

**А.В.ВЕНЦОВ, В.Б.КАСЕВИЧ**

# **ПРОБЛЕМЫ ВОСПРИЯТИЯ РЕЧИ**



**ИЗДАТЕЛЬСТВО С.-ПЕТЕРБУРГСКОГО УНИВЕРСИТЕТА**

**С.-ПЕТЕРБУРГ**

**1994**

© Издательство С.-Петербургского  
университета, 1994  
© А.В.Венцов, В.Б.Касевич, 1994

Восприятие звучащей речи, которому посвящена настоящая монография, представляет собой бурно развивающуюся область исследований, объединяющую лингвистов, психологов, физиологов, специалистов по компьютерным системам, по искусственному интеллекту и др. Между тем в отечественной литературе за последние 10–15 лет появились, кажется, всего две монографические публикации, специально посвященные проблемам, связанным с восприятием речи (Галунов, Родионов, 1988; Штерн, 1992). Можно отметить, что и в целом после значительного подъема, которому мы обязаны фундаментальными результатами в данной сфере (Чистович и др., 1965; Чистович и др., 1976 и многие другие работы), в дальнейшем наметился продолжающийся и поныне определенный спад. Такого рода ситуацию едва ли можно считать естественной, учитывая огромную теоретическую и практическую значимость проблемы.

Исследование восприятия звучащей речи – это изучение того, какие операции производит человек, чтобы от акустического речевого сигнала перейти к некоторому символическому представлению (подробнее см. гл. II). Хорошо известно, как широко дебатировались проблемы восприятия в общей психологии. Различные школы по-разному решают вопрос о том, каким образом стимул, обладающий определенной физической природой, отражается посредством внутреннего ментального представления (в свою очередь имеющего физический субстрат – определенное состояние нервной системы человека). В нашем случае сложность задачи существенно возрастает. Процесс интерпретации речевого сигнала с неизбежностью опосредуется культурной – в широком смысле – системой координат, присущей человеку.<sup>1</sup> Достаточно упомянуть, что физи-

---

<sup>1</sup> Для человеческого восприятия в целом характерна опосредованность культурой, но в случае языка – основного носителя культуры – эта опосредованность проявляется наиболее ярко.

чески один и тот же стимул может получить разное истолкование и соответственно представление в зависимости от языка, носителем которого является субъект восприятия. Коль скоро владение языком – ярчайшее отличие человека, выделяющее его из среды всех живых существ, то и восприятие речи, будучи наиболее «человеческим» из всех видов восприятия, принадлежит к самым интересным и сложным аспектам анализа человеческой деятельности.

Несомненна и практическая важность исследования восприятия речи. Назовем лишь некоторые области потенциального применения результатов перцептивных речевых исследований. Прежде всего это широко признанная и разрабатываемая в мощных исследовательских центрах США, Западной Европы, Японии, отчасти и СНГ область автоматического распознавания речи. Автоматическое распознавание речи с помощью компьютерных устройств позволило бы решить такие проблемы, как речевой ввод данных в компьютер, печать «с голоса» («автоматическая пишущая машинка»), речевое управление машинами и механизмами и т.п. Стоит подчеркнуть, что именно учет результатов анализа естественного «человеческого» восприятия речи, указывая на наиболее обещающие стратегии автоматического распознавания, одновременно демонстрирует принципиальные ограничения, налагаемые на любую специализированную систему такого рода (подробнее см. об этом в гл. I и II).

Другая область, развитию которой должен способствовать анализ процессов речевосприятия, – это изучение патологии речи и слуха. Существует целый ряд гипотез, теорий, по-разному объясняющих причины и механизмы таких распространенных последствий травм, сосудистых и опухолевых поражений, как афазия, агнозия и пр. Имея в своем распоряжении действующую модель восприятия речи, исследователь может целенаправленно вводить в нее те или иные «поломки», соответствующие принятым в логопатологии концепциям, и сравнивать изменившиеся параметры функционирования модели с тем, что предсказывается данной концепцией. Тем самым модель восприятия речи может служить эффективным средством верификации концепций речевых расстройств.

Не менее полезными такие модели могут оказаться и при диагностике и исследовании поражений периферической части слуховой системы (нейросенсорной тугоухости и т.п.), а также при разработке методов и технических средств их реабилитации,

ибо обязательным элементом моделей восприятия речи должна быть функциональная модель периферического слухового анализатора человека.

Наконец, работа в области искусственного интеллекта тоже немыслима без моделирования восприятия речи и, шире, речевой деятельности как таковой (а проблема восприятия речи, в свою очередь, не может быть решена в полном объеме вне рамок искусственного интеллекта).

К сожалению, не обладая экономическими, организационными и техническими возможностями, которые имеются в распоряжении специальных научно-исследовательских центров, мы не могли претендовать на создание действующей модели восприятия речи. Вместе с тем все исследования, отраженные в настоящей монографии, проводились «с прицелом» на разработку такой модели. Глава II монографии посвящена описанию общих принципов, на которых, по убеждению авторов, должна основываться модель, имитирующая процессы восприятия речи. В остальных главах книги представлены результаты частных исследований, призванные наполнить конкретным содержанием те или иные блоки, модули общего «каркаса», описанного в гл. II.

Хотелось бы отметить условия, которые нам казались необходимым соблюдать при работе над данной темой. Коль скоро нас интересует адекватное воспроизведение соответствующей разновидности человеческой деятельности, мы не имеем права игнорировать какие бы то ни было звенья, аспекты, процедуры и операции при переходе от акустического сигнала к символьному представлению заданного вида. Например, исследователь волен отстаивать концепцию, в соответствии с которой план выражения воспринимаемых речевых единиц описывается в терминах артикуляторных жестов, акустические же характеристики не играют в этом никакой роли (см. об этом в гл. I). Однако, отражая модельными средствами реальный процесс речевосприятия, мы не вправе «опускать» ту стадию, когда в распоряжении слушающего еще нет другого источника информации, кроме самого акустического сигнала (плюс, естественно, знание языка, фоновые знания и т.п.).

Хотя данный конкретный исследователь или исследовательский коллектив, конечно, может избирательно концентрировать свои усилия в той или иной подобласти, отвечающей одному из фрагментов общей модели речевосприятия, сам процесс исследования должен направляться некоторой гипотезой

об общем строении модели. Последняя может быть реализована лишь компьютерными средствами. Существуют два принципиальных источника ее верификации: поведение носителей языка, прежде всего в экспериментальных, т.е. контролируемых, условиях, и воспроизведение процесса речевосприятия с использованием разработанной компьютерной модели «вместо» человека.

Еще одно важное обстоятельство заключается в том, что необходимо учитывать различие перцептивных стратегий в зависимости от типа языка. В существующей литературе обычен подход, когда исследователи стремятся обнаружить некоторые универсальные свойства модели, присущие «абстрактному» носителю языка. Но в действительности, кроме универсальных стратегий, важное место должно принадлежать стратегиям, обусловленным характеристиками используемого языка. Например, если мы утверждаем, что важным перцептивным «ключом» является тип ритмической структуры, определяемый прежде всего словесным ударением, то нельзя не учитывать, что достаточно широко распространены языки, лишенные словесного ударения как особого фонологического средства: для таких языков необходимо искать какие-то другие «ключи». Поэтому в нашей работе представлены результаты экспериментов, выполненных не только на материале русского языка, но и языков иной типологии: агглютинативных сингармонистических (киргизский, бурятский), тональных (вьетнамский, китайский) и др.

Зависимость перцептивных стратегий от типа языка есть, в сущности, частный случай самообучаемости перцептивной системы: такие, а не иные стратегии используются потому, что человек «отрабатывал» соответствующие системы в раннем онтогенезе в данной языковой среде. Эти рано сформировавшиеся установки выступают относительно жесткими. Однако одновременно, и тоже в раннем онтогенезе, человек (ребенок) не может не формировать и более гибкие стратегии, которые позволяют ему подстраиваться под индивидуальные или групповые характеристики говорящих: позволяют адекватно интерпретировать высказывания, принадлежащие мужчинам, женщинам, детям со всем диапазоном их индивидуальных различий. Нельзя исключить и того, что эти последние стратегии являются генетически заданными (врожденными), и именно таким образом обеспечивается сама возможность самообучения. Данный аспект непосредственно не отражен в наших материалах, но сам прин-

цип самообучаемости системы мы стремились не упустить из виду (ср. некоторые замечания в гл. I, с. 30 и сл., а также гл. IV и «Заключение»).

Условия успешности моделирования речевосприятия, бегло очерченные выше, достаточно разнородны и, что важно, отражают междисциплинарный характер рассматриваемой области исследования. Невыполнение такого рода условий во многих конкретных работах нередко объясняется тем, что, скажем, лингвист не учитывает должным образом тех аспектов, которые связаны с работой слухового анализатора человека, а физиолог (или психолог) не уделяет необходимого внимания языковым закономерностям высоких уровней. В своей работе мы старались не допустить «дискриминации» разнородных составляющих перцептивного процесса, хотя, разумеется, не могли анализировать все составляющие с равной степенью детальности. Многое, и притом существенное, осталось почти не затронутым; в частности, мы смогли лишь бегло коснуться важнейшей проблемы использования интонационных характеристик как «ключа» для перцептивной интерпретации высказываний (Hart e. a, 1990).

## Современные модели восприятия речи: критический обзор

В настоящее время известно около 15 моделей восприятия речи, которые отличаются как общей «идеологией» – принципами, заложенными в основу соответствующих разработок, так и детальностью, богатством (или бедностью) привлекаемого экспериментального материала, степенью ориентированности на реальные процессы и стратегии, присущие воспринимающему естественную речь человеку, наконец, уровнем формализованности и пригодности для воплощения в виде компьютерных программ или специализированных устройств.

Относительно последних двух пунктов стоит сказать несколько подробнее. На первый взгляд, степень гуманитарной адекватности и степень компьютерной адекватности – как можно условно обозначить отмеченные два свойства моделей восприятия речи – противоречат друг другу: в первом случае исследователь стремится максимально воспроизвести в модели все реальные процессы, которые обеспечивают восприятие речи человеком, вне зависимости от возможности/невозможности, легкости/трудности реализации данной модели в виде компьютерной программы; во втором случае основным требованием к модели выступает именно ее компьютерная реализуемость – при условии, разумеется, что компьютерная модель способна адекватно сопоставлять естественному звучащему тексту запись заданного вида. При ориентации на компьютерную адекватность как основную цель моделирования исследователи нередко отклоняются тем или иным образом от требования включать в модельное представление те и только те процедуры, которыми, как можно думать, реально пользуется человек.

В известной степени противопоставление, очерченное выше, реально, ср. (Касевич, 1977; 1986). Но только лишь в известной степени. Дело в том, что компьютерную реализуемость модели не следует считать неким «внешним» условием или же аспектом чисто прикладного характера. Модель интересующего

нас вида не может не быть действующей, ибо динамическая природа – неотъемлемое свойство ее естественного прототипа. Поскольку о механической действующей модели восприятия речи по понятным причинам говорить не приходится, единственный возможный вариант действующей модели с необходимостью оказывается компьютерным.

Из сказанного следует, что именно компьютерное моделирование есть тот инструмент, который сущностно адекватен задаче: исследованию реальных процессов восприятия речи. Выше мы говорили о «процедурах, которыми, как можно думать, реально пользуется человек» при восприятии речи. На существование таких процедур нам обычно «намекают» экспериментальные данные, именно они позволяют нам думать, что человек прибегает к соответствующим стратегиям. Но лишь встроив формализованные аналоги этих стратегий в (компьютерную) модель, мы получаем возможность проверить, как они функционируют и функционируют ли вообще в процессе достижения заданного результата. Тем самым без модельного «проигрывания» разных вариантов решения полное описание перцептивного процесса оказывается невозможным.

Иначе говоря, критерии гуманитарной и компьютерной адекватности выступают в качестве взаимно дополнительных. Первый требует, чтобы исследователь исходил из некоторых общих представлений о том, как может функционировать человек в решении перцептивных задач, а также из конкретных экспериментальных данных, релевантных для данной области. Второй дает в руки исследователя незаменимый способ объединения различных данных в целостную действующую систему и проверки их на функциональность, совместимость и эффективность в достижении заданного результата.

Ниже мы кратко рассмотрим основные из представленных в современной литературе моделей восприятия речи.<sup>1</sup>

**Неомоторная модель и теория прямого восприятия.** Специалистам в области восприятия речи хорошо известна так называемая моторная теория восприятия речи, вызвавшая в свое время оживленную дискуссию (Чистович и др., 1965; Галунов и

---

<sup>1</sup> При написании настоящего раздела были использованы, кроме конкретных работ соответствующих авторов, материалы обзоров Д.Клатта (Klatt, 1989), С.Голдингера и др. (Goldinger e. a., 1990), Дж.Альтманна (Altmann, 1990) и К.Фаулера (Fowler, 1991).

др., 1967; Liberman е. а., 1967 и многие другие работы). Примерно через 20 лет после публикации первых работ, излагавших принципы этой теории, появилась ее новая версия (Liberman, Mattingly, 1986), которую мы и назвали здесь неомоторной моделью (теорией) восприятия речи.

Согласно неомоторной теории, восприятие речи осуществляется в терминах артикуляторных жестов, ассоциированных с акустическими сегментами. Предполагается, что соответствующие системы человека включают специализированный модуль, структура и основные правила функционирования которого предопределены генетически. Этот модуль обладает способностью «видеть» артикуляторные жесты «за» акустическими характеристиками фонетических сегментов. Существенно, что процедуры, используемые данным модулем, позволяют устанавливать не столько реальные артикуляции, вызвавшие конкретное звучание, сколько типовые артикуляторные жесты, или артикуляторные намерения, стандартно ассоциированные с акустическими характеристиками того или иного типа (ср. известное понятие «звукового намерения» у Бодуэна де Куртенэ). Высказывается также предположение о том, что артикуляторные жесты, реализующиеся как движения губ, языка и его подструктур, мягкого неба, гортани с голосовыми связками, составляют систему врожденных естественных категорий, в равной степени обслуживающих порождение и восприятие речи и соотносимых с традиционными дифференциальными (различительными) признаками, трактуемыми универсалистски (Browman, Goldstein, 1985). Именно эти жесты и воспринимаются непосредственно.

Здесь неомоторная теория сближается с воззрениями школы так называемого прямого восприятия (Гибсон, 1988). Дж. Гибсон и его последователи настаивают на том, что восприятие призвано прежде всего обнаруживать события реальной действительности. В частности, при зрительном восприятии зрительный анализатор человека непосредственно взаимодействует с ближайшей средой – световым потоком, модулированным в силу его взаимодействия с теми или иными предметами, но воспринимает человек не эту оптическую среду, а те дистальные события, которые определяют ее свойства, т.е. предметы и их взаимодействие. Ближайшая среда оказывается «прозрачной» для восприятия. Понимаемое таким образом восприятие и называют прямым. Говоря об общих принципах теории, К.Фаулер подчеркивает, что прямой характер восприятия предполагает

отсутствие посредующих когнитивных процессов, которые заключались бы в проверке гипотез, выводе следствий: эти процессы могут быть источником ошибок (Fowler, 1986, p. 4); см. также (Fowler, 1990a; 1990b) и другие работы.

Перенося эти представления на восприятие речи, К. Фаулер и др. исходят из того, что применительно к речи дистальными событиями, подлежащими прямому восприятию, следует считать артикуляторные события, происходящие в голосовом тракте говорящего. Что же касается акустических характеристик речевого сигнала, то они выполняют роль непосредственной, ближайшей среды, которая несет на себе отпечаток сформировавших ее артикуляторных событий. Акустическая среда «приобретает структуру благодаря динамическим изменениям в голосовом тракте, и она способна передавать эту структуру слуховой перцептивной системе, тем самым сообщая соответствующие информационные характеристики слушающему, который обладает к ним восприимчивостью» (Fowler, 1986, p. 6). Акустическая среда с этой точки зрения столь же «прозрачна» для слухового восприятия, сколь оптическая среда — для восприятия зрительно-го.

И авторы неомоторной теории, и последователи теории прямого восприятия видят главное различие между двумя подходами в том, что для первых реальна специальная обработка акустического сигнала, которая «принимает во внимание как анатомо-физиологические, так и фонетические ограничения на работу артикуляторов» (Fowler, 1986, p. 7), в то время как для вторых акустические характеристики вообще не являются самостоятельным объектом анализа, а только лишь носителем информации о породивших их артикуляторных событиях (см. также ниже).

По мнению Д. Клатта (Klatt, 1989, p. 178), неомоторная теория предлагает «обладающую привлекательностью философию», однако еще весьма далеко до сколько-нибудь разработанной модели восприятия речи, основанной на соответствующих представлениях.

Мы добавили бы к этому, что и сама по себе «философия» неомоторной теории и теории прямого восприятия едва ли безупречна.

Вызывает сомнение исходная посылка, заключающаяся в том, что перцептивная активность воспринимающего речь человека направлена на выяснение структуры событий, послуживших источником речевого звука. По-видимому, здесь имеет место

более или менее механический перенос на речь действительно характерных для человека (и других живых существ) принципов реакции на неречевую звуковую информацию. Когда человек имеет дело с неречевыми звуками (точнее, со звуками, которые не являются экспонентами знаков), то его, конечно, интересует не акустическая структура этих звуков, а та информация о звучащем предмете, которую они содержат. Когда мы слышим, например, звук шагов, то не собственные его характеристики для нас важны, а именно то, что это – звук шагов (не случайно мы и говорим – в полном соответствии с теорией прямого восприятия! – *слышать шаги*). Иначе говоря, релевантна всегда информация «второго порядка»: не о самом звуке, а о той «внезвуковой» и «звучковой» информации, которую из него можно извлечь.<sup>2</sup>

Для речевого звука и, шире, любого звука, формирующего экспонент знака, ситуация иная. Число «порядков», надстраивающихся в этом случае над характеристиками звука как такового, многократно возрастает: в типичном случае человека интересует лишь «поток смыслов», моделирующих внеязыковую действительность, а все остальные структуры – в случае естественного языка звуковая, морфологическая и синтаксическая как минимум – должны быть в принципе столь же «прозрачны» для речевого восприятия, сколь та же оптическая среда при зрительном восприятии предметов, ср. (Ohala, 1986; Morish, 1990). В сущности, «прозрачен» для восприятия и сам текст: обычно слушающего интересует не текст как таковой, а описываемая им ситуация. В то же время собственные признаки всех структур звучащего текста могут стать объектом самостоятельного анализа в силу того, что от типа структуры зависит смысловая интерпретация. Звуковые же структуры – по крайней мере какие-то их аспекты – должны анализироваться, поскольку невозможно миновать акустическую информацию в качестве первичного источника.

К.Фаулер не учитывает различия между природными и культурными «модусами» восприятия, ср. (Nearey, 1990). Будет ли Фаулер утверждать, что человек, воспринимающий музыку, извлекает из «звуковой материи» информацию о сложных жестах, скажем, пианиста (или о колебательных движениях струн

---

<sup>2</sup> В то же время нельзя не учитывать, что и в подобных случаях невозможно опустить стадию обработки акустического сигнала с его признаками как такового.

рояля)? Процессы культуры приводят к тому, что поведенчески самостоятельную ценность для человека приобретают и сами по себе признаки среды, которые в других, докультурных и внекультурных отношениях носили бы преимущественно служебный характер.

**Модель анализа через синтез.** Ранняя версия модели анализа через синтез, которая появилась в конце 50-х – начале 60-х годов, была не чем иным, как моторной теорией восприятия речи: анализ речевого сигнала, согласно соответствующим представлениям, осуществлялся путем восстановления, или синтеза, породивших этот сигнал артикуляторных характеристик (Bell e. a., 1961; Stevens, Halle, 1964). Предполагалось, в частности, что тем самым преодолевалась трудность, связанная с высокой степенью вариабельности акустических параметров речевого сигнала; иначе говоря, принималось, что артикуляторные параметры звуков речи должны быть более стабильными, константными, нежели акустические. Впоследствии это предположение не подтвердилось.

Более поздние версии модели анализа через синтез уже не ограничивались оперированием информацией, относящейся исключительно к фонетическому аспекту речи. На первый план вышли процедуры, связанные с выдвижением гипотез относительно воспринимаемых слов с последующей верификацией этих гипотез. Синтезу в этой версии подлежали, таким образом, уже не фонемы или фонемоподобные единицы, а целостные слова, анализ же главным образом заключался в проверке того, насколько объективные характеристики речевого сигнала соответствуют выдвинутой гипотезе. И здесь одним из основных доводов выступали трудности, связанные с высокой степенью вариабельности речевого сигнала, с типичностью редукции, вплоть до нулевой, и т. п. Как пишет Д.Клатт (Klatt, 1989, p. 181), эта «модель анализа через синтез ... подверглась критике с двух точек зрения: никто не предложил удовлетворительного объяснения тому, каким образом процесс начинается (т.е. откуда появляются гипотезы относительно слов); такая модель может потребовать слишком большого объема когнитивной обработки [данных] – принятия чересчур большого числа последовательных решений в единицу времени...»

Дальнейшее развитие представления, связанные с анализом через синтез, получили в работах В.Зью, Д.Хаттенлохера и др. (Huttenlocher, Zue, 1984; Zue, 1986). Важнейшим новшест-

вом стало положение о гиперпризнаках (robust features) типа «сильный фрикативный», «слабый фрикативный», «передний гласный», «смычный», «ударный», «безударный». В типичном случае одного такого признака достаточно для характеристики того или иного речевого сегмента, сама же модель анализа через синтез предполагает три основные стадии: (1) предварительная сегментация акустического сигнала и приписывание сегментам гиперпризнаков; (2) обращение к словарю для получения некоторого множества слов (как считается, не слишком большого), которые удовлетворяют данному набору гиперпризнаков; (3) поиск слова внутри установленного множества, что осуществляется за счет более детального фонетического анализа его элементов, опять-таки с использованием принципа анализа через синтез.

Хотя Зью и его коллеги исходили из убеждения, что вводимые ими гиперпризнаки в значительной степени позволяют преодолеть вариабельность, неопределенность речевого сигнала благодаря своей устойчивости по отношению к контексту и относительной независимости от конкретного диктора, попытки приложения этих идей к построению реальных моделей распознавания речи оказались не слишком успешными (Huttenlocher, 1986). Гласс и Зью предложили еще одну модификацию подхода, когда гиперпризнаки – оценка изменений в слуховом спектре (т.е. в спектральной картине на выходе слуховой модели) – используются и для сегментации, а затем вычлененным сегментам присваиваются характеристики в терминах 10 спектральных эталонов; далее полученные словарные кандидаты сопоставляются с гипотезами, основанными на языковых лексических вероятностях (Glass, Zue, 1987). Утверждается, что данная модификация приводит к улучшению распознавания, хотя количество ошибок и сохраняющаяся степень неопределенности все еще велики.

Клатт высказывает сомнение в потенциальной эффективности этой модели. Гиперпризнаки дают слишком большой разброс возможных ответов как для сегментации, так и для идентификации слов. При оперировании соответствующими правилами на материале изолированных слов множество слов, описанных в терминах гиперпризнаков, в среднем составляет 21 единицу для словаря объемом 20 тыс. слов. Но при переходе к связному звучащему тексту с фонетической неопределенностью его границ (особенно при учете внешнего сандхи) и высокой вариабельностью фонетического облика слова мощ-

ность множества слов, выделенных по гиперпризнакам, может составить миллионы элементов (Harrington, Johnstone, 1987).<sup>3</sup> По мнению Клатта, даже введение узкой фонетической транскрипции, учитывающей редукцию, тип реализации звуков внутри слога (и тем самым позволяющей извлекать информацию о словесных границах), едва ли позволило бы существенно снизить неопределенность, которую сопровождает использование гиперпризнаков (Klatt, 1989, p. 182).

### **Модели, использующие дифференциальные признаки.**

Модели, подобные описанным выше, в сущности разрабатывались как альтернативные по отношению к тем представлениям, которые опирались на традиционные лингвистические концепции с их понятием дифференциального признака фонемы. Существуют и модели, непосредственно исходящие из лингвистического понятия дифференциального признака и пытающиеся «буквально» реализовать теоретические предложения, содержащиеся в трудах Р.Якобсона и его соавторов, или более поздние версии Хомского, Халле и др., ср., например, (Pisoni, Luce, 1987).

Надо сказать, однако, что соответствующие модели – по вполне понятным причинам – уделяют значительно больше внимания процедурам обработки акустического сигнала. Если собственно лингвистические описания либо вообще плохо отличают содержание дифференциального признака от акустических или акустико–артикулярных параметров, либо неопределенно говорят об акустических (акустико–артикулярных) коррелятах дифференциального признака, то в модельных представлениях фигурируют стадии, ответственные за ступенчатый переход от «сырого» акустического материала к некоторой символической записи.

Первая такая стадия – это обычно этап слуховой обработки, или периферического анализа. По мнению Д.Клатта, результатом периферического анализа должны быть, как минимум, два вида представления: динамическая спектрограмма в координатах «время – частота – усредненная частота разрядов нейрона» (an average-firing-rate representation of the short-time

---

<sup>3</sup> Как можно понять, в последнем случае имеются в виду наборы не слов, а словосочетаний: одна и та же фонетическая цепочка членится на те или иные последовательности слов. В противном случае утверждение о миллионах слов-вариантов – что превышает объем обычного словаря – становится непонятным.

spectrum) и динамический синхронный спектр (synchrony spectrum), получаемый за счет синхронной импульсации «нейронов» в некоторых частотных каналах анализатора. Предполагается, что использование в модели стадии периферического описания облегчит задачу последующих модулей, поскольку в результате инвариантные акустические характеристики фонетических признаков предстанут в «очищенном» виде, а функционально иррелевантная вариабельность будет снята (Klatt, 1989, p. 183).

Следующая стадия – использование детекторов акустических параметров. Детекторы оценивают не абсолютные, а относительные признаки акустического сигнала, которые, как предполагается, характеризуются большей инвариантностью применительно как к разным контекстам, так и к индивидуальным дикторам. Детекторы акустических параметров мыслятся, вероятно, как врожденные компоненты соответствующих механизмов человека. Они определяют сравнительно простые параметры слухового представления сигнала независимо от того, речевой это сигнал или неречевой; примерами служат детекторы изменений в сигнале, детекторы спектральных переходов, спектральных пиков, формантных частот, движения формант во времени, детекторы присутствия голоса, выделения частоты основного тона, форманты назализации.

Представление речевого сигнала, выработанное на стадии периферического анализа, является объектом анализа модуля детекторов дифференциальных признаков. Некоторые авторы – см., например (Stevens, 1986), склонны и дифференциальные признаки считать универсальными, а соответствующие детекторы, вероятно, врожденными. Однако Клатт предпочитает подход, согласно которому дифференциальные признаки носят внутриязыковой характер (что, конечно, гораздо лучше согласуется с богатейшими данными фонологических исследований). Детектор дифференциальных признаков имеет дело с набором значений, являющихся результатом работы детекторов акустических параметров; в свою очередь результат работы детектора дифференциальных признаков – это решение о том, какому дифференциальному признаку отвечает данный акустический параметр или их набор.

Предполагается, что ответ детектора носит в принципе двоичный характер: фиксируется наличие либо отсутствие признака в данный момент времени. Однако допускаются и вероятностные решения, когда информация, полученная от детекторов акустических параметров, позволяет следующему модулю оце-

нить лишь вероятность присутствия тех или иных дифференциальных признаков в соответствующей точке временного континуума. При работе со звучащим текстом вероятностные оценки гораздо реалистичнее; уместность их использования определяется также и тем обстоятельством, что окончательные решения на ранней стадии распознавания речи затрудняют процедуры последующей коррекции, использование которых практически неизбежно в типичной ситуации неопределенности акустического сигнала.

В моделях описываемого класса работа детекторов обоих типов осуществляется параллельно, поэтому следующий этап анализа – это линеаризация<sup>4</sup> фонологического представления, которое должно принять вид последовательности экспонентов словарных единиц. Д.Клатт, говоря об этом этапе, допускает разные возможности, отчасти вслед за авторами, разрабатывающими соответствующие модели. Наиболее экономным он признает использование последовательности признаковых матриц, которые могут образовывать решетки в случае неоднозначности решений, предлагаемых детекторами. Некоторые авторы при линеаризации фонологического представления используют информацию о слогеделении, а также о внутренней структуре слога, но не вполне ясно, как именно соответствующие структуры находят свое место в модели, работающей с реальным речевым сигналом. Наконец, вслед за теориями типа ауто-сегментной фонологии допускается вариант представления, когда разные дифференциальные признаки относятся к разным уровням, соотносящимся по специальным правилам, а единая (единственная) цепочка дискретных линейных единиц, строго упорядоченных во времени, просто отсутствует, ср. (Stevens, 1986).

В любом случае признаковые матрицы, присутствующие в том или ином виде во всех подтипах обсуждаемой модели, характеризуются вероятностной природой; применительно к задачам линеаризации такая оценка отражает вероятность присутствия данного сегмента (обычно фонемы) в данной по-

---

<sup>4</sup> Здесь и далее термины «линеаризация», «линейный», «нелинейный» относятся к задаче получения описания речи в виде линейно упорядоченной последовательности символов (знаков), а не к линейности/нелинейности возможных преобразований акустического сигнала в слуховой системе, отражающей особенности амплитудных характеристик соответствующих элементов системы.

зиции. Обычный пример – преобразование последовательности типа [æ̃t], установленной по показаниям работы детекторов, анализирующих английскую речь, в фонологическую цепочку /ænt/: обнаруженная в сигнале назализация гласного перекодируется как информация о наличии носового согласного, степень же назализации находит свое отражение в приписывании вероятности присутствия согласного.

К сожалению, сколько-нибудь детальных сведений относительно этой очень важной проблемы – как перейти от нелинейного признакового пространства к линейным последовательностям фонологических единиц – в описаниях не содержится. Не говорится и о том, используется ли в процедурах линейаризации информация о соотношенности формантных значений, спектральных срезов с осью времени.

Следующий этап в работе модели – обеспечение доступа к словарю: информация, полученная на выходе модуля, ответственного за линейаризацию фонологического представления, используется для поиска в пространстве словаря таких лексических единиц, которые наилучшим образом соответствовали бы полученным последовательностям признаковых матриц. Предполагается, что одновременно может использоваться информация о просодических характеристиках, а также о семантических и синтаксических ожиданиях (заметим, что источник этой информации в моделях обсуждаемого типа остается неясным, а роль ее – скорее вспомогательной, ср. с нашим подходом в гл. II). Механизм действия модели на этой стадии заключается прежде всего в том, что особые процедуры должны оценивать вероятности совпадения тех или иных фонемных цепочек с экспонентами словарных единиц. Неопределенность решений в таких случаях может быть весьма значительной, особенно если учесть, что границы между словами по большей части неизвестны и, кроме того, в обрабатываемых цепочках вполне вероятны как недостающие, так и «лишние» фонемные сегменты.

Заключая обзор «признаковых» моделей, Д.Клатт пишет: «Вопросы, которые возникают в связи с полными характеристиками алгоритмов, извлекающих [из сигнала] признаки, не имеют простых ответов. Безусловно, не внушают оптимизма весьма ограниченные возможности систем распознавания, базирующихся на дифференциальных признаках, и алгоритмов их извлечения, о которых до сих пор сообщалось в литературе. Вполне возможно, что дифференциальные признаки составляют

существенный аспект восприятия [речи], но едва ли стратегии их оценки будут простыми. К тому же я не думаю, что для лексического поиска окажется оптимальной последовательность признаковов матриц, которые носят характер фонологической или широкой фонетической» записи» (Klatt, 1989, p. 188).

**Модель LAFF.** Некоторые недостатки «признаковых» моделей ставит своей целью преодолеть другая модель, которая также оперирует дифференциальными признаками, но с использованием иных стратегий. Это так называемая модель лексического доступа через признаки (Lexical Access From Features, LAFF), разрабатываемая главным образом К.Стивенсом (Stevens, 1986).

Основные особенности обсуждаемой модели заключаются в следующем. Используемые признаки трактуются как универсальные и бинарные. Каждому признаку сопоставлены некоторые акустические корреляты, поиск которых в речевом сигнале осуществляется в пределах отрезков, определяемых независимо. Так, если модель «имеет дело» со слоговым пиком, то для этого отрезка оцениваются корреляты признаков «высокий», «низкий», «задний», «огубленный», «напряженный», «назальный», «несжатогортанный (spread glottis)», «сжатогортанный (constricted glottis)»; если же вместо слогового пика представлены отрезки, характеризующиеся резкими спектральными изменениями, шумовыми составляющими, то вступают в дело другие признаки со своими коррелятами – «непрерывный», «сонорный», «согласный», «звонкий», «яркий (strident)», «корональный», «передний», «распределенный», «боковой». Поиск акустических коррелятов признаков, таким образом, заранее сориентирован на участки речевого сигнала, отвечающие неко-

Таблица 1.1. Исходная признаковая матрица для словарного описания экспонента слова *rawp*

Признаки	р	о	п
Высокий	–	–	–
Низкий	–	+	–
Задний	–	+	–
Назальный	–	–	+
Несжатогортанный	+	–	–
Сонорный	–	+	+
Звонкий	–	+	+
Яркий	–	–	+
Согласный	+	–	+
Корональный	–	–	+
Передний	+	–	+
Непрерывный	–	+	–

торым требованиям; возможны и такие участки, которые, не отвечая соответствующим критериям, вообще не анализируются (don't-care regions).

Таблица 1.2. Признаковая матрица экспонента слова *rawп*, модифицированная для облегчения сопоставления с набором признаков, полученным действием перцептивной модели

Признаки	р	о	п
Высокий		–	
Низкий		+	
Задний		+	
Назальный			+
Несжатогортанный		+	
Сонорный	–		
Звонкий	–		
Яркий			
Согласный	+		+
Корональный	–		+
Передний	+		+
Непрерывный	–		–

Признаки в принципе не группируются таким образом, чтобы в итоге получались единицы, упорядоченные во времени. Иначе говоря, этап линейаризации здесь отсутствует. Лексический поиск осуществляется посредством прямого сравнения признакововых матриц, полученных путем

обработки акустического сигнала, с аналогичными матрицами, характеризующими лексические единицы в словаре. При этом предполагается, что в самом словаре лексическим единицам (их экспонентам) сопоставлены модифицированные варианты матриц, специально предназначенные для облегчения сравнения. В таблицах показано, как отличается исходная признаковая матрица английской словарной единицы *rawп* (табл. 1.1) от ее же варианта, используемого для сравнения с ним набора признаков, полученного действием перцептивной модели (табл. 1.2) (по: Klatt, 1989).

Как можно видеть, в модифицированной «квазиперцептивной» (предположительно ориентированной на восприятие) матрице не проставлены значения признаков, которые можно считать перцептивно иррелевантными. Кроме того, данный вариант матрицы призван не только дать сам набор признаков, но и указать на то, «где» следует искать акустические корреляты этих признаков. Так, признак «несжатогортанный» указывает на придыхательность начального сегмента в слове *rawп*, но придыхание акустически реализуется после раскрытия смычки согласных и перед началом включения голосового источника при переходе к гласному, поэтому знак «плюс», отвечающий данно-

му признаку, помещается между столбцами для [p] и [ɔ]. Аналогичным образом, наличие/отсутствие звонкости, как и придыхания, оценивается лишь в контексте отрицательного значения признака «сонорность», что и объясняет расположение в матрице соответствующих знаков. Местоположение знака «плюс» для признака «назальность» между столбцами для [ɔ] и [n] связано с тем, что в звучащем тексте носовой согласный может быть реализован в виде назализации конечной фазы предшествующего гласного.

Значения признаков, принадлежащие одному и тому же столбцу матрицы, образуют особый уровень (tier), и полное описание складывается в результате взаимодействия разных уровней. Можно сказать, что тем самым отсутствие информации о временных соотношениях дискретных единиц в составе экспонента лексической единицы в известной степени компенсируется внутриматричными указаниями на положение акустических ключей (коррелятов), ответственных за те или иные признаки, относительно исходной речевой волны. (Мы воздерживаемся от обсуждения самой используемой системы признаков).

Предполагается, что поиск в словаре осуществляется путем прямого сличения набора признаков, извлеченного из акустического сигнала, с модифицированной «квазиперцептивной» матрицей, которая выступает присловной характеристикой каждой единицы в составе словаря. Используется также дополнительная информация о сравнительных весах разных признаков. Обращение к признаку, которому приписан меньший перцептивный вес, приводит к повышению «стоимости» решения при том, что в модель встроен механизм, отбирающий решения, самые низкие по «стоимости». Впрочем, детальные разработки, касающиеся такого рода механизмов, пока отсутствуют.

