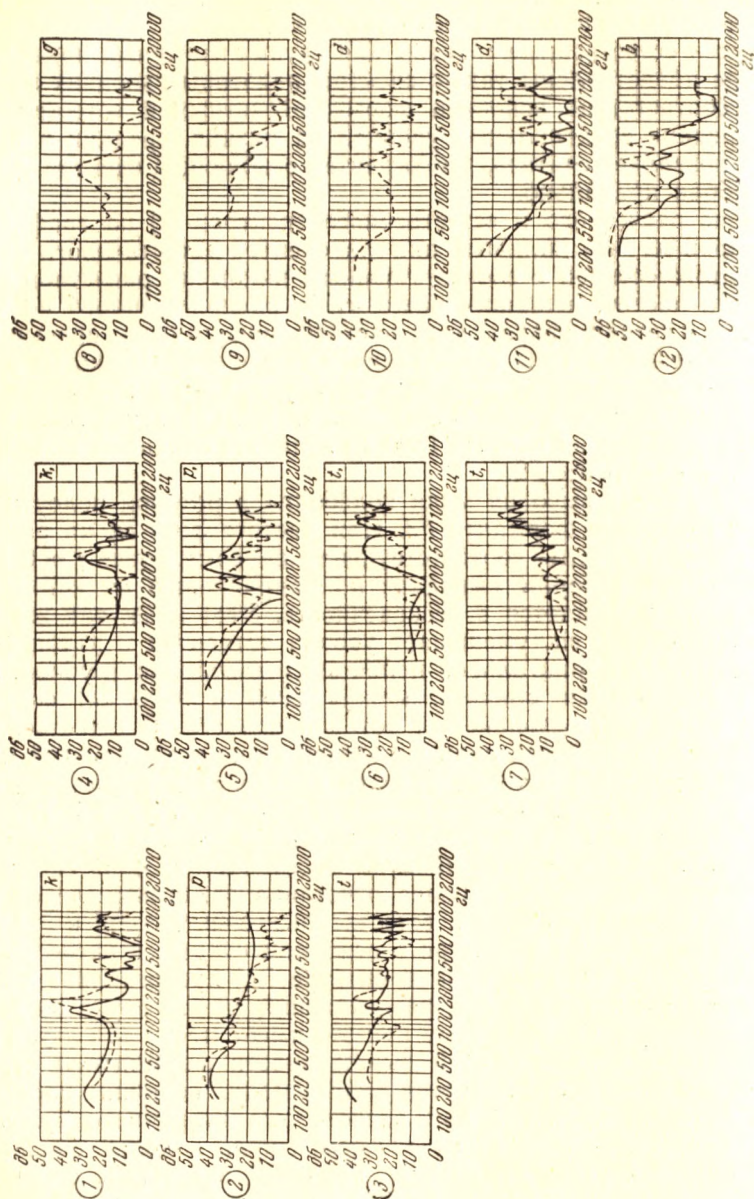


Рис. 67. Расчетные и измеренные спектры для [k], [p], [t], [k'], [p'], [t'] и измеренные спектры для [g], [b], [d] и [g'], [b'], [d']. Во всех расчетах был принят источник с огибающей спектра — 6 дБ/окт и внутренним сопротивлением 0,25 рс, за исключением [t'] для которого спектр источника был равномерный и сопротивление источника 2рс, т. е. источник был такой же, как для фрикативных. Момент снятия спектрального разреза — первые 10 мсек после взрыва, за исключением [t'] (графики 6 и 7) и [d'] (график 11), где измерения относятся к фрикативному интервалу, следующему через 20 мсек после взрыва. Пунктирная кривая для [b], относится к моменту времени 10 мсек после взрыва, т. е. к первому периоду основного тона голоса в последующем гласном.

частот. У губных упомянутые два пика не так близки, как у нёбных, и их общая энергия меньше. Кроме того, отличительные признаки содер-
жятся, конечно, и в переходных процессах.

Расчетный спектр [t] не очень хорошо совпадает со спектром естественного звука, будучи сходным с ним только по общему характеру распределения энергии в спектре. В спектре естественного звука первый нуль на 850 *гц* и слабая форманта на частоте 1100 *гц*, лежащие ниже F_2 с частотой, равной 1600 *гц*, связаны с наличием добавочной системы резонаторов, а именно либо носовой полости, либо, что более правдоподобно, трахеи, влияние которой было отмечено при анализе фрикативных. Эти детали спектра не случайны; они отмечались и в более ранних исследованиях шведских согласных [46]. На графике 6 рис. 67 показаны результаты первой попытки расчета спектра [t.] на основе функции площади, представленной на рис. 66. Если удвоить площадь задней части нёбного сужения, сохранив минимальные ее размеры на протяжении 2 *см* передней части сужения, то совпадение кривых становится значительно лучше, как видно на графике 7. По-видимому, это в большей мере соответствует артикуляторной конфигурации естественного звука.

Как видно из рис. 95, спектральные разрезы были получены для трех участков звука [t.] — одного в интервале взрыва, другого в фрикативном интервале, начинающемся на 20 *мсек* позднее, и последнего — в смешанной фрикативно-аспирированной шумовой части звука. За основу сравнения с расчетными данными был взят второй интервал. Все три спектральных разреза характеризуются высоким уровнем в области 6000—9000 *гц*. Огибающая спектра для первого интервала в среднем горизонтальна, т. е. спектральный уровень постоянный; заметны только весьма слабые следы формантной структуры, если не считать F_2 на 1900 *гц* и дополнительной форманты на частоте 1300 *гц*. В фрикативном интервале наблюдается в среднем подъем огибающей спектра на 6 *дб/окт* вплоть до частоты основного пика при 7500 *гц*. Наконец, в третьем интервале огибающая спектра вновь становится в общем горизонтальной, вследствие подъема уровней в области ниже 5000 *гц* и некоторого ослабления высоких частот. В этом интервале четко выявляется формантная структура, характеризующаяся наличием в среднем одного пика на каждые 1000 *гц* вплоть до частоты 8000 *гц*. Такая характеристика, типичная для «аспираторного» интервала, объясняется, как указывалось раньше, более открытой артикуляцией, которая приводит к более тесной связи между различными частями речевого тракта, а также возможным появлением одного или нескольких добавочных шумовых источников в речевом тракте, вероятнее всего у голосовых связок.

Повышение интегральной интенсивности, начинающееся в момент взрыва и достигающее максимума в середине фрикативного интервала, является типичным для сильно и напряженно артикулируемых зубных взрывных, но редко встречается в нёбных напряженных взрывных. Это, по-видимому, объясняется сравнительно медленным движением языка от смычки к оптимальному раскрытию. Такой же характер повышения интенсивности наблюдался для данного испытуемого в интервалах взрыва и фрикации звука [k].

Очень важный вопрос, поднятый Халле и др. [86], относится к явлениям быстрого изменения частоты формант. Очевидно, что при очень быстрых изменениях происходит рассредоточение энергии в спектре, проявляющееся в виде расширения, понижения уровня и даже

расщепления пиков; это явление выражено тем сильнее, чем больше длительность анализируемого отрезка звука, или тем меньше, чем шире полоса фильтра анализатора. Подобный эффект может быть причиной некоторого выравнивания спектра губных и зубных смычных в первом шумовом интервале. Правда, в последней группе звуков подавление формантной структуры вызывается, возможно, в большей мере активным сопротивлением сужения. У нёбных и велярных при умеренном изменении площади сужения скорость изменения частоты основного резонанса, зависящего от ротовой полости, невелика. Как видно из спектра [k], отклик речевого тракта на начальной стадии переходного процесса характеризуется преобладанием хорошо очерченного пика, который сохраняется для этого звука в течение большей части шумового интервала.

§ 3. Идеализированные модели фрикативных и взрывных

Из сказанного в предыдущем разделе ясно, что существует определенная возможность предсказать спектральный состав акустической картины фрикативных, аффрикат и взрывных звуков по физиологическим данным, характеризующим артикуляцию. Приведенные выше детальные данные полезно дополнить упрощенным рассмотрением, позволяющим выявить роль резонаторов речевого тракта в формировании основных характеристик каждого из трех следующих классов звуков, образующихся при расположении шумового источника в месте артикуляторного сужения или возле него:

а) губных и губно-зубных;

б) зубных;

в) всех других звуков, место артикуляции которых более отодвинуто назад.

Эти три группы будут представлены фонетическими символами [ptk] для взрывных и [fsʃ] для фрикативных звуков. Символ [ʃ] обозначает здесь любой фрикативный, противопоставленный губным и зубным по признаку большего объема передней полости. На упрощенной модели будет рассмотрено также возможное значение изменений местоположения источника звука для различения этих звуков при постоянной конфигурации полости. Акустической проблемой, представляющей особый интерес, является влияние палатализации на спектральный состав шумового интервала в звуках речи.

Попытка максимального упрощения структуры тракта представлена идеализированными моделями, показанными на рис. 68. Полости, расположенные позади артикуляторного сужения, здесь отброшены, однако их влияние также будет учтено в дальнейшем. Следует отметить, что эффективная акустическая граница между системами передней и задней полостей лежит не в центре сужения, а ближе к его заднему концу, в идеальном случае — в месте наиболее резкого изменения площади поперечного сечения.

Поскольку система передней полости должна включать все важнейшие элементы, которые определяют характеристики звуков, необходимо ввести в расчеты полный импеданс сужения, определяющийся его эффективной длиной.

Физический механизм, обуславливающий образование турбулентных звуков как в упрощенной механической модели, так и в естественном речевом тракте, не настолько ясен, чтобы можно было вполне обоснованно указать точное положение и протяженность источника по оси

системы полостей (см. § 2 Приложения II). Как видно из результатов расчета, приведенных в предыдущем разделе, целесообразно, тем не менее, относить источник к месту образования вихрей, т. е. турбулентной циркуляции, а именно к тем участкам тракта, где площадь изменяется наиболее быстро, по преимуществу в узких проходах. В модели с профилем прямоугольной формы источники могут находиться как у входа, так и у выхода модели.

Точное положение источника в узком канале не имеет существенного значения, если длина канала мала по сравнению с представляющими интерес длинами волн; это справедливо, например, для губного

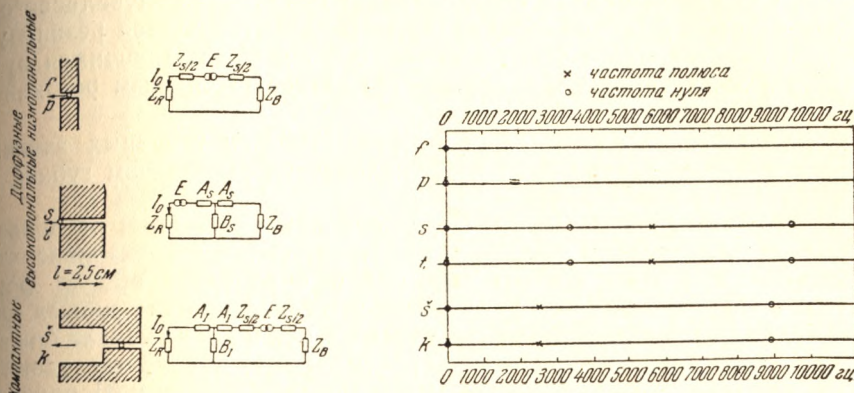


Рис. 68. Сильно упрощенные модели полостей, эквивалентные им схемы и расположение нулей (O) и полюсов (X) для губных, зубных и альвеолярных взрывных и фрикативных. Альвеолярным соответствует более отодвинутое назад место артикуляции и большая, чем у зубных и губных, полость перед ним. Источник в случае зубных и губных расположен у переднего конца модели или вблизи него; небольшое различие в положении источника при этом несущественно. Зубные по сравнению с губными характеризуются наличием резонирующего прохода существенной длины позади источника.

сужения. Аналогично, положение источника в узком, но не обязательно коротком канале, расположенном позади полости достаточно больших размеров, как это встречается в велярных или альвеолярных, не очень существенно, так как форма спектра обуславливается в этом случае главным образом основным резонансом передней полости. Однако ни то, ни другое из этих условий не выполняется в случае зубных звуков. В какой мере расчеты спектров таких звуков, как [s] и [t], могут дать достаточно показательные результаты, зависит в большой мере, хотя и не полностью, от надлежащего выбора положения источника; расположение его у зубов дает наиболее отчетливое представление о характере спектра.

Для большей наглядности расчетов в упрощенной модели не была учтена неглубокая полость, расположенная перед зубами. Влияние на расчетный спектр [s] этой и других деталей конфигурации системы полостей было рассмотрено в предыдущем разделе.

На рис. 68 показаны эквивалентные схемы трех моделей и распределение полюсов и нулей на вещественной оси плоскости комплексной частоты, определяющее фильтровую функцию этих моделей при отсутствии потерь. Специфические свойства каждой из трех моделей следующие:

А. Губные. Спереди или сзади основного сужения нет резонаторов, имеющих существенное значение. Источник звука находится вблизи от центра сужения. В результирующем спектре модели отсутствуют полюсы и нули, т. е. резонансы и антирезонансы в области 0—10 000 гц.

Губные и губно-зубные характеризуются, таким образом, огибающей с равномерным распределением энергии по спектру. Заметим, что такой вывод является чрезмерно упрощенным представлением спектра губно-зубных, так как не учитывает основного резонанса артикуляторного сужения и маленькой полости, образуемой губами перед верхними резцами. Резонанс этой полости может иметь достаточно низкую частоту, около 6000—7000 гц; резонанс губной полости не принимается во внимание ниже и при описании зубных звуков.

Б. Зубные. Впереди от места основного сужения нет резонатора существенных размеров. Большая часть сужения расположена позади источника. Суженный проход обуславливает появление полуволнового резонанса на частоте около 5500 гц. Спектр характеризуется резким увеличением спектрального уровня выше 4000 гц, усугубляющимся нулем на частоте 3500 гц, который обусловлен четвертьволновым резонансом узкого канала позади источника.

Основной спектральный пик на 4000—7000 гц у зубных звуков в действительных условиях речи еще усиливается резонансом губной полости, рассмотренным выше.

В. Велярные, нёбные и др. Перед основным сужением расположен резонатор относительно больших размеров. Основной резонанс соответственной полости главным образом и определяет общую форму спектра: он расположен ниже по частоте, чем главный пик у зубных, даже в том случае, если длина полости такая же, как и длина канала при зубных, поскольку задний конец полости почти закрыт; для зубных задний конец канала можно считать открытым. В моделях для [k] и [t], приведенных на рис. 68, это сводится к различию между четвертьволновым и полуволновым резонансами. Между спектральными характеристиками зубного и передненёбного звуков имеется еще одно различие; оно состоит в том, что главный пик у второго звука в общем расположен близко к частоте, на которой у первого звука находится основной нуль.

Предельно упрощая, можно принять следующие характеристики резонансных систем для разных звуков:

[t] — резонансных частот нет или они лежат очень высоко;

[s] — высокая резонансная частота или, точнее, спектр в форме частотной характеристики высокочастотного фильтра с высокой граничной частотой;

[š] — одна резонансная частота, лежащая ниже, чем для [t] и [s]. Соответственные артикуляторные особенности можно описать так:

[s] и [t] — отсутствие большого резонатора впереди от места сужения;

[š] — артикуляторное сужение, образующее короткую трубу, открытую с обоих концов;

[š] — резонатор относительно большого объема впереди от места артикуляции.

Можно было бы классифицировать рассматриваемые звуки по положению источника, т. е. определять [s] и [t] как более передние, чем [š]. Однако такое определение не вполне однозначно, поскольку различия в размерах полостей сами по себе определяют различие этих звуков. Так, звук [š] может быть образован, по крайней мере на модели, с зубным положением источника; вместе с тем в случае [s] возможно расположение источника в месте максимального сужения прохода между языком и губами.

Рис. 69 иллюстрирует влияние добавления задних полостей на распределение полюсов и нулей. Полная длина от закрытого конца задней полости до излучающего отверстия принята равной 15 см. На рисунке приведены модели с задней полостью как в виде одного отрезка трубы, так и в виде двух отрезков труб, что является имитацией палатализации. Для воспроизведения палатализации в губных передняя часть модели, имитирующей ротовую полость, также принималась в виде двух отрезков труб разных диаметров.

Приведенные на рисунке распределения полюсов и нулей относятся, в частности, к взрывным; ввиду этого в распределение введен полюс, отмеченный крестиком немного выше начала координат, характеризующий источник в виде ступенчатой функции, т. е. с наклоном огибающей спектра — 6 дБ/окт. Два кружка у начала координат соответствуют

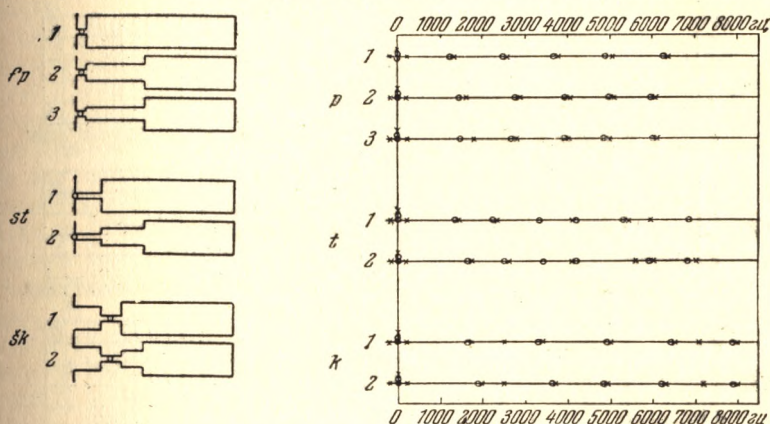


Рис. 69. Влияние конфигурации задней полости на расположение нулей и полюсов в моделях губных, зубных и небных звуков, согласно моделям, приведенным на рис. 68. Полости обозначены номерами 1, 2, 3 в порядке возрастания соответствия их условиям палатализации.

нулю, связанному с излучением, и первому нулю на частоте, при которой обращается в бесконечность импеданс, взятый по направлению от источника к голосовой щели, т. е. первому нулю функции $H_z(s)$ в формулах (III.61) и (III.63). Два крестика по обеим сторонам начала координат соответствуют сопряженной паре полюсов, характеризующих основной резонанс всего речевого тракта, т. е. относящихся в первой форманте $F1$. Поскольку затухание для резонанса, отвечающего $F1$ в глухих звуках, обычно больше критического, то эти две точки в действительности лежат на отрицательной вещественной оси, т. е. на вертикальной оси упрощенной диаграммы; однако здесь это обстоятельство не учитывается. Суммарное влияние на форму спектра тесно сгруппированных у начала координат особых точек эквивалентно влиянию трех полюсов минус два нуля, т. е. одного полюса, обуславливающего в среднем спадание огибающей спектра на 6 дБ/окт в интервале взрыва у взрывных звуков.

В пояснение к этому формальному описанию при помощи нулей и полюсов можно добавить, что очень узкое артикуляторное сужение при отсутствии в нем потерь ведет себя как большая индуктивность импеданса источника, так как один из двух сопряженных нулей, связанных с $F1$, нейтрализуется нулем $H_z(s)$, расположенным в начале координат. Из этого следует, что влияние изменений площади сужения на уровень интенсивности более высоких формант может рассматриваться

либо как непосредственный эффект изменения импеданса источника согласно формуле (2.51), либо как эффект изменения частоты F_1 в зависимости от того, что более удобно для рассуждений.

Спектры для моделей [p], [t] и [k] имеют те же свободные полюсы и нули, что и для моделей на рис. 68, в которых не учитываются задние полости. Однако при учете их добавляются пары полюс — нуль, характеризующие сильно нейтрализованную резонансную структуру для задних полостей. Палатализация, как это видно из сравнения спектров для моделей 2 и 1 или, еще более отчетливо, из сравнения спектров для моделей 3 и 1, всегда приводит к повышению частоты F_2 . При губных, по мере сужения «ротовой» трубы, все более четко разделяются F_2 и связанный с ней нуль. В известной мере это относится также и к F_3 .

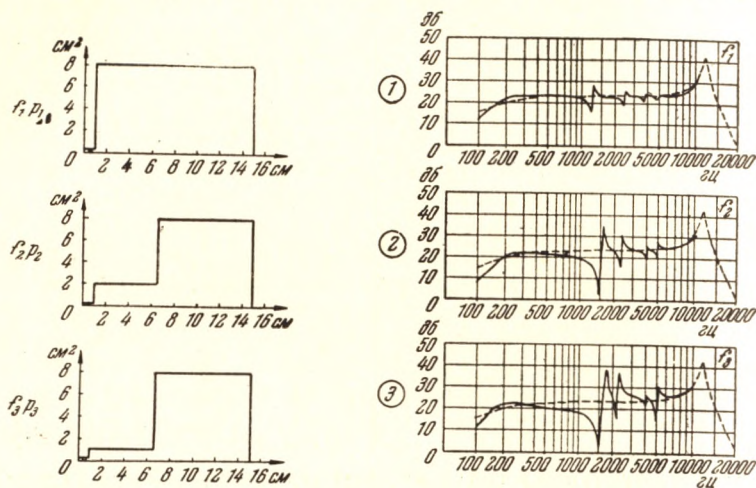


Рис. 70. Функции площади и спектры моделей губных звуков рис. 68 при источнике с равномерным спектром и сопротивлением $0,25\text{рс}$, расположенном на расстоянии $0,5\text{ см}$ от излучающего конца. Сопротивление голосовой щели принято равным 5рс . По мере увеличения палатализации F_2 и F_3 , а также нуль, расположенный непосредственно перед F_2 , становятся все более отчетливо выраженными.

для губных и в меньшей степени к F_2 и F_3 для зубных. Вообще говоря, подчеркнутость F_2 и F_3 достигает наибольшей степени в том случае, когда импеданс нёбного участка ротовой полости становится соизмеримым с импедансом сужения, в котором расположен источник. Для этого последнее сужение должно быть относительно коротким и широким или нёбный участок — длинным и узким. Таким образом, условия, при которых появляется формантная структура, аналогичны условиям для шума аспирации, рассмотренным в § 2 Приложения II.

На рис. 70 приведены огибающие спектра для длительных звуков, создаваемых тремя моделями с губным отрезком, а также функции площади для этих моделей, характеризующие размеры резонаторов. Штриховыми кривыми на графиках представлены огибающие в том случае, когда влияние задних полостей не учитывается. Можно видеть, что остаточная формантная структура сплошной кривой накладывается весьма симметрично на штриховую огибающую. Постоянство в среднем спектрального уровня обусловлено постоянным уровнем спектра источника, принятого для длительных фрикативных. Можно видеть, что F_2 повышается с 1300 гц на графике 1 до 1700 гц на графике 3, соответ-

ствующем максимальной степени палатализации. Уровни интенсивностей второй и третьей формант увеличиваются на 10 дБ вследствие разделения соответственных пар полюс-нуль. Для звуков [f.] и [p.] нуль, лежащий чуть ниже $F_2 = 1500$ гц, у нашего испытуемого расположен в области 1300—1500 гц как и в расчетных, так и в естественных спектрах (см. соответственно рис. 62 и 67); это было детально рассмотрено выше. Детальная структура спектра, рассчитанного для упрощенной модели рис. 69, более походит на спектр естественного звука, приведенный на рис. 62, чем полученная расчетом по рентгенографическим данным. Это обусловлено, очевидно, тем, что испытуемый слишком плотно прижимал язык к нёбу во время рентгеносъемки и тем самым чересчур сужал нёбный проход.

На рис. 71 показаны функции площади и соответственные спектры упрощенных моделей для щелевых с задней полостью, состоящей из одного отрезка трубы и спектры тех же моделей без задней полости. Дополнительно график 1 дает представление о том влиянии, которое оказывает пренебрежение фильтрующим действием того сужения, где находится источник в модели палатализованных звуков. Фильтрующие свойства сужения почти не сказываются на главном пике на частоте 2500 гц, но следующий пик, на 7000 гц, сдвигается к 8000 гц. Последнее соответствует $3/4$ -волновому резонансу передней полости, если ее рассматривать как трубу, открытую спереди и закрытую сзади. Этот сдвиг сопровождается уменьшением уровня второго пика, что обусловлено тесной зависимостью этого резонанса от передней полости, вследствие чего он сильнее демпфируется сопротивлением излучения, возрастающим с увеличением частоты.

Дополнительные изменения, которые вносит наличие задней трубы в моделях нёбных, губных и зубных звуков, показаны на графиках 2, 3 и 4; они в основном не отличаются от рассмотренных выше. Основной пик нёбных звуков соответствует F_3 , но на частоте 1700 гц видны следы пика F_2 . Из-за весьма малого импеданса сужения, принятого для всех расчетов равным 0,25 рс, виден главный нуль зубных звуков, лежащий на частоте 3500 гц. Дополнительные формантные пики имеют весьма малую величину по сравнению с уровнем основного пика зубных звуков, лежащего на частоте 7000 гц.

Эффект добавления более сложной, характерной для палатализованных звуков, задней полости показан для трех взрывных на рис. 72; при построении кривых этого рисунка был принят источник со спадающим огибающей спектра 6 дБ на октаву. Как и следовало ожидать, F_2 и F_3 для палатализованных губных и зубных взрывных более выражены, чем для тех же, но непалатализованных звуков, спектры которых были приведены на рис. 71. Влияние палатализации в модели губных звуков было рассмотрено выше в связи с рис. 70. Как было показано, относительная подчеркнутость F_2 и F_3 в шумовом интервале фриктивных и взрывных может являться вторичным признаком палатализации наряду с влиянием палатализованного звука на формантные переходы в прилегающем гласном, характер которых определяется высокими локусами F_2 и F_3 .

На модели нёбных звуков влиянием такой идеализированной палатализации можно пренебречь. Однако в действительности палатализация [k] связана с уменьшением размеров переднего резонатора. Поэтому основной резонанс сдвигается так, что в F -картине велярного [k] он идентифицируется как F_2 , а для препалатального [k] — как F_3 или F_4 .

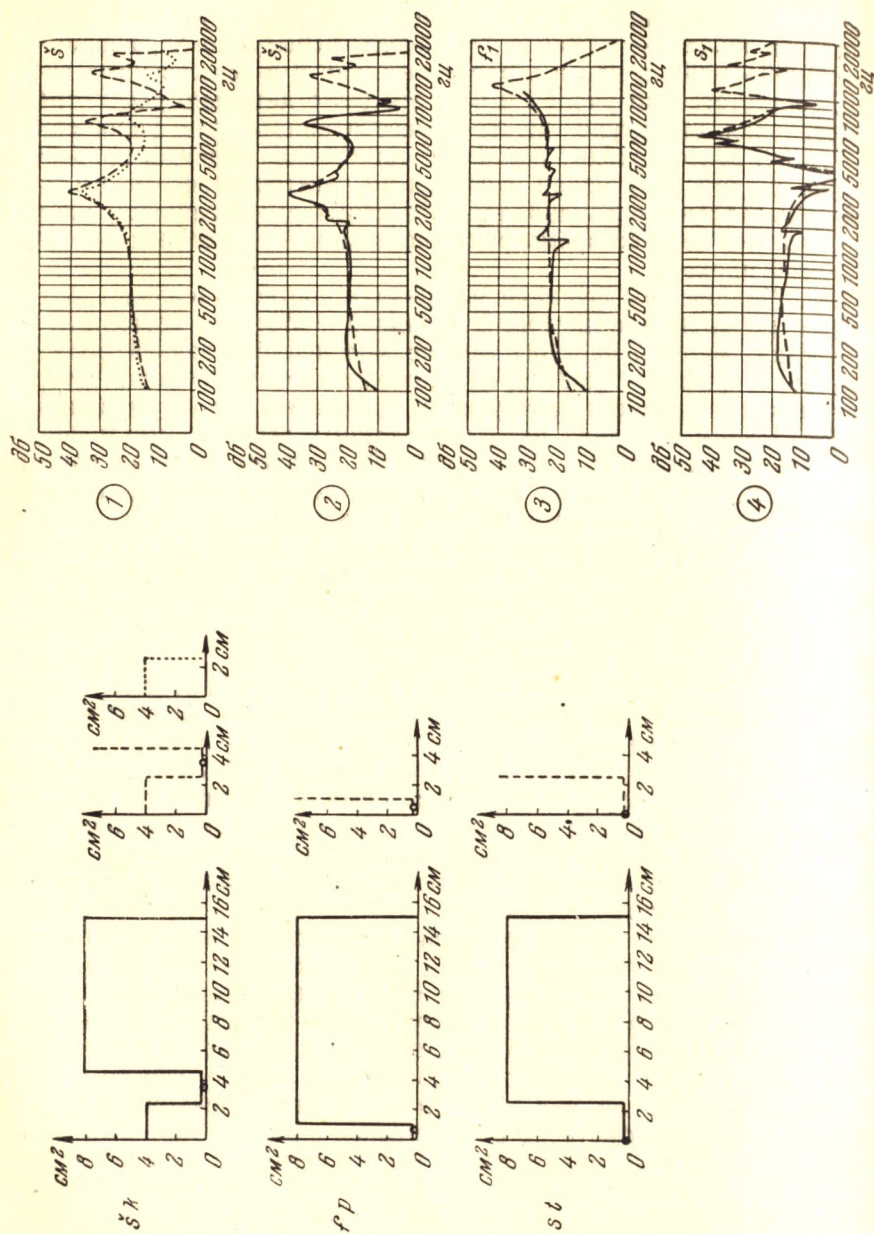


Рис. 71. Влияние добавления отрезка трубы в качестве задней полости к модели рис. 68, рассматриваемым как переносная полость. Для небоях показано также влияние устранения фильтрующего эффекта, обусловленного сужением проходом. Характеристики источника, как на рис. 70.

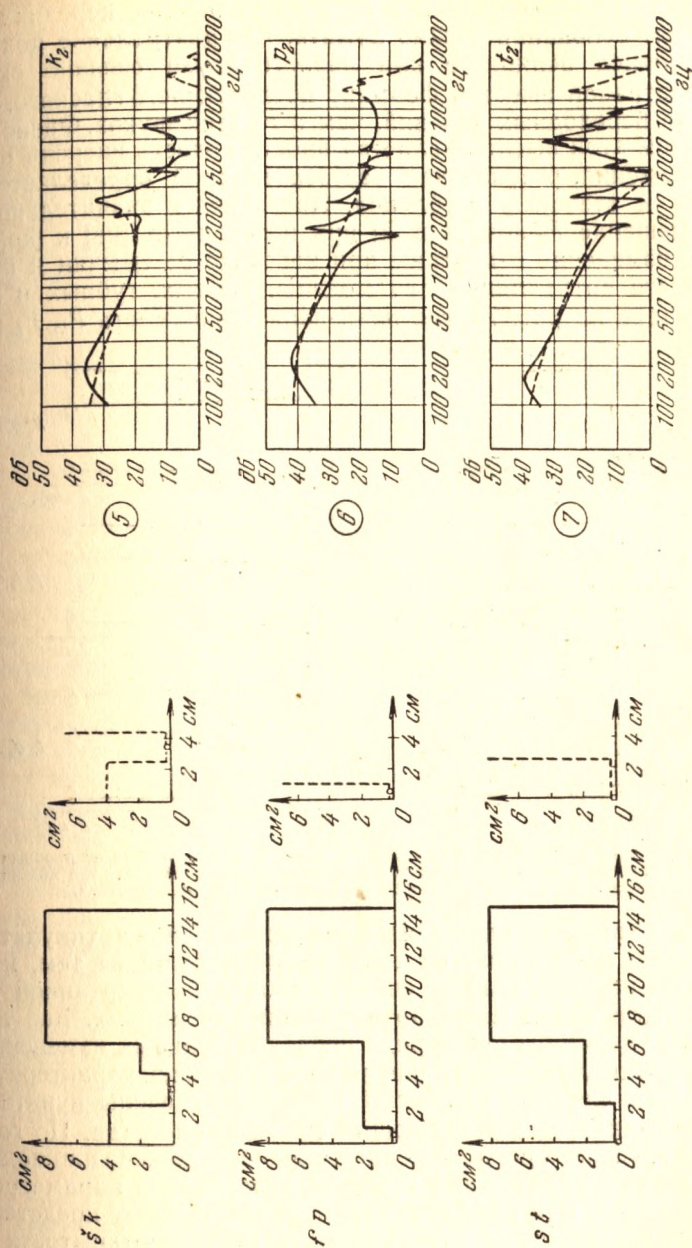


Рис. 72. Влияние добавления задней полости, характерной для палатализации, в моделях нёбных, губных и зубных, приведенных на рис. 68. Спектр источника — с наклоном отбояющей — 6 дБ/окт.

Эффективная «средняя» высота совокупности F_2 и более высоких формант для задних гласных близка к F_2 , а для наиболее передних — к F_3 или F_4 . На этом основании можно думать, что центральное положение главного резонанса велярного или нёбного звука связано скорее с большой эффективной высотой формант последующего гласного, особенно в интервале переходных явлений, чем с F_2 этих звуков. Физиологическим коррелятом этой высоты является просто полость впереди от артикуляторного сужения, резонансная частота которой определяется объемом полости и степенью огубления. Последняя в значительной мере определяет смещение F_2 -локуса в [k] при переходе от [ka] к [ko].

Типичное изменение F -картины при переходе от смычки к полному раствору в месте артикуляции для моделей нёбных, губных и зубных смычных показано на рис. 73, где приведена зависимость F_1 , F_2 и F_3 от

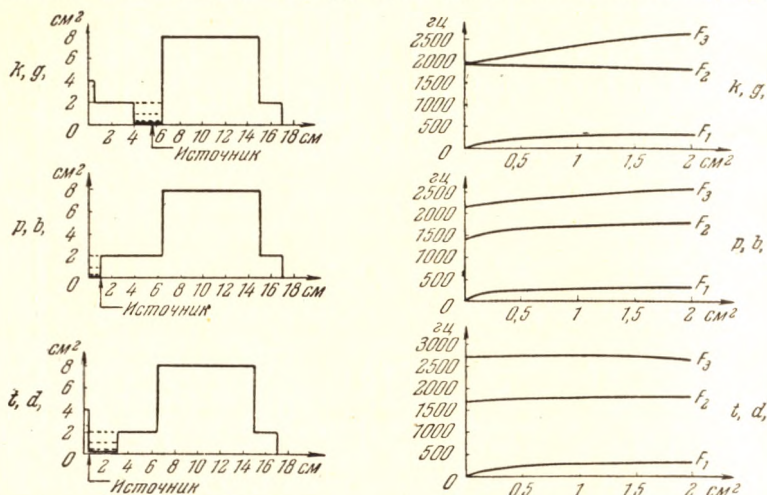


Рис. 73. F -картины моделей нёбных, губных и зубных в зависимости от площади поперечного сечения в проходе, соответствующем месту артикуляции. Все три модели отличаются только местоположением и размерами суженного прохода.

изменяющейся во времени площади поперечного сечения артикуляторного сужения. Модели, показанные на этом рисунке, подобны тем, которые были разобраны раньше, если не считать добавления гортанной трубки и более сдвинутого назад места артикуляции для нёбных. Все три модели построены так, чтобы после ряда последовательных изменений площади в месте артикуляции они приняли конфигурацию, характерную для передних гласных. Ступени изменений площади отмечены пунктирными линиями на графиках, представляющих функции площади. По горизонтальной оси на графиках, расположенных справа, отложена площадь поперечного сечения артикуляторного сужения, являющаяся параметром, зависящим от времени. Эти графики, таким образом, дают представление об изменении F -картины в зависимости от степени открытости звука.

При полной смычке кривые для F_2 и F_3 модели нёбных сходятся на частоте 1900 гц. Это вызвано совпадением частот основного резонанса передней полости и полуволнового резонанса задней. В случае губных как F_2 , так и F_3 имеют в начале более низкие частоты ($F_2 = 1400$ гц, $F_3 = 2200$ гц). F -картина в интервале смычки зубных ближе к F -картине прилежащего гласного, для которого $F_2 = 1700$ гц и $F_3 = 2700$ гц. Предельная F -картина при полном растворе, характери-

зующаяся частотами $F_1 = 400$ гц, $F_2 = 1750$ гц и $F_3 = 2600$ гц, соответствует слегка огубленному полуоткрытому переднему гласному.

Сравнение с аналогичными переходными F -картинами, наблюдаемыми в спектрограммах естественной речи, показывает, что скорость изменения площади поперечного сечения после момента взрыва лежит в пределах от 5 до 20 см²/сек; это означает, что при желании преобразовать шкалу площади сужения на рис. 73 в шкалу времени мы должны были бы изменить масштаб по оси абсцисс в 2—8 раз.

Изменения F_2 во время перехода для губных взрывных в основном заканчиваются, когда площадь губного сужения достигает 0,2 см². В процессе последующего открытия индуктивное сопротивление губного сужения становится малым по сравнению с таковым для нёбно-язычного прохода и потому не оказывает существенного влияния на полуволновой резонанс фаринкса, определяющий F_2 . Однако увеличивающаяся степень открытия губ обуславливает дальнейший заметный сдвиг вверх по частоте резонанса ротовой полости и тем самым F_3 . В [k] и [g] расхождение частот F_2 и F_3 , следующее за взрывом, продолжается вплоть до последующего гласного. Эти примеры показывают, что соотношения между степенью артикуляторного открытия и характером переходных процессов не так просты. Скорость движения артикуляторных органов также бывает различной. Открытие губного прохода происходит обычно с большим ускорением, чем понижение языка при замыкании нёбной или велярной смычки. Поэтому можно часто видеть, что основная часть перехода в случае губных заканчивается в пределах 15 мсек.

Изменение местоположения источника внутри модели голосового тракта при неизменной конфигурации полостей не влияет, очевидно, на частоты резонансов до тех пор, пока сохраняется один и тот же импеданс сужения. Активная составляющая импеданса сужения, так же как и параметры возможных источников шума, в основном определяется постоянным потоком воздуха; поэтому изменения местоположения источника представляют интерес только с точки зрения исследования поведения модели при различных допущениях. В связи с этим может возникнуть вопрос, могут ли небольшие изменения уровня формант, обусловленные изменением функции нулей при смещении источника, но при сохранении одной и той же конфигурации полостей, вызвать столь большое изменение формы спектра, что оно привело бы к различной фонематической оценке звука. Для окончательного ответа на этот вопрос нужно провести опыты по восприятию синтетической речи, хотя имеющиеся данные говорят о вероятности положительного ответа.

Один из примеров подобных небольших изменений в процессе синтеза взрывных представлен на рис. 74 расчетными данными для модели из двух отрезков труб. Приведенные на этом рисунке спектры были получены путем численного расчета по формулам гл. 4 и были опубликованы ранее [48].

В расчетах учитывалось только демпфирование, которое обусловлено собственным активным сопротивлением звеньев аналога и сопротивлением излучения. Добавочное сопротивление, вносимое источником, таким образом, не принималось во внимание. Модели звуков [k], [g] и [d], [t] были идентичны по конфигурации и соответствуют довольно большому артикуляторному раствору. Нули модели [g] и [k] расположены на частотах $\frac{nc}{2l_2} = n \times 1850$ гц, где l_2 — длина задней полости,

равная 9,5 см. Первый из этих нулей соответствует нулевой частоте; следующий, на частоте 1850 *гц*, оказывает, как обычно, нейтрализующее влияние на вторую форманту. Третья форманта, на частоте 2500 *гц*, образующая основной пик, отвечает полуволновому резонансу переднего отрезка, а пик на 5000 *гц* обусловлен цельноволновым резонансом этого отрезка. Такие же форманты отчетливо выражены и в спектрах [d] и [t]; однако здесь они не являются доминирующими ввиду наличия промежуточных формант, чередующихся с нулями. Так как источник в модели зубных звуков занимает крайнее переднее положение, нули и полюсы находятся в среднем на одинаковых расстояниях. Эту статистическую закономерность, так же как и ряд специфических спектральных особенностей для зубных и нёбных, можно видеть в соответствующих спектрах естественных звуков, приведенных в правой части рисунка. Эти спектры относятся к смешанной фрикативно-аспираторной фазе шведских взрывных в конечном положении.

Модель губных звуков является самой простой из всех возможных. Тем не менее и она воспроизводит некоторые особенности спектров естественных звуков, такие, как последовательное уменьшение амплитуд трех первых формант и наличие высокочастотной формантной области, лежащей выше 5000 *гц*.

§ 4. Заключение относительно характеристик источника для фрикативных и взрывных

Расчеты, проведенные в предыдущих разделах, показывают, что, несмотря на недостаток данных о физической природе источников турбулентных звуков, метод эквивалентных схем может быть использован для фонетического обоснования теории фрикативных и взрывных. Существенные особенности их могут быть получены, так же как и для гласных, на основе рассмотрения характеристик источника и фильтра, причем в ряде случаев спектральные характеристики этих звуков могут быть предсказаны не хуже, чем для гласных. Недостаточно определенными в расчетах остаются данные, касающиеся положения и спектра источника турбулентного звука.

Особенно хорошее совпадение между расчетным спектром и спектром естественного звука [x] обусловлено достаточно точным расположением источника в центре сужения во время спонтанных движений артикуляторных органов, вызываемых бернуллиевыми силами вытекающего воздуха. Спектр источника со спаданием огибающей 6 *дб* на октаву, дающий правильные результаты для [x] до частоты 3000 *гц*, может быть обусловлен толчками вырывающихся при этом порций воздуха.

Возможно, что в случае [š], помимо источников в области язычного сужения, имеется дополнительный зубной источник; для звука [šč] наличие такого источника столь же вероятно.

Для губно-зубных звуков источник, по-видимому, расположен вблизи от верхних передних зубов; однако если добавляется сильная палатализация и увеличение губного прохода, может возникнуть соизмеримый с первым нёбный источник. Вторичный нёбный или гортанный источник не является обязательным для появления соответствующей F-картины в спектре звука, так как она определяется главным образом величиной артикуляторного раствора; поэтому трудно установить непосредственно по спектрограмме наличие подобных дополнительных

источников. Источник в голосовой щели, так же как и губной источник, подчеркивает F_1 , а нёбный источник приводит к некоторому подчеркиванию F_3 или F_4 и к нейтрализации F_2 при условии, что этот источник не распределен по всей длине нёбного прохода. Впрочем, сказанное относительно нёбного источника остается в силе, хотя и в меньшей степени, и в том случае, когда комбинируются нёбное сужение и губной источник. При этом вполне определенным признаком, свидетельствующим о наличии губного источника, является нуль, лежащий приблизительно на 1300 $гц$.

Импеданс источника, строго говоря, относится к фильтровой функции, но его активная составляющая зависит от постоянного потока воздуха и потому в условиях турбулентности пропорциональна скорости потока, как это показано в § 2 Приложения II. Отношение активной составляющей импеданса сужения к реактивной R_s/L_s заметно возрастает с увеличением потока. Хотя первичной характеристикой источника в эквивалентной схеме является напряжение $E_s(\omega)$, но объемная скорость потока

$$U_s(\omega) = \frac{E_s(\omega)}{R_s + j\omega L_s}, \quad (10.1)$$

и, таким образом, звуковое давление в излучаемой волне зависит также и от импеданса сужения. Как показано в § 6 Приложения III, граничная частота импеданса сужения $F_c = \frac{R_s}{2\pi L_s}$ в условиях турбулентности почти не зависит от площади сужения:

$$F_c = \frac{v}{2\pi l}, \quad (10.2)$$

где v — скорость частиц воздуха и l — длина сужения. В нормальных условиях образования зубных или нёбных фрикативных при $l = 2,5$ см и $v = 3000$ см/сек граничная частота равна приблизительно 200 $гц$, но сдвигается к 500 $гц$ при более коротком губном проходе.

Пока эта частота выше, чем F_1 , демпфирование речевого тракта для первой форманты больше критического, что и было принято в настоящем анализе.

Отсюда видно, что при определении спектра источника, как разности между кривой спектральных уровней естественного звука и частотной характеристикой фильтровой функции речевого тракта, на низких частотах возникает неуверенность в правильности выбранного для модели речевого тракта активного сопротивления. Если при расчетах эта величина преуменьшена или преувеличена индуктивность, то на низких частотах будут преуменьшены и спектральные уровни напряжения, создаваемого источником. Принятое в §§ 2, 3 гл. 10 значение R/L дает только правильный порядок величины и может отличаться от действительного значения вдвое. Поэтому в настоящей работе не было смысла определять спектр источника для частот ниже 300—500 $гц$.

На частотах выше 5000 $гц$ размеры речевого тракта, перпендикулярные к направлению потока воздуха, не всегда столь малы, чтобы можно было пренебрегать снижением уровней спектра, обусловленным поперечными модами колебаний. Влияние такого рода колебаний относится в большей мере к палатализованным ретрофлексным, чем к зубным. В дальнейшем это влияние не принималось во внимание.

Вместо того чтобы приводить отдельно данные об огибающих спектра источника, принятых для каждого из звуков, целесообразно представить эти данные в следующем компактном виде:

Огибающая спектра звукового давления источника	Пригодна для следующих звуков в указываемом диапазоне частот:
Спектр со спаданием 6 <i>дб/окт</i> (т. е. интегрированный белый шум для длительных звуков и ступенчатая функция для переходных звуков) Спектр со спаданием 0 <i>дб/окт</i> (т. е. спектр с постоянным спектральным уровнем)	[x] 300—4000 <i>гц</i>. [f] 800—10 000 <i>гц</i>. Лучшее совпадение получается при спадании огибающей спектра источника, равном 3 <i>дб/окт</i>. [i,] 400—8000 <i>гц</i>. В фазе взрыва и шумовом интервале максимальной интенсивности для звуков [k], [k,], [p], [p,] 300—8000 <i>гц</i>. [š] 300—6000 <i>гц</i>, предполагая апикулярный источник. В области 300—800 <i>гц</i> данные не вполне достоверны. [š] 300—2000 <i>гц</i>, предполагая зубной источник. Такой спектр мало вероятен. Выше 2000 <i>гц</i> следует принимать спадание огибающей в 12 <i>дб</i> на октаву. [šš] 300—3000 <i>гц</i>; при зубном источнике выше 3000 <i>гц</i> следует принимать спадание огибающей в 6 <i>дб</i> на октаву. [šš] 1500—8000 <i>гц</i> при расположении источника в области сужения. [s] 800—4000 <i>гц</i>; выше 4000 <i>гц</i> следует принимать спадание огибающей 6 <i>дб</i> на октаву. [s,] 1000—8000 <i>гц</i>. [t] и [t,] 500—8000 <i>гц</i> для фриктивного интервала.

Следует отметить, что эти данные относятся к тем частным отрезкам естественной речи, которые указаны на спектрограммах, приведенных в Приложениях.

Возможная ошибка в оценке наклона огибающей спектра источника довольно велика и может достигать 3 *дб/окт*; важнейшей причиной отклонений расчетных спектров от спектров естественных звуков являются неизбежные различия в расположении пиков огибающей спектра, обусловленные разницей в артикуляции при обычной речи и при снятии рентгенограмм, а также неточностью определения размеров речевого тракта и величины активных элементов в эквивалентных схемах.

По-видимому, нет существенной разницы в среднем наклоне огибающей спектра между звонкими и глухими взрывными для интервала, соответствующего взрыву. Согласно Халле, Хьюзу и Рэдли [86] в напряженных английских взрывных огибающая спектра падает менее круто, чем в ненапряженных; такая же тенденция наблюдается и для шведских взрывных (см. рис. 74).

Данные для длительных зубных и нёбных фриктивных звуков хорошо сходятся с данными, полученными Гейнцем [96] для спектра механических моделей источника, имевших длину 1 см и площадь сужения 0,03 см². Он нашел, что спектральный уровень почти постоянен от 1500 до 4000 *гц* и что выше 4000 и ниже 1500 *гц* наклон огибающей составляет 12 *дб/окт*¹⁾.

На вопрос о размерах сужения, оптимальных с точки зрения возбуждения тех или иных турбулентных звуков, пока нет удовлетвори-

¹⁾ По данным MIT спектр источника даже более равномерен (личное сообщение К. Н. Стивенса).

тельного ответа. Предварительные исследования показали, что площадь сужения в $0,1 \text{ см}^2$ дает правильный порядок величины; представляется, что импеданс сужения в области частот главного спектрального пика ведет себя по преимуществу как большая индуктивность. Поэтому следует ожидать, что площадь, обеспечивающая максимум создаваемого источником давления (т. е. напряжения в эквивалентной схеме), меньше, чем площадь, необходимая для максимальной отдачи звука.