По мнению Д.Клатта, к сильным сторонам модели LAFF, разрабатываемой К.Стивенсом, принадлежит именно то, что он обходится без сложных процедур линеаризации фонологического представления, да и вообще по существу без разграничения представлений разного уровня абстрактности. Кроме того, в идеале процедуры, предусмотренные данной моделью, должны обеспечить ситуацию, когда детекторы признаков без промежуточных перекодировок дают непосредственно те характеристики, на основании которых осуществляется лексический поиск в словаре; детекторы в этом случае оценивают некоторые инва-

рианты отношений между параметрами, а также те или иные количественные соотношения.

В то же время модель далеко не снимает общие трудности, связанные с самим понятием признака (в частности, вопрос о взаимозависимости признаков, об использовании правил взаимодействия акустических ключей и др.). Недостатками, с точки зрения Клатта, следует считать также отказ от вероятностных оценок в пользу оценок по принципу «да/нет» или наряду с ними, равно как и общую картину, при которой оказывается, что описание звуковой стороны лексических единиц применительно к речевосприятию является менее абстрактным по сравнению с тем же описанием, ориентированным на речепорождение.

Суммируя обсуждение модели LAFF, Клатт отмечает, что Стивенс предпринял радикальный отход от традиционных представлений, связанных с использованием признаков в моделях восприятия речи. Результатом явилось радикальное упрощение перцептивных процедур, при использовании которых небольшого набора признаков достаточно для перехода от акустического речевого сигнала к словарным единицам языка. Однако существующие системы распознавания вряд ли обещают достижение этой цели в ближайшем будущем на путях, предполагаемых моделью LAFF (Klatt, 1989, p. 191–192).

Добавим, со своей стороны, что надежда на простые решения, которые обеспечивали бы эффективную идентификацию языковых единиц по небольшому набору признаков, может оказаться вполне реалистичной, если поиск осуществляется в критически ограниченном (за счет использования просодической информации, синтаксических и семантических ожиданий и т.п., см. об этом гл. II) подпространстве словаря.

Что касается гетерогенности систем представления для звукового аспекта порождения и восприятия речи, то Клатт безусловно прав, когда склоняется к более традиционным взглядам, согласно которым обе стороны речевой деятельности – речепорождение и речевосприятие – обслуживаются одной и той же системой фонологических единиц. Однако это относится именно к фонологическим системам, прежде всего к системам фонем. Если же говорить о фонологическом представлении языковых единиц (их экспонентов) и под степенью абстрактности понимать реально степень подробности, необходимой для реализации перцептивных или же моторных задач, то ситуация будет существенно иной. При восприятии речи, особенно в условиях информационно насыщенной среды, сильно сужающей

поле выбора, может оказаться достаточным и весьма примитивный набор признаков – т.е. экспоненту лексической единицы в целом достаточно будет приписать сколь угодно простую фонологическую характеристику, в этом смысле предельно абстрактную (ср., например, достаточность признака назальности для идентификации слова нет в условиях, когда поле поиска сужено до выбора между да и нет). В то же время при порождении речи описание звуковой стороны языковой единицы всегда будет максимально подробным (если не считать автоматических коррекций в силу коартикуляции и т.п., которые могут реализовываться, в терминах Бернштейна, по «принципу цепочки»): говорящий не может «опустить» какую бы то ни было характеристику звукового оформления, предписанную системой и нормой (Касевич, 1983).

Иначе говоря, уровень подробности (абстрактности?) записи словарных единиц в моделях порождения и восприятия речи действительно может различаться в зависимости от аспекта речевой деятельности – речепорождения или речевосприятия, но меньшая степень детализированности ассоциируется с восприятием, а не порождением речи (ср. выше).

**Модель LAFS.** Рассматриваемая модель лексического доступа через спектры (Lexical Access From Spectra, LAFS) принадлежит Д.Клатту (Klatt, 1980; 1989). Автор выделяет 4 положения, отличающие его модель от других систем, моделирующих восприятие речи (Klatt, 1989, p. 192193):

(1) исходным моделируемым явлением речи выступает фонетический переход; предполагается, что любой фонетический переход может быть представлен с любой степенью точности через некоторый набор спектров, образующих последовательность, или же с помощью нескольких таких наборов, являющихся альтернативными;

(2) описанные подобным образом участки речевого сигнала непосредственно сличаются со словарными единицами; иначе говоря, каждой единице ментального лексикона должен быть сопоставлен ее спектральный эталон, хранящийся в долговременной памяти;

(3) фонетическая вариабельность гласных и согласных как внутри слова, так и на словесных границах отражается в модели с помощью декодирующей сети, в которой учтены спектральные характеристики всех возможных сочетаний данного языка;

(4) в модели используются процедуры, оценивающие функцию сходства между фонетическими (спектральными) характеристиками входного сигнала и характеристиками словарных спектральных эталонов; результатом действия процедур, основанных на представлении о такого рода метрическом пространстве, выступает список словарных единиц, ранжированных по величине функции сходства относительно спектральных характеристик входного сигнала.

Первое из приведенных положений Д.Клатт иллюстрирует с помощью примера перехода от начального согласного [t] к гласному [a] в английском языке. Характеристики этого перехода аппроксимируются с помощью последовательности из 5 статических спектров, представляющих собой огибающую энергии в соответствующих критических полосах (*static critical-band spectra*). Первый из последовательности спектров отражает период паузы (т.е., артикуляторно, смычки согласного), второй – взрыва, третий – аспирации (придыхания), четвертый – момент включения голосового источника, пятый – квазистационарного участка гласного (*vowel midpoint*).

Декодирующая сеть, о которой говорится в третьем из приведенных положений, строится в терминах дифонов – участков речевой цепи от «центра» согласного до «центра» гласного или, наоборот, в слогах СГ или ГС соответственно – и переходов от одного дифона к другому. Каждый дифон с его внутренним переходом отражается в модели самостоятельным спектральным эталоном, при этом учитываются все варианты, возникающие в тех или иных контекстах. Если система «работает» с единицами, принадлежащими словарю, т.е. со знакомыми словами, то обращения к фонетическим признакам, равно как и к индивидуальным сегментным единицам, не происходит, все процедуры имеют дело лишь со спектральными характеристиками как таковыми.

Декодирующая сеть, выработанная для входного речевого сигнала, «налагается» на словарную сеть для оценки меры их близости. Словарная сеть генерируется в три этапа. Первый этап заключается в построении фонологических деревьев, которые отражают синтагматические цепочки фонем, образующие экспоненты словарных единиц. Эти деревья могут пересекаться, иметь общие части – поддеревья, поскольку слова могут совпадать (начальными) частями своих экспонентов, ср. [kən] в *control* и *convert*; см. также рис. 1.1, А.

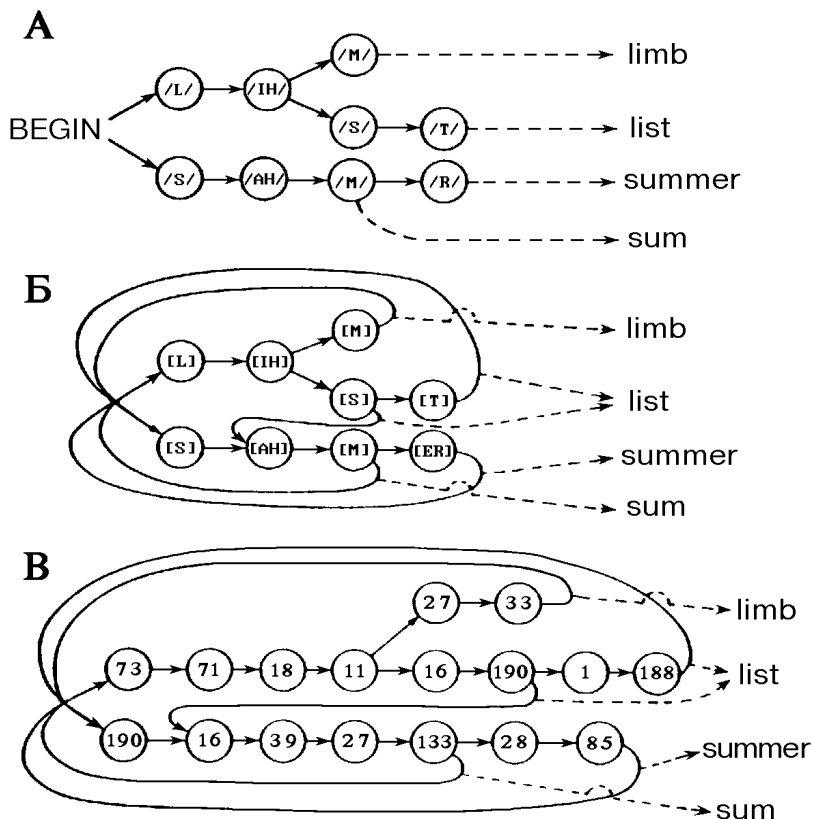


Рис. 1.1. Этапы фонетического оформления слов *limb*, *list*, *summer*, *sum* в модели LAFS (по: Klatt, 1989, р. 195).

А – представление экспонентов слов посредством фонологического дерева; Б – преобразование фонологического дерева в фонетическую сеть с учетом вариантов произнесения; В – преобразование фонетической сети в сеть спектральных эталонов (обозначенных условными номерами).

Следующий этап состоит в превращении фонемного дерева в фонетическую сеть, где узлами выступают уже не фонемы, а их фонетические корреляты с учетом всех возможных вариантов, в том числе и тех, которые появляются на словесных границах (рис. 1.1,Б).

Наконец, третий этап заключается в том, что символам фонетической (узко-фонетической) транскрипции — именно так можно интерпретировать узлы фонетических сетей — сопоставляются последовательности спектральных эталонов, взятых из универсального словаря последних, где им условно присвоены порядковые номера (рис. 1.1, В). Именно с такой сетью, образованной последовательностью спектральных эталонов, сличается в принципе аналогичная декодирующая сеть, выработанная в результате анализа входного акустического сигнала.

Сам автор, Д.Клатт, следующим образом оценивает сравнительные достоинства и недостатки собственного подхода (Klatt, 1989, p. 194–195).

Положительные стороны модели он видит в следующем:

(1) используемые процедуры не требуют предварительной сегментации, что являлось бы потенциальным источником ошибок;

(2) модель обходится без принятия решений относительно дифференциальных признаков или фонетических сегментов, что также является источником ошибок и одновременно сопряжено с утратой полезной информации о характеристиках аллофонов в различных контекстах, включая просодические;

(3) не требуется предположений, связанных с признанием инвариантности фонетических единиц; любые фонетические сегменты, их последовательности могут рассматриваться как обладающие собственной структурой;

(4) спектральные эталоны для слов могут модифицироваться в результате обучения системы, при этом перенастройка выполняется текущим образом и результаты ее не обязательно закрепляются, т.е. опасности сверхгенерализации не возникает;

(5) оценка подобия, устанавливающая меру близости данного сегмента речевого сигнала единице словаря, едина для всей сети, ее использование основывается на понятии унифицированной функции спектрального сходства, служащей для измерения воспринимаемого «расстояния» между спектрами и допускающей вероятностное выражение;

(6) модель исключает опасность принятия ранних решений, которые затрудняли бы внесение коррекций в случае необходимости, ибо уже самое первое решение принимается применительно к отождествлению со словарной единицей.

Говоря о недостатках модели, Д.Клатт отмечает, что:

(1) возможно, не удастся разработать достаточно эффективные процедуры оценки меры сходства между спектрами,

чтобы они «справлялись» со всеми типами вариабельности речевого сигнала, которые можно наблюдать в реальных фонетических исследованиях;

(2) трудности могут возникнуть в связи с задачей приведения спектров к некоторому нормальному виду для преодоления вариабельности, связанной с индивидуальными особенностями разных носителей языка;

(3) некоторые типы вариабельности, такие, как разная степень распространения назализации на гласный, соседствующий с носовым согласным, можно отразить в модели лишь путем введения семейства альтернативных спектральных эталонов, но такой путь практически невыполним, если учесть огромное число степеней свободы, присущее артикуляторам, ответственным за данный процесс;

(4) трудно представить себе процессы, с помощью которых в модели можно было бы отразить все типы фонетического взаимодействия, представленные на границах слов.

В своем обзоре Д.Клатт рассматривает некоторые возможности преодоления ограниченности собственной модели, на чем мы не будем останавливаться. Отметим лишь, что и те свойства модели LAFS, которые выделяются в качестве ее достоинств, вообще говоря, не во всех своих пунктах выглядят таковыми. Скажем, отказ от сегментации представлен как средство избежать ошибок, связанных с неверным членением акустического сигнала. Но чем операция сегментации отличается от операции идентификации, которая тоже носит с неизбежностью вероятностный характер и в этом смысле равным образом связана с риском ошибки? Можно утверждать, что «удачная» сегментация – там, где она реальна, – резко повышает шансы адекватной идентификации, которая выступает конечной целью перцептивных процедур.

В целом понятно стремление Клатта по существу свести весь перцептивный процесс к одной процедуре – сличению наблюдаемого спектра со спектром словарной единицы как целостного образования. Однако если говорить о модели восприятия речи (а не о распознающей системе, не связанной требованием воспроизвести ее естественный прототип), то положение о целостности слова, его экспонента едва ли следует трактовать столь прямолинейно. Действительно, есть основания говорить об общем «фонетическом облике» слова, см. об этом (Зиндер, Касевич, 1989). Но фонетическая целостность слова проявляется в наличии признаков, свойственных ему как таковому (акцент-

ный контур, сингармоническая модель, признаки начала, конца, середины слова), а также в возможности достраивать до целого облик слова по части его признаков. Иначе говоря, речь должна идти о стратегиях холистского восприятия в противоположность элементаристскому, а не о спектральной картине слова в целом.

Наконец, к этому надо добавить, что если допустить практически неограниченный объем памяти и неограниченное быстроедействие, то, возможно, модель типа LAFS действительно окажется в состоянии осуществлять поиск слов в словаре на основании сличения целостных спектров. Но вряд ли можно утверждать, что это будет именно модель восприятия речи человека. Можно представить себе, что естественные системы человека прибегают к генерированию спектров, отвечающих словам, для сличения с ними спектров входного сигнала: способность такого рода предполагается тем тривиальным обстоятельством, что слушающий является потенциальным говорящим, и при таком допущении LAFS предстала бы вариантом модели анализа через синтез. Ее особенность заключалась бы в холистско-фонетической природе перцептивных процедур. Но трудно признать, что в словаре хранятся спектральные эталоны слов как таковых.

Впрочем, само это положение тоже не вполне ясно. С одной стороны, в изложении Клатта содержатся утверждения о том, что модель как будто бы осуществляет сличение спектра сигнала со спектром целостного слова. С другой стороны, детали описания указывают, скорее, на то, что речь идет о спектральных характеристиках таких явлений – смычка, взрыв, аспирация и т.п., которые вполне можно истолковать в качестве коррелятов «обычных» дифференциальных признаков. В этом случае основное отличие модели LAFS от других заключается, пожалуй, в достаточно последовательном неразличении того, что традиция противопоставляет как фонологические и фонетические аспекты звуковой стороны языка и речи. Например, в описании английского слога [tə] используется признак «придыхательность» (см. выше), который для английской системы не является фонологическим, дифференциальным.<sup>5</sup>

---

<sup>5</sup> В то же время не приходится отрицать, что в определенных условиях восприятия речи придыхательность действительно может приобретать роль перцептивного сигнала глухости согласного.

В модели LAFS фонемы практически не играют какой бы то ни было роли в перцептивном процессе (хотя, надо заметить, модель содержит подсистему SCRIBER, основная задача которой – посегментный анализ незнакомых слов).

**Моделирование слуха.** Первоначально модели этого класса ограничивались лишь уровнем первичного спектрального анализа речевого сигнала. При их разработке использовались результаты психоакустических (психофизических) экспериментов, полученные методами маскировки. Очевидно, что параметры таких моделей отражали как особенности собственно периферического спектрального анализатора слуховой системы, так и алгоритмы принятия решений испытуемыми. Вряд ли целесообразно использовать подобные модели в качестве блока первичного анализа сигнала в полной функциональной модели восприятия речи: некоторые процедуры обработки сигнала окажутся учтенными в модели дважды – и в характеристиках первичного анализатора, и в правилах принятия решений.

С этой точки зрения более адекватными представляются модели, основанные на данных нейрофизиологии и учитывающие гидромеханические характеристики «улитки» внутреннего уха. Подобного рода модель в течение почти двух десятилетий разрабатывается и используется для анализа экспериментальных сигналов в лаборатории физиологии речи Института физиологии им. И.П.Павлова РАН (Чистович и др., 1986). В ней, помимо характеристик «улитки», учтены эффекты двухтонового подавления и периферической кратковременной адаптации. Последующая обработка полученного с помощью модели представления речевого сигнала включает сегментацию (структура и параметры блока выбраны на основе психоакустических данных) и выделение локальных (по частоте) параметров спектра.

Аналогичные модели используются многими зарубежными исследователями. Некоторые из них расширены за счет введения блоков, воспроизводящих дальнейшие преобразования сигнала: латеральное торможение (Shamma, 1988) и получение синхронного спектра (Sachs e. a., 1982; Seneff, 1988). В последние годы (видимо, благодаря значительному увеличению мощности персональных ЭВМ) появились модели с элементами «нейронных сетей», воспроизводящие функциональные свойства не только «улитки» и волосковых клеток, но и нейронов кохлеарных ядер (Visual Representations ..., 1993), что существенно

расширяет возможности моделей в плане выделения признаков для первичного субъективного описания сигналов.

Спектральное представление сигнала на выходе таких моделей характеризуется тем, что в области первой форманты выделяются изолированные гармоники, особенно для женских голосов с высокой частотой основного тона. При использовании традиционного описания сегментов речевого потока в терминах формантных частот это создает дополнительную проблему оценки положения первой форманты по амплитудам соответствующих гармоник. Однако результаты психоакустических исследований (Чистович, Шупляков, 1971) показывают, что при восприятии гласных за первую форманту слушатель принимает именно максимальную гармонику в соответствующей частотной области.

**Квазинейронные модели.** Первые попытки построения квазинейронных моделей (не применительно к восприятию речи) появились уже в конце 40-х годов (Hebb, 1949), интерес к ним особенно усилился начиная с 60-х годов (Rosenblatt, 1962). Авторы этих моделей стремились к своего рода бионической адекватности моделей, т.е. к воплощению в моделях именно тех принципов, которые лежат в основе работы человеческого мозга и, шире, нервной системы человека (или других живых существ, в зависимости от решаемой задачи). Исходя из того, что для мозга характерны специализированные структуры, а также не менее характерна способность к обучению – формированию новых функциональных структур, квазинейронные системы старались конструировать таким образом, чтобы использовались либо предпрограммированные для тех или иных целей простые механизмы, либо структуры, способные к самообучению по особым алгоритмам. На развитие соответствующих представлений сильно повлияли и результаты в области искусственного интеллекта, согласно которым мозг достаточно плохо приспособлен к выполнению логических операций, обладает сравнительно низким быстродействием, но в то же время функционирует весьма эффективно в решении задач, связанных с параллельной обработкой больших массивов информации, с операциями классификации и сравнения.

Первые квазинейронные модели, основанные на алгоритмах самообучения, оказались способными лишь к классификации некоторых простых структур по признакам сходства между ними (Rosenblatt, 1962). В логико-математических терминах это озна-

чает, что такого рода модели были эффективны лишь в обучении отношениям, описываемым с помощью первопорядковых предикатов.

Применительно к материалу, с которым имеет дело система восприятия речи, Д.Клатт приводит целый ряд примеров, когда логики первопорядковых предикатов недостаточно; в данном случае это означает, что одинаковые признаки могут потребовать разной интерпретации, а разные – одинаковой, если учесть контекст того или иного рода. В числе примеров, которые представлены в виде таблицы (Klatt, 1989, р. 204, табл. 3), фигурируют следующие:

- десинхронизация шумового и голосового источника, служащая признаком разграничения звонких и глухих смычных, почти одинакова для слогов [pa] и [gi]; иначе говоря, интерпретация акустического «ключа» для одного признака – звонкости – требует учета значений других признаков, в данном случае – места образования;

- коартикуляционное огубление [s] перед губными гласными в речи мужчин создает акустический эффект, который делает спектральные характеристики [s°] примерно такими же, как у [s] перед негубными гласными в речи женщин; таким образом, учитывая коартикуляционный эффект при классификации согласных, необходимо одновременно принимать во внимание еще один фактор – пол диктора;

- спектр взрыва [g] перед передними гласными близок к спектру взрыва [d] перед огубленными гласными, т.е. и здесь собственные признаки согласных при их группировке предполагают обращение к идентичности последующего гласного;

- по правилам теории «локуса» начальная частота формантного перехода от шумного согласного к гласному определяется не только типом самого согласного, но и типом последующего гласного;

- величины формант для гласных обнаруживают зависимость от длительности последних, что, в свою очередь, существенным образом связано с позицией соответствующего слова во фразе и другими факторами; иначе говоря, и в данном случае классификация, группировка гласных по их формантным величинам требует механизма, который помогал бы «отстроиться» от возмущений, вносимых влиянием синтаксической позиции, просодического выделения с целью эмфазы и т.д.

К примерам Клатта, часть из которых мы привели в силу их достаточной красноречивости, можно было бы добавить, ко-

нечно, еще многие (например, близость спектральных характеристик [а] между мягкими и [е] между твердыми в русском языке). По существу, здесь – на материале построения моделей нового типа – речь идет о старых проблемах, которые известны столь же давно, сколь существует фонология: фонетически разные сегменты могут быть тождественными фонологически (функционально) и наоборот.

В квазинейронных моделях, разрабатываемых как самообучающиеся системы, одним из формальных средств, которые предназначены для преодоления «первопорядковой ограниченности», выступает введение так называемых промежуточных единиц (hidden units). Содержательно использование промежуточных единиц означает появление в модели элементов внутреннего представления отображаемых явлений: если в ранних версиях, которые осуществляли группировку тех или иных явлений, объектов путем их прямого сличения (фактически – наложения), вход и выход были соединены непосредственно, то здесь вход и выход опосредованы промежуточным отображением на код внутренних для модели единиц.<sup>6</sup>

Простейшую модель с использованием промежуточных единиц Д.Румельхарт, Дж.Хинтон и Р.Уильямс демонстрируют на формальном аналоге операции строгой дизъюнкции, т.е. исключающего «или»

(Rumelhart e. a., 1988). Авторы со ссылкой на работу М.Минского и С.Пейперта о перцептронах (Minsky, Papert, 1969) отмечают, что при формальной обработке обычной матрицы истинности для строгой

Таблица 1.3. Матрица истинности для операции строгой дизъюнкции (1 «истинно», 0 «ложно»).

Вход	Выход
00	-> 0
01	-> 1
10	-> 1
11	-> 0

дизъюнкции средствами самообучающейся модели, которая «умеет» только группировать структуры по их внешнему сходст-

<sup>6</sup> Нетрудно видеть в этих представлениях сходство с эволюцией бихевиористских моделей в психологии: если в раннем, ортодоксальном бихевиоризме поведение описывалось схемой «стимул – реакция», то в необихевиоризме связь реакции со стимулом опосредована внутренними промежуточными переменными, отражающими состояние (в широком смысле) организма.

ву, результат оказывается парадоксальным: структуры, обладающие наименьшим сходством, должны получить одинаковое отображение и наоборот (табл. 1.3).

Введение всего одной промежуточной единицы, как на рис. 1.2, дает работающую систему, которая воспроизводит функционирование «реле», реализованного в квазинейронной сети и основанного на отношении строгой дизъюнкции. Поясним работу этой системы (Rumelhart e. a., 1988, p. 319–320). Числа, помещенные в кружочках, отражают условные величины, функционирующие в качестве пороговых; так, пороговое значение 0,5 для выходного узла системы (логически отвечающего истинностному значению строгой дизъюнкции) означает, что он активируется, срабатывает (логически – приобретает значение «истинно»), когда общий уровень активации, передающейся выходному узлу в силу активации входных узлов и/или промежуточного (промежуточной единицы), превышает  $+0,5$ . Численные величины на стрелках схемы указывают на веса, задающие уровень активации соответствующих узлов сети.

Таким образом, когда активируется, «включается» (логически – приобретает значение «истинно») любой из входных узлов, что дает количественно уровень в  $+1$ , активируется и выходной узел (принимает значение «истинно»). Когда же в состоянии активации находятся оба входных узла, они соединяются с выходом через промежуточный узел. В этом случае «уровень возбуждения», достигающий выходного узла, не превышает порогового ( $1,5 - 2$  или  $1,5 - 2 + 1 + 1$ , если допустить параллельную, минуя промежуточную единицу, связь входных узлов с выходными). Иначе говоря, в указанных условиях выходной узел не активируется, или, в логических терминах, ему соответствует значение «ложно».

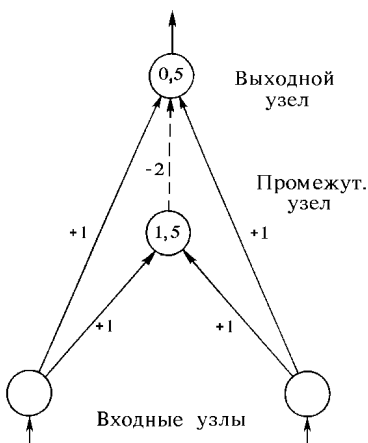


Рис. 1.2. Модель операции строгой дизъюнкции с использованием промежуточного узла (промежуточной единицы) (по: Rumelhart e. a., 1988).

Такого же типа схемы, только, разумеется, значительно более сложные, используются в квазинейронных моделях восприятия речи. Сущность их заключается в том, что самообучающаяся система учится соотносить с входными сигналами функционально, поведенчески адекватные реакции с учетом многообразных зависимостей значимости этих сигналов от контекста и прочих факторов. Стабилизировавшаяся система такого рода функционирует в многопризнаковом пространстве, для которого она устанавливает достаточно сложную систему имплицативных связей («если  $X$ , то  $Y$ »), оцениваемых к тому же количественно через придание им соответствующих весов.

Д.Клатт отмечает, что подход авторов квазинейронных моделей не слишком отличается от известного по статистическим моделям распознавания образов (Klatt, 1989). Это, по-видимому, некоторое преувеличение, ибо статистические модели в типичном случае не прибегают ни к алгоритмам самообучения, ни к внутреннему представлению сигнала, что существенно для моделей квазинейронного типа.

**Модель взаимной активации.** Эта модель, известная больше под названием TRACE, принадлежит Дж.Элману и Дж.Мак-Клелланду (Elman, McClelland, 1986; Frauenfelder, Peeters, 1990). В некоторых важных отношениях основные принципы, на которых базируется данная модель, напоминают о концепциях, известных по фонологической литературе; в частности, можно усмотреть некоторые аналогии с теорией стратификационной фонологии С.Лэма (Lamb, 1970; 1972; см. также Касевич, 1983).

Структура модели TRACE представляет собой иерархически устроенную сеть, узлами которой выступают дифференциальные признаки, фонемы и слова. Признаки многозначны, каждому из значений признака соответствует детектор обнаружения соответствующих акустических параметров. Все признаки, фонемы, слова могут находиться в состоянии активации той или иной степени, которая в модели оценивается условным интервалом от  $-0,3$  до  $1$ . Внутри этих пределов степень активации каждой конкретной единицы определяется ее взаимодействием с другими единицами как того же, так и других уровней, а активация дифференциальных признаков – главным образом параметрами входного акустического сигнала. Сила, с которой одна единица воздействует на другую, повышая или понижая ее активированность активированностью собственной, зависит от связи

между ними, также оцениваемой количественно; сила этой связи для каждой индивидуальной пары (например, «признак Q – фонема X», «фонема X – фонема Y») принадлежит к числу (перенастраиваемых) параметров модели.

Для входного акустического сигнала определяется квантование на элементарные временные отрезки; для каждого такого отрезка модель устанавливает значения всех признаков. Принимается, что фонеме отвечает последовательность из шести элементарных отрезков, однако признаковая информация о каждой фонеме интегрируется относительно каждых трех элементарных отрезков, что обеспечивает перекрытие «фонемных отрезков» во времени, предположительно моделируя процессы коартикуляции.

Признаковый узел, пришедший в состояние возбуждения (активированности) в силу воздействия акустического сигнала, приводит в то же состояние все обладающие им фонемы; например, обнаружение в сигнале «звонкости» активирует все звонкие фонемы. В свою очередь, активируются слова, экспоненты которых содержат ту или иную фонему; уровень их активированности зависит от степени активации фонемы и силы связи между данной фонемой и словом. Возможно и обратное, когда слово, активированное в силу действия каких-то факторов высших уровней, например высокой частотности, активирует соответствующие фонемы.

Важной особенностью модели TRACE является использование не только возбуждения, активации, но и торможения (lateral inhibition). Например, если детектор, специализированный на выявлении признаков гласных, фиксирует наличие характеристики переднего гласного, то в интервале, отвечающем длительности предшествующего согласного, подавляется («тормозится») функционирование всех детекторов установления места образования, кроме тех, что специализированы на опознании согласных перед передними гласными. Одно слово может подавлять уровень активированности другого. Длинные слова имеют более высокие шансы на распознавание, поскольку, обладая более высоким суммарным уровнем активации, успешнее подавляют «соперничающие» единицы.

Фонема или слово оказываются идентифицированными, когда уровень их возбуждения принимает значение, превосходящее степень активированности всех других фонем или слов для данного временного отрезка речевого потока. Предлагаются количественные формулы, по которым вычисляется вероятность

выбора той или иной единицы при данных параметрах системы и входного сигнала.

Модель TRACE не предусматривает специальной процедуры сегментации; принимается, что начало слова может соответствовать любой идентифицированной фонеме, так что система исследует вероятности поступления на вход нового слова в каждой «фонемной» точке речевой цепи. Такой подход, по мнению К.Фаулер, с которым следует согласиться, является неэкономным и вообще едва ли реалистичным (Fowler, 1991, p. 179).

Модель TRACE принадлежит к числу наиболее разработанных, имеются компьютерные версии, в той или иной степени воплощающие принципы этой модели. Тщательное исследование Э.Бард (Bard, 1990) как будто бы свидетельствует об избыточности механизма торможения, без использования которого модель дает те же результаты.

Близка к TRACE модель динамической сети Д.Норриса (Norris, 1990), использующая, подобно квазинейронным сетям, понятие промежуточных (hidden) единиц. Норрис опирается на работу М.Джордана, в которой моделируются процессы порождения речи. В системе Джордана имеются входные узлы, или единицы планирования (plan units), которые соединены с промежуточными узлами, а уже последние соединены с выходными. Помимо этого, выходные узлы связаны с текущими узлами (state units); последние отражают текущее состояние системы и тоже обладают связями с промежуточными узлами. Соответственно состояние системы на выходе определяется активацией не только со стороны входа (через посредство промежуточных узлов), но также предыдущими ее состояниями, информацию о которых хранят текущие узлы. В варианте Норриса планирующие узлы Джордана берут на себя функцию входа, который осуществляет прием акустической информации, выходом же выступают идентифицированные лексические единицы.

Важной особенностью этой модели Норриса является ее способность к самообучению и учету временных параметров, что требуется, в частности, при изменении темпа речи.

**Модель с использованием логики размытых (нечетких) множеств.** В настоящей модели, основным автором которой является Д.Массаро (Massaro, 1987 и другие работы), большое внимание (пожалуй, большее, чем в прочих моделях) уделяется

формальному механизму, с помощью которого осуществляется переход от акустических признаков, обнаруживаемых в речевом сигнале, к сегментным единицам (фонемам или слогам). Особенностью модели является также использование не только акустической, но и визуальной информации (наблюдаемых губных артикуляций и т.п.).

Процедуры, описываемые Д.Массаро и его соавторами, относятся только к фонетической (в широком смысле) стороне языка и речи. Выделяются две основные стадии перцептивного процесса. На первой, как и во многих других моделях, в акустическом речевом сигнале детектируются акустические признаки. На второй с помощью установленной метрики оценивается степень близости данного акустического признака к его прототипическому, эталонному значению, ассоциированному с теми или иными фонемами (слогами). Указанная степень близости принимает вид числа в интервале от 0 до 1, т.е. от полного несовпадения до идеального совпадения; иначе говоря, используется логика размытых, или нечетких множеств, согласно которой вместо ответа «да/нет» на вопрос о принадлежности данного элемента некоторому множеству мы оцениваем меру принадлежности множеству (при вероятностной трактовке, принимаемой, впрочем, не всеми, – вероятность вхождения элемента в множество). Численное выражение меры близости выступает истинностным значением (вместо значений «истинно» и «ложно» в двузначной логике). Перемножая истинностные значения, соответствующие данным акустическим признакам, мы получаем ранг предпочтительности: величину, определяющую близость набора признаков тому идеальному, эталонному их комплексу, который отвечает той или иной фонеме или слогу. Деление истинностных значений на сумму рангов предпочтительности для всех фонем (или слогов) языка дает квазивероятностную оценку каждого из решений относительно идентичности фонемы (слога), принимаемого моделью.

Модель Массаро обнаружила достаточно высокую эффективность на материале, сильно ограниченном, впрочем, качественно и количественно (некоторые типы английских открытых слогов). Клатт приводит фактические примеры, по его мнению, демонстрирующие материал, плохо поддающийся обработке процедурами Массаро (Klatt, 1987;1989). Опыты с синтезированной речью показывают, что акустические признаки, для мужского голоса обычно ответственные за наализацию – увеличение полосы частот, соответствующей F1, и возрастание интен-

сивности 1-й гармонике, для женского голоса оказываются связанными с фарингализацией (breathiness), характеризующейся специфическим шумом в частотной области выше 2 кГц. Но в английском языке (Клатт обсуждает эксперименты, выполненные на материале английского языка) нет фарингализованных фонем, следовательно, отсутствует прототип, эталон, относительно которого оценивалась бы фарингализация. Таким образом, указанное реальное соотношение признаков с поправкой на пол диктора не может быть учтено в модели Массаро.

Тем не менее в целом модель Массаро, по мнению Клатта, превосходит все существующие модели, пользующиеся обращением к признакам и сегментным единицам (Klatt, 1989, p. 209).

**Модель «перцептивной стрелки».** Настоящая модель отражена в нескольких публикациях Дж.Миллера и его соавторов (Miller, 1982; Miller e. a., 1980; Miller, Jongman, 1987). Модель содержит три основных компонента.

Первый компонент моделирует слуховую систему человека, его задача – устанавливать спектральные характеристики речевого сигнала. Результатом выступает лоцирование («расположение») сигнала в  $n$ -мерном пространстве. Для гласных это пространство трехмерно, где в качестве трех его осей используются отношения формант:  $F1/kF0$ ;  $F3/F2$ ;  $F2/F1$ .

Второй компонент представлен механизмом, основной принцип работы которого реализуется как вращение «перцептивной стрелки» в указанном  $n$ -мерном пространстве. Стрелка пробегает значения, соответствующие данному сигналу, реагируя на них с постоянной времени 60–180 мс. Обладая инерцией и упругостью, «стрелка» при внезапной смене направления изменения параметров входного сигнала еще продолжает движение в прежнем направлении и «проскакивает» положение, соответствующее моменту этого внезапного изменения. Так, повышение  $F2$  после взрыва в слове [ba] приведет к тому, что «стрелка» укажет на большее, чем на стационарном участке гласного, значение  $F2$ , тогда как понижение  $F2$  в слове [da] приведет к обратному результату. После усреднения на интервале гласного (за счет большой постоянной времени) эффект коартикуляции будет нейтрализован; вернее, таково намерение автора модели, однако нет уверенности в том, что этот результат достижим столь простыми средствами.

Третий компонент модели ответствен за принятие решения относительно характера речевого сегмента; для гласных такой сегмент считается детектированным, когда движение «перцептивной стрелки» характеризуется медленной скоростью или сменой направления. По существу, как можно понять, речь идет о том, чтобы «отстроиться» от возмущений в сигнале и определить параметры квазистационарного участка с использованием избранной метрики. Вызывает, однако, сомнения возможность установить надежные численные константы, позволяющие различать события, относящиеся к коартикуляции, от информационно независимых его изменений. Кроме того, сомнительна и сама по себе абсолютность этого различия. Хорошо известно, например, что для губных мягких согласных русского языка сигналом мягкости выступает именно коартикуляционный эффект – [i]–образный тип формантного перехода от согласного к гласному. Утрата этой информации, будучи невозможной, сделает невозможной идентификацию согласного – хотя идентификация гласного без такой «помехи» окажется более простой задачей.

**Логогенная модель.** Представления, связанные с настоящей моделью, в течение ряда лет разрабатывались Дж. Мортонем (Morton, 1969; 1979; 1982). Логогеном автор назвал информационно–поисковую структуру, сопоставленную каждому слову словаря воспринимающей системы. Структура включает все сведения об этом слове: его значение, морфологические, словообразовательные и синтаксические функции и признаки, фонологическую и орфографическую структуру, а также ранг частотности. Относительно соответствующей информации, предположительно уникальной для каждого слова, оцениваются все доступные признаки и параметры воспринимаемого фрагмента текста. Если общая сумма параметров превышает некоторый критический порог, логоген активизируется, что и соответствует моменту опознания ассоциированного с ним слова.

В логогенной модели, таким образом, предусматривается взаимодействие всех источников знания уже на самом раннем этапе перцептивного процесса. Одновременно и в интерактивном режиме взаимодействуют признаки разных уровней – от собственно фонетических до синтактико–семантических и частотных. Поскольку в логоген «встроены» и признаки потенциальных контекстов слова, возможно использование и этого параметра; в результате слова, в большей степени предсказываемые

контекстом, опознаются быстрее, их логогены раньше достигают порога возбуждения. Аналогичным образом слово, которому соответствует более высокий ранг частотности, обладает логогеном, устанавливающим более низкий порог активации, т.е. для опознания этого слова требуется меньше признаков других уровней и меньше времени.

Логогены представляются, скорее, как пассивные устройства наподобие фильтров, «просеивающие» поступающую на вход информацию, но не предназначенные для ее активного извлечения из сигнала и, шире, текста.

Рассматривая логогенную модель и достаточно высоко оценивая лежащие в ее основе идеи, С.Д.Голдингер и др. отмечают, что «к сожалению, подобно многим обсуждаемым теориям, логогенная теория носит крайне неопределенный характер. В лучшем случае эта теория помогает нам составить общее представление о том, как работает интерактивная система, как можно учесть эффект частотности, но в ней почти ничего не говорится о том, как именно интегрируются типы информации, получаемые от акустико-фонетического источника знания и источников знания о высших [языковых] уровнях, как протекает процесс распознавания во времени, какова структура словаря» (Goldinger е. а., 1990, р. 37). С этими упреками можно в целом согласиться, добавив, однако, что ответом на все заданные вопросы может служить лишь функционирование реальной модели, построенной в соответствии с данными – или какими-то иными – теоретическими принципами.

**Кортежные модели.** Под данным условным наименованием мы объединили модели, авторы которых разрабатывают систематические процедуры последовательного сужения класса поиска. Вероятно, наиболее известна модель когорты (MarslenWilson, Welsh, 1978 и другие работы). Согласно соответствующим авторам, распознавание слов производится «слева направо»: начальная цепочка входного сигнала проверяется на совпадение с аналогичными цепочками имеющих словарных единиц. В разных работах по-разному оценивается, каким должен быть объем цепочки, с которой начинается процесс распознавания слова: иногда это минимум в одну-две фонемы (Tyler, Wessels, 1983), иногда такая цепочка описывается как состоящая из 3–5 фонем или приравнивается к первому слогу слова (Taft, Hambly, 1986; Segui е.а., 1990; Shillcock, 1990),

иногда же длительность цепочки оценивается во временных параметрах.

Перцептивная система активирует все слова словаря, поддерживающие данную начальную цепочку, которые и формируют класс выбора (поиска). Одна из возможностей ограничить класс – в идеале свести его к единице, т.е. распознать слово, заключается в продолжении анализа: если оказывается, что цепочка становится уникальной, решение принимается без дальнейшего обращения к фонетическим характеристикам слова; такую точку, после которой принимается решение о данном слове, называют «точкой распознавания». Установление точки распознавания может основываться и на отрицательном материале. Например, при восприятии фразы *Tell the gardener to plant some more tulips* точка распознавания для /tɛl/ устанавливается при формировании цепочки /tɛld/, поскольку в английском языке нет слова, начинающегося на /tɛld/ (Cole, Jakimik, 1980).

Положение о том, что перцептивная обработка слов осуществляется именно в направлении «слева направо», проверялось в экспериментах по семантической активации (semantic priming). Оказалось, что начальная цепочка *capt* английского языка активирует одновременно лексемы *guard* и *ship*, поскольку *capt* может быть началом как слова *captive*, так и слова *captain*. В отличие от этого, если испытываемым предъявляются конечную цепочку, например *attle*, ср. *battle*, активации лексем типа *war* не происходит.

В более поздней версии модели Марслен–Вильсона (Marslen-Wilson, 1987; 1990) допускается вероятностная оценка совпадения анализируемой начальной цепочки с соответствующими цепочками словарных единиц; это делает стратегию более гибкой и не приводит к ситуации, когда неверная оценка одной-единственной начальной фонемы или ее вырожденная, ошибочная реализация делают идентификацию слова невозможной – при том, что, согласно имеющимся данным, ошибочное произнесение, например *grocodile*, отнюдь не исключает адекватного восприятия слова *crocodile* (Taft, Hambly, 1986). Учитываются также частотность слова и контекст: сужение класса поиска осуществляется за счет отбрасывания низкочастотных слов и слов, не соответствующих контексту.

В ряде работ показано, что обязательное прекращение фонетического анализа после точки распознавания является слишком сильным условием. Соглашаясь с ним, невозможно, в частности, объяснить распознавание незнакомых слов, уже нача-

по которых характеризуется уникальным фонемным составом. Кроме того, при данном подходе оказывается, что суффиксы и окончания, нормально находящиеся «справа» от точки распознавания, выпадают из анализа, и идентификация многих, даже большинства, дериватов и словоформ становится невозможной (Shillcock, 1990). Соответственно М.Тафт и Г.Хамбли, Р.Шиллкок предлагают свои модифицированные модели (Taft, Hambly, 1986; Shillcock, 1990).

Другие авторы используют понятия, близкие к кортежу (когорте) Марслен–Вильсона, имея в виду классы поиска, организованные по другим принципам – не по совпадению начальной цепочки в экспонентах слов. Так, Дж.Худ и Дж.Пул (Hood, Poole, 1980) в экспериментах по восприятию на фоне шума обнаружили, что высокочастотные слова вопреки ожиданиям не всегда характеризуются лучшей разборчивостью. Д.Пизони и его соавторы (Pisoni e. a., 1985) обследовали список слов, использовавшихся в экспериментах Худа и Пула, с точки зрения фонетических характеристик соответствующих лексических единиц. Было установлено, что 25 слов экспериментальной программы, которые отличала самая низкая разборчивость, в словаре объемом 20 тыс. слов имеют максимальное число квазиомонимов. Этот аспект, вообще говоря, обращал на себя внимание исследователей довольно давно: чем больше у слова квазиомонимов, тем хуже оно воспринимается. Из чего следует, очевидно, что классом поиска, кортежем (когортой) может быть и множество квазиомонимов.

Те же авторы (Luce e.a., 1990) расширяют этот принцип, предлагая *окрестностную модель* (Neighborhood–Activation Model). В данной модели также происходит параллельное использование информации, относящейся к сигналу и характеристикам словарных единиц. Здесь, однако, кортеж составляется словами, совпадающими не обязательно начальными цепочками, но любыми характеристиками с точки зрения «фонетического сходства». Другая важная особенность заключается в том, что идентификация слова ставится в зависимость от объема класса поиска – в отличие от модели когорты, где этот параметр считается иррелевантным. Э.Бард предполагает, что полученный в экспериментах эффект объема когорты (кортежа) может в действительности объясняться влиянием частотности: в более обширном классе поиска вероятнее наличие высокочастотных слов, которые выступают «сильными соперниками» идентифицируемого слова (Bard, 1990, p. 200).

Черты модели когорты и логогенной модели совмещает *ограничивающая модель* (checking model; Norris, 1986). Центральным понятием этой модели выступает перенастраиваемая перцептивная установка (criterion bias): собственно перцептивный – как можно понять, скорее сенсорный – анализ акустического материала предоставляет в распоряжение слушающего набор лексических единиц, а установка, сформированная на базе знания частотностей слов, конкретного контекста, ограничивает этот набор, способствуя принятию решений об одних словах и отбраковывая другие, снимая неоднозначности различного происхождения. Достигается это за счет снижения порога для принятия решения относительно вероятных слов и поднятия порога для слов с низкой вероятностью.

Как можно видеть, многие модели, и прежде всего кортежные, делают акцент на работе со словарем: распознавание слова – это отождествление входного сигнала с той или иной единицей словаря. Такой подход, во многом базирующийся на нисходящих – «сверху вниз» – процедурах, полностью оправдан. Однако К.Фаулер, рецензируя коллективную работу «Cognitive models of speech perception», справедливо замечает, что ни в одной из рассматриваемых там моделей не представлен «a front end», т.е. «компонент, в котором признаки, фонемы или какие-либо другие единицы извлекались бы из сырого акустического сигнала...» (Fowler, 1991, p.183). Она же указывает, что нельзя не учитывать процессы усвоения языка, когда распознаются новые слова, добавляясь к еще только формирующемуся словарю (то же можно сказать о словах, новых для взрослого носителя языка).

Здесь фактически присутствуют две проблемы. Первая – это выделение психоакустических процедур, которые лишь поставляют материал – еще достаточно «сырой» – для принятия фонологических решений. Никакие стратегии обращения к словарю невозможны с использованием самих по себе результатов работы соответствующего модуля. Вторая проблема – различение фонологических и лексических решений. Из признания важности обращения к словарю никак не следует, что распознавание слов невозможно помимо словаря: безусловно, возможно и в ряде случаев необходимо фиксирование структурированных фонологических последовательностей, которым сопоставлены либо не сопоставлены (когда это новое, знакомое слово) определенные значения.

К сожалению, в очень содержательных, богатых экспериментальными данными и тонкими наблюдениями работах, посвященных проблемам идентификации слов (кроме уже обсуждавшихся, можно упомянуть статьи Becker, 1980; Cairns, Hsu, 1980; Forster, 1981; Gordon, Caramazza, 1985; Seidenberg, Tanenhaus, 1986 и многие другие работы), недостаточно четко трактуется центральное понятие лексического доступа, доступа к словарю, лексической идентификации (lexical access). Предлагаются понятия постлексического (post-access) механизма, различаются предлексические процессы, сам лексический доступ и постлексические процессы, см., например, (Seidenberg, Tanenhaus, 1986, p. 138–139), где под предлексическими процессами понимается «анализ сенсорных стимулов». Однако важно сознавать, что анализ сенсорных стимулов вообще не может иметь отношения к лексике, он лишь дает, как, в сущности, уже говорилось, описание сигнала в некоторой метрике. Когда авторы кортежных моделей говорят о сравнении цепочек фонем с «сенсорным входом», тем более, когда говорят о «словах ментального лексикона с той же ... сенсорной последовательностью» (Tyler, Wessels, 1983, p. 409), это нельзя трактовать иначе, как недоразумение. В словаре лексем заведомо невозможны единицы сенсорного уровня, а «сенсорный вход» не несет фонологической информации как таковой.

Идентификация слова может быть лишь двух видов: (а) фонологическая, когда устанавливается звуковая последовательность, соответствующая правилам данного языка (при акустическом сигнале, вырожденном в той или иной степени, мы можем иметь дело с классом словоподобных единиц, между которыми надлежит сделать выбор, или же со звуковой цепочкой, не полностью идентифицированной с фонологической точки зрения), и (б) полная идентификация, когда речевая последовательность акустических событий отождествляется в итоге с некоторой словарной единицей, и таким образом воспринятому, идентифицированному слову приписывается фонологической, грамматической, семантической и иной информации ровно столько, сколько «имеет» в словаре конкретного носителя языка данная единица.

Реалистичной представляется точка зрения тех исследователей, которые предполагают наличие множественных источников информации, среди которых может быть и тип начальной подцепочки, и тип подцепочки конечной, и акцентный контур слова (см. особенно: Касевич и др., 1990б; Slowiaczek, 1990), и

ударный гласный, и гиперпризнаки, отражающие способ артикуляции (Pisoni e. a., 1985). Пизони и его соавторы показали, в частности, что на материале словаря в 125 тыс. слов использование начальной и конечной подцепочек для формирования класса поиска дает примерно одинаковые результаты (Pisoni e. a., 1985).

Положение о множественности источников информации, отказ от поиска какой-то единственной процедуры, обеспечивающей доступ к словарю, носит принципиальный характер.

Именно на сочетании представлений об относительной автономности различных блоков модели, их иерархическом соотношении и систематическом взаимодействии основан подход авторов настоящей монографии (см. об этом гл. II).

Что же касается моделей, рассмотренных выше, то ни одна из них, конечно, не претендует на полное воспроизведение процесса восприятия речи, ибо ни одна не дает системы формализованных правил преобразования речевого сигнала на всех этапах его обработки в слуховой системе – от исходных акустических параметров до символического представления. Некоторые из этих моделей являются, скорее, специализированными инженерными системами, предназначенными для решения ограниченного круга задач, связанных с автоматическим распознаванием речи.

Вместе с тем каждая содержит элементы, которые гипотетически могли бы быть частью такой естественной системы и могут быть использованы при создании полной функциональной модели восприятия речи. Эти элементы (уровни обработки, блоки, модули, субмодули) образуют две большие группы, различающиеся функционально:

(1) элементы, осуществляющие переход от акустического (внешнего) представления речевого сигнала к субъективному (внутреннему) параметрическому представлению;

(2) элементы, производящие на основании полученного внутреннего представления описание сигнала в терминах единиц того языка, носителями которого являются говорящий и слушающий.

Обработка сигнала в блоках первой группы производится только «снизу вверх», и ее принципы, по-видимому, одинаковы для любых акустических сигналов. Блоки второй группы связаны между собой перекрестными прямыми и обратными связями, обработка в них производится не только «снизу вверх», но и

«сверху вниз», т.е. используются процедуры анализа и «анализа через синтез».

Таким образом, правила, по которым внешнее представление сигнала преобразуется во внутреннее, и соответственно структура этого представления определяются только анатомо-морфологическим строением соответствующих отделов слуховой системы. Принципы же создания символического описания сигнала (принятия решений о дискретных языковых единицах) зависят также от свойств обрабатываемого сигнала и прижизненно складывающихся связей в высших отделах мозга. Собственно говоря, именно здесь и может быть проведена граница, разделяющая полную функциональную модель восприятия речи на две достаточно независимые подмодели. Казалось бы, эта относительная независимость обеспечивает возможность их раздельного исследования и разработки алгоритмов для последующего включения в модель. Однако на практике это оказывается невозможным, так как по вполне понятным причинам мы лишены прямого доступа к внутреннему представлению сигнала в естественной системе восприятия речи.

С этим связана едва ли не главная проблема, возникающая при разработке и практической реализации полной модели, проблема адекватного ее тестирования, т.е. проверки заложенных в модель алгоритмов на их соответствие тому, что имеет место в реальной естественной системе. Оказывается, что практически ни один из элементов модели не может быть исследован, реализован и тестирован изолированно от целостной модели всей системы.

Еще одна серьезная проблема, которую большинство авторов описанных выше моделей стараются обойти тем или иным способом, – это проблема сегментации непрерывного речевого потока на последовательность дискретных отрезков, поддающихся описанию в терминах языковых единиц разного уровня (см. об этом ниже, с. 65 и сл.).

### Интегративная модель восприятия речи: возможные подходы

#### ПРОБЛЕМА ВОСПРИЯТИЯ В ГНОСЕОЛОГИИ И ПСИХОЛОГИИ



Несмотря на всю специфичность процесса восприятия речи, он не может не быть частным случаем по отношению к восприятию как таковому. Именно поэтому представляется необходимым рассмотреть хотя бы в самом общем виде, как изучаются в науке проблемы восприятия. Соответствующая проблематика очевидным образом принадлежит области психологии, но существует и давняя традиция обсуждения этих вопросов в философии (что лишь отчасти объясняется неразграничением философии и психологии в эпоху, когда традиция зарождалась), и полностью отвлечься от нее не представляется возможным.

Тесная связь перцептологической проблематики с философской (гносеологической, эпистемологической) объясняется тем, что познание окружающего мира предполагает его восприятие, последнее же невозможно без определенной познавательной базы (ср. ниже).

Едва ли нужно оговаривать, что в настоящей работе нет места, как нет равно возможности и необходимости, для полного освещения современного состояния общей проблемы восприятия, тем более ее истории. Мы обратимся лишь к тем идеям в философии и психологии, которые в наибольшей степени существенны для представлений, лежащих в основе развиваемого нами подхода.

Как хорошо известно, соответствующие вопросы в европейской традиции рассматривались уже античными философами. В античные времена сформировались два основных подхода – эмпирический и рационалистический, которые в дальнейшем определяли сущность философских взглядов в области гно-

сеологии, а тем самым в большой степени и теории восприятия. Представители эмпиризма (Демокрит, Эпикур и др.) делали акцент на чувственных «впечатлениях» от объектов окружающей действительности как главном (или даже единственном) источнике познания, рационалисты (Платон и др.) отводили главную роль внутренним (ментальным) процессам и представлениям. В знаменитой теории Платона познание предстает как процесс, в ходе которого душа (≈разум) «припоминает» при встрече с объектом действительности то, что ей известно от рождения. Иначе говоря, объекты действительности выступают как некие несовершенные отражения врожденных универсальных идей, которые одни только и являются подлинными реальностями.

В новое время рационалистическая традиция представлена прежде всего Р.Декартом. Несмотря на дуализм, который традиционно подчеркивается в литературе, Декарт в области гносеологии безусловный приоритет отдавал активности разума, также, подобно Платону, постулируя существование врожденных идей. «Я не могу сомневаться в том, - писал Декарт, - что во мне имеется некоторая пассивная способность к восприятию, т.е. к приятию и опознаванию идей чувственных объектов; но эта способность была бы для меня бесполезной, и я никак не мог бы ею пользоваться, если бы во мне, или же еще где-то, не было бы одновременно другой, активной, способности, дающей возможность формировать и порождать эти идеи» (цит. по: Gardner, 1985, p. 52).

Представители эмпиризма, и прежде всего Дж.Локк, во главу угла ставили чувственные впечатления, утверждая их уникальность в качестве источника знания и настаивая на том, что нет ничего в разуме, чего бы не было прежде в ощущениях. Локк, как известно, считал разум *tabula rasa*, на которой чувственный опыт «пишет» те или иные «идеи». В дальнейшем имеет место достаточно сложный процесс образования «сложных идей» из «простых», обобщения и абстракции, установления связей и отношений между идеями, в результате чего и возникает все богатство образно-мыслительной деятельности человека.

И.Канту принадлежат фундаментальнейшие концепции, задачей которых, в частности, было преодоление противопоставления эмпиризма и рационализма, апостериоризма и априоризма. Кант признает, что ощущения человека вызываются существующими независимо объектами внешней действительности. Но

эти объекты – знаменитые кантовские «вещи в себе» – являются лишь источником ощущений, взаимодействие с ними органов чувств создает мир феноменального, мир же ноуменального, сущностного недоступен восприятию и познанию. Познание есть результат наложения на мир феноменального категорий и схем, принадлежащих человеческому рассудку. Поэтому суждения человека одновременно синтетичны, ибо они дают новое знание, и априорны, ибо это знание порождается в силу функционирования собственных механизмов рассудка.

Помимо пространства и времени, выступающими по Канту априорными формами чувственности, он выделяет 12 категорий, которые вслед за Аристотелем (вкладывая в термин, однако, несколько иное содержание) Кант называет «категориями мысли»; к последним принадлежат количество (единичность, множественность, всеобщность), качество (реальность, отрицание, предельность), отношение (субстанция, акциденция, причинность, взаимодействие), модальность (возможность, существование, необходимость) и др.

Поскольку категории достаточно удалены от тех феноменов, с которыми имеет дело познающий рассудок, Кант вводит еще один уровень взаимодействия рассудка и опыта – уровень (трансцендентных) схем. Именно схемы непосредственно налагаются на чувственный опыт, организуя его и структурируя. Схемы создаются на основе категорий, но они приближены к реальному опыту и несут соответственно черты образной природы. Возможны схемы разной степени абстрактности: и число есть схема, через которую к чувственному опыту прилагается категория количества, и, скажем, ощущения от такого объекта внешней действительности, как дом, организуются и структурируются определенным образом за счет наложения схемы «абстрактного дома».

Говоря о синтетических ощущениях, Кант, как известно, различал синтез схватывания и синтез воспроизведения, подчеркивая, что в основе того и другого лежит «синтез воображения». «...В нас есть деятельная способность синтеза этого многообразного, – писал Кант, – которую мы называем воображением; его деятельность, направленную непосредственно на восприятия, я называю схватыванием» (Кант, 1964, с. 713). И, что очень знаменательно, в примечании Кант добавлял: «Что воображение есть необходимая составная часть самого восприятия, об этом, конечно, не думал еще ни один психолог» (там же).

По-видимому, еще достаточно долго и после Канта «ни один психолог» не учитывал в должной мере роль активной компоненты в процессах восприятия. В настоящее время это уже почти общее место, однако соотношение активных и пассивных процедур, их тип, сравнительная роль понимаются разными авторами неодинаково.

Х.Гарднер в своей монографии, посвященной истории «когнитивной революции» в психологии, выделяет два основных подхода к интерпретации психических процессов: «молекулярный» и «молярный» (Gardner, 1985). Под первым имеется в виду акцент на исследовании элементарных событий в психических процессах, таких, как, например, формирование единичных ассоциаций, образующихся в течение кратких временных интервалов; для второго подхода типично внимание к более «глобальным» процессам, которые требуют больших периодов времени и предполагают обращение к понятиям наподобие схем, фреймов, стратегий. Гарднер склонен сближать «молекулярный» подход с признанием восходящего характера психических процессов – «снизу вверх», а молярный – с предпочтением нисходящего характера – «сверху вниз», хотя он и оговаривается, что эти две пары понятий нельзя считать идентичными (Gardner, 1985, p. 97). Он также замечает, что «молекулярный» и восходящий подходы близки традициям эмпиризма, а молярный и нисходящий – рационализма (там же).

Мы не можем здесь заниматься специальным анализом содержания названных подходов, тем более в историческом аспекте.<sup>1</sup> Вместе с тем есть основания согласиться с возможностью разграничения аспектов (скорее аспектов, нежели подходов, ср. ниже, особенно с. 61), которые мы предпочли бы обозначить, в согласии с одной из существующих традиций, как элементаристский и холистский. Эти последние действительно тесно связаны с представлениями о психических процессах, и

---

<sup>1</sup> Заметим лишь, что едва ли оправданно класть в основу разграничения разных подходов временной масштаб анализируемых явлений, как это делает Гарднер. Одно и то же явление может помещаться (или не помещаться) в разный контекст – это и будет, в сущности, различием в подходах, в связи с чем, конечно, уместно вспомнить знаменитое противопоставление «анализа по элементам» и «анализа по единицам» у Л.С.Выготского. Что же касается постулируемой связи между рационализмом и признанием нисходящего характера психических процессов, то для парадигмы, где рационализм противопоставляется не эмпиризму, а романтизму, представления о нисходящем направлении будут тяготеть именно к последнему.

прежде всего восприятия, как имеющих восходящую и нисходящую направленность соответственно.

Элементаризм предполагает, что перцепт строится «снизу вверх» в процессе организации, структурирования результатов отражения, относящегося к простейшим событиям, признакам, элементам воспринимаемого объекта (как можно видеть, здесь действительно представлен подход, близкий к традициям Локка и других сторонников философского эмпиризма). В конечном счете, по-видимому, перцепт оказывается «вырастающим» из ощущений, из первичного сенсорного материала.

Холистические концепции во главу угла кладут признание целостности воспринимаемого объекта в качестве неперменного условия отражения его качественного своеобразия. Они настаивают на том, что без гипотезы относительно природы целого восприятие частей невозможно, хотя бы потому, что части (элементы) могут быть структурированы самыми различными способами в зависимости от того, какое целое они образуют. Знаменитые парадоксы: «целое больше суммы своих частей» и «познание целого невозможно без познания его частей, но последнее немыслимо без знания о целом», которые были подмечены уже Аристотелем, а в особенно заостренной форме сформулированы Шеллингом, здесь как бы снимаются в пользу признания приоритетности целого и целостности. Как хорошо известно, именно холистская направленность характерна для гештальтпсихологии.

Холистский подход не может описывать восприятие как автономный процесс – независимый от фоновых знаний человека, его модели мира и т.п. Этот подход естественным образом тяготеет к акцентированию нисходящих процедур, протекающих по принципу «сверху вниз». Что же касается соотношения целостного представления объекта и отражения его частей, конкретных признаков и т.д., то, например, в концепции Л.М.Веккера (1974) построение перцепта рисуется как его поэтапная конкретизация, переход от топологии к метрике, осуществляющийся посредством закономерно сменяющих друг друга стадий. Заметим, что при таком подходе перцептивный процесс носит строго последовательный характер, ср. также (Ганзен, 1974); это, впрочем, справедливо и для большей части элементаристских концепций.

Не занимаясь более обсуждением общих проблем восприятия, обратимся непосредственно к восприятию речи, где мы с неизбежностью встречаем практически все вопросы, кото-

рые дебатировались в эпистемологии, гносеологии и психологии, плюс, естественно, целый ряд аспектов, специфичных для языка и речи.

## ОСНОВНЫЕ ПРИНЦИПЫ ПРЕДСТАВЛЕНИЯ ПРОЦЕССОВ РЕЧЕВОСПРИЯТИЯ

Начнем с замечаний терминологического характера. В некоторых традициях принято различать «восприятие языка» и «восприятие речи». Как пишет автор одного из последних обзоров, «традиционно восприятие речи и восприятие языка были двумя различными областями исследования. ... Первое было связано с восприятием таких единиц, как фонема или слог, а второе занималось словосочетаниями и предложениями. Однако в последнее десятилетие или более ... эта картина стала меняться» (Flores d'Arcais, 1989, p. 98). В отечественной литературе такого противопоставления, кажется, никогда не существовало. В то же время здесь важно различать два аспекта, лишь один из которых носит более терминологический характер.

Мы имеем в виду, что говорить о «восприятии языка» вообще едва ли законно. Объектом перцептивных процессов выступает отнюдь не языковая система сама по себе, но именно речь (текст) как продукт ее деятельности.

Что же касается противопоставления процессов, связанных с восприятием фонем или слогов, с одной стороны, и словосочетаний, слов, тем более предложений, с другой, то различие между ними действительно есть, оно носит принципиальный характер. Восприятие фонем, слогов, равно как и тонов, акцентных контуров, интонационных типов не предполагает понимания, в то время как восприятие предложений, текстов, в известной степени слов и словосочетаний, в типичной ситуации именно понимание и предполагает. Условно говоря, в первом случае мы находимся в пространстве «между» акустикой/психоакустикой и языком, во втором – «между» языком и энциклопедическими знаниями.

Обсуждая этот вопрос, мы сталкиваемся с необходимостью уточнения самого понятия и термина «восприятие». Автор уже цитировавшегося обзора замечает, что «в зависимости от области исследования и интересов исследователя термин "восприятие" был синонимичным по отношению к идентификации, распознаванию (*recognition*), различению, пониманию и усвоению (*comprehension*). В исследованиях по восприятию речи этот

термин покрывал собой почти любые сенсорные и перцептивные операции, а в психолингвистике он использовался для обозначения таких разных процессов, как распознавание слов, сегментация речевого сигнала, определение сходства между двумя языковыми структурами и даже понимание связного текста» (Flores d'Arcais, 1989, p. 97).

Мы, со своей стороны, предпочли бы следующую формулировку: восприятие речи есть приписывание языковой структуры речевому сигналу. Как можно видеть, данная формулировка акцентирует скорее результат, нежели процедурный аспект перцептивного процесса. Она в значительной степени снимает то противопоставление требующих/не требующих понимания перцептивных операций, о котором говорилось выше. В процессе восприятия речи слушающий (нас интересует в первую очередь восприятие речи звучащей) должен интерпретировать речевой сигнал в терминах соответствующей языковой системы. Последняя обладает, как минимум, номинативным (словарным плюс, возможно, словообразовательным), фонологическим, морфологическим, синтаксическим и семантическим компонентами с собственными единицами и правилами. Следовательно, поступающий речевой сигнал должен быть перекодирован слушающим в структуру, охарактеризованную через признаки, принадлежащие всем или некоторым из названных выше компонентов. Так, результатом восприятия может быть установление лишь фонологической структуры, например фонемной цепочки и места ударения, без отождествления полученной единицы с одной из словарных единиц, как это и имеет место реально при восприятии бессмысленных, новых слов или имен. В других случаях результатом работы перцептивных механизмов будет структура, сформированная словами, отождествленными с соответствующими словарными единицами, которые (слова) входят в определенные синтаксические и семантические отношения друг с другом, а сами обладают, в свою очередь, внутренней структурой, морфологической и, в плане выражения, фонологической.

Иначе говоря, возможна разная «глубина» восприятия, определяемая мерой, полнотой использования языковой системы для внутреннего представления речевого сигнала. Кроме того, возможны разные пути, разные процедуры достижения данного перцептивного результата: так, установление фонемной цепочки, выступающей экспонентом слова, может быть получено в силу анализа акустической картины, обрабатываемой с

целью извлечения информации о коррелятах тех или иных дифференциальных признаков, но может быть и побочным результатом действия механизмов предсказания, которые прогнозируют, иногда с вероятностью, равной единице, появление конкретного слова в данной точке речевой цепи. В типичном случае для воспринимающего речь человека нет субъективного различия между разными путями формирования языкового перцепта.

Здесь мы вплотную подходим к вопросу о структурировании перцептивного механизма и его процедур. Однако сначала представляется необходимым выделить доперцептивный, или, условно, психоакустический механизм, функционирование которого составляет необходимую предпосылку работы перцептивного.

Психоакустический блок, или модуль, не обладает специфичностью ни по отношению к конкретному языку, ни по отношению к языку вообще. Он, в сущности, совпадает с анализатором, генетически предназначенным для обработки слуховой информации, т.е. со слуховым анализатором – вернее, с сенсорными (рецепторными), т.е. низшими, отделами последнего. Несколько более подробно отдельные проблемы, связанные с работой данного модуля, будут обсуждаться ниже. Сейчас же ограничимся констатациями общего порядка: механизмы психоакустического модуля измеряют частотные, временные, энергетические параметры звукового сигнала в их соотношении. Эти механизмы преднастроены (предпрограммированы), их деятельность поддается описанию посредством соответствующих функций. Едва ли возможно говорить об их обучаемости в собственном смысле слова, хотя, безусловно, можно говорить об обучаемости систем, «считывающих» показания психоакустических механизмов и структурирующих их определенным образом.

На выходе психоакустического модуля формируется описание исходного речевого сигнала в виде последовательности дискретных сегментов речевого потока, еще не имеющих размерности фонем. Эти сегменты образуются в результате работы процедур микросегментации (см. гл. III, с. 66), и каждый из них описывается набором признаков с указанием их «весов» (вероятностей). По своему содержанию они, видимо, могут совпадать с выходными сигналами «детекторов акустических признаков» (обсуждение проблемы исследования детекторов см. в гл. IV).

Сами перцептивные механизмы несомненно включают блок, или модуль, непосредственно работающий с информацией, соответствующей выходу психоакустического модуля. Иначе говоря, выход психоакустического модуля доступен для целого ряда когнитивных систем, в том числе и для перцептивной речевой системы, для одного из ее модулей. Последний включается в работу, когда принимается решение о том, что звуковой сигнал является речевым. Будем говорить о таком модуле как о фонетико-фонологическом. Он включает некоторый набор субмодулей. Их точный состав не может быть пока определен.

Во входных субмодулях данного модуля принимаются решения о фонемах, дискретной последовательностью которых может быть описан данный участок речевого потока. Для этого на каждом из сегментов используются комбинации признаков, полученные в психоакустическом модуле, а также данные для нескольких последовательных сегментов. При принятии решений учитываются вероятности признаков, так что первыми выдаются наиболее вероятные фонемные решения.

Если результаты лексического поиска и работы последующих уровней анализа при полученной таким образом фонемной последовательности окажутся неудовлетворительными и возникнет необходимость «вернуться» к уровню создания фонемного описания, чтобы попытаться получить более подходящую цепочку фонем в качестве описания данного сигнала, то будут повторно задействованы входные субмодули; при этом исходная информация для них (выходные сигналы психоакустического модуля) должна храниться в памяти и не может быть изменена процедурами «сверху вниз».

Ясно, однако, что в любом случае мы имеем дело с многоканальной системой, пронизанной прямыми и обратными связями, которые позволяют ей функционировать в интерактивном режиме (ср. ниже).

В отличие от психоакустического модуля, фонетико-фонологический складывается прижизненно, нормально в раннем онтогенезе. Его механизмы, будучи сформированными, тоже в известном смысле преднастроены: они осуществляют в пространстве наличного сенсорного материала активный поиск информации, которая может быть перекодирована в некоторое символическое представление, отвечающее системе данного языка.

В частности, для каждой фонемы конкретного языка должен существовать, вероятно, ее перцептивный эталон. Последний задает конфигурацию признаков, выделяемых психоакустическим модулем, а также допустимые пределы их варьирования относительно заданных условий (высоты голоса, темпа речи, позиции и т.п.) и относительно друг друга. Метрика перцептивного эталона позволяет системе использовать функцию сходства для определения меры близости конфигурации признаков, выделяемых психоакустическим модулем, к конфигурации эталонной. Эта операция необходима, поскольку реальные параметры речевого сигнала сплошь и рядом отличаются высокой степенью неопределенности. Гиперпризнаки, о которых говорилось выше в связи с обсуждением одной из современных моделей восприятия речи (с. 13 и сл.), возможно, целесообразнее интерпретировать как такую ситуацию, когда значение функции сходства позволяет отнести данную конфигурацию признаков к двум или более эталонам с равной вероятностью.

Правда, смысл введения понятия гиперпризнаков в этой модели был связан не с тем, что слушающий не может использовать соответствующую информацию, а, скорее, с тем, что он не должен («не обязан») ее использовать: гиперпризнаки, по мысли авторов модели, позволяют достичь идентификации языковых единиц более экономными средствами. Но мы сейчас обсуждаем фактически максимальные возможности каждого из модулей. Функционирование субмодулей, обладающих низким иерархическим статусом, может действительно в конкретных условиях оказываться избыточным, тем не менее все они необходимы для максимального использования всех ресурсов системы в минимально благоприятных условиях.

Выход фонетико-фонологического модуля представлен цепочкой дискретных единиц-символов, в типичном случае фонем.<sup>2</sup> Как следует из сказанного выше, такая цепочка в достаточно большом числе случаев (хотя какая бы то ни было статистика нам не известна) не может фигурировать в качестве «готового» экспонента той или иной языковой единицы: часть членов этой цепочки может характеризоваться лишь мерой близости к

---

<sup>2</sup> Необходимость оговорки вызвана тем обстоятельством, что существуют языки, фонологические системы которых описываются в терминах не фонем, а других единиц — силлабем или силлаботмем, см. об этом (Касевич, 1983).

фонеме А или же фонемам А/В (А/В/С...), а какая-то часть элементов цепочки может вообще отсутствовать (в сопоставлении с намерением говорящего), причем возможен как вариант, когда такое отсутствие абсолютно (ничто в сформированной фонетико-фонологическом модулем цепочке не говорит о том, что налицо нулевая редукция), так и вариант «относительного» отсутствия, когда фиксируется наличие сегмента, но не его качество. Последний случай можно интерпретировать как вычисление такой функции сходства, которая устанавливает равнозначную отнесенность данной конфигурации признаков к любому из перцептивных эталонов, отвечающих фонемам данного языка. При другом подходе можно говорить о фиксировании единственного признака [+ сегментный], ср. некоторые системы признаков, используемые в генеративной фонологии.

До сих пор речь фактически шла о функционировании субмодуля фонетико-фонологического модуля, который (субмодуль) ответствен за принятие решений относительно сегментных единиц – фонем (вопрос о слогах, также являющихся сегментными единицами, мы сейчас специально затрагивать не будем, хотя информация о слоговой структуре экспонента языковой единицы также необходима, ср. ниже). Но одновременно нужно допустить существование некоторого набора субмодулей, связанных с установлением супraseгментных, или просодических, характеристик речи. Как минимум, это интонационный субмодуль (или, возможно, интонационные субмодули) и акцентный, а для тональных языков – тональный. Эти субмодули действуют параллельно сегментному, что, однако, никак не означает невозможности их иерархического соотношения (см. об этом ниже).

Разумеется, просодические субмодули также работают с информацией, полученной на выходе психоакустического модуля. Сформулируем здесь лишь несколько положений, связанных с возможными принципами функционирования акцентного субмодуля.