Типичная для [f], значительно меньшая по сравнению с [s] или [š] интегральная интенсивность может быть в значительной мере объяснена отсутствием резонанса передней полости в [f] в пределах частот до 8000 гц, или, если этот резонанс существует, более низкой добротностью Q , в сочетании с более крутым спаданием огибающей спектра источника. К чему приводит первое предположение, можно видеть на рис. 71. Второе предположение оправдано возможностью менее интенсивного шумообразования в случае весьма узкого, но растянутого в боковых направлениях прохода, как показано в § 2 Приложения II. Возможно также, что преграда, которую представляют собой зубы, является существенным фактором, обуславливающим большую интенсивность зубных и губно-зубных по сравнению с губными и большую подчеркнутость высоких частот в спектрах первых. В дополнение к опытам на моделях необходимо дальнейшее тщательное изучение уровней и спектров шума, условий протекания воздуха, дифференциальных сопротивлений, а также конфигурации речевого тракта при образовании турбулентных звуков.

ЧАСТЬ ТРЕТЬЯ

ВЫВОДЫ

ГЛАВА II

СЕГМЕНТАЦИЯ РЕЧИ И ОПРЕДЕЛЕНИЕ ЕЕ СТРУКТУРНЫХ ЭЛЕМЕНТОВ

Работа, результаты которой излагаются в настоящей книге, была посвящена главным образом теории образования звуков речи, в особенности аналитическим соотношениям между артикуляцией и акустической картиной, а также теории резонансных систем, являющихся приближением к системе резонаторов речевого тракта. В дополнение в Приложениях некоторое внимание уделено приборам и методам анализа и представления акустических аспектов речи, главным образом для иллюстрации использованной техники и дополнения расчетов характеристиками естественной человеческой речи.

Проведенное исследование имеет общий характер, и ссылки на звуки русской речи обусловлены только тем, что использованные для расчетов рентгенографические данные были получены для русского диктора. В данном разделе мы подытожим и обсудим полученные результаты с точки зрения специфических интересов фонетики.

Проблема сегментации речевого процесса на основании визуального анализа записей была затронута в § 2 гл. 1. Понятно, что акустические границы, обнаруживаемые на таких записях, могут быть соотнесены только с фонетической транскрипцией и не могут рассматриваться как границы фонем¹⁾. Фонетические (или акустические) границы лучше всего выявляются на спектрограммах в виде локализованных во времени изменений звуковой картины. Они делят речь на единицы, соответствующие звуку речи или части его. Так как любой звук речи можно рассматривать как результат фильтрации звука, создаваемого источником, целесообразно связывать границы с изменением или типа источника звука, или его интенсивности, или с быстрым изменением фильтровой функции голосового тракта, или, наконец, с одновременным изменением как источника, так и фильтра.

Примерами звуковых единиц, которые логично считать частями звуков речи, являются звонкий или глухой интервалы, соответствующие артикуляторной смычке и последующему шумовому интервалу взрывных звуков. Особенно важно иметь четкое определение основных единиц речи при измерении длительности взрывных. В положении перед гласными длительность шумового интервала можно определить как расстояние во времени между моментом раскрытия смычки и началом колебаний голосовых связок. Если голос и взрыв совпадают во времени и если, кроме того, существует тенденция к фрикативности, как

¹⁾ См., например, [204].

в русском [d], взрыв следует определять по длительности шумового интервала, видного на спектрограмме. Взрыв в свою очередь может быть разбит на последовательные частично перекрывающиеся интервалы раствора, фрикации и аспирации¹⁾ (см. § 1 гл. 1 и § 2 Приложения II). Однако только в редких случаях удастся установить точные границы этих интервалов. Другим примером расчленения на единицы, меньшие чем звуки речи, является появление видимой границы в том случае, когда момент включения голоса запаздывает. Это имеет место в плавных, начальная часть которых ассимилирована с предшествующим глухим согласным. Подобные более мелкие единицы являются естественными элементами, из которых складывается связная речь; необходимо, однако, определенным образом условиться о порядке соотношения их с теми или иными звуками речи. Фишер-Йоргенсен [57, 58] считает, что аспираторный интервал сильного взрывного звука может быть по артикуляторным признакам (т. е. по фильтровой функции) отнесен к последующему гласному; однако обычно он рассматривается как часть взрыва, что целесообразно с практической точки зрения, так как дает возможность охарактеризовать различие взрывных по признаку «напряженные — ненапряженные». У напряженных взрывных взрыв длительнее, если включить аспирацию в состав согласного (см. рис. 100 и замечания в конце настоящего раздела).

Структурное описание речи как последовательности фонем не находится в противоречии с тем обстоятельством, что речевые стимулы, определяющие идентификацию фонем слушающим, распределены в нескольких последовательных во времени минимальных акустических отрезках [83].

Единственное затруднение при этом возникает в том случае, когда исследователь ставит перед собой задачу определить или измерить длительности фонем. Конечно, возможно принять определенные условия, позволяющие попросту использовать исходную фонетическую транскрипцию в структурном плане путем необходимого расширения смысла некоторых знаков транскрипции. Такая логическая операция позволяет устранить взаимное перекрытие фонем во времени, однако оно сохраняется в физических речевых сигналах в том смысле, что один и тот же участок может нести информацию, относящуюся к двум последовательным фонемам.

Охарактеризовать количественно форму речевой волны нелегко. Выбору какой-либо системы описания должна предшествовать оценка точности, обеспечивающей наилучший компромисс между противоположными требованиями краткости и полноты. Требования эти различны в зависимости от целей, стоящих перед исследователем. Одной крайностью является изучение сигнала безотносительно к элементам сообщения, например при детальном описании акустической картины речи и ее связи с артикуляцией. В этом случае необходимо располагать совокупностью огибающих спектра, взятых для последовательных временных отрезков речи.

Другой крайностью является задача формулировки акустических коррелятов дифференциальных признаков с минимальной избыточностью, как это, например, попытались сделать Якобсон и др. [114] и Якобсон и

¹⁾ В терминологии некоторых авторов, например Шатца [179], понятие «взрыв» относится к раствору + фрикации. Другие под «взрывом» понимают весь шумовой интервал взрывного.

Халле [115]. Описание акустических особенностей напряженных форм такими словами, как «большое рассредоточение энергии по спектру и во времени», довольно неопределенно и не может быть использовано для практических целей, пока не уточнено количественно понятие «рассредоточения энергии» в виде определенных формул такого типа, как предложенные Халле [81, 82] (см. также [105]).

ГЛАВА 12

СВЯЗИ МЕЖДУ *F*-КАРТИНОЙ И АРТИКУЛЯЦИЕЙ

При выборе и сопоставлении первичных данных, получаемых на основании записей естественной речи, полезно, в дополнение к данным о количественном распределении энергии по спектру, определять и соответствующую *F*-картину (см. § 3 гл. 1).

Для гласных *F*-картиной является просто совокупность формантных частот с некоторыми специфическими ограничениями, необходимыми в случае наличия связи с носовой полостью. В более общей формулировке *F*-картина есть совокупность резонансных частот речевого тракта. Эта совокупность определяет существенные особенности спектра гласных и в известной мере спектра согласных и является хорошим коррелятом артикуляторных конфигураций речевого тракта. Обратный переход от спектральных данных к артикуляции до известной степени ограничен благодаря существованию компенсаторных форм артикуляции; только некоторые общие аспекты такой компенсации были рассмотрены в настоящей работе (см., например, §§ 1—3 гл. 4). Более глубокое понимание возможностей физиологической интерпретации спектрограмм требует выполнения большой работы по изучению коррелятивных связей с использованием кинорентгенографии речевого аппарата и спектрографии.

F-картина, регистрируемая непрерывно за время произнесения речевого отрезка, является обобщением понятий «локус» или «хаб», так как она включает не только F_2 , но также F_1 и F_3 и те из более высоких резонансных частот, которые могут быть необходимы в том или ином специальном случае. Формантные переходы в звонких интервалах, примыкающих к согласному, есть не что иное, как переход от *F*-картины, характерной для моментов максимального сужения при артикуляции согласного, к *F*-картине гласного.

Накопленные в фонетической литературе данные по речи относятся главным образом к вопросам образования речи и представляются обычно в форме описания положений артикуляторных органов; лишь в незначительной мере эти данные дополняются рентгенограммами и палатограммами. Однако для аналитического предсказания акустических характеристик звуков речи на основании данных о механизме их образования необходимо полное трехмерное описание геометрии полостей речевого аппарата. Последняя, конечно, коррелирована с положением артикуляторов, однако не следует думать, что место артикуляции, в традиционном фонетическом понимании этого термина, всегда совпадает с какими-либо акустически значимыми границами полостей речевого тракта (см. § 2 гл. 7).

Приближенное определение размеров полостей системой трех параметров, предложенное Стивенсом и Хаузом [194, 195] и использованное

нами в несколько иной форме (§ 3 гл. 4), с акустической точки зрения более рационально, чем традиционная классификация гласных по «высшей точке языка». Оно полезно также для соотнесения F -картины согласных с данными о конфигурации речевого тракта. Такими параметрами являются: 1) место и 2) площадь поперечного сечения язычного прохода (сужения) и 3) степень огубления. Эти параметры можно использовать по-разному. При рассмотрении губных звуков два параметра, относящиеся к язычной части тракта, можно, например, использовать для характеристики вторичного, т. е. внутреннего, места артикуляции. Различные диаграммы, связывающие F -картину с тремя параметрами такой модели, приведены в § 3 гл. 4. При использовании этих диаграмм может возникнуть необходимость в интерполяции; если общая длина речевого тракта данного диктора не совпадает с принятой на нашей модели, достаточно ввести линейный коэффициент пересчета шкалы, причем все форманты изменяются обратно пропорционально общей длине.

Один из выводов, которые позволяет сделать трехпараметрическое описание артикуляторных данных, заключается в том, что гласный [i] имеет ту же степень открытости и тот же эффективный язычный проход, что и [a], и что звуки могут быть расположены в порядке, определяемом местоположением этого прохода, например [æ], [a], [o], [u], [i], [e], [i]. Если артикуляция звука очень открытая с точки зрения эффективного язычного прохода, F -картина весьма близка к случаю нейтрального гласного, с частотами $F_1 = 500$ гц, $F_2 = 1500$ гц, $F_3 = 2500$ гц и т. д., независимо от местоположения язычного прохода, так что задание этого местоположения теряет смысл. Если имеется достаточно явно выраженный язычный проход, то оптимальным его положением для получения высокой F_1 является задняя часть модели, т. е. у фаринкса, а для получения высокой F_2 — середина передней половины модели, т. е. в области твердого нёба. F_1 понижается, а F_2 повышается, если сужается язычный проход, расположенный в передней половине модели. Обратное — в известных пределах — имеет место, если этот проход находится в задней половине модели.

Существует некоторое значение площади поперечного сечения прохода, оптимальное с точки зрения получения максимума F_1 , однако F_2 растет непрерывно с увеличением раствора в этом проходе. Сужение или удлинение губного прохода снижает частоты всех формант, представляющих интерес.

Вообще говоря, каждая из полостей, входящих в состав речевого тракта, оказывает в той или иной мере влияние на все резонансные частоты. Такую зависимость можно видеть непосредственно из рассмотрения номограммы для трехпараметрической модели. Если сдвиг язычного прохода вперед приводит, при прочих равных условиях, к повышению частоты форманты, то из этого можно сделать вывод, что она в большей мере обусловлена участком, расположенным перед язычным проходом, чем задней полостью. Таким образом, обусловленность формант той или иной из полостей изменяется в точках максимумов и минимумов этих кривых, как было отмечено Стивенсом и Хаузом [195]. Максимум одной из двух первых формант соответствует координате язычного прохода, близкой к координате минимума следующей по порядковому номеру форманты.

Специальные исследования были предприняты для решения вопроса о применимости моделей, основанных на двойном резонаторе Гельм-

гольца. Такая модель может быть использована для определения F_1 и F_2 задних гласных и гласного [i]. При расчете F_1 передних гласных весь речевой тракт может быть представлен в виде простого резонатора Гельмгольца. Передняя полость в [i] определяет в значительно большей степени F_3 , чем F_2 ; так, в русском гласном [i], который нами анализировался, F_2 определенно является полуволновым резонансом задней полости. Подробное рассмотрение коэффициентов, характеризующих связь частот формант с задней или передней полостью и губным или язычным проходом, было предпринято в § 2 гл. 7; полученные в результате данные сравнивались с расчетными данными, полученными на основе теории резонаторов Гельмгольца. При использовании электрического аналога LEA было установлено, что в задних гласных как передняя, так и задняя полости заметным образом влияют как на F_1 , так и на F_2 , и обычно постулируемая связь F_1 с задней, а F_2 с передней полостью оправдывается только в ограниченной степени. Однако для гласного [u] F_1 зависит от губного прохода в значительно большей степени, чем F_2 ; обратное справедливо для язычного прохода. Если вдобавок учитывать еще и объемы полостей, то можно сделать вывод, что для [u] частота F_1 связана главным образом с задней полостью и отверстием губ, а F_2 — с передней полостью и язычным проходом. Однако совокупное влияние передней полости и отверстия губ на F_1 в [u] больше, чем их влияние на F_2 , а совокупное влияние задней полости и язычного прохода на F_2 больше, чем их влияние на F_1 . С этой точки зрения F_2 в [u] можно считать формантой, обусловленной задним резонатором, а F_1 — передним.

Из проанализированных гласных классическая теория о связи F_1 с задней, а F_2 с передней полостью оказалась верной лишь для [i]. При отсутствии явно выраженного язычного сужения каждая форманта в равной мере зависит от всех частей речевого тракта, а для полуоткрытых передних гласных имеет место заметное влияние как передней, так и задней полостей на F_2 и F_3 ; F_2 становится в большей мере обусловленной передней полостью при сдвиге языка назад от того положения по отношению к нёбу, которое соответствует максимальному F_2 . Увеличение огубления влияет на связь формант с полостями так же, как смещение назад координаты язычного прохода. Поэтому, когда этот проход расположен спереди от палатального F_2 -максимума, добавочное огубление может изменить степень связи с передней полостью в пользу F_2 вместо F_3 . При фарингальном месте артикуляции огубление изменяет влияние передней полости так, что связь ее с F_3 усиливается, а с F_2 — уменьшается.

ГЛАВА 13

НЕКОТОРЫЕ АСПЕКТЫ ТЕОРИИ ДИФФЕРЕНЦИАЛЬНЫХ ПРИЗНАКОВ

Рассмотрим теперь некоторые аспекты теории дифференциальных признаков. В идеальном случае для всех признаков, которые могут независимо коммутироваться, должны быть независимыми и их акустические корреляты. Этот идеал, отнюдь не обязательный для лингвистического анализа, не может быть достигнут в такой системе гласных, в которой используют три или более неидентичных признака; действительно, гласные можно вполне удовлетворительно синтезировать,

используя только две форманты. При этом верхняя форманта с частотой F_{2e} заменяет, с точки зрения восприятия, вторую и более высокие форманты в передних гласных и совпадает с F_2 в задних гласных. Качество гласных определяется, таким образом, двумя переменными F_1 и F_{2e} .

Рассматривая три основных признака гласных, установленные Якобсоном и др. [114], можно ассоциировать увеличение компактности с F_1 , а уменьшение низкотональности — или с эффективной верхней формантной частотой F_{2e} , или с самой F_2 . Уменьшение бемольности (огубления) можно ассоциировать с $F_1 + F_{2e}$. При этом признак компактности становится акустически независимым от признака низкотональности. Более тесное смыкание губ или их вытягивание ведет, при прочих равных условиях, к понижению частот F_1 и F_{2e} , что является сдвигом к менее компактному и более низкотональному гласному в свете определений, приведенных выше. На структурном уровне наличием подобных взаимозависимостей можно пренебречь, но они нарушают принцип ортогональности на уровне звуковой субстанции. Такого вида нарушения обычны при анализе речи, однако они не вызывают затруднений, если известны взаимосвязи между признаками.

Теория дифференциальных признаков представляет собой не только мощный инструмент структурной лингвистики, главной задачей которой является построение максимально стройной системы, свободной от избыточности, но имеет также большое значение для техники, например как полезный принцип при построении пишущих машин, управляемых речью. Содержание понятия минимальной избыточности, однако, не одно и то же для инженеров и фонетистов. Прежде всего, для инженеров некоторая избыточность полезна, так как повышает надежность автоматического опознавания. Во-вторых, не всегда возможно с помощью каких-либо механических операций удалить избыточные признаки из акустических характеристик речи, не потеряв в той или иной мере значимой информации.

Возможны различные формулировки акустических коррелятов дифференциальных признаков и даже различный их выбор, так что система признаков, избранная структурной лингвистикой, может не совпасть с системой, принятой инженерами и фонетистами. Если фонемы сравниваются по принципу минимальных пар, достаточно использовать F_2 или F_{2e} как критерий низкотональности. Однако если в данном языке весь ряд задних гласных должен быть противопоставлен всем передним, пользуясь только акустическим критерием, более удобно принять в качестве такового $F_2 - F_1$ или $F_{2e} - F_1$ ¹⁾, проведя соответствующую линию раздела на диаграммах зависимости F_1 от F_2 или F_{2e} . Подобно этому, критическое значение $F_1 + F_{2e}$ можно использовать для разделения всех огубленных передних гласных и неогубленных.

Учитывая возможности, заложенные в параметрах $F_2 - F_1$ и $F_{2e} + F_1$, можно считать целесообразным принять их за основные координаты. Дополнительно параметр $F_2 - F_1$ можно использовать для разделения различных неогубленных передних гласных, а $F_2 + F_1$ — для разделения различных задних гласных.

¹⁾ $F_2 - F_1$ обозначает, например, F_2 минус F_1 и т. д. Эта схема классификации (см. формантные данные Фанта [49, 53]) приведена здесь главным образом в качестве повода для дальнейшей дискуссии; ее лингвистические достоинства еще нуждаются в оценке.

Двоичный признак «спектрального рассредоточения», определяемый параметром $F_2 - F_1$, может, таким образом, заменить признаки низкотональности и компактности и включить в то же время аспекты обоих этих признаков. Следует отметить, что параметр $F_2 - F_1$ более просто связан с трехпараметрической моделью речевого тракта, чем F_1 или F_2 порознь. Предельно противопоставленными неогубленными гласными являются средненёбное [i] и фарингальное [ɑ], образуемые оба с узким язычным проходом. Двумя группами, занимающими среднее положение, являются полуоткрытые и открытые передние гласные или незакрытые передние гласные и неоптимальные¹⁾ задние гласные; последние являются более открытыми с точки зрения ширины фарингального прохода, сочетаясь, возможно, с гортанным сужением ([æ] в противопоставлении [ɑ]) или более передним положением язычного прохода (сравните [ü] и [u], [ö] и [o]). Можно видеть, что переходы от узкой к широкой заднеязычной артикуляции связаны главным образом с увеличением F_2 , в соответствии с противопоставлением низкотональный — высокотональный звук.

С точки зрения восприятия $F_2 - F_1$ является критерием, позволяющим различать двухформантную структуру гласных от одноформантной. Аналогично параметр $F_{2e} + F_1$ отражает «среднюю» тональность гласного²⁾. Главным артикуляторным коррелятом $F_{2e} + F_1$ является огубление, однако смещение языка назад и увеличение объема ротовой полости также влияют на этот параметр.

Простейшая форма автоматического анализа должна, таким образом, использовать противопоставление друг другу рядов фонем во всех случаях, где это только возможно. Это более экономично, чем заставлять машину перестраивать критерии идентификации на основе последовательных решений, принимаемых в процессе анализа данного звукового интервала.

Теорию дифференциальных признаков иногда противопоставляют концепции о множественности характеристик, которые должны участвовать в акустическом описании дифференциального признака. Усложнение системы признаков может быть полезным в тех случаях, когда приходится рассматривать многообразие контекстов, в которых эти признаки встречаются, включая сюда различия между дикторами, вероятностные связи следования одних звуков за другими и специфические совокупности характеристик, сосуществующих в пределах изучаемого интервала. Однако существенный вопрос заключается в том, имело ли бы вообще смысл вводить понятие признака, если бы все характеристики звуков, которые могли бы быть так или иначе использованы для практических целей фонетики, изменялись бы в зависимости от контекста.

Следует учитывать два типичных случая в зависимости от того, являются ли различные характеристики всегда или почти всегда сосуществующими или же взаимно исключают друг друга. Представляется мало плодотворным обычный подход, когда из всех сосуще-

¹⁾ По определению, данному Якобсоном и Халле [115], оптимальным гласным является гласный, характеризующийся отсутствием явных ограничений во времени и наибольшим ограничением в частотной области; оптимальный согласный характеризуется крайним ограничением во времени при отсутствии явных ограничений в частотной области. (Прим. перев.)

²⁾ Если при воспроизведении магнитной записи уменьшить скорость носителя вдвое, то [a] превращается в [o], [e] — в [ö] и т. д.

ствующих характеристик выбирается одна на том основании, что эта характеристика сама по себе, без сопутствующих ей других характеристик, оказалась достаточной для идентификации и потому может рассматриваться как минимально необходимое условие. Опыты прослушивания систематически варьируемой синтетической речи, конечно, весьма ценны для установления относительной важности различных характеристик; однако, каковы бы ни были результаты таких опытов, остается желательным попытаться так сформулировать рассматриваемый признак в его акустическом проявлении, чтобы все характеристики логически входили в сжатое его определение, которое в идеале должно быть не более длинным, чем формулировка самой характеристики. Примером такой возможности является акустический коррелят ударения. Оказалось, что длительность при этом столь же важна, как и интенсивность [73]. Обе эти характерные для ударения особенности могут быть скомбинированы вместе в произведении, имеющем размерность энергии и представляющем более естественный коррелят ударения, чем одна интенсивность, не требуя в то же время усложнения формулировки признака ударности [53]. В некоторых случаях эти соображения применимы и к признаку напряженности.

Трудности, возникающие при формулировке основных аспектов таких признаков, как компактность, низкотональность, напряженность и т. д., которые относятся и к гласным и к согласным, очевидны, и на некоторые из них указывала Фишер-Йоргенсен [59]. Трактовка взаимноисключающих признаков как идентичных является, однако, не только средством экономии кода. Нельзя отрицать факт существования фонетического подобия между ними, хотя оно и не всегда настолько определено, чтобы какая-либо всеобъемлющая формулировка оказалась самоочевидной для любого контекста или оптимальной для всех контекстов. Это вполне естественно, поскольку степень подобия может быть различной. Один и тот же дифференциальный признак звука может быть сформулирован сжато и точно в той мере, в которой известны фонетические факты. Пока что акустический анализ речи находится еще на относительно ранней стадии развития.

ГЛАВА 14

ЗАМЕЧАНИЯ ОБ АКУСТИЧЕСКОЙ ПРИРОДЕ ДИФФЕРЕНЦИАЛЬНЫХ ПРИЗНАКОВ

Формулировки, которые дали в 1956 г. Якобсон и Халле, во многих отношениях лучше, чем те, которые Якобсон и др. выдвинули в 1952 г., но в них не хватает дополнительных разъяснений, которые облегчили бы фонетическую интерпретацию этих формулировок. Ниже приводятся, в свете результатов, полученных в этой работе, некоторые замечания к формулировкам Якобсона и Халле [115], цитируемым здесь:

1. Признак вокалические — невокалические формулируется следующим образом: «Акустически — наличие или отсутствие четко выраженной формантной структуры; генетически — источник возбуждения в основном или исключительно — одна только голосовая щель при наличии свободного прохода в речевом тракте».

Под «четкой формантной структурой» можно понимать физическую реализацию *F*-картины, например наличие *F*₁, *F*₂ и *F*₃ и преобладание

этой части формантной структуры над другими формантами или формантными областями, связанными с наложением шумовой картины в звонких фрикативных и взрывных или с явлениями назализации в носовых согласных. Из генетического описания, приведенного выше, можно видеть, что оно полностью применимо к различию между аспирацией и фрикацией в интервале взрыва взрывных, которое было рассмотрено в § 1 гл. 1 и в гл. 10. Однако [h]-звук, по сравнению с дающими подобную же *F*-картину гласными, является менее вокалическим, так как открытая голосовая щель вызывает сильное демпфирование *F*₁ у высокотональных и *F*₁ и *F*₂ у низкотональных звуков типа [h]. Подобные же явления имеют место при назализации.

2. Признак консонантные — неконсонантные сформулирован следующим образом: «Акустические — малая (противопоставляется большой) общая энергия, генетически — наличие или отсутствие преграды в речевом тракте».

Этот признак тесным образом связан с классическим разделением звуков речи на гласные и согласные, но он подразумевает не только это. Большая энергия (или, лучше, интенсивность, так как длительность неударных гласных может быть очень мала) акустически обусловлена степенью раствора — чем меньше раствор, тем более сильно выражено подавление высоких частот; этот эффект с точки зрения формантных представлений может быть описан как понижение *F*₁, влекущее за собой снижение общей интенсивности, как показано в § 1 гл. 1 и § 2 гл. 3.

Замечания Халле и др. [86], что согласные на спектрограммах ориентированы в вертикальном, а гласные — в горизонтальном направлении, приложимо к общему различию между переходными и длительными звуками, в особенности к внезапному сдвигу спектральной картины в случае разрыва артикуляторной смычки или быстрого движения артикуляторных органов к месту смычки. Однако разрыв непрерывности в спектрограмме всегда является признаком акустической границы между звуками. Такой разрыв может служить границей между длительным гласным и длительным согласным, но может также являться самостоятельным консонантным звуковым интервалом.

Квалификация гласных как вокалических и неконсонантных, плавных согласных — как вокалических и консонантных и остальных согласных — как невокалических и консонантных представляется хорошо обоснованной. Плавные, подобно гласным, отличаются отчетливой *F*-картиной, но в то же время сохраняют свойственный согласным элемент смычки, проявляющийся в наличии переходов на границах звука, а также в присущей всем согласным малой интенсивности.

Глайды более проблематичны. Звук типа [h] в английском и шведском языках имеет более задемпфированную первую форманту, чем плавные, и, таким образом, является невокалическим; он может также рассматриваться как неконсонантный в силу относительно большого артикуляторного прохода в речевом тракте выше голосовой щели. Однако последний признак приводит к трудностям на уровне акустической картины, так как [h] не противопоставлен ни одному из других звуков по признаку большей интенсивности. Смещение языка или губ от положения, характерного для [h], в положение частичного сужения сопровождается, как правило, появлением более интенсивного, чем у звука [h], «фрикативного» источника в сужении. Единственный неконсонантный признак [h] сводится к тому, что отсутствует разрыв непрерывности в граничном интервале, поскольку *F*-картина [h] близка к картине

последующего гласного. Скачок, соответствующий смене источника от аспирации при [h] к нормальной фонации при гласном, не очень отчетлив, особенно в интервокальном положении или в шепотной речи.

Слабый нефрикативный [j], плавно переходящий в последующий гласный, относится к глайдам и может рассматриваться так же, как и фонема [h] [81, 82]. Однако русские [j] представляют собой по большей части звонкие нёбные фрикативные звуки, причем фрикация особенно отчетлива в конечном положении [18]. Удельный вес фрикации в [j] примерно такой же, как и в [v], так что оба эти звука могли бы трактоваться аналогично. Определяя консонантность по признаку интенсивности, фонему [j], реализуемую в виде звонкого фрикативного, следовало бы классифицировать как невокалическую и консонантную, противопоставляя ее палатализованному [ʒ] (диззному). Если считать наличие фрикации несущественным, фонема [j] должна была бы быть охарактеризована как вокалическая и консонантная.

Тогда обе фонемы [v] и [j] были бы противопоставлены [l] и [r] как менее компактные, причем [v] отличалось бы низкотональностью по сравнению с высокотональным [l]. Такое решение привело бы, однако, к нежелательному разделению [v] и [f], хотя оно и представляется в известной мере справедливым на основании опытов, проводившихся на шведской синтезирующей установке OVE и на основании исследований Р. Якобсона [113]. Оказалось возможным имитировать [r] и [l] и начальные [v] и [j] значительно лучше, чем другие согласные, используя только голосовой источник и воссоздавая соответственную *F*-картину.

3. Признак компактный — диффузный определяется так: «Акустически — большая (или малая) концентрация энергии в относительно узкой центральной области спектра, сопровождающаяся увеличением (или уменьшением) общего количества энергии; генетически — расширяющаяся вперед или назад конфигурация речевого тракта. Различие определяется соотношением между объемом резонансной полости перед сужением и позади сужения».

В табл. 4 (стр. 119) были приведены данные о размерах речевого тракта, относящиеся к проанализированным в настоящей работе русским гласным. Отношение объемов пригодно как критерий различия между [a] и [o], [o] и [u], [e] и [i]; однако это же отношение может служить для различения низкотональных и высокотональных гласных. В ряду задних (низкотональных) гласных объем передней полости изменяется меньше, чем объем задней; обратное справедливо для передних, высокотональных, гласных, в которых различные степени компактности связаны в первую очередь с изменениями объема передней полости и площади язычного прохода.

Термины «расширяющийся вперед» и «расширяющийся назад» непосредственно приложимы к описанию идеализированных «рупорных» моделей речевого тракта, приведенных на рис. 19. Эти модели дают акустические корреляты к избирательному сдвигу F_1 .

Следует отметить, что если определять признак компактности на артикуляторном уровне независимо от наличия огубления, то это приводит к соответственным ограничениям на акустическом уровне; а именно, изменения F_1 , обусловленные огублением, приходится считать несущественными для компактности. В действительности увеличение понимаемой таким образом компактности и уменьшение бемольности (т. е. огубления) типично в ряду задних гласных [u], [o], [a]. Какой из этих признаков считать главным, зависит от особенностей данного языка и

от критерия эффективности классификации, принятого в структурной лингвистике. С другой стороны, если определять компактность на акустическом уровне, приняв за критерий значение F_1 , придется считать огубление одним из артикуляторных способов реализации надлежащего эффекта. Такое решение не ограничивает практического использования признака бемольности, так как при акустическом определении последней учитываются все форманты.

Компактность в согласных, т. е. акустические и артикуляторные особенности, типичные для нёбных и велярных, в случае фрикативных находятся в простой связи с размерами полости перед артикуляторным сужением; эта связь была детально рассмотрена в §§ 2 и 3 гл. 10. Если полость настолько мала, что ею можно пренебречь, то компактность определяется самим сужением, как в губных и зубных звуках. Конфигурация ротовой полости при этом подобна рупору, обращенному раствором внутрь.

Высокий F_2 -локус типичен только для нёбных. У велярных F_2 ниже ввиду большего размера и большей длины передней полости, особенно при наличии коартикуляции с последующим задним гласным. Однако большая часть сдвига F_2 -локуса вниз в ряду [ga] [go] [gu] обусловлена увеличением степени огубления. Общим для акустических признаков звуков [g] или [k] плюс гласный может считаться, как было установлено ранее, концентрация энергии в области F_{2e} последующего гласного. Артикуляторным инвариантом является наличие полости перед местом сужения, а энергетические соотношения в спектре выявляются в момент взрыва, если он есть, и в начале переходного участка¹⁾. Среди взрывных и фрикативных степень концентрации энергии в спектре является основной характеристикой компактности, причем возможны значительные различия частоты основной форманты. Последнее лучше согласуется с использованием в качестве параметра гласных разности $F_{2e} - F_1$, а не F_1 ; более подробно этот вопрос рассмотрен в гл. 13. Отношение объемов передней полости к задней одинаково характерно для обоих этих параметров.

Некоторое значение для различения между [k] и [p] имеет еще ряд факторов, а именно: большая энергия [k], обусловленная наличием переднего резонатора; меньшее изменение резонансных частот в первом переходном интервале после взрыва, что связано с меньшим влиянием на резонансную частоту передней полости опускания языка по сравнению

¹⁾ Сообщения об опытах, проделанных в лабораториях Хаскинс [33] по восприятию синтетических слогов согласный — гласный, составленных из двухформантного гласного и начальных формантных переходов, воспроизводящих звонкий смычный, показывают, что имеет место большой сдвиг F_2 -локуса по частоте при переходе от [ga] к [gɔ], отделяющий группу [i], [e], [ɛ], [a] от группы [ɔ], [o], [u]. Эти локусы, рассмотренные Либерманом [132], тесно связаны с локусами в артикуляторном смысле, о которых идет речь у нас, но не идентичны им. Так как в синтетических стимулах, использовавшихся в этих опытах, отсутствует третья форманта и взрыв согласного, можно предполагать, что F_2 -переходы в них должны быть значительно подчеркнуты или же в каких-то других отношениях должны отличаться от переходов, встречающихся в человеческой речи. Акустические характеристики звуков не изменяются скачком от [ga] к [gɔ]. Однако хорошо известно, что близость локусов F_2 и F_3 в [g] исчезает при переходе от сочетания с задним гласным к сочетанию с передним. Как указала Фишер-Йоргенсен [59], наблюдаемые изменения являются естественными результатами коартикуляции; эти изменения были детально рассмотрены Стивенсом и Хаузом в 1956 г. Заключение Либермана [132] о том, что [g] в [ga] и [gɔ] воспринимается как одна фонема не по их акустическому подобию, а за счет подобию в их артикуляции, не представляется достаточно обоснованным. Артикуляция и акустические характеристики никогда не идут разными путями.

с раскрытием губ; наконец, наличием одной только форманты у [k] по сравнению со многими формантами у [p], что определяется различными положениями источника в том и другом случае и, следовательно, различным видом функции нулей.

4. Признак низкотональности — высокотональности определяется следующим образом: «Акустически — концентрацией энергии в нижней (или верхней) части спектра; генетически — периферическое или медиальное образование; периферические фонемы (велярные и губные) обладают более обширным и менее четко очерченным резонатором, чем соответствующие медиальные фонемы (нёбные и зубные)».

Термин «резонатор», приведенный выше, относится к ротовой полости. Артикуляторное описание наиболее высокотональных гласных как средних (с сужением в середине ротовой полости и потому скорее среднёнёбных) относится к положению языка, характерному для [i], и хорошо согласуется с номограммой трехпараметрической модели¹⁾, приведенной в § 3 гл. 4.

В случае согласных соотношения более сложны. Рассматривая консонантный интервал взрывных и фрикативных, следует отметить, что оптимальным, с точки зрения наиболее высокочастотного шума, местом артикуляции является таковое при зубных, а не при нёбных звуках. При наличии небольшого переднего резонатора имеет место более эффективное образование высокочастотного шума, чем при отсутствии такого резонатора, но увеличение длины переднего резонатора, сопровождающееся сдвигом места артикуляции назад, снижает высоту фрикативного шума.

Размеры передней полости являются, таким образом, фактором, от которого зависит как компактность, так и низкотональность и бемольность. Если две передние полости с относящимися к ним сужениями имеют одинаковую длину, то та из них, у которой площадь поперечного сечения менее отличается от площади поперечного сечения в сужении, дает менее компактный и более высокотональный спектр (см. данные, приведенные в § 3 гл. 10).

Отличия *F*-картины для низкотональных — высокотональных согласных состоят в понижении — повышении локусов F_2 или F_3 или того и другого. Последний признак важен для различения палатализованных [m] и [n], как видно из рис. 76.

5. Признак бемольный — простой определяется теми же авторами следующим образом: «Акустически бемольные фонемы, в противоположность соответствующим простым, характеризуются сдвигом вниз или ослаблением ряда высокочастотных компонент, генетически — первые (суженная щель) фонемы, в противоположность вторым (расширенная щель), образуются с уменьшенным передним или задним отверстием ротового резонатора и сопровождаются веляризацией, которая расширяет ротовой резонатор».

К этой формулировке можно добавить, что удлинение или уменьшение губного прохода всегда сдвигает вниз не только высокие форманты, но и F_1 .

¹⁾ Конфигурация ротовой полости не обязательно является важнейшим признаком, отличающим [a] от [æ]. С точки зрения размеров фаринкса приведенное выше положение справедливо в обратном смысле, т. е. звук является типично низкотональным (т. е. имеет минимальное значение $F_2 - F_1$ в трехпараметрической модели), если задняя часть языка заполняет фарингальную полость и язычный проход расположен в центре фаринкса (сравните обычные аспекты низкотональности и компактности).

Как установлено Якобсоном и др. [114], бемольность генетически связана с огублением, фарингализацией или ретрофлексией. Влияние первых двух артикуляторных переменных на F -картину гласного видно из номограмм для трехпараметрической модели, приведенных в § 3 гл. 4. Из них можно видеть, что сужение язычного прохода ведет к сдвигу F_2 по шкале частот вниз только в том случае, когда этот проход расположен в задней полости модели, т. е. в области фаринкса. Если он расположен у язычка, его сужение не оказывает влияния; обратное имеет место, если проход образуется в полости рта.

Эффект ретрофлексии не отражается в этих номограммах, однако хорошо известно, что он приводит к низкой F_3 , типичной для ретрофлексного звука [г] и ретрофлексной модификации гласных в американско-английском языке. Если ретрофлексия осуществляется в сильно продвинутом вперед месте, например в альвеолярных, то может случиться, что F_4 или даже F_5 берут на себя роль F_3 , особенно у мужчин с очень длинным речевым трактом.

Влияние фарингализации на F -картину согласных качественно такое же, как и рассмотренное выше для гласных в тех случаях, когда фаринкс становится вторичным местом артикуляции. Однако это обстоятельство вызывает лишь небольшие изменения в спектре консонантного шумового интервала фрикативных и взрывных, так как их характеристики определяются в основном конфигурацией речевого тракта в области первичного места артикуляции и перед ним.

Лабиялизация и ретрофлексия, с другой стороны, изменяют эффективную величину передней полости, где происходит фильтрация шума, и тем самым оказывают явно выраженное влияние как на его спектр, так и на F -картину, как можно видеть из сравнения рис. 7 и 8 в статье Якобсона и др. [114].

6. Признак дизный — простой определен так: «Акустически дизные фонемы, в отличие от соответствующих простых, характеризуются сдвигом вверх некоторых из высокочастотных компонент; генетически — дизные фонемы (расширенная щель), в противоположность простым (суженная щель), образуются при увеличенном фарингальном проходе, т. е. расширенном заднем отверстии ротового резонатора; сопутствующая палатализация ограничивает и отчетливо выделяет полость рта».

Признак дизный — простой в формулировке, данной выше, противоположен признаку бемольный — простой по размерам фаринкса. Признак бемольности содержит элемент, общий с признаком низкотоности — высокотоности; этим элементом является наличие большой передней полости. Фонетически признак дизный — простой связан с палатализацией, типичный эффект которой выражается в повышении F_2 и F_3 по мере того, как язык приближается к положению, характерному для образования [i]. Это можно видеть из рис. 75, относящегося к F -картинам русских согласных, вычисленным на основании использованных в этой работе рентгенограмм.

Различие между русскими твердыми и мягкими согласными на этом рисунке в некоторых случаях больше, чем это следует из результатов анализа связной речи того же диктора, показанного на рис. 76. Представленные на этом рисунке F -картины построены по приведенным в приложении спектрограммам и относятся ко всему интервалу звука [а], следующего за каждым из согласных. Следует отметить, что звуки, характеризующиеся одним и тем же местом артикуляции, т. е. все губные

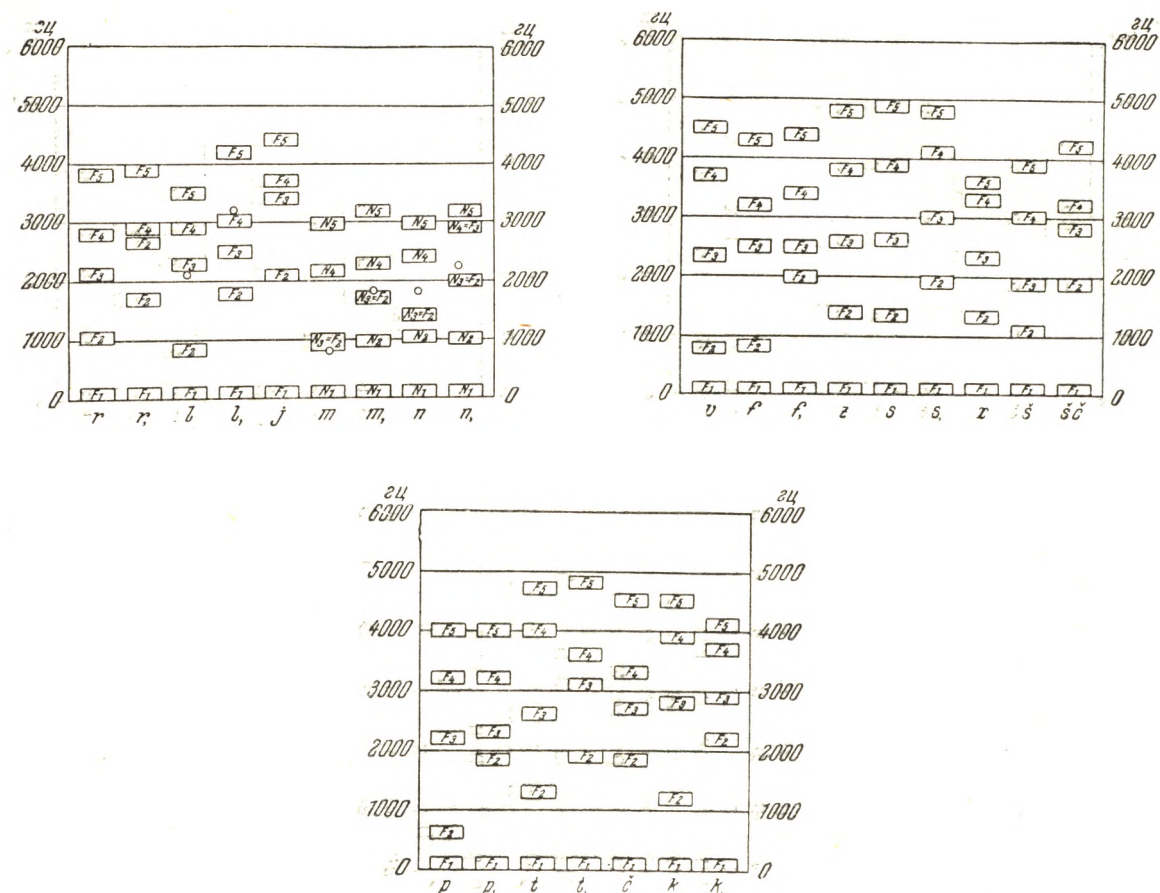


Рис. 75. Расчетные F -картины русских согласных, исследованных в настоящей работе. Дифференцировка их по значению F_1 не проводится. Нули носовых и плавных отмечены кружками.

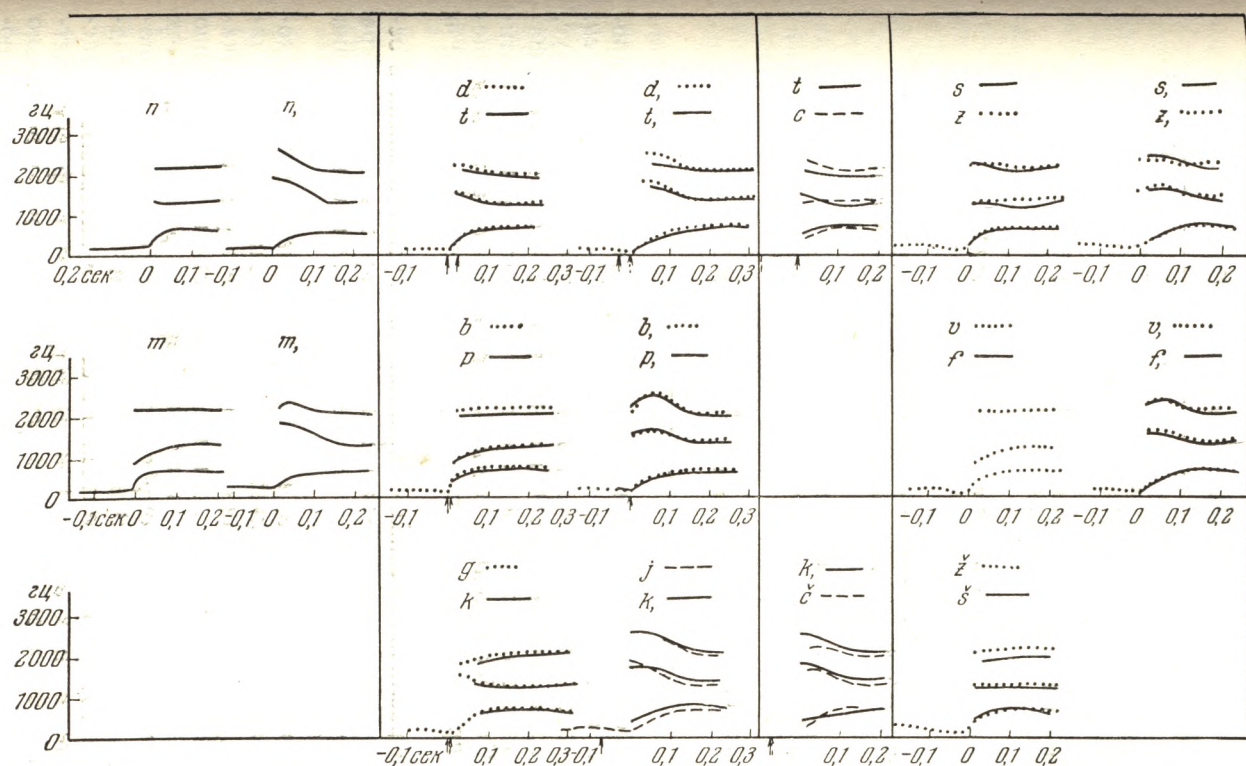


Рис. 76. Формантные переходы в гласном [а], следующем за каждым из указанных на рисунке согласных. Изменения F -картины построены по данным спектрограмм, приведенных в Приложении. Звонкие и глухие (фонематически не всегда являющиеся минимальными парами) объединены попарно и совмещены во времени в соответствии с их артикуляцией, определяющей временные изменения F -картины. Моменты, которые считались началом движения артикуляторных органов после затвора (в случае взрывных — в самом начале взрыва), отмечены стрелками на оси времени. Как можно видеть, для того, чтобы начало гласных в [d, a] и [t, a] совпало с артикуляторной точки зрения в обоих случаях, [t] должно начинаться раньше. Это обусловлено более длительным фрикативным интервалом в [t]. Противоположное явление можно наблюдать при сопоставлении [d] и [t]. Это не является типичным и может быть объяснено тем, что для русских взрывных фонематически значимо противопоставление звонкий — глухой, а не сильный — слабый. [g] и [k] начинаются одновременно, что указывает на одинаковую скорость движения артикуляторов в течение первых 70 мсек после взрыва, являющегося глухим для [k].

или все палатализованные губные, обнаруживают почти идентичные переходы в F -картине. Данные для взрывных и фрикативных можно суммировать следующим образом:

	Частота форманты	F_2 , гц	F_3 , гц
Простые губные	Расчетная	800	2250
	Измеренная	920	2025
Диезные губные	Расчетная	1800	2400
	Измеренная	1820	2400
Простые зубные	Расчетная	1370	2640
	Измеренная	1440	(Данные неточны)
Диезные зубные	Расчетная	1900	3050
	Измеренная	1700	(Данные неточны)
Простое [к]	Расчетная	1200	2800
	Измеренная	1425	2600
Диезное [к,]	Расчетная	2200	2900
	Измеренная	1800	2700

Разница наибольшая для губных, где сдвиг F_2 при переходе от простых к диезным фонемам достигает величины, близкой к 1000 гц. В этом случае у простых фонем наблюдается особенно низкий F_2 -локус, обусловленный возможностью приближения задней части языка к фаринксу; это придает особо сильную бемольность простому члену противопоставления. Подобное усиление бемольности типично также для членов такого же противопоставления по признаку диезные — простые в ряду плавных.

Акустически различие диезные — простые для русского языка в основном связано с изменениями F -картины, поскольку общая структура системы резонаторов в месте и перед местом первичной артикуляции не изменяется очень сильно.