Обычно принимается, что информация об ударении – это фактически информация об акцентном контуре слова. Из этого следует, что для получения соответствующих характеристик необходимо обладать сведениями о членении звучащей речи на слова. Однако, как будет показано ниже (гл. III), информация о сегментации речевого потока на слова никоим образом не обеспечивается одними или даже преимущественно фонетическими признаками. Между тем есть основания полагать, что

место ударения можно установить и без знания о том, где проходят границы слов, т.е. в определенном смысле эта задача может решаться в пределах фонологического (фонетико-фонологического) компонента.

Соответственно получение информации об ударении может пониматься как задача детектирования ударных слогов не по отношению к границам слова, а, скорее, по отношению друг к другу: определяется момент, соответствующий первому ударному слогу, второму, третьему и т.д. Это связывает нахождение акцентных характеристик с проблемой речевого ритма (см. об этом гл. III, с. 98 и сл.).

Мы отвлекаемся от вопросов, относящихся к типу признаков, интерпретируемых акцентным субмодулем как проявление ударности, от проблемы вариативности этих признаков, прежде всего под влиянием интонации. Результатом работы субмодуля является расстановка меток, соответствующих ударным слогам. В принципе можно было бы говорить и об ударных гласных, однако имеются данные, согласно которым ударность может проявляться и в качественных и/или количественных признаках соответствующих согласных. Из этого, равно как и из некоторых других соображений, следует, что ударность может выступать как характеристика слога в целом.<sup>3</sup> В свою очередь это означает, что детектирование ударных слогов предполагает доступность информации о слоговой структуре речи. Вероятно, установление такой структуры есть задача отдельного субмодуля.

Если в распоряжении системы имеются данные о составе цепочек сегментных единиц (с учетом сделанных выше оговорок относительно достаточно типичной неполноты таких цепочек), а также о местоположении ударных слогов, то необходима еще информация о членении речевого потока на слова, чтобы система могла обратиться к процедуре отождествления соответствующих цепочек с единицами словаря. Заметим сразу же, что в языках с фиксированным ударением обнаружение ударных слогов естественным образом обеспечивает одновременно установление словесных границ (Касевич и др., 1990б). Но и в языках с разноместным ударением, к которым принадлежит

---

<sup>3</sup> Проблема качественного и количественного изменения характеристик ударного слога в целом и составляющих его элементов (по отношению к безударным) в процессе порождения речи, а особенно роли этих изменений при восприятии ударности требует специального обсуждения и исследования.

русский, выявление ударных слогов есть установление числа слов, что также содействует обнаружению словесных границ, особенно если учесть статистические закономерности местоположения ударного слога в слове.

Проблема сегментации звучащего текста на слова будет обсуждаться в разделе гл. III, специально посвященной этому вопросу (гл. III, с. 65 и сл.). Сейчас же обратимся к механизмам словарного поиска.

Если словарь представить себе как некоторую упорядоченную совокупность словоформ (хотя это не очень реалистическая гипотеза, ср. ниже<sup>4</sup>) и исходить из такой ситуации, когда фонемная цепочка, отвечающая слову, «заполнена» однозначно и без пропусков, то проблема словарного поиска решается тривиально. Идентификация данной цепочки, охарактеризованной просодически, в качестве экспонента некоторого слова (словоформы) реализуется в результате последовательного просмотра всего корпуса словаря. Как вполне понятно, такой процесс сплошного сканирования множества словарных единиц носит громоздкий характер. Процедуры этого типа, возможно, используются в ситуациях, когда распознаванию подлежат новые слова, а также слова, появление которых в данной точке речевой цепи характеризуется низкой вероятностью.

Более обычны ситуации, когда восприятие речи организуется встречной активностью субъекта восприятия. Использованию активных стратегий восприятия способствует то обстоятельство, что сам словарь пронизан множественными связями между его единицами – связями семантическими, грамматическими, фонетическими (фонологическими). Иначе говоря, упорядоченность словаря реализуется как наличие в его составе пересекающихся групп слов, каждая из которых объединена теми или иными признаками. Контекст речевого акта, установки слушающего, предтекст позволяют достаточно узко отграничить, прежде всего с семантической точки зрения, тот подсловарь, единицы которого слушающий «рассчитывает» встретить в воспринимаемом тексте. Поэтому и сканирование, о котором говорилось выше, обычно осуществляется в рамках не всего словаря, а одного из подсловарей.

---

<sup>4</sup> Если для входа в словарь при лексическом поиске в числе прочих параметров используется акцентный (ритмический) контур, то в словаре должны быть представлены, как минимум, те словоформы одного слова, которые имеют разную акцентную (ритмическую) структуру.

Меньший объем словаря (подсловаря) имеет одним из своих результатов снижение требований к точности и полноте описания входных единиц с фонетической (фонологической) точки зрения. В этих условиях и возникает возможность использования гиперпризнаков, понимаемых как признаки единиц супрафонемного формата, в типичном случае – слов. К числу гиперпризнаков, трактуемых указанным образом, принадлежат, как можно думать, акцентный контур слова, т.е. информация о числе его слогов и месте ударения, а также тип начальной подцепочки.

Любой подсловарь дополнительно стратифицирован в силу разной частотности входящих в него слов. Частотность слова также принадлежит к его гиперпризнакам.

Все используемые в процессе восприятия признаки (гиперпризнаки) иерархизированы за счет приписывания им определенных весов. Это означает, что при идентификации входной фонемной цепочки в качестве экспонента слова  $X$  или  $Y$  по признакам (гиперпризнакам)  $a, b, c, \dots$ , когда последние отсылают эту цепочку к разным словарным адресам – например, признак  $a$  к слову  $X$ , а признаки  $b$  и  $c$  – к слову  $Y$ , отдается предпочтение тому признаку, которому в системе отвечает больший вес. Поскольку идентификацию слов уместно понимать как активирование связей, принадлежащих ментальному лексикону, можно сказать, что признаки, которым отвечает меньший вес, достаточно часто вообще не используются: уровня активации, достигаемого за счет использования признаков высокого иерархического статуса (с большим весом), оказывается достаточно, чтобы имела место идентификация входной цепочки в качестве слова  $X$ .

Вес, приписываемый признаку, может носить постоянный и переменный характер. По-видимому, постоянно высокое значение весов свойственно тем признакам, которые отличает существенная помехоустойчивость. К последним принадлежат прежде всего просодические признаки. Частотность как таковую тоже представляется уместным отнести к признакам с высоким весовым значением, однако конкретное распределение слов по частотности будет зависеть от выбора подсловаря. В условиях шепотной речи, например, естественным будет снижение веса признака «звонкость/глухость», который в русском и ряде других языков может фигурировать как гиперпризнак, характеризующий более или менее протяженные фонемные цепочки.

Можно предположить, что взаимодействие признаков (гиперпризнаков) в процессе лексического поиска формально описывается с помощью правил теории размытых, или нечетких множеств, как это делает Массаро для идентификации слогов и фонем.

В другой редакции это будет утверждением о том, что контекст, частотность слова и другие факторы повышают или снижают требования к количеству и качеству акустической информации, необходимой для принятия решения об идентификации слова, ср. (Norris, 1986) и другие работы (гл. I, с. 42 и др.): высокочастотные слова, слова, предсказываемые контекстом с достаточно высокой вероятностью, позволяют обходиться меньшим числом признаков, их менее выраженными акустическими коррелятами. При прочих равных условиях такие слова распознаются быстрее. В сущности, это проявление иерархически более высокого статуса таких признаков, как семантические и грамматические валентности, частотность и др. (Касевич и др., 1990б). Вопрос об автономности/интерактивности различных компонентов перцептивной системы оказывается в значительной степени искусственным, ср. (Seidenberg, Tanenhaus, 1986): модули и субмодули перцептивной системы функционируют независимо, но результаты их работы постоянно сопоставляются.

Лишь очень кратко упомянем о тех проблемах, которые принято относить к высшим уровням языка и речевой деятельности и которые связаны не со словарем, а с правилами построения текста и соответственно его анализа при восприятии. Здесь очень важны разного типа валентности. Если расслоение словаря на подсловари сужает класс поиска слова в процессе интерпретации входной цепочки за счет парадигматических связей слова, то валентности дают тот же эффект за счет синтагматических связей. Приведем простой пример. Если предтекст содержит (уже идентифицированное) слово *ограбить*, то весьма высока вероятность появления слова в винительном падеже со значением лица (лиц) или имущественного объекта (*ограбить банкира, богатых, квартиру, банк и т.п.*). Иначе говоря, распознавание слова *ограбить* приводит к активации подсловарей, принадлежащих определенному семантическому полю, а также предопределяет грамматическую форму, в которой должны появиться соответствующие слова. Такого рода сужение класса поиска объясняется именно тем, что слушающий – носитель языка обладает знанием валентностей языковых еди-

ниц, и это знание используется как источник наложения ограничений на семантические и формальные свойства языковых единиц, которые могут быть в тексте после и/или до единицы X.

Из сказанного следует, в частности, что требуют оговорок традиционные формулировки наподобие той, которую мы использовали чуть выше в связи с разграничением сферы словаря языковых единиц и сферы правил, регулирующих использование этих единиц – т.е., практически, грамматики. При всей обоснованности такого разграничения необходимо признать, что уже в словаре при лексической единице должна быть записана достаточно детализированная информация, отражающая валентности этой единицы, равно как и ее частотность, формо- и словообразовательные потенции и т.д.

Для информации о формо- и словообразовательных потенциях требуется, в свою очередь, чтобы словарная единица была представлена как некая формальная структура с выделением блоков, релевантных для действия правил формо- и словообразования. Разные блоки и в процессе восприятия речи будут обрабатываться отдельно. Здесь возникает еще один источник максимального «распараллеливания» процесса восприятия, в данном случае выражающегося в том, что лексическая информация, а также разные виды грамматической информации обрабатываются отдельно и в значительной степени одновременно. В связи с этим заслуживают внимания экспериментальные результаты ряда авторов (Taft, Forster, 1975 и др.), которые указывают на такой процесс распознавания, в частности префигированных слов:

- «(1) установи, что в слове присутствует префикс и удали его для [дальнейшего] лексического поиска;
- (2) найди в словаре основу;
- (3) когда основа найдена, восстанови префикс и зафиксируй слово. Распознавание [слова] осуществляется на данной стадии» (Emmorey, Fromkin, 1989, p. 136).

Использование подобных и близких к ним процедур как раз и дает возможность представлять словарь, участвующий в процессах речевосприятия, не как множество словоформ, отражающих полную парадигму каждого слова (поскольку все члены парадигмы могут встретиться в тексте), а как множество основных, и притом внутренне структурированных словоформ, к которым по определенным правилам можно свести реальные словоформы текста.

Разумеется, префиксация, на которой сосредоточили свое внимание упомянутые выше авторы, – лишь частный, притом не самый распространенный способ передачи грамматической информации. Суть в том, что система восприятия речи должна располагать отдельными модулями и субмодулями для обработки грамматической информации: на основе анализа определенных сегментов текста, фонемных подцепочек, эти модули формируют гипотезы относительно синтаксической структуры предложения, а также целого ряда признаков наподобие грамматического времени, вида, числа и т.п. Вес, приписываемый этим гипотезам, должен быть достаточно высоким, ибо грамматическая структура служит источником существенных ограничений на вид структуры семантической, установление которой и является конечной целью процесса речевосприятия.

В то же время сама гипотеза о грамматической структуре верифицируется не только относительно результатов работы фонетико-фонологического модуля с его субмодулями, когда, например, информация о наличии вопросительного слова сопоставляется с типом интонации и, если рассогласования этих двух источников нет, принимается решение о коммуникативном типе предложения. Гипотеза о грамматической структуре высказывания должна одновременно находиться под контролем семантической гипотезы, или смысловых ожиданий слушающего, которые формируются, как уже упоминалось, его собственными установками, ситуацией, контекстом и т.п.

С этой точки зрения процесс речевосприятия не только завершается «семантической стадией», но и начинается с таковой. Учитывая данное обстоятельство, равно как и неизбежное несовершенство самого акустического речевого сигнала (который в типичном случае просто не содержит информации, необходимой и достаточной для его декодирования), следует эксплицитно предположить, что автономная модель восприятия речи – автономная в том смысле, что она не использует информации, помимо содержащейся в сигнале и в обрабатывающей сигнал языковой системе, – просто невозможна. Задача создания полной и универсальной модели восприятия речи в этом смысле эквивалентна созданию искусственного интеллекта, ср. (Касевич, 1974).

В то же время, как уже отмечалось, следует принимать во внимание, что возможна разная «глубина» восприятия и что само по себе восприятие не предполагает с необходимостью понимания. В начале этого раздела восприятие было опреде-

лено как приписывание речевому сигналу языковой структуры. Чтобы осуществить эту задачу, не всегда необходимо строить ментальную модель (Johnson-Laird, 1983), которая была бы фрагментом модели мира слушающего. Например, человек обычно может записать на слух не очень длинное предложение (которое субъективно может быть представлено как последовательность единиц, числом не превосходящих объем оперативной памяти), даже не понимая его смысла, например, когда предложение отражает какую-то информацию из специфической сферы, чуждой слушающему. Это как раз и означает, что акт восприятия имел место, поскольку языковая структура речевому сигналу приписана, при этом приписанная семантическая структура носит сугубо внутриязыковой характер, т.е. отражает лишь словарные значения лексических единиц (если слова вообще знакомы слушающему) и связанную с ними информацию из области грамматической семантики.

Строго говоря, непосредственной задачей специалиста, моделирующего восприятие речи и не рискующего при этом вторгаться в область искусственного интеллекта, могло бы быть воспроизведение именно процедур такого рода.

### Взаимодействие компонентов и уровней в перцептивных процессах сегментации и идентификации языковых единиц

#### ПЕРЦЕПТИВНАЯ СЕГМЕНТАЦИЯ\*

Понятие перцептивной сегментации. Любой текст представляет собой структурированную целостность. Сама по себе структурированность еще не предполагает с необходимостью линейной упорядоченности и расчлененности. Однако применительно к звучащему тексту такая необходимость практически очевидна: звучащий текст разворачивается в одномерном времени и воспринимается текущим образом; трудно представить себе, чтобы текст не был организован в некоторые линейные кванты-последовательности и эта организация не служила бы опорой для его восприятия. Отсюда и следует, что, воспринимая речь, человек должен каким-то образом ее членить, сегментировать.

Проблема перцептивной сегментации речи имеет уже определенную историю, заслуживающую специального изучения. Начнем, однако, с некоторых аспектов, которые полезно различать, говоря о сегментации.

Прежде всего необходимо разграничивать перцептивную сегментацию и сегментацию как исследовательскую процедуру – часть фонологического анализа. В последнем случае фонолог, устанавливая систему фонем данного языка, должен дать ответ на вопрос о моно-/полифонемности любого сегмента; например, если в текстах на этом языке встречаются сегменты типа [au] или [ts], необходимо определить, имеем мы дело с моно-

---

\* Настоящий раздел написан при участии Н.А.Слепокуровой, ср.: Kasevich e. a., 1993.

фонемными образованиями /au/ и /с/ соответственно или же с сочетаниями фонем /au/ и /ts/. Не ответив на этот вопрос, мы не сможем описать систему фонем языка.

В отличие от этого, понятие перцептивной сегментации предполагает, что носитель языка, владеющий, естественно, языковой системой, должен в процессе восприятия речи представить спектрально-временной континуум акустического речевого сигнала в терминах уже известных ему дискретных языковых единиц. Иначе говоря, можно усмотреть два основных различия между «аналитической» (исследовательской) и перцептивной сегментацией: (1) в первом случае мы выясняем, каков вид фонологической системы и какие разрешенные последовательности фонем имеются в изучаемом языке, во втором – под какую из разрешенных в языке фонемных последовательностей можно подвести воспринимаемый отрезок звучащего текста; (2) для «аналитической» сегментации любая встречающаяся в текстах последовательность должна быть истолкована как та или иная фонема или фонемная цепочка – иначе мы не сможем определить состав фонем в языковой системе, перцептивная же сегментация необходима лишь постольку, поскольку ее результаты влияют на идентификацию языковых единиц.

Все то же самое можно повторить и применительно к единицам более высоких уровней. В лингвистике в течение многих десятилетий обсуждается проблема слова, понимаемая, в частности, и как определение его границ в тексте, или, иначе, как вычленение слова из текста. Здесь тоже есть «аналитический» и перцептивный аспекты. Для первого существенно, какие текстовые последовательности однословны и, следовательно, входят в словарь языка, а какие выступают в качестве словосочетаний, формирующихся из словарных единиц. Для второго аспекта, опять-таки, вопрос о членении на слова значим в той мере, в какой от результата сегментации зависит структурирование, а в конечном счете – семантизация текста, ср. традиционные примеры наподобие *девицу вели – девиц уве-ли* (Лисенко, 1966).

В дальнейшем, говоря о сегментации, мы будем иметь в виду только перцептивную сегментацию. Можно различать по крайней мере два ее типа: микро- и макросегментацию. В первом случае речь будет идти о сегментации непрерывного акустического сигнала (точнее, его внутреннего субъективного представления) на некоторые минимальные отрезки. Природа последних не может считаться вполне ясной: возможно, это не-

посредственно отрезки «фонемной протяженности», возможно – любые другие отрезки, информация о свойствах которых в идеальном случае необходима и достаточна для установления фонемной структуры данного фрагмента текста. В результате такой сегментации появляется возможность измерять длительность вычленившихся отрезков, оценивать степень синхронности появления в речевом потоке тех или иных событий, наконец, идентифицировать эти отрезки или их конфигурации (или же, наоборот, их подструктуры) с элементами соответствующих систем и подсистем, которыми располагает реципиент текста – носитель языка, тем самым ставя в соответствие данному отрезку тот или иной фонемный символ. Результатом такой первичной обработки, включающей сегментацию, является новое описание сигнала в виде последовательности (фонемных) символов.

Что касается макросегментации, то здесь мы имеем в виду членение текста на значимые единицы разного формата и функционального статуса. Естественно, что в первую очередь это будут слова, синтагмы и фразы (clauses).

В настоящей работе мы исследовали исключительно макросегментацию звучащего текста, которая в целях краткости именуется ниже «сегментацией». В терминах А.Мартине (хотя содержательно эта проблема обсуждалась уже Бодуэном) можно сказать, что нас интересовало «первое членение».

Как уже отмечалось в гл. I, большинство существующих моделей восприятия речи не предусматривает специальных процедур сегментации, а авторы некоторых моделей, как, например, модели TRACE, настаивают на ее принципиальной избыточности; обычно принимается, что сегментация выступает побочным продуктом идентификации (см. об этом, например: Лисенко, 1966; Клаас, 1966).<sup>1</sup> Вместе с тем, даже если согласиться с типичностью невыделения особых процедур сегментации в перцептивных процессах, нельзя не учитывать ситуации овладения языком, когда «единственный способ усвоить, какие по-

---

<sup>1</sup> Любопытно, что при моделировании зрительных процессов, где, в отличие от речи, мы имеем дело с нелинейными, трехмерными объектами, в качестве первого шага называют именно сегментацию – «разложение [обозреваемой] сцены на сегменты, или области, которые могут служить кандидатами на роль осмысленных фрагментов [воспринимаемой] картины» (Arbib, 1986, p. 224). Там же, впрочем, указывается, что задача сегментации неразрешима за счет чисто формальных средств и требует привлечения «программ высших уровней».

следовательности являются словами, включает сегментацию сигнала, основывающуюся на информации, которая содержится в нем самом» (Fowler, 1991, p. 181; ср. также: Mehler e.a., 1990, p. 236-237).

В литературе не всегда дифференцируется сегментация на синтагмы, предложения, сверхфразовые единства и вычленение слов. Между тем даже априори можно утверждать, что вероятность наличия в речи специальных указателей на границы между предложениями существенно выше, чем вероятность появления маркеров словесных границ.

Ниже мы в первом приближении обсудим основные факторы, которые потенциально могут оказаться ответственными за обнаружение в тексте словесных границ. При этом мы будем принимать во внимание и границы между синтагмами и предложениями, учитывая, что всякая такая граница - это одновременно и граница между словами (хотя обратное, разумеется, неверно).

Упомянутые факторы - пограничные сигналы - можно условно разделить на несколько основных групп.

Первая группа - фонологические пограничные сигналы, которые в свою очередь делятся на просодические и сегментные. К просодическим обычно относят паузы. Как минимум, необходимо уточнить: наличие паузы говорит о присутствии словесной границы (если это не пауза хезитации, ср. *пере ... - как это говорится?* - ... *форматировать*), но отсутствие паузы не указывает на отсутствие словесной границы.

В качестве просодического пограничного сигнала может функционировать и ударение. Фиксированное ударение, если оно падает на слог, положение которого относительно начала/конца слова строго определено, однозначно указывает на словесную границу. Нефиксированное, разноместное ударение - как в русском языке - не дает информации о словесных границах, хотя число ударений (ударных слогов) говорит о числе слов в данном фрагменте текста.

Среди работ, специально исследующих роль словесной просодики в процессах сегментации, следует упомянуть обширный цикл публикаций А.Катлер и ее соавторов (Cutler, Norris, 1988; Cutler, 1989; 1990 и др.).

Катлер и Норрис разрабатывают представления об особой «стратегии метрической сегментации»: предполагается, что

сильные слоги,<sup>2</sup> в качестве которых выступают слоги с нередуцированными гласными, служат сигналами начала слова (в английском языке). Определив начало слова, распознающая система может более эффективно использовать стратегии подбора словесных кандидатов на базе процедур, предлагаемых такими моделями, как TRACE или, в особенности, модель когорты. А.Катлер ссылается на результаты, полученные в компьютерных экспериментах Э.Бриско, где в одинаковых условиях испытывались программы, построенные по принципу инициирования лексического поиска в каждой «фонемной точке» (ср. модель TRACE), на каждой слоговой границе, с отсчетом от начала предложения и на каждом сильном слоге. Именно последний вариант, соответствующий стратегии метрической сегментации, дал наилучшие результаты (Cutler, 1990, p. 110-111).

В пользу взглядов, развиваемых Катлер и др., говорят и статистические данные: при разных способах подсчета, от 75 до 90% полнозначных слов английского словаря, включая односложные, начинаются с сильного слога; близки и данные по текстовым выборкам. Начальный слабый слог, т.е. слог с редуцированным, смещенным к центру гласным, обычно шва, характерен для служебных слов (Cutler, Carter, 1987; Cutler, 1990).

Другие авторы склонны акцентировать то обстоятельство, что стратегия метрической сегментации будет слишком, по их мнению, часто - примерно в 16% случаев - давать неверные словесные границы, особенно если учесть «встроенные» псевдослова типа *bone* в *trombone* (Shillcock, 1990). Э.Бард вообще подвергает сомнению гипотезу метрической сегментации, считая, что наличие в речевой цепи слабого слога приводит к слабой активации большого числа лексических кандидатов, а наличие сильного - к сильной активации небольшого числа. Иначе говоря, с ее точки зрения, противопоставление сильных и слабых слогов имеет своим результатом не разные типы сегментации, а различные механизмы формирования класса поиска и отбора внутри такого класса (Bard, 1990).

---

<sup>2</sup> Катлер различает лексическую просодику - чередование ударных и безударных слогов - и метрическую просодику - чередование сильных и слабых слогов. Заметим, что несколько раньше, применительно к порождению речи, аналогичное разделение предлагали Дж.Аллен и С.Хокинс, противопоставляя слоговой вес и акцентуацию, где под слоговым весом имелась в виду степень редукции (Allen, Hawkins, 1980). Ср. также различие «нексуса» и «курсуса» у Э.Пальгрэма (Pulgram, 1970).

Вполне понятно, - об этом говорит и А.Катлер, - что в таких языках, как французский, где нет аналогичной дихотомии «сильный слог/слабый слог», стратегия метрической сегментации вообще неприменима.

Сегментно-фонологические пограничные сигналы - это прежде всего присутствие фонем или их сочетаний, которые приурочены к началу/концу или середине слова. Например, в русском языке согласный й никогда не встречается в сочетании с последующим согласным в начале слова; иначе говоря, обнаруживая такое сочетание (ср. *шайба, тайга, балалайка, пойма, сайра, пайщик* и т.п.), мы всегда можем определить, что это не начало слова. Сочетания *йб, йг, йк* и пр. служат, таким образом, отрицательными пограничными сигналами. В отличие от этого, сочетания *взбр, взгл* и некоторые другие выступают положительными сигналами начала слова. В ряде работ говорят, вслед за Трубецким (1960), об афонематических пограничных сигналах, под которыми понимаются аллофоны, приуроченные к началу/концу или середине слова, ср. так называемое темное и светлое [I] в английском языке. Однако здесь возникают и теоретические трудности, связанные с зависимостью субфонологической единицы, аллофона, от грамматического контекста (Касевич, 1986), и определенные сомнения в том, что человек реально использует такого рода информацию в процессе перцептивной сегментации. К сегментно-фонологическим пограничным сигналам относят и синтагматическое «чередование» сингармонистических моделей, см. об этом с. 113 и сл.

В числе других признаков, способствующих сегментированию звучащего текста, называют увеличение длительности слов перед паузой и, шире, синтаксической границей, использование гортанной смычки (Lehiste, Wang, 1977; Lehiste, 1979a; 1979b; Harris e. a., 1981; Umeda, Quinn, 1981; Kreiman, 1982; Shen, 1993).

**Грамматические** пограничные сигналы представлены главным образом морфемами и их сочетаниями, которые регулярно употребляются в начале или конце слова. Естественно, что префиксы, особенно если они по своему фонологическому облику отличаются от морфем иных позиционных типов, служат указанием на начало слова, а флексии - на его конец.

Просодические и грамматические факторы действуют и опосредованно - путем указания на границы синтагм и предложений. Определенные интонационные типы, как известно, оформляя синтагмы и предложения, тем самым указывают на их

границы. Любая такая граница является одновременно границей слова. Точно так же существуют многочисленные и разнообразные грамматические средства, опираясь на которые можно определить границы именных и глагольных групп (синтагм), а также предложений. Например, в очень многих языках – тюркских, монгольских, тибето-бирманских и др. – наличие в тексте финитной формы глагола говорит о конце одного предложения и начале другого, а следовательно, и о словесной границе. Границы синтагм могут также «вычисляться», исходя из знания формально-грамматических валентностей слов различных классов и подклассов. Так, зная возможности распространения именной группы, а также то обстоятельство, что ее левая граница может задаваться предлогом, мы получаем полезную информацию о потенциальной сегментации предложения. Естественно, что в конкретном тексте появление сегментационных границ, определяемых на основании знания грамматических валентностей, зависит от грамматических правил, по которым функционируют соответствующие слова и конструкции данного языка. Например, в английском языке выбор между *but*, *when* или *then* не связан жестко с возможностями дальнейшего продолжения – после любого из этих слов может следовать, скажем, *I shot the cat*. В отличие от этого, в немецком языке аналогичный выбор между *aber*, *als* или *dann* предопределяет линейную структуру предложения: *aber ich habe die Katze erschossen, als ich die Katze erschossen habe* и *dann habe ich die Katze erschossen*. Соответственно для немецкого текста трудно ожидать наличия сегментирующей границы после указанных служебных слов, в то время как для английского текста ее вероятность более высока (Butcher, 1981, p. 58-59).

Многие авторы подчеркивают примат структурно-лингвистической информации над фонетической в обеспечении сегментации речевого сигнала, ср., например: «Вполне вероятно, что, скорее, знание лингвистической структуры дает слушающему возможность воспринимать границу в речевом сигнале, а не наличие [физического] сигнала позволяет воспринимать структуру» (Henderson, Nelms, 1980, p. 149). Вместе с тем в литературе отмечается, что паузы между предложениями (clauses) способствуют лучшему структурированию текста, в отличие от пауз между словами внутри предложений, которые, наоборот, могут деструктурировать речь (Goldman-Eisler, 1972; Reich, 1980).

Семантические факторы, способствующие сегментации, также относятся главным образом к сфере валентностей, на этот раз семантических, содержательных: поскольку существуют смысловые ограничения на сочетаемость слов, слушающий может предсказывать тот или иной тип распространения данной единицы и, следовательно, потенциальные словесные границы.

Валентности, как формально-грамматические, так и содержательные, - это основной источник распределения вероятностей переходов в речевой цепи. В общем случае наличие сегментационных границ должно ожидать там, где слушающий с наименьшей вероятностью способен предсказать появление последующих языковых единиц на базе знания предыдущих, т.е. в точках наибольшей неопределенности внутритекстового перехода того или иного типа (ср. Umeda, Quinn, 1981). Чем лучше знакомы слушающему тема текста, установки говорящего и т.п., тем легче ему осуществлять предсказание и тем, вероятно, реже будет он «расставлять» сегментационные границы.

Уже из этого очень беглого обзора видно, что проблема сегментации намного сложнее, чем ее часто представляют в литературе. Прежде всего, в процессах макросегментации следует различать стратегии, в большей/меньшей степени опирающиеся на информацию, непосредственно представленную в акустическом сигнале. Так, ударный слог, выделяющийся соответствующими акустическими параметрами, может - в условиях фиксированного ударения - непосредственно указывать на словесную границу. В отличие от этого, границы, устанавливаемые на основании валентностных предсказаний, имеют своим источником не параметры сигнала, а знание системных закономерностей, которым владеет носитель языка. Пограничные сигналы, в качестве которых выступают те или иные морфемы, их сочетания носят промежуточный характер: с одной стороны, они присутствуют в тексте, подлежащем сегментации, с другой же, необходимо привлечение высоких уровней языковой системы, чтобы идентифицировать данные участки текста в качестве соответствующих морфологических единиц.

Можно сказать, что во всех случаях сегментация неотделима от идентификации, различия же заключаются в том, сколько и каких именно уровней и компонентов языковой системы должно быть вовлечено в действие, чтобы идентифицированный элемент текста мог служить пограничным сигналом и, шире, участвовать в сегментации. Для ударения достаточно уста-

новить, что, скажем, данный мелодический перепад следует трактовать как акустический коррелят ударности, т.е. идентифицировать ударный слог, с автоматическими следствиями для задач сегментации. В случае морфемного пограничного сигнала мы должны не только идентифицировать некоторый фрагмент текста, как, например, /ах/ в русском языке, но и каким-то образом определить, что это - окончание существительного, тем самым фиксируя словесную границу «справа» от /ах/. Использование валентностных признаков требует, естественно, полной идентификации слов и словосочетаний, ибо только в этом случае к ним можно применить характеризующие их правила распространения, известные носителю языка.

Из предпринятого обзора следует также, что в сегментации необходимо различать два аспекта. С одной стороны, существует «принудительная» сегментация: слушающий должен выделить в тексте отрезки, которые он мог бы интерпретировать как синтактико-семантические структуры. Последние заведомо не могут быть опознаны в целом через отождествление с единицами словаря, ибо словарей такого рода в распоряжении слушающего нет и не может быть. В этом случае именно сегментация как таковая выдвигается на первый план, а выделяемыми отрезками, подлежащими структурированию, служат синтагмы и предложения.

С другой стороны, существует и сегментация на слова, которые выступают структурообразующими элементами синтагм, а «через» синтагмы и предложений. В этом случае сегментация и идентификация могут осуществляться в значительной степени параллельно, чему способствует возможность идентификации слов через их сличение с единицами словаря.

Наконец, мы видим также, сколь значительной может быть роль типологического своеобразия языка в определении стратегий, используемых в процессе сегментации; особенно ярко это проявилось при обсуждении проблем, связанных с местом словесной просодики в разбиении текста на слова.

Из всего этого следует, что проблемы сегментации должны исследоваться экспериментально на конкретном языковом материале. Один из таких возможных экспериментальных подходов представлен в последующих разделах настоящей главы.

**Планирование и проведение эксперимента.** В настоящем разделе описываются эксперименты, в ходе которых изучалось,

на какие отрезки членит звучащий текст слушающий.<sup>3</sup> Экспериментальный материал представлял собой запись читаемого текста длительностью от 5 до 8 минут в зависимости от модификации последнего (см. ниже).

Испытуемые (обычно 10 человек) слушали текст через громкоговоритель, им предлагалось полностью и с максимальной возможной точностью записать текст в бланках протоколов. При этом испытуемым разрешалось останавливать магнитофон, пользуясь клавишей «пауза», в любых местах, но не разрешалось перематывать ленту назад, т.е. не допускалось повторное прослушивание.

Каждый отрезок текста, записываемый в бланк протокола, испытуемый фиксировал с красной строки под отдельным номером. Одновременно экспериментатор отмечал границы отрезков по нажатиям клавиши «пауза» в собственном экземпляре текста. Моменты остановки магнитофона трактовались как границы между выделяемыми отрезками текста. Иначе говоря, материалы эксперимента должны были дать в наше распоряжение данные о том, где находятся границы перцептивных квантов, которые мы условимся называть перцептивными синтагмами, или просто синтагмами.<sup>4</sup> Далее следовало изучить, какова природа перцептивных синтагм и каковы те факторы, за счет которых осуществляется разбиение текста на синтагмы.

В качестве текста использовался отрывок из газетной статьи, посвященной ценообразованию; отрывок был прочитан специалистом в области экономики (что обеспечивало адекватную паузировку и интонационное оформление), аудиторами же выступали филологи – студенты, аспиранты и преподаватели С.-Петербургского университета. Семантическая чуждость воспринимаемого текста для аудиторов, надо полагать, снижала функциональную нагрузку семантических факторов, ответственных за сегментацию (и, шире, обеспечивающих восприятие текста), одновременно повышая нагрузку прочих факторов.

---

<sup>3</sup> Предварительные результаты некоторых экспериментов, относящихся к этой программе, уже публиковались (Касевич и др., 1990а; 1993). Близкий к нашему экспериментальный подход представлен в работах О.Ф.Кривновой (1986, 1989).

<sup>4</sup> Употребление термина «синтагма» здесь условно: реальные отрезки могли быть как неодносинтагменными (это проявлялось, в частности, и в возможности паузы внутри отрезка), так и частями синтагм (см. ниже).

В наших экспериментах, проведенных на материале русского языка, предъявление тестового материала осуществлялось в трех режимах: аудирование исходного, интактного текста (контроль, далее «эксперимент I»)<sup>5</sup>, аудирование того же текста, лишенного пауз (эксперимент II), и, наконец, аудирование беспаузального текста, подвергшегося одновременно семантической дезорганизации за счет «перетасовки» его синтагм (эксперимент III).

В эксперименте I была использована оригинальная запись текста в том виде, как его прочел диктор (1837 слогов, темп чтения - средний, постоянный - 4,4 слога/с). Для последующих экспериментов необходимо было создать на основе оригинала звучащие тексты, модифицированные в соответствии с требованиями конкретного исследования. С этой целью исходный звучащий текст был введен в ЭВМ (ЕС1036) через 10-разрядный АЦП с частотой дискретизации 10 кГц. Затем с использованием системы интерактивного анализа речевых сигналов (Венцов и др., 1990) в тексте были выделены все участки (синтагмы), ограниченные паузами, и из них составлен рабочий архив. Всего в нем оказалось 149 синтагм общей длительностью 3381104 отсчетов (338,11 с), а суммарная длительность пауз составила 79,18 с.

Запись беспаузального варианта текста была получена из его оцифрованной компьютерной версии с выводом через ЦАП. Удаление пауз привело к сокращению общей длительности текста на 19%.

Семантическая дезорганизация была достигнута, как сказано, путем изменения порядка следования синтагм на более или менее случайный. В качестве синтагмы в данном случае принимался отрезок текста между двумя паузами. При этом обращалось внимание на сохранность грамматической приемлемости результирующих последовательностей, равно как и на приемлемость их интонационного оформления. Последнее, впрочем, не могло быть достигнуто полностью, так как после удаления пауз некоторые участки речевой цепи отличались ощутимыми мелодико-энергетическими перепадами, которые по условиям нашего эксперимента не подлежали корректировке.

---

<sup>5</sup> Данные по интактному тексту уже использовались в указанных выше публикациях (Касевич и др., 1990а, 1993). В настоящей работе произведена лишь дополнительная их обработка по единой для всех трех экспериментов методике.

Приведем пример фрагмента экспериментального текста, полученного в результате семантической дезорганизации:

*«Здесь остро стоит вопрос а также эффективность разниц неоправданно высоких затрат, с оборота и платежей в бюджет из прибыли. Дотирование следует как основное энерго- и котельно-печное топливо на их конечной продукции.»*

Объемы исходного и беспаузального текстов были одинаковыми: 1837 слогов. Семантически дезорганизованный текст

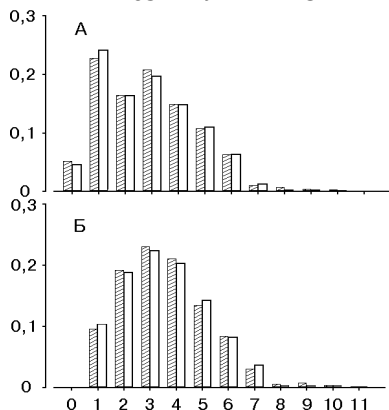


Рис. 3.1 Распределение грамматических (А) и фонетических (Б) слов в тексте по их длине.

Заштрихованные колонки - оригинальный текст; светлые - семантически дезорганизованный текст.

По оси абсцисс - длина слов в слогах; по оси ординат - относительная частота появления слова указанной длины.

имел несколько больший объем: 1963 слога. Частоты появления слов разной длины (в слогах) в этих текстах приведены на рис. 3.1, А.

Поскольку при обработке результатов экспериментов нас интересовали сопоставление записанного текста (результат его слухового восприятия) с исходным и анализ возможных стратегий фонетической интерпретации звучащего текста, вся статистическая обработка велась с учетом фонетических слов. Для этого односложные (и не образующие слога) союзы и предлоги (и, а, о, с, к, в, но, на, не, за, из, до, по, от, об, при, над, для, без) присоединялись к последующему слову, а частицы (бы) - к предшествующему. В исходном и беспаузальном текстах

таких фонетических слов было 511, а в дезорганизованном - 555. Распределение фонетических слов по их длине (в слогах) приведено на рис. 3.1, Б.

На рис. 3.2 представлено распределение длин синтагм (в фонетических словах), выделяемых диктором. Там же приведены аналогичные распределения для синтагм, на которые

разбил экспериментальный текст в его оригинальном графическом облике один из авторов работы. Эти синтагмы, которые мы называем здесь «лингвистическими», вычленились интуитивно, без сознательного следования тем или иным лингвистическим представлениям о природе синтагмы. Разбиение на «лингвистические» синтагмы было осуществлено двояким образом: при максимальном и минимальном дроблении текста. Очевидно, что дикторское членение текста оказывается промежуточным между указанными максимальным и минимальным вариантами.

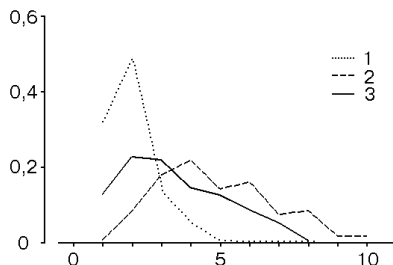


Рис. 3.2. Гистограммы длины «лингвистических» синтагм при максимальном (1) и минимальном (2) дроблении текста в сопоставлении с дикторским (3) вариантом членения.

По оси абсцисс - длина синтагмы в фонетических словах; по оси ординат - относительная частота встречаемости синтагм указанной длины.

**Роль пауз в сегментации текста.** Под паузой здесь и далее понимается (отличная от смычной фазы глухих согласных физически - по длительности - и функционально) фонетическая пауза, т.е. реальный перерыв в фонации. Таких фонетических пауз в тексте было 148; их распределение по длительности см. в табл. 3.1.

Таблица 3.1. Распределение по длительности пауз разного типа  
Длительность паузы, с

Тип паузы	0-0,5	0,5-1,0	1,0-1,5	1,5-2,0	2,0-2,5	2,5-3,0
«Точка»	—	18	10	—	2	1
«Запятая»	16	13				
Проч. граммат.	23	15	1			
Немотивир.	45	5	—	1		

Заполненные паузы хезитации в тексте отсутствовали (точнее, те немногие, что обнаружили, были очищены и приведены к нормальной длительности при подготовке текста к аудированию).

Заметим, однако, что фонетические паузы составляют лишь часть от общего числа акустических событий текста, которые с функциональной и перцептивной точек зрения могли бы интерпретироваться в качестве пауз. Выше уже говорилось, что эквивалентом паузы могут выступать тип мелодического перехода, изменение темпа и длительности «граничных» гласных и согласных и некоторые другие акустические события. Поэтому реальное число случаев совмещения перцептивных границ с паузами, понимаемыми не чисто фонетически, а функционально, должно - или, во всяком случае, может - оказаться заметно большим. Точная количественная оценка такого расхождения невозможна ввиду отсутствия данных по мелодическим, энергетическим и временным коррелятам паузы, экспертные же оценки ненадежны, так как трудно определить, как часто эксперт будет обозначать паузу, основываясь исключительно на семантико-грамматической информации.

В эксперименте I (запись интактного текста) в среднем 74,9% (от 66,4 до 88,9% у разных испытуемых) границ из фиксированных аудиторами путем остановки магнитофона совпали с паузами. Из этого можно было бы заключить, что пауза является достаточно сильным признаком для перцептивной сегментации, если бы не следующие достаточно серьезные соображения.

Чтобы использовать паузу как признак границы, надо иметь временной критерий (эталон), позволяющий отличить паузу-границу от паузы - фазы смычки глухого согласного. Если такой критерий существует, решение о наличии границы может быть принято только по истечении этого времени от момента начала паузы.

Здесь следует сделать оговорку методического характера. Как было уже сказано, экспериментатор вел параллельный протокол, проставляя в нем границы в момент нажатия испытуемым клавиши «пауза». Данные двух протоколов - испытуемого и экспериментатора - могли расходиться. Дело в том, что за время, проходящее от момента принятия решения о завершении перцептивной синтагмы до фактического выключения звука (остановки магнитофона), испытуемый мог - если все это время не приходилось на паузу - услышать некоторый дополнительный отрезок, являющийся началом следующей синтагмы. Длительность этого отрезка определяется временем реализации простой двигательной реакции (сгибание пальца) и задержкой выключения магнитофона. Конкретное его речевое

наполнение зависит от длительности паузы и темпа дикторского произнесения текста, а для беспаузального текста - только от темпа.

Испытуемый мог использовать разные стратегии работы с такого рода «промежуточным» отрезком текста: либо вообще игнорировать его, когда такой отрезок оказывался вне фокуса внимания испытуемого (впрочем, случаев пропуска «промежуточных отрезков» наблюдалось чрезвычайно мало), либо, как бы занеся отрезок в буфер памяти, затем приписать его к только что закончившейся синтагме, или, удержав, начать с него запись следующей синтагмы. Легко предположить, что выбор одной из двух последних стратегий определялся главным образом семантической и структурной уместностью отнесения отрезка «налево» или «направо». Отмечались также случаи лексического предсказания, когда испытуемый записывал то, чего еще не мог услышать, но смог предсказать на основании текущего анализа услышанного отрезка и уже образовавшегося у него подсловаря для данного текста. Например, в одном из опытов испытуемый остановил магнитофон, услышав отрезок РОЗ, записал же он синтагму, заканчивающуюся словами РОЗНИЧНЫХ ЦЕН.

Во всех этих случаях имела место своего рода редакция собственно перцептивных решений, частичное нарушение принципа, согласно которому должно было осуществляться сканирование текста только в одном направлении (и достижение чего должна была гарантировать инструкция, запрещающая перемотку ленты). Поскольку нас интересовал именно текущий перцептивный анализ звучащего текста и соответственно момент принятия решения относительно конца синтагмы, то расхождения между протоколами испытуемого и экспериментатора решались в пользу последнего. При статистической обработке итоговых протоколов приходилось учитывать возможность (или необходимость) систематического сдвига полученных границ влево.

С учетом такой задержки остановки магнитофона реальное выключение звука может приходиться (при не очень длительных паузах) на начало следующего слова. Но таких случаев в эксперименте I практически не отмечено (всего 3 случая на 1352 синтагмы). Это может быть либо результатом неточности в методике проведения эксперимента (экспериментатор не был ориентирован на максимально точную фиксацию в протоколе момента выключения звука; в пользу такого вывода говорит и тот

факт, что случаи остановки на следующем за границей синтагмы слове не отмечены даже при полном отсутствии паузы в этом месте), либо следствием того, что аудиторы при определении границы ориентировались не на факт появления перерыва в звучании, а на какое-то предшествующее этому событие. Событие, вызвавшее решение о конце синтагмы, могло возникнуть в

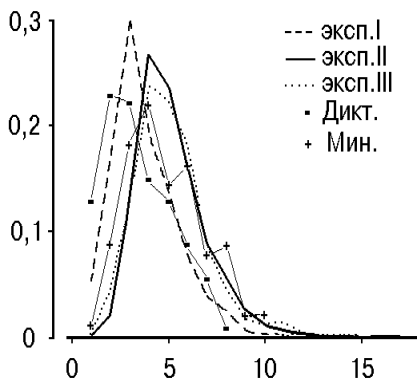


Рис. 3.3. Гистограммы распределения перцептивных синтагм по их длине, построенные по результатам записи интактного (I), беспаузального (II) и семантически дезорганизованного (III) текстов, в сравнении с распределением «лингвистических» синтагм при минимальном дроблении текста (Мин) и дикторской сегментацией (Дикт).

По оси абсцисс — длина синтагмы в фонетических словах; по оси ординат — относительная частота появления в записях синтагм соответствующей длины.

но при этом структура их распределений по длине оказывается достаточно близкой к той, что характеризует допустимые «лингвистические» синтагмы (рис. 3.3). При дальнейшем усложнении задачи — семантической дезорганизации беспаузального текста — происходит только вполне объяснимое снижение словесной разборчивости (с 91,6 до 85,3%).

Из этого можно заключить, что наличие/отсутствие (фонетических) пауз, даже в сопровождении частичной десемантизации, похоже, не влияет катастрофическим образом на резуль-

результате и текущего анализа просодических признаков в сигнале, и семантико-грамматического анализа. Наличие же паузы только облегчало задачу, подтверждая ожидаемое появление границы. Учет указанных проблем и определил выбор типа экспериментального материала и методики в экспериментах II и III.

Отметим прежде всего, что удаление из звучащего текста всех пауз описанного выше типа не привело к существенному ухудшению результатов работы испытуемых: словесная разборчивость снизилась всего на 1,5% (с 93,1 до 91,6%) в эксперименте II. Изменились параметры выделяемых синтагм (наиболее частыми стали синтагмы в 4-5 слов вместо 3 в эксперименте I),

таты восприятия звучащего текста. Близкие к этому данные, относящиеся, правда, лишь к перцептивной релевантности пауз, в литературе имеются. Так, М.Мирон и Э.Браун сообщают, что полное устранение пауз в речи нормального темпа приводит лишь к пренебрежимо малому (надо полагать, статистически незначимому) снижению разборчивости на 0,01% (Miron, Brown, 1971). Т.Казанав-Делюфе обнаружила, что во фразах типа *I'd like you to meet my aunt Tessie*, где последнее имя, в зависимости от типа интонации и отсутствия/наличия паузы, может быть либо приложением к существительному *aunt*, либо обращением, релевантность паузы для снятия неоднозначности проявляется лишь в условиях сглаживания «других компонентов просодической структуры» (Casanave-Delyfer, 1988).

Обратимся к вопросу о соотношении «весов» просодических и семантико-грамматических признаков. В эксперименте I границы выделенных аудиторами синтагм совпадают с паузами, сделанными диктором при чтении, в 74,9% случаев. Причем паузы разного типа обладают разной сегментирующей «силой» (табл. 3.2).

Таблица 3.2. Влияние семантико-грамматических и просодических факторов на перцептивную сегментацию естественного звучащего текста

Тип паузы	Кол-во в дикт. тексте	Из них ≤250 мс		Аудиторское членение для 10 исп.	
		кол-во	%	кол-во	%
«Точка»	32	—	—	315	98,4
«Запятая»	31	9	29	246	79,3
Проч.граммат.	43	13	30	274	63,7
Немотивир.	38	21	55,5	167	43,9

Естественно, что паузы, сопровождающие «точку» - абсолютный конец фразы - и имеющие наибольшую длительность, почти всегда используются аудиторами при сегментации. Завершение фразы, кроме того, всегда отмечено и интонационно. Естественно также, что «прочие грамматические» - реально на границах «малых» синтагм - и немотивированные паузы обладают наименьшим «весом»: из приведенных выше данных следует, что текст допускает достаточно разнообразное членение на «лингвистические» синтагмы, и, по-видимому, диктор и аудиторы широко пользовались такой возможностью.

Результат сопоставления количества (в процентах) пауз короче 250 мс с долей пауз, не использованных аудиторами для

сегментации, не исключает того, что 250 мс являются тем самым пороговым значением-эталоном, которым руководствуется слушающий при принятии решения о фиксировании границы (возможно, это связано именно с тем, что физические паузы меньшей длительности могут соответствовать фазе смычки глухого согласного; мы воздержимся здесь от обсуждения временных аспектов навязанной испытуемому дополнительной процедуры - остановок магнитофона).

Удаление пауз из звучащего текста (эксперимент II) привело к резкому сокращению (до 37,7%, от 26,5 до 54,4% у разных испытуемых) числа случаев, когда граница синтагмы приходится на паузу (на место в тексте, где должна была начаться пауза) (табл. 3.3).

Поскольку физические паузы в тексте отсутствовали, совпадение границ синтагм с «началом» несуществующей паузы можно было бы отнести на счет влияния «функциональных» пауз и принять относительное число таких совпадений за «вес» этих пауз.

Таблица 3.3. Влияние семантико-грамматических и просодических факторов на перцептивную сегментацию беспаузального звучащего текста

Тип паузы	Кол-во в дикт. тексте	Из них		Аудиторское членение для 9 исп.	
		≤250 мс			
		кол-во	%	кол-во	%
«Точка»	32	—	—	116	40,3
«Запятая»	31	9	29	96	34,4
Проч.граммат.	43	13	30	103	26,6
Немотивир.	38	21	55,5	40	11,7

Факт совпадения границы синтагмы с «паузой», если только экспериментатор не допустил ошибки, отмечая это место на слух, означает, что магнитофон остановился точно в момент окончания слова, предшествовавшего паузе в исходном тексте. Принимая во внимание сказанное выше о задержках выключения, момент принятия решения о конце синтагмы (или, скорее, принятия решения о том, что уже услышанного текста вполне достаточно для его правильной записи и что еще не звучавшая часть не окажет существенного влияния на результат) должен приходиться не на конец слова, а совпадать с некоторым событием «внутри» этого слова. Если допустить, как уже говорилось,

что в процессе сегментации испытуемые пользуются «функциональной паузой», выраженной, например, мелодическими изменениями, то естественно считать, что такое событие будет приурочено к ударному слогу (гласной), на котором эти изменения обычно реализуются.

В тех случаях, когда ударный слог оказывается последним в слове, после которого останавливался магнитофон, предполагаемые просодические изменения вряд ли были использованы испытуемым, так как решение о конце синтагмы он мог принять еще до того, как их услышал. В этих случаях испытуемый, по всей вероятности, ориентировался на семантико-грамматические признаки, и их следовало бы исключить из оценки «веса» «функциональных» пауз.

С другой стороны, в ряде случаев до остановки магнитофона испытуемый «успевал» услышать 1-2 слога следующего за удаленной паузой слова. С известной долей уверенности можно считать, что в этих случаях испытуемый действительно принимал решение о границе синтагмы в момент окончания слова, предшествовавшего паузе в оригинальном тексте, и, следовательно, мог использовать для этого все просодические признаки, реализованные во время его произнесения.

С учетом приведенных соображений («функциональные») паузы, которые можно считать признаком синтагматической границы, в беспausalном и семантически дезорганизованном текстах (в последнем случае на них приходится 47,5% границ, от 27 до 66,7% у разных испытуемых) имеют разный «вес» (табл. 3.4).

В целом «вес» пауз («функциональных») как признака границы не очень высок, но он еще и не постоянен, а изменяется в зависимости от условий, в которых испытуемому при-

Таблица 3.4. Относительная частота совпадения границ перцептивных синтагм с «паузами» разного типа, проценты

Тип паузы	Эксперимент II	Эксперимент III
«Точка»	54,9	57,7
«Запятая»	39,4	46,0
Проч.граммат.	24,3	23,7
Немотивир.	14,0	28,0

ходится работать с сигналом. Когда беспausalный текст был семантически дезорганизован, снижалась роль семантико-грамматических факторов в сегментации текста, и испытуемые чаще обращались к иным (мелодическим и пр.) средствам. Сильнее всего это сказалось на «немотивированных» паузах, ко-

торые, не будучи обусловлены семантико-грамматической структурой текста, все же возникали в дикторском произнесении этого текста.

Вопрос о том, какие именно просодические средства используются при такой сегментации, требует специального анализа и исследования.

### **Факторы, определяющие объем перцептивных синтагм.**

На рис. 3.3 для каждого из трех экспериментов приведены сводные результаты по всем испытуемым. Результаты эксперимента I обнаруживают хорошее совпадение с распределением длительностей синтагм в дикторском прочтении текста. Результаты экспериментов II и III для данного аспекта практически совпадают: распределения длин синтагм (в фонетических словах) оказываются почти одинаковыми, и количество синтагм длиной 9 и менее фонетических слов (предельный объем оперативной памяти по Миллеру) составляет 97,5 и 96,1% соответственно. Но с дикторским членением на синтагмы они не совпадают, а оказываются ближе к распределению «лингвистических» синтагм при минимальном дроблении текста.

Таблица 3.5. Распределение испытуемых по числу вычлняемых синтагм

Эксперим.	170-150	150-130	130-110	110-90	90-70	70-50
I	3	3	3	—	1	—
II	—	1	3	3	1	1
III	—	2	6	1	0	1

При этом во всех экспериментах ярко проявилась индивидуализация стратегий, используемых испытуемыми при перцептивной сегментации текста. Одни испытуемые склонялись к членению текста на более, другие - на менее крупные синтагмы. Минимальное число синтагм, на которые испытуемые сегментировали текст, было 89, 69 и 70 в экспериментах I, II и III соответственно, максимальное - 169, 140 и 143 (см. табл. 3.5). Бросается в глаза практически постоянное, примерно 1:2, отношение минимального и максимального числа синтагм в протоколах аудиторов (при том, что в разных экспериментах участвовали не совпадающие по составу бригады испытуемых).<sup>6</sup> Отметим, что

<sup>6</sup> Заметим, что то же соотношение получено в аналогичных экспериментах на материале китайского и японского языков (Касевич и др., 1993).

пословной записью текста испытуемые пользовались только в эксперименте I (5,3% всех синтагм), хотя такая стратегия могла оказаться предпочтительной при заданной инструкции (как можно точнее записать услышанный текст).<sup>7</sup> В каждом эксперименте обнаруживались различия между испытуемыми, что позволило выделить характерные группы, главное различие между которыми состоит в степени выраженности максимума на распределении выделяемых ими синтагм по их длине и его положению на шкале длин синтагм (в числе слов).

Эти различия, можно думать, имеют прямое отношение к объему оперативной памяти. Предположим, что длина прослушанной части текста (части текста между двумя последовательными нажатиями клавиши «пауза», границами) непосредственно зависит от объема оперативной памяти, характеризующего знаменитым «магическим числом» Миллера  $7 \pm 2$ . Тогда следует ожидать, что длина синтагмы в основном не будет превосходить максимальный объем оперативной памяти, т.е. 9 единиц (в данном случае - слов), и доля синтагм, отвечающих этому условию, может служить характеристикой поведения испытуемых данной группы.

Соответствующие данные представлены ниже, и они подтверждают сделанное предположение.

Материалы табл. 3.6 показывают также соотношение используемого объема оперативной памяти со словесной разборчивостью, под которой понимается отношение числа правильно записанных слов к общему числу слов в тексте.

Результаты эксперимента I, кажется, дают основание предположить, что перегрузка оперативной памяти (увеличение доли синтагм, объем которых превышает 9 слов) приводит к снижению разборчивости, т.е. к ухудшению качества работы испытуемого. Такой вывод, казалось бы, подтверждается и результатом работы одного испытуемого (группа I) в эксперименте III: работая со сверхдлинными синтагмами (средняя длина 8,26 при диапазоне от 6,8 до 10 слов), он плохо справился с заданием (доля правильно записанного текста - 74,9%). Более того, у этого испытуемого обнаруживается четкая связь между изменением длины синтагмы в ходе опыта и изменением разборчивости (рис. 3.4).

---

<sup>7</sup> Вероятно, это связано и с тем, что в отсутствие достаточно продолжительной паузы практически невозможно остановить магнитофон точно на границе между двумя словами.

Таблица 3.6. Зависимость между объемом используемой оперативной памяти и точностью записи звучащего текста

Экспер.Группа	Число испыт.	Доля синт. с №9, %	Словесн. разб., %
I	I	1	94,4
	II	5	99,2
	III	4	100,0
II	I	1	82,6
	II	2	95,9
	III	2	97,6
	IV	4	100,0
III	I	1	62,8
	II	2	98,7
	III	4	99,2
	IV	3	100,0

Однако один из испытуемых (группа I) в эксперименте II тоже работал с длинными синтагмами (средняя длина 7,7 при диапазоне от 6,3 до 9 слов), но достиг при этом высшей точности работы (разборчивость 96,1%), а испытуемые группы IV в экспериментах II и III при отсутствии какой-либо

перегрузки оперативной памяти обнаруживают существенно худшее качество работы (меньшую разборчивость).

Видимо, объем вычленяемых синтагм испытуемые выбирают без оглядки на возможности оперативной памяти, а из ка-

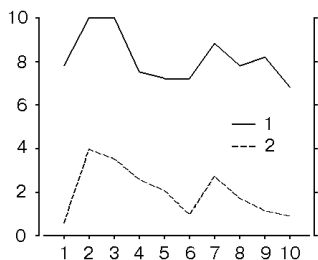


Рис. 3.4. Изменение стратегии членения текста на перцептивные синтагмы и точности работы у исп. М.

1 — средняя длина синтагмы в фонетических словах; 2 — словесная разборчивость.

По оси абсцисс — «время» от начала опыта в десятых долях от полного числа синтагм в записях данного испытуемого; по оси ординат — число фонетических слов (слева) и процент правильно записанных слов (справа).

ких-то других соображений (лингвистической их значимости, скорости работы и т.д.). Но в то же время они способны под- сознательно оценивать успешность своей работы (точность записи текста, словесную разборчивость), и если оценка оказывается неудовлетворительной (в частности, и из-за перегрузки оперативной памяти), испытуемые выбирают синтагмы меньшей длины. Некоторые из них (группа IV) делают это заранее, не надеясь на свои способности, другие изменяют тактику в ходе опыта, о чем уже говорилось выше.

По-видимому, объем опера- тивной памяти данного инди-

видуума ставит предел объему перцептивной синтагмы. Однако, во-первых, перцептивная синтагма может быть и меньше, если действие того или иного семантического, грамматического, фонетического фактора диктует фиксирование сегментирующей границы еще до насыщения оперативной памяти. В частности, даже при минимально дробном членении использованного нами текста длина синтагм не превышает 10 слов, а число синтагм с 9-ю и менее словами составляет 98,1%. Диктор же членил текст на синтагмы не длиннее 8 слов.

Во-вторых, этот предел не абсолютен, ибо перцептивная синтагма может и превышать объем оперативной памяти в терминах слов, если уровень избыточности текста позволяет переходить на код более высоких и соответственно более крупных единиц - вплоть до предложений. Именно эти «ножницы» по существу и вызывают к жизни вопрос о механизмах сегментации в качестве особой проблемы - иначе он сводился бы полностью к вопросу об объеме оперативной памяти.

Говоря об избыточности текста как регуляторе выбора кода перцептивных единиц, мы имели в виду всю совокупность языковых и неязыковых факторов в их взаимодействии. О переплетении этих факторов также говорят наши данные: так, сегментируя беспаузный текст на синтагмы большой длины, диктор из группы I в эксперименте II достиг словесной разборчивости в 96,1%, а при работе с семантически дезорганизованным текстом диктору из группы I в эксперименте III удалось достичь лишь разборчивости в 74,9%.

Назовем «коэффициентом согласия» показатель, который определяется отношением числа границ, совпавших у 60 и более процентов испытуемых, к общему числу перцептивных границ, т.е. границ между перцептивными синтагмами в тексте. Для эксперимента I коэффициент согласия принимает значение 0,31, для эксперимента II - 0,14, для эксперимента III - 0,23. Детальное распределение границ в тексте с точки зрения их совпадения у разных испытуемых представлено на гистограммах рис. 3.5.

Оценка согласованности действий испытуемых может интересовать нас в качестве показателя выраженности в звучащем тексте признаков, используемых при сегментации.<sup>8</sup> Если в

---

<sup>8</sup> Следует только учитывать, что, оперируя перцептивными синтагмами разного объема, испытуемые, которые предпочитают менее дробное членение, могут просто не использовать часть признаков текста, «пропуская» их и тем самым уменьшая величину коэффициента согласия.

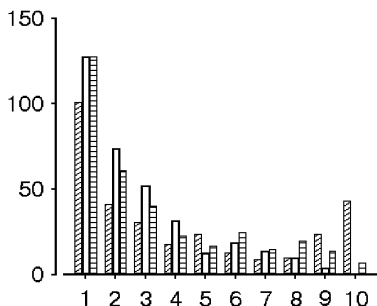


Рис. 3.5. Число синтагматических границ, совпадающих у нескольких испытуемых в разных экспериментах.

Колонки с косой штриховкой - эксперимент I, светлые - эксперимент II, с горизонтальной штриховкой - эксперимент III.

По оси абсцисс - число испытуемых с совпадающим положением синтагматических границ; по оси ординат - количество совпадающих границ у данного числа испытуемых.

меток в разных местах.

Существенная степень «навязанности» способа сегментации проявляется в близости распределения длин синтагм в дикторском варианте прочтения текста и в эксперименте I (рис. 3.3).

В эксперименте I аудиторы чаще всего членили текст на синтагмы длиной 3 слова, а в экспериментах II и III - на синтагмы в 4-5 слов (рис. 3.3). Приведенные данные выглядят, в известной степени, парадоксальными. При ухудшении условий восприятия - устранении пауз (в эксперименте III сопровождаемом еще и семантической дезорганизацией) - происходит некоторое увеличение объема перцептивной синтагмы, чего, казалось бы, естественнее ожидать при благоприятных условиях, таких, как привычность тематики, минимальность уровня помех и т.п.

Эти результаты, однако, поддаются, как представляется, объяснению. И устранение пауз, и семантическую дезорганизацию текста можно рассматривать как операции, приводящие к

данном месте текста могут быть использованы паузы - сами по себе или в комплексе с семантико-грамматическими признаками, то следует ожидать большей согласованности в постановке меток. Если преобладающими были семантико-грамматические признаки и правила, то можно ожидать и большую, и меньшую согласованность в постановке меток сегментации в зависимости от того, какими признаками и правилами пользовался конкретный испытуемый. Вполне возможно, однако, что и при малой согласованности, полученной по формальному правилу несовпадения меток у разных испытуемых, последние тем не менее работали по одному и тому же правилу, но структура текста допускала при этом постановку

частичному снятию ограничений на возможные способы структурирования текста, что прямо связано с его членением. Действительно, в указанных экспериментальных условиях меньше факторов, обуславливающих то, а не иное членение, в отсутствие же ограничений человеку свойственно стремиться к максимальному укрупнению единиц восприятия. Именно это, можно думать, и отражают наши данные.

Полной ясности здесь, конечно, нет уже потому, что отсутствует определенность в понимании верхнего предела укрупнения единиц восприятия. С одной стороны, сама кратковременная память, согласно существующим представлениям, разбивается на ряд временных интервалов, с которыми соотносят не только способность удерживать запоминаемый материал в течение определенного времени, но и тот или иной объем этого материала; так, говорят, что объем первичной памяти, соответствующей одному из интервалов памяти кратковременной, составляет 2-3 слова (Психологический словарь, с. 250). С другой стороны, число Миллера  $7 \pm 2$  характеризует как будто бы кратковременную (оперативную) память как таковую, без подразделения на интервалы, и при этом относится, скорее всего, к последовательности независимых единиц.

Когда единицы не являются независимыми, объем запоминаемого материала может увеличиваться, см., например, (Клаас, 1966), но не вполне ясно, где в таких случаях мы имеем дело только лишь с предсказуемостью единиц, входящих в цепочку, за счет данного распределения вероятностей, а где взаимозависимость членов цепочки переходит в новое качество, когда цепочка трактуется в

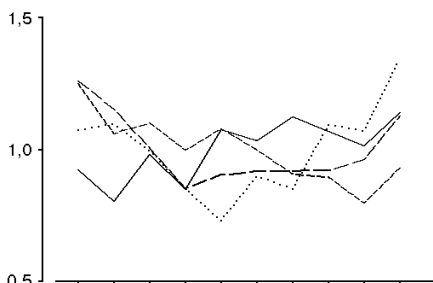


Рис. 3.6. Динамика изменения длины перцептивных синтагм в ходе опыта по членению интактного текста (эксперимент I).

Данные разных групп испытуемых, объединенных общностью стратегии. По оси абсцисс — «время» от начала опыта в десятичных долях от полного числа синтагм, записанных данным испытуемым; по оси ординат — отношение средней длины синтагмы на данном интервале к средней длине синтагмы в опыте.

целом как единица кода более высокого уровня - например, не цепочка фонем или слогов, а слово. Без разрешения такого рода общих проблем трудно рассчитывать на исчерпывающее объяснение результатов, подобных полученным в нашем эксперименте.

Чуть ниже мы вернемся к обсуждению этого вопроса, сейчас же обратимся к еще одной группе данных, полученных в эксперименте, - к динамике изменения объема перцептивных синтагм (рис. 3.6-3.8).

Для эксперимента I в целом характерно некоторое снижение объема вычленяемых синтагм по завершении примерно первой четверти текста. Такая форма экспериментальной кривой достаточно естественно объясняется сменой стратегии: стремление оперировать максимально крупными единицами восприятия уступает место более дробному квантованию, когда испытуемый убеждается в известной затрудненности текста и при этом в равной степени не готов ни жертвовать уровнем разборчивости (о разборчивости см. ниже), ни «перенапрягать» оперативную память.<sup>9</sup>

Для эксперимента II у разных аудиторов отмечается значительная индивидуализация стратегий перцептивной сегментации текста в ее динамике. Лишь для 3 из 10 испытуемых экспериментальная кривая характеризуется общим нисходящим движением, отличным, впрочем, от полученного в эксперименте I. Примерно половина испытуемых обнаруживает тенденцию к заметному, подчас скачкообразному, изменению объема синтагм в процессе аудирования. По-видимому, здесь в большей степени, чем при сравнении средних объемов синтагм и позиции перцептивных границ, проявляется относительная небезразличность для перцептивной сегментации наличия фонетических пауз. Их отсутствие в определенной степени дезорганизует процесс восприятия за счет эффекта

---

<sup>9</sup> Здесь хотелось бы привести данные, полученные в аналогичном эксперименте на материале других языков (Касевич и др., 1993), где ярко проявляется, с одной стороны, независимость типа стратегии от языка, с другой же - необходимость учета языковой типологии для адекватной интерпретации экспериментальных результатов. Описанная выше форма кривой была практически идентичной для русского, китайского и японского языков. Однако она резко и непредсказуемо отличалась на материале японского языка, если объем японских синтагм измерялся в слогах. Типичная форма кривой «возвращалась», когда измерение объема японских синтагм осуществлялось в терминах мор (или слов), что вполне объяснимо, ибо мора - «в ущерб» слогу - выступает основной единицей фонетической организации речевого потока в японском языке (Касевич и др., 1990б).

обманутого ожидания: текущее предсказание, основанное на совокупности признаков, «обещает» наличие паузы, и отсутствие последней, неподтверждение текущего прогноза вносит элемент хаоса в перцептивный процесс.

В эксперименте III один из испытуемых обнаруживает трудно поддающуюся объяснению

тенденцию к оперированию сверхкрупными синтагмами, объемом до 10 слов на участках, соответствующих отрезку от исхода первой минуты до 2,5 минут от начала звучания (1/10 - 3/10 текста от его начала). Остальные испытуемые образуют две группы, внутри которых динамика вычленения перцептивных синтагм достаточно близка. В первой группе отсутствует выраженная тенденция к возрастанию или уменьшению объема перцептивных синтагм в пространстве текста, аудиотексты второй группы характеризуются преимущественно нисходящим рисунком экспериментальной кривой при высоком (9 слов у всех испытуемых) ее начале.

Как можно видеть, в целом условия эксперимента III применительно к данному, динамическому, аспекту привели, скорее, к более унифицированным стратегиям испытуемых. И здесь мы должны вернуться к обсуждению тезиса о роли ограничений,

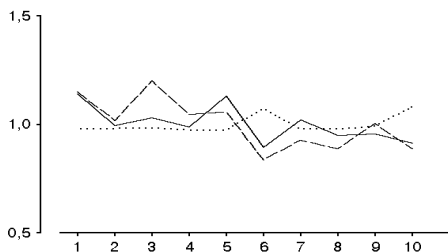


Рис. 3.7. Изменение длины перцептивной синтагмы в ходе опыта по членению для беспаузального текста (эксперимент II). Обозначения те же, что на рис. 3.6.

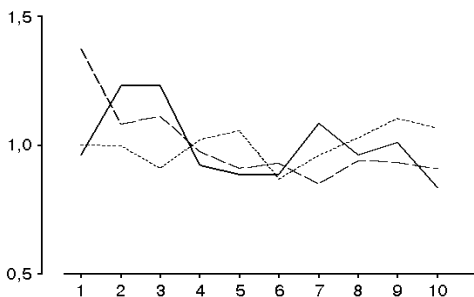


Рис. 3.8. Изменение длины перцептивной синтагмы в ходе опыта по членению для беспаузального семантически дезорганизованного текста (эксперимент III). Обозначения те же, что на рис. 3.6.

налагаемых различными параметрами текста на способ его членения и, шире, структурирования. В эксперименте II ситуация восприятия оказывается более конфликтной, нежели в эксперименте III. В первом из них действуют более или менее согласованно все факторы, опираясь на которые человек осуществляет перцептивное членение текста, - кроме отсутствующих фонетических пауз, что, как уже говорилось, создает эффект обманутого ожидания. Во втором семантическая рассогласованность синтагм оказывается «в согласии» с отсутствием пауз, дезорганизация семантики находит свою параллель в дезорганизации паузального аспекта просодики, и реакции испытуемых в этих условиях должны приближаться к максимальному использованию объема оперативной памяти в процессе сегментации текста.

Вместе с тем неверным было бы заключить, что один лишь объем оперативной памяти как таковой выступает в указанных условиях в качестве определяющего фактора (этому не соответствуют, по-видимому, и количественные показатели). Воспринимая текст, даже лишенный семантической связности, испытуемые все же структурируют его в соответствии с доступными им «оставшимися» параметрами - грамматическими характеристиками, включая формальные валентности слов, и внутри-синтагменными лексико-семантическими валентностями. Причем эти факторы могут оказываться даже более сильными, чем акустические, связанные с фонологической идентичностью гласных и согласных, когда между ними возникает конфликт. Так, в одном из протоколов отмечается запись

**ЦЕНА ДОЛЖНА УЧИТЫВАТЬ ВСЮ ОБЩЕСТВЕННО  
НЕОБХОДИМУЮ**

вместо

**ЦЕНА ДОЛЖНА УЧИТЫВАТЬ ВСЕ ОБЩЕСТВЕННО  
НЕОБХОДИМЫЕ,**

при том, что акустическая определенность гласной [э] не вызывает сомнения. Такого рода замен, когда грамматическое или лексико-грамматическое решение «подавляет» фонетико-фонологическое, в протоколах испытуемых не столь уж мало.

Заметим, что в аналогичных экспериментах на материале бурятского языка (данные С.П.Маккуойд еще не опубликованы) испытуемые нередко заменяли в своих протоколах заимствованные русские слова их исконно бурятскими эквивалентами. Иначе говоря, объектом «текущего слежения» для испытуемых и здесь выступал прежде всего смысл сообщения.

**«Поверхностное» восприятие.** Из представленных выше более или менее частных результатов и их обсуждения можно вывести, среди прочих, некоторые следствия достаточно общего характера. Восприятие речи может обеспечиваться действием языкового механизма, которое носит по преимуществу формальный характер. Если считать, что критерием успешного восприятия служит воспроизведение текста с заданной точностью, то - при определенной сохранности сегментно-фонологических средств, от чего мы сейчас в значительной степени отвлекаемся, - некоторого минимума просодических признаков и грамматико-семантической структурированности достаточно, чтобы текст был воспринят. Хотя испытуемые, что явствует из самоотчетов, вполне сознают семантическую ущербность экспериментального текста (впрочем, уже исходный, контрольный текст оценивался испытуемыми как «не очень понятный»), они способны достаточно точно воспроизводить его в записи. Словесная разборчивость при записи семантически дезорганизованного текста оказалась не намного меньше, чем при записи исходного текста (85,3 и 93,1% соответственно). Для сравнения: в беспаузальном тексте она составила 90,4%, а с учетом фонетически допустимых замен (т.е. там, где акустический сигнал допускал различную интерпретацию) - 91,6%.