Степень подобных первичных артикуляторных изменений в условиях действительной речи не может быть определена по данным настоящей работы, но расчетные данные, приведенные в § 3 гл. 10, показывают, что изменение конфигурации задней полости по направлению к палатализованному состоянию ведет при прочих равных условиях не только к подъему F_2 и F_3 , но сопровождается и определенным возрастанием интенсивности F_2 и F_3 . Эти форманты, типичные для переходных характеристик, должны выявляться и в шумовом интервале согласного. Анализ естественных звуков показывает, что этот эффект отчетливо выражен в губных согласных. В случае зубных он относительно мал в интервале чистой фрикации и более заметен в интервале заключительной аспирации.

7. Признак напряженный — ненапряженный определен Якобсоном и Халле следующим образом: «Акустически — большее (против меньшего) общее количество энергии в соединении с большим (против мень-

шего) рассредоточением энергии по спектру и во времени; генетически — большая (против меньшей) деформация речевого тракта по сравнению с положением покоя. Роль мышечного напряжения языка, стенок речевого тракта и голосовой щели нуждается в дальнейшем изучении».

«Рассредоточение энергии во времени» означает просто длительность звука. «Рассредоточение энергии по спектру» надо рассматривать как отклонение от нейтральной *F*-картины в случае гласных и как большую частотную протяженность и интенсивность фрикативного шума напругенных взрывных, что способствует «заполнению» высокочастотной части спектра. Правда, такая интерпретация не очень хорошо обоснована.

Длительность, как простой и весьма существенный признак, заслуживает большего внимания, чем ей обычно уделяется. В случае взрывных под длительностью надо понимать длительность взрыва, а не смычки. Если различия *p—b*, *t—d* или *k—g* есть различия типа «сильный — слабый», то неизбежно должна иметь место большая длительность интервала взрыва напругенных взрывных, считая от первого момента взрыва до появления голоса. Это, по-видимому, определяется в первую очередь состоянием голосовых связок в начале взрыва или сразу же после него. Как видно из расчетов, дополненных прямыми измерениями взрывных для звуков шведской речи (§ 2 Приложения II), постоянная времени процесса выравнивания давления при растворе смычки, а тем самым и длительность протекания струи воздуха, зависит прежде всего от объема участвующего в этом процессе воздушного резервуара.

Когда голосовые связки сближаются в момент начала их колебаний, этот объем ограничен только ротовой полостью, так что постоянная времени на порядок величины меньше, чем в том случае, когда в состав этого резервуара входит и объем легких. Статическое давление воздуха перед раствором может и не быть больше в случае сильного взрывного, чем в случае слабого; поэтому неправильно было бы полагать, что задержка момента начала вибрации голосовых связок вызывается большим избыточным давлением, увеличивающим время, необходимое для выравнивания давлений. Дополнительным фактором, приводящим к увеличению энергии сильных взрывных, является тенденция к аффрикации или, скорее, фрикации вследствие замедленного перехода артикуляторных органов от более закрытого положения в более открытое, соответствующее последующему звуку. Появление фрикативности, вызванной замедленным раскрытием сужения, свойственно не только глухим звукам. Она встречается как избыточная характеристика русских палатализованных звуков, и наблюдается даже в звонком [d_l], где колебания связок продолжают в течение всего звука, включая и смычку.

Если степень раствора в месте сужения достигает относительно большой величины до появления голоса, имеет место аспирированный конец взрыва. Характеристики возникающего при этом звука [h] отличаются от характеристик фрикативных интервалов большей подчеркнутостью низкочастотных формант, относящихся к *F*-картине, чем тех, которые лежат в более высокой области частот.

Из приведенного выше рассмотрения можно видеть, что удлинение интервала взрыва в аспирированных напругенных взрывных является в основном результатом задержки появления голоса, в отличие от одновременного появления голоса и взрыва в ненапругенных взрывных. Традиционное фонетическое противопоставление звонкий — глухой сохраняет при этом известный смысл даже тогда, когда в результате

исчезновения голоса в момент смычки различие сводится к противопоставлению напряженный — ненапряженный. «Оглушение» обычно не бывает полным при конечном положении [b], [d] или [g]; в этом случае укорочение периода смычки и сопутствующее этому удлинение предшествующего гласного квалифицируется как различие по признаку звонкий — глухой [137]. Однако все же и в этом случае на кривых интенсивности или на спектрограммах можно наблюдать большую энергию для интервалов взрыва [k], [p] и [t].

Обычно предполагается, что результатом мышечного напряжения, с которым по традиции ассоциируются напряженные взрывные, является удлинение фрикативного интервала полусмычки. Однако в случае напряженных гласных нет основания предполагать, что мышечное напряжение сказывается в заметной степени на демпфирующих свойствах стенок полостей и тем самым оказывает влияние на ширину полосы формант. Напряженность и большая длительность обуславливают артикуляцию, более удаленную от нейтрального положения; таким образом, наблюдающиеся различия в ширине формант следует считать просто результатом различной степени раствора.

Подводя итог, можно указать на следующие факторы, которые могут привести к различию между напряженными и ненапряженными звуками в процессе их образования: 1) избыточное статическое давление перед взрывом, 2) момент, когда происходит смыкание голосовых связок, необходимое для образования последующего звонкого звука, если он имеется, и 3) скорость артикуляторных движений и в особенности время пребывания артикуляторов в положении, оптимальном для образования фрикативного шума.

Первый из этих факторов влияет на уровень энергии при всех обстоятельствах, но не является обязательным в начальном положении согласного. В этом положении наибольшую роль играет второй фактор. Он в первую очередь определяет степень аспирации. Третий фактор является решающим для длительности фрикативного шума. Если длительность достаточно велика, то этот фактор является также основным генетическим фактором, обуславливающим аффрикацию, и поэтому входит и в признак «резкости», которым аффриката [č] отличается от взрывного [k] и аффриката [c] от взрывного [t].

8. Признак резкий — нерезкий определяется так: «Акустически — высокая интенсивность в противоположность низкой интенсивности шума; генетически — наличие острой или закругленной кромки».

Исходя из приведенных выше соображений о напряженности, в этом определении следовало бы учесть фактор длительности шума, позволяющий отличить аффрикаты от взрывных [81, 82]. Относительная интенсивность различных шумовых источников речи требует дальнейшего исследования.

ПРИЛОЖЕНИЯ

I. АНАЛИЗ РЕЧЕВОЙ ВОЛНЫ

§ 1. Измерение интенсивности

Любое описание речевого сигнала, вне зависимости от того, является оно осциллографическим или спектрографическим, требует определения амплитуд или интенсивностей. Под амплитудой при этом могут подразумеваться как мгновенные, так и усредненные во времени значения давления, объемной скорости, скорости частиц в определенной точке звукового поля, а также значения токов и напряжений на выходе микрофона. Понятие амплитуды распространяется и на численные значения усиленной, отфильтрованной или каким-либо иным способом преобразованной электрической «копии» речевого сигнала.

Интенсивность звука определяется звуковой энергией, проходящей за единицу времени через единицу площади. В системе CGS, широко принятой в литературе и использованной в настоящей работе, единицей интенсивности является $\text{эрг/сек} \cdot \text{см}^2$. Этому соответствует 10^{-7} вт/см^2 . Для свободно распространяющейся плоской или сферической звуковой волны интенсивность выражается так:

$$W = \frac{P^2}{\rho c} \frac{\text{эрг} \cdot \text{сек}^{-1}}{\text{см}^2}, \quad (1.1)$$

где P — эффективное значение звукового давления в дин/см^2 , ρ — плотность среды в г/см^3 , c — скорость распространения звука в см/сек . Произведение ρc представляет собой удельное акустическое сопротивление среды и равно для воздуха $41,4 \text{ дин сек/см}^3$ при температуре 20°C и $40,0 \text{ дин сек/см}^3$ при 35°C ; последнее значение наиболее отвечает условиям распространения звука в полостях речевого аппарата.

Интенсивность речевых сигналов почти никогда не измеряется непосредственно. При помощи микрофона измеряется звуковое давление, а интенсивность, если это нужно, определяется по формуле (1.1).

Данные о значениях звукового давления или интенсивности обычно представляются в логарифмической шкале, т. е. в децибелах (дб) по отношению к начальному логарифмическому уровню, принимаемому за 0. Стандартное звуковое давление P_0 , соответствующее этому нулю, принимается равным $0,0002 \text{ дин/см}^2$, чему соответствует интенсивность W_0 , близкая к 10^{-16} вт/см^2 .

Логарифмический уровень интенсивности определяется равенством

$$L = 10 \log_{10} \left(\frac{W}{W_0} \right) \text{ дб}. \quad (1.2)$$

Логарифмический уровень звукового давления будет

$$L = 20 \log_{10} \left(\frac{P}{P_0} \right) \text{ дб}. \quad (1.3)$$

Эти уровни равны между собою при условии, что P_0 связано с W_0 соотношением (I.1), так же как и P и W .

В акустике, в частности при определении чувствительности микрофонов, за нулевой уровень звукового давления иногда принимается величина $P_0 = 1 \text{ дин/см}^2$. Эта величина соответствует приблизительно среднему давлению при речи на расстоянии 30 см от рта говорящего.

Оценивая пригодность того или иного прибора для измерения интенсивности, следует принимать во внимание следующие его свойства:

- 1) Частотную коррекцию, если таковая предусмотрена.
- 2) Тип выпрямления: а) одно- или двухполупериодное; б) линейное, квадратичное или промежуточное между этими двумя видами.
- 3) Время интегрирования и вид весовой функции, входящей в интеграл.
- 4) Амплитудную характеристику, которая может быть линейной, логарифмической или промежуточной по степени амплитудной компрессии.

Эти свойства целесообразно было бы рассматривать с точки зрения соответствия их свойствам слуха. Однако соотношения между физическими характеристиками сложных, изменяющихся во времени звуков и вызываемыми такими звуками ощущениями громкости весьма сложны. Вряд ли они могли бы быть реализованы в измерителе громкости, приемлемом для эксплуатации. Кроме того, психоакустические законы восприятия таких стимулов, как звуки речи, недостаточно хорошо изучены. Технические соображения по вопросу стандартизации измерителей уровня интенсивности звука приведены у Харди и др. [88].

В состав измерителей интенсивности звука часто включают специальный каскад с особой формой частотной характеристики, имеющей целью придать больший удельный вес тем спектральным составляющим, которые расположены в средней (от 1000 до 4000 гц) области частот.

Измерители подобного типа, применяемые для определения интенсивности шумов, описаны, например, у Беранека [7]; в них предусматриваются три вида частотной коррекции, позволяющие придать прибору одну из трех частотных характеристик, для краткости называемых далее шкалами A , B и C .

Шкала A предназначается для измерения звуков, уровень интенсивности которых лежит в пределах 20—55 дб; реализующая такую характеристику корректирующая схема имеет максимальное значение коэффициента передачи при 2500 гц и затухание на низких частотах, достигающее 25 дб при 100 гц. Небольшим уменьшением коэффициента передачи на высоких частотах в случае шкалы A можно пренебречь, если иметь в виду применение прибора для фонетических целей.

Шкалу B рекомендуется применять при измерении уровней, лежащих в пределах $L = 55—85 \text{ дб}$. Эта шкала является средней между шкалой A и равномерной шкалой C , и притом ближе к последней; для нее подавление низких частот по отношению к частоте 1000 гц составляет всего 6 дб.

Диапазон интенсивностей человеческой речи составляет около 30 дб. Гласные, находящиеся под главным ударением, имеют в среднем уровень примерно 65 дб на расстоянии 1 м от рта говорящего [40]; глухие согласные в среднем на 20 дб менее интенсивны [46].

Из этого можно было бы сделать вывод, что при измерении интенсивности глухих согласных следует пользоваться шкалой A , а при измерении интенсивности гласных — шкалой B ; однако в ряде случаев

критерий громкости, исходя из которого выбраны эти характеристики, нельзя считать решающим. Иногда желательно получить данные об интенсивности, которые связаны не столько с фонетическим качеством гласных, т. е. с их F -картиной, сколько с речевым усилием говорящего. В этом случае предпочтительнее пользоваться шкалой S , с тем чтобы не ослаблять влияние первой форманты, в основном определяющей суммарную громкость. Еще правильнее было бы применить интегрирование исходного процесса, что равносильно подъему низких частот. Идеальным явилось бы полное подавление формант, с тем чтобы восстановить характеристики самого источника. Возможность применения такого метода в настоящее время исследуется.

Тип выпрямления менее важен, чем частотная характеристика прибора. Предпочтительно двухполупериодное выпрямление, которое не приводит к потере данных либо об отрицательной, либо о положительной полуволне. Разница в величинах этих полуволн значительна, в частности для низких мужских голосов, особенно на границах раздела звуков речи. Эта разница становится еще более значительной, если характеристика выпрямителя нелинейная.

О разнице в показаниях прибора при использовании квадратичного и линейного выпрямления можно судить по результатам суммирования группы спектральных составляющих; в случае квадратичного выпрямления результат точно равен сумме квадратов составляющих.

Независимо от типа выпрямления приборы всегда можно проградуировать так, чтобы при синусоидальной форме волны он давал одно и то же показание; однако при измерении интенсивности сложного комплекса составляющих, подобного речи, проградуированный таким образом прибор с линейным выпрямителем дает в среднем показания, меньшие на 0—3 дБ по сравнению с прибором, обладающим квадратичным выпрямителем. Наибольшие отклонения наблюдаются, если звук содержит много незначительно различающихся по амплитуде составляющих.

Простой опыт для проверки закона суммирования состоит в том, что к прибору подводятся две синусоидальные составляющие одинаковой амплитуды; квадратичный прибор покажет на 3 дБ больше, чем для одной составляющей, а линейный — на 2 дБ. Такая же разница в 1 дБ между показаниями наблюдается и при измерении шума ([7], стр. 453; см. также комментарии Сноу [186]).

При отсутствии устройства, извлекающего квадратный корень, прибор с квадратичным выпрямителем дает показания, соответствующие непосредственно интенсивности. Вероятно, квадратичное выпрямление в большей мере соответствует свойствам слуха, чем линейное; однако разница не очень велика и, как правило, может быть учтена при градуировке прибора¹⁾. Нужно учитывать, кроме того, что показания линейного прибора зависят от фазовых соотношений между составляющими [79]. Несмотря на все это, в большинстве случаев практически предпочтительнее линейное выпрямление, так как оно проще осуществимо и требует меньшей степени компрессии в конечной ступени прибора.

Если осциллограф включается непосредственно на выход выпрямляющего каскада измерителя интенсивности, минуя усредняющую

¹⁾ Квадратичное выпрямление, конечно, не является идеальным с точки зрения слухового восприятия; прибор, измеряющий пиковые значения, может быть не менее ценен, поскольку пиковые значения являются коррелятами громкости сложного широкополосного звука.

схему, то получается осциллограмма мгновенной интенсивности. Она отличается от обычной осциллограммы тем, что все отрицательные отклонения от нулевой линии изменяют свой знак; в случае квадратичного выпрямления имеет место также расширение шкалы амплитуд по мере увеличения отклонений. Очевидно, что такая осциллограмма практически мало пригодна, так как в ней отражены многие лишние структурные детали. Усреднение или сглаживание, выполняемое низкочастотным фильтром, расположенным после выпрямителя, аналогично определению площади, заключенной между несглаженной кривой мгновенной интенсивности и нулевой линией, за определенный промежуток времени, например за один период основной частоты голоса. Такая графоаналитическая операция предполагает использование прямоугольной весовой функции; при этом любой ординате в пределах выбранного отрезка времени придается одинаковый вес, т. е. одинаковая значимость. Однако подобная характеристика усреднения не характерна для слухового восприятия, так же как и для любого реального фильтра. И в том и в другом случае доля, вносимая в окончательный результат предшествующими участками процесса, определяется некоторой «функцией памяти». Эта функция памяти вдобавок, вероятно, зависит от типа стимула и от контекста.

В силу этого каждая точка усредненной кривой интенсивности должна соответствовать сумме мгновенных значений интенсивности для некоторого временного отрезка речевой волны, не имеющего резких границ. Эффективная величина этого отрезка, называемая временем интегрирования или усреднения, может быть определена как площадь, заключенная между кривой «функции памяти» и осью абсцисс, деленная на ординату центра тяжести этой площади [128]. Для большого класса низкочастотных фильтров LC абсцисса центра тяжести недалека от абсциссы, отвечающей максимальному значению. Определяемое ею расстояние во времени относительно момента наблюдения называется «временем задержки» и является одной из временных характеристик интегрирующей схемы. В классе сглаживающих схем, к которым относятся простые низкочастотные фильтры RC и которые характеризуются экспоненциальной функцией памяти

$$h(t) = h_0 \cdot e^{-t/RC}, \quad (1.4)$$

пиковое значение «функции памяти» $h_{\max}$ имеет место в момент наблюдения; время задержки, в соответствии с определением Лоурента [128], равно $T_d = RC$ и время усреднения равно $T_a = eRC$. В общем случае время усреднения имеет порядок

$$T_a = \frac{1}{2B}, \quad (1.5)$$

где B — граничная частота низкочастотного фильтра.

Вообще говоря, время задержки не играет никакой роли при измерении интенсивности, за исключением того случая, когда сопоставляются данные, записанные на многоканальном осциллографе после различных преобразований; при этом, очевидно, следует сопоставлять синхронизованные временные отрезки, т. е. должны учитываться времена задержки, свойственные разным видам преобразований.

Если вместо осциллографической записи значения интенсивности отсчитываются по отклонению стрелки прибора, интегрирование происходит в силу механической инерции подвижной системы прибора.

Выпрямитель и низкочастотный фильтр RC с большой постоянной времени могут быть рассчитаны таким образом, чтобы прибор давал пиковые значения измеряемой величины. Показания такого прибора достигают максимального значения очень быстро, за время порядка периода основной частоты речевого звука, но спадают медленно. Поэтому прибор в течение достаточно большого времени не реагирует на пиковые значения амплитуды меньшие, чем то пиковое значение, которым был обусловлен первоначальный заряд емкости в схеме. В отрезке связанной речи длительностью около 3 сек наивысший мгновенный пик примерно на 20 дБ превышает среднее за длительное время; это среднее получается путем деления величины энергии, соответствующей всему отрезку, на его длительность [40].

Инерцию слухового восприятия, которую Джуз [121] обозначает термином «*smear*» (размазывание), нельзя достаточно точно описать, используя постоянные времена указанного выше типа, однако можно полагать, что время задержки и время усреднения лежат в пределах 20—200 мсек. Эксперименты по восприятию тональных импульсов разной длительности показали, что увеличение длительности импульса свыше примерно 0,18—0,25 сек не дает прироста ощущения громкости. Удлинение импульса сверх этой величины дает только ощущение большей протяженности, тогда как при оценке громкости более коротких импульсов играют роль как интенсивность, так и длительность импульса (см., например, [4]). Относительное значение каждого из этих факторов изменяется с изменением уровня стимула [156] и различно для различных испытуемых [77]. При восприятии коротких шумовых импульсов время усреднения составляет около 50 мсек [151].

В вопросе о постоянной времени слуха многое еще остается неясным. По-видимому, различные аспекты речевого процесса воспринимаются с различной инерцией слуха, так что не представляется возможным охарактеризовать восприятие речи одной постоянной времени. Часть информации достигает высших центров очень быстро, по коротким и прямым нервным путям, другие данные подвергаются сложным преобразованиям, происходящим со значительной задержкой во времени и усреднением. Так, время восприятия для звуковых импульсов большой интенсивности меньше, чем для слабых импульсов.

Прибор для измерения интенсивности типа стандартного VU -метра имеет время усреднения около 250 мсек, что в среднем несколько больше длительности звонкой части слога; этой части слога соответствуют показания на 2 дБ большие, чем среднее за длительное время [16]. Однако, учитывая неопределенность постоянной времени для слуха, нецелесообразно использовать в приборах для измерения интенсивности слишком большое время интегрирования. При достаточно малой постоянной времени прибора всегда возможно путем графического интегрирования получить результат, соответствующий желаемой степени «размазывания». Надо, однако, принимать во внимание то обстоятельство, что если время интегрирования будет выбрано меньше $1/F_0$, где F_0 — основная частота голоса, то на записи кривой интенсивности будет наблюдаться волнистость, соответствующая этой частоте. Такая волнистость не является помехой, пока она мала по сравнению со средней интенсивностью, она может быть даже полезной, так как позволяет определить основную частоту F_0 . Время интегрирования 10 мсек, соответствующее согласно формуле (I.5) сглаживающему фильтру с граничной частотой 50 гц, является удовлетворительным. Оно находится

в соответствии с определением «средней мощности речи», введенным Флетчером [70]. Время интегрирования в 20 мсек почти полностью исключает волнистость, обусловленную основной частотой голоса, даже для наиболее низкочастотных отрезков речи, но дает чрезмерно большое размытие, которое не позволяет правильно судить о характере нарастания и спада интенсивности за время раскрытия смычки у взрывных.

При оценке сглаживающего устройства следует учитывать, помимо времени интегрирования, и форму функции памяти. Если таким устройством является пассивный фильтр RC , то функция памяти отличается значительно большей скоростью нарастания, чем спада. Если на такой фильтр попадает настолько короткий импульс (например, в момент взрыва короткого взрывного звука), что длительность его меньше времени усреднения, то результирующая кривая будет отражать только функцию памяти самого фильтра¹⁾.

Градуировка регистрирующего прибора дается в виде таблицы или в виде графика, позволяющего переводить амплитуды записываемой прибором кривой, отсчитываемые в миллиметрах, в уровни звукового давления в децибелах.

Градуировка в децибелах является естественно обоснованной в том случае, если в оконечном звене прибора, после интегрирующей схемы, предусмотрено логарифмирующее устройство; однако подобная градуировка возможна и в том случае, когда шкала прибора линейна. Можно использовать также некоторый компромисс между логарифмическим и линейным представлением; такая смешанная шкала реализуется при помощи очень простого компрессора, который дает большие и средние амплитуды в логарифмическом, а малые — в линейном масштабе. Кстати говоря, такая шкала отражает свойства слуха даже лучше, чем чисто логарифмическая, так как дифференциальный порог по интенсивности, т. е. наименьшая воспринимаемая разность уровней интенсивности в $дб$, больше у порога слышимости, чем при более высоких уровнях. Величины дифференциального порога (ДП) лежат в пределах 0,5—5 $дб$. Для гласных ДП, по данным Фланагана [65, 66], близок к 1 $дб$. Согласно Стивенсу [197], увеличение уровня интенсивности звука на 10 $дб$ ведет к удвоению ощущения громкости. Эти цифры дают возможность оценить точность, необходимую при получении данных об интенсивности звука.

В качестве слухового коррелята удара было предложено использовать величину площади, охватываемой слоговым пиком на записи при линейной шкале амплитуд. Получаемая при этом величина является мерой как интенсивности, так и длительности [46, 53]²⁾. Целесообразность использования такого параметра обосновывается, с одной стороны, данными о том, что громкость зависит не только от интенсивности, но и от длительности звука, с другой же стороны — опытами по анализу и синтезу речи, согласно которым одного только укорочения длительности достаточно для изменения восприятия ударенности. Пред-

¹⁾ Практика Лаборатории передачи речи показала целесообразность применения фильтров LRC с фазовой компенсацией, имеющих выше граничной частоты затухание 18 $дб/окт$. Функция памяти такого фильтра достаточно симметрична и обеспечивает отсутствие заметного всплеска. Расчетные данные такого фильтра приведены в другой работе автора [55].

²⁾ Другими коррелятами удара являются увеличение основной частоты голоса F_0 и увеличенный контраст между гласным и согласным в ударном слоге [53].

лагаемый параметр, для которого рекомендуется наименование «индекс импульса», имеет размерность энергии только при квадратичной характеристике выпрямителя. При линейном выпрямлении индекс имеет, строго говоря, размерность звукового давления, умноженного на время, хотя в общем энергетический смысл его сохраняется и в этом случае. Практически целесообразно индекс импульса выражать в дб , т. е. исходить из относительной величины площади. При квадратичном законе выпрямления увеличение интенсивности слога на 3 дб , т. е. удвоение амплитуды, эквивалентно с точки зрения предлагаемого параметра удвоению длительности слога. В случае линейного закона выпрямления и линейной шкалы прибора увеличению интенсивности будет придаваться меньший вес, так как удвоению амплитуды будет соответствовать увеличение на 6 дб . Однако пока еще нельзя дать уверенных рекомендаций относительно наиболее подходящего вида шкалы и целесообразной формы представления данных о величине индекса ввиду недостаточности накопленного опыта; пока представляются практически более приемлемыми линейный закон выпрямления и линейная шкала записи на регистрирующем приборе.

§ 2. Измерение спектра и формы волны

Используя существующую в настоящее время технику спектрографического и осциллографического анализа речи, можно получить следующие данные:

- а) мгновенную амплитуду;
- б) интенсивность, усредненную за небольшой промежуток времени порядка $10\text{--}20 \text{ мсек}$;
- в) спектральный состав в одной из двух возможных форм: в виде ряда Фурье, т. е. амплитуд и фаз основной частоты и гармоник; в виде амплитуд и фаз на выходах достаточно большого числа полосовых фильтров, перекрывающих весь представляющий интерес диапазон частот.

Под фазой здесь подразумевается мгновенная фаза $\varphi_m(t)$ напряжения $v_m(t)$ на выходе фильтра номера m при частоте, равной средней частоте фильтра $F_m = \frac{\omega_m}{2\pi}$. Это напряжение определяется формулой

$$v_m(t) = V_m(t) \cos [\omega_m t + \varphi_m(t)], \quad (\text{I. 6})$$

в которой $V_m(t)$ характеризует амплитудную информацию, причем предполагается, что $V_m(t)$ и $\varphi_m(t)$ мало изменяются за время одного периода $2\pi/\omega_m$.

Колебание с частотой F_m можно рассматривать как несущую, изменяющую амплитуды и фазы которой содержат информацию о сигнале. После выпрямления и сглаживания остается только амплитудная информация $V_m(t)$. При этом наблюдаемая величина $V_m(t)$ является средним значением амплитуды за время $1/B_m$, где B_m — ширина полосы анализирующего фильтра. Время это обусловлено интегрирующим действием фильтра и отличается от времени интегрирования низкочастотного фильтра, определяемого формулой (I.5), только отсутствием множителя $1/2$.

Фазовая информация имеет весьма ограниченное значение для восприятия речи и, как правило, не сохраняется в существующих анализаторах. Следует в связи с этим отметить, что фазовая информация

связана с амплитудной и может быть восстановлена при наличии последней. Частично справедливо и обратное, так что частотно-фазовый анализ может до некоторой степени заменить частотно-амплитудный [104]. В силу этой взаимосвязи данные о фазах имеют по преимуществу теоретическое значение. Большой зависящий от частоты сдвиг фаз в передающей или регистрирующей системе приводит к задержке во времени прихода спектральных составляющих; если, кроме того, зависимость сдвига фаз от частоты нелинейна, имеет место различная задержка для различных частотных участков спектра.

Для очень коротких щелчков зависимость интенсивности от частоты такая же, как и для белого шума, а именно уровень интенсивности в частотной полосе, равной единице, один и тот же для любой части спектра. Однако спектр щелчка отличается от спектра случайного белого шума по соотношению фаз спектральных составляющих: для первого фазы всех составляющих одинаковы, для второго — распределены по случайному закону. Отсюда можно было бы сделать вывод о необходимости измерения фаз; однако в силу приведенных выше соображений совершенно достаточно ограничиться данными об амплитудах и временах прихода спектральных компонент с шириной полосы B и длительностью $1/B$, которые распределены в плоскости частота — время при спектрографическом представлении. В случае воздействия бесконечно короткого импульса отклик фильтра совпадает с его функцией памяти, эффективная длительность которой равна $1/B$. Только в том случае, если длительность шума меньше $1/B$, спектрограф не сможет отличить щелчок от шумового импульса.

Графический анализ по Фурье, или так называемый гармонический анализ, в настоящее время используется очень редко, так как анализ с помощью фильтров значительно проще и быстрее приводит к цели. На более ранних этапах развития экспериментальной фонетики гармонический анализ сыграл большую роль [26, 189, 131, 187, 201]. Такому анализу эквивалентен анализ при помощи идеального фильтра, имеющего бесконечно узкую полосу пропускания. Представление процесса при помощи спектральных составляющих, определенных в частотно-временной области, не только представляет собою наиболее подходящий способ описания работы спектрографа; оно лучше, чем ряды или интеграл Фурье, отражает особенность слуха [74, 75]. Однако значения эффективной ширины полосы и постоянной времени требуют уточнения с точки зрения особенностей восприятия тех или иных сторон спектральной картины; то же относится и к форме функции памяти (см. § 1 Приложения I).

Джуж считал более правильным применение для спектрального анализа узкополосных фильтров и обосновывал это характерным для слуха, по его данным, временем «размазывания», равным 50 мсек. Надо, однако, учитывать, что эффект «размазывания» при таком анализе зависит не только от свойств фильтров, но и от свойств остальных элементов системы, следующих за фильтром; именно эти элементы, при соответственном их расчете, могут определять основную часть времени интегрирования. Представляется более вероятным, что слуховой аппарат работает, как широкополосный анализатор, выходы «фильтровых каналов» которого выпрямляются, сглаживаются и взаимно коррелируются в нервной системе [135]; в результате трех последних операций время интегрирования значительно превосходит величину, обратную ширине полосы самих «фильтров» периферического органа и

составляет 20—200 мсек. Имеются также указания на то, что передача информации по многочисленным нервным путям характеризуется различными постоянными времени [74]. Указанные особенности спектрального анализа относятся к слуховому восприятию качественных, или тембровых, характеристик различных звуков речи. Восприятие основной частоты, с которой связаны интонационные характеристики речи, обусловлено если не полностью, то во всяком случае частично иным слуховым процессом. Если использовать по-прежнему аналогию свойств слухового аппарата со свойствами системы широкополосных фильтров, то процесс этот представляет собою подсчет числа максимумов энергии за единицу времени во всех или некоторых фильтровых каналах, или, иными словами, определение периодичности огибающей в этих каналах. Следует, конечно, подчеркнуть, что аналогия с системой фильтров не претендует на то, чтобы объяснить действительный механизм слухового восприятия; тем не менее для понимания некоторых особенностей работы слухового аппарата такая аналогия полезна.

В одной из разновидностей приборов для спектрального анализа, относящихся к типу спектрографов, амплитудная информация в различных частотных полосах выдается почти одновременно во всех полосах за счет быстрого сканирования их; к такой разновидности относится спектрограф RIT [200]. В другой разновидности приборов того же типа используется повторное воспроизведение анализируемого процесса, причем в составе анализатора имеется только один фильтр, средняя частота которого сдвигается небольшими ступенями по окончании каждого цикла воспроизведения; эта разновидность представлена «Сонаграфом» фирмы Кэй Электрик, прообразом которого является звуковой спектрограф лабораторий фирмы Белл [6].

«Сонаграф» дает трехмерное, в координатах время — частота — интенсивность, представление процесса длительностью до 2,4 сек; такое представление называется «спектрограммой». Горизонтальная ось на спектрограмме является осью времени, вертикальная — осью частот, а интенсивность различных спектральных составляющих представлена степенью почернения; в приборе предусмотрена возможность получения в желаемый момент времени так называемых «разрезов», под которыми подразумевается зависимость уровня амплитуды, или, точнее, уровня интенсивности, от частоты. При получении как спектрограммы, так и разреза может быть использован либо узкополосный фильтр, для которого $B=45$ гц, либо широкополосный, для которого $B=300$ гц. Спектральный разрез дает спектр некоторого отрезка процесса за время, определяемое временем усреднения интегрирующих элементов схемы, включая и полосный фильтр; положение этого отрезка во времени определяется моментом, соответствующим его середине. В случае узкополосного фильтра величина отрезка определяется в основном шириной полосы фильтра и составляет примерно 30 мсек. При широкополосном фильтре величина отрезка определяется главным образом интегрирующими элементами, следующими после фильтра, и приблизительно равна 5 мсек.

Примеры спектрограмм, полученных с помощью «Сонаграфа», представлены на рис. 77 (вклейка между стр. 240 и 241). Результаты узкополосного анализа представлены на рисунке спектрограммой А и разрезом С. Такой анализ выявляет отдельные гармоники, образующие в формантных областях на спектрограмме тонкую структуру почти горизонтальных линий; на разрезе гармоники представлены отдельными

пиками. Ширина линий на спектрограмме, представляющих гармоники, равна ширине полосы пропускания фильтра.

Частоту форманты, ее ширину и уровень можно найти, если на спектральном разрезе провести плавную огибающую, охватывающую отдельные гармоники в областях спектральных максимумов [52]. Обычно частота форманты определяется как частота, соответствующая максимуму на огибающей; однако такое определение не является достаточно однозначным. Трудности возникают при сильно несимметричной форме огибающей в формантной области и при высокой основной частоте голоса F_0 . Предложение Поттера и Стейнберга [175] считать за частоту форманты центр тяжести площади, охватываемой огибающей в области форманты, хотя может быть и отражает в некоторой мере свойства слуха, однако не решает задачу. Более показательные для описания F -картины данные могут быть получены, если определять непосредственно резонансы фильтровой функции речевого аппарата (§ 2 гл. 2 и § 2 гл. 3). Уровень формант также обычно определяется по огибающей как уровень звукового давления в децибелах¹⁾ в точке максимума. Уровень форманты можно также определить как сумму интенсивностей гармоник, входящих в состав формантной области. Если основная частота голоса F_0 больше, чем ширина форманты, то разница между определенным таким образом уровнем интенсивности и уровнем максимума огибающей не превышает 1 дб. Эти соотношения будут более подробно рассмотрены в отдельной работе. Для определения ширины полосы форманты нужно провести прямую, параллельную оси частот, на 3 дб ниже максимума огибающей. Тогда ширина полосы равна разности частот, соответствующих точкам пересечения этой прямой с огибающей.

На широкополосной спектрограмме, если только основная частота голоса не очень велика, отдельные гармоники не выявляются. Это объясняется тем, что фильтр с широкой полосой пропускает сразу несколько гармоник. Однако, поскольку большая ширина полосы B означает малое время интегрирования $1/B$, спектрограмма отражает колебания интенсивности с периодом основной частоты, возникающие в результате суммирования пропускаемых фильтром гармоник. Эти колебания дают на спектрограмме для низкого мужского голоса вертикальные штрихи, расстояние между которыми соответствует периоду колебаний голосовых связок. Каждый толчок воздуха, создаваемый гортанью при работе голосовых связок, возбуждает затухающие колебания в полостях речевого аппарата.

С математической точки зрения форманта, представленная в частотной области, идентична затухающему колебанию во временном представлении. Если бы это было достаточно ясно пятьдесят лет назад, общепринятая классическая теория образования звуков речи Гельмгольца и теория Германна-Виллиса не противопоставлялись бы друг другу; на это было указано еще Рэлеем [176], далее по этому вопросу см. [24, 202].

Кажущаяся ширина форманты при широкополосном анализе представляет собой сумму действительной ширины форманты и ширины полосы пропускания фильтра анализатора, причем последняя

¹⁾ Петерсон и Барни [167] используют термин «амплитуда» для обозначения величины формантного пика в децибелах. Однако если только определены единица и метод измерения, само наименование получаемой величины не представляет особенно существенным.

значительно превосходит первую, во всяком случае для первых трех формант.

Если средняя основная частота голоса F_0 у данного диктора высока и такого же порядка, что и ширина полосы фильтра анализатора, то спектрограмма покажет отдельные гармоники в тех интервалах наибольших F_0 , которые соответствуют интонационному повышению основной частоты голоса, но в интервалах низких F_0 чисто формантная структура может оказаться смешанной с тонкой структурой спектра в виде отдельных гармоник. Чтобы в подобных случаях избежать недоразумений, можно рекомендовать снять дополнительно спектрограмму при скорости носителя вдвое меньшей по сравнению со скоростью при записи анализируемого отрезка речи; при этом F_0 будет снижено в два раза, и формантная структура выявится значительно отчетливее; правда, одновременно с F_0 будут снижены вдвое и частоты формант, что равносильно понижению частотной избирательности анализатора, так что форманты, расположенные недалеко друг от друга, могут оказаться недостаточно ясно разделенными.

Последние разработки в области приборов для получения спектральных данных описаны в одной из прежних работ автора [53]. Одним из очень точных, но трудоемких способов получения спектральных разрезов является снятие осциллограмм на выходах некоторого числа полосных фильтров, перекрывающих нужный диапазон частот, или же на выходе одного фильтра, средняя частота которого смещается после каждого цикла воспроизведения.

Этот способ был использован для получения разрезов, приведенных ниже; с такими разрезами и сравнивались результаты расчетов. Такой же способ ранее был успешно использован при анализе взрывных звуков речи [46].

Одним из давно применяемых методов спектрального анализа речи является анализатор, основанный на принципе «ищущего тона». Такой анализатор обеспечивает постоянство ширины полосы пропускания, центр которой как бы непрерывно скользит вдоль всего исследуемого диапазона частот. Приборы, основанные на этом принципе и использованные, например, Совиярви [187, 188] и Барчинским и Тинхаузом [2], требовали для проведения анализа около двух минут. Поддержание

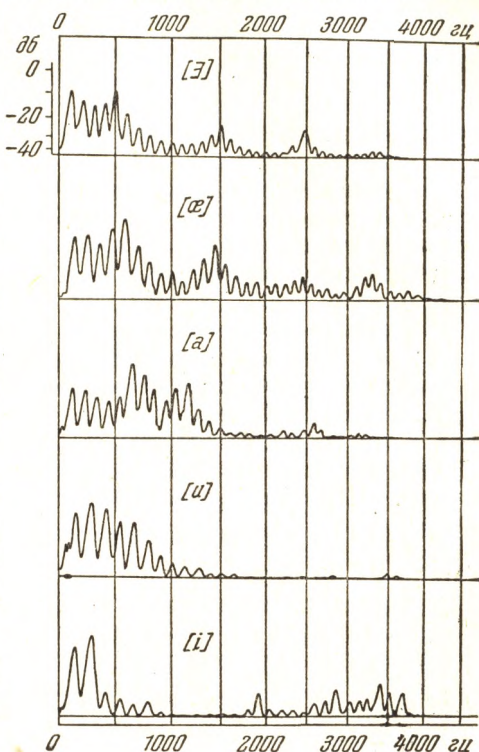


Рис. 78. Спектры синтезированного нейтрального гласного и некоторых естественных протяжно произнесенных гласных, полученные при помощи гетеродинного анализатора. Диктор — швед Г. Ф. Ширина полосы анализатора 31 гц. Скорость изменения несущей — 4000 гц за 3 сек. Запись на «Миннографе», без подъема высоких частот. Видны отдельные гармоники и формантная структура спектра.

стационарного звука в течение столь длительного времени было возможно только для хорошо тренированных певцов. В настоящее время [149, 53] этот метод улучшен настолько, что отпала необходимость затягивать произнесение звука на время большее, чем теоретически необходимый минимум; этот минимум равен отношению всего используемого диапазона частот к квадрату ширины полосы применяемого фильтра. Так, анализ в диапазоне частот до 8000 *гц* требует всего 2 *сек* при фильтре с шириной полосы 63 *гц*, и 0,5 *сек* при фильтре с полосой 125 *гц*. Уменьшать время анализа более чем до 0,5 *сек* нецелесообразно, так как при этом тонкая временная структура последовательных периодов основной частоты голоса начинает сильно искажать результаты анализа.

Примеры анализа гласных, приведенные на рис. 78, иллюстрируют хорошее качество спектральных разрезов¹⁾, получаемых при использовании на выходе анализатора пишущего осциллографа с высокой собственной частотой. Подобный метод особенно подходит для исследования качества голоса.

С технической точки зрения целесообразным является определение частоты и ширины полосы форманты не по спектральным данным, а по временным. Если дана осциллограмма речевой волны, предварительно отфильтрованной так, чтобы выделить данную форманту, то частоту форманты можно измерить по величине периода полученной кривой. Начальная амплитуда этой кривой является мерой интенсивности форманты. Ширина форманты определяется на основании равенства $B = \frac{\sigma}{\pi}$, где $1/\sigma$ — отрезок времени, в течение которого колебания затухают на 8,6 *дб*, т. е. в *e* раз (рис. 79). Получаемые таким образом данные имеют то преимущество, что соответствуют определению *F*-картины. Эти данные в общем хорошо согласуются с получаемыми по спектральным разрезам; однако ширина формант в последнем случае получается больше, чем в первом, который с физической точки зрения дает более правильные данные, поскольку разложение в ряд Фурье

¹⁾ Кривые на рис. 78 были получены без подъема на высоких частотах. Было установлено, что простой высокочастотный фильтр *RC*, характеристика которого определяется нулем при 200 *гц* и полюсом при 5000 *гц* на отрицательной вещественной оси, дает практически достаточную коррекцию падения высоких частот, типичного для большинства звуков речи, и позволяет поэтому лучше видеть высокочастотные форманты.

Достоинство такой коррекции — ее простота и симметричность относительно 1000 *гц*; она представляет собой компромисс между задачей скомпенсировать неравномерность спектра речи за длительное время и в то же время учесть частотную характеристику восприятия громкости; указанный фильтр на частотах ниже 3000 *гц* дает кривую, близкую к кривой равной громкости 40 фон. По нашим данным она лучше выравнивает спектр речи, чем подъем высоких частот, принятый в «Сонаграфе». Для представления типовых спектров желательно было бы также пользоваться шкалой «МЭЛ» [199], с тем чтобы зрительное восприятие спектрального распределения учитывало важность различных частот для слуха. Линейная шкала слишком преувеличивает удельный вес высоких частот, а логарифмическая — низких. Шкала Кенига [122], являющаяся приближением к шкале «МЭЛ», линейна до 1000 *гц*, а выше — логарифмична. Лучшим приближением к шкале «МЭЛ» явилась бы шкала, определяющаяся соотношением $x = k \log \left(1 + \frac{f}{1000} \right)$, где *x* — координата по оси частот в сантиметрах [46]. Ее преимущество — отсутствие разрыва непрерывности при 1000 *гц*. Так как большинство приборов для спектрального анализа имеет частотную шкалу либо линейную, либо логарифмическую, то во избежание кропотливого труда по пересчету при построении спектральных распределений, обычно сохраняют исходную частотную шкалу, принятую в данном приборе.

или анализ при помощи фильтров обуславливают усреднение во времени и тем самым увеличение ширины полосы.

Основная частота голоса, согласно формуле (1.2), равна $1/T_0$, где T_0 — основной период, измеряемой по осциллограмме или же по расстоянию между вертикальными штрихами на широкополосной спектрограмме. Для определения основной частоты голоса можно использовать также любую из гармоник на узкополосной спектрограмме, если известен номер этой гармоники; наконец, можно определить основную частоту при помощи устройств для автоматического выделения основного тона. Прообразом применяемых в настоящее время для этой цели приборов является схема Грюцмахера и Лотермозера [78]; она дает запись основной частоты сонорных участков речи. Подобные приборы,

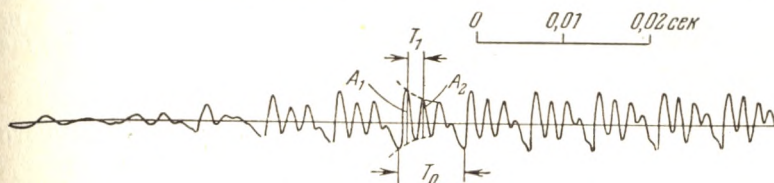


Рис. 79. Определение основной частоты голоса F_0 , а также частоты F_1 и ширины B_1 первой форманты по четвертому периоду осциллограммы звука [э] в [да]. Запись осциллограммы на быстродействующем пишущем осциллографе («Миниографе») при половинной скорости воспроизведения на магнитофоне. Для вычислений используются формулы $F_0 = \frac{1}{T_0}$, $F_1 = \frac{1}{T_1}$ и $B_1 = \frac{\lg A_1/A_2}{T_1 \pi \lg e}$. В данном частном случае $F_0 = 126$ гц, $F_1 = 520$ гц и $B_1 = 55$ гц.

однако, не всегда обеспечивают достаточную надежность полученных результатов.

Примем далее следующие обозначения:

F_n — частота форманты номера n в гц.

B_n — ширина полосы форманты номера n в гц.

L_n — уровень форманты номера n в дб.

F_n — общее обозначение для форманты номера n безотносительно к ее параметрам. Таким образом, первая форманта обозначается как $F1$, в то время как частота первой форманты обозначается через F_1 .

F_0 — основная частота голоса в гц.

$F0$ — общее обозначение основной частоты голоса без указания ее численного значения.

Естественные пределы изменения основной частоты голоса и частот формант для сонорных неносовых звуков речи при произнесении их средним мужским голосом следующие:

$$F_0 \quad 60 - 240 \text{ гц}$$

$$F_1 \quad 150 - 850 \text{ »}$$

$$F_2 \quad 500 - 2500 \text{ »}$$

$$F_3 \quad 1500 - 3500 \text{ »}$$

$$F_4 \quad 2500 - 4500 \text{ »}$$

Для женских голосов основная частота голоса выше в среднем на октаву, а частоты формант — только на 17% (см. [167, 51]). Для детей в возрасте около десяти лет форманты в среднем на 25% выше, чем для взрослого мужчины, а основная частота в среднем равна 300 гц.

Индивидуальные отклонения от приведенных средних величин могут быть весьма значительными.

При среднем интегральном уровне речи уровень первой форманты, измеренный на расстоянии 1 м от диктора, составляет 60—65 дБ (при нулевом уровне 0,0002 бара, см. [46, 40]). Уровни более высоких формант понижаются с повышением номера форманты; зависимость уровней этих формант от их частоты будет рассмотрена ниже.

Ширина полосы первых двух формант составляет 30—100 гц, но увеличивается с повышением номера форманты; так, значения B_3 и B_4 находятся в пределах 40—200 гц. Эти данные получены на основе осциллографических записей и значительно меньше значений ширины полосы, полученных Богертом [17]. Эта систематическая разница, как было указано выше, обусловлена техникой измерения. Более поздние данные Хауза и Стивенса [101] и Ван-ден-Берга [8] хорошо согласуются с приведенными выше.

§ 3. Спектрограммы речевого материала, использованного для сравнения с результатами расчета согласных

Рис. 80—98 (вклейка между стр. 240 и 241) содержат акустические характеристики односложных слов типа согласный + [a], произнесенных диктором, для которого снимались рентгенограммы. Запись была проведена в заглушенной камере и включала слоги:

Рис. 80. [ma] [m,a]
Рис. 81. [na] [n,a]
Рис. 82. [la] [l,a]
Рис. 83. [ra] [r,a]
Рис. 84. [ja]
Рис. 85. [fa] [f,a]
Рис. 86. [va] [v,a]

Рис. 87. [sa] [s,a]
Рис. 88. [za] [z,a]
Рис. 89. [xa]
Рис. 90. [ša] [šča]
Рис. 91. [ža]
Рис. 92. [ča] [ca]

Рис. 93. [pa] [p,a]
Рис. 94. [ba] [b,a]
Рис. 95. [ta] [t,a]
Рис. 96. [da] [d,a]
Рис. 97. [ka] [k,a]
Рис. 98. [ga]

Эти слоги произносились один за другим со скоростью приблизительно два слога в секунду.

Анализ сводился к получению спектрограмм (с помощью «Сонаграфа»), кривых изменения интенсивности во времени при различных типах предварительной фильтрации и спектральных разрезов для представляющих интерес моментов звучания согласных. Амплитудно-частотные зависимости построены на основании записи с помощью пишущего осциллографа «Мингограф» напряжения на выходе анализатора гармоник с шириной полосы пропускания его фильтра 150 гц¹⁾. Получение этих зависимостей требует большего времени, чем анализ при помощи «Сонаграфа», но дает более определенные данные об интенсивности. Именно по таким записям можно было обнаружить F2 и F3 звонкой смычки; поэтому они и были использованы в качестве контрольных данных при установлении практической возможности расчета спектров речи на основе рентгенограмм.

Момент времени, для которого производилось определение спектрального разреза, выбирался для некоторых заранее установленных точек на временной шкале спектрограммы. Для анализа глухих взрывных выбиралась три различные точки (или, вернее, три отрезка вре-

¹⁾ В начальной стадии этой работы участвовали Халле и Юз.

мени) в пределах интервала раствора, обозначенные I, II и III. Они соответствуют моментам взрыва, фрикации (фрикативный интервал) и аспирации; см., например, [t] и [k]. Обычно первые или последние два интервала более или менее взаимно перекрывались.