Достаточно высокая разборчивость, полученная на материале семантически дезорганизованного текста, и подтверждает тезис о преимущественно формальном характере перцептивного механизма. Вероятно, в обычных условиях восприятие речи - его, условно, «первый слой» - также осуществляется столь же формально. Когда для слушающего реакция на слышимый текст не является личностно релевантной, он может ограничиться - и нередко ограничивается - формальным, поверхностным, «чисто языковым» восприятием. Модель мира, с необходимостью вовлекаемая человеком в любой процесс восприятия, здесь используется лишь постольку, поскольку она отражена в формальных структурах языка.<sup>10</sup>

В связи со сказанным возникает еще один вопрос, до сих пор нами не обсуждавшийся. Испытуемый получал инструкцию максимально точно записывать те синтагмы, на которые он чле-

---

<sup>10</sup> Иначе говоря, не следует абсолютизировать в целом справедливый тезис о необходимости достаточно полного объема внеязыковых знаний для адекватного восприятия текста, ср.: «...мы не можем ... понимать язык без понимания вещей, на которые направлено наше внимание» (Полани, 1985, с. 149).

нил текст, между тем есть основания предполагать, что в естественной ситуации производится текущее перекодирование поверхностно-текстовых структур в глубинно-семантические. При восприятии семантически дезорганизованного текста сама проблема семантической структуры очевидным образом специфична: либо испытуемый «смиряется» с некоторой мозаичной семантикой, когда границы смысловых структур совпадают с границами синтагм, либо он просто отказывается от собственно семантического кодирования, не идя дальше поверхностных структур (впрочем, как мы видели выше, говорить о полном исключении семантического аспекта в стратегиях испытуемых эксперимента III едва ли возможно).

Здесь мы сталкиваемся с проблемой, от специального обсуждения которой вынуждены отказаться, ограничившись очень кратким ее рассмотрением. Проблема заключается в следующем: каковы признаки перцептивных синтагм с лингвистической точки зрения и какой языковой механизм, кроме самих процедур сегментации и идентификации, следует постулировать, чтобы объяснить именно данный, а не какой-либо иной, набор лингвистических признаков?

Указанная проблема обсуждается в целом ряде работ, опубликованных главным образом в 60 - 70-е годы. Некоторые авторы, по-видимому, считают основной единицей перцептивной сегментации предложение (clause), ср. выше, с. 71. Однако наши материалы, как нетрудно видеть, плохо согласуются с этим тезисом - и даже априори трудно ожидать, что предложение, которое может обладать весьма сложной внутренней структурой и неопределенно большой длительностью, является идеальным кандидатом на роль типичной перцептивной синтагмы.

Согласно самой распространенной точке зрения, целесообразно исходить из того, что перцептивная синтагма (в нашей терминологии) «кончается» там, где слушающему наиболее удобно перекодировать данную составляющую высказывания в некоторую абстрактную целостную единицу, освобождая при этом соответствующий процессор для восприятия последующей синтагмы (Fodor *е. а.*, 1974). Поскольку большинство исследователей проблемы исходило из генеративистских представлений о синтаксисе и семантике высказывания, в экспериментах проверялась гипотеза, согласно которой границы между перцептивными синтагмами должны отвечать переходам от одной глубинной структуры к другой, см., например, (Bever *е. а.*,

1969). Дж.Кэрролл и М.Таненхаус резонно замечают по этому поводу, что уже сама посылка сомнительна: так, например, в высказывании наподобие *Fleeing was John's favorite strategy...* составляющей *fleeing* соответствует самостоятельная глубинная структура, однако выделение *fleeing* в качестве отдельной перцептивной синтагмы крайне маловероятно (Carroll, Tanenhaus, 1978). Не вполне подтвердилась в эксперименте и гипотеза, согласно которой перцептивные синтагмы совпадают с непосредственно составляющими - поверхностно-синтаксическими единицами формального синтаксиса (Chapin e. a., 1972).

Кэрролл и Таненхаус, подобно ряду авторов, исходят из того, что граница перцептивной синтагмы должна создавать условия для перекодирования последней в глубинную структуру. Отсюда они делают вывод, что субъективная граница вероятнее там, где синтагма в большей степени завершена, - содержит, в частности, информацию о предикате, субъекте, объекте, позволяя тем самым перекодирование в отдельную пропозицию (с этим в известной степени перекликается трактовка единиц, выделяемых по интонационному контуру, как «sense groups», см., например, Cruttenden, 1970). Гипотеза Кэрролла и Таненхауса в целом подтверждается тщательно проведенными экспериментами (Carroll, Tanenhaus, 1978). Вместе с тем экспериментальные материалы, приведенные в этой главе, показывают, что не следует абсолютизировать роль семантических характеристик воспринимаемого текста, несмотря на всю их важность. Слушающий осуществляет непрерывное слежение за семантическим развитием воспринимаемого текста, по мере накопления информации осуществляя свертывание («архивирование») семантики до вида, удобного для хранения (отдельный вопрос - природа типичных семантических структур, реализующихся при этом; вряд ли это просто набор пропозиций, ср. Flores d'Arcais, 1974). Однако в то же время слушающий не может не проводить текущего анализа формально-синтаксической структуры текста, без чего невозможен и анализ семантический. С этой точки зрения, как уже отмечалось, особенно важен учет валентностей слов в составе высказываний: насыщение валентностей во многом предопределяет прохождение границ перцептивной синтагмы. Соответственно границы вполне реальны и там, где не приходится говорить о «полных» пропозициях.

Подобные стратегии особенно типичны для «поверхностного» восприятия с его обедненной семантикой. Ставя вопрос более широко, следует допустить, что не существует одной-

единственной содержательной и/или формальной лингвистической характеристики перцептивных синтагм – они изменяются в зависимости от избранной стратегии восприятия.

Как уже не раз отмечалось, процесс восприятия и его «поверхностный» вариант – «формально-языковое» восприятие – с необходимостью носит многоканальный характер. Конфликтные ситуации, о которых говорилось выше, как раз и отражают существование и взаимодействие таких каналов. Возможно, наиболее удачным образом эти аспекты отражает разрабатываемая в последние годы модель параллельных распределенных процессоров, или коннекционистская модель (Parallel ..., 1988).

Те различные каналы получения информации о структуре текста, которые, судя по всему, использовали наши испытуемые в процессе перцептивной сегментации (паузировка, интонация, грамматические, лексико-семантические признаки), тоже есть основание отнести к работе самостоятельных процессоров, которые либо поддерживают друг друга, либо входят в конфликт; следует только уточнить, что как взаимная активация, так и эффект подавления связаны не с процессорами как таковыми, а с их определенными состояниями, в той или иной степени совместимыми или несовместимыми.

Лишь детальный учет всех факторов и стоящих за ними процессоров в их взаимодействии может дать убедительную картину работы перцептивных механизмов – в частности, и для решения проблемы сегментации речи.

Резюмируем кратко основные результаты предпринятого нами анализа перцептивной сегментации.

1. Пауза как перерыв фонации (отличный от смычной фазы глухих согласных) в типичном случае служит не столько в качестве непосредственной опоры перцептивной сегментации, сколько как сигнал, подтверждающий релевантность акустических событий, потенциально интерпретируемых как метка синтагматической границы и во времени предшествующих паузе.

2. Для перцептивной сегментации весьма характерна индивидуальная вариабельность стратегий, в частности, проявляющаяся в выделении синтагм существенно разного объема у разных носителей языка.

3. Объем перцептивной синтагмы зависит от ряда факторов. В целом слушающий следует рисунку членения текста, предлагаемому говорящим, при этом он «связан» объемом собственной оперативной памяти. Сокращение длины выделяе-

мой синтагмы может вызываться невозможностью получить требуемую точность восприятия при объединении двух синтагм в одну из-за перегрузки оперативной памяти и одновременно невозможностью распространения данной синтагмы за счет соседней без потери связности семантического квантования. Увеличение длины выделяемой синтагмы происходит тогда, когда благодаря предсказуемости их составляющих «крупные» синтагмы становятся самостоятельными единицами особых кодов.

4. В отсутствие языковых ограничений объем перцептивных синтагм стремится к совпадению с максимально допустимым объемом оперативной памяти. Типичны стратегии, когда слушающий начинает с попытки оперирования максимальными (для него) синтагмами, в дальнейшем же перцептивное квантование стабилизируется на некоторой средней отметке. Максимизирующие стратегии могут иметь своей оборотной стороной снижение точности восприятия.

5. Даже очень значительная степень дезорганизации текста, формальной и семантической, позволяет сохранить некоторый приемлемый уровень адекватности восприятия; это свидетельствует о наличии у человека механизма «поверхностного» восприятия речи.

## РОЛЬ РИТМИЧЕСКИХ СТРУКТУР В СЕГМЕНТАЦИИ И ИДЕНТИФИКАЦИИ ЯЗЫКОВЫХ ЕДИНИЦ\*\*

Различные аспекты речевой ритмики изучаются, как известно, давно. Давно используется в фонетике понятие «ритмическая структура», хотя эта категория не вполне ясным образом соотносится с понятием речевого ритма: дело в том, что традиционная ритмическая структура – чаще всего, характеристика слова, в то время как речевой ритм обнаруживается на отрезке не меньшем, нежели высказывание. Положение, согласно которому ритмическая структура есть единица, регулярное воспроизведение которой в тексте и создает речевой ритм, нуждается в оговорках (см. об этом ниже, а также в монографии, посвященной ударению и тону: Касевич и др., 1990б).

---

\*\* Исследования, отраженные в настоящем разделе, выполнены при участии канд. филол. наук Е.М.Шабельниковой (С.-Петербургский университет).

Ритмика всегда исследовалась главным образом применительно к речепроизводству, что вполне естественно: речепроизводство (порождение речи) «поверхностно» реализуется как двигательная активность, а последняя у человека (и у всех живых организмов) неизбежно в той или иной степени ритмизована. Как пишет Кристина Цалтьери (Zaltieri, 1987), ритм составляет «всегда сопровождающий нас, неустрашимый фон всего нашего существования (*shadowy, omnipresent background of our existence*)».

Определения ритма, вообще говоря, не существует. Обычно полагают, что ритм реализуется как более или менее регулярное воспроизведение некоторых структур. И. Фонадь выделяет два аспекта ритмизованности: повторяемость того или иного события, признака и т.п. и чередование моментов напряжения и спада (Fonagy, 1982). В сущности, это можно рассматривать как некоторую интерпретацию нестрогого определения, приведенного выше: аспект повторяемости соотносится с воспроизведением как структуры в целом, так и составляющих ее элементов, а чередование моментов напряжения и спада можно толковать как отражение иерархичности структуры; коль скоро структура (с неизбежностью) иерархична, продвижение от «зачина» структуры к ее ядру в известном смысле создает напряжение, в то время как после достижения ядра наступает спад.

Как Фонадь, так и Цалтьери (равно и другие авторы), обращают внимание на то, что аспект воспроизводимости, повторяемости, глубоко присущей ритму, оборачивается повышением избыточности текста.<sup>11</sup> Ведь один из основных способов повышения избыточности заключается именно в многократном повторении информации.

Обращаясь к понятию избыточности, мы фактически от характеристики процессов порождения речи переходим к процессам ее восприятия, т.е. становимся главным образом на точку зрения приема информации. В то же время возникает, по-видимому, вопрос: несет ли вообще ритмическая упорядоченность речи информацию и, если да, то какую? В принципе мож-

---

<sup>11</sup> В сущности эта же мысль содержится в ярком описании ритма у М.М.Бахтина: «... действительное, роковое, рискованное абсолютное будущее преодолевается ритмом, преодолевается самая граница между прошлым и будущим (и настоящим, конечно) в пользу прошлого; смысловое будущее как бы растворяется в прошлом и настоящем...» (Бахтин, 1979, с. 103).

но было бы посчитать, что уже сама по себе «неизбежность» ритмизации, о которой говорилось выше, делает ритм афункциональным и, следовательно, не несущим какой бы то ни было информации.

До сих пор мы оставляли в стороне вопрос о том, какова же природа воспроизводимых структур, являющихся ритмообразующими элементами текста. Не предпринимая здесь подробного обсуждения этого вопроса (Антипова, 1984; Касевич, 1988; Касевич и др., 1990б), ограничимся краткой формулировкой нашего общего подхода. Мы исходим из существования системы речевых ритмов, которые одновременно реализуются в тексте: текст ритмически упорядочен применительно к слогам, словам, синтагмам, высказываниям, сверхфразовым единствам, которые и выступают разноуровневыми ритмообразующими структурами (элементами). Почти априори ясно, что такого рода разные ритмы должны отличаться как качественно (по способу проявления ритмизации), так и количественно (по степени ритмизации).

Последнее обстоятельство, быть может, особенно важно. В чем бы и каким бы образом ни проявлялась ритмизация текста, очевидно, что человек нормально не изъясняется ритмической прозой. Иначе говоря, как общая степень ритмизованности текста, так и выраженность в текстах разных ритмов могут варьировать, вероятно, в достаточно широких пределах.

Нас будет интересовать главным образом «словесный ритм» в акцентных языках типа русского, т.е. ритм, создаваемый воспроизведением слов как ритмических структур, в тексте на таком языке, в котором существует разноместное словесное ударение.

Ограничение сферы рассмотрения «словесным» ритмом относительно упрощает задачу, поставленную выше: ответить на вопрос, несет ли ритм - в данном случае словесный - информацию и какую именно (а также, как эта информация передается).

Уже приходилось (Касевич и др., 1990б) в этой связи цитировать Дж.Аллена и С.Хокинса: «Слушающий, похоже, предвидит, когда [в речевой цепи] возникает ударение, и сосредоточивает внимание на этих моментах. [...] Ритмическая структура тем самым обеспечивает полезную перцептивную избыточность в речи путем снижения неопределенности в определении моментов, когда могут произойти важные (с артикуляторной точки зрения) события» (Allen, Hawkins, 1980, p. 229). Однако едва ли

функцию ритмизованности (на уровне словесного строения высказывания, текста) следует сводить к облегчению детектирования артикуляторных событий: слушающего интересуют не артикуляторные события сами по себе и даже не ударение как таковое - ему необходимо вычлениить и идентифицировать значимые языковые единицы, в данном случае слова.

Именно с этой точки зрения ритм не афункционален, он несет определенную информацию, которая заключается в том, что «воспроизведение» слова - разумеется, другого, следующего слова в составе высказывания (текста) - делает появление этого слова предсказуемым. Имплицитное владение знанием о типичном акцентном рисунке высказывания способствует вычлениению слов из речевой цепи.

Что же представляет собой такой типичный акцентный рисунок? В сущности, речь должна идти о некотором пределе в распределении ударений в пространстве текста, к которому (пределу) стремится любой экземпляр последнего. По имеющимся данным (Liberman, Prince, 1977), в разных акцентных языках типичный акцентный рисунок в пределе составляет воспроизведение ударного слога через каждые 3-4 безударных. Иначе говоря, межакцентные интервалы стремятся к уравниванию, характеризуясь при этом протяженностью в 3-4 слога.

Представление о регулярных межакцентных интервалах как о некотором пределе позволяет, думается, наиболее адекватным образом подходить к описанию реальных речевых процессов в сфере как порождения, так и восприятия речи. В этой связи необходимо хотя бы вкратце остановиться на такой важной проблеме, как соотношение ритмического и синтактико-семантического членения текста.

В различных моделях синтеза речи (не очень удачно называемых моделями «текст-речь») используется метод порождения просодических структур на основе одной лишь или преимущественно синтаксической информации, хотя существуют и модели, где просодика генерируется главным образом как последовательность некоторых ритмических структур, более или менее независимых от структур синтаксических. Наилучший же результат достигается там, где оба подхода совмещаются (Bailly, 1989). И это не удивительно. Следует учитывать как относительную независимость ритмического и синтактико-семантического членения текста, так и их неизбежное взаимодействие, когда просодическая структура выступает как результат своего рода сплетения двух (а на самом деле - и более) динамических процессов. Ре-

чевые структуры со всеми признаками, включая и фонологические, в известном смысле налагаются на ритмические «волны», порождаемые - назовем это так - независимым генератором сложных ритмов. Последние, имеющие отчетливые биологические истоки, не являются тем не менее исключительно врожденными и тем самым независимыми от данного языка, они обладают конкретно-языковой спецификой, которая должна составить предмет специального изучения.

Чем более выражена скоррелированность синтактико-семантического членения с ритмическим - чем совершеннее эвритмия, тем, в принципе, выше уровень избыточности текста и соответственно легче его восприятие.

Вместе с тем применительно к слову картина более сложна. Идеальному (предельному) распределению ударений, о котором говорилось выше, препятствует объективная разноразность слов и разноместность ударения. Относительная «отдельность» ритма по отношению к членению на слова хорошо видна в поэзии, где границы стоп не столь уж редко не скоррелированы с границами слов. Однако именно в поэзии реализуется максимально равномерное распределение ударений.

Сказанное выше подводит нас к такой постановке вопроса: является ли сама по себе регуляризация ритма фактором, способствующим восприятию речи? Именно материал поэзии может дать ответ на этот вопрос.

Мы предприняли экспериментальное изучение означенной проблемы. Текстовым материалом экспериментов послужили два стихотворения: одно, написанное белым стихом, «Осада Памбы» А.К. Толстого (первые 34 строки), другое - рифмованным, «В парке» Дм.Кедрина. Стихотворения были трансформированы таким образом, чтобы нарушить их ритмическую структуру. В результате каждое стихотворение использовалось в двух вариантах: исходном (оригинальном) и «прозаизированном». Варианты были абсолютно тождественны с точки зрения лексики, морфологии и синтаксиса как такового - включали тот же набор словоформ в тех же синтаксических отношениях; отличались они исключительно порядком слов (т.е. линейным синтаксисом и отчасти коммуникативной организацией - актуальным членением), за счет чего и достигался эффект деритмизации или, точнее, снижения степени ритмизации. Приведем примеры, демонстрирующие тип использованных текстов:

Девять лет дон Педро Годец,  
По прозванию Лев Кастильи,  
Осаждает замок Памбу,  
Молоком одним питаюсь.

Дон Педро Годец, по прозванию Лев Кастильи, девять лет осаждает замок Памбу, питаюсь одним молоком.

Старинной купаленки шаткий настил,  
Бродя у пруда, я ночью потрогал.  
Под этими липами Пушкин грустил,  
На этой скамеечке сиживал Гоголь.

Я потрогал ночью, бродя у пруда, шаткий настил старинной купаленки. Под этими липами грустил Пушкин, Гоголь сиживал на этой скамеечке.

Оба варианта каждого текста были записаны на магнитную ленту в исполнении диктора, имеющего навык как записи в студийных условиях, так и публичного чтения стихов. Далее, чтобы затруднить аудиторам доступ к информации, связанной с высокими языковыми уровнями, и тем самым повысить функциональную нагрузку фонетических признаков, на запись обоих вариантов двух экспериментальных текстов был наложен белый шум при соотношении сигнал/шум, равном 0 дБ.

В качестве аудиторов выступали 10 человек - преподаватели, сотрудники и студенты СПбГУ (8 женщин и 2 мужчин, в возрасте от 18 до 38 лет). Тексты предъявлялись аудиторам через громкоговорители в следующей последовательности: прозаический вариант стихотворения «Осада Памбы» (далее - текст 1П), его оригинальная версия (текст 1С), прозаический вариант стихотворения «В парке» (текст 2П), его оригинальная версия (текст 2С) - все на фоне шума, а затем, в той же последовательности, контрольные варианты.

Приведем основные количественные характеристики экспериментальных текстов. Текст 1 состоял из 112 слов (270 слогов). Текст 2 насчитывал 89 слов (229 слогов). Таким образом, 10 аудиторов прослушали в шуме по 402 слова (998 слогов) каждый.

Исходные тексты и их прозаические версии были разбиты на синтагмы, после которых при прослушивании текстов делались паузы по 20 с, позволяющие аудиторам записать предъявленный материал. В стихотворных текстах за синтагму принималась каждая стихотворная строка (34 и 20 для текстов 1 и 2 соответственно). В прозаических вариантах выделялось несколько отличающееся число синтагм - 30 и 21 соответственно.

Испытуемым предлагалось наиболее полно записать тексты, не допуская в них пропусков. В инструкции оговаривалась возможность записи «чего-то максимально похожего по звучанию, пусть даже в ущерб смыслу». Разрешалось неразличимые части текста передавать с помощью последовательностей типа «та-та-та» с оставлением пробелов между «словами» и обозначением ударения.

Выше был поставлен вопрос о том, зависит ли успешность восприятия текста от степени его ритмизации, которая в текстах стихотворных заведомо выше, чем в их прозаических аналогах. Успешность восприятия естественно толковать как уровень разборчивости – словесной и слоговой. Именно эти характеристики экспериментальных текстов прежде всего устанавливались в наших опытах (табл. 3.7). Данные табл. 3.7 свидетельствуют о том, что повышение степени ритмизации по крайней мере может приводить к возрастанию разборчивости: для текста 1 словесная разборчивость оригинальной версии превышает величину того же показателя для прозаизированного варианта на 23,8%, а слоговая разборчивость – на 21,3%. Что же касается текста 2, то здесь превышение значительно менее заметное: 6,1% для словесной разборчивости и почти совпадающие величины (40,4 и 39,8%) для слоговой.

Приходится признать, что мы затрудняемся дать объяснение различию в данных, относящихся к двум экспериментальным текстам. В принципе можно было бы ожидать для текста 2 более ярко выраженного

эффекта ритма, ибо в стихотворении Дм. Кедрина, помимо метра, представлен дополнительный ритмизирующий фактор – рифма. Тем не менее следует предположить существование неизвестного нам дополнительного фактора (см., впрочем, ниже, с. 107-108), который нивелирует восприятие двух версий экспериментального текста. В дальнейшем все наше обсуждение будет базироваться на данных, относящихся к тексту 1, хотя данные по тексту 2 будут также фигурировать в приводимых таблицах и графиках.

Таблица 3.7. Разборчивость экспериментальных текстов, проценты

Текст	Словесная	Слоговая
1С	49,2	52,5
1П	25,4	31,2
2С	39,0	40,4
2П	32,9	39,8

Каков же реальный механизм повышения разборчивости текста при возрастании степени его ритмической упорядоченности? На какие именно параметры текста опирается человек, когда он демонстрирует лучшее восприятие более ритмизированного текста по сравнению с менее ритмизованным?

Будем исходить из того, что восприятие речи предполагает осуществление двух сопряженных операций: сегментации и идентификации. Сегментация во многом, хотя и не единственно, определяется обращением к ударению. Это относится, конечно, к сегментации на слова. Как хорошо известно, число ударений соответствует числу фонетических слов, которые в русском языке в большинстве случаев (хотя статистика неизвестна) совпадают с флективными словоформами. Ударения тем самым могут выступать метками, способствующими сегментации.

Иначе говоря, при восприятии речи полезной процедурой оказывается поиск ударений. Если ударения распределены в пространстве текста (высказывания, синтагмы) более или менее случайным образом (полностью случайного, неупорядоченного распределения, впрочем, практически не бывает), то детектировать их сложнее, чем в ситуации упорядоченного, регулярного размещения.

Уместно ожидать, отсюда, что в стихотворном варианте будет более высоким процент слов с позицией ударения, адекватной по отношению к предъявленному тексту. Именно такова картина, полученная на материале текста 1 (табл. 3.8).

Таблица 3.8. Адекватное обнаружение позиции ударения, проценты

	Без учета Текстзаударной части слова	Без учета предударной части слова
1С	85,1	77,9
1П	63,2	58,8
2С	80,0	72,2
2П	73,4	65,2

Как можно видеть, в табл. 3.8 отдельно показаны данные по сохранности места ударения, когда не принимается во внимание заударная часть слова, и также данные при условии «игнорирования» предударной части; поясним причины, делающие целесообразным такой двойной счет.

Упорядочение ритмики интересующего нас типа реализуется как выравнивание межакцентных интервалов - промежутков между двумя ближайшими ударениями; есть основания полагать, что длительность таких интервалов определяется

главным образом в терминах слогов (см., например, работу А.Катлер (Cutler, 1980)). Но межакцентный интервал неоднороден: он складывается из заударной части  $n$ -го слова и предударной части  $(n+1)$ -го слова. Как собственно фонетически, так и перцептивно указанные слагаемые межакцентного интервала неравноценны. С перцептивной точки зрения предударная часть слова, можно утверждать, обладает большим «весом». Особая релевантность ударения, ударного слога порождает феномен «ожидания ударения» (Касевич, 1977), т.е. большей сфокусированности субъекта перцептивного процесса на обработке того участка, который предшествует ударению, после чего возможна известная релаксация; это, по-видимому, хорошо соотносится с чередованием моментов напряжения и спада, которое И.Фонадь вводит в определение ритма (см. об этом выше).

Положение о перцептивной неравновесности предударной и заударной частей слова можно интерпретировать и как вопрос о том, что является точкой отсчета для определения позиции ударения: начало слова или конец слова? Известно, что фонологические правила фиксирования ударения могут основываться на обоих вариантах: в одних языках ударение приурочено к первому слогу слова (чешский, венгерский и др.), в других - к последнему (например, фарси, французский) или предпоследнему (польский и т.п.). Но как будто бы нет данных о том, к каким стратегиям прибегает человек, определяя место словесного ударения в тексте при восприятии речи, когда ударение разноместно и надо обнаружить его местонахождение. Выдвижение начала слова на роль точки отсчета, относительно которой определяется позиция ударения, кажется предпочтительным даже вне связи с тем, что говорилось выше о соотношении предударной и заударной частей в языках типа русского: начало первого слова задается началом фонации или паузой, разделяющей высказывания, синтагмы в пределах текста. Можно было бы вспомнить в этом контексте и о явлении инициальной интенсивности, которая характеризует первый слог слова и тем самым маркирует его начало. Иначе говоря, существуют, в принципе, независимые маркеры начала слова.

Может показаться, вероятно, что наше рассуждение впадает в порочный круг: с одной стороны, мы задаемся вопросом о сегментирующей, квантующей роли словесного ударения в восприятии речи, т.е. вопросом об обнаружении границ слов на основании информации о местонахождении ударного слога, с другой же - обсуждаем возможность перцептивного фиксиро-

вания ударения на базе информации о границах слова. Но если здесь и есть ситуация порочного круга, то она присутствует, скорее, в самом процессе восприятия речи, который определенно не следует законам формальной логики. Разные модули перцептивного механизма параллельно обрабатывают разные параметры сигнала, постоянно обмениваясь информацией (см. гл. II). Поэтому интересующая нас проблема соотношения позиции ударения и границ слова оборачивается двумя аспектами: скоррелированность указанных двух параметров и их взаимные веса. Последний аспект и вводит элемент иерархии, поскольку приписывание данному параметру большего веса по отношению к другому как раз и означает, что в типичной для речи ситуации неопределенности предпочтительнее основываться на этом параметре (обычно как более доступном) для определения некоторого другого (ср. гл. II). Из этого не следует, что обладающий меньшим весом параметр (фактор) полностью лишен перцептивной самостоятельности и обнаружим лишь по показаниям связанного с ним иерархически старшего. Иерархически старший параметр может, в свою очередь, проверяться иерархически младшим, и если информация о последнем, оказавшись в условиях конкретного коммуникативного акта более доступной, «пересиливает», то вполне возможен рекурсивный возврат к обладающему системно большим весом фактору с целью корректировки информации о нем.

В нашем контексте мы исходим из иерархического старшинства фактора ударности, от которого в процессе речевосприятия, как мы полагаем, зависит (вероятностное) обнаружение словесных границ, прежде всего начала слова. Одновременно мы принимаем, что существуют собственные признаки начала слова, пусть относительно слабо выраженные, с которыми может коррелировать информация об ударности.

Данные табл. 3.8 как будто бы дают косвенные основания для подтверждения такого рода представлений. В протоколах наших испытуемых более высокий процент слов с адекватно определенным местом ударения обнаруживается тогда, когда не учитывается сохранность числа слогов в заударной части слов, по сравнению с таким методом подсчета, когда не принимается во внимание сохранность числа слогов в предупредной части. Это хорошо согласуется с ранее опубликованными данными по восприятию фильтрованной речи (Касевич и др., 1990а).

Таким образом, в утверждение о том, что речевой ритм может создаваться выравниванием межакцентных интервалов, следует, вероятно, внести поправку: нарушения равенства межакцентных интервалов могут не являться таковыми с перцептивной точки зрения, если они возникают за счет разного числа слогов, принадлежащих заударным частям соседних слов.

В языках с разноместным ударением типа русского поиск границы между словами осуществляется в пределах межакцентного интервала плюс один из ударных слогов, «окаймляющих» этот интервал. Наши данные на материале текста 1 показывают, что в условиях повышенной степени ритмизации, когда межакцентные интервалы выравнены, вероятность адекватного установления словесных границ также повышается: для стихотворного текста процент «разборчивости» словесных границ на 22% выше по сравнению с тем же показателем для прозаического варианта текста.

Для нашего материала объяснение может оказаться довольно простым. Использувавшееся в эксперименте стихотворение А.К.Толстого «Осада Памбы» написано хореическим размером (с пиррихиями). Хорей - наиболее естественный из метров, наиболее органичный. Именно хореический характер носят спонтанные квазистихотворения у детей раннего возраста («экикики» К.Чуковского). Уже эта органичность делает хореически организованный текст наиболее благоприятным для восприятия. Но главное, как можно думать, заключается в том, что при использовании хорея - двусложного размера - межакцентный интервал составляет ровно один слог. Следовательно, абсолютное число вариантов, между которыми надо выбрать при определении места прохождения словесной границы, здесь предельно сужено. Тем самым облегчается задача определения точек словоделения.<sup>12</sup>

Не исключено, что именно в этом кроется тот неизвестный фактор, который обусловил различия в данных, относящихся к текстам 1 и 2. Стихотворение Дм. Кедрина «В парке» написано амфибрахией - трехсложным размером, где межакцентные интервалы двусложны. Соответственно можно высказать пред-

---

<sup>12</sup> Мы оставляем в стороне вопрос обо всех прочих источниках информации, используемых для сегментации на слова (пограничные сигналы в виде типичных аналутных фонемосочетаний, пограничные морфемы и т.д. и т.п.).

положение (кажущееся довольно естественным) о том, что с увеличением межакцентного интервала вероятность обнаружения словесных границ снижается даже при высокой степени ритмической упорядоченности текста.

Итак, данные по обнаружению словесных границ показывают, что уровень разборчивости связан с доступностью информации о словесных границах. В предыдущих публикациях, основанных на материале изучения фильтрованной речи (Касевич и др., 1984), была эмпирически выведена количественная закономерность, ставящая словесную разборчивость в зависимость от «разборчивости» словесных границ:

$$R = - 0,0064967 \times B^2 + 1,913 \times B - 49,404,$$

где R – словесная разборчивость, B – «разборчивость» словесных границ. Данные, полученные в настоящем эксперименте, уже не укладываются в ту же формулу, хотя достаточно определенная тенденция к установлению аналогичного соотношения между словесной разборчивостью и процентом случаев адекватного обнаружения межсловесных границ наблюдается.

По-видимому, для разного материала (в экспериментах с фильтрованной речью изучалось восприятие изолированных словосочетаний разного объема), разных экспериментальных режимов соотношение разборчивости и адекватного детектирования словесных границ может характеризоваться не совпадающими численными выражениями. Мы не будем выводить отдельную формулу, связывающую разборчивость и успешность сегментации применительно к данному материалу и экспериментальному режиму.

До сих пор мы говорили о том влиянии ритмики на разборчивость, которое связано с лучшими условиями для проявления сегментирующей функции словесного ударения в ситуации высокой ритмической упорядоченности. Но «хорошая» ритмика может и более непосредственно улучшать разборчивость. В предшествующем изложении речь шла об ударениях как некоторых абстрактных метках, расставленных в пространстве высказывания. Но каждое ударение в русском языке выделяет соответствующий слог, а в нем – гласную. При повышенной ритмизованности разборчивость таких гласных должна (или, во всяком случае, может) возрастать в силу того, что четкий ритмический рисунок фокусирует внимание воспринимающе-

го речь субъекта на точках максимальной выделенности, совпадающих с ударными гласными.

Данные, полученные на материале текста 1, подтверждают сформулированное предположение; разборчивость ударных гласных составила:

текст	1С	71,7%
	1П	47,8%
	2С	60,8%
	2П	59,0%.

Иначе говоря, при восприятии стихотворного текста разборчивость увеличивается на 23,9% по сравнению с прозаизированной версией текста.

Как мы видели, ритмическая упорядоченность текста влияет на словесную разборчивость опосредованно, через облегчение установления межсловных границ, т.е. способствуя сегментации, и непосредственно, через улучшение разборчивости (ударных) гласных, т.е. создавая благоприятные условия для идентификации. Существует еще один параметр, который в известном смысле охватывает оба аспекта - сегментацию и идентификацию. Это акцентный контур слова. Опознание акцентного контура - самостоятельного признака слова, одной из просодий последнего - связано с сегментацией по той простой причине, что «начало» и «конец» акцентного контура совпадают, естественно, с началом и концом слова. В то же время определение акцентного контура имеет прямое отношение к идентификации слова, ибо существуют перцептивные стратегии грубой классификации слов по их акцентным контурам (что в информационно насыщенной среде может оказаться достаточным для идентификации слова без дальнейшего анализа плана выражения).

В протоколах испытуемых мы изучали также «разборчивость» акцентных контуров, т.е. процент случаев, когда оставался сохранным акцентный контур слова безотносительно к сохранности самого слова; иначе говоря, адекватная запись слова, запись другого слова с тем же акцентным контуром или структура типа та-та-та с адекватно проставленным ударением и верным числом слогов не различались. «Разборчивость» акцентных контуров в текстах разных типов оказалась следующей:

текст	1С	48,8%
	1П	16,0%
	2С	36,0%
	2П	27,6%.

«Разборчивость» акцентных контуров очень низка для прозаического текста и сравнительно высока для стихотворного, разница составляет 32,8% на материале текста 1 (между прочим, здесь мы впервые замечаем более или менее заметное расхождение и для текста 2).

Данные по адекватному определению акцентных контуров, возможно, особенно показательны и в то же время представляют известные трудности для интерпретации. С одной стороны, как сказано, акцентный контур – интегральная характеристика фонетического облика слова; определив правильно акцентный контур, мы тем самым установили длину слова в слогах и место ударения, а значит, отнесли слово к соответствующему акцентному классу. Знаменательно, что в наших данных процент словесной разборчивости и процент случаев сохранения акцентного контура для стихов практически совпадают, а для прозы близки.

С другой стороны, не вполне ясно соотношение природы акцентных контуров в стихотворной и прозаической речи и отсюда их сравнительная роль в двух типах текстов, различающихся степенью ритмической упорядоченности. Дело в том, что для прозы минимальной ритмической единицей (не считая слога) является фонетическое слово, которое в языках типа русского совпадает с ритмической структурой (ритмической группой, тактом) и характеризуется акцентным контуром. Для стиха же минимальной ритмической единицей выступает стопа, которая сплошь и рядом не совпадает с фонетическим словом, и, таким образом, возникает почва для известной интерференции собственно языкового и стихотворного ритмов, где последнему принадлежит главенствующая роль. И тем не менее, как демонстрируют наши данные, в идентичных условиях стихотворная речь обеспечивает лучшее членение текста на собственно языковые ритмические единицы – фонетические слова и лучшее опознание их акцентных контуров. Вопрос этот заслуживает дальнейшего теоретического и экспериментального изучения.

Наконец, в наших материалах мы исследовали также динамику разборчивости различных единиц и признаков на протяжении прозаических версий текстов. При этом мы основывались на следующих соображениях. Прозаические тексты, использовавшиеся в эксперименте, обладали внутренней неоднородностью с точки зрения ритмической упорядоченности. По понятным причинам невозможно было преобразовать стихотворения таким образом, чтобы ни одна синтагма итогового текста

не сохраняла бы исходного размера; например, заключительные (из использовавшихся в эксперименте) строки стихотворения «Осада Памбы» оставались практически без изменений: «... целых девять лет не ели, ровно ничего не ели, кроме только молока». Таким образом, в прозаических текстах обнаруживались «островки стиха». Можно было предположить, что на границах таких «островков» уровень разборчивости склонен колебаться – скорее всего, падать: если улучшение разборчивости в условиях более высокой ритмической упорядоченности объясняется прежде всего избыточностью, которую создает регулярное воспроизведение ритмических структур, то изменение ритма, пусть даже «в лучшую сторону», должно, наоборот, ухудшать разборчивость, коль скоро при этом нарушается «ритмическая инерция» восприятия.