Производилась синхронная запись следующих функций времени. 1) Осциллограмма с частотной характеристикой (шкалой) В (см. § 1), т. е. с подавлением низких частот. Верхняя граничная частота определялась свойствами «Мингографа» и была равна приблизительно 800 гц.

2)—6) Зависимость интенсивности от времени. Кривые (2)—(6) были получены при эффективном времени усреднения 10 мсек, что соответствует граничной частоте сглаживающего фильтра 50 гц. Использовалось двухполупериодное линейное детектирование. Шкала амплитуд для (2) и (3) логарифмическая, для (4), (5) и (6) линейная, во всех случаях представлена в дб. При получении зависимостей (2)—(6) была использована аппаратура со следующими частотными характеристиками.

2) Измеритель уровня интенсивности (шумомер) с частотной характеристикой (шкалой) А (см. § 2). Эта шкала дает в основном подавление низких частот.

3) Высокочастотный фильтр с граничной частотой 1500 гц.

4) Полосный фильтр с полосой пропускания 1400—1800 гц.

5) Полосный фильтр с полосой пропускания 2800—3600 гц.

6) Высокочастотный фильтр с граничной частотой 4000 гц.

Представленный материал позволяет сделать ряд интересных выводов¹⁾, например дает возможность сравнить ход огибающей интенсивности во времени для взрывных, аффрикат и фрикативных. Взрывные характеризуются относительно коротким интервалом взрыва или по крайней мере коротким фрикативным участком в интервале времени раствора. Длительность у аффрикат больше, чем у взрывных, но меньше, чем у длительных. Взрыв для взрывных характеризуется значительно более коротким периодом нарастания, чем спадания. Почти такие же длительности можно наблюдать у аффрикат. В случае длительных в начальном положении нарастание интенсивности длится в течение значительной части всего звука вследствие возрастания избыточного давления, связанного с началом дыхательного импульса (см. рис. 100). Сводку количественных данных, характеризующих периоды нарастания, установившегося состояния и спадания, дал Халле [82].

II. ХАРАКТЕРИСТИКИ ИСТОЧНИКА

§ 1. Голосовой источник

При образовании сонорных звуков первичным источником энергии является работа дыхательных мышц, создающих при сокращении избыточное давление в легких и, как следствие этого давления, поток воздуха; последний периодически изменяется по величине в результате

¹⁾ Очень слабая третья форманта у гласных, видимая на некоторых спектрограммах, является индивидуальной особенностью диктора. Как показано в § 2 гл. 8, имеется теоретическая возможность того, что предельно малая степень назализации вызовет это явление. Третья форманта растет по интенсивности при большой связи с носовой полостью или при отсутствии такой связи, сравните [la] с [pa] и [za] (см. рис. 82, 81 и 88).

периодического, с частотой основного тона голоса, расхождения и смыкания голосовых связок. С акустической точки зрения работу связок нельзя трактовать по аналогии с колеблющейся мембраной; связки только модулируют постоянный поток воздуха при выдохе, но не генерируют звуковых колебаний сколько-нибудь заметной интенсивности путем прямого преобразования механических колебаний в звуковые.

Что касается самого механизма возбуждения колебаний голосовых связок, то, согласно Юссону [106], к мышцам гортани подаются от центральной нервной системы нервные импульсы с частотой следования, равной основной частоте голоса. Эти импульсы и вызывают периодическую деформацию голосовых связок, синхронно с частотой их поступления.

Однако после экспериментов Смита, который успешно имитировал человеческий голос на моделях, после экспериментальных и теоретических работ Смита [185] и Ван-ден-Берга [9—12, 14, 15], а также электромиографических исследований Фааборг-Андерсена [43] представляется, что теория Юссона дает мало вероятное объяснение механизма образования голоса. Гораздо проще объяснить работу голосовых связок как механической системы, если учитывать переменную силу, которая представляет собой смену избыточного давления, действующего на сомкнутые голосовые связки, и отрицательного давления, возникающего при протекании воздуха через открытые голосовые связки. Появление отрицательного давления, обусловленного гидродинамическим эффектом Бернулли, является достаточной причиной перехода первоначально раздвинутых голосовых связок к сближению, без всякого участия мышечной деятельности. Так происходит дело при мягком приступе в грудном регистре, как это показала скоростная киносъемка голосовых связок, выполненная в лабораториях фирмы Белл (см. [56, 185]). При твердом приступе связки первоначально сомкнуты и раздвигаются напором воздуха. Будучи раздвинуты, они возвращаются в сомкнутое состояние не только вследствие эффекта Бернулли, но и в результате наличия упругих сил. Более ранние механические теории учитывали только упругие возвращающие силы. Однако наличие эффекта Бернулли само по себе достаточно для объяснения работы голосовых связок и в грудном регистре.

При низкой основной частоте голоса раскрытие голосовой щели при выходе голосовых связок из сомкнутого состояния постепенно распространяется вверх и наружу. Согласно Фарнсворту [56] и Смигу [185] обратный процесс, т. е. смыкание голосовых связок, также начинается с нижней их части. Разность фаз между перемещениями верхнего и нижнего краев связок становится все менее заметной по мере повышения основной частоты, что обусловлено увеличением жесткости голосовых связок и уменьшением действующей массы, участвующей в колебаниях. В регистре фальцета колеблются в основном верхние края связок. Увеличение жесткости голосовых связок осуществляется растяжением их видимой части; при этом, однако, наибольшая длина голосовой щели, открытой для прохода воздуха, уменьшается [185].

Голосовая щель имеет эффективную длину порядка 12 мм при грудном регистре, а ширина ее при умеренном голосовом усилии составляет около 2,5 мм. Проход между связками, закрывающийся в фазе смыкания, имеет размер около 2—5 мм, считая в направлении потока воздуха.

Эксперименты на модели, имевшие целью установить величину эффекта Бернулли, были выполнены Ван-ден-Бергом и др. [15]. В этих экспериментах на макете гортани было определено пространственное распределение давлений вдоль воздушного потока при различных величинах площади поперечного сечения голосовой щели. При очень малой ширине голосовой щели сопротивление ее достаточно для того, чтобы сохранялось положительное давление; при ширине больше 0,2 мм среднее избыточное давление становится отрицательным, по величине составляет около четверти подсвязочного давления и достигает половины последнего в точке максимума. Из этих экспериментов можно заключить, что отрицательное давление в фазе раскрытия голосовой щели прямо пропорционально подсвязочному давлению и по величине достаточно для возникновения заметного всасывающего эффекта.

Упрощенный анализ движения голосовых связок, без учета их упругости, показывает, что время, необходимое для того, чтобы связки разошлись и затем пришли в исходное состояние, обратно пропорционально квадратному корню из подсвязочного давления, прямо пропорционально квадратному корню из колеблющейся массы связок и, наконец, тому минимальному расстоянию между связками, при котором среднее давление в голосовой щели становится отрицательным. Отсюда видно, что увеличение подсвязочного давления влияет на основную частоту голоса так же, как влияло бы увеличение упругости связок.

При речи увеличение интенсивности голоса вызовет, таким образом, повышение основной частоты при условии, что не будет иметь место нормальная компенсация такого повышения вследствие уменьшения натяжения голосовых связок. Явление это легко продемонстрировать, если нажать на живот человека, поющего какую-либо определенную ноту [14].

Связанное с тем же механизмом возбуждения колебаний связок, но противоположное по характеру явление состоит в понижении основной частоты в моменты смычки звонких взрывных. Если вследствие сопротивления воздушному потоку, которое оказывает артикуляторное сужение в верхней части речевого тракта, уравнивается заметная часть подсвязочного давления, то это приводит к уменьшению эффекта Бернулли в голосовой щели. Вызванное этим уменьшение восстанавливающей силы, противодействующей раскрытию голосовой щели, увеличивает максимальную ширину последней; в результате увеличивается время восстановления сомкнутого состояния связок. При мягком приступе или отступе, вследствие неполного возвращения голосовых связок назад, может произойти неполное смыкание; такое же неполное смыкание может быть и особенностью голоса вообще данного лица. В частности, сиплый голос обусловлен тем, что между черпаловидными хрящами, которые нормально должны быть плотно сведены даже в моменты расхождения голосовых связок, остается некоторая утечка.

Вообще полное смыкание голосовых связок не является необходимым условием возникновения голосового источника, так как любое периодическое возмущение постоянного воздушного потока является источником звука; неполная модуляция этого потока только уменьшает интенсивность источника.

Спектр голосового источника можно рассчитать по данным о подсвязочном давлении во время фонации и о переменной проводимости голосовой щели, зависящей от изменения во времени размеров прохода между связками. В электрическом аналоге давлению, создаваемому

легкими, эквивалентно постоянное напряжение, последовательно с которым включены импеданс, соответствующий механическому импедансу голосовой щели, и входной импеданс речевого тракта со стороны голосовой щели. Предполагая, что импеданс всей расположенной ниже связок системы мал по сравнению с импедансом голосовой щели, можно приближенно считать подвязочное давление постоянным. Кроме того, надо учитывать, что и импеданс речевого тракта оказывает известное влияние на поток воздуха через голосовую щель.

Для того чтобы получить отчетливое разграничение источника и фильтра, целесообразно в эквивалентной схеме представить давление в легких как генератор постоянного тока, а импеданс голосовой щели включать параллельно входному импедансу речевого тракта. Такое представление не вполне точно из-за нелинейности импеданса голосовой щели; однако оно лучше отражает действительное положение дела, чем часто принимаемое предположение о бесконечно большом импедансе источника, так как при этом не нарушаются граничные условия. Эксперименты, выполненные на модели Ван-ден-Бергом и др. [15], показывают, что сопротивление R_F голосовой щели потоку воздуха, зависящее от площади поперечного сечения щели A и скорости частиц $v = \frac{u}{A}$, может быть разбито на две части:

$$R_F = R_L + R_T,$$

причем R_L пропорционально A^{-3} и не зависит от величины потока, а R_T пропорционально A^{-1} и v . Первое из этих сопротивлений представляет собой не что иное, как сопротивление весьма узкой щели при ламинарном потоке, и равно

$$R_L = \frac{12\mu l}{A^3/b^2} \text{ дин} \cdot \text{сек/см}^5, \quad (\text{II. 1})$$

где μ — коэффициент вязкости, равный $1,84 \cdot 10^{-4}$.

Поперечное сечение голосовой щели при этом принимается прямоугольным с площадью A и размерами сторон a и b . Глубина щели равна l .

Когда площадь щели достигает примерно $1/6$ от ее максимальной величины A , второй член R_T становится равным первому, а при дальнейшем увеличении площади становится преобладающим. Сопротивление R_T зависит от наличия вихревых потерь и равно $7/8$ от R_B (см. ниже), а последнее определяется кинетической составляющей давления в гидродинамическом уравнении Бернулли:

$$p = \frac{\rho v^2}{2}, \quad (\text{II. 2})$$

где p — перепад давления в сужении. Из этого условия находим сопротивление:

$$R_B = \frac{p}{u} = \frac{\rho v}{2A} = \frac{\rho u}{2A^2}. \quad (\text{II. 3})$$

Наряду с сопротивлением следует учитывать и реактивную составляющую импеданса голосовой щели:

$$L_q = \frac{\rho l}{A}. \quad (\text{II. 4})$$

Объемная скорость источника $u(t)$ после этого определяется тем, что противодействие, создающееся на импедансе источника, равно постоянному подсвязочному давлению p_q , т. е.

$$R_F(t) \cdot u(t) + L_q(t) \frac{du(t)}{dt} = p_q. \quad (\text{II. 5})$$

Если пренебречь той составляющей R_F , которая выше была обозначена как R_L , то $R_F \approx R_B$; подставляя выражение последнего из (II.3), получим:

$$\frac{\rho}{2} \frac{u^2(t)}{A^2(t)} + \frac{\rho l}{A(t)} \cdot \frac{du(t)}{dt} = p_q, \quad (\text{II. 6})$$

или

$$\frac{\rho}{2} v^2(t) + \frac{\rho l}{A(t)} \cdot \frac{d[v(t) \cdot A(t)]}{dt} = p_q. \quad (\text{II. 7})$$

Решение этого дифференциального уравнения довольно сложно, если только в его левой части первый или второй член не является заведомо преобладающим. В первом случае скорость частиц $v(t)$ постоянна и равна $\sqrt{2p_q/\rho}$, откуда

$$u(t) = A(t) \sqrt{2p_q/\rho}, \quad (\text{II. 8})$$

и поток, создаваемый голосовым источником, пропорционален площади голосовой щели.

Во втором случае, т. е. если можно пренебречь активной составляющей, скорость частиц равна

$$u(t) = \frac{p_q}{\rho l} \int A(t) dt. \quad (\text{II. 9})$$

Наличие интеграла в этом выражении показывает, что наклон огибающей спектра потока, создаваемого источником, на 6 дБ/окт больше, чем в случае чисто активного сопротивления.

При обычной речи переход от условий преобладания активной составляющей импеданса к условиям преобладания реактивной имеет место примерно на частоте 2000 гц¹⁾. Получающееся при этом дополнительное подавление высоких частот частично уравнивается более крутыми фронтами голосового импульса, обусловленными переходом от ламинарного к турбулентному течению.

Что касается той области размеров площади голосовой щели, для которой соблюдаются условия ламинарного течения, то здесь сопротивление воздушному потоку, видимо, не так быстро падает с увеличением площади, как следовало бы из наличия A в степени -3 в формуле Ван-ден-Берга и др. [15]. Это можно объяснить тем, что при грудном регистре дополнительно к изменению длины происходит и изменение ширины голосовой щели.

Если принять, что активная составляющая импеданса определяется в основном турбулентностью, то получается хорошее соответствие с результатами измерения. Согласно данным Чибя и Каджияма [24] расход воздуха при средней интенсивности голоса и основной частоте

¹⁾ Более детальное рассмотрение этой задачи при аналогичном подходе было выполнено Фланаганом [67]. Приводимые им данные подтверждают изложенное выше.

$F_0 = 144$ гц составляет $140 \text{ см}^3/\text{сек}$ при подвязочном давлении $p_q = 16 \text{ см}$ водяного столба, или $16 \cdot 980 = 15700 \text{ дин}/\text{см}^2$. По этим данным, согласно формуле (II.2), находим, что скорость частиц равна $5200 \text{ см}/\text{сек}$ и средняя площадь голосовой щели $0,027 \text{ см}^2$. Согласно данным этих авторов пикфактор, т. е. отношение максимального потока к среднему, равен 3,75; отсюда максимальная площадь голосовой щели получается равной $0,10 \text{ см}^2$. Принимая далее эллиптическую форму голосовой щели, найдем, что наибольшая ширина ее при длине 12 мм равна $1,1 \text{ мм}$; это дает правильный порядок величины, но несколько меньше, чем в действительности. Ван-ден-Берг, принимая $p_q = 10 \text{ см}$ водяного столба и расход $150 \text{ см}^3/\text{сек}$, получает в результате подобных же расчетов максимальную ширину голосовой щели $1,3 \text{ мм}$. Однако и эта теоретическая величина почти вдвое меньше величины, которую дает измерение на снимках, выполненных в лабораториях фирмы Белл. Пренебрежение индуктивностью в этих расчетах может до некоторой степени объяснить это расхождение.

В фазе раскрытия голосовых связок на сопротивлении голосовой щели расходится часть колебательной мощности форманты. Средняя амплитуда колебаний, соответствующая этой мощности потерь, мала по сравнению со средней амплитудой потока воздуха, протекающего через голосовую щель; для частот, соответствующих более высоким формантам, относительная величина амплитуды колебаний еще существенно уменьшается в силу падения огибающей спектра голосового источника к высоким частотам.

Из этого следует, что для оценки демпфирования, определяющего ширину формант, следует исходить из дифференциального сопротивления $R_D = \frac{\Delta p}{\Delta v A}$. Продифференцировав формулу (II.2), можно видеть, что дифференциальное сопротивление R_D равно удвоенной величине сопротивления постоянному потоку $R_B = \frac{p}{v A}$, а именно [208]:

$$R_D = \frac{\Delta p}{\Delta v A} = \frac{2p}{v A^2}. \quad (\text{II. 10})$$

Это соотношение справедливо до тех пор, пока сопротивление потоку остается пропорциональным объемной скорости. Подставляя сюда приведенные выше численные значения, можно видеть, что среднее дифференциальное сопротивление равно $2 \cdot 15700/140 = 225 \text{ ак. ом}$, или 5 рс .

Для того чтобы при таком сопротивлении сужение, находящееся выше голосовой щели, могло оказать существенное влияние на величину потока, оно должно быть больше, чем это обычно имеет место для гласных. Это справедливо и для большинства звонких согласных.

Максимальная реакция системы, расположенной выше голосовой щели, на голосовые связки имеет место во время звонкой смычки, предшествующей взрыву звонких взрывных; в это время в ротовые полости магнетается воздух, заставляя раздаваться их стенки.

Помимо потери давления в голосовой щели, связанного с сопротивлением расположенного выше ее источника, наличие артикуляторного сужения приводит к понижению F_1 . Если вследствие этого F_1 совпадает с F_0 , появляется увеличенная активная нагрузка на голосовые связки входным импедансом речевого тракта; этот импеданс на частоте форманты является активным и достигает большой величины.

Это явление было детально исследовано Ван-ден-Бергом [9], однако вопрос о связанных с ним изменениях спектра источника требует дальнейшего изучения. Можно ожидать, что в спектре источника при этом будут подчеркнуты низкие частоты.

Пульсации потока у входа в сужение, расположенное выше голосовой щели, можно считать в основном синфазными с последовательностью импульсов воздушного потока, возникающих в последней. Эти импульсы, однако, несколько запаздывают и сглаживаются вследствие подавления высоких частот после основного резонанса речевого тракта, т. е. F_1 .

Мгновенное значение сопротивления потоку в артикуляторном сужении достигает максимума в моменты пиковой величины потока и может быть найдено по формуле (II.3), помня, конечно, что форма прохода и состояние поверхностей могут оказать влияние на результат; вопрос о степени этого влияния рассмотрен ниже, в § 2. Пиковое значение воздушного импульса, возникающего в голосовой щели, соответствует минимальному ее сопротивлению в момент наибольшего раскрытия; в тот же момент сужение имеет наибольшее сопротивление, так как скорость потока в нем максимальна. Все это приводит к тому, что на этом участке периода основного тона имеет место максимальное демпфирование форманты (см. § 6 Приложения III).

Колебательная составляющая объемной скорости в артикуляторном сужении на частоте первой форманты меньше или сравнима по величине с пульсациями воздушного потока, причем разность их уровней уменьшается с понижением F_1 .

То обстоятельство, что в соотношении между объемной скоростью в губном проходе и давлением в звуковом поле входит множителем частота, позволяет дать оценку спектра потока в сужении для частот, равных или лежащих ниже частоты первой форманты. Найти этот спектр можно путем интегрирования известного спектра давления в звуковом поле перед говорящим, т. е. введения поправки, составляющей -6 дб/окт.

Если колебательная объемная скорость на частоте форманты мала по сравнению со средним значением объемной скорости потока, дифференциальное сопротивление может достигнуть величины, вдвое превышающей сопротивление постоянному потоку.

Имеющиеся данные о площади голосовой щели и об изменениях потока воздуха в пределах основного периода были использованы для определения спектра голосового источника. Результаты приведены на рис. 99; первые две кривые на этом рисунке построены на основании данных скоростной киносъемки, выполненных в лабораториях фирмы Белл. Каждый период на кривых построен по тридцати последовательным кадрам. Огибающие спектра построены путем разложения этих кривых в ряд Фурье при помощи механического анализатора гармоник, причем принято, что поток пропорционален площади голосовой щели. Как можно видеть из рисунка, огибающая спектра спадает к высоким частотам примерно на 12 дб/окт. При меньшей интенсивности голоса это спадание для частот выше 500 гц еще больше. Это понятно, поскольку в этом случае переход от фазы раскрытия к фазе смыкания голосовой щели происходит менее резко.

Время пребывания голосовой щели в открытом состоянии, отнесенное к периоду колебания голосовых связок, влияет в первую очередь на низкочастотную часть спектра. Эта часть спектра соответствует колебательной составляющей, имеющей частоту, равную основной частоте

голоса, и постоянной составляющей потока, определяемой количеством выдыхаемого за период воздуха. На основании прежних измерений спектров сонорных звуков известно, что среднее спадание огибающей спектра значительно изменяется при изменении интенсивности голоса. Приблизительно можно считать, что изменение интегрального уровня речи при постоянной высоте голоса, вызывающее повышение уровня первой форманты на 10 дб, приводит к повышению уровня на основной частоте

голоса на 4 дб. Обратное соотношение имеет место при уменьшении интегрального уровня речи.

Эти данные в общем согласуются с приведенными на рис. 99, за исключением соотношений между составляющими выше и ниже 1000 гц. Предположение об изменении характера импеданса голосовой щели от существенно активного при высоком уровне речи, когда поток турбулентный, к существенно индуктивному при пониженном уровне, когда поток ламинарный, могло бы частично объяснить наблюдающиеся расхождения с приведенными выше данными. Действительно, такое изменение характера импеданса достаточно для того, чтобы увеличить спадание уровня на 6 дб/окт. Однако это лишь предположение: необходимы дальнейшие эксперименты и расчеты, для того чтобы создать основания для более строгой теории.

На рис. 99 приведена для сравнения огибающая спектра голосового источника, использованная Стивенсом и др. [196] в их более ранних работах, с аналогом речевого тракта MIT.

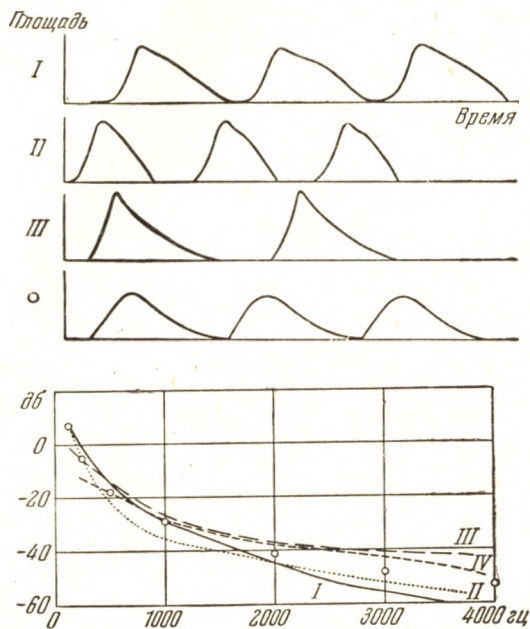


рис. 99. Полученные расчетом формы волны и огибающие спектра голосового источника. Кривые I и II построены на основании измерений площади голосовой щели на снимках голосовых связок, выполненных в лабораториях Белл. Кривая I относится к низкому звуку малой интенсивности. Кривая II — к низкому звуку большой интенсивности. Кривая III рассчитана по данным о зависимости площади голосовой щели от времени, приведенным у Чибба и Каджияма [24]. Кривая, отмеченная кружками, представляет собой экспоненциальную кривую при критическом демпфировании, характеризующуюся наклоном огибающей спектра -12 дб/окт и принятой за стандарт в расчетах, приведенных в настоящей работе. Для сравнения показана также огибающая спектра источника, использованного Стивенсом и др. [196] в их экспериментах с аналогом речевого тракта (кривая IV).

Она отличается от других меньшим спаданием огибающей в диапазоне частот 100—1000 гц. Такой спектр придает голосу более резкую окраску, чем спектры с большим преобладанием низких частот.

На том же рисунке приведен спектр источника со спаданием огибающей в 12 дб/окт , принятый за стандарт в расчетах, выполненных в настоящей работе. Имеются данные, указывающие на то, что такой спектр обеспечивает хорошую естественность речи, хотя уровни третьей и более высоких формант, возможно, оказываются все же несколько преувеличенными.

Миллером [154] и Чжаном [20] были описаны опыты по определению формы кривой воздушного потока в голосовой щели. В этих работах был

использован метод устранения формантных пиков антирезонансными контурами¹⁾). Предварительные результаты наших собственных экспериментов в этом направлении показывают, что такой метод, несмотря на косвенный его характер, позволяет достаточно детально исследовать свойства потока в голосовой щели. Обнаруженные нами изменения формы кривой подтверждают приведенные выше общие положения. Остается исследовать роль пиковых значений потока и переходных процессов при открытии и смыкании голосовой щели в формировании спектра источника для разных категорий голосов.

Коэффициент полезного действия «генератора» голоса весьма невелик. По Ван-ден-Бергу [13] величина этого коэффициента, определяемая как отношение излучаемой акустической мощности к произведению избыточного подсвязочного давления на объемную скорость потока воздуха, выдыхаемого при фонации, изменяется от $0,45 \cdot 10^{-5}$ при звуковом давлении 55 дб на расстоянии 25 см от рта до $45 \cdot 10^{-5}$ при увеличении звукового давления на том же расстоянии на 40 дб. Ван-ден-Берг нашел также, что объемная скорость линейно зависит от подсвязочного давления и что при постоянной основной частоте голоса звуковое давление излучаемого звука возрастает пропорционально квадрату объемной скорости воздушного потока.

Из цитированных выше данных об изменениях спектра, связанных с увеличением уровня речи при постоянной F_0 [46], следует, что изменения амплитуды на основной частоте голоса приблизительно определяются корнем квадратным из изменений амплитуды интегрального звукового давления. Поскольку изменения среднего значения потока, или, иначе говоря, нулевой гармоники спектра голосового источника, должны быть подобны изменениям амплитуды на основной частоте голоса, можно сделать тот вывод, что повышение к. п. д. при повышении интенсивности речи обусловлено в основном уменьшением относительной длительности импульсов воздушного потока и связанным с этим более резким нарастанием или спадом кривой или обострением ее пика²⁾).

§ 2. Турбулентный и импульсный источники

Имеющиеся в настоящее время сведения о физических характеристиках шумовых источников менее полны, чем для голосового источника. Длительные звуки шумового характера обусловлены турбулентным источником, расположенным в суженном проходе речевого тракта или вблизи от него. Сжатие потока в таком проходе увеличивает скорость протекания воздуха в нем, что связано с возникновением циркуляций и вихрей, имеющих частично статистический характер. Место их возникновения в сужении зависит как от величины потока, так и от геометрической формы сужения и условий на поверхности его стенок. При ударе струи о препятствие может возникнуть турбулентный источник.

¹⁾ Эксперименты того же рода, выполненные Лауренсом (Англия) с помощью разработанного им «речевого микроскопа», показывают, что показатель затухания (ширина полосы) первой форманты изменяется с изменением периода основной частоты голоса (личное сообщение).

²⁾ Ван-ден-Берг считает более существенным фактором изменение фильтровой функции, а не формы кривой, как это принято здесь. Он подчеркивает наличие корреляции между уровнем речи и величиной раскрытия рта. Однако подобные артикуляторные изменения должны были бы быть очень велики для объяснения имеющей место зависимости, а это привело бы к существенному повышению F_1 и изменению вследствие этого фонетического качества звука.

более интенсивный, чем источник, возникающий в самом проходе. Роль такого препятствия играют, например, верхние резцы при образовании зубных согласных.

Образование вихрей у свободно обтекаемого препятствия при малых скоростях потока может привести к возникновению свистящих шумов квазипериодического характера, так называемых «эоловых звуков» или «тонов лезвия» [150]. Тенденция к периодичности усиливается, если такой источник расположен в полости резонатора. При небольших скоростях потока резонатор может воздействовать на источник таким образом, что периодичность вихреобразования будет определяться собственной частотой резонатора. На резонансной частоте импеданс резонатора со стороны источника является активным и малым по величине; наличие потока воздуха связано с появлением отрицательного сопротивления, что и приводит к возбуждению незатухающих колебаний в системе. Такой именно механизм имеет место при образовании свиста, создаваемого губами или зубами. Однако при увеличении скорости потока периодичность нарушается и свист переходит в случайный шум. Акустический эффект в этом случае определяется уже свойствами источника, а не резонатора.

Основное значение при рассмотрении турбулентного течения имеет число Рейнольдса

$$Re = \frac{vh}{\nu}, \quad (II. 11)$$

представляющее собой безразмерную величину, пропорциональную скорости частиц v см/сек и эффективной ширине прохода h см. Постоянная $\nu = 0,15$ см²/сек есть кинематическая вязкость, определяемая как отношение вязкости к плотности газа.

В зависимости от геометрии прохода и свойств ограничивающей его поверхности существует определенное критическое значение числа Рейнольдса, при достижении которого начинается вихреобразование.

В относительно коротком суженном проходе сопротивление турбулентному потоку в большей мере зависит от минимальной площади поперечного сечения прохода, чем от его длины. Согласно измерениям Гейнца [95] главная часть перепада давления в узкой трубке длиной 5 см и диаметром 0,25 см, играющей роль сужения в механической модели речевого тракта, имеет место у заднего ее конца, в который ударяет струя воздуха. Внутри трубки давление падает значительно медленнее. Эти данные с точностью до 10% совпадают с теоретическими выводами, относящимися к перепаду давления в трубках, длина которых достаточна для обеспечения однородности потока. Согласно Шлихтингу [180] и Экку [41] перепад давления на единицу длины при таких идеальных условиях равен

$$\frac{\Delta p}{\Delta l} = \frac{\lambda}{d} \cdot \frac{\rho}{2} v^2, \quad (II. 12)$$

где $\lambda = 0,3164 Re^{-1/4}$, $Re = \frac{vd}{\nu}$ и d — диаметр. Отсюда видно, что перепад давления пропорционален $v^{7/4}$. При ламинарном потоке

$$\frac{\Delta p}{\Delta l} = \frac{64}{Re d} \frac{\rho v^2}{2} = 32 \frac{\rho \nu}{d^2}. \quad (II. 13)$$

Эта формула могла бы быть получена из формулы (II.1).

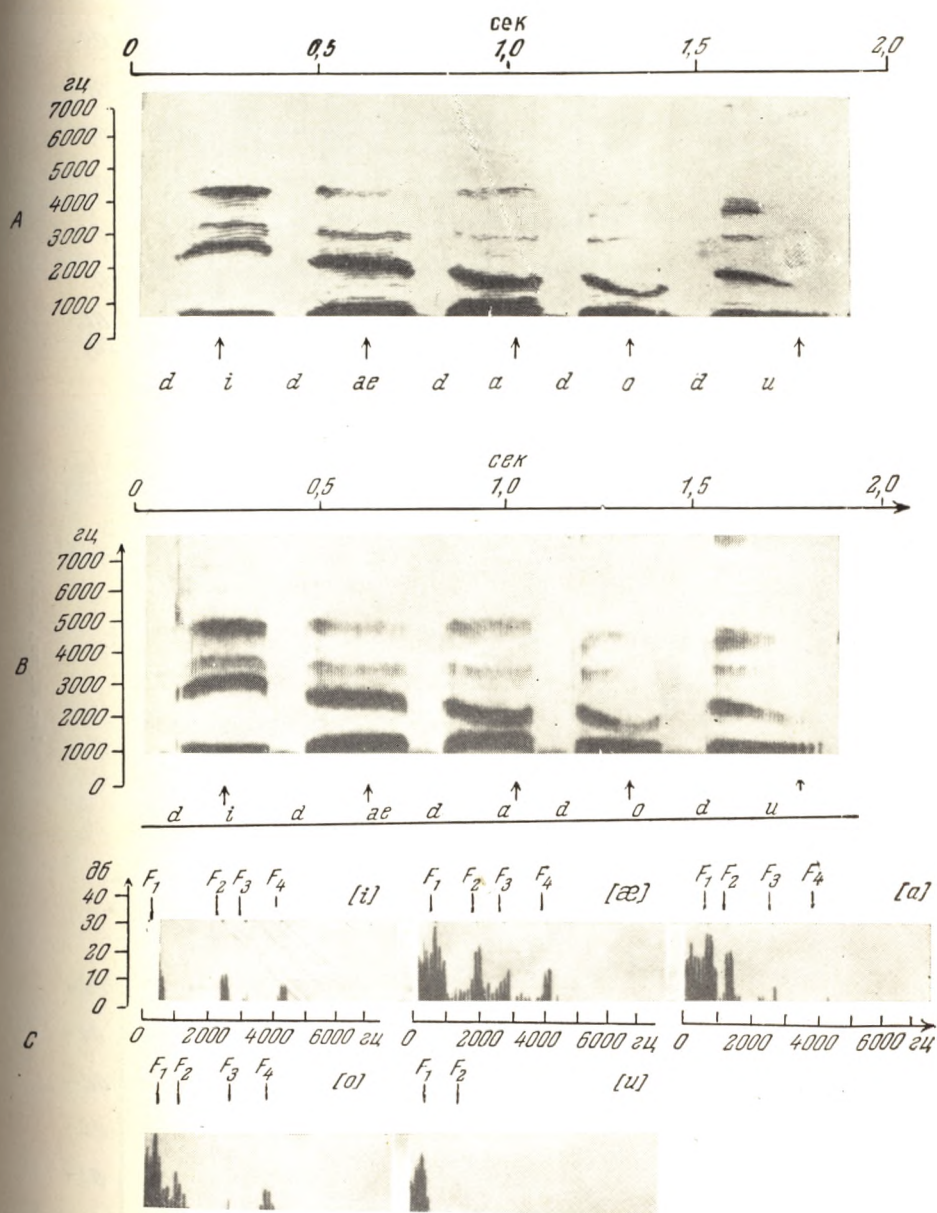
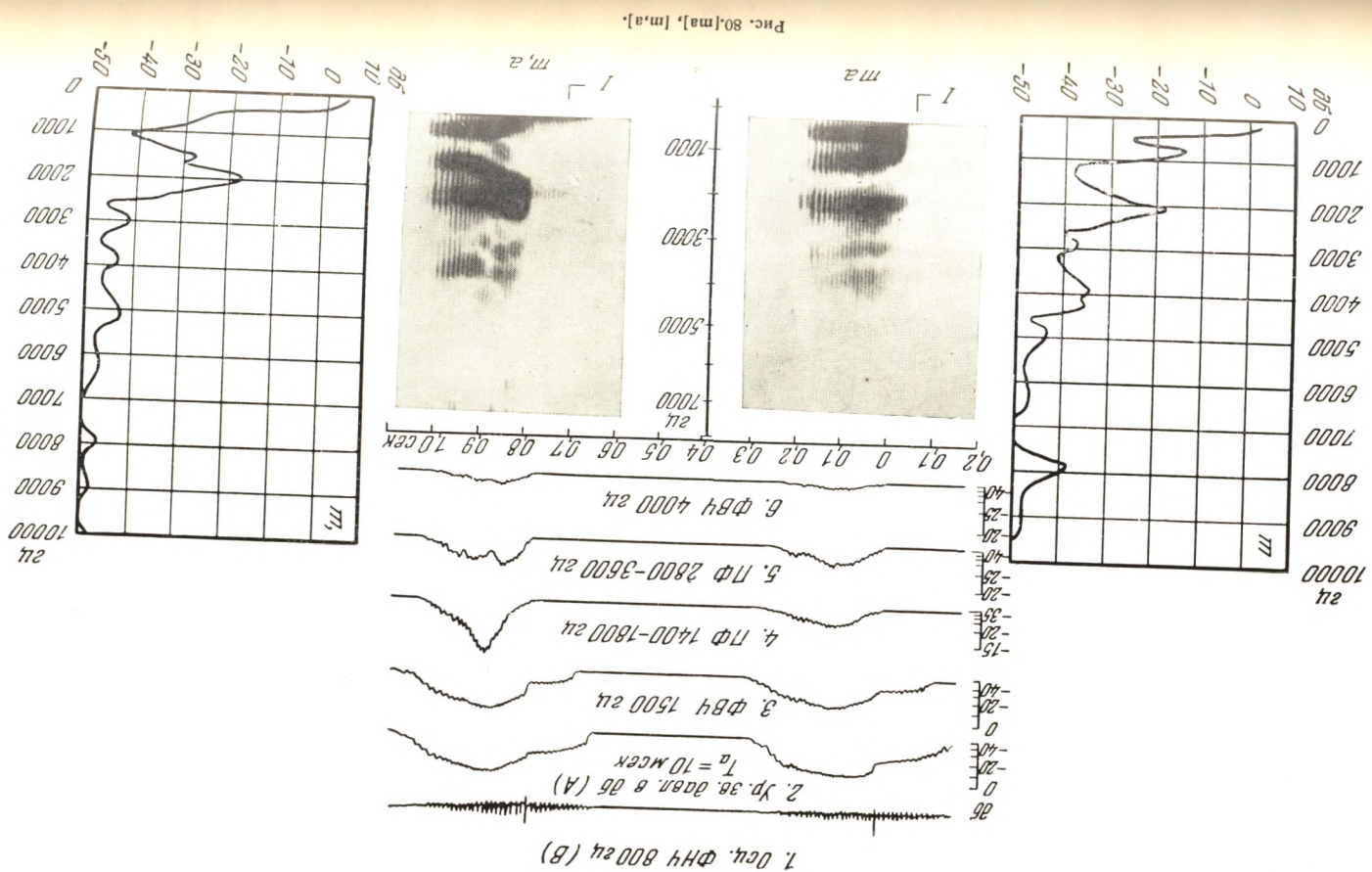
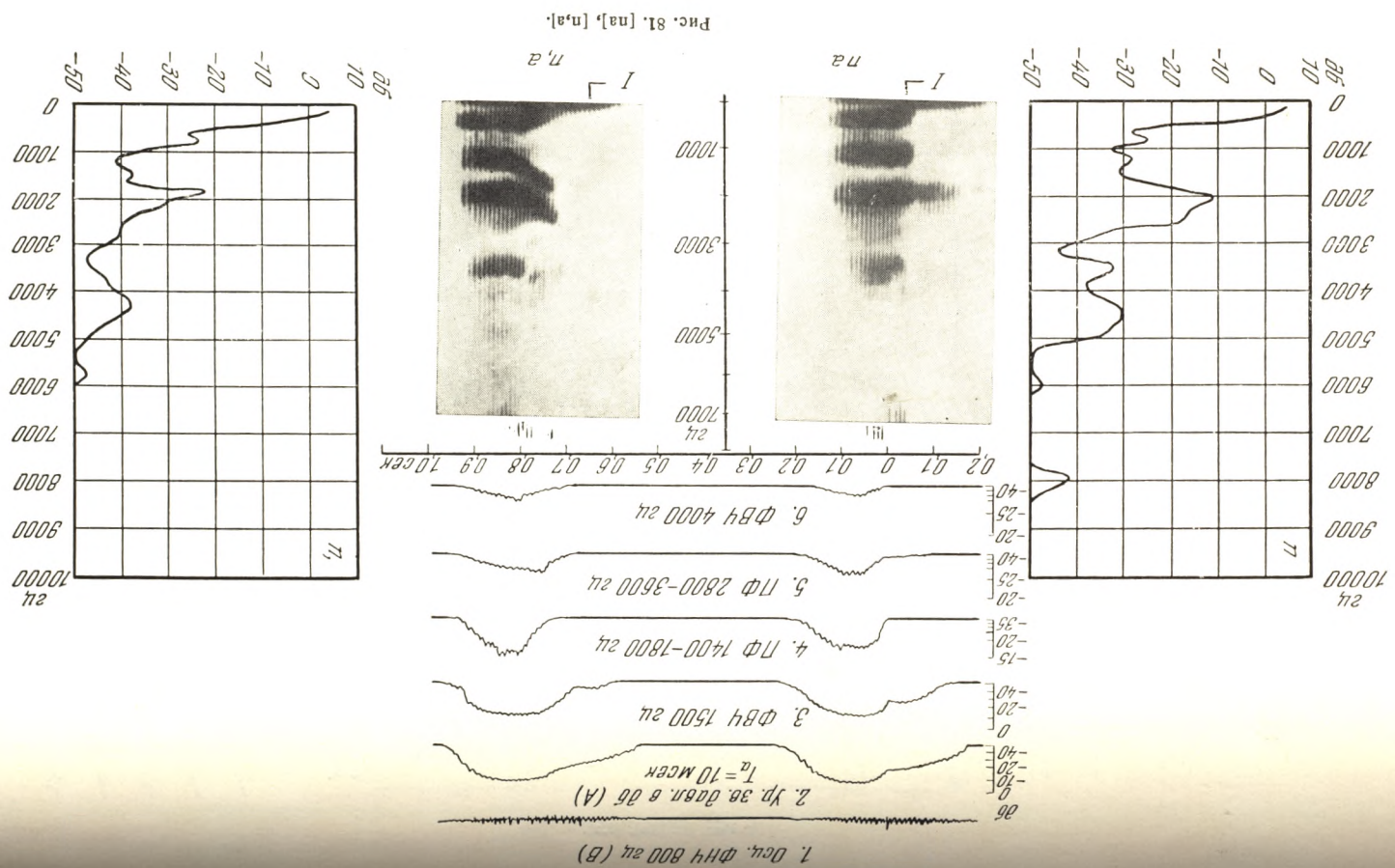
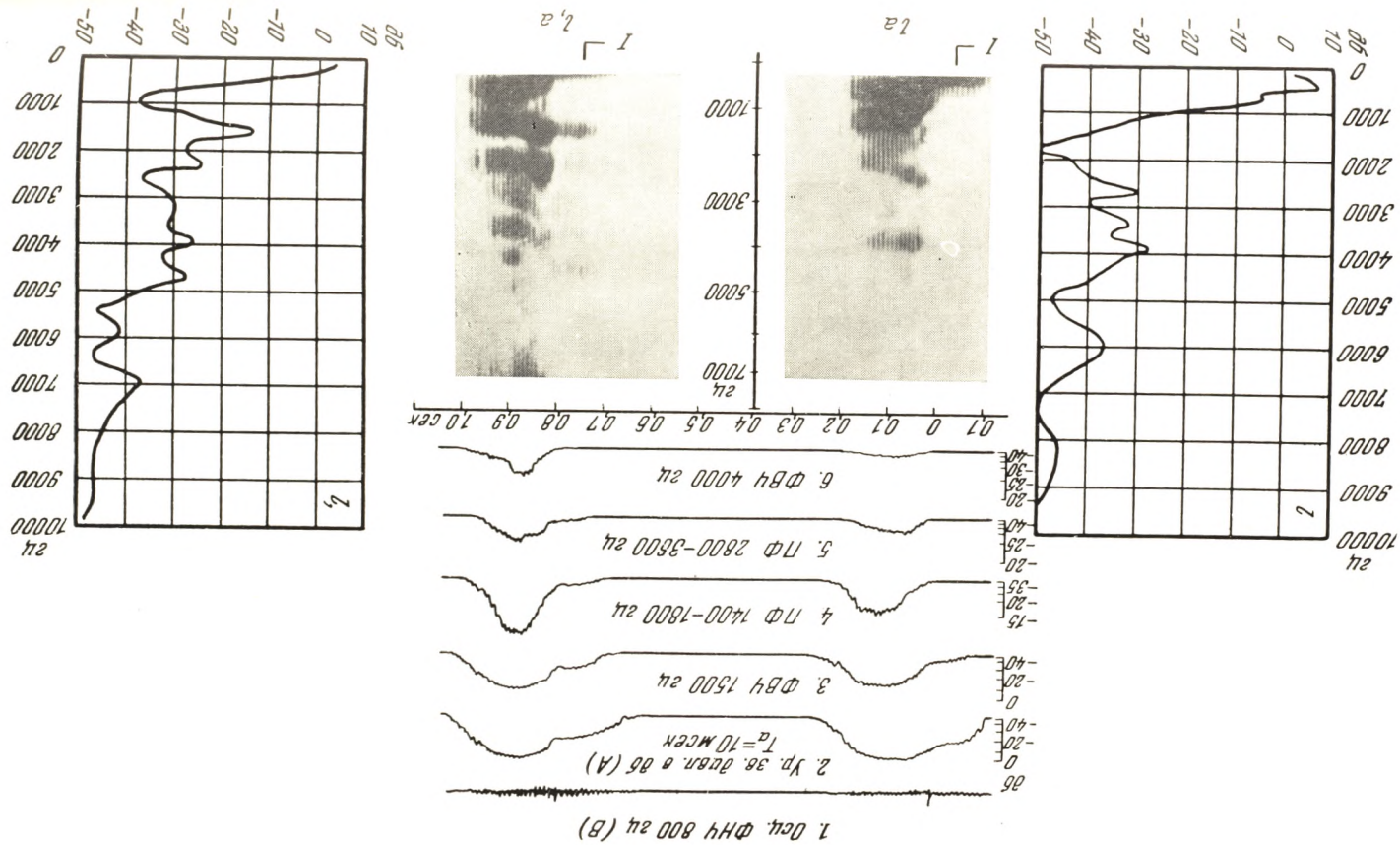
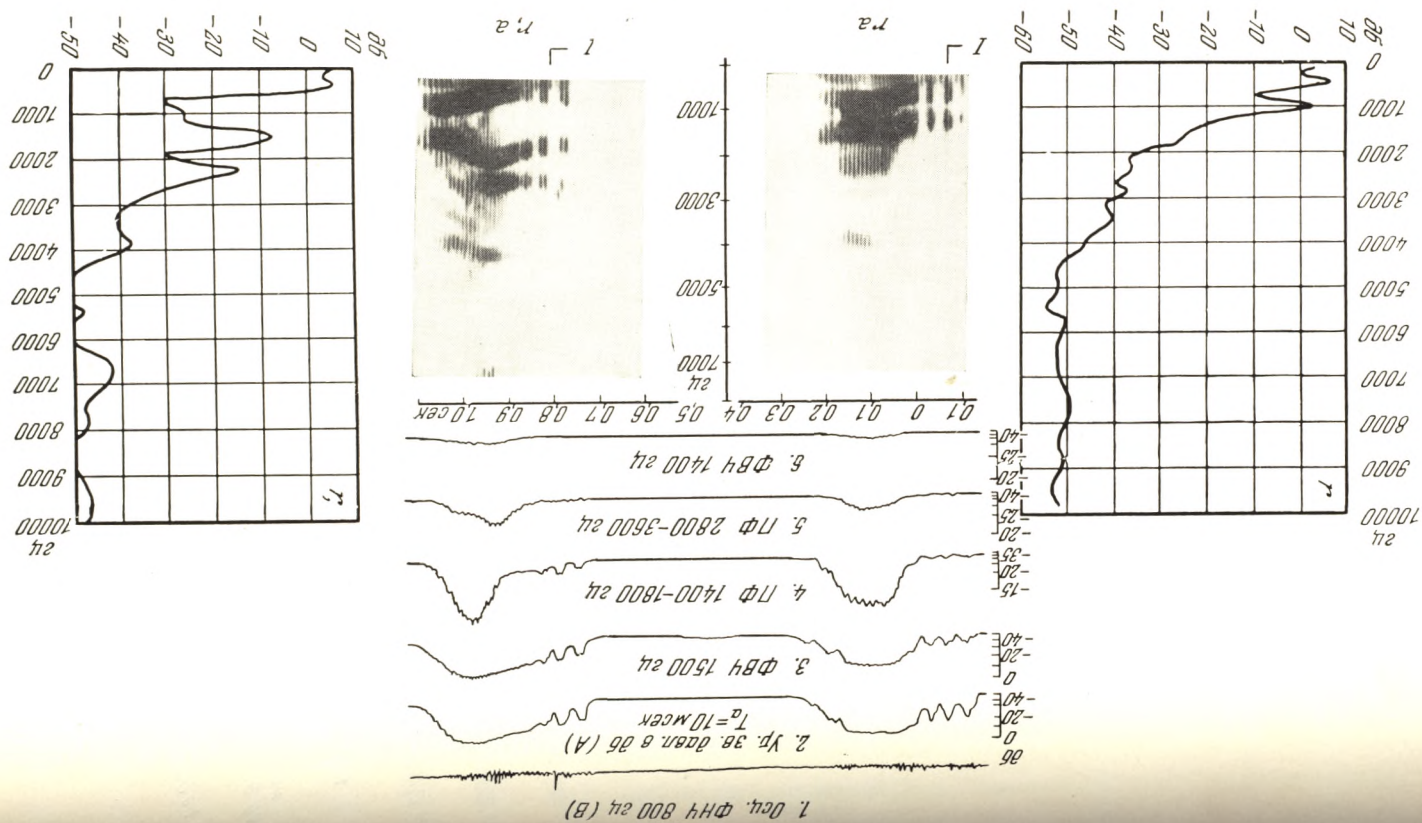
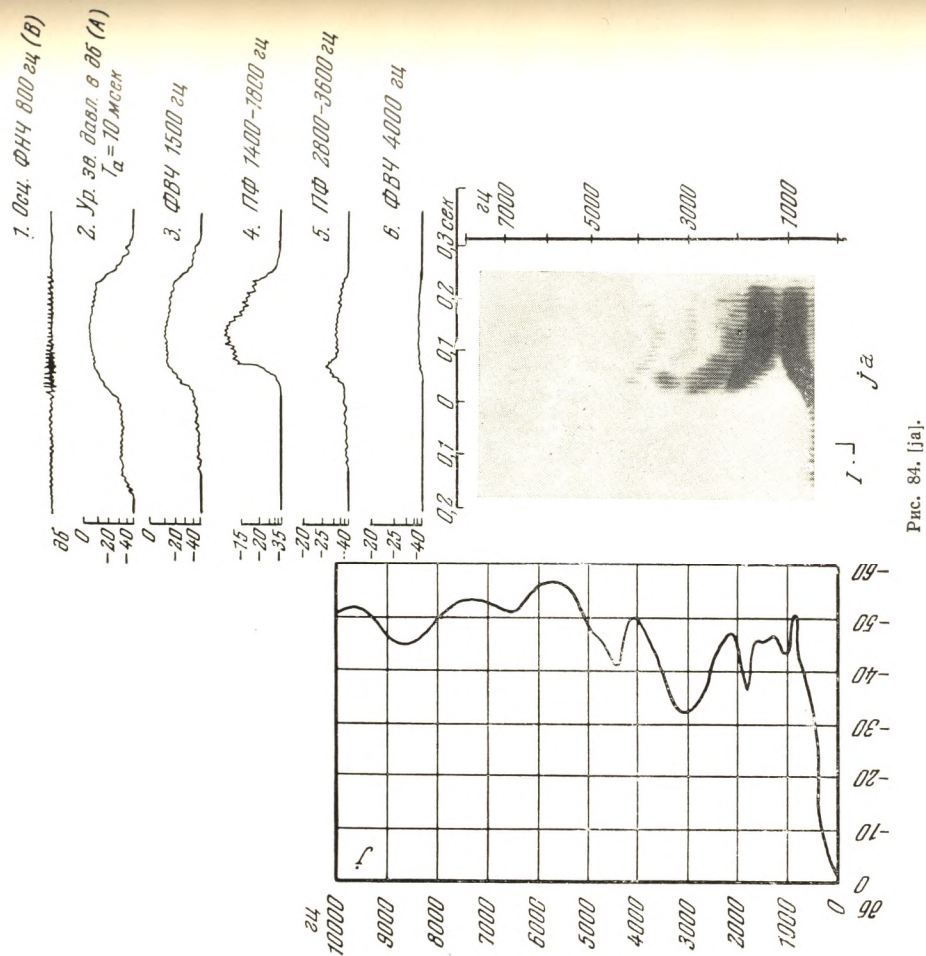
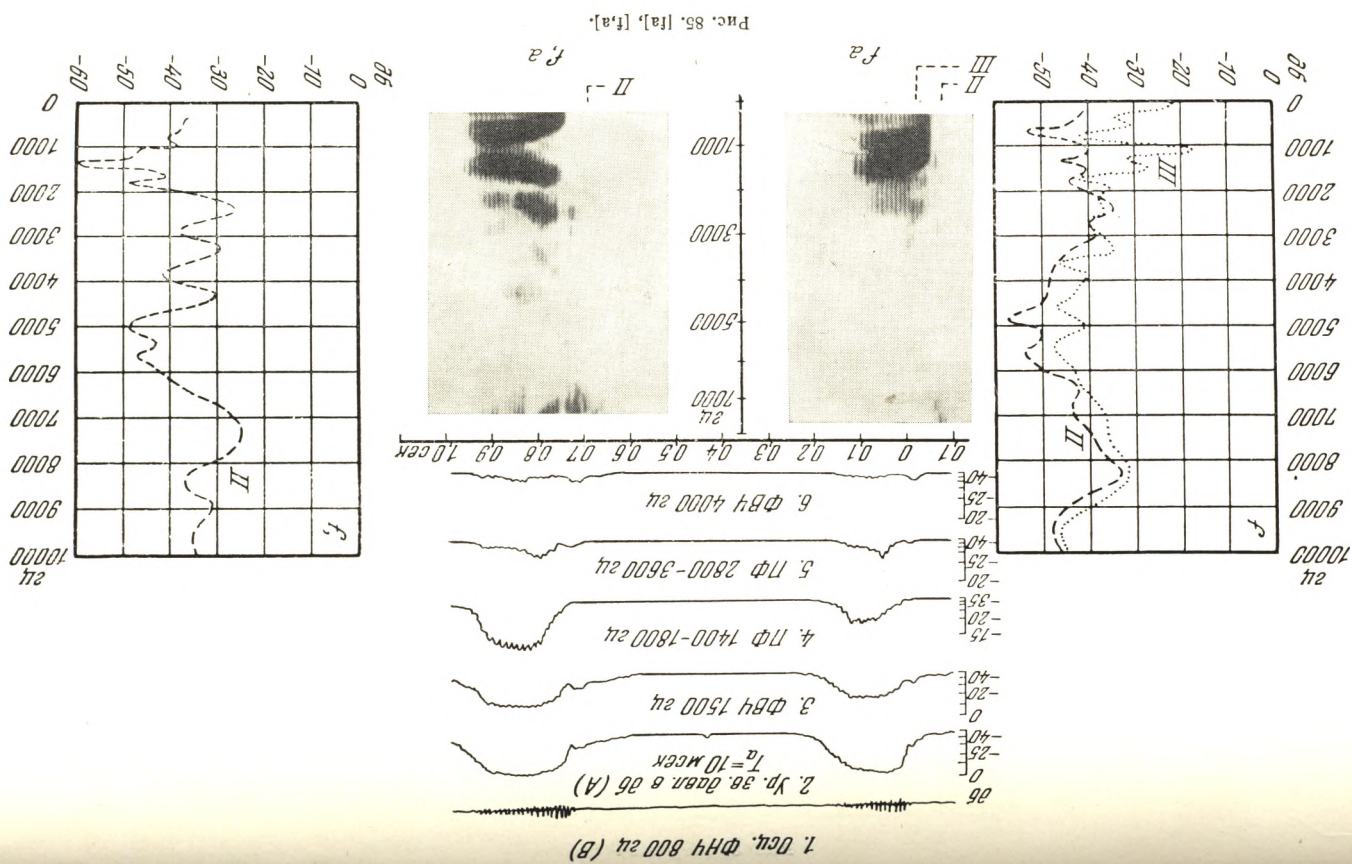
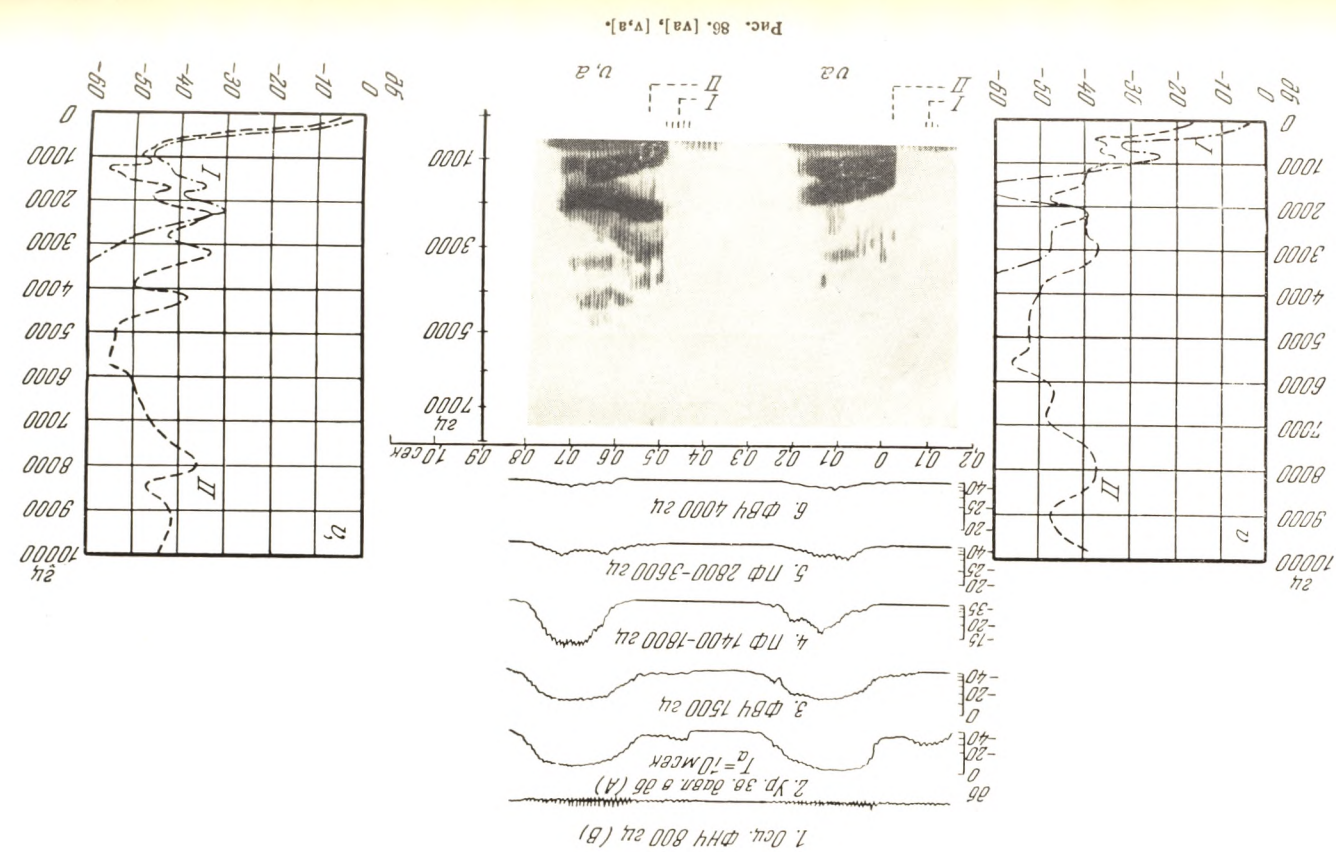
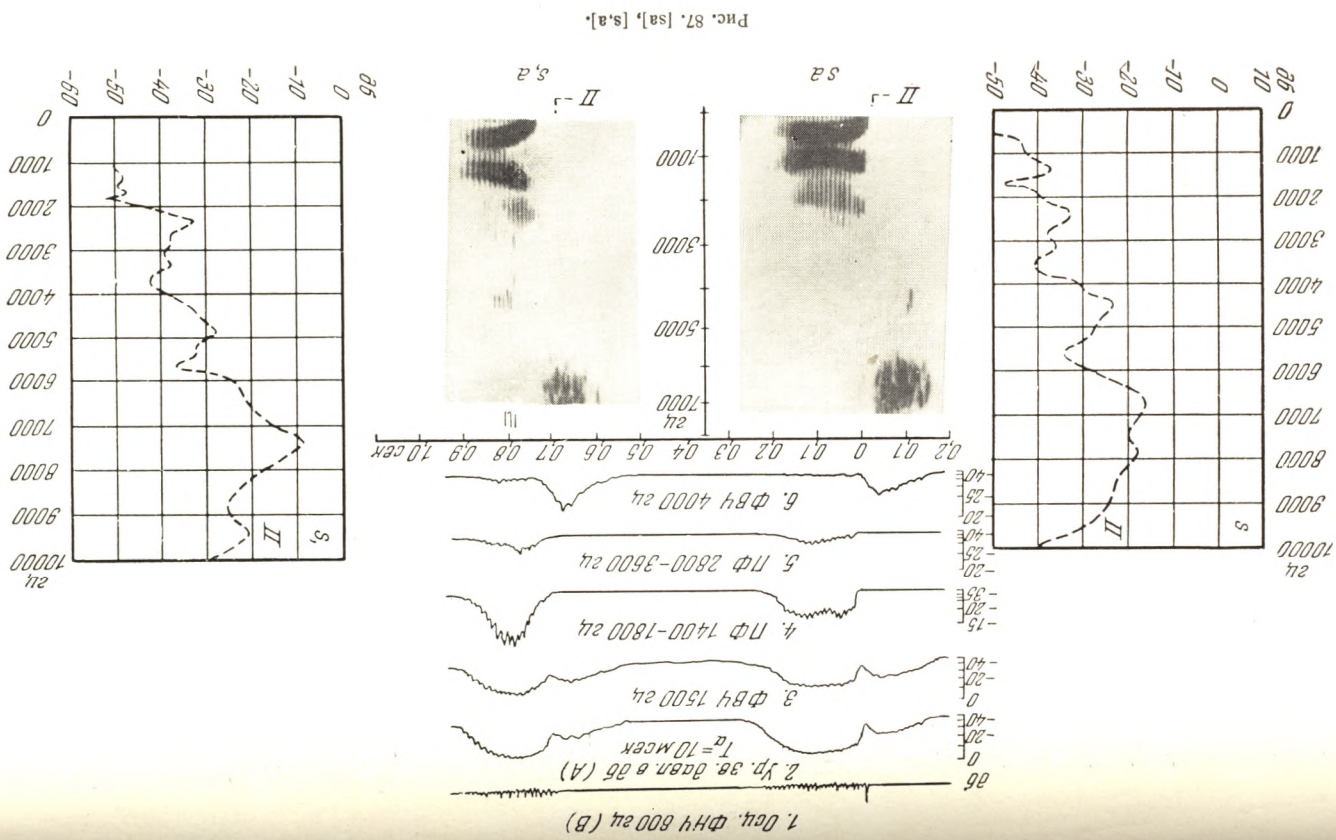


Рис. 77. Спектрограммы, полученные при помощи «Сонаграфа»: А — узкополосная, В — широкополосная, С — спектральный разрез. Речевой материал [didædadodu]; диктор — американец. Моменты снятия спектральных разрезов отмечены стрелками на оси времени каждой из спектрограмм.









1. Осы, ФНЧ 800 гц (В)

2. Ур. зв. даял. в дб (А)
 $T_d = 10 \text{ мсек}$

3. ФВЧ 1500 гц

4. ПФ 7400-7800 гц

5. ПФ 2800-3600 гц

6. ФВЧ 4000 гц

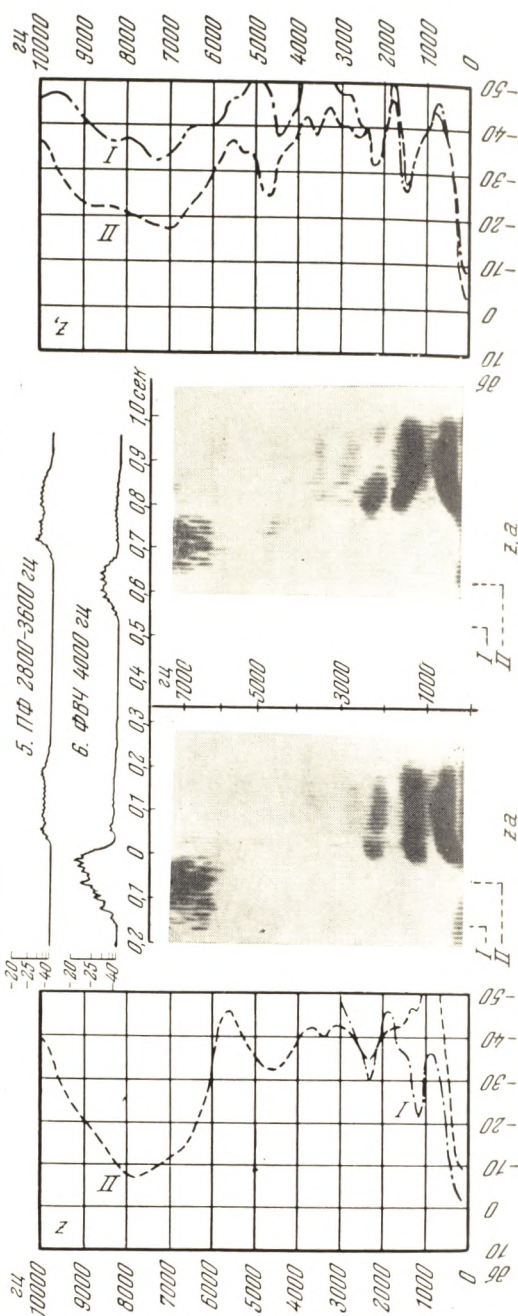
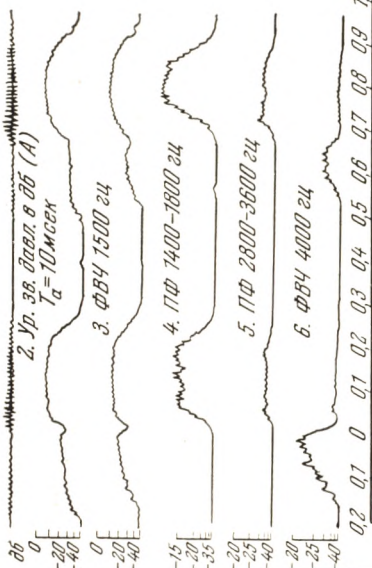


Рис. 88. [za], [za].

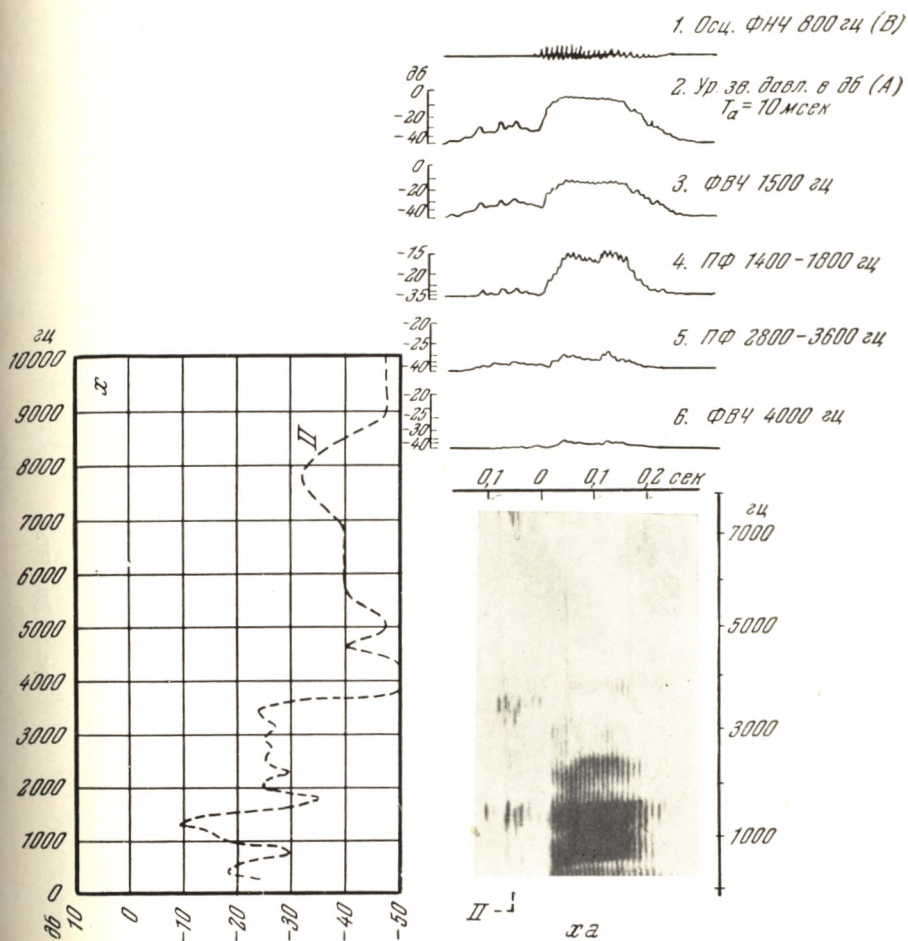
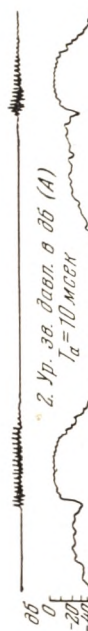
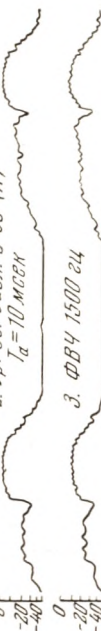


Рис. 89. [ха].

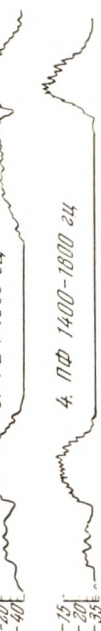
1. $\sigma_{\text{св}}$ $\Phi_{\text{НЧ}}$ 800 $\sigma_{\text{св}}$ (В)



2. Ур. св. вост. в $\sigma_{\text{св}}$ (А)



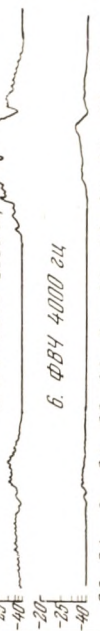
3. $\Phi_{\text{ВЧ}}$ 1500 $\sigma_{\text{св}}$



4. $\Pi\Phi$ 1400-1800 $\sigma_{\text{св}}$



5. $\Pi\Phi$ 2800-3600 $\sigma_{\text{св}}$



6. $\Phi_{\text{ВЧ}}$ 4000 $\sigma_{\text{св}}$

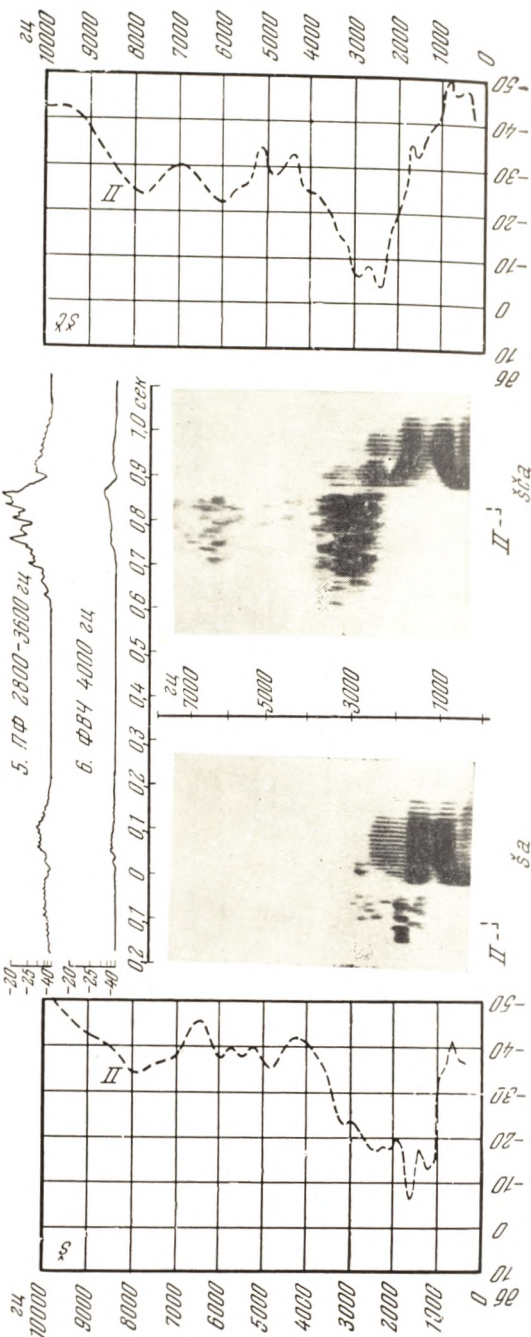
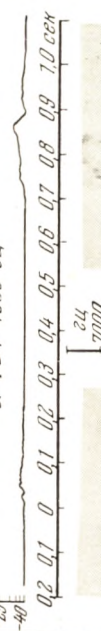
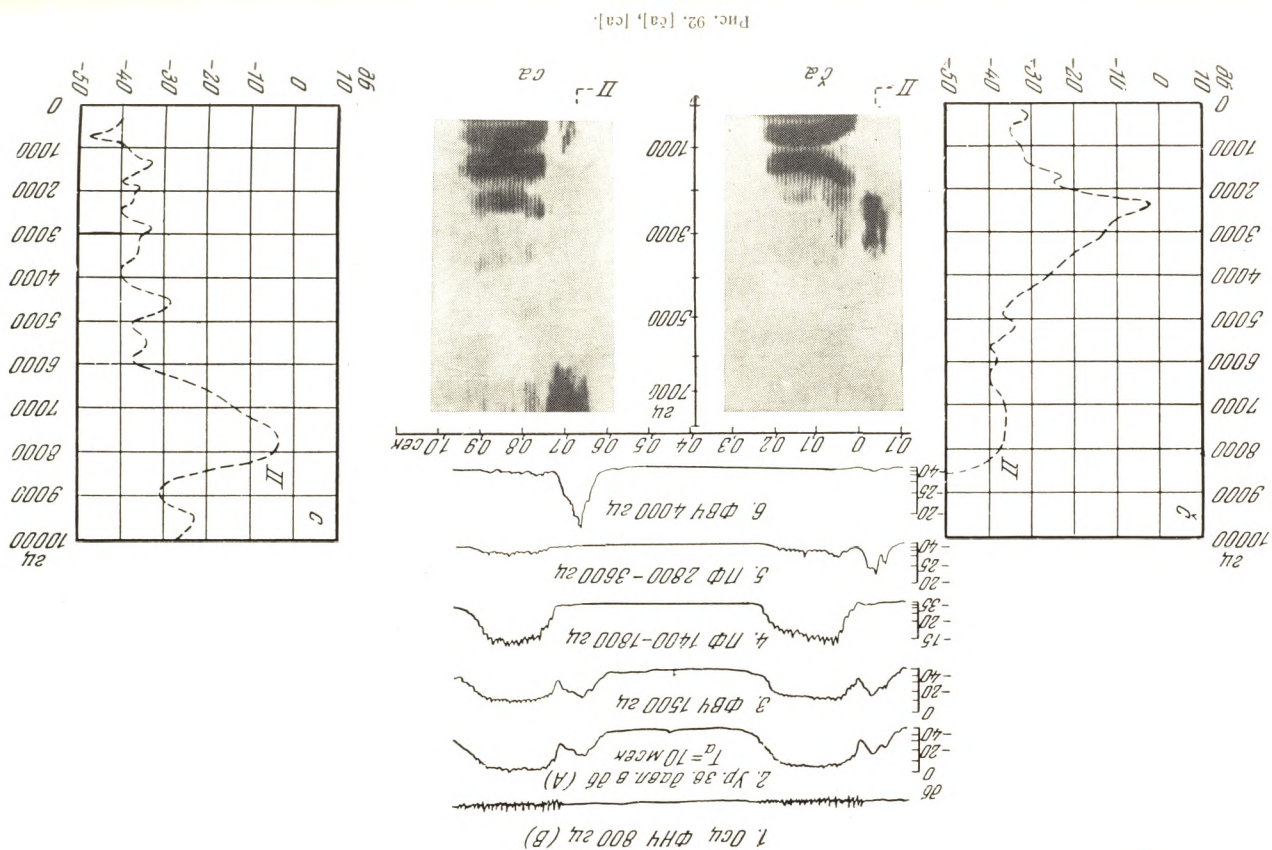
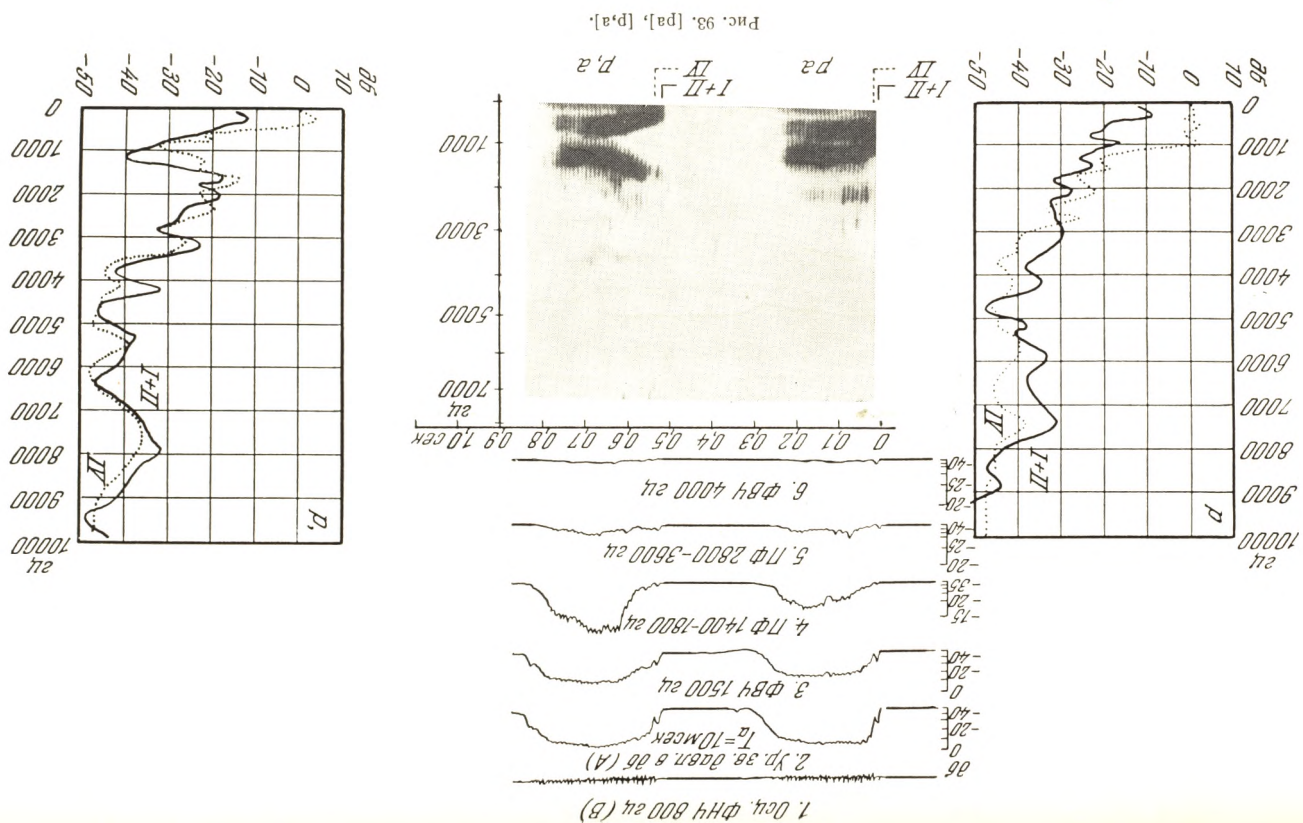
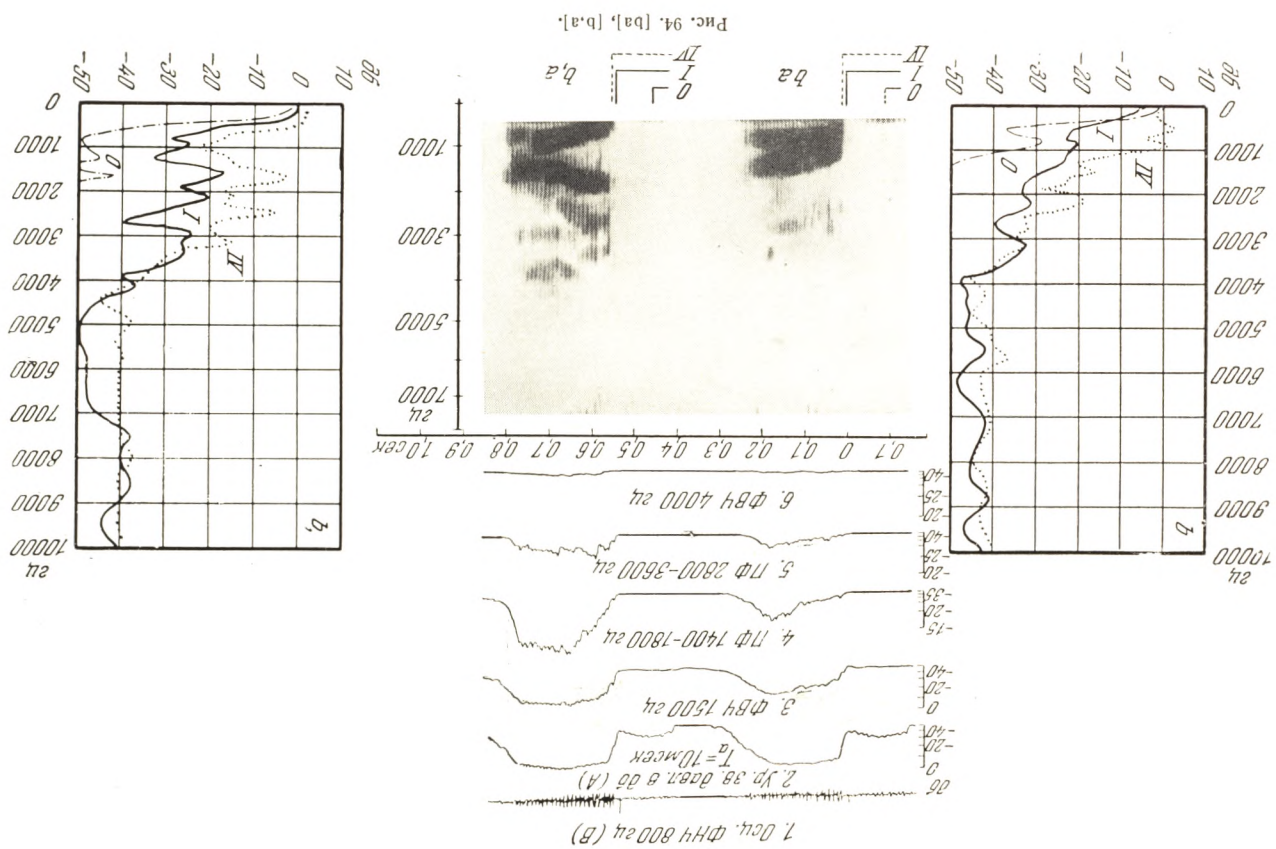
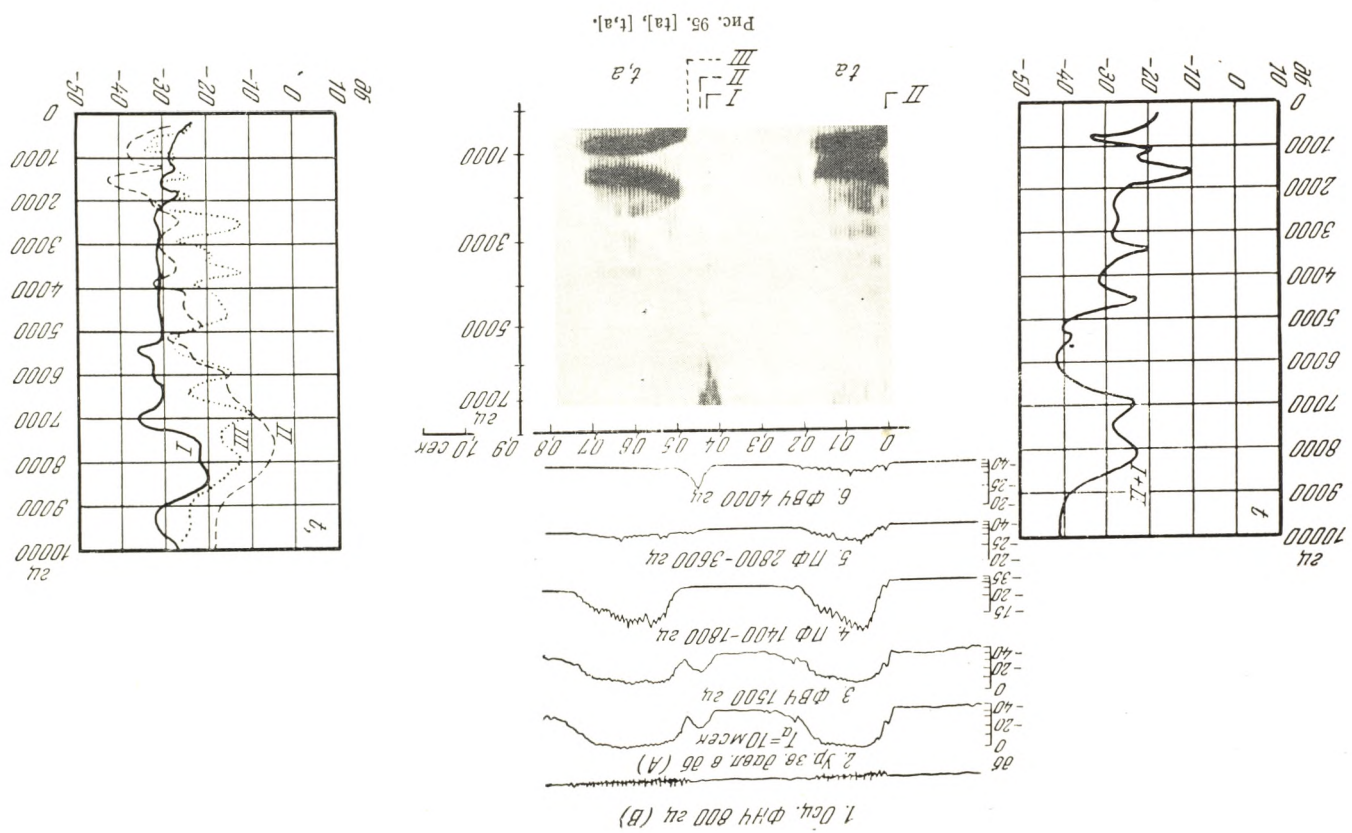
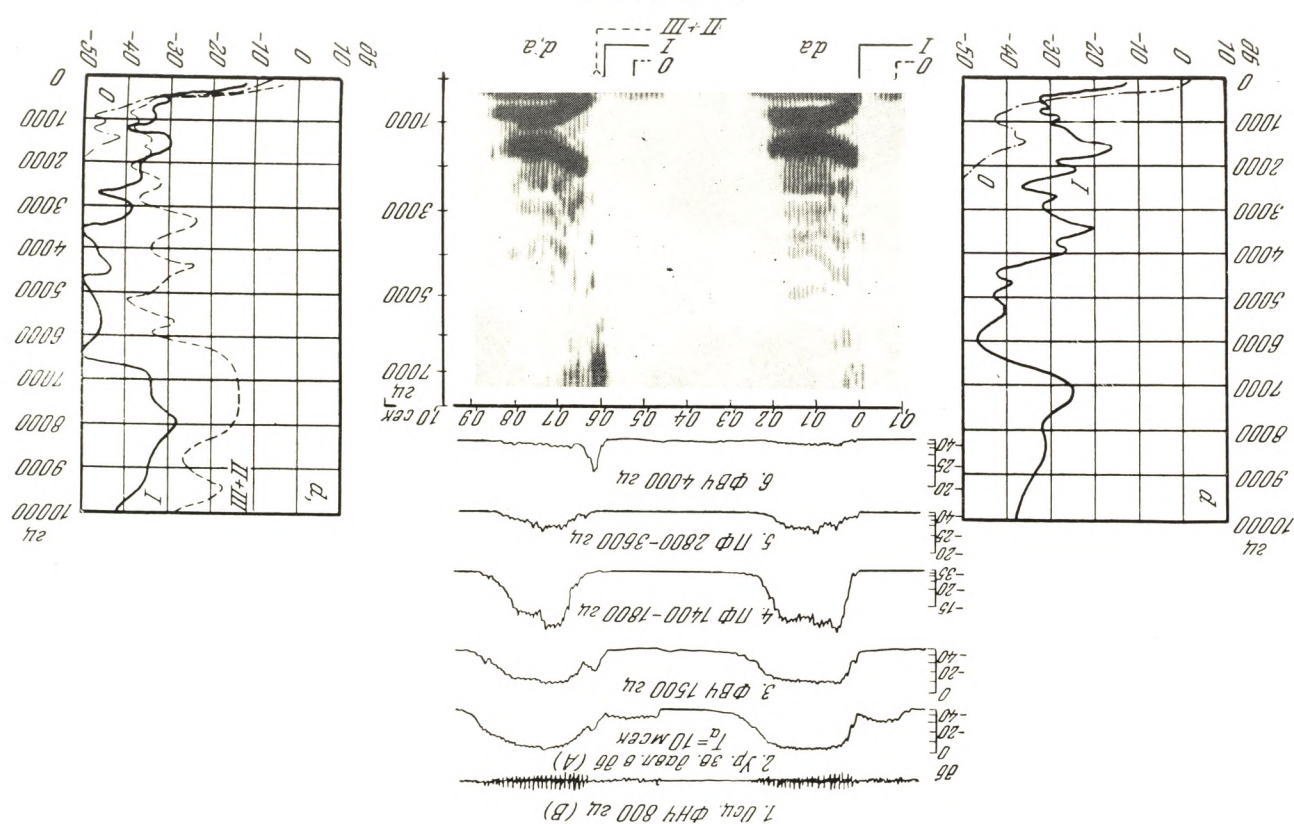
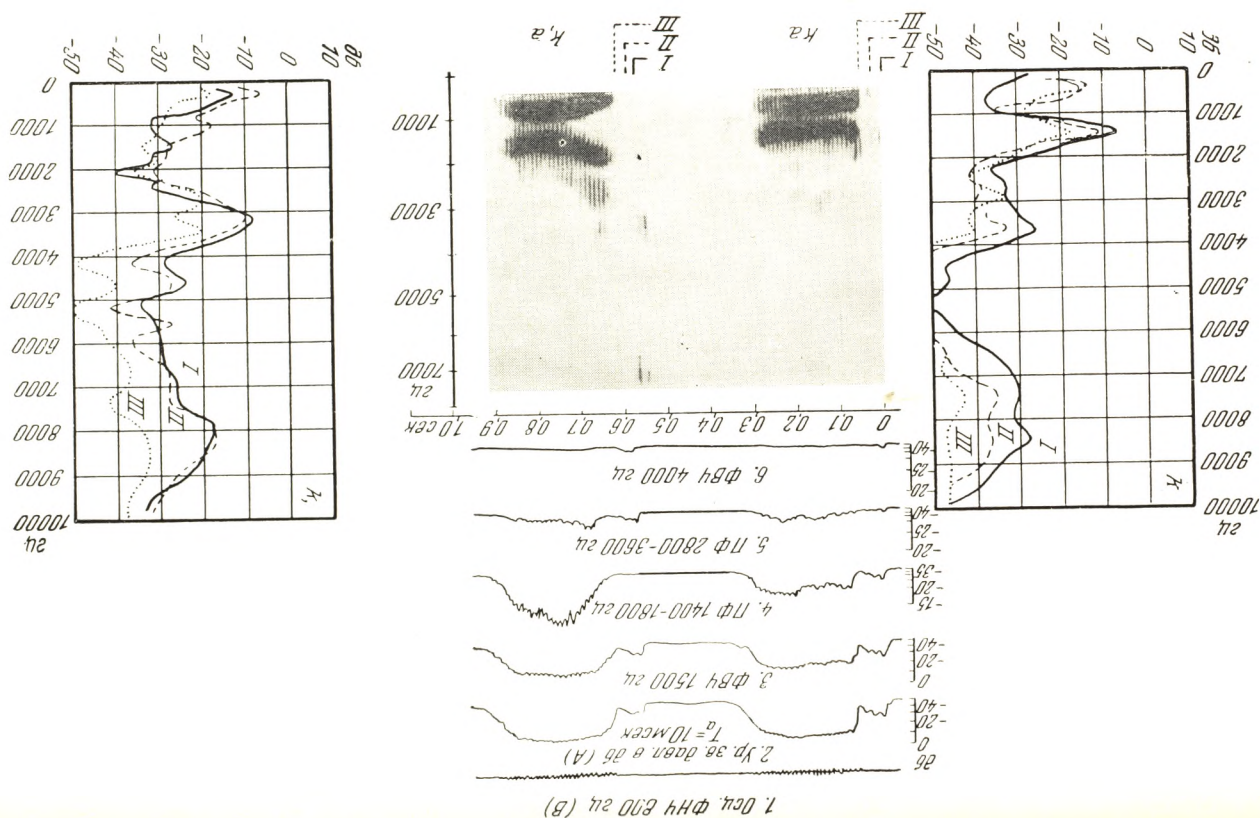


Рис. 90. [са], [с'а].







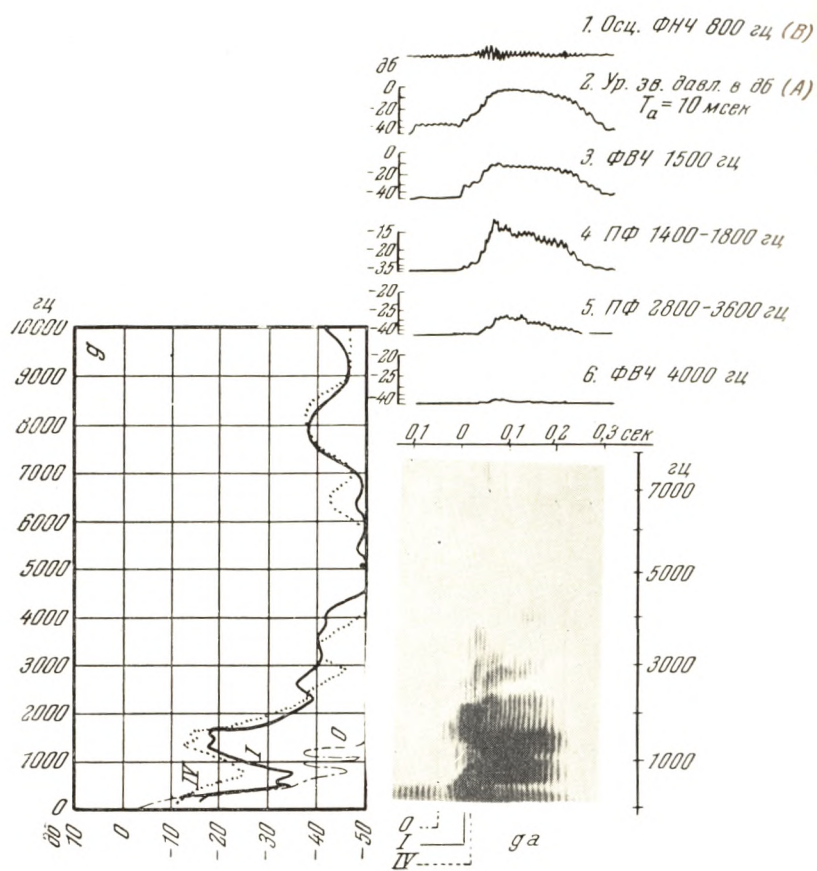


Рис. 98. [га].

Зависящий от квадрата скорости перепад давления у входного отверстия определяется соотношением

$$p = \frac{\rho}{2} \left(\frac{u}{AK} \right)^2, \quad (\text{II. 14})$$

где K — постоянная, A — площадь отверстия, u — объемная скорость. Средняя скорость частиц меньше, чем их скорость на оси потока, вследствие сжатия последнего. Постоянная K , согласно Вестервельту [208], учитывает частично этот эффект, частично же — потери на трение. В своих экспериментах с малыми отверстиями он нашел, что эта постоянная имеет величину порядка 0,7—0,9. Величину того же порядка получил и Гейнц [95] в опытах с трубками, имитирующими сужение, для пропорционального потоку члена в общем сопротивлении. Однако следует иметь в виду, что при речеобразовании артикуляторное сужение редко имеет резкие границы, поэтому более вероятно, что K ближе к единице, чем приведенная выше величина; таким образом, значения этой постоянной, полученные на моделях гортани в опытах Ван-ден-Берга и др. [15], более соответствуют реальным условиям.

Майер-Эпплер [150] проводил эксперименты с трубами из пластического материала, имевшими сужение; результаты этих экспериментов были подтверждены контрольными опытами на человеке, произносившем три фрикативных согласных: [s], [ʃ] и [f]. Им было показано, что существует критическое значение числа Рейнольдса Re_c , при котором начинается возбуждение шумового источника. Звуковое давление шума P_i , измеренное на расстоянии l от диктора, определяется соотношением

$$P_i = \alpha (Re^2 - Re_c^2), \quad (\text{II. 15})$$

где α — постоянная, а $Re_c \approx 1800$ для моделей из пластика и меньше при образовании естественных звуков.