С целью проверки сформулированного предположения мы определяли разборчивость различных единиц и признаков отдельно для синтагм, не сохранивших исходного размера, и синтагм, выступающих как «островки стиха».

Экспериментальные материалы не подтверждают сделанное предположение. Разборчивость не изменяется сколько-нибудь заметно в зависимости от того, исследуются ли синтагмы с нейтральным ритмическим рисунком «обычной» прозы или же «островки стиха». Возможно, объяснение следует искать в небольшой (за немногими исключениями) протяженности синтагм, сохранивших исходный ритм, и их случайном расположении в тексте. «Ритмическая инерция» создается, можно думать, характером начала текста, а для перцептивной перестройки требуются ритмически ярко выраженные отрезки, способные составить конкуренцию сложившейся инерции. Кроме того, перцептивно «сильнее» снижение ритмической упорядоченности, нежели ее возрастание.

Между тем динамика разборчивости языковых единиц и признаков в пространстве текста отнюдь не характеризуется устойчивостью. Для всех без исключения параметров обычны резкие колебания разборчивости вдоль временной оси. На рис. 3.9 в качестве иллюстрации показан типичный график, отражающий динамику разборчивости ударных гласных.

Какой-либо закономерности в колебаниях разборчивости обнаружить не удастся. Вполне возможно, что флуктуации разборчивости связаны лишь с естественной вариабельностью речевого сигнала при постоянстве шумового фона. Здесь есть непрямая методическая проблема, которая, как нам представляется,

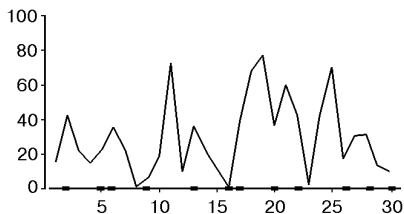


Рис. 3.9. Динамика разборчивости ударных гласных при восприятии текста 1П.

Зачерненный прямоугольник на оси абсцисс — указатель положения «островка стиха».

По оси абсцисс — текущий номер синтагмы; по оси ординат — разборчивость, проценты.

заслуживает внимания экспериментаторов, работающих в области речевых исследований.

Заключая раздел, необходимо признать, что вопросов в нем содержится едва ли не больше, нежели ответов. Вместе с тем мы получили экспериментальное подтверждение по крайней мере двум обстоятельствам: 1) возможна прямая зависимость разборчивости — т.е., в конечном счете, восприятия речи — от уровня ритмической организа-

ции текста; 2) в процессе речевосприятия имеет место сложное взаимодействие процедур сегментации — установления границ, в нашем случае межсловных, — и идентификации языковых единиц, слов, причем для реализации обеих процедур используется информация о ритмическом строении текста.

Как в парадигматике, так и в синтагматике всегда наблюдается сочетание системного и асистемного; один из синтагматических аспектов этого сочетания — соотношение ритмического и аритмического. Баланс между ними — важное прагматическое свойство текста. Возможна (и необходима) типология текстов по их ритмическим признакам. Необходимо также дальнейшее изучение сложного взаимопереплетения и взаимодействия различных единиц и признаков речи с ее ритмическими характеристиками с точки зрения как порождения, так и восприятия речи.

## РОЛЬ СИНГАРМОНИСТИЧЕСКИХ МОДЕЛЕЙ В СЕГМЕНТАЦИИ И ИДЕНТИФИКАЦИИ\*\*\*

Весьма обширная литература по сингармонизму исследует эту проблему почти исключительно с точки зрения внутрисистемных соотношений: изучается, каким образом наличие и тип сингармонизма сказываются на связях и отношениях, действительных для фонем в рамках системы и в пределах словоформы. Значительно меньше внимания уделяется деятельности аспекту, т.е. функционированию правил сингармонизма в процессах речевой деятельности. Такая недооценка динамической стороны языковых явлений (типичная для большинства направлений лингвистики) фактически приводит к игнорированию самого существенного в языковом механизме, который никоим образом не является «вещью в себе и для себя», будучи предназначен для передачи информации в ходе коммуникативного акта.

Как неоднократно отмечалось, правила, оперирующие языковыми единицами, естественным образом различаются в зависимости от сферы применения – речепроизводства или речевосприятия. Правила сингармонизма также целесообразно рассматривать отдельно, применительно к порождению и восприятию речи.

С точки зрения речепроизводства (порождения речи) сущность сингармонизма заключается, главным образом, в том, что сингармонический язык предоставляет в распоряжение говорящего ограниченное число моделей для фонологической организации словоформ; эти модели устроены иерархически: в составе словоформы выделяется элемент, позиция и/или морфологическая и фонологическая природа которого обуславливает признаки других элементов (слогов, морфем) в рамках той же словоформы. Под упомянутыми признаками имеются в виду дифференциальные признаки гласных, иногда и согласных, входящих в экспонент словоформы.

Иначе говоря, сингармонизм можно рассматривать в качестве морфофонологического аналога механизма согласования в морфосинтаксических процессах. Наличие согласования в грамматике ведет к ограничению числа моделей, действитель-

---

\*\*\* В основу настоящего раздела положены экспериментальные данные, полученные канд. филол. наук М.И. Момунбаевой (Киргизский университет).

ных для синтагм определенного типа. Например, если прилагательное-определение согласуется с существительным-определяемым в роде, числе и падеже, как это имеет место в русском языке, то определительная словоформа должна воспроизводить категориальные признаки определяемой. Аналогично и для сингармонизма: если правила сингармонизма заключаются в том, что гласные аффиксов уподобляются гласной корня по ряду, то это и означает, что в фонологии аффиксов воспроизводятся категориальные признаки фонологии корня. В обоих случаях имеем ограничения на сочетаемость: морфологических признаков в рамках синтаксического единства и фонологических признаков в рамках единства морфологического соответственно. Подобно тому как говорящий при наличии морфосинтаксического механизма согласования не волен выбирать форму слова, выступающего в качестве определительной или сказуемой синтаксемы, отсутствует и «право выбора» варианта аффикса, подчиняющегося закономерностям сингармонизма - морфофонологического согласования.

Мы отвлекаемся от обсуждения разного рода особых и сложных случаев, когда действие сингармонизма обнаруживает те или иные отступления, когда выделяются гласные или согласные, блокирующие распространение морфофонологического согласования, и т.п. (Касевич, 1986). В плане речепорождения важнейшее следствие наличия в языке правил сингармонизма заключается, повторим, в том, что говорящий должен «уложиться» в одну из имеющихся в системе морфофонологических моделей. Выход за пределы данного множества моделей есть одновременно выход за пределы норм соответствующего языка; он может и маркировать слово в качестве неассимилированного заимствования (разные сингармонические языки отличаются степенью толерантности по отношению к словам-заимствованиям, нарушающим правила сингармонизма).

Слова в словаре сингармонического языка сгруппированы в классы, отвечающие действительным для этого языка морфофонологическим моделям. Как известно, сингармонические языки с грамматической точки зрения обычно принадлежат к агглютинативному типу, который предпочитает аффиксам квазиаффиксы - незначительные морфемы, обладающие смешанными признаками служебных слов и аффиксов. Квазиаффиксы равным образом входят в словарь и также распределены по морфофонологическим классам: как неварьирующие, но требующие согласования от последующего (предыду-

щего) квазиаффикса, варьирующие и одновременно задающие варьирование, варьирующие, но не требующие уподобления от «соседей», и т.д. и т.п. Каждая конкретная модель словоформы («квазисловоформы») тем самым выводима как результат взаимодействия признаков класса основы и классов (квази)аффиксов.

Последние, помимо морфонологического класса, обладают, естественно, и признаками класса морфосинтаксического, как правило, соотнесенного с позицией в составе словоформы. Поскольку позиция в типичном случае связана с активностью/пассивностью морфемы в процессах сингармонизма - диктует эта морфема морфонологическое согласование или же сама должна «подстраиваться» под признаки «соседа», - морфосинтаксическая и морфонологическая классная принадлежность обнаруживают непосредственную связь. В итоге порождение агглютинативной сингармонической словоформы оказывается результатом взаимодействия закономерностей, следующих из вхождения соответствующих морфем в определенным образом связанные морфонологические и морфосинтаксические классы. Особенности как знаменательных, так и служебных морфем, взаимодействующих в рамках указанного механизма, с формальной точки зрения исчерпываются классной принадлежностью и отсюда чувствительностью к правилам линейного расположения и морфонологического согласования.

Со времен Бодуэна де Куртенэ принято считать, что сингармонизм обеспечивает связанность морфем в составе словоформы (правда, Бодуэн говорил о слогах в слове). Как можно видеть из изложенного выше, этот тезис нуждается в уточнении и развитии: речь должна идти именно о моделях, охарактеризованных одновременно грамматически и фонологически, которые регулируют заполнение синтаксических позиций в составе высказываний.

Такая постановка вопроса помогает легко перейти от рассмотрения сингармонизма в плане речепроизводства к анализу роли этого явления в восприятии речи. Восприятие речи предполагает, среди прочего, идентифицирование определенных сегментов текста со словарными единицами. Распределение последних по формальным морфонологическим классам сужает класс поиска, т.е. облегчает поиск, делает его оперативнее. Определив модель слова, слушающий тем самым одновременно отсекает все варианты, не согласующиеся с данной моделью. Поскольку в составе модели фонологическое оформ-

ление аффиксов частично предсказуемо, знание модели позволяет получить грамматическую информацию о словоформе наиболее экономным образом.

Иначе говоря, ассоциированность любой словоформы сингармонического языка с одной из существующих в системе морфонологических моделей позволяет производить грубую категоризацию словоформ путем упрощенного доступа к источникам грамматической информации и отнесения всей словоформы к классу, внутри которого осуществляется дальнейший поиск.

Однако в восприятии речи процедуры идентификации неразрывно связаны с процедурами сегментации, или делимитации. Выше говорилось о том, что со словарными единицами отождествляются те или иные сегменты текста. Чтобы операция отождествления стала реальной, сегменты должны быть выделены из текста. И к этой задаче сингармонистические модели также имеют самое непосредственное отношение. Само существование в системе разных моделей имеет своим следствием вероятность соположения в тексте словоформ, реализующих эти разные модели. Иначе говоря, в тексте следует ожидать наличия «точек», соответствующих линейному переходу от модели одного типа к модели другого, и эти «точки», естественно, отвечают границам между словоформами в тексте, т.е. обеспечивают сегментацию.

Вполне понятно, что нет и не может быть общего правила, согласно которому в тексте регулярно чередовались бы словоформы, принадлежащие разным сингармонистическим моделям. Употребление той или иной словоформы в каждой данной позиции диктуется семантикой, синтаксисом, стилистическими требованиями, частотными характеристиками и т.п., так что априорная оценка вероятности соположения в тексте словоформ, реализующих одинаковые/разные морфонологические модели, едва ли возможна. Одной лишь информации о числе таких моделей в системе явно недостаточно.

Из сказанного следует, что вопрос о том, каков реальный вклад сингармонистических моделей, их текстового распределения в задачу сегментации текста, носит эмпирический характер и должен, соответственно, исследоваться экспериментально.

Именно с этой целью был проведен эксперимент на материале киргизского языка, в котором испытуемым предлагалось фиксировать на письме синтагмы разного объема, запи-

санные на магнитную ленту в условиях частотной фильтрации. Воспроизводились условия серии экспериментов, где на материале разных языков (русского, японского, таджикского, французского) была выявлена зависимость разборчивости языковых единиц от доступности делимитативных сигналов, обеспечиваемых, главным образом, просодическими маркерами соответствующих языков (Касевич и др., 1990б).

Экспериментальный материал представлял собой 180 словосочетаний - именных фраз с ядром-существительным и различными определениями к нему, например *Башталгыч мектептердин мугалимдери* 'Учителя начальной школы'. Число слогов - от 5 до 12. Представленность словосочетаний с разным числом слогов и слов была примерно одинаковой, записывались они в случайном порядке в исполнении диктора, владеющего нормативным произношением киргизского литературного языка. Фильтрация осуществлялась с помощью спектрометра звуковых частот с полосой пропускания 906-1141 Гц.<sup>13</sup> Отфильтрованная запись была предъявлена бригаде auditors, состоявшей из 10 студентов Киргизского университета, чьим родным языком был киргизский. Прослушивание осуществлялось через громкоговорители. Инструкция предлагала испытуемым записать услышанное максимально полно, без пропусков, подставляя вместо плохо слышимых слов, если таковые будут, их наиболее близкие по звучанию аналоги. Отказы не допускались.

В протоколах испытуемых анализировались словесная и слоговая разборчивость, разборчивость гласных и согласных, а также сохранность позиции межсловных границ вне зависимости от адекватности/неадекватности распознавания слов, разделяемых этими границами.

Следует признать, что сформулировать «нулевую гипотезу», которая предсказывала бы результаты эксперимента в интересующих нас аспектах, достаточно непросто. Выше отмечалось, что в проведенных ранее аналогичных экспериментах аудиторы, по всей вероятности, ориентировались на просодические - акцентуационные - признаки, именно за счет информации об ударении выделяя слова в условиях труднодоступности информации о сегментных единицах, гласных и согласных. Для

---

<sup>13</sup> По техническим причинам установленный режим фильтрации оказался нарушенным, однако полоса пропускания не изменялась в течение всего процесса фильтрации.

киргизского языка наличие словесного ударения как самостоятельной фонологической категории не может считаться самоочевидным. Представляется вполне вероятным, что киргизский, подобно некоторым (многим) сингармонистическим языкам, относится к классу анацентных языков. Часть функций, которые в акцентных языках свойственны ударению, берет на себя сингармонизм. Эта одна из причин распространенности трактовки, согласно которой сингармонизм сближается со сферой просодики или даже прямо включается в последнюю.

Не обсуждая соответствующих концепций, отметим, что с точки зрения реализационных, «субстратных» средств сингармонизм далек от просодических явлений. Ударение, в частности, реализуется прежде всего за счет мелодики и/или длительности – именно эти фонетические характеристики обладают повышенной помехоустойчивостью в сравнении с признаками гласных и согласных, что и объясняет возможность их использования в затрудненных условиях восприятия. Что же касается сингармонизма, то, какова бы ни была функционально-системная природа этого явления, фонетически он реализуется именно как тип распределения гласных, иногда (в частности, и в киргизском языке) также согласных, – т.е. именно тех единиц, признаки которых в типичном случае наиболее уязвимы при наличии помех, обладая пониженной устойчивостью.

Соответственно в нашем эксперименте на материале киргизского языка информация о типе сингармонистического оформления слова и о консонантно-вокальном наполнении последнего как будто бы не может быть разграничена, реализуясь одними и теми же средствами.

Обратимся, однако, к экспериментальным данным. Средняя словесная разборчивость оказалась весьма низкой – 16,7%. Это очень близко к данным по словесной разборчивости, полученным в тех же условиях на материале русского языка (17,3%), и значительно уступает разборчивости японских, таджикских и французских слов (42, 65,2 и 52,6% соответственно). Низкая словесная разборчивость на материале русского языка, если следовать гипотезе об обращении к просодической информации в условиях дефицита информации о гласных и согласных, объясняется типом русского ударения: нефиксированность его позиции позволяет определить число слов, но не границы между ними, отсюда и ошибки в распознавании слов. О чем говорит низкая разборчивость киргизских слов?

Кажется естественным предположить, что при использовании киргизского материала, где пограничными сигналами могут служить переходы от одной морфонологической модели к другой, именно эти сигналы становятся труднодоступными, поскольку, как сказано выше, вводимая в сигнал помеха - в отличие от ситуации с русским материалом - не действует избирательно. В этом случае экспериментальный параметр, отражающий сохранность в ответах аудиторов позиции словесных границ, должен быть достаточно низким. В свою очередь, трудности в обращении к информации о межсловных границах, т.е. к информации о сегментации словосочетаний, как раз и должны вызывать снижение словесной разборчивости.

Однако это предположение подтверждается не полностью. Указанный параметр действительно оказался невысоким - 61,9% в среднем,<sup>14</sup> но он заметно превышает значение параметра, полученное в тех же условиях на русском материале, - 40,4%. Если для русского, японского и таджикского языков удалось эмпирически выявить строгую количественную зависимость между словесной разборчивостью и «разборчивостью» словесных границ (см. с. 108) - что и доказывает гипотезу об обусловленности словесной разборчивости адекватностью сегментации, то киргизские данные в выведенное соотношение определенно не укладываются.

Для объяснения экспериментально установленного соотношения, которое на киргизском материале характеризуется низкой словесной разборчивостью при «недостаточно низком» показателе сохранности словесных границ, нужно снова обратиться к анализу природы сингармонизма и возможных механизмов его функционирования в речевой деятельности, в частности в процессах восприятия речи.

Наличие в языке сингармонизма ведет к тому, что в типичном случае словоформа в целом характеризуется сквозным дифференциальным признаком, реже двумя, который (которые) можно обнаружить у всех гласных словоформы. Создается эффект своего рода «фонологического плеоназма».

---

<sup>14</sup> Заметим, что если бы в киргизском языке существовало фиксированное словесное ударение, падающее на конечный слог слова (как чаще всего считается), то сохранность словесных границ, однозначно определяемых по месту ударения, была бы гораздо выше. Именно такова картина для таджикского языка с его ударением, закрепленным за последним слогом слова, где в тех же экспериментальных условиях словесные границы были адекватно определены испытуемыми в 83,7% случаев.

Плеоназм принадлежит к тем средствам, которые повышают избыточность текста. Этот эффект мы по существу и наблюдаем на материале сингармонистических языков: воспроизводимость, повторяемость дифференциального признака на участке текста, соответствующего словоформе, облегчает распознавание гласных, а отсюда и самой словоформы.

Практически это означает, в частности, что распознавание хотя бы одной гласной из экспонента словоформы (или даже одного из вокальных признаков, если это признак, участвующий в создании сингармонистической модели) уже в сильнейшей степени ограничивает возможности «продолжения» этой словоформы от данной гласной как «вправо», так и «влево». В результате вполне реальна ситуация, когда в затрудненных условиях, в условиях неполной информации о гласных и согласных происходит замена словоформы с сохранением ее сингармонистической модели. Легко видеть, что процесс указанного типа должен приводить к достаточно низкой словесной разборчивости при сравнительно высоком параметре сохранения межсловных границ, если последние поддерживаются сингармонизмом как делимитативным средством.

Именно таковы наши данные. Сформулированные выше предположения подтверждаются также и типом замен словоформы в протоколах аудиторов. Всего, как явствует из изложенного выше, замены слов отмечались в 83,3% случаев. Из них 61,7% приходится на замены словоформ, в которых сохранялся ряд гласных по отношению к исходному слову, например *Эми кайталан айтпаймын* 'Теперь я уже не повторяю' вместо *Элдин талкаланыш каманы* 'Разгромленный народ'; 54,6% - на замены с сохранением признака огубленности, ср.: *Кергезмеге коюлган сут* 'Молоко, выставленное на витрину' вместо *Бетелкеге коюлган сут* 'Молоко, налитое в бутылку'; и 58,2% - с сохранением как ряда, так и огубленности (неогубленности).<sup>15</sup>

Знаменательно, что как раз на те случаи, когда замена словоформы не сопровождается полной заменой сингармонистической модели, удерживая один или оба дифференциальных признака, приходится и основное число случаев сохранения позиции словесных границ (89,2%).

---

<sup>15</sup> В киргизском языке, как известно, согласование гласных в пределах словоформы осуществляется как по ряду, так и по огубленности. Есть и более сложные зависимости, в том числе и нарушающие основное правило, на чем мы не можем останавливаться (см., например: Johnson, 1980).

Мы уже отмечали, что делимитативную роль сингармонизма естественнее всего представлять себе как следствие соположения в речевой цепи словоформ, принадлежащих к разным морфонологическим моделям. Действительно, смена моделей как будто бы должна оптимальным образом содействовать сегментации текста на словоформы. Выше мы, однако, выдвинули по существу предположение, согласно которому сам факт сингармонистической оформленности слова способствует его вычленению из текста вне зависимости от контрастирования/неконтрастирования по отношению к морфонологической модели ближайших «соседей». Насколько правомерно выдвинутое предположение?

Тенденцию, которая заключалась бы в естественном квантовании текста на словоформы уже в силу соответствия каждой из них той или иной морфонологической модели, конечно, нельзя абсолютизировать. Однако есть основания считать, что действие сингармонизма, даже не создавая пограничных сигналов как таковых, способствует использованию в качестве пограничных сигналов других языковых средств. Мы имеем в виду аффиксы (квазиаффиксы), которые в языке типа киргизского, во-первых, занимают фиксированную позицию в словоформе, а во-вторых, обладают в значительной степени предсказуемым – именно благодаря закономерностям сингармонизма – фонологическим составом. Как мы уже видели, избыточность, предсказуемость естественным образом улучшает восприятие. В данном случае, если наше предположение верно, правила сингармонизма действуют не непосредственно, а опосредованно, облегчая распознавание аффиксов (квазиаффиксов), фиксированность позиции которых позволяет им функционировать в качестве пограничных сигналов.

Неоднозначность ситуации с использованием делимитативных потенций сингармонистических моделей обнаруживается и при рассмотрении наших экспериментальных данных (см. рис. 3.10). Лишь для пяти-, шестисложных словосочетаний, вне зависимости от числа слов, отмечается гораздо лучшее сохранение в ответах аудиторов тех словесных стыков, которые характеризуются переходом от одной сингармонистической модели к другой. Из более длинных последовательностей только трехсловные сочетания, состоящие из 9–10 слогов, обнаружили ярко выраженную закономерность – улучшение идентификации сингармонически контрастных словесных границ по сравнению с неконтрастными (рис. 3.10,Б). В остальных случаях контраст-

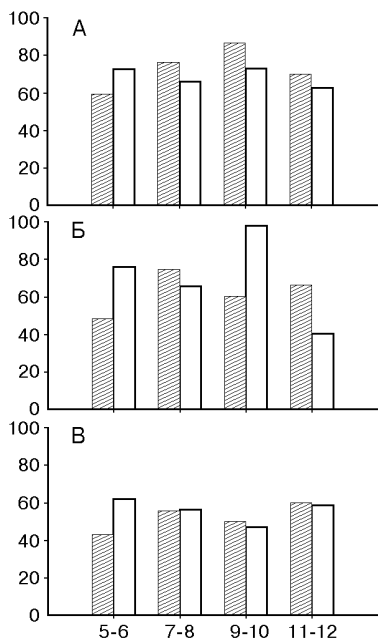


Рис. 3.10. Разборчивость словесных границ, контрастных и неконтрастных по сингармонизму.

А: 2-словные синтагмы; Б: 3-словные; В: 4-словные.

Заштрихованные столбики – контрастные, светлые – неконтрастные границы.

По оси абсцисс – число слогов в синтагме; по оси ординат – разборчивость в процентах.

ность / неконтрастность стыка либо вообще не влияет на его перцептивную сохранность, либо неконтрастность обнаруживает даже более высокие показатели по сравнению с контрастностью.

Добавим к этому, что 68,5% из всех адекватно опознанных границ характеризовались наличием аффиксов.

Мы не будем анализировать все возможные комбинации факторов, которые, предположительно, могли бы объяснить разные перцептивные механизмы сегментации текста (в нашем случае – словосочетаний). Упомянем, что среди всех межсловных стыков экспериментальной программы 48,4% приходилось на сингармонистически контрастный тип, и 51,6% – на неконтрастный.<sup>16</sup>

В целом же вырисовывается картина, которую можно охарактеризовать как комплексное использование ряда факторов, которые, комбинируясь различным образом, обеспечивают перцептивную сегментацию текста. Сингармонизм, как можно предпо-

ложить, является одним из таких факторов. Сингармонистическая

<sup>16</sup> Чтобы оценить, хотя бы приблизительно, типичность такого соотношения, мы произвели аналогичные подсчеты на материале текста объемом в 10 страниц, или более 2,5 тыс. словоформ (Айтматов Ч.Т. Гульсарат. Фрунзе, 1978). В указанном отрезке текста обнаружилось 50,5% контрастных стыков и 49,5% – неконтрастных.

модель выполняет в процессах речевосприятия двоякую функцию.

С одной стороны, как уже говорилось выше, фиксирование такой модели относит слово к некоему классу, что облегчает идентификацию слова (в информационно насыщенной среде идентификация слова с точностью до класса может оказаться достаточной для реализации доступа к этому слову в словаре). Об указанной роли сингармонистической модели говорят данные по предпочтительности замены слов с сохранением модели.

С другой стороны, оформленность слова сингармонистической моделью способствует его вычленению из текста. Контрастность моделей соседних слов обеспечивает сегментацию лишь с выполнением некоторых дополнительных условий, на нашем материале это ограничения со стороны длины сегментируемой цепочки в терминах слогов. В остальных случаях закономерности сингармонизма участвуют в выполнении делимитативной функции, скорее всего, не непосредственно, а облегчая функционирование других пограничных сигналов, среди которых важнейшими, можно думать, являются позиционно закрепленные аффиксы (квазиаффиксы).

Мы заметили лишь некоторые аспекты экспериментально-го исследования вклада, который вносит сингармонизм в стратегии речевосприятия. Важный, с нашей точки зрения, результат состоит в том, что можно говорить о типологии стратегий, скоррелированной с типологией языка, и о комплексном характере факторов, к которым относится и сингармонизм, взаимодействующих - с переменными весами - в процессах речевой деятельности.

## О ПРЕДЕЛАХ АВТОНОМНОСТИ ПРОСОДИЧЕСКИХ ЯВЛЕНИЙ В ПЕРЦЕПТИВНЫХ ПРОЦЕССАХ

В многочисленных экспериментах показано, что просодическая информация может обрабатываться воспринимающим речь человеком независимо от информации о сегментных единицах. В частности, не раз демонстрировалось то обстоятельство, что при восприятии речи на фоне тех или иных помех аудиторы в типичном случае правильно воспроизводят ритмическую структуру слова, заменяя, иногда полностью, консонантно-вокальный состав последнего, см., например, (Чистович и др., 1965). Указанный феномен объясняют большей помехоустойчи-

востью просодических характеристик речи в сравнении с устойчивостью к помехам гласных и особенно согласных, что, в свою очередь, связывается как с различиями в соответствующих акустических явлениях, так и со свойствами слухового анализатора человека.

Одновременно – и независимо от экспериментальных штудий – в ряде направлений современной теоретической фонологии проводится мысль об автономности просодических характеристик языковых единиц по отношению к сегментным составляющим их экспонентов. В особенности на этом настаивает теория аутосегментной фонологии, согласно которой структурно-фонологическая характеристика высказывания заключается в порождении независимых цепочек, принадлежащих автономным ярусам (*tiers*); одна из цепочек отражает последовательность сегментных единиц, другая же, например, слоговых тонов или индексов ударенности. Полная фонологическая характеристика слова (или другой языковой единицы) есть результат установления соответствия между разными ярусами, интегрирования принадлежащей им информации (Goldsmith, 1979; Ewen, 1986 и др.).

Иначе говоря, и экспериментальные результаты, и фонологические формализмы указывают на высокую степень взаимной автономности просодической и сегментной информации. В то же время, если заниматься изучением реальных механизмов речевой деятельности, желательно представлять себе более точно именно степень и характер связанности – либо несвязанности – указанных явлений. Можно заранее предположить, что для разных языков и даже для разных фрагментов одной и той же фонологической системы решение этого вопроса будет разным; например, если ударение падает только на закрытые слоги или слоги с долгим гласным, то взаимозависимость акцентных и сегментных характеристик будет больше по сравнению с ситуацией, где такого ограничения нет. Но существуют и аспекты более общего порядка, где различия между фонологическими системами и подсистемами, обусловленные типологически, едва ли могут быть значительными. Один из таких аспектов исследовался авторами на материале вьетнамского языка, где предмет изучения формулировался так: возможно ли при восприятии речи полное отвлечение фонологически релевантного тона от его сегментного базиса – слога?

Как известно, вьетнамский – тональный язык. Каждый слог вьетнамского языка получает фонологически релевантное про-

содическое оформление – произносится в одном из шести тонов. Слогов вне тонального оформления не существует, равно как не существует тональных характеристик, не ассоциированных с каким-либо слогом. Имеются ограничения на сочетаемость тонов и типов слогов (см., например, Гордина, Быстров, 1984), что составляет отдельную проблему (см. ниже, с. 132).

Исходный экспериментальный материал<sup>17</sup> представлял собой 12 предложений, записанных в исполнении диктора – носителя стандарт-

ного варианта вьетнамского языка, уроженца г. Ханоя. Каждое предложение состояло из шести односложных слов.

Распределение тонов в корпусе экспериментального материала не было сбалансированным ни в целом, ни применительно к позиции в составе предложения. Количественные характеристики представленности тонов показаны в табл. 3.9.

В экспериментальной программе на каждый коммуникативный тип – повествовательный, вопросительный, побудительный – приходилось по 4 предложения. Вся программа была записана дважды.

Экспериментальный материал вводился в ЭВМ через 10–разрядный АЦП с частотой дискретизации 20 кГц. Программными средствами в экспериментальных предложениях выделялись все периоды основного тона, после чего оригинальный сигнал на каждом периоде заменялся двумя полными периодами синусоиды, а участки глухих согласных – паузами. Полученные в результате сигналы сохраняли только тонально-мелодические характеристики экспериментальных предложений. Сигналы выводились из ЭВМ через 12–разрядный ЦАП и записывались на магнитную ленту. Каждое предложение записы-

Таблица 3.9. Распределение тонов по позициям (в абсолютных числах)

Тон	Номер позиции					
	1	2	3	4	5	6
1	8	7	4	5	6	10
2	–	3	–	3	4	1
3	1	–	–	–	–	–
4	–	–	3	1	1	1
5	2	1	5	3	1	–
6	1	1	–	–	–	–

<sup>17</sup> В подборе материала, проведении экспериментов и обработке данных участвовали аспирантка Т.И.Тагиева и студент А.С.Шиканов (С.-Петербургский университет).

валось двукратно, все 24 предложения, учитывая повторения, были записаны в случайном порядке.

Серии экспериментальных сигналов, представляющих собой некоторое модулированное «мычание», предъявлялись 8 испытуемым, уроженцам северных провинций Вьетнама. В инструкции предлагалось прослушать каждый из разделенных паузами сигналов дважды, затем остановить магнитофон и сочинить любое предложение, мелодика которого «ложилась» бы на прослушанное звучание. Время для составления предложения не оговаривалось.

Иначе говоря, нашей задачей была проверка возможности «буквального прочтения» гипотезы о максимальной автономности сегментных и супraseгментных средств языка. По некоторым данным, существуют реальные ситуации, когда носитель языка ставится в условия, тождественные тем, в которых работали наши испытуемые. Так, в Японии известна игра, в ходе которой на флейте «высвистывается» мелодия, имитирующая акцентный контур слова, и играющим предлагается отгадать слово, отвечающее данному акцентному контуру (Рыбин, личное сообщение).

Задача, поставленная перед аудиторами, воспринималась ими как весьма трудная, многочисленными были попытки отказов. Основная трудность, с которой пришлось столкнуться уже на стадии обработки результатов эксперимента, заключалась в том, что лишь в 17% случаев предложения, составленные испытуемыми, совпадали с исходными по числу слогов. Хотя традиционно считается, что из всех тональных языков вьетнамский едва ли не минимально обнаруживает количественную и качественную редукцию слогов (Sprigg, 1976), в потоке речи слоги, конечно, не следуют друг за другом как блоки равной протяженности, с одинаковой проявленностью тональных характеристик. Соответственно в условиях эксперимента, когда у испытуемых отсутствовала возможность извлечения информации о слоговых границах, два и более слогов вполне могли восприниматься в качестве одного и, наоборот, один – в качестве двух; типичнее оказалось сокращение общего числа слогов в предложении, т.е. перцептивное укрупнение составляющих слоговой структуры экспериментальных предложений.

В результате определение адекватности воспроизведения тона для всех слогов в предложении оказалось возможным на материале лишь 17% от общего количества слогов программы; для остальных 83% слогов соответствующие заключения дела-

лись лишь применительно к тем случаям, когда эти слоги выступали в качестве начальных или конечных в предложении и тем самым задавалась однозначно по крайней мере одна из слоговых границ.

В табл. 3.10 приведены результаты опознания слоговых тонов (прочерки указывают на отсутствие слога в данном тоне в соответствующей позиции, ср. табл. 3.9).

Как можно видеть, в среднем наиболее устойчивыми оказываются 1-й и 2-й тоны, а среди позиций – конечная. Когда совпадают две характеристики – 2-й тон и конечная позиция – опознание тонов оказывается, учитывая условия эксперимента, поразительно успешным, достигая 94%. В целом же на материале вьетнамского языка мы не получаем картины, когда одних чисто просодических конфигураций оказывалось бы достаточно для опознания отвечающих им функциональных единиц – тонов, если эти последние даны в полном отвлечении от своего сегментного субстрата.

Здесь уместно вспомнить, что с функциональной точки зрения тоны суть характеристики слогов (точнее, слогоморфем, но на этом вопросе мы здесь не будем останавливаться, см. Касевич, 1983). Иначе говоря, тоны в тональном языке существуют именно и только для того, чтобы обеспечивать просодическое оформление слогов, за счет чего умножаются диакритические, дистинктивные потенции силлабария, ограниченного по своему объему в силу допустимости лишь простых слоговых структур относительно бедного состава. Уже из этого следует, по-видимому, что автономность тона по отношению к слогу должна иметь свои пределы: тон может быть независим по отношению к данному слогу (хотя бы в том смысле, что один и тот же тон сочетается со слогами разных типов, а при восприятии речи реально опознание тона без идентификации слога), но он не может быть независимым по отношению к слогу как таковому.

Таблица 3.10. Опознание тонов слогов по позициям (эксперимент 1), проценты

Тон	Номер позиции						Средн.
	1	2	3	4	5	6	
1	83	5	9	16	17	63	32
2	—	15	—	0	16	94	31
3	0	—	—	—	—	—	0
4	—	—	10	0	0	38	12
5	13	25	6	3	0	—	9,4
6	6	0	—	—	—	—	3
Средн.	25,5	11,3	8,3	4,8	8,3	65	

Исходя из сказанного, а также учитывая результаты описанного выше эксперимента, мы провели еще один эксперимент, являвшийся модификацией предыдущего. В сигналы, использовавшиеся для аудирования, были введены паузы, позиции и длительности которых соответствовали расположению и продолжительности согласных (теперь уже всех, включая звонкие) в исходных предложениях. Тем самым формировалась квазислоговая структура экспериментальных стимулов: тоны были по-прежнему лишены своего естественного сегментного субстрата, но они получали сферу реализации, ограниченную слогоподобными границами. Возникла возможность фиксировать начальную и конечную точки тональной кривой, нормально ассоциированные с границами слога; при

Таблица 3.11. Оpozнание тонов слогов по позициям (эксперимент 2), проценты

Тон	Номер позиции						Средн.
	1	2	3	4	5	6	
1	94	52	66	77	59	61	68,3
2	—	65	—	33	44	94	59
3	19	—	—	—	—	—	19
4	—	—	27	6	12	44	22,5
5	44	63	14	34	63	—	43,6
6	0	6	—	—	—	—	3
Средн.	39,3	46,5	36	37,5	44,5	66,3	

этом надо учесть, что именно «приграничные» значения акустических параметров — коррелятов признаков тона — особенно важны для перцептивной идентификации последнего (Касевич и др., 1990б).

Подготовленные указанным способом стимулы были также предъявлены бригаде auditors из 8 человек (состав бригады, к сожалению, практически полностью отличался от того, что использовался в первом из экспериментов); инструкция испытуемым оставалась прежней.

Результаты опознания тонов в ходе второго эксперимента представлены в табл. 3.11.

Сопоставление данных таблиц 3.10 и 3.11 показывает, что введение в экспериментальные стимулы дополнительных слогоподобных границ резко повышает все показатели. При этом общие тенденции сохраняются: по-прежнему в восприятии лидируют 1-й и 2-й тоны, а также конечная позиция. Так же, как в первом эксперименте, хуже всего воспринимаются 3-й и 6-й тоны, которые в системе выступают как максимально маркированные (обладающие сложными признаковыми характери-

ками) и отличаются самой низкой частотностью и в словаре, и в тексте (Гордина, Быстров, 1984). Нужно также отметить, что в обоих экспериментах ошибки в восприятии тонов, как правило, не затрагивали регистр: замены осуществлялись внутри подкласса тонов верхнего (1-й, 3-й и 5-й) или нижнего (2-й, 4-й и 6-й) регистров.