При наличии турбулентности перепад давления p_d в сужении, определяющийся приближенно формулой (II.12), пропорционален квадрату скорости частиц u . Однако при изучении образования звуков речи удобнее говорить об избыточном давлении позади сужения, чем о скорости частиц. Поэтому выражение (II.15) лучше представить в виде

$$P_i = k_1 h^2 p_d - k_2, \quad (\text{II. 16})$$

где k_1 и k_2 — постоянные. Отсюда видно, что звуковое давление шума пропорционально превышению избыточного давления или квадрату скорости частиц над определенной пороговой величиной, а также пропорционально квадрату эффективной ширины сужения h . Если сужение имеет эллипсоидальную форму, эффективная ширина h приблизительно равна $4A/S$, где S — периметр, A — площадь эллипса. В частном случае, когда эллипс обрамляется в круг, h равна диаметру; для эллипса с большим эксцентриситетом h приближается к величине меньшего диаметра эллипса, умноженной на $\pi/2$. Таким образом, при заданных избыточном давлении и эффективной ширине сужения расход воздуха значительно меньше для круговой формы сужения, чем для эллиптической с большим эксцентриситетом; число Рейнольдса, а значит и интенсивность шума, наоборот, больше для сечения круговой формы, чем для эллиптической, при условии постоянства мощности выдоха. Эта мощность равна произведению избыточного давления на объемную скорость. Значительная величина отношения длины к ширине сужения

может явиться поэтому одним из факторов, определяющих малую интенсивность межзубного фрикативного [θ].

Приведем несколько цифр, характеризующих порядок рассматриваемых переменных. Согласно Майер-Эпплеру [150] минимальное избыточное давление ¹⁾, необходимое для образования фрикативных [s], [ʃ] и [i], равно соответственно $p_d = 1; 1,5$ и 6 см водяного столба; этому соответствуют скорости частиц 1300, 1600 и 3100 см/сек и объемные скорости 130, 250 и 500 см³/сек. При сильном выдохе избыточное давление имеет величину около 20 см водяного столба, скорость частиц — 5000 см/сек.

Если предварительно закрытый проход в речевом тракте раскрывается постепенно, то для поддержания постоянного избыточного давления позади прохода необходимо увеличение потока воздуха в соответствии с изменением площади сужения в том его месте, где она минимальна. Это требует увеличения мощности, расходуемой дыхательными мышцами; последняя, однако, имеет предел, чем и обуславливается максимальная величина потока при заданном избыточном давлении. С другой стороны, создаваемое при выдохе максимальное избыточное давление уменьшается с увеличением площади сужения. Наибольшие значения давления имеют место при смычке, достигая величины порядка 130 см водяного столба.

Эти ограничения, налагаемые на переменные, входящие в формулу (II.16), приводят к существованию определенной оптимальной эффективной ширины. Дальнейшее увеличение h приводит к уменьшению произведения $h^2 p_d$, а значит и уровня шума.

Из того обстоятельства, что интенсивность шума источника зависит от величины потока, следует, что в звонких фрикативных шум должен быть модулирован по интенсивности с частотой основного тона голоса. Иными словами, звуковое давление шумовых составляющих флюктуирует, причем величина флюктуаций определяется превышением квадрата скорости частиц над пороговым значением; последнее определяется критической скоростью, необходимой для возникновения турбулентности. Спектрограммы для фрикативных звонких согласных были приведены выше. На широкополосных спектрограммах видно, что вертикальные штрихи в области высокочастотных формант, например, для звука [z], появляются синхронно с вертикальными бороздами, образующими тонкую структуру формантных полос F_1, F_2, F_3 у сонорных звуков.

Фрикативный участок взрывного звука тождествен фрикативному длительному с той же артикуляцией, но только короче; последнюю часть шумового интервала, предшествующую появлению голоса в аспирированных взрывных, следует считать коротким звуком [h], как уже отмечалось выше. Вообще говоря, характер нарастания и спада турбулентного звука, так же как и его длительность, существенно зависят от скорости движения органов артикуляции и от того, сколько времени они остаются в положении, соответствующем эффективному шумообразованию. Дополнительным фактором, влияющим на длительность звуков данного типа, является положение этого звука в пределах дыхательного цикла.

¹⁾ Существование порога того же порядка величины, так же как и правильность уравнения (II.12), были в последнее время подтверждены собственными экспериментами автора настоящей работы. Нужно, однако, заметить, что при построении зависимости давления от потока в логарифмических масштабах координат порог мало заметен.

У взрывных звуков временные характеристики потока, возникающего при выходе воздуха, накопленного за время смычки, зависят, кроме указанных выше факторов, еще и от объема полостей, в которых происходит накопление воздуха, а также от начального избыточного давления. Процесс «взрыва» аналогичен процессу разряда конденсатора через нелинейное сопротивление, величина которого является функцией времени.

Этот вопрос рассматривал К. Н. Стивенс [191] на основе закона Бойля и закона сохранения массы. Он получил уравнение

$$\frac{d}{dt} V \left(1 + \frac{p(t)}{P_0} \right) = -u(t), \quad (\text{II. 17})$$

где P_0 — атмосферное давление, V — объем позади сужения. При этом предполагается, что дыхательные мышцы только уравнивают образовавшееся избыточное давление, но не затрачивают дополнительной энергии на вытеснение воздуха.

При этом условии количество воздуха, выдыхаемого при размыкании смычки, мало, так что можно считать объем V приблизительно постоянным. Если, сверх того, принять, что в первом приближении, при достаточно больших скоростях потока, перепад давления $p(t)$ и скорость частиц $v(t)$ связаны формулой (II.2), то мы получим простое дифференциальное уравнение

$$dv = -\frac{P_0}{\rho V} A(t) dt. \quad (\text{II. 18})$$

Полагая, что наименьшее сечение в суженном проходе изменяется от 0 до максимальной величины A_0 по экспоненциальному закону, т. е. что $A(t) = A_0(1 - e^{-t/t_i})$, получим решение уравнения (II.18) в виде

$$v_0 - v(t) = \frac{P_0 A_0 t}{\rho V} \left[1 - \frac{t_i}{t} (1 - e^{-t/t_i}) \right], \quad (\text{II. 19})$$

где $v_0 = \sqrt{2p_0/\rho}$ — начальная скорость частиц в момент $t = 0$, p_0 — начальное избыточное давление. В момент $t = t_i$ площадь $A(t)$ достигает относительной величины, равной $1/e$, или 37% от предельного значения A_0 , а перепад давления составляет только 37% от того значения, которое имело бы место в идеальном случае изменения площади от 0 до A_0 при $t = 0$. Если время размыкания t_i пренебрежимо мало, то скорость частиц $v(t)$ падает линейно со временем, перепад же давления $p(t)$ и, следовательно, звуковое давление должны падать по параболическому закону. В случае конечного t_i и малого A_0 время нарастания шума будет того же порядка, что и t_i ; действительно, пока h возрастает быстрее, чем падает v , увеличение ширины сужения h будет являться определяющим фактором изменения интенсивности шума (см. формулу (II.16)). Длительность периода нарастания для зубных оказывается больше, чем для всех других взрывных.

На последней стадии выравнивания давлений течение становится ламинарным и вследствие постоянства сопротивления поток спадает по экспоненциальному закону. В качестве приближенной меры длительности турбулентного интервала можно принять время, необходимое для падения скорости частиц и вместе с тем избыточного давления, до нуля,

пренебрегая, таким образом, наличием ламинарного потока в заключительном периоде. Согласно формуле (II.19) это время равно

$$t_T = \frac{\rho V}{P_0 A_0} \sqrt{2P_0/\rho}. \quad (\text{II. 20})$$

Если голосовая щель открыта, то объем V , являющийся объемом легких, составляет примерно 4000 см^3 .

Рассмотрим образование сильного глухого взрывного звука. Примем, что для такого звука избыточное давление составляет 6 см водяного столба и площадь поперечного сечения прохода равна $A_0 = 0,1 \text{ см}^2$, что соответствует зубному взрывному. Используя эти величины, получим время спадания звука равным $t_T = 130 \text{ мсек}$, если допустить, что площадь возрастает скачком. Согласно формуле (II.2) идеально линейное спадание начинается со скорости 3200 см/сек . Расчеты показывают, что большое сопротивление сужения вызывает демпфирование, значительно превышающее критическую величину для $F1$. Постоянная R/L для сужения дает время нарастания порядка 1 мсек , т. е. намного меньше того времени, которое в обычных условиях речеобразования требуется для того, чтобы удалить язык от места смычки. Через 50 мсек скорость частиц v достигает величины 1300 см/сек ; согласно Майер-Эпплеру [150] это является, по-видимому, нижним пределом скорости, при которой еще возможно образование зубного шума. В этот момент количество израсходованного воздуха составит только 11 см^3 . Как можно видеть из спектрограмм, а также из рис. 100 (стр. 246), приведенные временные данные дают правильный порядок величин.

Длительность спадания для взрывного звука, как отметил Стивенс [191], очень сильно зависит от того, в каком состоянии находится голосовая щель при раскрытии смычки. При закрытой голосовой щели, принимая $V = 70 \text{ см}^3$ и $p_i = 3 \text{ см}$ водяного столба, что, вероятно, справедливо для слабых звонких взрывных, воспользоваться формулой (II.19) уже нельзя, так как начальная скорость, равная при этих условиях примерно 800 см/сек , едва ли достаточна для возникновения турбулентности. Демпфирование при этом, в силу малого объема, ниже критического, так что спадание звука имеет колебательный характер, определяющийся в основном первой формантой; общая длительность спадания — порядка 5 мсек .

Противопоставление взрывных по признаку напряженный — ненапряженный (*tense* — *lax*), иногда рассматриваемое как противопоставление сильный — слабый (*fortis* — *lenis*), определяется длительностью шумового интервала. Интересно отметить, что необходимое различие в этой длительности может быть обеспечено, как показывает уравнение (II.20), либо открытием или соответственно закрытием голосовой щели, либо уменьшением или соответственно увеличением скорости возрастания площади поперечного сечения в артикуляторном сужении, либо, наконец, увеличением или соответственно уменьшением избыточного давления позади места смычки. Любая из этих возможностей позволяет в нужной степени варьировать время спадания звука.

Если в момент раскрытия появляется дополнительно импульс выдоха, то это может вызвать удлинение периода спада звука в силу поддержания избыточного давления приблизительно на первоначальном уровне. Если при этом площадь поперечного сечения прохода остается

достаточно малой, образуется аффрикация, а если ширина прохода быстро достигает величины, превосходящей критическую, появляется явно выраженный аспираторный интервал.

Турбулентный шум при образовании взрывных является вторичным эффектом. Импульс воздушного потока при раскрытии смычки сам по себе может рассматриваться как источник; в электротехнике такой импульсный источник представляется внезапно приложенной к цепи постоянной э. д. с. В электрической эквивалентной схеме речевого тракта этому соответствует замыкание ключа, расположенного в месте артикуляторного сужения. Тогда часть речевого тракта, соответствующая сужению и задней полости, может быть заменена двухполюсником, состоящим из последовательно включенных постоянной э. д. с., ключа, импеданса сужения и входного (со стороны сужения) импеданса задней полости. Этот двухполюсник образует одну из ветвей эквивалентной схемы, а постоянная э. д. с. в этой ветви соответствует избыточному давлению, образующемуся во время смычки. Подобный источник характеризуется во временном представлении ступенчатой функцией, а в спектральном — спаданием огибающей спектра на 6 дБ/окт. Поскольку в действительности время раскрытия смычки конечно, крутизна спада будет еще больше.

С акустической точки зрения начальное избыточное давление можно рассматривать как равномерно распределенный в задних полостях источник [109]. Однако приведенный выше электрический эквивалент, который легко получить на основании теоремы Тевенена, более удобен для расчета частотной характеристики системы.

Правильность предположения о том, что в качестве характеристики импульсного источника следует принять ступенчатую функцию, положенного в основу приведенных выше расчетов, подтверждается результатами, полученными в настоящей работе.

Спектры турбулентных источников звука менее ясны; однако результаты наших расчетов согласуются с данными, полученными на моделях в работах MIT (см., например, [96]). По этим данным основная часть шумового спектра, создаваемого турбулентным источником, плоская с некоторым падением к низким и высоким частотам.

Импеданс источника оказывает определенное влияние на спектр образующегося звука; влияние это особенно заметно на низких частотах. Для взрывных и длительных фрикативных импеданс источника включает индуктивность сужения, в котором возникает шум, и активную составляющую. Последняя, согласно формуле (II.10), в первом приближении равна удвоенному сопротивлению потока; см. также формулы (II.12) — (II.14).

Синхронные записи интенсивности звука и избыточного давления при произнесении взрывных и фрикативных¹⁾ показаны на рис. 100. Интервал взрыва короче для начального [g], чем для [k], почти отсутствует у начальных [b] и [d] и имеет заметную величину для [p] и [t]. Несмотря на то, что избыточное давление в момент взрыва оглушенного [d] больше, чем для [t], время спада интенсивности у [d] мало и не превосходит, а может быть и меньше, времени интегрирования зонда, применявшегося для измерения давления (15 мсек). Для начальных же

¹⁾ Такие записи интенсивности и избыточного давления выполнялись Фишер-Йоргенсен (личное сообщение). Малеко [142] применял носовой зонд (катетер) для кимографической записи давления. Стэтсон [190] использовал ротовой зонд.

[p], [t], [k] длительность спада составляет примерно 70 мсек; только на этом расстоянии от момента взрыва появляются колебания голосовых связок. Для конечных [p], [t], [k] спадание давления от начальной величины 5—7 см водяного столба продолжается около 100—200 мсек.

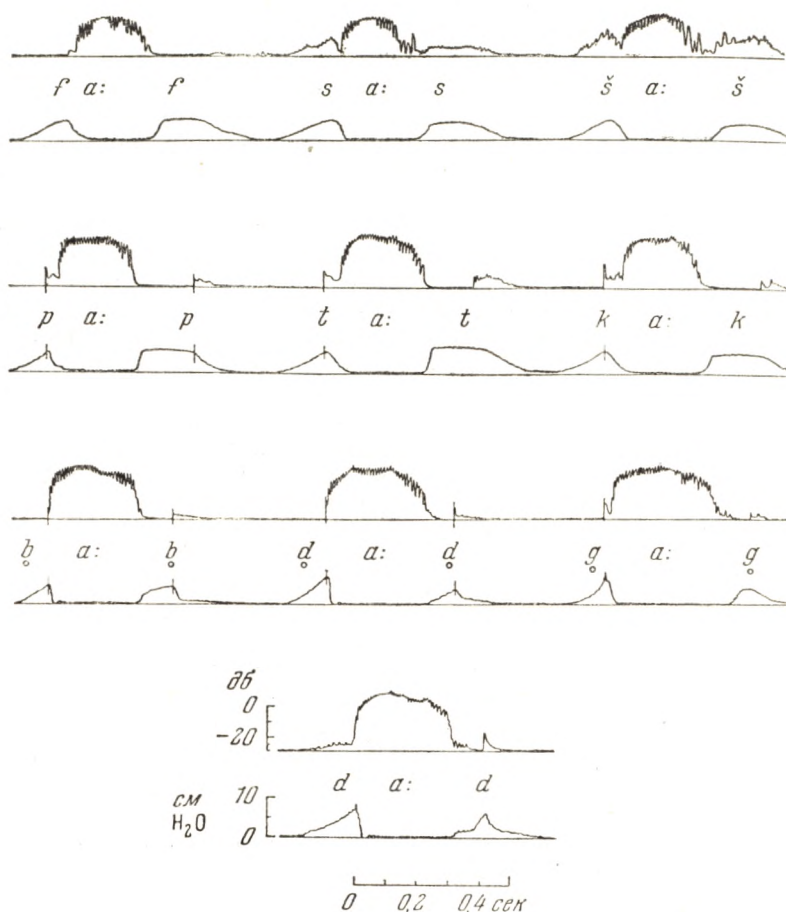


Рис. 100. Синхронные осциллографические записи (на «Мингографе»). Верхняя кривая: интенсивность звука (частотная характеристика А, сжатая амплитудная шкала). Нижняя кривая: давление в области фаринкса (получена при помощи носового катетера). Кривые снимались при произнесении односложных слов типа согласный—[a:]—согласный, где согласный—один из звуков [f], [s], [ʃ], [p], [t], [k], [b], [d], [g]. Слова произносились диктором-шведом. Согласные [b], [d], [g] представляют собой глухие варианты соответственных звуков, т. е. произнесены при отсутствии голоса во время смычки.

Длительность смычки у конечных взрывных вдвое больше для [p], [t] и [k], чем для [b], [g], [d], однако отрезок времени от момента начала гласного до момента взрыва почти одинаков для всех этих звуков.

Ход изменения интенсивности фрикативных во времени повторяет ход изменений давления и имеет порог, соответствующий давлению порядка 1 см водяного столба. Интенсивность [f], как правило, мала по сравнению с [s] и [ʃ].

III. АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПРОСТЫХ МОДЕЛЕЙ РЕЗОНАТОРОВ ПРИМЕНИТЕЛЬНО К МЕХАНИЗМУ РЕЧЕОБРАЗОВАНИЯ

§ 1. Одиночный резонатор Гельмгольца

А. Одно отверстие

Теоретической моделью, широко используемой для описания резонансных явлений при образовании звуков речи, является резонатор Гельмгольца. Существует обширная литература, относящаяся к физическим свойствам этого класса резонаторов. Затухание и концевая поправка для простых резонаторов подробно рассматривались Нильсеном [157] и Ингардом [108], причем последний особое внимание уделил нелинейным эффектам. Однако ни простые, ни связанные резонаторы, имеющие размеры, при которых необходимо учитывать конечную скорость распространения звука, не изучены в такой же степени. Задачей настоящего раздела является по возможности полное описание простых резонаторов с сосредоточенными постоянными (типа резонаторов Гельмгольца) и с распределенными постоянными (типа линий передачи).

При математическом рассмотрении резонаторов мы будем пользоваться представлением их в виде эквивалентных схем. Одной из основных задач анализа таких схем является определение частот резонансов и антирезонансов, а также постоянных затухания, т. е. в конечном итоге комплексных нулей и полюсов системы. Первым шагом для решения этой задачи является получение в явной форме выражения для функции передачи схемы, определяющей отношение тока на выходе, соответствующем ветви с импедансом излучения, к току или напряжению, создаваемому источником. После этого комплексная частота $s = \sigma + j\omega$ заменяется в этом выражении на круговую частоту $j\omega$ и вещественная и мнимая части его приравняются порознь нулю.

Функция передачи одиночного резонатора, показанного на рис. 101, имеет вид

$$H(s) = \frac{U_0(s)}{U_1(s)} = \frac{1}{1 + RG + s(LG + RC) + s^2LC} = \frac{\omega_{01}^2}{(s - \hat{s}_1)(s - \hat{s}_1^*)}. \quad (\text{III. 1})$$

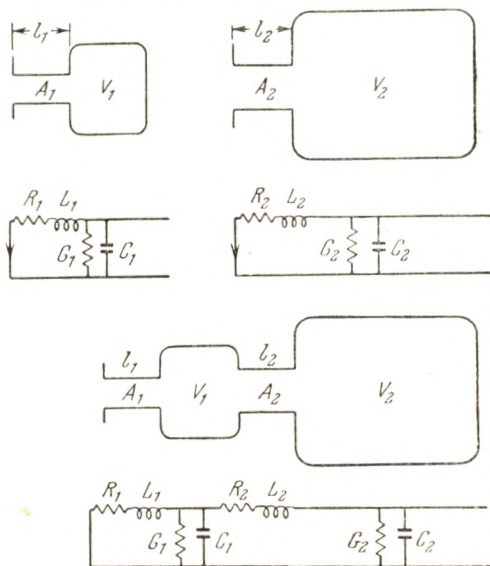


Рис. 101. Классический двоянный резонатор Гельмгольца и его эквивалентная электрическая схема.

1. Если затухание меньше критического, то

$$\left. \begin{aligned} \frac{1+RG}{LC} &> \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2, \\ \hat{s}_1, \hat{s}_1^* &= -\frac{1}{2} \left(\frac{R}{L} + \frac{G}{C} \right) \pm j \sqrt{\frac{1+RG}{LC} - \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2}, \\ \hat{s}_1 &= \sigma_1 + j\omega_1, \quad \hat{s}_1^* = \sigma_1 - j\omega_1. \end{aligned} \right\} \quad (\text{III. 2})$$

Резонаторы речевого тракта обычно характеризуются большим $Q = \frac{\omega_0}{2\sigma}$, что означает $RG \ll 1$ и $\omega_0^2 = \frac{1}{LC} \gg \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2$. В этом случае частота резонанса и ширина резонансной кривой имеют значения

$$\left. \begin{aligned} F_1 &= \frac{1}{2\pi} \omega_1 = \frac{1}{2\pi \sqrt{LC}}, \\ B_1 &= \frac{-1}{\pi} \sigma_1 = \frac{R}{2\pi L} + \frac{G}{2\pi C}. \end{aligned} \right\} \quad (\text{III. 3})$$

Переходя к акустическим параметрам, т. е. подставляя $L = \frac{\rho l_e}{A}$ и $C = \frac{V}{\rho c^2}$, найдем:

$$F_1 = \frac{c}{2\pi} \sqrt{\frac{A}{l_e V}}, \quad (\text{III. 4})$$

как это было показано выше.

Из этого равенства видно, что частота резонанса обратно пропорциональна квадратному корню из объема V и эффективной длины l_e и прямо пропорциональна квадратному корню из площади поперечного сечения A .

Зависимость ширины B резонансной кривой, определяющейся вторым из выражений (III.3), от параметров схемы мы рассмотрим на примере изменения свойств резонатора вследствие уменьшения площади поперечного сечения его горла.

Если считать сопротивление R пропорциональным $f^{\frac{1}{2}} A^{-3/2}$, что соответствует классическому случаю потерь на вязкость (см. формулу (2.13), то ширина резонансной кривой будет изменяться пропорционально $f^{-1/2}$ или пропорционально $A^{-1/4}$, т. е. обратно пропорционально квадратному корню из радиуса горла, если объем полости резонатора и длина горла остаются постоянными.

Если считать R не зависящим от частоты, то ширина полосы будет изменяться пропорционально $1/f$, т. е. Q будет пропорционально f^2 ; влияние излучения существенно только для частот выше 1000 гц и возрастает пропорционально A и f^2 при постоянном объеме полости. Поэтому Q должно достигать максимума при некоторой промежуточной частоте, как это нашел Ингард [108]. Вместе с тем имеется и некоторое определенное значение радиуса горла, при котором ширина резонансной кривой минимальна, что можно видеть из приведенных выше данных. Более детальное рассмотрение вопросов демпфирования будет дано в § 6, А.

Демпфирование очень низких формант, таких, как $F1$ для звонких согласных, теоретически может быть выше критического; последнее определяется условиями $\hat{s}_1 = \hat{s}_1^*$ и $\omega_1 = 0$.

В некоторых случаях может оказаться целесообразным использовать аналогию с согласованно нагруженным полужвеном низкочастотного фильтра-прототипа, для которого $\sigma_1 = \omega_1$.

2. Если затухание больше критического, то

$$\left. \begin{aligned} \frac{1+RG}{LC} &< \frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2, \\ \hat{s}_1, \hat{s}_1^* &= -\frac{1}{2} \left(\frac{R}{L} + \frac{G}{C} \right) \pm \sqrt{\frac{1}{4} \left(\frac{R}{L} + \frac{G}{C} \right)^2 - \frac{1+RG}{LC}}. \end{aligned} \right\} \quad (\text{III. 5})$$

Таким образом, корни вещественны и не равны друг другу, причем один из них увеличивается, а другой уменьшается при возрастании потерь. Если основным фактором, обуславливающим демпфирование, являются потери в элементе схемы R , причем вносимое им затухание велико, то полюсы имеют место при $-\frac{R}{L}$ и $-\frac{1}{RC}$.

Подобное сверхкритическое затухание, вероятно, имеется для первой форманты звонкой смычной. Измерение частот полюсов при таких условиях, очевидно, весьма затруднительно.

Б. Два отверстия

Добавление в эквивалентной схеме ветви $R_2 + sL_2$ параллельно емкости вызывает повышение резонансной частоты и появление полюса α_1 и нуля α_2 на отрицательной вещественной оси. Принимая потери малыми, получим:

$$H(s) = \frac{L_2(s - \alpha_2)}{(L_1 + L_2)(s - \alpha_1)} \cdot \frac{(\omega_1^2 + \sigma_1^2)}{(s - \hat{s}_1)(s - \hat{s}_1^*)}, \quad (\text{III. 6})$$

где

$$\begin{aligned} \alpha_2 &\approx -\frac{R_2}{L_2}, \quad \alpha_1 \approx \frac{R_1 + R_2}{L_1 + L_2}, \\ \omega_1 &\approx \sqrt{\frac{L_1 + L_2}{L_1 L_2} \cdot \frac{1}{C}}, \\ \sigma_1 &\approx -\frac{1}{2} \left(\frac{G}{C} + \frac{R_2}{L_2} \cdot \frac{L_1}{L_1 + L_2} + \frac{R_1}{L_1} \cdot \frac{L_2}{L_1 + L_2} \right), \\ \hat{s}_1, \hat{s}_1^* &= \sigma_1 \pm j\omega_1. \end{aligned}$$

После перехода к акустическим параметрам выражение для частоты резонанса можно представить в виде

$$f_1 = \frac{c}{2\pi} \sqrt{\frac{1}{V} \left(\frac{A_1}{l_{1e}} + \frac{A_2}{l_{2e}} \right)}, \quad (\text{III. 7})$$

где l_{2e} — длина, A_2 — площадь поперечного сечения второго горла. Резонансная частота растет с увеличением индексов проводимости A/l каждого из отверстий резонатора.

Если потери распределены равномерно между первым и вторым горлом, т. е. $\frac{R_1}{L_1} = \frac{R_2}{L_2}$, то выражение для функции передачи такое же, как в случае простого резонатора, как можно видеть из замечаний по вопросу о назализации, приведенных в гл. 8.

Если источник рассматривается как генератор заданной э. д. с., включенной последовательно с $R_2 L_2$, то единственное изменение, которое

должно быть сделано в выражении для функции передачи системы, сводится к замене множителя $(s - \alpha_2)$, характеризующего нуль, постоянной $1/L_2$:

$$H(s) = \frac{1}{R_1 + R_2} \cdot \frac{\alpha_1}{s - \alpha_1} \cdot \frac{(\omega_1^2 + \sigma_1^2)}{(s - \hat{s}_1)(s - \hat{s}_1^*)} = \\ = \frac{1}{L_1 + L_2} \cdot \frac{1}{s - \alpha_1} \cdot \frac{(\omega_1^2 + \sigma_1^2)}{(s + \sigma_1)^2 + \omega_1^2}. \quad (\text{III. 8})$$

§ 2. Сдвоенный резонатор Гельмгольца

Сдвоенный резонатор без потерь был рассмотрен впервые Рэле-ем (1877). Он широко использовался в фонетической литературе, например при рассмотрении зависимости частот первых двух формант гласных от размеров полостей речевого аппарата Пэджетом [160], Крендаллом [27], Чиба и Каджияма [24], Эсснером [42] и Джюзом [121].

Очевидное ограничение применимости теории сдвоенного резонатора к исследованию речеобразования связано с тем, что размеры речевого тракта, как правило, соизмеримы с длинами волн $\lambda = \frac{c}{f}$ в интересующем нас диапазоне частот и могут считаться достаточно малыми только для низких частот, удовлетворяющих неравенству $f < \frac{c}{8l}$; однако ограничиться только этой областью частот при изучении речеобразования явно невозможно. Кроме того, теория сдвоенного резонатора предполагает наличие явно выраженного горла как на выходе резонатора, так и в месте соединения двух полостей. Первые две форманты задних гласных, а также первая форманта не слишком открытых передних могут быть приближенно найдены при использовании модели в виде простого или сдвоенного резонатора Гельмгольца. Однако расчет F_2 передних гласных на основе теории таких резонаторов может привести к полному несоответствию с действительностью; это было показано в § 3 гл. 7.

Схема, приведенная на рис. 101 (внизу) с добавлением ветви R_3L_3 параллельно C_2 , представляет собой полную эквивалентную схему сдвоенного резонатора с горлом у каждой из обеих полостей. Коэффициент передачи по току для этой схемы

$$H(s) = \frac{L_3}{(L_1 + L_2 + L_3)} \cdot \frac{(s - \alpha_2)}{(s - \alpha_1)} \cdot \frac{(\omega_1^2 + \sigma_1^2)(\omega_2^2 + \sigma_2^2)}{[(s + \sigma_1)^2 + \omega_1^2][(s + \sigma_2)^2 + \omega_2^2]}, \quad (\text{III. 9})$$

где

$$\alpha_2 \approx \frac{-R_3}{L_3}, \quad \alpha_1 \approx \frac{-(R_1 + R_2 + R_3)}{(L_1 + L_2 + L_3)},$$

$$\omega_{1,2} = \left[\frac{\omega_a^2}{2} \pm \left(\frac{\omega_a^4}{4} - \omega_b^4 \right)^{\frac{1}{2}} \right]^{\frac{1}{2}},$$

$$\omega_a^2 = \omega_1^2 + \omega_2^2 = \frac{1}{L_1 C_1} + \frac{1}{L_2 C_2} + \frac{1}{L_3 C_2} + \frac{1}{L_2 C_1},$$

$$\omega_b^4 = \frac{L_1 + L_2 + L_3}{L_1 L_2 L_3 C_1 C_2} = \omega_1^2 \omega_2^2,$$

$$\sigma_1 \approx \frac{1}{2(\omega_2^2 - \omega_1^2)L_1C_1} \left\{ \frac{R_1}{L_1} \left(\frac{L_1}{L_2} + \frac{L_1C_1}{L_2C_2} + \frac{L_1C_1}{L_3C_2} - \omega_1^2L_1C_1 - \frac{L_1/L_3}{\omega_1^2L_2C_2} \right) + \right. \\
+ \frac{R_2}{L_2} \left(1 + \frac{L_1C_1}{L_3C_2} - \omega_1^2L_1C_1 - \frac{L_2/L_3}{\omega_1^2L_2C_2} \right) + \\
+ \frac{R_3}{L_3} \left(1 + \frac{L_1}{L_2} + \frac{L_1C_1}{L_2C_2} - \omega_1^2L_1C_1 - \frac{1}{\omega_1^2L_2C_2} \right) + \\
\left. + \frac{G_1}{C_1} \left(\frac{L_1C_1}{L_3C_2} + \frac{L_1C_1}{L_2C_2} - \omega_1^2L_1C_1 \right) + \frac{G_2}{C_2} \left(1 + \frac{L_1}{L_2} - \omega_1^2L_1C_1 \right) \right\}. \quad (\text{III. 10})$$

Выражение для σ_2 можно получить, заменив здесь ω_1 на ω_2 в выражении, заключенном в фигурные скобки. Влияние α_1 и α_2 заметно только на очень низких частотах, как это имело место в § 1, Б. Рассмотрим соотношения между элементами эквивалентной схемы и резонансными частотами сдвоенного резонатора.

Резонансная частота переднего резонатора, взятого изолированно, равна

$$F_{01} = \frac{1}{2\pi\sqrt{L_1C_1}}. \quad (\text{III. 11})$$

Аналогично резонансная частота заднего резонатора будет

$$F_{02} = \frac{1}{2\pi\sqrt{L_2C_2}}. \quad (\text{III. 12})$$

Ветвь L_3R_3 представляет собой импеданс голосовой щели, который имеет большую величину. Если индуктивность голосовой щели L_3 конечна, то F_{02} возрастает:

$$F_{02} = \frac{1}{2\pi\sqrt{L_2C_2}} \cdot \sqrt{\frac{L_2+L_3}{L_3}}. \quad (\text{III. 13})$$

В дальнейшем ветвь L_3R_3 или вообще не будет учитываться, или же будет учитываться как увеличение проводимости ветви L_2R_2 путем подключения параллельно к ней большого импеданса. Если $L_3+L_2 \gg L_1$, анализ можно проводить, не считаясь с наличием ветви L_3R_3 .

Резонансные частоты всей системы, при наличии связи между двумя ее частями, обозначим через F_1 и F_2 , причем примем $F_2 > F_1$; F_{01} не связано с F_{02} никакими условиями. Из формулы (III.10)

$$\left. \begin{aligned} F_1^2 + F_2^2 &= \frac{1}{4\pi^2} \left(\frac{1}{L_1C_1} + \frac{1}{L_3C_2} + \frac{1}{L_2C_1} \right), \\ F_1^2 F_2^2 &= \left(\frac{1}{4\pi^2} \right)^2 \frac{1}{L_1C_1L_2C_2} = F_{01}^2 F_{02}^2. \end{aligned} \right\} \quad (\text{III. 14})$$

Введем коэффициенты, учитывающие отношение индуктивностей:

$$k_L = \frac{L_1}{L_2} = \frac{l_{1e}A_2}{l_{2e}A_1} \quad (\text{III. 15})$$

и отношение емкостей

$$k_C = \frac{C_1}{C_2} = \frac{V_1}{V_2}. \quad (\text{III. 16})$$

Последний коэффициент представляет собою не что иное, как отношение объемов полостей обеих частей.

Резонансные частоты могут быть после этого представлены в виде

$$F_{1,2} = \frac{1}{4\pi^2L_1C_1} \left[\frac{1}{2}(1 + k_L + k_Lk_C) \mp \sqrt{\frac{1}{4}(1 + k_L + k_Lk_C)^2 - k_Lk_C} \right]. \quad (\text{III. 17})$$

Отсюда видно, что при введении связи между резонаторами более высокая резонансная частота увеличивается, а более низкая — уменьшается. Однако произведение резонансных частот для связанных резонаторов остается таким же, как и для несвязанных. На рис. 102 приведен коэффициент сдвига резонансных частот $\frac{F_{01}}{F_1} = \frac{F_2}{F_{02}}$

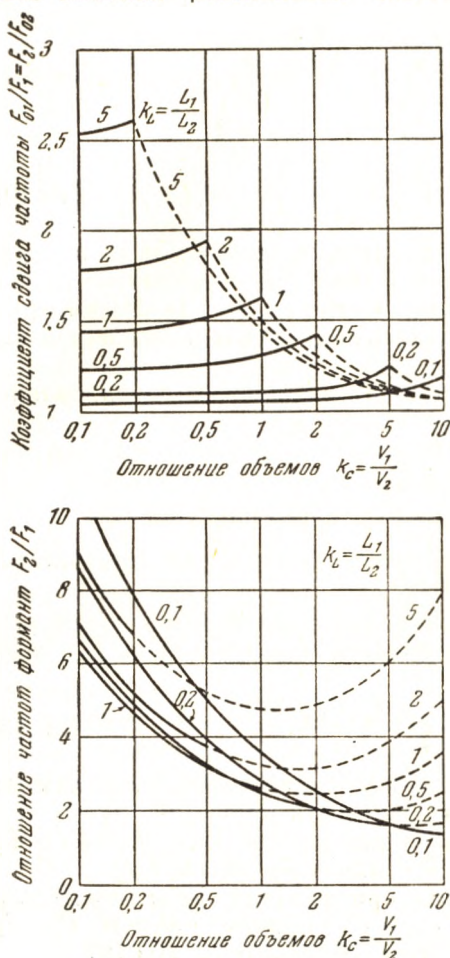


Рис. 102. Номограммы для определения частот первой форманты F_1 и второй форманты F_2 и их отношения для двоянного резонатора Гельмгольца. F_{01} означает более низкую, F_{02} — более высокую резонансную частоту каждого из входящих в состав системы простых резонаторов при отсутствии связи между ними. Следует отметить, что частоту резонанса передней полости могут представлять как F_{01} , так и F_{02} . Далее V_1 и V_2 — объемы передней и задней полостей; $L_1 = \frac{l_1}{A_1}$ и $L_2 = \frac{l_2}{A_2}$ — индуктивности горловины переднего и заднего резонаторов. Таким образом, $k_L = \frac{L_1}{L_2} = \frac{l_1 A_2}{l_2 A_1}$. После численного определения F_{01} и F_{02} F_1 и F_2 могут быть получены по верхнему графику, показывающему, насколько сдвигаются резонансы при наличии связи между полостями.

= 630 гц и $F_2 = 1000$ гц, необходимо обеспечить отношение $V_1/V_2 = 5$. Такое большое отношение объемов не может быть получено путем уменьшения размеров передней и увеличения размеров задней по-

уменьшается. Однако произведение резонансных частот для связанных резонаторов остается таким же, как и для несвязанных. На рис. 102 приведен коэффициент сдвига резонансных частот $\frac{F_{01}}{F_1} = \frac{F_2}{F_{02}}$ в случае $F_{02} > F_{01}$ (сплошные кривые на рисунке) или $\frac{F_{02}}{F_1} = \frac{F_2}{F_{01}}$ в случае $F_{02} < F_{01}$ (штриховые кривые). Используя эти кривые, легко определить резонансные частоты двоянного резонатора; для этого находят численно резонансные частоты F_{01} и F_{02} каждого из резонаторов при отсутствии связи между ними, после чего при помощи коэффициентов, приведенных на рис. 102, определяются резонансные частоты связанной системы. Внизу на том же рисунке приведено отношение F_2/F_1 . Если F_2 значительно отличается от F_1 или же значительно отличаются объемы полостей, частотные сдвиги невелики.

Приведенные здесь соотношения дают представление о тех артикуляторных условиях, которые обуславливают малую величину F_2/F_1 , имеющую место для задних гласных [о], [э], [а]. Для этого оба резонатора должны иметь примерно одинаковые величины несвязанных резонансных частот, т. е. в пределе должно иметь место условие $F_{01} = F_{02}$; при этом как один, так и другой резонатор одинаково влияют на F_1 и F_2 . Чтобы F_2/F_1 достигло значения, близкого к 1, отношение объемов резонаторов должно быть возможно больше, что достигается уменьшением объема задней полости путем приближения задней части языка к стенке фаринкса. Чтобы получить отношение $F_2/F_1 = 1,6$, как это имеет место, например, для гласного [а], для которого $F_1 =$

лости. Теоретически остаются две возможности объяснить наблюдаемое отношение частот формант. Одной из них является учет влияния конечной индуктивности голосовой щели L_3 ; она может играть известную роль, если голосовая щель остается открытой в течение достаточно большой части периода основной частоты голоса, при слабом воздушном потоке, когда возрастание сопротивления голосовой щели за счет нелинейных эффектов минимально. Увеличение проводимости голосовой щели приводит к тому, что F_2 повышается больше, чем F_1 .

Второй возможностью является допущение некоторой связи с носовой полостью; это, как показывают расчеты, может дать аналогичный эффект.

Некоторый общий интерес представляет исследование вопроса о том, в какой мере каждый из резонансов сдвоенного резонатора обусловлен тем или иным из входящих в его состав единичных резонаторов.

Связь эта не так проста, как это обычно предполагается фонетической теорией, в которой F_1 считается обусловленной задней, а F_2 — передней полостью.

Приведенный ниже математический анализ этого вопроса относится только к моделям, состоящим из двух резонаторов Гельмгольца. Сначала, однако, следует несколько уточнить терминологию. Термином «полость», часто встречавшимся выше, будем обозначать отдельный одиночный резонатор в целом, т. е. собственно полость и относящееся к ней горло. Задняя полость входит в дальнейшие формулы в виде ее объема V_2 или его электрического эквивалента, емкости C_2 . Сужение, соединяющее заднюю и переднюю полости, будем считать горлом заднего резонатора; оно входит в формулы в виде эквивалентной индуктивности L_2 . Каждая из частот формант F_1 и F_2 связана со всеми четырьмя элементами эквивалентной схемы L_1 , L_2 , C_1 и C_2 или с геометрическими параметрами акустической системы l_1/A_1 , l_2/A_2 , V_1 и V_2 .

Дифференцируя выражение (III.14), можно определить, какое приращение получают F_1 и F_2 при уменьшении величины каждого из четырех элементов схемы сдвоенного резонатора. Результаты анализа можно представить в виде сводки:

Изменяющийся элемент, выраженный через		Относительный сдвиг частоты	
акустические параметры	электрические параметры	F_1	F_2
V_1	C_1	$\frac{1}{2} A$	$\frac{1}{2} B$
V_2	C_2	$\frac{1}{2} B$	$\frac{1}{2} A$
l_1/A_1	L_1	$\frac{1}{2} C$	$\frac{1}{2} D$
l_2/A_2	L_2	$\frac{1}{2} D$	$\frac{1}{2} C$

Здесь

$$\left. \begin{aligned} A &= -2 \frac{\partial F_1}{F_1} \frac{C_1}{\partial C_1} = -2 \frac{\partial F_2}{F_2} \frac{C_2}{\partial C_2} = \frac{F_{02}^2 - F_1^2}{F_2^2 - F_1^2}, \\ B &= -2 \frac{\partial F_2}{F_2} \frac{C_1}{\partial C_1} = -2 \frac{\partial F_1}{F_1} \frac{C_2}{\partial C_2} = \frac{F_2^2 - F_{02}^2}{F_2^2 - F_1^2}, \\ C &= -2 \frac{\partial F_1}{F_1} \frac{L_1}{\partial L_1} = -2 \frac{\partial F_2}{F_2} \frac{L_2}{\partial L_2} = \frac{F_2^2 - F_{01}^2}{F_2^2 - F_1^2}, \\ D &= -2 \frac{\partial F_2}{F_2} \frac{L_1}{\partial L_1} = -2 \frac{\partial F_1}{F_1} \frac{L_2}{\partial L_2} = \frac{F_{01}^2 - F_1^2}{F_2^2 - F_1^2}. \end{aligned} \right\} \quad (\text{III. 18})$$

Постоянные A , B , C и D могут изменяться от 0 до 1; первый случай соответствует полной независимости частот формант от того или иного из элементов системы, второй — максимальной зависимости. Условие $A=C=0$, $B=D=1$ характеризует переход к одиночному резонатору. Изменение входящей в его состав индуктивности или емкости на 1% вызывает сдвиг резонансной частоты на 0,5%.

В общем случае

$$A + B = C + D = 1. \quad (\text{III. 19})$$

Это показывает, что сумма относительных приращений той и другой резонансной частоты всегда равна половине относительного уменьшения, вызвавшего это приращение элемента системы. Зависимость резонансных частот от параметров резонатора симметрична, т. е. влияние относительного изменения C_1 на F_1 такое же, как C_2 на F_2 , а влияние изменения C_1 на F_2 такое же, как C_2 на F_1 . То же справедливо и для влияния изменений L_1 и L_2 .

В частном случае, когда передняя и задняя полости настроены при отсутствии связи на одинаковые частоты, т. е. когда $F_{01} = F_{02}$ или $\frac{1}{L_1 C_1} = \frac{1}{L_2 C_2}$, постоянные A , B , C , D принимают вид

$$\left. \begin{aligned} A &= D = \frac{1}{1 + F_2/F_1}, \\ B &= C = \frac{1}{1 + F_1/F_2}, \end{aligned} \right\} \quad (\text{III. 20})$$

т. е.

$$A + B = C + D = 1. \quad (\text{III. 21})$$

Из этих соотношений видно, что в данном частном случае относительный сдвиг F_1 и F_2 , обусловленный одновременным уменьшением как объема полости, так и отношения $1/A$ для одного из резонаторов на 1%, равен +1%. Это означает, что передний и задний резонаторы в одинаковой мере влияют на F_1 и F_2 , если в несвязанном состоянии их резонансные частоты были одинаковы. Вместе с тем первый резонанс в несколько большей степени зависит от объема заднего резонатора и параметров горла переднего, а второй — от объема переднего и параметров горла заднего. Влияние на F_1 и F_2 объемов переднего и заднего резонаторов распределяется между ними все более равномерно по мере сближения F_1 и F_2 , вызванного изменением размеров резонаторов. Сравнение рассчитанных по формуле (III.18) коэффициентов, характе-

ризирующих зависимость формантных частот от параметров резонаторов, с измеренными при помощи электрического аналога речевого тракта дано в § 3 гл. 7.

Зависимость формант от каждого из резонаторов, входящих в состав сдвоенного, выявляется и в величинах затухания, вносимого в каждый из резонансов активными элементами отдельных частей системы; количественной оценкой этого затухания является ширина резонансов.

Влияние отдельных активных элементов сдвоенного резонатора на ширину каждого из двух резонансов может быть определено из формулы (III.10), если не учитывать влияния ветви L_3R_3 в явной форме. Принимая обозначения, введенные в формулах (III.18), можно получить следующие выражения для ширины резонансов:

$$\left. \begin{aligned} B_1 &= \frac{G_1}{2\pi C_1} A + \frac{G_2}{2\pi C_2} B + \frac{R_1}{2\pi L_1} C + \frac{R_2}{2\pi L_2} D, \\ B_2 &= \frac{G_1}{2\pi C_1} B + \frac{G_2}{2\pi C_2} A + \frac{R_1}{2\pi L_1} D + \frac{R_2}{2\pi L_2} C. \end{aligned} \right\} \quad (\text{III. 22})$$

Из этих выражений следует, что демпфирование каждого из резонансов пропорционально отношениям R/L и G/C соответствующих ветвей схемы, с теми же коэффициентами пропорциональности A , B , C и D , которые характеризуют относительное влияние этих ветвей на частоту каждого из резонансов. Таким образом, оценка влияния на резонансы отдельных частей системы на основе критерия затухания в точности совпадает с оценкой их по частотному критерию; этого и следовало ожидать, исходя из энергетических соображений.

Если резонансы далеки один от другого, скажем $F_2 \gg F_1$, и отношение $\frac{F_2}{F_{01}} = \frac{F_{02}}{F_1}$ близко к 1, коэффициенты A и C очень малы, а B и D близки к 1; это означает, что F_1 определяется в основном задним, а F_2 — передним, меньшим по размерам резонатором.

Из эквивалентной схемы сдвоенного резонатора Гельмгольца можно также найти отношение звуковых давлений в передней и задней полостях; оно равно

$$\frac{P_1}{P_2} \underset{(f=F_1)}{=} 1 - \frac{F_1^2}{F_{02}^2} \quad (\text{III. 23})$$

при частоте первого резонанса и

$$\frac{P_1}{P_2} \underset{(f=F_2)}{=} 1 - \frac{F_2^2}{F_{02}^2} \quad (\text{III. 24})$$

при частоте второго, более высокого, резонанса.

Так как F_2 всегда выше, а F_1 ниже, чем F_{02} , то из этих выражений видно, что звуковые давления в двух полостях имеют одинаковый знак при частоте F_1 и противоположный при частоте F_2 .

Если частота несвязанного резонанса F_2 заднего резонатора близка к F_1 , то можно видеть, что при частоте F_1 давление в задней полости больше, чем давление в передней, а при частоте F_2 имеет место обратное.

Результаты детальных расчетов для моделей с распределенными постоянными, лучше отображающими действительный речевой тракт, будут приведены в § 4.

§ 3. Труба постоянного сечения как акустический резонатор

Модель речевого тракта в виде одиночного или двоянного резонатора Гельмгольца имеет ограниченное значение не только в силу относительно больших размеров речевого аппарата человека, но и потому, что действительная его конфигурация часто существенно отличается от подобной теоретической модели. Так, площадь ротового отверстия иногда может быть равна или даже больше площади поперечного сечения передней части ротовой полости.

Связанное с этим отсутствие явно выраженного горла у переднего резонатора является, однако, еще не самым решающим ограничением применимости указанной модели; существеннее то, что у заднего резонатора горло может либо вовсе отсутствовать, либо, как в случае ряда передних гласных, так органически входить в состав передней полости, что разделение их становится весьма условным и теряет смысл. Наконец, могут встретиться случаи появления более двух сужений по длине тракта, как, например, для гласных и согласных, характеризующихся комбинированной ретрофлексной и заднеязычной артикуляцией.

Для того чтобы сохранить в максимальной степени информацию, которая содержится в картине артикуляции данного звука, приходится прибегать к приближенному представлению истинной конфигурации речевого тракта при помощи достаточно большого числа цилиндрических отрезков.

В то же время сведение числа таких отрезков к разумному минимуму полезно для выявления наиболее существенных особенностей речеобразования, так как в этом случае возможно получить законченное математическое выражение для функции передачи. Такое представление физически более оправдано, чем представление речевого тракта двоянным резонатором, хотя бы потому, что оно не ограничивает формантную картину только двумя формантами. Правда, при рассмотрении вопроса о зависимости первой или второй форманты от параметров полостей, оба подхода можно использовать с одинаковым успехом, поскольку одиночный резонатор Гельмгольца можно считать частным случаем системы цилиндрических отрезков труб. Для этого частного случая ниже будет приведено общее решение при условии малых потерь; подобная модель удобна для изучения вопросов, связанных с демпфированием формант. Предварительно, однако, полезно рассмотреть одну цилиндрическую трубу при произвольных нагрузках на ее концах.

Труба, имеющая круговую или какую-либо иную, но постоянную по длине трубы форму поперечного сечения, ведет себя как электрическая линия вплоть до частоты $f = 20\,000/d$, где d — наибольший из поперечных размеров трубы в см. При более высоких частотах могут возникать явления поперечных стоячих волн, что дает эффект, подобный шунтированию полости некоторой параллельно включенной системой. Однако при анализе речи эти явления можно, как правило, не учитывать.

Коэффициент передачи по току, связывающий ток на коротко замкнутом конце линии, соответствующем открытому концу трубы, с током, создаваемым источником с большим внутренним сопротивлением

на другом конце, равен

$$\frac{U_0}{U_q} = H(s) = \frac{1}{\operatorname{ch} \Gamma(s)} = \frac{1}{\operatorname{ch} [al + sl/c]} = \frac{K}{\prod_{n=1}^{\infty} (1 - s/s_n)(1 - s/s_n^*)}, \quad (\text{III. 25})$$

где, как и раньше, l — длина линии, c — скорость звука и α — коэффициент затухания, определяющийся соотношением

$$\alpha = \alpha_R + \alpha_G = \frac{1}{2c} \left(\frac{R}{L} + \frac{G}{C} \right), \quad (\text{III. 26})$$

где R , L , G , C — сопротивление, индуктивность, проводимость и емкость на единицу длины линии.

Полюсы определяются комплексными частотами $s = \sigma + j\omega$, при которых $H(s)$ обращается в бесконечность. Это означает, что

$$\operatorname{ch} \Gamma(s) = 0, \quad \text{или} \quad \Gamma(s) = \pm j \frac{\pi}{2} (2n - 1), \quad (\text{III. 27})$$

$$\omega_n = \pm \frac{\pi}{2} (2n - 1) \frac{c}{l}, \quad \sigma_n = -(\alpha_R + \alpha_G) c. \quad (\text{III. 28})$$

Отсюда находим частоту и ширину полосы резонансов:

$$F_n = (2n - 1) \frac{c}{4l}, \quad B_n = \frac{c(\alpha_R + \alpha_G)}{\pi}. \quad (\text{III. 29})$$

Если потери малы, то

$$K = \frac{1}{\operatorname{ch} al} \approx 1. \quad (\text{III. 30})$$

Назовем звук, создаваемый идеализированным речевым трактом, имеющим постоянное сечение и длину 17,6 см, или, точнее, $c/2000$ см (где c — скорость звука), идеальным нейтральным гласным [а]. Частоты формант этого звука определяются из соотношения

$$F_n = (2n - 1) \cdot 500 \text{ гц}, \quad (\text{III. 31})$$

откуда $F_1 = 500$ гц, $F_2 = 1500$ гц, $F_3 = 2500$ гц и т. д.

Представление коэффициента передачи, в соответствии с формулой (III. 25), в виде бесконечного произведения можно приближенно заменить произведением g множителей, где $g = 2, 3$ или 4, соответствующих первым g формантам, и остаточным коэффициентом $k_{rg}(s)$, как это показано в § 3 гл. 2 (формула (2.43))

$$\frac{U_0}{U_q} = \frac{k_{rg}(s)}{\prod_{n=1}^g (1 - s/\hat{s}_n)(1 - s/\hat{s}_n^*)}. \quad (\text{III. 32})$$

Значение множителя k_{rg} , соответствующее трубе постоянного сечения, может быть принято и для любой конфигурации резонатора, так как влияние более высокочастотных полюсов на нижележащие определяется в основном средней их плотностью на оси $j\omega$; в силу этого общая длина является наиболее важным параметром.

Замена бесконечного числа сомножителей конечным числом дает хорошие результаты практически до частоты, соответствующей $(g-1)$ -му полюсу.

Условие резонанса для трубы постоянного сечения, закрытой или почти закрытой на обоих концах, определяется равенством

$$Z \operatorname{cth} \Gamma(s) = \infty. \quad (\text{III. 33})$$

В случае трубы, открытой на обоих концах или заканчивающейся элементами с относительно большой площадью поперечного сечения, резонанс определяется равенством

$$Z \operatorname{th} \Gamma(s) = 0, \quad (\text{III. 34})$$

которое идентично равенству (III.33). Решение этих уравнений при отсутствии потерь будет:

$$\omega_n = (n-1) \frac{\pi c}{l_e}, \quad (\text{III. 35})$$

или

$$F_n = (n-1) \frac{c}{2l_e}, \quad \text{где } l_e = l + l_t. \quad (\text{III. 36})$$

Полуволновые резонансы в такой трубе зависят от ее эффективной длины l_e , т. е. от ее действительной длины l с добавлением дополнительной длины l_t , представляющей собой концевую поправку. То же относится и к четвертьволновому резонатору, каковым является труба, открытая на одном конце и закрытая на другом (см. формулу (III. 33)).

В самом общем случае следует вводить две концевые поправки, учитывающие условия на обоих концах трубы. Если открытый конец снабжен сферическим экраном радиусом 9 см, концевая поправка, согласно формуле (2.17), равна $l_t = 0,8 \left(\frac{A}{\pi} \right)^{1/2}$. При отсутствии экрана, по Морзу [155], $l_t = 0,6 \left(\frac{A}{\pi} \right)^{1/2}$. Если открытый конец при отсутствии плавного перехода нагружен на достаточно большую полость, следует пользоваться формулой (2.18).

Неполное закрытие конца трубы соответствует нагрузке конца линии большим импедансом. Акустически такая нагрузка может быть представлена короткой открытой трубкой длиной l_g и площадью поперечного сечения A_g , причем $A_g \ll A$. Концевая поправка отрицательна и равна

$$l_t = - \frac{c^2 A_g}{\omega^2 A l_g}. \quad (\text{III. 37})$$

Эта поправка исчезает на высоких частотах; область ее применения ограничена частотами, для которых соблюдается условие $\omega L_g \gg Z$. Если необходимо учитывать, кроме массы, также потери в нагрузочной трубке, т. е. сопротивление R_g , включенное последовательно с L_g , то при условии $\omega L_g > Z$ или $R_g > Z$

$$l_t = - \frac{c Z L_g}{R_g^2 + \omega^2 L_g^2}, \quad (\text{III. 38})$$

где $L_g = \frac{\rho l_g}{A_g}$. Длина l_g трубки, при которой ее можно считать короткой, определяется условием $\frac{\omega l_g}{c} < 1$ и условием $\frac{\operatorname{th} \omega l_g}{c} \approx \frac{\omega l_g}{c}$.

Если $\omega L_g \ll Z$, то концевая поправка положительна, так же как и в случае сопротивления излучения, и равна

$$l_t = \frac{l_g A}{A_g}. \quad (\text{III. 39})$$

Если нагрузочная трубка закрыта на удаленном конце, то она представляет собою емкостную нагрузку, равную

$$C_g = \frac{l_g A_g}{\rho c^2}, \quad (\text{III. 40})$$

при условии $\omega l_g/c < 1$. Концевая поправка в этом случае положительна и равна

$$l_t = \frac{l_g A_g}{A}. \quad (\text{III. 41})$$

Наконец, если при закрытом конце нагрузочной трубки ее емкостное сопротивление мало, т. е. если ее можно рассматривать как замкнутую полость длиной l_g при площади поперечного сечения $A_g \gg A$, то концевая поправка отрицательна и равна

$$l_t = \frac{-c}{Z\omega^2 C_g} = \frac{-c^2 A}{\omega^2 A_g l}. \quad (\text{III. 42})$$

С концевой поправкой всегда связано наличие активных элементов, например, сопротивления излучения или вязкого сопротивления (см. формулу (2.14)). Влияние таких элементов можно учесть, увеличив коэффициент затухания основной трубы на величину

$$\alpha_t = \frac{R_T}{Zl_e}, \quad (\text{III. 43})$$

если R_T представляет собой последовательно включенное малое сопротивление, или на величину

$$\alpha_t = \frac{ZG_T}{l_e}, \quad (\text{III. 44})$$

если $1/G_T$ представляет собой большое параллельно включенное сопротивление. Поэтому в случае нагрузки с большим импедансом следует представить ее в виде параллельно включенного элемента.

Описанные преобразования не изменяют фазового угла характеристического импеданса, т. е. α_t делится поровну между коэффициентом α_G , обусловленным параллельным сопротивлением потерь, и коэффициентом α_R , обусловленным последовательным сопротивлением потерь. Однако приведение реактивной составляющей нагрузки к основной трубе путем изменения постоянной распространения вызывает некоторое, хотя и незначительное, перераспределение потерь. Это можно видеть из нижеследующего. Характеристики трубы как передающей системы при условии малых потерь определяются выражениями (см. формулу (2.3))

$$\left. \begin{aligned} Z &= Z_0 \left[1 + j \frac{c}{\omega} (\alpha_G - \alpha_R) \right]; \\ \Gamma &= l \left[\alpha_R + \alpha_G + j \frac{\omega}{c} \right]; \\ Z_0 &= \frac{\rho c}{A}. \end{aligned} \right\} \quad (\text{III. 45})$$

Входной импеданс Z_i , т. е. импеданс на конце трубы, противоположном нагруженному импедансом Z_T , можно представить в хорошо известной форме

$$\left. \begin{aligned} Z_i &= Z \operatorname{th} (\Gamma + \Gamma_t) = Z \operatorname{th} \Gamma_e, \\ \Gamma_t &= \operatorname{arth} \frac{Z_T}{Z}. \end{aligned} \right\} \quad (\text{III. 46})$$

Если $\frac{Z_T}{Z} < 1$, то удобнее другое выражение:

$$\left. \begin{aligned} Z_i &= Z \operatorname{cth} (\Gamma + \Gamma_t), \\ \Gamma_t &= \operatorname{arth} \frac{Z}{Z_T}. \end{aligned} \right\} \quad (\text{III. 47})$$

Общее выражение для добавочного члена постоянной распространения имеет вид

$$\Gamma_t = l_t \left(\mp \alpha_R \pm \alpha_G + \frac{j\omega}{c} \right) + (l + l_t) \alpha_t. \quad (\text{III. 48})$$

Формулы для различных видов нагрузки приведены в следующей таблице:

Нагрузка	Выражение для импеданса нагрузки	Концевая поправка l_t	Изменение собственного затухания при распространении	Добавочное затухание при распространении α_t
А. Малая индуктивность	$Z_T = R_T + j \frac{\omega \rho l_g}{A_g}$	$\frac{l_g A}{A_g}$	$l_t (-\alpha_R + \alpha_G)$	$\frac{R_T}{Z(l + l_t)}$
В. Большая индуктивность	$\frac{1}{Z_T} = G_T + \frac{A_g}{j\omega \rho l_g}$	$\frac{-c^2 A_g}{\omega^2 A l_g}$	$l_t (\alpha_R - \alpha_G)$	$\frac{G_T Z}{(l + l_t)}$
С. Малая емкость	$\frac{1}{Z_T} = G_T + j \frac{\omega l_g A_g}{\rho c^2}$	$\frac{l_g A_g}{A}$	$l_t (-\alpha_R + \alpha_G)$	$\frac{G_T Z}{l + l_t}$
Д. Большая емкость	$Z_T = R_T + \frac{\rho c^2}{j\omega l_g A_g}$	$\frac{-c^2 A}{\omega^2 A_g l_g}$	$l_t (\alpha_R - \alpha_G)$	$\frac{R_T}{Z(l + l_t)}$

В зависимости от того, является ли условие на конце трубы, противоположном Z_T , условием холостого хода или короткого замыкания, частота резонансов определяется соотношениями

$$\Gamma_e(s) = j \frac{\pi}{2} (2n - 1) \quad (\text{III. 49})$$

или

$$\Gamma_e(s) = j\pi (n - 1). \quad (\text{III. 50})$$

Выражения показателя затухания одинаковы как для четвертьволнового резонанса (формула (III. 49)), так и для полуволнового

(формула (III.50)). Для различных видов нагрузки эти выражения следующие:

Условия нагрузки	Показатель затухания σ	
A	$-c \left[\alpha_R \left(1 - \frac{2l_t}{l_e} \right) + \alpha_G + \alpha_t \right],$	}
B	$-c \left[\alpha_R \left(1 - \frac{2l_t}{l_e} \right) + \alpha_G + \alpha_t \left(1 - \frac{2l_t}{l_e} \right) \right],$	
C	$-c \left[\alpha_R + \alpha_G \left(1 - \frac{2l_t}{l_e} \right) + \alpha_t \right],$	
D	$-c \left[\alpha_R + \alpha_G \left(1 - \frac{2l_t}{l_e} \right) + \alpha_t \left(1 - \frac{2l_t}{l_e} \right) \right].$	

(III. 51)

Они могут быть получены в результате соответствующих приближений по общим формулам, относящимся к системе, состоящей из двух отрезков труб (см. ниже формулу (III.83)). Из приведенных выражений можно видеть, что если l_t мало, показатель затухания равен приближенно

$$\sigma = -c(\alpha_R + \alpha_G + \alpha_t). \quad (\text{III. 52})$$

Отрицательный знак концевой поправки l_t для трубы, нагруженной на полость большого объема, показывает, что концевая поправка l_t вызывает повышение частот резонансов. Однако нагрузочную полость можно считать емкостью только приближенно; для получения большей точности следует учитывать последовательную индуктивность, равную одной трети массы воздуха в полости, а также дополнительную индуктивность, определяемую по формуле (2.60) и обусловленную излучением. Из всего этого следует, что при нагрузке типа D концевую поправку было бы правильней учитывать в виде

$$(l_t)_D = l_g \frac{A}{A_g} \left[\frac{1}{3} - \frac{c^2}{\omega^2 l_g^2} + \frac{0,48 A_g}{l_g A^{1/2}} \left(1 - 1,25 \left(\frac{A_g}{A} \right)^{1/2} \right) \right]. \quad (\text{III. 53})$$

Элементы с очень малым или с очень большим активным или реактивным сопротивлением, расположенные на том или другом конце трубы, влияют на показатель затухания σ почти независимо друг от друга, и потому результаты этого влияния можно суммировать.

Ясно, что учет импеданса нагрузки путем соответствующего изменения постоянной передачи основной трубы ограничивает возможности использования эквивалентной схемы. Хотя полюсы при таком учете получаются правильно, нули могут определяться неверно, например нули импеданса того конца трубы, на котором включено Z_T . При расчетах передачи по речевому тракту пересчет импеданса к постоянным распространения трубы допустим в тех случаях, когда рассчитываются коэффициенты передачи по току или переходные проводимости от входа к выходу, т. е. отношения объемной скорости на выходе к объемной скорости (для источника с заданной объемной скоростью) или к давлению (для источника с заданным давлением) на входе системы.

Возможность включения импеданса излучения, а в случае большого раскрытия губ и импеданса самого выходного отверстия, в состав характеристических параметров трубы, изображающей полость рта, особенно полезна в тех случаях, когда речевой тракт рассматривается как

сочетание отрезков труб. Аналогично сопротивление голосовой щели, а при более высоких частотах и ее индуктивность, могут быть учтены путем изменения параметров заднего отрезка. В области частот ниже третьей форманты участок тракта, соответствующий гортани, может быть учтен путем увеличения длины фаринкса на $l_g A_g / A$, где A — площадь поперечного сечения фаринкса, а $l_g A_g$ — объем гортани.

Коэффициент передачи трубы, связывающий ток U_T на ее конце, нагруженном малым импедансом Z_T , с входным током U на другом конце трубы, равен

$$\frac{U_T}{U} = \frac{1}{\operatorname{ch} \Gamma + \frac{Z_T}{Z} \operatorname{sh} \Gamma}, \quad (\text{III. 54})$$

или

$$\frac{U_T}{U} = \frac{\operatorname{ch} \Gamma_t}{\operatorname{ch} (\Gamma + \Gamma_t)}, \quad \text{где } \Gamma_t = \operatorname{arth} \frac{Z_T}{Z}. \quad (\text{III. 55})$$

Числитель $\operatorname{ch} \Gamma_t$ приблизительно равен 1, если $Z_T \ll Z$. Аналогичные соотношения справедливы при расчете напряжения на конце трубы, нагруженном большим импедансом.

Приведенные выше преобразования не ограничены только случаями предельных значений импеданса нагрузки Z_T . Примером подобного преобразования, при котором нагрузка не является предельно малой, является внесение импеданса излучения $R_0 + j\omega L_0$ в параметры переднего отрезка трубы при частотах $f > 3000 \left(\frac{10}{A_0} \right)^{1/2}$. В этом случае нельзя считать R_0 малым по сравнению с Z_0 ; тогда собственными потерями в переднем отрезке трубы можно пренебречь. Если, кроме того, $\alpha_G = \alpha_R$, то точное выражение для приращения постоянной распространения, обусловленного излучением, будет иметь вид

$$\begin{aligned} \Gamma_e &= \frac{j}{2} \operatorname{arctg} \frac{2\omega L_0 / Z}{1 + \left(\frac{R_0}{Z} \right)^2 + \left(\frac{\omega L_0}{Z} \right)^2} + \frac{1}{2} \operatorname{arth} \frac{2R_0 / Z}{1 + \left(\frac{R_0}{Z} \right)^2 + \left(\frac{\omega L_0}{Z} \right)^2} = \\ &= \frac{j\omega l_t}{c} + l\alpha_t. \end{aligned} \quad (\text{III. 56})$$

Отношение R_0/Z здесь может быть как меньше, так и больше 1. Показатель затухания для полюсов после этого находится приближенно как

$$\sigma_n = -c\alpha_t. \quad (\text{III. 57})$$

Одновременное возрастание $R_0(\omega)$ и ωL_0 при увеличении ω исключает возможность точного согласования по импедансу, и потому показатели затухания при достаточно большой длине трубы остаются все же малыми по сравнению с частотами полюсов вплоть до довольно высоких частот. Это можно видеть на рис. 5, где даже высокие резонансы не демпфируются полностью.

Если бы имело место условие $L_0 = 0$, $R_0 = Z$, т. е. нагрузка была бы точно согласована с характеристическим импедансом линии, α_t обратилось бы в бесконечность, а показатель затухания σ_n был бы равен $-\infty$; на частоты же полюсов ω_n чисто активная нагрузка не оказывает влияния.

§ 4. Системы из четырех отрезков труб. Уравнения преобразования для произвольного расположения источника

Свойства сложных систем резонаторов, состоящих из отрезков труб неравного сечения, определяются приведенными ниже трансцендентными соотношениями. В случае четырех таких отрезков коэффициент передачи, связывающий объемную скорость на выходе, т. е. через ротовое отверстие, и объемную скорость в голосовой щели, может быть представлен в виде

$$\frac{U_0}{U_q} = H_p(s) = \frac{1}{\operatorname{ch} \Gamma_1 \operatorname{ch} \Gamma_2 \operatorname{ch} \Gamma_3 \operatorname{ch} \Gamma_4 (AB + CD)}, \quad (\text{III. 58})$$

где

$$A = \left(1 + \frac{A_2}{A_1} \operatorname{th} \Gamma_2 \operatorname{th} \Gamma_1\right), \quad B = \left(1 + \frac{A_4}{A_3} \operatorname{th} \Gamma_4 \operatorname{th} \Gamma_3\right), \\ C = \frac{A_3}{A_2} \left(\operatorname{th} \Gamma_2 + \frac{A_2}{A_1} \operatorname{th} \Gamma_1\right), \quad D = \left(\frac{A_4}{A_3} \operatorname{th} \Gamma_4 + \operatorname{th} \Gamma_3\right).$$

При этом предполагается, что $\frac{Z_m}{Z_{m-1}} = \frac{A_{m-1}}{A_m}$, т. е. фазовые углы характеристических импедансов всех отрезков считаются равными.

Полагая, что потери малы, мы получаем выражение

$$H_p(s) = \frac{1}{\prod_{n=1}^{\infty} (1 - s/\hat{s}_n)(1 - s/\hat{s}_n^*)}. \quad (\text{III. 59})$$

Это выражение означает, что функция передачи полностью определяется сопряженными полюсами и становится равной 1 при нулевой частоте. Функция передачи, связывающая ток на выходе системы с напряжением источника $E_s(s)$, включенного в произвольном месте в одно из продольных плеч эквивалентной схемы, имеет вид

$$\frac{U_0(s)}{E_s(s)} = H(s) = H_p(s) H_z(s), \quad (\text{III. 60})$$

где $H_p(s)$ — функция передачи, аналогичная определяемой формулой (III. 59), а $H_z(s)$ — функция нулей, учитывающая и постоянно присутствующие факторы.

Если источник включен последовательно с импедансом излучения, то

$$H_z(s) = \frac{\operatorname{sh} \Gamma_1 \operatorname{sh} \Gamma_2 \operatorname{sh} \Gamma_3 \operatorname{sh} \Gamma_4 (ef + gh)}{Z_4}, \quad (\text{III. 61})$$

где

$$e = \frac{A_1}{A_3} \left(1 + \frac{A_3}{A_4} \operatorname{cth} \Gamma_4 \operatorname{cth} \Gamma_3\right), \quad f = \operatorname{cth} \Gamma_2 + \frac{A_2}{A_1} \operatorname{cth} \Gamma_1, \\ g = \frac{A_1}{A_2} \left(1 + \frac{A_2}{A_1} \operatorname{cth} \Gamma_1 \operatorname{cth} \Gamma_2\right), \quad h = \frac{A_3}{A_4} \operatorname{cth} \Gamma_4 + \operatorname{cth} \Gamma_3.$$

При очень низких частотах, пренебрегая потерями, находим:

$$\lim_{\omega \rightarrow 0} H_z(\omega) = j\omega \sum_{v=1}^4 \frac{l_v A_v}{\rho c^2}, \quad (\text{III. 62})$$

что соответствует формуле (2.49).

В общем случае

$$H_z(s) = sC_b \prod_{n=1}^{\infty} (1 - s/\bar{s}_n)(1 - s/\bar{s}_n^*), \quad (\text{III. 63})$$

где C_6 — емкость, соответствующая полному объему полостей, расположенных позади источника.

Если источник находится на расстоянии l_s впереди от переднего конца отрезка 2, единственное изменение, которое надо ввести в выражение функции, описываемой формулой (III. 61), сводится к замене Γ_1 на $(l_s/l_1)\Gamma_1$. Подобного рода подстановки можно использовать при любом расположении источника. При этом элементы, расположенные перед источником, не входят в выражение для функции нулей. Если, например, источник находится в середине отрезка 2, следует положить $\Gamma_1=0$ и заменить Γ_2 на $\Gamma_2/2$. Если источник расположен на границе между отрезками 1 и 2, получим следующее выражение:

$$H_z(s) = \frac{\text{sh } \Gamma_2 \text{ sh } \Gamma_3 \text{ sh } \Gamma_4 \left(e^{\frac{A_2}{A_1}} + h \text{cth } \Gamma_2 \right)}{Z_4}. \quad (\text{III. 64})$$

Остановимся еще на резонаторе, состоящем из двух отрезков труб; такой резонатор может являться в ряде случаев полезной моделью речевого тракта. Функцию полюсов для этого резонатора легко получить, если положить $\Gamma_3=\Gamma_4=0$:

$$H_p(s) = \frac{1}{\text{ch } \Gamma_1 \text{ ch } \Gamma_2 \left[1 + \frac{A_2}{A_1} \text{th } \Gamma_1 \text{th } \Gamma_2 \right]}. \quad (\text{III. 65})$$

Если на основе такой модели рассматривать согласные, то надо учитывать, что источник может быть расположен в любом ее месте. Если он находится у излучающего конца, то

$$H_z(s) = \frac{\text{sh } \Gamma_1 \text{ sh } \Gamma_2 \left(\text{cth } \Gamma_1 + \frac{A_1}{A_2} \text{cth } \Gamma_2 \right)}{Z_2}; \quad (\text{III. 66})$$

если же он расположен на границе между отрезками 1 и 2, то

$$H_z(s) = \frac{\text{sh } \Gamma_2}{Z_2}. \quad (\text{III. 67})$$

Рассмотрим более подробно простейший случай, когда $A_2=A_1$ ¹⁾. Пренебрегая концевой поправкой, учитывающей излучение, найдем:

$$H_p(s) = \frac{1}{\text{ch } (\Gamma_1 + \Gamma_2)}, \quad (\text{III. 68})$$

$$H(s) = H_p(s) H_z(s) = \frac{\text{sh } \Gamma_2}{Z_2 \text{ch } (\Gamma_1 + \Gamma_2)}. \quad (\text{III. 69})$$

Учитывая нагрузку, представленную импедансом излучения, получим:

$$H_p(s) = \frac{1}{\text{ch } (\Gamma_1 + \Gamma_2) + \left(\frac{R_0}{Z_1} + s \frac{\Gamma_0}{Z_1} \right) \text{sh } (\Gamma_1 + \Gamma_2)}, \quad (\text{III. 70})$$

или

$$H_p(s) = \frac{\text{ch } \Gamma_0}{\text{ch } (\Gamma_1 + \Gamma_2 + \Gamma_0)}, \quad (\text{III. 71})$$

где

$$\Gamma_0 = \text{arth} \left(\frac{R_0(s)}{Z_1} + s \frac{L_0}{Z_1} \right),$$

¹⁾ Модель, состоящая из одного отрезка трубы, была использована Ингардом [109] при рассмотрении образования взрывных. Метод анализа при помощи эквивалентных схем, примененный здесь, приводит к тем же результатам, но в более компактной форме. Он был использован независимо Гейнцем [97] при изучении образования фриктивных на модели, состоящей из двух отрезков труб.

как это было показано при анализе свойств одиночного отрезка трубы. В большинстве интересующих нас приложений допустимо низкочастотное приближение, при котором Γ_0 мало, так что $\text{sh } \Gamma_0 \approx 1$.

Предположим, что в момент $t=0$ замыкается ключ, расположенный между передним и задним отрезками трубы; начальное избыточное давление в заднем отрезке положим равным Δp . Свойства источника в этом случае описываются ступенчатой функцией, выражение которой через комплексную частоту имеет вид

$$E(s) = \frac{\Delta p}{s}, \quad (\text{III. 72})$$

т. е. характеризуется спектром, огибающая которого имеет наклон -6 дб/окт. Спектр мощности излучаемого звука будет

$$W(\omega) = R_0(\omega) U_0^2(\omega) = E^2(\omega) R_0(\omega) H^2(\omega) = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\text{sh}^2 \Gamma_2 \text{ch}^2 \Gamma_0}{\text{ch}^2(\Gamma_1 + \Gamma_2 + \Gamma_0)} \quad (\text{III. 73})$$

и модуль его

$$|W(\omega)| = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\sin^2 \varphi_2}{\left[\left(1 - \frac{\omega L_0}{Z_1} \text{tg } \varphi\right)^2 + \frac{R_0^2(\omega)}{Z_1^2} \text{tg}^2 \varphi\right] \cos^2 \varphi}, \quad (\text{III. 74})$$

где

$$\varphi_1 = \frac{\omega l_1}{c}, \quad \varphi_2 = \frac{\omega l_2}{c} \quad \text{и} \quad \varphi = \frac{\omega(l_1 + l_2)}{c}.$$

Сравнительно простая форма, которую имеет формула (III.69), сохраняется и для выражения (III.74) в случае низких частот. В этом случае приближенное выражение имеет вид

$$|W(\omega)| = \left(\frac{\Delta p}{\omega}\right)^2 \left(\frac{A_2}{\rho c}\right)^2 R_0(\omega) \frac{\sin^2 \varphi_2}{\cos^2 \varphi_e + R_1^2(\omega)/Z_1^2}, \quad (\text{III. 75})$$

где $l_e = (l_1 + l_2 + l_i)$ — эффективная длина трубы и $\varphi_e = \frac{\omega l_e}{c}$.

Ширина полосы резонансов равна

$$B_n = \frac{-\sigma_n}{\pi} = \frac{c R_0(\omega)}{\pi l_e Z_1}, \quad (\text{III. 76})$$

а частоты их

$$F_n = \frac{\omega_n}{2\pi} = (2n - 1) \frac{c}{4l_e}. \quad (\text{III. 77})$$

Нули, т. е. антирезонансы, имеют место при частотах

$$F_m = \frac{\omega_m}{2\pi} = \frac{mc}{2l_2}. \quad (\text{III. 78})$$

В практических приложениях целесообразнее рассматривать не общую излучаемую энергию, а звуковое давление в некоторой точке акустического поля, создаваемого говорящим. Из формул (2.55), (III. 60) и (III. 72) следует, что для любой сложной системы резонаторов частотная зависимость звукового давления на выходе системы при воздействии на вход источника давления в виде ступенчатой функции, определяется равенством

$$P(\omega, l_d) = \frac{\Delta p}{4\pi l_d} K_T(\omega) H_z(\omega) H_p(\omega). \quad (\text{III. 79})$$

Здесь $H_z(s)$ и $H_p(s)$ — функции нулей и полюсов, имеющие в случае системы из четырех отрезков труб форму, определяемую формулами (III. 58) и (III. 61). $K_T(\omega)$ представляет собой частотную поправку, учитывающую направленность излучения и возрастание $R_0(\omega)$ быстрее, чем пропорционально ω^2 .

Результаты расчетов спектров взрывных, выполненных ранее [48] при использовании модели речевого тракта из двух отрезков труб (формулы (III. 65), (III. 66), (III. 79)), в предположении малости внутренних потерь, приведены в гл. 10 (см. рис. 74). Коэффициенты затухания при этом определялись по эмпирической формуле

$$\alpha = 0,007 \left(\frac{\pi}{A} \right)^{\frac{1}{2}} \text{ непер/см}, \quad (\text{III. 80})$$

полагая $\alpha_G = \alpha_R = \frac{\alpha}{2}$. Такое идеальное распределение потерь в параллельных и последовательных элементах, конечно, в известной мере гипотетично. Для идеальной трубы с жесткими стенками $\alpha_G \approx 0,45 \alpha_R$, как было показано в § 1 гл. 2. Следует отметить, что коэффициенты затухания для полостей речевого тракта в 2—8 раз больше.

§ 5. Демпфирующее влияние потерь в последовательных и параллельных элементах резонатора, состоящего из двух отрезков труб

Приведенный ниже анализ соотношений между параметрами системы, состоящей из двух отрезков труб, и значениями показателя затухания и частотами полюсов предполагает малость потерь. Однако он является достаточно общим и потому может служить в качестве математической основы для исследования влияния потерь в параллельных и последовательных элементах модели речевого тракта в будущем.

Частоты полюсов можно определить из приближенного уравнения

$$\frac{A_2}{A_1} \operatorname{tg} \frac{\omega l_1}{c} = \operatorname{ctg} \frac{\omega l_2}{c}. \quad (\text{III. 81})$$

Показатели затухания находятся обычным способом, путем введения комплексной частоты $s = j\omega + \sigma$. В результате характеристическое уравнение

$$\frac{Z_1(j\omega, \sigma)}{Z_2(j\omega, \sigma)} \operatorname{th} \Gamma_1(j\omega, \sigma) + \operatorname{cth} \Gamma_2(j\omega, \sigma) = 0 \quad (\text{III. 82})$$

разделяется на вещественную и мнимую части; каждая из этих частей порознь приравнивается нулю, что дает необходимые два уравнения. Решение может быть представлено в виде

$$\sigma = -c [k_{R1}\alpha_{R1} + k_{G1}\alpha_{G1} + k_{R2}\alpha_{R2} + k_{G2}\alpha_{G2}], \quad (\text{III. 83})$$

причем суммы коэффициентов в этом выражении, относящиеся к членам, характеризующим последовательные потери, и к членам, характеризующим параллельные потери, равны единице, т. е.

$$k_{R1} + k_{R2} = k_{G1} + k_{G2} = 1. \quad (\text{III. 84})$$

Обозначим для краткости $\frac{\omega l_1}{c} = \varphi_1$ и $\frac{\omega l_2}{c} = \varphi_2$. Тогда эти коэффициенты

могут быть представлены в виде

$$\left. \begin{matrix} k_{R1} \\ k_{G1} \end{matrix} \right\} = \frac{1 + \operatorname{tg}^2 \varphi_1 \pm \operatorname{tg} \varphi_1 / \varphi_1}{1 + \frac{A_1 l_2}{A_2 l_1} + \left(1 + \frac{l_2 A_2}{l_1 A_1}\right) \operatorname{tg}^2 \varphi_1}, \quad (\text{III. 85a})$$

$$\left. \begin{matrix} k_{R2} \\ k_{G2} \end{matrix} \right\} = \frac{1 + \operatorname{ctg}^2 \varphi_2 \pm \operatorname{ctg} \varphi_2 / \varphi_2}{1 + \frac{A_2 l_1}{A_1 l_2} + \left(1 + \frac{l_1 A_1}{l_2 A_2}\right) \operatorname{ctg}^2 \varphi_2}. \quad (\text{III. 85б})$$

Знак минус у входящих в эти выражения членов с двойным знаком относится к k_{G1} и k_{R2} .

Соотношение (III. 84) указывает на то, что демпфирующее влияние потерь в обоих отрезках труб носит взаимный характер. Значения коэффициентов в выражении (III. 83) характеризуют степень зависимости резонансов от каждого из отрезков труб, в полной аналогии с выражением (III. 22). Для идеального резонатора Гельмгольца $k_{R1} = k_{G2} = 1$ и $k_{R2} = k_{G1} = 0$. В случае резонанса стоячей волны, определяемой в основном отрезком 2, коэффициенты приближаются к значениям $k_{R1} = k_{G1} = 0$ и $k_{R2} = k_{G2} = 1$.

Приведенные в таблице в § 3 формулы, относящиеся к одиночной трубе при большой или малой индуктивной нагрузке (случаи А и В), могут быть получены из формул (III. 85), если положить $l_1 \ll l_2$ и $\operatorname{tg} \varphi_1 = \varphi_1$.

В случае очень узкого горла резонатора имеет место условие $\frac{A_2}{A_1} \gg 1$; кроме того, положим $\frac{l_2 A_1}{l_1 A_2} < 1$, как для простого резонатора. Допустим далее, что длина отрезка 2 такова, что следует считаться с образованием в нем стоячей волны. Тогда частота основного резонанса в первом приближении будет

$$\omega \approx c \left(\frac{A_1}{A_2 l_1 l_2} \right)^{1/2}, \quad (\text{III. 86})$$

как и должно быть. Показатель затухания будет равен

$$\sigma \approx -c \left[\frac{2\alpha_{R1} + \alpha_{R2} l_2 A_1 / l_1 A_2}{2 + l_2 A_1 / l_1 A_2} + \alpha_{G2} \right]. \quad (\text{III. 87})$$

Из этого выражения видно, что последовательно включенные сопротивления основного отрезка трубы, определяющие коэффициент α_{R2} , становятся при указанных выше предположениях малыми.

Если $\frac{l_2 A_1}{l_1 A_2} \ll 1$, труба эквивалентна емкости и показатель затухания

$$\sigma = -c (\alpha_{R1} + \alpha_{G2}). \quad (\text{III. 88})$$

Второй и более высокие резонансы обуславливаются стоячими волнами. Вследствие большого импеданса горла уравнение (III. 81) сводится к

$$\frac{A_2}{A_1} \frac{\omega l_1}{c} = \frac{1}{\omega l_2 / c - (n-1)\pi}. \quad (\text{III. 89})$$

Отсюда

$$\omega_n = \pi (n-1) \frac{c}{l_{2e}}, \quad (\text{III. 90})$$

причем

$$l_{2e} = l_2 \left(1 - \frac{1}{\frac{A_2}{A_1} \frac{\omega^2 l_1 l_2}{c^2}} \right), \quad (\text{III. 91})$$

что согласуется с полученной ранее концевой поправкой для случая большой индуктивной нагрузки (см. формулу (III. 48)).

Для точного определения показателя затухания следует использовать формулу (III. 85). Если можно с достаточной точностью считать $l_{2e} = l_2$, то показатель затухания определится из упрощенной формулы

$$\sigma_n = -c \left[\alpha_{R1} \frac{2 + \left(\frac{l_1}{l_2} \right)^2 (n-1)^2 \pi^2}{(n-1)^2 \pi^2 l_1 A_2 / l_2 A_1} + \alpha_{R2} + \alpha_{G2} \right]. \quad (\text{III. 92})$$

Если $(n-1) \pi \frac{l_1}{l_2} < 1$, то коэффициент при α_{R1} обратно пропорционален квадрату частоты резонанса. Отсюда видно, что влияние последовательного сопротивления в горле значительно меньше, чем влияние такого же сопротивления, но включенного как в случае открытой трубы, т. е. являющегося либо непосредственно, либо последовательно с небольшой индуктивностью, нагрузкой отрезка 2 трубы. Применительно к речеобразованию из этих соотношений может быть сделан вывод, что при огублении влияние сопротивления излучения на вторую и более высокие форманты существенно уменьшается. Отношение показателей затухания, соответствующих случаям малой и большой индуктивной нагрузки, при постоянном α_{R1} можно получить из формулы (III. 85) или (III. 48); это отношение равно

$$\frac{\sigma_{ns}}{\sigma_{nl}} = (n-1)^2 \pi^2 \left(\frac{l_1}{l_2} \right)^2 \left(\frac{A_2}{A_1} \right)^2. \quad (\text{III. 93})$$

Длина l_1 здесь включает концевую поправку для горла резонатора. Когда эффективная длина l_1 не очень мала по сравнению с l_2 , множитель $\left[(n-1) \pi \frac{l_1}{l_2} \right]^2$ в приведенном выражении может становиться больше единицы. При этих условиях вклад α_{R1} в затухание имеет минимальное значение:

$$\sigma_n = -c \left[\alpha_{R1} \frac{l_1 A_1}{l_2 A_2} + \alpha_{R2} + \alpha_{G2} \right]. \quad (\text{III. 94})$$

Если физическая длина l_1 относительно велика, нагрузка минимальна, когда фазовый сдвиг удовлетворяет условию $\frac{\omega l_1}{c} = \frac{(2n-1)\pi}{2}$. Нагрузка становится максимальной, когда $\omega l_1 / c$ близко к $n\pi$, т. е. когда резонансная частота нагружающей трубы близка к резонансной частоте основной трубы. Эффект трансформации импеданса при этом является максимальным. Входной импеданс линии без потерь, замкнутой на удаленном конце на малое сопротивление R , выражается следующим образом:

$$Z_i = Z \operatorname{th} \left(\frac{j\omega l}{c} + \operatorname{arth} \frac{R}{Z} \right); \quad (\text{III. 95})$$

следовательно, при условии $\frac{\omega l}{c} = m\pi$

$$Z_i = R, \quad (\text{III. 96})$$

а при условии $\frac{\omega l}{c} = \frac{(2n-1)\pi}{2}$

$$Z_i = \frac{Z^2}{R}. \quad (\text{III. 97})$$

Вообще говоря, затухание, обусловленное влиянием одной трубы в сложной системе на соседнюю, может быть рассчитано путем определения импеданса нагрузочной трубы, измеренного со стороны главной полости. Если этот импеданс имеет достаточно большую величину, т. е. если площадь отрезка, рассматриваемого в качестве нагрузки, меньше площади основного, он может быть представлен в виде параллельного соединения реактивного и активного элементов. Эти элементы учитываются тогда концевой поправкой, которая сравнительно мало изменяется в области частот, соответствующей резонансному пику.

Аналогично этому, если отрезок трубы, замыкающий основную, имеет площадь больше, чем у основного, этот отрезок можно рассматривать как нагрузку с малым импедансом, состоящим из последовательно включенных активного и реактивного сопротивлений. Реактивное сопротивление учитывается путем введения концевой поправки, а активное — путем изменения коэффициента затухания так, как это было показано выше.

§ 6. Сводка расчетных формул для определения затухания в резонаторах из двух отрезков труб; приложение к модели речевого тракта

В настоящем разделе приводится сводка формул, используемых при расчете ширины резонансных пиков в идеальных резонаторах, состоящих из одного или двух отрезков труб, и дается оценка этих формул; заданными считаются размеры отрезков и параметры диссипативных элементов.

Сопротивление излучения, как показывает формула (2.16), содержит коэффициент $K_S(f)$ (см. рис. 5), учитывающий отклонение $R_0(f)$ от пропорциональности f^2 , в предположении, что излучатель помещается в сферическом экране. Классические потери в трубе с жесткими стенками разбиты на потери в последовательном сопротивлении, обусловленные вязкостью и характеризуемые величиной α_R , и потери в параллельном сопротивлении, обусловленные теплопроводностью и характеризуемые величиной α_G (см. формулу (2.13)).

Ширина полосы B_n резонанса номера n определяется величиной σ_n , представляющей собой вещественную часть соответственного полюса функции передачи, характеризующей речевой тракт, как это показано в § 2 Приложения I и в формуле (2.61):

$$B_n = -\frac{\sigma_n}{\pi}. \quad (\text{III. 98})$$

Когда демпфирование какого-либо резонанса обусловлено несколькими диссипативными элементами, влияния их суммируются линейно, если соседний резонанс не слишком близок к рассматриваемому и ширина резонансов не слишком велика. Эти условия, являющиеся по существу условиями малости потерь, обычно выполняются для первых трех формант звуков речи.

Теория, относящаяся к системам с жесткими стенками, может быть проверена измерениями на механических моделях, результаты которых

приведены ниже в § 6, В. Теоретические расчеты при этом оправдываются с хорошей степенью точности, однако результаты их справедливы для реальных условий речеобразования только частично. Действительно, в этих условиях преимущественную роль играют другие источники потерь и в формулы следовало бы включить характеристики поверхностей и стенок полостей речевого тракта, которые, однако, недостаточно хорошо известны. Хотя в настоящей работе и приведены теоретические основы расчета влияния голосовой щели, колебания стенок полостей и турбулентных потерь, обусловленных наличием постоянного потока воздуха, однако необходима еще систематическая экспериментальная проверка соответствующих постоянных для речевого тракта.

Относительное значение различных диссипативных элементов на основе их теоретической оценки рассмотрено в § 4 гл. 7 в связи с расчетами для моделей речевого тракта, соответствующих образованию шести русских гласных.

А. Резонатор Гельмгольца

При рассмотрении резонатора как системы с сосредоточенными постоянными мы имеем в соответствии с формулой (III.3):

$$B = \frac{R}{2\pi L_e} + \frac{G}{2\pi C} = B_R + B_G, \quad (\text{III. 99})$$

где $L_e = \rho l_e / A$ — эффективная индуктивность горла резонатора, $C = V / \rho c^2$ — емкость полости, имеющей объем V . Резонансная частота равна

$$F = \frac{1}{2\pi} \left(\frac{1}{L_e C} \right)^{1/2} = \frac{c}{2\pi} \left(\frac{A}{l_e V} \right)^{1/2}. \quad (\text{III. 100})$$

1) Сопротивление излучения. Доля, вносимая сопротивлением излучения в ширину резонанса, равна $R_0 / 2\pi L_e$ и может быть представлена следующим образом:

$$B_{R_0} = \frac{c A^2 K(f)}{8^2 \pi^2 V l_e^2} = 45 \left(\frac{A}{l_e} \right)^2 \left(\frac{10}{V} \right) K_S(f), \quad (\text{III. 101})$$

или, если исключить объем резонатора V ,

$$B_{R_0} = \frac{f^2 A K_S(f)}{2 c l_e} = 14 \left(\frac{f}{1000} \right)^2 \left(\frac{A}{l_e} \right) K_S(f). \quad (\text{III. 102})$$

Если, напротив, исключить проводимость горла резонатора A / l_e , где l_e — эффективная длина и A — площадь поперечного сечения горла, то

$$B_{R_0} = \frac{2\pi f^4 V K_S(f)}{c^3} = 4,5 \left(\frac{f}{1000} \right)^4 \left(\frac{V}{10} \right) K_S(f). \quad (\text{III. 103})$$

Демпфирующее влияние излучения в простых резонаторах пренебрежимо мало на низких частотах; из формулы (III. 103) можно видеть, что ширина полосы при постоянном объеме полости пропорциональна четвертой степени собственной частоты резонатора. Согласно формуле (III. 101), затухание возрастает как квадрат отношения A / l_e и изменяется обратно пропорционально объему V в том случае, если эти параметры изменяются независимо друг от друга. При больших значениях площади поперечного сечения горла A резонатор уже нельзя рассматривать как систему с сосредоточенными постоянными.

2) Потери на трение в горле резонатора в идеальном случае. Для резонатора Гельмгольца составляющая ширины резонанса, определяемая потерями на трение, согласно формулам (III.99) и (2.13), равна

$$B_{Rf} = S_A \left(\frac{\mu f}{\rho A} \right)^{1/2} = 13 \left(\frac{f}{1000A} \right)^{1/2} S_A, \quad (\text{III. 104})$$

или, исключая f ,

$$B_{Rf} = \left(\frac{\mu c}{2\pi\rho} \right)^{1/2} \left(\frac{1}{l_e A V} \right)^{1/4} S_A = 9,7 \left(\frac{100}{A l_e V} \right)^{1/4} S_A. \quad (\text{III. 105})$$

Если, напротив, исключить A , то

$$B_{Rf} = \frac{c}{2\pi} \left(\frac{\mu}{\rho l_e V f} \right)^{1/2} = 7,3 \left(\frac{1000}{f} \right)^{1/2} \left(\frac{100}{V l_e} \right)^{1/2} S_A. \quad (\text{III. 106})$$

При этом предполагается, что длина горла l столь велика, что различие концевых поправок, обусловленных активным сопротивлением и индуктивностью, можно пренебречь. Коэффициент формы

$$S_A = \frac{S}{(4\pi A)^{1/2}} \quad (\text{III. 107})$$

равен 1 для круговой формы сечения и близок к 2 для случая сильного сужения в речевом тракте (отношение ширины к высоте 9). По-видимому, в приведенные выше соотношения должен был бы входить множитель, характеризующий поверхность; следует также учитывать амплитудную зависимость затухания для низкочастотных резонансов, проявляющуюся уже при довольно умеренных уровнях интенсивности звука (см. формулу (2.15)). Формула (III. 105) показывает, что затухание обратно пропорционально корню четвертой степени из площади горла, длины горла и объема полости резонатора¹⁾.

Уменьшение площади горла в 4 раза снижает частоту резонанса на октаву, причем ширина резонансной кривой увеличивается на 40%, т. е. согласно закону $f^{-1/2}$. Этим в известной мере можно объяснить обратную зависимость ширины полосы от частоты первой форманты, если последняя ниже 300 *гц*. Впрочем, колебания стенок и нелинейно зависящее от величины потока сопротивление в сужении могут дать такой же эффект.

3) Влияние турбулентного потока воздуха. Шунтирующее влияние голосовой щели. Исследование сопротивления потоку в отверстиях и трубах с более или менее резко очерченными границами (см. § 2 Приложения II) показывает, что основная часть сопротивления, обусловленная турбулентностью, сосредоточена у входа в суженный проход, где линии тока сходятся; поэтому для сравнительно коротких труб ($l < 5$ см) это сопротивление не зависит от длины прохода. Из формулы (2.15) следует, что дифференциальное сопротивление при наличии постоянного потока равно приблизительно $R_D = \rho v / A$, где v — скорость постоянного потока воздуха, A — площадь сужения. Так как индуктивность последнего равна $L = \rho l / A$, то соответственная составляющая ширины основного резонанса

$$B_{D1} = \frac{R_D}{2\pi L} = \frac{v}{2\pi l}. \quad (\text{III. 108})$$

¹⁾ Вопрос о демпфировании в речевом тракте рассматривался Ван-ден-Бергом [8]. Его утверждение, что ширина полосы изменяется обратно пропорционально третьей степени радиуса горла, справедливо в том случае, если сохраняется постоянство частоты резонанса. Для этого необходимо, чтобы объем полости уменьшался одновременно с сужением горла, так, чтобы компенсировать обусловленное этим увеличение его индуктивности.

Соответствующая этому значению постоянная времени $1/\pi B_{D1}$ того же порядка, что и время, необходимое для переноса частицы в постоянном воздушном потоке от одного конца сужения до другого.

Уменьшение площади сужения A при постоянной величине объемной скорости u вызывает увеличение B_{D1} , пропорциональное A^{-1} или f^{-2} , в отличие от пропорциональности $f^{-1/2}$ в случае потерь на вязкость (III. 106). Демпфирование, обусловленное турбулентностью постоянного воздушного потока, оказывает, по-видимому, существенное влияние на ширину первой форманты звонких длительных, артикуляция которых характеризуется площадью сужения A около $0,2 \text{ см}^2$ и меньше и объемной скоростью более $100 \text{ см}^3/\text{сек}$. Связанное с турбулентностью демпфирование изменяется периодически, следуя за пульсациями постоянного воздушного потока. Поскольку величина последних ненамного превышает амплитуду колебания $F1$, можно полагать, что формула (III. 108) дает преувеличенные значения. Кроме того, зависимость рассматриваемого затухания от формы сужения и от характера образующих его поверхностей недостаточно ясна, так что приведенной здесь упрощенной теорией следует пользоваться с осторожностью. При очень сильном сужении, в особенности при открытой голосовой щели и связанном с этим увеличении эффективного объема, демпфирование первой форманты может превзойти критическое и B_{D1} в формуле (III. 108) в этом случае представляет собой не что иное, как граничную частоту импеданса сужения (см. § 4 гл. 10).

Влияние сопротивления голосовой щели $R_q = \frac{\rho v_q}{A_q}$ на затухание форманты $F1$ при наличии турбулентности, если принять для этой форманты модель резонатора Гельмгольца, можно видеть из равенства

$$B_{q1} = \frac{1}{2\pi R_q C_2} = \frac{c^2 A_q}{2\pi v_q V_2}, \quad (\text{III. 109})$$

где V_2 — объем речевого тракта, A_q и v_q — площадь голосовой щели и скорость частиц в ней. Последнюю можно считать приближенно постоянной, так что периодические изменения B_{q1} обусловлены только изменениями A_q . Вызываемые этим изменения в величине потерь энергии в голосовой щели происходят синхронно с изменением потерь в сужении, хотя и с задержкой во времени и искажениями формы кривой голосовых импульсов.

Составляющая B_{q1} общей ширины резонанса равна всего 12 гц, т. е. мала при нормальных условиях образования голоса. Однако при образовании глухих турбулентных длительных звуков открытая голосовая щель создает свободное сообщение с подгортанной полостью, что вызывает значительные потери энергии для тех формант, которые обусловлены в наибольшей мере задней полостью. Считая импеданс трахеи чисто активным и равным $R_t = \rho c/A_t$, найдем, что влияние его на демпфирование форманты $F1$ определяется величиной

$$B_{t1} = \frac{1}{2\pi R_t C_2} = \frac{A_t c}{2\pi V_2}, \quad (\text{III. 110})$$

которая при $V_2 = 70 \text{ см}^3$ и $A_t = 3 \text{ см}^2$ составляет 240 гц. Вообще говоря, импеданс подгортанной области следует рассматривать как шунтирующий импеданс, включенный в том конце системы полостей, где расположена голосовая щель; этот импеданс состоит из вещественной части R_t

и мнимой X_t . Влияние подобного шунтирующего импеданса рассмотрено было в § 5.

Влияние активной нагрузки R_t на ширину кривой резонанса, обусловленного возникновением стоячей волны в фарингальной полости, при длине последней l_2 , согласно формулам (III.44), (III.52) и (III.98), определяется выражением

$$B_{t2} = \frac{Z_2 c}{\pi R_t l_2} = \frac{\rho c^2}{\pi R_t V_2} = \frac{A_t c}{\pi V_2}, \quad (\text{III. 111})$$

т. е. вдвое больше, чем для первой форманты.

4) Потери, обусловленные теплопроводностью и колебаниями стенок полостей. Потери, обусловленные теплопроводностью стенок полости резонатора Гельмгольца, при площади поперечного сечения горла A_1 значительно меньшей, чем сечение полости A_2 , малы по сравнению с потерями на вязкость в горле. Из формул (2.13) и (III.99), принимая одинаковый коэффициент формы для полости и для горла, найдем, что составляющая ширины резонанса, обусловленная потерями на теплопроводность, равна

$$B_G = 0,45 \left(\frac{A_1}{A_2} \right)^{1/2} B_R, \quad (\text{III. 112})$$

где B_R — составляющая, обусловленная потерями на вязкость в горле. Используя формулы (2.12) и (III.99), можно получить общее выражение

$$B_G = \frac{S(\kappa - 1)}{A} \left[\frac{K_h f c}{4\pi \rho C_p} \right]^{1/2} = 5,8 \left(\frac{f}{1000 A_2} \right)^{1/2} S_{A_2}, \quad (\text{III. 113})$$

где S_{A_2} — коэффициент формы основной полости (см. формулу (III.107)). Формула (III.113) сходна с формулой (III.104). Потери на теплопроводность нужно учитывать только при детальном анализе затухания в системах с жесткими стенками. Потери, вызываемые колебаниями стенок полостей речевого тракта могут, согласно Ван-ден-Бергу [8], отразиться на демпфировании только первой форманты. Параллельно включенная активная часть $1/G'_w$ импеданса стенок, рассчитанная на единицу длины, определяет составляющую ширины первой форманты для передних гласных в соответствии с выражением

$$B_{G_w} = \frac{G'_w}{2\pi C'_2}, \quad (\text{III. 114})$$

где C'_2 — емкость, рассчитанная на единицу длины фаринкса. Ван-ден-Берг оценивает эту составляющую величиной около 25—50 гц при 300 гц; она убывает пропорционально $f^{-2,5}$. Теоретически на этот вид потерь может оказать влияние напряженность мышц горла, однако вряд ли это влияние имеет сколько-нибудь заметное значение.

Б. Стоячие волны в трубах

1) Демпфирование в открытой трубе, обусловленное излучением. Согласно формулам (2.16), (III.43), (III.52) и (III.98) составляющая ширины резонанса, обусловленная импедансом излучения, равна

$$B_{R_0} = \frac{R_0 c}{Z l_e \pi} = \frac{f^2 A}{c l_e} K_S(f) \quad (\text{III. 115})$$

при условии $R_0/Z < 0,5$.

Приведенное к эффективной длине 17,6 см и площади поперечного сечения 8 см², это выражение будет иметь вид

$$B_{R_0} = 29 \left(\frac{f}{1500} \right)^2 \frac{A}{8} \cdot \frac{17,6}{l_e} K_S(f). \quad (\text{III. 116})$$

При частоте второго резонанса 1500 гц величина $K_S(f) = 1,6$, так что $B_{R_0} = 46$ гц. Для основного резонанса при 500 гц $B_{R_0} = 3,9$ гц. Те же результаты получим и по формуле (III.102) для резонатора Гельмгольца, если переднюю половину трубы считать горлом, а заднюю — полостью.

Возрастание сопротивления излучения, пропорциональное квадрату частоты (и даже более быстрое, что учитывается коэффициентом $K_S(f)$), приводит к быстрому увеличению демпфирования для более высоких формант. Так, $B_{3R_0} = 137$ гц, а $B_{4R_0} = 224$ гц. При этом, однако, следует учитывать, что наличие индуктивности излучения приводит к изменению степени согласования между импедансом трубы и ее нагрузкой (см. § 3 приложения III).

Далее, демпфирование пропорционально отношению A/l_e . Поэтому короткие передние резонаторы речевого аппарата более чувствительны к влиянию излучения, чем полностью открытый неартикулированный речевой тракт.

2) Демпфирование стоячих волн в трубе с малым отверстием, обусловленное излучением. Если труба с поперечным сечением A_2 и длиной l_2 заканчивается коротким горлом, имеющим эффективную длину l_{le} и весьма малую площадь поперечного сечения A_1 , то при условии $l_{le}/c < 0,5$ и $l_1 A_2 / l_2 A_1 > 5$ приводимая ниже расчетная формула дает точность порядка 10%; эта формула может быть выведена из формул (III.48) и (III.92):

$$B_{R_0} = \frac{c A_1^2}{4 \pi^2 l_{le}^2 l_2 A_2} K_S(f). \quad (\text{III. 117})$$

Частотная зависимость R_0 здесь оказывается скомпенсированной уменьшением связи между сопротивлением излучения и основной полостью трубы. Для расчетных целей формулу удобно представить в виде

$$B_{R_0} = 90 \left(\frac{A_1}{l_{le}} \right)^2 \left(\frac{10}{l_2 A_2} \right) K_S(f). \quad (\text{III. 118})$$

Так же, как и для случая основного резонанса простого резонатора (формула III.101), составляющая ширины резонансной кривой здесь уменьшается как квадрат «параметра огубления» l_{le}/A_1 ; этим объясняется тот факт, что ширина полосы формант для огубленных гласных меньше, чем для неогубленных.

3) Демпфирование стоячих волн, обусловленное влиянием стенок полостей. Пренебрегая поправками (формула III.48), т. е. принимая оба конца трубы или абсолютно закрытыми или полностью открытыми, и полагая

$$B_x = \frac{c}{\pi} (\alpha_R + \alpha_G) = \frac{R}{2\pi L} + \frac{G}{2\pi C} = B_R + B_G, \quad (\text{III. 119})$$

где R , L , G и C — распределенные постоянные, рассчитанные на единицу длины, можно, используя формулу (2.13), определить составляющую B_R ширины резонанса, обусловленную последовательным сопротивлением потерь, и составляющую B_G , обусловленную параллельным сопротивлением.

Пренебрегая влиянием колебаний стенок на α_G , получим нормированные выражения для B_a , B_G и B_R в виде

$$B_a = 18,5 \left(\frac{f}{1000A} \right)^{1/2} S_A; \quad B_G = 0,31 B_a; \quad B_R = 0,69 B_a. \quad (\text{III. 120})$$

Следует отметить, что в выражения для B_R и B_G входит один и тот же коэффициент формы S_A , несмотря на различие в физической природе этих составляющих; кроме того, составляющие эти не зависят от длины трубы. Сходство расчетных формул (III.121), (III.113) и (III.104) определяется одинаковой формой выражений (III.99) и (III.119) ¹⁾.

В. Эксперименты и расчеты, относящиеся к резонаторам с жесткими стенками, состоящим из одного и двух отрезков труб

1) Труба, открытая на одном конце. Расчеты были получены частоты и ширина резонансных пиков для идеальной цилиндрической трубы с жесткими стенками, имеющей площадь поперечного сечения 8 см^2 и длину $16,4 \text{ см}$, заделанную в сферический экран диаметром 18 см ; результаты расчета сравнивались с данными измерений, выполненных на латунной трубе в деревянной сфере. Открытый конец трубы заканчивался плоским фланцем диаметром $7,5 \text{ см}$, так что исследованная экспериментально модель несколько отличалась от расчетной.

Маленький конденсаторный микрофон, расположенный на закрытом конце трубы, являлся излучателем звука. Вся модель подвешивалась свободно в заглушенной камере.

На расстоянии 25 см от открытого конца трубы помещался миниатюрный измерительный микрофон. Измерение частот резонанса и частот, соответствующих понижению уровня на 3 дБ относительно резонансного пика, производилось декадным частотомером. Измерения производились при температуре 21°C , т. е. при скорости звука $c = 34\,400 \text{ см/сек}$.

Результаты расчетов и измерений приведены в табл. 12. Из таблицы видно, что измеренные частоты резонансов хорошо совпадают с расчетными; различие не превосходит ошибки экспериментального определения частоты. Совпадение для ширины резонансных пиков также достаточно хорошее, чем подтверждается и правильность частотной зависимости $K_S(\omega)$, найденной теоретически для поршня круговой формы на поверхности сферического экрана.

Отклонение расчетных данных от измеренных положительно и не превосходит 10% . Это расхождение можно считать малым, учитывая неполное соответствие модели условиям расчета.

2) Модель резонатора из двух отрезков труб. Теория резонатора, состоящего из двух отрезков труб, проверялась сравнением расчетных и измеренных величин на модели, соответствующей речевому тракту при произнесении гласного [I]²⁾; параметры переднего отрезка

¹⁾ Последние расчеты Хауза и Стивенса [102], в которых были использованы данные Франка (1951) об удельном сопротивлении стенок полостей речевого тракта, равном $z_w = (r_w + j\omega l_w)$, где $r_w = 200$ и $l_w = 0,02$, дали следующие значения ширины полосы формант нейтрального гласного: $B_{1G} = 70 \text{ гц}$ при $F_1 = 500 \text{ гц}$, $B_{2G} = 40 \text{ гц}$ при $F_2 = 1500 \text{ гц}$ и $B_{3G} = 20 \text{ гц}$ при $F_3 = 2500 \text{ гц}$. Полученное значение B_{1G} , таким образом, вдвое больше, чем рассчитано Ван-ден-Бергом [8] для модели [I]. Это расхождение не может быть обусловлено различием в типах резонаторов, хотя первая форманта [I] является типичным гельмгольцевым резонансом, а первая форманта нейтрального гласного — четвертьволновым резонансом стоячей волн.

²⁾ Фонетическое качество — слегка огубленное.

Таблица 12

Частота резонанса F , гц		Расчетная ширина резонансного пика, обусловленная			Полная ширина резонансного пика B , гц	
расчетная	измеренная	вязкостью и теплопроводностью B , гц	излучением		расчетная	измеренная
			K_3	B_{R_0} , гц		
486	485	4,5	1,1	3,4	7,9	8,0
1459	1459	7,9	1,6	43,6	51,5	44,0
2445	2434	10,2	1,7	133,0	143,0	128,0
3444	3442	12,1	1,45	225,0	237,0	228,0

$A_1 = 1 \text{ см}^2$, $l_1 = 6 \text{ см}$, заднего — $A_2 = 8 \text{ см}^2$, $l_2 = 7,98 \text{ см}$. Как и в предыдущем примере, модель помещалась в сферический экран.

Концевая поправка, учитывающая индуктивность излучения, составляет 0,45 см; аналогичная поправка для внутреннего конца переднего отрезка, вычисленная по формуле (2.17), равнялась 0,27 см. Эффективная длина переднего отрезка с учетом поправок равна, таким образом, 6,72 см. Частота основного резонанса, по полной формуле (III.81), равна $F_1 = 255 \text{ гц}$.

Если воспользоваться формулой для простого резонатора Гельмгольца (формула (III.100)), то расчет дает величину несколько большую, а именно $F_1 = 264 \text{ гц}$, однако если в этой формуле к индуктивности горла добавить половину индуктивности заднего отрезка, основанием для чего может служить приближенная эквивалентная схема линии на низких частотах, то расчетная величина понижается до $F_1 = 247 \text{ гц}$. Значительно лучшие результаты получаются, если в качестве добавочной индуктивности учитывать только треть длины заднего отрезка, как это подсказывает разложение $\text{ctg } \omega l_2/c$ в степенной ряд; в этом случае $F_1 = 253 \text{ гц}$.

Приближенное значение $F_2 = c/2l_2 = 2150 \text{ гц}$, соответствующее полуволновому резонансу заднего отрезка, не очень далеко от более точного расчетного значения, согласно формуле (III.81) равного $F_2 = 2045 \text{ гц}$. Частота третьего резонанса $F_3 = 2664 \text{ гц}$ близка к полуволновому резонансу переднего отрезка $c/2l_1 = 2560 \text{ гц}$; наконец, частота четвертого резонанса 4290 гц очень близка к величине $c/l_2 = 4310 \text{ гц}$.

Расчетные и измеренные величины частот резонансов и ширины резонансных пиков приведены в табл. 13.

Таблица 13

	Частота резонанса F , $гц$			Расчетная ширина резонансных пиков, обусловленная										Общая ширина B , $гц$	
	расчетная	измеренная		излучением		вязкостью переднего отрезка		теплопроводностью переднего отрезка		вязкостью заднего отрезка		теплопроводностью заднего отрезка		расчетная	измеренная
				B_{R_0} , $гц$	k_{R_0}	B_{R_1} , $гц$	k_{R_1}	B_{G_1} , $гц$	k_{G_1}	B_{R_2} , $гц$	k_{R_2}	B_{G_2} , $гц$	k_{G_2}		
F_1	255	253		0,3	0,96	6,2	0,96	0,1	0,03	0,1	0,045	1,0	0,97	7,7	8,6
F_2	2045	2047		5,5	0,19	3,5	0,19	2,3	0,28	5,8	0,81	2,1	0,72	18,7	26
F_3	2664	2653		40,1	0,80	16,9	0,80	7,3	0,74	1,5	0,20	0,9	0,26	70,7	74
F_4	4290	4289		16,0	0,12	3,1	0,12	1,7	0,14	8,3	0,88	3,7	0,86	32,8	41

Как можно видеть из таблицы, для частот резонансов совпадение расчетных и измеренных данных очень хорошее. Ошибку в $+1\%$ для F_1 и $+0,4\%$ для F_3 можно объяснить уменьшением эффективной площади поперечного сечения переднего отрезка трубы, определяющимся уравнением (2.11). Расчет поправок, учитывающих это уменьшение, дает значения соответственно $-1,2\%$ и $-0,4\%$.

Для тех резонансных пиков, ширина которых обусловлена в большей степени потерями в переднем отрезке трубы, измеренные значения B_1 и B_3 весьма близки к расчетным; для B_2 измеренная величина на 40% больше расчетной. В литературе данных об исследовании подобных моделей нет. Однако можно отметить как общую тенденцию, что демпфирование, обусловленное вязкими потерями, больше того, которое дает расчет в предположении совершенно жестких стенок.

Все полученные расчетом составляющие ширины резонансов меньше, чем дает расчет для одного из двух отрезков, рассматриваемого отдельно. Величины $k_v = \frac{\pi B_v}{c\alpha_v}$, где $v = R1, G1, R2$ или $G2$ характери-

зуют доли, вносимые соответствующими коэффициентами затухания $\alpha_{R1}, \alpha_{G1}, \alpha_{R2}, \alpha_{G2}$, связанными с разными видами потерь. Как было отмечено выше, величины k_v показательны с точки зрения зависимости резонансов от свойств отдельных элементов системы. Большое k_{G1} для третьего резонанса указывает на его связь с образованием стоячей волны в переднем отрезке. Аналогично большое k_{G2} для F_2 связано с возникновением стоячей волны в заднем отрезке. Между величинами k_v , как можно видеть, соблюдается соотношение (III.84), а именно $k_{R1} + k_{R2} = k_{G1} + k_{G2} = 1$.

При возникновении стоячей волны у закрытого конца трубы получается максимум давления и минимум скорости частиц, у открытого же конца, наоборот, скорость максимальна, а давление минимально. Если длина волны меньше удвоенной длины трубы, в ней возникают дополнительные узлы и пучности (см. § 4 гл. 7).

На частотах формант сонорных звуков скорость у голосовой щели хотя и конечна, но мала, и велика в отверстии рта. Распределение скорости по длине речевого тракта на частоте второй форманты имеет дополнительный минимум, помимо расположенного у голосовой щели. На частоте третьей форманты возникают два таких минимума, на четвертой — три и т. д. (см. § 4 гл. 7).

Если длина отрезка или полости речевого тракта велика по сравнению с длиной волны, то для потерь, зависящих от величины давления и определяющих α_G , и для потерь, зависящих от величины скорости и определяющих α_R , справедливо соотношение $k_G \approx k_R$. При частоте резонанса, соответствующего стоячей волне, как k_G , так и k_R близки к 1. Однако при основной частоте резонатора, состоящего из двух отрезков труб, для основной полости k_G велико и близко к 1, но k_R мало; обратное соотношение имеет место для k_G и k_R отрезка, играющего роль горла.

ЛИТЕРАТУРА

1. Ayers E. W., Address given at the S. R. D. E. colloquium, 1955, Ministry of Supply, Christchurch, England. S. R. D. E. Rep., No. 1100, 28—32 (1956).
2. Barczinski L., Tnienhaus E., Klangspektren und Lautstärke deutscher Sprachlaute. Arch. Néerland. Phon. Exp. 11, 47—68 (1935).
3. Bayston T. E., Campanella S. J., Continuous analysis speech band-width compression system. J. Acoust. Soc. Am. 29, 1255 (A) (1957).
4. von Békésy G., Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. Phys. Z. 30, 721—745 (1929).
5. von Békésy G., The structure of the middle ear and the hearing of one's own voice by bone conduction. J. Acoust. Soc. Am. 21, 217—232 (1949).
6. Bell Telephone Laboratories, Technical aspects of visible speech. Bell Telephone System, Monograph B-1415 (1946). J. Acoust. Soc. Am. 17, 1—89 (1946).
7. Beranek L. L., Acoustic Measurements (New York, 1949).
8. van den Berg Jw., Physica van de stemvorming, met toepassingen, diss., Rijks-universiteit te Groningen ('s-Gravenhage, 1953).
9. van den Berg Jw., Sur les théories myo-élastique et neuro-chronaxique de la phonation. Rev. de Laryng. 74, 494—511 (1954).
10. van den Berg Jw., Über die Koppelung bei der Stimmbildung. Z. Phonet. 8, 5/6, 281—293 (1954—1955).
11. van den Berg Jw., Calculations on a model of the vocal tract for vowel /i/ (meat) and on the larynx. J. Acoust. Soc. Am. 27, 332—337 (1955).
12. van den Berg Jw., On the rôle of the laryngeal ventricle in voice production. Folia Phoniatica 7, 57—69 (1955).
13. van den Berg Jw., Direct and indirect determination of the mean subglottic pressure. Folia Phoniatica 8, 1—24 (1956).
14. van den Berg Jw., Subglottic pressures and vibrations of the vocal folds. Folia Phoniatica 9, 65—71 (1957).
15. van den Berg Jw., Zantema J. T., Doornenbal Jr., On the air resistance and the Bernoulli effect of the human larynx. J. Acoust. Soc. Am. 29, 626—631 (1957).
16. Boeryd A., Undersökning av taleffekten (volymen) från en telefonapparat som funktion av telefonförbindelsens kvalitet. Examensarbete i telegrafi och telefoni, Royal Institute of Technology, Stockholm (1957).
17. Bogert B. P., On the bandwidth of vowel formants. J. Acoust. Soc. Am. 25, 791—792 (1953).
18. Boyanus S. C., A Manual of Russian Pronunciation, 2nd ed. (London, 1944).
19. Broch O., Slavische Phonetik (Heidelberg, 1911).
20. Chang S.-H., Two schemes of speech compression system. J. Acoust. Soc. Am. 28, 565—572 (1956).
21. Chang S.-H., Stubbs H. L., Dolanský L. O., Wiren J., Denes P., Howard C. R., Carrabes M. J., Visual message presentation. Northeastern Univ., Electronics Res. Lab., Scientific Report, No. 5 (AFCRC-TN-56-582) (1956).
22. Cherry C., On Human Communication (London, 1957).
23. Cherry C., Halle M., Jakobson R., Toward a logical description of languages in their phonemic aspects. Language 29, 34—46 (1953).
24. Chiba T., Kajiyama M., The Vowel—Its Nature and Structure (Tokyo, 1941).
25. Cooper F. S., Delattre P. C., Liberman A. M., Borst J. M., Gerstman L. J., Some experiments on the perception of synthetic speech sounds. J. Acoust. Soc. Am. 24, 597—606 (1952).

26. Crandall I. B., Sounds of speech. *Bell System Techn. J.* 4, 586—626 (1925).
27. Crandall I. B., Dynamical study of the vowel sounds. *Bell System Techn. J.* 6, 100—116 (1927).
28. David Jr. E. E., Signal theory in speech transmission. *Bell Telephone System, Monograph* 2831 (1956).
29. David Jr. E. E., Artificial Auditory Recognition in Telephony. *IBM J. of research and development* 4, 294—309 (1958).
30. Delattre P., The Physiological interpretation of sound spectrograms. *PMLA*, LXVI, 864—875 (1951).
31. Delattre P., Les attributs acoustiques de la nasalité vocalique et consonantique. *Studia Linguistica*, VIII, 103—109 (1954).
32. Delattre P., Liberman A. M., Cooper F. S., Voyelles synthétiques à deux formantes et voyelles cardinales. *Le Maître Phonétique* 96, 30—36 (1951).
33. Delattre P., Liberman A. M., Cooper F. S., Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am.* 27, 769—773 (1955).
34. Delattre P., Liberman A. M., Cooper F. S., Gerstman L. J., An experimental study of the acoustic determinants of vowel color. *Word* 8, 195—210 (1952).
35. Dieth E., *Vademekum der Phonetik* (Bern, 1950).
36. Dudley H., Remaking speech. *J. Acoust. Soc. Am.* 11, 165—177 (1939).
37. Dudley H., Fundamentals of speech synthesis. *Bell Telephone System Monograph* 2648 (1956).
38. Dudley H., Riesz R. R., Watkins S. S. A., A synthetic speaker. *J. Franklin Inst.*, 227, 739 (1939).
39. Dunn H. K., The calculation of vowel resonances and an electrical vocal tract. *J. Acoust. Soc. Am.* 22, 740—753 (1950).
40. Dunn H. K., White S. D., Statistical measurements on conversational speech. *J. Acoust. Soc. Am.* 11, 278 (1940).
41. Eck B., *Technische Strömungslehre* (Berlin, 1944).
42. Essner C., Recherches sur la structure des voyelles orales. *Arch. Néerland. Phon. Exp.* 20, 40—77 (1947).
43. Faaborg-Andersen K., *Electromyographic Investigation of Intrinsic Laryngeal Muscles in Humans* (Copenhagen, 1957).
44. Fano R. M., The information theory point of view in speech communication. *J. Acoust. Soc. Am.* 22, 691—696 (1950).
45. Fant C. G. M., Analys av de svenska vokalljuden. *L M Ericsson protokoll H/P* 1035 (1948).
46. Fant C. G. M., Analys av de svenska konsonantljuden. *L M Ericsson protokoll H/P* 1064 (1949).
47. Fant C. G. M., Transmission properties of the vocal tract. *MIT Acoustics Lab. Quarterly Progress Rep.*, July—Sep. 20—23 (1950).
48. Fant C. G. M., Transmission properties of the vocal tract. Part II. *MIT Acoustics Lab. Quarterly Progress Rep.*, Oct.—Dec. 14—19 (1950).
49. Fant C. G. M., Transmission properties of the vocal tract with application to the acoustic specification of phonemes. *MIT Acoustics Lab. Techn. Rep.*, No. 12 (1952).
50. Fant C. G. M., Speech communication research. *IVA* 24 (Royal Swedish Academy of Engineering Sciences), 331—337 (1953).
51. Fant C. G. M., Discussion of paper read by G. E. Peterson at the 1952 Symposium on the Applications of Communication Theory. *Communication Theory*, ed. W. Jackson, 421—424 (London, 1953).
52. Fant C. G. M., On the predictability of formant levels and spectrum envelopes from formant frequencies. For Roman Jakobson, 109—120 (s'-Gravenhage, 1956).
53. Fant C. G. M., Modern instruments and methods for acoustic studies of speech. *Royal Inst. of Technology, Div. of Telegraphy-Telephony, Rep.*, No. 8 (1957). *Proc. of the VIII Int. Congr. of Linguists* 1957, 282—358 (Oslo, 1958). Also published in *Acta Polytechnica Scandinavica*, No. 1, 1—81 (1958).
54. Fant C. G. M., Acoustic theory of speech production. *Royal Inst. of Technology, Div. of Telegraphy-Telephony, Rep.*, No. 10. Presented in mimeographed form (1958).
55. Fant C. G. M., Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics*, No. 1, 3—108 (1959).
56. Farnsworth D. W., High speed motion pictures of the human vocal cords. *Bell Telephone Lab., Record* 18, 203 (1940).
57. Fischer-Jørgensen E., Acoustic analysis of stop consonants. *Miscellanea Phonetica*, II, 42—59 (1954).
58. Fischer-Jørgensen E., The commutation test and its applicability to phonemic analysis. For Roman Jakobson, 140—151 (s'-Gravenhage, 1956).

59. Fischer-Jørgensen E., What can the new techniques of acoustic phonetics contribute to linguistics? *Proc. of VIII Int. Congr. of Linguists 1957*, 433—478 (Oslo, 1958).
60. Fischer-Jørgensen E., Die Bedeutung der funktionellen Sprachbeschreibung für die Phonetik. *Phonetica*, Suppl. ad Vol. 4, 7—28 (1959).
61. Fischer-Jørgensen E., Tybjærg Hansen A., An electrical manometer and its use in phonetic research. *Phonetica* 4, 43—53 (1959).
62. Flanagan J. L., A speech analyzer for a formant-coding compression system. MIT Acoustics Lab. Scientific Rep., No. 4 (AFCRC-TN-55-793) (1955).
63. Flanagan J. L., Difference limen for vowel formant frequency. *J. Acoust. Soc. Am.* 27, 613—617 (1955).
64. Flanagan J. L., Band width and channel capacity necessary to transmit the formant information of speech. *J. Acoust. Soc. Am.* 28, 592—596 (1956).
65. Flanagan J. L., Estimates of the maximum precision necessary in quantizing certain 'dimensions' of vowel sounds. *J. Acoust. Soc. Am.* 29, 533—534 (1957).
66. Flanagan J. L., Note on the design of 'terminal-analog' speech synthesizers. *J. Acoust. Soc. Am.* 29, 306—310 (1957).
67. Flanagan J. L., Some properties of the glottal sound source. *J. of Speech and Hearing Research* 1, 99—116 (1958).
68. Flanagan J. L., A resonance-vocoder and baseband complement: A hybrid system for speech transmission. *IRE WESCON Convention Record* 3, 5—16 (1959).
69. Flanagan J. L., House A. S., Development and testing of a formant-coding speech compression system. *J. Acoust. Soc. Am.* 28, 1099—1106 (1956).
70. Fletcher H., *Speech and Hearing* (New York, 1929; later edition 1953).
71. Forchhammer J., Die Sprachlaute in Wort und Bild (Heidelberg, 1942).
72. Forchhammer J., Almindelig talelaere, Chapter XII in *Nordisk Laerebog for Talepaedagoger*, Almindelig Del, 218—242 (1954).
73. Fry D. B., Duration and intensity as physical correlates to stress. *J. Acoust. Soc. Am.* 27, 765—768 (1955).
74. Gabor D., Theory of communication. *J. Inst. Elect. Engrs.* 93, Part III, 429—457 (1946).
75. Gabor D., A summary of communication theory. *Proc. of the 1952 Symposium on the Applications of Communication Theory*. *Communication Theory*, ed. W. Jackson, 1—24 (London, 1953).
76. Ganeshsundaram P. C., A cascade modulation theory of speech formants. *Z. f. Phonet.* 10/1, 1—7 (1957).
77. Garner W. R., The loudness and loudness matching of short tones. *J. Acoust. Soc. Am.* 21, 398—403 (1949).
78. Grützmaier M., Lottermoser W., Über ein Verfahren zur trägheitsfreien Aufzeichnung von Melodiekurven. *Akust. Z.*, 242—248 (1937).
79. Haase K. H., Vilbig F., Errors in spectrum analysis by a set of narrow band selecting filters. AFCRC, Bedford, Communications Lab. (AFCRC-TR-56-121) (1956).
80. Hála B., *Nature Acoustique des Voyelles*. *Acta Universitatis Carolinae* (Prague, 1956).
81. Halle M., The strategy of phonemics. *Word* 10, 197—209 (1954).
82. Halle M., The Russian Consonants. A Phonemic and Acoustical Investigation, Dr. Phil. thesis, Harvard University (1954).
83. Halle M., Review of «Manual of Phonology» by C. F. Hockett, in *J. Acoust. Soc. Am.* 28, 509—511 (1956).
84. Halle M., In defense of the number two. *Studies Presented to Joshua Whatmough on His Sixtieth Birthday*, 65—72 ('s-Gravenhage, 1957).
85. Halle M., The Sound Pattern of Russian ('s-Gravenhage, 1959).
86. Halle M., Hughes G. W., Radley J. P., Acoustic properties of stop consonants. *J. Acoust. Soc. Am.* 29, 107—116 (1957).
87. Halsey R. J., Swaffield J., Analysis-synthesis telephony, with special reference to the vocoder. *Proc. Inst. Elect. Engrs.* 95, 391—411 (1948).
88. Hardy H. C. and others, Symposium on sound level meters. *J. Acoust. Soc. Am.* 29, 1330—1341 (1957).
89. Harris C. M., A study of the building blocks in speech. *J. Acoust. Soc. Am.* 25, 962—969 (1953).
90. Harris C. M., A speech synthesizer. *J. Acoust. Soc. Am.* 25, 970—975 (1953).
91. Harris Z. S., *Methods in Structural Linguistics* (Chicago, 1951).
92. Hartley R. V. L., Transmission of information. *Bell System Techn. J.* 7, 535 (1928).
93. Hattori S., Yamamoto K., Fujimura O., Nasalization of vowels and nasals. *Bull. of the Kobayasi Inst. of Phys. Res.* 6, 226—235 (1956).

94. Heffner R. M. S., General Phonetics (Madison, 1949).
95. Heinz J. M., Fricative consonants. MIT Acoustics Lab. Quarterly Rep., Oct. — Dec., 5—7 (1956).
96. Heinz J. M., Fricative consonants. MIT Acoustics Lab. Quarterly Rep., April — June, 1 (1957).
97. Heinz J. M., A terminal analog of fricative consonant articulation. MIT Acoustics Lab. Quarterly Rep., July — Sep., 1—3 (1957).
98. Hockett C. F., Manual of Phonology, Indiana Univ. Publications in Anthropology and Linguistics, No. 11 (Bloomington, 1955).
99. House A. S., Analog studies of nasal consonants. J. of Speech and Hearing Disorders 22, 190—204 (1957).
100. House A. S., Stevens K. N., Analog studies of the nasalization of vowels. J. of Speech and Hearing Disorders 21, 218—232 (1956).
101. House A. S., Stevens K. N., Measurements of the transient response of the vocal tract. MIT Acoustics Lab. Quarterly Rep., July — Sep., 3—5 (1957).
102. House A. S., Stevens K. N., Estimation of formant band widths from measurements of transient response of the vocal tract. J. of Speech and Hearing Research 1, 309—315 (1958).
103. Howard C. R., Speech analysis-synthesis scheme using continuous parameters. J. Acoust. Soc. Am. 28, 1091—1098 (1956).
104. Huggins W. H., A phase principle for complex-frequency analysis and its implications in auditory theory. J. Acoust. Soc. Am. 24, 582—589 (1952).
105. Hughes G. W., Halle M., Spectral properties of fricative consonants. J. Acoust. Soc. Am. 28, 303—310 (1956).
106. Husson R., Etude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée, thèse, l'Univ. de Paris (Paris, 1950).
107. Husson R., Stemmehandsvibrationernes fysiologi (translation by M. Kloster-Jensen). Nord. Tidskrift for Tale og Stemme 16, 49—73 (1956).
108. Ingård U., On the theory and design of acoustic resonators. J. Acoust. Soc. Am. 25, 1037—1067 (1953).
109. Ingård U., On the spectra of explosive speech sounds. MIT Acoust. Lab. Quarterly Rep., July — Sep. 13—15 (1956).
110. Jakobson H., Information and the human ear. J. Acoust. Soc. Am. 23, 463—471 (1951).
111. Jakobson R., Observations sur le classement phonologique des consonnes. Proc. of the III Int. Congr. of Phonetic Sciences, Ghent 1939, 34 (1939).
112. Jakobson R., Kindersprache, Aphasie und allgemeine Lautgesetze. Språkvetenskapl. Sällskapet i Uppsala Förhandl. (1940—1942).
113. Jakobson R., Die Verteilung der stimmhaften und stimmlosen Geräuschlaute im Russischen. Festschrift für Max Vasmer, 199—202 (Berlin, 1956).
114. Jakobson R., Fant C. G. M., Halle M., Preliminaries to speech analysis. The distinctive features and their correlates. MIT Acoustics Lab. Techn. Rep., No. 13 (1952); 3rd printing.
115. Jakobson R., Halle M., Fundamentals of Language ('s-Gravenhage, 1956).
116. Jakobson R., Halle M., Phonology in relation to phonetics. Manual of Phonetics, 215—251 (Amsterdam, 1957).
117. Jassem W., The phonology of Polish stress. Word 15, 252—269 (1959).
118. Jones D., An Outline of English Phonetics (Leipzig, 1934).
119. Jones L. G., Acoustic Patterns of the Russian Vowels, Dr. Phil. thesis, Harvard University (1952).
120. Jones L. G., The Vowels of English and Russian: An Acoustic Comparison. Word 9, 354—361 (1953).
121. Joos M., Acoustic Phonetics. Language 24, 1—136 (1948).
122. Koenig W., A new frequency scale for acoustic measurements. Bell Telephone Lab., Record 27, 299—301 (1949).
123. Koneczna H., Zawadowski W., Obrazy Rentgenograficzne Glosek Rosyjskich (Warszawa, 1956).
124. Küpfmüller K., Warns O., Sprachsynthese aus Lauten. Nachrichtentechn. Fachber. 3, 28—31 (1956).
125. Ladefoged P., Broadbent D. E., Information conveyed by vowels. J. Acoust. Soc. Am. 29, 98—104 (1957).
126. Laurent T., Matematisk behandling av kontinuerligt inhomogena ledningar medelst ekvivalenter samt exempel på metodens användning för olika praktiska problem. Tekn. Medd. K. Telegrafstyrelsen, 113—133 (1940).
127. Laurent T., On kontinuerligt inhomogena ledningar. Tekn. Medd. K. Telegrafstyrelsen, 186 (1940).

128. Laurent T., Delay time and transient time in electrical filters with phase distortion. Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory, ed. W. Jackson, 310—313 (London, 1953).
129. Laurent T., Vierpoltheorie und Frequenztransformation (Berlin, 1956).
130. Lawrence W., The synthesis of speech from signals which have a low information rate. Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory, ed. W. Jackson, 460—469 (London, 1953).
131. Lewis D., Vocal resonance. J. Acoust. Soc. Am. 8, 91 (1936).
132. Liberman A. M., Some results of research on speech perception. J. Acoust. Soc. Am. 29, 117—123 (1957).
133. Liberman A. M., Delattre P. C., Gerstman L. J., Cooper F. S., Tempo of frequency change as a cue for distinguishing classes of speech sounds. J. of Experimental Psychology 52, 127—137 (1956).
134. Licklider J. C. R., Basic correlates of the auditory stimulus. Handbook of Experimental Psychology, 985—1039 (New York, 1951).
135. Licklider J. C. R., On the process of speech perception. J. Acoust. Soc. Am. 24, 590—594 (1952).
136. Licklider J. C. R., Miller G. A., The perception of speech. Handbook of Experimental Psychology, 1040—1074 (New York, 1951).
137. Lisker L., Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33, 42—49 (1957).
138. Lotz J., Speech and language. J. Acoust. Soc. Am. 22, 712—717 (1950).
139. Lotz J., The structure of human speech. Transactions of the New York Academy of Sciences, Ser. II, 16, No. 7, 373—384 (1954).
140. Lundell J. A., Etude sur la prononciation russe (Stockholm, 1890).
141. MacMillan A. S., Kelemen G., Radiography of the supraglottic speech organs. A. M. A. Archives of Otolaryngology 55, 681—682 (1952).
142. Malécot A., An experimental study of force of articulation. Studia Linguistica, IX, 35—44 (1955).
143. Malécot A., Acoustic cues for nasal consonants. Language 32, 274—284 (1956).
144. Malmberg B., Le problème du classement des sons du langage et quelques questions connexes. Studia Linguistica, VI, 1—56 (1952).
145. Malmberg B., Distinctive features of Swedish vowels; some instrumental and structural data. For Roman Jakobson, 316—321 ('s-Gravenhage, 1956).
146. Mandelbrot B., An informational theory of the statistical structure of language. Proc. of the 1952 Symposium on the Applications of Communication Theory. Communication Theory, ed. W. Jackson, 486—500 (London, 1953).
147. Mason W. P., Electromechanical transducers and wave filters (New York, 1948).
148. Menzerath P., de Lacerda A., Koartikulation, Steuerung und Lautabgrenzung (Berlin—Bonn, 1933).
149. Meyer-Eppler W., Die Schwingungsanalyse nach dem Suchton-Verfahren. Archiv der Elektr. Übertragung 4, 331—338 (1950).
150. Meyer-Eppler W., Zum Erzeugungsmechanismus der Geräuschlaute. Z. für Phonetik 7, No. 3/4, 196—212 (1953).
151. Miller G. A., The perception of short bursts of noise. J. Acoust. Soc. Am. 20, 160—170 (1948).
152. Miller G. A., Language and Communication (New York, 1951).
153. Miller G. A., Nicely P. E., An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27, 338—352 (1955).
154. Miller R. L., Nature of the vocal cord wave. J. Acoust. Soc. Am. 28, 159 (1956).
155. Morse Ph. M., Vibration and Sound (New York, 1948).
156. Munson W. A., The growth of auditory sensation. J. Acoust. Soc. Am. 19, 584—591 (1947).
157. Nielsen A. K., Acoustic resonators of circular cross-section and with axial symmetry. Trans. Dan. Acad. Techn. Sci. 10, 9—70 (1949).
158. Ochiai Y., Fukumura T., Nakatani K., Timbre study of nasals, Part II. Memoirs of the Faculty of Engineering, Nagoya University 9, 160—173 (1957).
159. O'Connor J. D., Gerstman L. J., Liberman A. M., Delattre P. C., Cooper F. S., Acoustic cues for the perception of initial (w, j, r, l) in English. Word 13, 24—43 (1957).
160. Paget Sir R., Human Speech (London, 1930).
161. Parmenter C. E., Treviño S. N., Vowel positions as shown by X-rays. The Quarterly J. of Speech, XVIII, 351—369 (1932).
162. Peterson G. E., The phonetic value of vowels. Language 27, 541—553 (1951).
163. Peterson G. E., Application of information theory to research in experimental phonetics. J. Speech and Hearing Disorders 17, 175 (1952).

164. Peterson G. E., The information bearing elements of speech, *J. Acoust. Soc. Am.* **24**, 629—637 (1952).
165. Peterson G. E., An oral communication model. *Language* **31**, 414—427 (1955).
166. Peterson G. E., Fundamental problems in speech analysis and synthesis. *Proc. of the VIII Int. Congr. of Linguists 1957*, 267—281 (Oslo, 1958).
167. Peterson G. E., Barney H. L., Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* **24**, 175—184 (1952).
168. Peterson E., Cooper F. S., Peakpicker: A band-width compression device. *J. Acoust. Soc. Am.* **29**, 777 (1957).
169. Pierce J. R., David Jr. E. E., *Man's world of sound* (New York, 1958).
170. Pollack I., The information of elementary auditory displays. *J. Acoust. Soc. Am.* **24**, 745—749 (1952).
171. Pollack I., The information of elementary auditory displays, II. *J. Acoust. Soc. Am.* **25**, 765—769 (1953).
172. Pollack I., Ficks L., Information of elementary multidimensional auditory displays. *J. Acoust. Soc. Am.* **26**, 155—158 (1954).
173. Polland B., Hála B., *Les radiographies de l'articulation des sons tchèques* (Prague, 1926).
174. Potter R. K., Kopp A. G., Green H. C., *Visible Speech* (New York, 1947).
175. Potter R. K., Steinberg J. C., Toward the specification of speech. *J. Acoust. Soc. Am.* **22**, 807—820 (1950).
176. Raleigh Lord, *Theory of Sound* (London, 1896).
177. Rösler G., Über die Vibrationsempfindung. *Literaturdurchsicht und Untersuchungen im Tonfrequenzbereich. Z. f. exper. u. angew. Psych.* **4**, 549—602 (1957).
178. Russel G. O., *The Vowel* (Columbus, 1928).
179. Schatz C. D., The role of context in the perception of stops. *Language* **30**, 47—56 (1954).
180. Schlichting H., *Grenzschicht-Theorie* (Karlsruhe, 1951).
181. Shannon C. E., Prediction and entropy of printed English. *Bell System Techn. J.* **30**, 50—64 (1951).
182. Shannon C. E., Weaver W., *The Mathematical Theory of Communication* (Urbana, 1949).
183. Smith S., Analysis of vowel sounds by ear. *Arch. Neerland. Phon. Exp.* **XX**, 78—96 (1947).
184. Smith S., Vocalization and added nasal resonance. *Folia Phoniatica* **3**, 165—169 (1951).
185. Smith S., Remarks on the physiology of the vibrations of the vocal cords. *Folia Phoniatica* **6**, 166—178 (1954).
186. Snow W. B., Rectification in the sound level meter. *J. Acoust. Soc. Am.* **29**, 1338 (1957).
187. Sovijärvi A., *Die gehaltenen, geflüsterten und gesungenen Vokale und Nasale der finnischen Sprache* (Helsinki, 1938).
188. Sovijärvi A., Die wechselnden und festen Formanten der Vokale erklärt durch Spektrogramme und Röntgenogramme der finnischen Vokale. *Proc. of the III Int. Congr. of Phonetic. Sciences*, Ghent, 1938.
189. Steinberg J. C., Application of sound measuring instruments to the study of phonetic problems. *J. Acoust. Soc. Am.* **6**, 16—24 (1934).
190. Stetson R. H., *Motor Phonetics* (Amsterdam, 1951).
191. Stevens K. N., Stop consonants. *MIT Acoustics Lab. Quarterly Rep.*, Oct. — Dec., 7—8 (1956).
192. Stevens K. N., Research on speech synthesis. *MIT Acoustics Lab. Scientific Rep.*, No. 17 (AFCRC-TN-58-140) (1958).
193. Stevens K. N., Toward a model for speech recognition. *J. Acoust. Soc. Am.* **32**, 47—55 (1960).
194. Stevens K. N., House A. S., Development of a quantitative description of vowel articulation. *J. Acoust. Soc. Am.* **27**, 484—493 (1955).
195. Stevens K. N., House A. S., Studies of formant transitions using a vocal tract analog. *J. Acoust. Soc. Am.* **28**, 578—585 (1956).
196. Stevens K. N., Kasowski S., Fant C. G. M., An electrical analog of the vocal tract. *J. Acoust. Soc. Am.* **25**, 734—742 (1953).
197. Stevens S. S., Calculation of the loudness of complex noise. *J. Acoust. Soc. Am.* **28**, 807—832 (1956).
198. Stevens S. S., Davis H., *Hearing* (New York, 1938; 1947).
199. Stevens S. S., Volkman J., The relation of pitch to frequency. *Amer. J. Psychology* **53**, 329—353 (1940).

200. Sund H., A sound spectrometer for speech analysis. Transactions of the RIT, No. 112 (1957).
 201. Tarnóczy T., Resonance data concerning nasals, laterals and trills. Word 4, 71—77 (1948).
 202. Trendelenburg F., Einführung in die Akustik, Zweite Auflage, Die menschliche Stimme, 138—150; Physikalische Eigenschaften natürlicher Schallvorgänge, 359—362 (Berlin, 1950).
 203. Truby H. M., A note on visible and indivisible speech. Proc. of the VIII Congr. of Linguists 1957, 393—400 (Oslo, 1958).
 204. Truby H. M., Acoustico-cineradiographic analysis considerations with especial reference to certain consonantal complexes. Acta Radiologica, Suppl. 182, 1—227 (1959).
 205. Tuller W. G., Theoretical limits on the rate of transmission of information. Proc. I.R.E. 37, 468 (1949).
 206. Vilbig F., Haase K. H., Some systems for speech-band compression. J. Acoust. Soc. Am. 28, 573—577 (1956).
 207. Weibel E. S., Vowel synthesis by means of resonant circuits. J. Acoust. Soc. Am. 22, 858—865 (1955).
 208. Westervelt P. J., Sieck P. W., The correlation of nonlinear flow and differential resistance for sharp-edged circular orifices. MIT Acoustics Lab. Quarterly Progress Rep., Apr. — June, 24—28 (1950).
 209. Witting C., Physical and functional aspects of speech sounds with special application to standard Swedish. Uppsala Universitets Årsskrift 1959:7, 1—151 (1959).
-