Разумеется, полученных данных недостаточно для обоснованного сопоставления перцептивных особенностей разных тонов вьетнамского языка в зависимости от позиции и других факторов. Для этого нужно было бы иметь специальную программу, сбалансированную по всем параметрам, с учетом системной маркированности/немаркированности тонов, их словарной и текстовой частотности и т.д. Но это и не было задачей проделанных экспериментов. Эксперименты должны были показать – и показали – пределы автономности тона с перцептивной точки зрения. При всей бесспорной независимости тональных характеристик от соответствующего сегментного субстрата, полностью лишенная дискретизации последовательность тональных посылок обнаруживает достаточно низкую воспринимаемость. В то же время само по себе введение дискретизации, имитирующей слоговое членение экспериментальных квазипредложений, приводит к резкому повышению разборчивости тонов. Иными словами, минимальная опора на сегментный – или даже квазисегментный, как в нашем случае – базис слогового формата есть необходимое условие для перцептивного использования тона.

Тем самым, как можно думать, распространенный в современной литературе тезис о полной взаимной автономности сегментных и супraseгментных явлений оказывается серьезно поколебленным.

К сказанному надо добавить, что слогоподобная дискретизация экспериментальных стимулов привела также к улучшению воспринимаемости коммуникативного типа. Если в первом эксперименте средняя разборчивость интонационного типа составила 40%, то во втором этот показатель был равен 60%.

Приведенные данные также есть основания считать достаточно демонстративными. Интонационные характеристики, будучи связаны с высказыванием и его подструктурами (если последние имеются) в целом, дальше «отстоят» от слогов, нежели тоны. Тем не менее, как видим, слоговая дискретизация оказывается небезразличной и для воспринимаемости интонации.

Выше исследовался вопрос о том, каким должен быть минимальный объем информации о сегментной структуре высказывания, чтобы доступной для обработки оказалась информация просодическая. Но, как уже упоминалось, есть и более частный аспект проблемы, который заключается в том, что в тональных языках обычны ограничения на сочетаемость определенных тонов с тем или иным типом слога, т.е. ограничения на сочетаемость просодических и сегментных единиц. Вполне понятно, что наличие дистрибутивных закономерностей указанного типа не может быть безразличным для перцептивных процессов. В ситуации спонтанной речи, да и просто в интонационно оформленных и ритмически организованных высказываниях, когда тоны могут подвергаться значительной модификации, вполне возможны и конфликтные ситуации, в условиях которых, например, мелодические характеристики данного конкретного слога указывают на один тон, а сегментный состав слога – на другой.<sup>18</sup> Ниже излагаются предварительные результаты эксперимента, связанного с изучением названной проблемы<sup>10</sup> и выполненного на том же речевом материале, что уже описывался выше.

Были использованы те же 12 шестисложных фраз вьетнамского языка в исполнении того же диктора.

С помощью аппаратно-программных средств ЭВМ каждая фраза членилась на слова, которые затем выводились из ЭВМ через 12-разрядный цифро-аналоговый преобразователь. Из записанных на магнитную ленту слов образовывался экспериментальный набор сигналов. Опыты проводились методом диктанта, при котором испытуемый записывал каждый предъявленный сигнал. Каждое слово в наборе «диктовалось»

---

<sup>18</sup> Поскольку слог в языках Китая и Юго-Восточной Азии обычно представляет собой морфему и возможны ограничения на сочетаемость тонов с морфемами некоторых классов, то возникает дополнительная проблема :интерференции: фонологических и грамматико-семантических признаков, ср., например, (Fox, Unkefer, 1985).

<sup>10</sup> К настоящему времени сложилась уже достаточно развитая традиция изучения восприятия тонов, главным образом на материале китайского языка с его диалектами и тайского языка. Мы не можем здесь, к сожалению, анализировать соответствующую литературу, см. (Румянцев, 1991; Gandour, 1979; Gandour, 1983; Chuang e.a., 1972) и другие работы.

только один раз, но при этом повторялось трижды. В опытах принимали участие 26 испытуемых – носителей языка.

Для анализа полученных результатов все использованные в опытах сигналы были разделены на два класса: с неизменным на отрезке финали периодом основного тона (О.Т.) и с изменяющимся на этом отрезке периодом. Схематические периодограммы сигналов из этих классов (зависимость длительности периода О.Т. от номера периода) приведены на рис. 3.11, А, Б и В соответственно. Судя по результатам экспериментов, резкое изменение длительности периодов в конце финали, связанное с образованием смычки последующего глухого согласного (рис. 3.11, А), не воспринимается испытуемыми как изменение высоты, и потому подобное изменение мелодического контура не приводит к нарушению его субъективной стационарности.

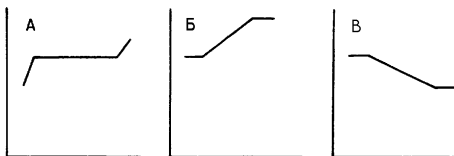


Рис. 3.11. Схематическое представление мелодических контуров, характеризующих тоны в слогах вьетнамского языка: зависимость длительности периода основного тона от его положения в контуре.

А – 1-й тон; Б – 2-й или 4-й тон; В – 5-й тон.

Для сигналов первого класса – стационарных – следовало ожидать преобладания ответов «1-й тон», однако для части сигналов преобладающими оказались ответы «5-й тон» или «6-й тон». Чтобы оценить характер перехода от одного типа ответов к другому, все сигналы этого класса были ранжированы в порядке убывания частоты ответов «1-й тон». Как следует из рис. 3.12 (внизу), убывание частоты ответов «1-й тон» сопровождается возрастанием суммарной частоты ответов «5-й тон» и «6-й тон» при мало изменяющейся частоте появления ответов с остальными тонами. Оказывается, что только неизменности периодов О.Т. недостаточно для восприятия данного сигнала как произнесенного в 1-м тоне. Анализ ответов испытуемых показывает, что 88% ответов «5-й тон»: или «6-й тон» приходится на случаи, когда в конце слова воспринимается глухой смычный согласный. Известно, что такие согласные могут реализоваться только в 5-м или 6-м тонах (Гордина, Быстров, 1984).

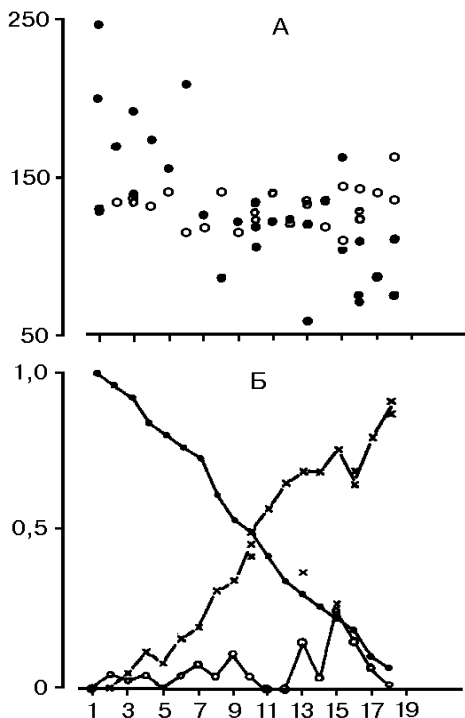


Рис. 3.12. Связь временных и мелодических характеристик слога с восприятием тона.

А – частота О.Т. (кружки) и длительность финали (точки); Б – относительная частота ответов «1-й тон» (точки), «5-й» и «6-й» (крестики), «остальные тоны» – кружки.

По оси абсцисс – «ранг» сигналов в порядке убывания частоты ответов «1-й тон»; по оси ординат: А – Гц и мс; Б – относительные единицы.

Естественно поэтому предположить, что сигналы со стационарным О.Т. относятся слушателями к 5-му или 6-му тону потому, что этого требуют результаты фонетической интерпретации сегментного состава этих сигналов (с глухим смычным согласным в конце) и известные испытуемым закономерности фонологической системы. Отсюда вытекает необходимость поиска тех признаков в сигнале, которые приводят к восприятию или собственно 5-го или 6-го тонов, или глухих смычных согласных в абсолютном завершении слова. На рис. 3.12 (вверху) для каждого ранга приведена длительность финали, т.е. части слога за вычетом начального согласного, и средняя частота О.Т. Нетрудно убедиться, что регистровые изменения вряд ли являются одним из искоемых признаков,

поскольку частота О.Т. мало изменяется от ранга к рангу. Что касается длительности, то возрастающей частоте ответов «5-й –

6-й тон» соответствует заметное уменьшение длительности финали слов.

Сигналы с изменяющимся во времени О.Т. могут быть описаны мелодическими контурами, схематически представленными на рис. 3.11,Б и В. Особенностью этих контуров является то, что в них обязательно присутствует участок с положительным или отрицательным наклоном аппроксимирующей прямой. Для

количественного описания таких контуров можно использовать либо величину наклона прямой на участке изменения О.Т., либо перепад длительности периода О.Т. на этом участке. Применительно к использо-

ванным нами сигналам лучшие результаты дает применение именно перепада длительности периода на участке изменения. Допустимость использования такого параметра подтверждают данные о восприятии изменений высоты (Люблинская, 1971).

Очевидно, что сигналы с положительным перепадом длительности периода (с увеличивающейся длительностью периода, т.е. с понижающимся тоном) должны чаще относиться слушателями к 4-му или 2-му тонам, а с отрицательным перепадом — к 5-му тону. Из рис. 3.13 следует, что мелодические контуры с большим положительным перепадом длительности периода О.Т. чаще интерпретируются слушателями как относящиеся к 4-му тону, при уменьшении величины перепада возрастает частота ответов «2-й тон», а при дальнейшем его уменьшении преобладающими становятся ответы «любой тон, кроме 2-го и 4-го». По-видимому, при некоторой величине изменения

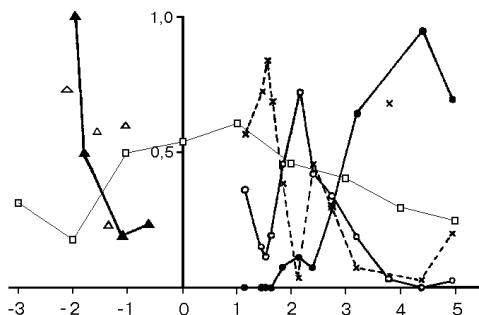


Рис. 3.13. Изменение оценок слогового тона в зависимости от величины перепада длительности периода О.Т. на участке слога.

Точки — «4-й тон», кружки — «2-й тон», крестики — «любой, кроме 2-го и 4-го», зачерненные треугольники — «5-й тон», квадраты — «1-й тон», светлые треугольники — ответы «5-й тон» для слогов с конечным глухим смычным согласным.

По оси абсцисс — перепад длительности периода О.Т. в мс (положительный для понижающегося тона); по оси ординат — относительная частота ответов.

периода О.Т. в сигнале мелодический контур воспринимается уже как стационарный, а отнесение данного сигнала к 1-му, 5-му или 6-му тонам производится на основании других признаков сигнала.

При увеличении отрицательного перепада наблюдается тенденция к увеличению числа ответов «5-й тон», хотя в части сигналов это сопровождается восприятием в конце слова глухого смычного согласного.

Отметим также довольно частые случаи, когда сигнал, имеющий формально (по результатам измерения на периодограмме) большой перепад длительности периода О.Т., воспринимается слушателями в 1-м тоне, для которого характерен стационарный мелодический контур. Если учесть, что ответы испытуемых основываются на внутреннем – субъективном – описании мелодического контура, а не на физическом описании, наблюдаемом на периодограммах, подобные «парадоксальные» результаты могут помочь в поиске тех экспериментальных процедур, которые позволят найти правила перехода от физического описания контура к субъективному и проверить их. При этом необходимо будет учитывать как влияние контекста (считая и нулевой), так и разные факторы, относящиеся к высшим уровням языковой системы, в том числе распределение лексических вероятностей.

Изложенные выше результаты являются предварительными, а предполагаемые «кандидаты» в воспринимаемые признаки тонов требуют дальнейшего строгого количественного исследования.

## РОЛЬ СЛОВАРЯ В ПЕРЦЕПТИВНЫХ ПРОЦЕССАХ. КЛЮЧЕВЫЕ СЛОВА

Вопрос о словаре и его роли в процессах восприятия речи уже затрагивался в общем виде в гл. II. Здесь мы несколько конкретизируем высказанные ранее положения и приведем данные, полученные при изучении словаря русского языка с точки зрения перцептивно релевантных признаков словарных единиц.

Перцептивные аспекты проблемы словаря обычно рассматриваются с двух точек зрения: с одной стороны, обсуждается структура словаря, с другой – процедуры отождествления текстовых единиц с единицами словарными. Пожалуй, не совсем точно приравнять разграничение указанных двух

аспектов (как это иногда делается) к различению декларативных и процедурных знаний соответственно. Важнейший компонент процесса восприятия речи – это, действительно, процедуры идентификации определенных сегментов звучащей речи с известными слушающему словарными единицами. Столь же очевидно, что словарь обладает определенной внутренней организацией. Однако неверно было бы думать, что эти два аспекта полностью независимы. В известном смысле организация словаря отражает процедуры, с помощью которых и осуществляется переход от текста к словарю и, далее, к языковой характеристике текста, в конечном счете – к его пониманию. Ведь словарь, как и вся языковая система, частью которой он является, не столько имеет самостоятельную ценность, сколько обслуживает коммуникативный процесс.

Более точно ситуацию можно описать следующим образом. Функции словаря многообразны. Так, есть все основания говорить о его номинативной функции, которая заключается в том, что словарь прямо отражает категоризацию объектов и явлений действительности, выработанную соответствующим культурно-историческим сообществом и закрепленную, в целях поддержания культурной традиции, в системе лексических единиц. Здесь явным образом пересекаются когнитивная и коммуникативная предназначенность языка.

Коммуникативная предназначенность языка вызывает к жизни две другие функции словаря: продуктивную и перцептивную. Иначе говоря, словарь обслуживает и порождение, и восприятие речи. Вполне очевидно, что он должен быть «приспособлен» для эффективного выполнения этих функций, т.е. обладать такой внутренней организацией, которая максимально способствовала бы успешному протеканию соответствующих процессов. Отсюда и следует, что знание внутренней организации словаря – знание декларативного типа – не может не быть в значительной степени знанием о том, как используются в речевой деятельности словарные единицы, т.е. знанием процедурного типа.

Мы не будем обсуждать тип организации словаря применительно к его номинативной и продуктивной функциям, как не будем и заниматься вопросом о том, насколько различается устройство словаря, рассматриваемое с точки зрения его различных функций. Нас будут интересовать лишь перцептивные аспекты проблемы.

Прежде всего полезно различать «общий» словарь и словарь «текущий». Дело в том, что человек, воспринимающий речь, как отчасти уже упоминалось, оперирует, вероятно, не всем («общим») словарем, а некоторым подсловарем, формируемым применительно к данному коммуникативному акту. Именно такой подсловарь мы называем «текущим» словарем.

Говоря о том, что «текущий» словарь «формируется», мы, пожалуй, не очень точно отражаем соответствующий процесс. Точнее, вероятно, было бы утверждать, что «общий» словарь разбит на потенциальные «текущие» по тематическому принципу, примерно так же, как это имеет место для словарей идеографического, или тезаурусного типа, создаваемых лексикографами. В процессе восприятия речи один из тематических подсловарей может активироваться, в результате чего и появляется возможность обращаться к «текущему» словарю. Уровень активации не используемых в данный момент подсловарей, будучи существенно ниже, как бы временно выводит их из игры, тем самым поле поиска словарных единиц существенно сужается, делая перцептивный процесс более экономным, менее времязатратным.

За счет чего же происходит активация потенциального подсловаря и перевод его в ранг словаря «текущего»? Многочисленные эксперименты по праймингу в его различных модификациях показали, в частности, что, «когда такому слову, как *няня*, [в эксперименте] предшествует семантически связанное с ним слово *доктор*, время принятия решения о слове (время, требующееся для того, чтобы решить, является ли [звуковая последовательность] *няня* словом) меньше, чем в случае, когда [стимулу] *няня* предшествует не связанное с ним слово. Одним из возможных механизмов, объясняющих такой результат, является "распространение активации" через пути, объединяющие тесно связанные слова, что и способствует более быстрому распознаванию таких связанных слов» (Emmorey, Fromkin, 1989, p. 143).

Иначе говоря, при появлении в «перцептивном поле» слушающего любого из слов, принадлежащих к семантической сети, которую образуют содержательно соотнесенные слова, активируется вся сеть. Активированная сеть и соответствует «текущему» словарю.

По-видимому, следует предусмотреть и два других способа формирования «текущего» словаря. Один заключается в том, что семантическую сеть активирует сама проблемная

ситуация. Другой состоит в возможности образовывать в процессе восприятия новые сети, т.е. устанавливать новые связи, вводить в словарь новые единицы и т.п. Это – один из элементов самообучения системы. Если в условиях, когда уже существующая семантическая сеть активируется самой проблемной ситуацией, становление «текущего» словаря осуществляется фактически до акта восприятия речи, то формирование нового подсловаря, новой сети возможно лишь по мере накопления информации о соответствующей предметной области – или по крайней мере о ее вербальном отражении – в процессе восприятия текста.

В связи с обсуждением проблемы «текущего» словаря естественно возникает вопрос о ключевых словах: не являются ли элементы такого словаря ключевыми словами по отношению к данному тексту? Представляется, однако, что прямое отождествление единиц двух типов неправомерно. Если считать, что «текущий» словарь есть результат активации уже существующей семантической сети (оставляем в стороне вариант, когда словарь формируется по мере знакомства с текстом), то практически неизбежна ситуация, при которой «текущий» словарь существенно шире по объему, чем набор ключевых слов, отвечающих заданному тексту. Совпадение возможно лишь тогда, когда восприятию подлежит некоторый универсум текстов, исчерпывающе отражающих соответствующую предметную область.

Можно сказать, вероятно, что набор ключевых слов для заданного текста представляет собой подмножество из совокупности словарных единиц, которые принадлежат «текущему» словарю, активированному текстом.

Такой подход, как представляется, проливает дополнительный свет на давно дебатлируемые вопросы о соотношении ключевых слов и такого параметра лексических единиц, как частотность. Ключевые слова по определению являются наиболее информативными, а информативность связывают с частотностью. Так, Л.М.Финк пишет: «Пусть вероятность  $i$ -го слова в словаре равна  $p_i$ . Средняя собственная информация или энтропия на одно слово (если не учитывать связей между словами) равна  $\sum p_i \log(1/p_i)$ , где суммирование производится по всем словам языка. Вклад  $i$ -го слова в эту сумму определяется величиной  $p_i \log(1/p_i)$ , которая имеет максимум при  $p_i = 1/e$ , а с уменьшением значения  $p_i$  монотонно убывает. Так как в любом естественном языке все  $p_i \ll 1/e$ , то отсюда следует,

что наибольшая часть информации связана с наиболее частыми словами в полном соответствии с практикой» (Финк, 1978, с. 250). Все это безусловно верно, однако, во-первых, необходимо различать частотность слова в словаре (или подсловаре) и его же частотность в конкретном тексте; во-вторых, если оставить в стороне случай с формированием нового «текущего» словаря по ходу восприятия текста, то «текущий» словарь активируется либо априорно по отношению к тексту, либо в самом начале процесса восприятия, в то время как набор ключевых слов, в особенности если выводить его по данным частотности, формируется апостериорно. Решение вопроса как раз и заключается в том, что «текущий» словарь задает широкую тематику, всю предметную область, к которой относится данный текст, а набор ключевых слов очерчивает в ней определенную подобласть. «Текущий» словарь находится «у истоков» воспринимаемого текста, набор же ключевых слов, или, вернее, их структура, выступает как результат процесса восприятия. Набор ключевых слов – это семантика текста, свернутая до иерархической структуры его тем и подтем (и в этом смысле лишь часть семантической структуры, ибо, владея информацией о теме, но не зная ремы, мы не можем претендовать на владение семантикой текста).

В тексте, конечно, представлена не структура ключевых слов, а их последовательность, разделенная словами более низкой информационной значимости. Соответственно возникает вопрос о том, возможно ли вычленение ключевых слов текста по некоторым формальным признакам. Если такая операция возможна, то она должна способствовать повышению эффективности восприятия, поскольку тем самым для слушающего облегчается задача сужения предметной области, задаваемой «текущим» словарем, до семантического пространства данного текста. Языки с этой точки зрения существенно различаются. Как хорошо известно, достаточно распространены языки, в которых существуют специальные грамматические средства для выделения слова-темы, как, например, *—wa* в японском языке. Как было сказано выше, ключевые слова в типичном случае ассоциируются с темами высказываний, содержащихся в тексте, поэтому маркеры, обслуживающие тему, обычно выделяют одновременно ключевые слова.

По данным А.Финн (Finn, 1984), в японском языке топикализация (выделение темы) сопровождается также использованием просодических средств: после служебного слова *—wa*

обычна пауза, а перед —*ва* наблюдается мелодический перелом, причем длительность паузы и крутизна мелодического перелома дополнительно служат средством различения нейтральной и контрастной темы.

Здесь мы встречаемся, с одной стороны, с семантическими категориями, связанными с понятием темы (ср. категорию контрастности), с другой — с ролью просодических средств для вычленения темы и слов родственной семантики, а тем самым и для вычленения ключевых слов. По-видимому, одним из первых вопрос о систематической связи просодических структур текста и его информационных структур, в которые входит и тема, поставил Дж. Хэллидэй (Halliday, 1970). Он сформулировал положение, согласно которому текст членится на отрезки (*chunks*), в которых одна часть, факультативная, отвечает данному, а другая, обязательная, — новому, причем новое выделяется просодически — повышением мелодики. Новое, в частности, — это впервые для данного текста введенная тема, поэтому повышение мелодики также может служить маркером для обнаружения ключевого слова, причем именно тогда, когда оно употребляется впервые, а потому особенно важно. Дальнейшие экспериментальные исследования в основном подтвердили утверждения Хэллидэя. В частности, было показано, что первая — по порядку введения — из тождественных тем действительно получает просодическую выделенность, при повторной же номинации в пределах того же текста выделенность утрачивается. Экспериментальные материалы показали также, что возможно маркирование информационно наиболее значимых единиц текста не только повышением, но и понижением мелодики (Terken, 1980; Brown, 1983).

Выше упоминалось о роли паузы как средства обнаружения тематических слов (и их подвидов), а следовательно, в типичном случае и ключевых слов. Данные, полученные нами на материале обследования китайских научно-технических текстов, показали, что выбор паузы определенной длительности и позиции по отношению к слову позволяет опознавать в слитном звучащем тексте слова, принадлежащие к ключевым или близкие к ним по семантике. В нашем случае пауза в интервале 520-730 мс служила пусковым сигналом для выделения слова, находящегося слева от паузы. Указанным способом из текста с продолжительностью звучания около 20 мин было автоматически вычленено 61 слово, из которых 31 слово безусловно принадлежало к ключевым, а среди оставшихся 30—ти 11 можно

было отнести к «информационно полезным» словам, т.е. к таким, которые, можно полагать, принадлежат к «текущему» словарю, но не отражают специфику данного текста. Для предметной области «полупроводники – импульсная техника», с которой мы работали, приведем такие примеры, как *фумайчун* «отраженный импульс» в качестве ключевого слова и *босинту* «график» – в качестве «информационно полезного» слова.

Таким образом, даже с опорой на очень простые формальные признаки удается вычленить единицы, среди которых почти 70% составляют ключевые и «информационно полезные» слова. Близкие к этому результаты были получены также на материале японских и арабских научно-технических текстов.

Как показывают результаты ряда экспериментов, в том числе и наши данные, по набору ключевых слов, вычленившихся из текста, испытуемые обычно в состоянии предложить реферат последнего, если порядок предъявления ключевых слов воспроизводит их последовательность в тексте, см. особенно (Мурзин, Штерн, 1991). Иначе говоря, порядок расположения ключевых слов в известной степени компенсирует отсутствие рем (ср. выше), т.е. как бы берет на себя их функцию (Мурзин, Штерн, 1991).

Если организация словаря и текста с точки зрения представленности ключевых слов релевантна для семантического «полюса» перцептивного механизма, то для фонологического «полюса», как легко предвидеть, существенным должно оказаться распределение фонологических признаков, среди которых может быть длина слова в терминах слогов и/или фонем, акцентный контур, просодии-гиперпризнаки и др. По каждому из признаков можно выделить классы эквивалентности, играющие роль в процессах восприятия речи.

По данным Д.Пизони и др., на которые ссылается Дж.Альтманн, уже учет длительности английских слов в терминах фонем позволяет сократить объем класса поиска со 126 тыс. слов (полный объем использованного словаря) до 6342 слов, т.е. до 5% от массива словаря. Если к этому показателю добавить только лишь информацию, позволяющую различать гласные и согласные, объем класса уменьшается до 0,1% от массива словаря (Altmann, 1990b, p. 212).

Есть и данные, показывающие сужение класса поиска за счет использования информации об ударении: в английском языке класс эквивалентности для слов с одинаковым акцентным контуром составляет 15 – 19% от общего объема словаря, см.

об этом (Altmann, 1990b, p. 214). Заметный эффект дает и обращение к гиперпризнакам. Если для описания экспонентов английских слов использовать только признаки «гласный», «носовой», «сильный ффрикативный», «слабый ффрикативный», «смычный» и «плавный», когда, например, слово *piston* транскрибируется как /смычный, гласный, сильный ффрикативный, смычный, гласный, носовой/, то средний объем класса поиска составляет 0,01% от всего словаря, а 32% слов получают единственно возможную фонемную интерпретацию (Altmann, 1990b, p. 214-215).

Это все довольно грубые и не слишком достоверные оценки, поскольку не принимаются во внимание ни частотность фонем (и слов), ни слоговая структура, ни целый ряд других немаловажных факторов. Существуют и более тонкие оценки эффекта, который достигается благодаря использованию глобальных фонологических характеристик слова в словаре и их статистического распределения (Altmann, 1990b).

Для русского языка, насколько нам известно, аналогичные подсчеты в сколько-нибудь серьезных масштабах до сих пор не производились. Ниже приводятся некоторые предварительные материалы, полученные нами на базе компьютерной версии «Частотного словаря современного русского литературного языка» (Штейнфельдт, 1963).

Словарь содержит 2500 слов, отражающих 315279 реализаций в текстах из детской и юношеской литературы. Структурируя словарь, мы исходили из допущения, что для начального этапа обнаружения подходящего слова в словаре достаточно использовать первые 1–2 элемента дискретного внутреннего представления сигнала с последовательным сужением класса поиска (см. описание кортежных моделей в гл. II). При этом в качестве таких элементов могут быть приняты участки речевого сигнала фонемной протяженности, описанные в терминах крупных фонетических классов. Для характеристики последних мы выбрали категории, наиболее устойчивые к помехам: «глухие/звонкие», «смычные», «ффрикативные», «сонорные/шумные» и, внутри сонорных, «дрожащие» ([r]) (рис. 3.14). Кроме того, был использован и такой признак слова, как его ритмическая структура, в данном случае позиция (номер) ударного слога относительно начала слова. Для этого в словарь в качестве самостоятельных элементов были включены слова, представляющие единственное и множественное число соответствующих элементов исходного словаря и различающиеся ритми-

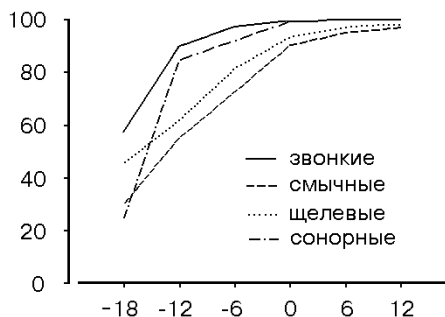


Рис. 3.14. Зависимость степени сохранения признаков крупных фонетических классов от уровня помех при восприятии начальных согласных.

По оси абсцисс — отношение сигнал/шум, дБ; по оси ординат — разборчивость признаков, проценты.

ческой структурой (ср. окно — окна). В результате рабочая версия словаря содержала 3092 единицы.

Исходная рабочая версия была транскрибирована с помощью программы, разработанной на кафедре прикладной лингвистики Московского государственного университета, и последующая работа проводилась с транскрибированной версией.

Приведем количественные данные об объемах «подсловарей», полученных в результате структурирования слова-

ря по фонологическим признакам начала слова (числитель каждой дроби дает количество слов в «подсловаре», а знаменатель — суммарную частотность соответствующих слов; первая колонка чисел — для слов с глухим началом, вторая — со звонким):

смычное начало	816/60967	346/33515
фрикативное	580/48550	413/53153
сонорное	—	373/44946
[r]	—	145/8947
гласное	—	419/64201
в целом по словарю	1396/110517	1696/204762

Возможности сокращения зоны поиска, которые дает выбранная схема структурирования словаря, можно видеть на примере слов, первый элемент которых описывается как «дрожащий» — [r]. Объем этого раздела словаря составляет 4,7% от исходной рабочей версии, а суммарная частотность — 2,8%. Если же в этом разделе выделить только слова с первым ударным слогом (их в нем всего 36), то зона поиска сужается до 1,2% (при частотности в 0,7% полного словаря).

Таким образом, использование только признаков крупных фонетических классов согласных и ритмической структуры позволяет существенно сузить область лексического поиска. Дальнейшее сужение возможно за счет признаков крупных фонетических классов гласных (например, «закрытый/открытый»). В рассмотренном выше случае это сужает зону поиска до 17–19 слов из исходных 3 тысяч. Если же использовать и грамматическую информацию (например, искать только существительные), то при открытом ударном гласном выбирать придется только из четырех слов: раз, радость, радости и радио. Очевидно, что знание предметной области соответствующего текста может свести выбор к одной единственной возможности – единичному слову.

#### ПСИХОАКУСТИКА И МОДЕЛИ ВОСПРИЯТИЯ РЕЧИ

Все, о чем говорилось в предыдущих главах, приводит к заключению, что при восприятии речи «слуховая» система человека работает в специфическом («речевом») режиме. При этом носители разных языков для решения похожих задач могут использовать разные характеристики акустического речевого сигнала и выделяемые из них субъективно-психологические признаки.

Точное количество и функциональные особенности всех образующих систему блоков (модулей) в данный момент назвать невозможно. Тем не менее есть основание говорить, по крайней мере, о двух группах функциональных модулей (подсистемах), предположительно различающихся отсутствием между ними связей «сверху вниз» (обратных связей). Эти модули как бы независимы друг от друга.

Независимы в том смысле, что результат работы каждого из них не сказывается на структуре и принципах работы другого. Это психоакустический модуль, производящий

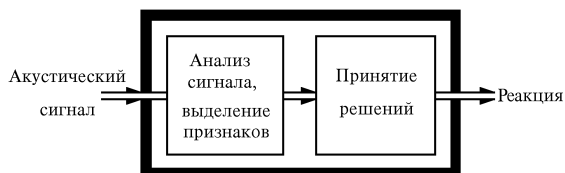


Рис.4.1. Схема условного деления системы восприятия речи на подсистему выделения признаков (психоакустический процессор) и подсистему принятия решений (фонетико-фонологический модуль).

первичную обработку акустического сигнала и выделение признаков, и подсистема принятия решений (рис. 4.1), включающая фонетико-фонологический модуль (см. гл. II).

При первичной обработке используются одинаковые принципы преобразования для всех акустических сигналов, с которыми приходится встречаться человеку, в том числе и для речевых: при этом внешний (физический) акустический сигнал преобразуется во внутреннее (субъективное) его представление и описывается изменяющимися во времени признаками, количество и конкретное содержание которых нам пока неизвестны.

Для принятия решений об общей стратегии работы с признаками и о выраженных в них конкретных особенностях обрабатываемого сигнала используется внутреннее представление акустического сигнала. По существу, может использоваться не один, а несколько модулей по числу классов акустических сигналов, требующих различающихся принципов обработки. Как минимум, речь может идти о двух разновидностях подсистемы: для обработки речевых и неречевых сигналов. Очевидно, одним из первых должно приниматься решение «речевой/неречевой».

Создание работоспособной функциональной модели восприятия акустических сигналов человеком, которая могла бы адекватно работать с естественными сигналами (речевыми в том числе), требует не только знания структуры образующих ее алгоритмов, но также их качественных и количественных параметров. Для исследования искусственных систем такой сложности с успехом применяется метод «черного ящика», когда об алгоритмах, реализуемых каждым из элементов системы, судят по результатам сопоставления множества его выходных сигналов с соответствующим множеством входных. К сожалению, мы не имеем доступа к выходным сигналам первого модуля, поскольку не знаем точного места их локализации в слуховой системе. Кроме того, по вполне понятным причинам мы лишены возможности непосредственно наблюдать работу отдельных нейронных структур человеческого мозга, которые, собственно говоря, и образуют то, что принято называть системой восприятия речи. Поэтому приходится прибегать к различного рода косвенным методам наблюдения и исследования слуховой системы человека.

В настоящее время практически используются два способа получения сведений о функционировании системы восприятия акустических сигналов и количественных данных для построения ее адекватной модели: нейрофизиологические (электрофизиологические) исследования свойств отдельных нервных

клеток (нейронов) и нейронных структур, образуемых этими клетками, и психоакустические исследования типа «стимул–реакция».

Данные, полученные в нейрофизиологических исследованиях, широко и с успехом используются при моделировании самых первых (периферических) этапов преобразования внешнего акустического сигнала в его внутреннее субъективное представление (Чистович и др., 1986; Seneff, 1988; Shamma, 1988). На этом этапе существенные для дальнейшей работы системы преобразования происходят либо в одиночных нейронах, либо в ограниченных ансамблях однотипных нейронов, так что конечный результат преобразования может быть предсказан на основании известных свойств одиночного нейрона, входящего в ансамбль. Не менее важно и то, что на этом уровне еще сохраняется некоторая возможность одновременно контролировать и входной и выходной сигналы исследуемого модуля, поскольку интенсивность перекрестных связей между образующими его элементами невелика.

В дальнейшем структура нейронных сетей становится настолько сложной, что исследователь оказывается в состоянии контролировать только их выходные сигналы. Причем ему приходится устанавливать их связь не с параметрами входных сигналов соответствующего модуля, а с параметрами внешнего акустического сигнала. Соответственно, речь может идти только о гипотетических структуре и алгоритмах работы такого модуля, конструируемых исходя из вероятных сигналов на его входе. Таким образом, к проблеме создания функциональной модели восприятия речи добавляется проблема адекватных методов исследования нейронных сетей, образующих ее природный прототип.

Кроме того, исследуя один нейрон или ансамбль нейронов, экспериментатор не может с уверенностью сказать, к какому функциональному уровню преобразования они относятся, поскольку не располагает точными данными о локализации анатомо-морфологических структур, производящих предусмотренные функциональной моделью преобразования. Возникает необходимость оценивать, какие параметры внешнего акустического сигнала могла бы выделять исследуемая часть нейронной сети, а затем проверять в психоакустическом эксперименте, в какой мере эти параметры используются человеком при восприятии речи. Следовательно, заранее невозможно утвер-

ждать, что исследуемый участок нейронной сети имеет отношение к системе восприятия речи.

Наконец, следует иметь в виду, что единственный доступный нам способ прямого наблюдения деятельности нейронных структур – это опыты на лабораторных животных, поэтому использование полученных при этом сведений для создания модели восприятия речи человеком предполагает предварительную проверку допустимости такого переноса. С этой целью данные нейрофизиологии сопоставляются с психофизическими данными, полученными на животных, а последние с аналогичными данными, полученными на человеке.

Еще сложнее обстоит дело, когда исследования модуля первичной обработки проводятся методами психофизики на человеке. Вполне очевидной особенностью этого модуля является полная неосознаваемость слушающим непосредственных результатов его деятельности. Эти результаты проявляются только в тех окончательных решениях, которые принимает на их основе другой модуль. Таким образом, исследователь «видит» результаты работы первого модуля только «сквозь» второй. Очевидно, в этих условиях невозможно судить о выходных сигналах первого модуля, не зная точных алгоритмов работы подсистемы принятия решений. Другими словами, мы приходим к необходимости работать с целостной моделью, чтобы получить представление о содержании и конкретных количественных параметрах отдельных ее частей.

Для изучения человека в психофизике разработаны и широко используются разнообразные методы количественного и качественного исследования. Представляется необходимым отметить некоторые характерные особенности психофизических методов, проанализировать влияние этих особенностей на результаты, получаемые при использовании таких методов, и оценить возможность применения психофизических данных о работе сенсорных систем (в особенности количественных параметров этих систем) в создаваемых функциональных моделях, прежде всего в моделях восприятия речи.

При ближайшем рассмотрении оказывается, что в любом психофизическом эксперименте (независимо от его конкретной методики) задача испытуемого может быть сведена к классификации предъявляемых стимулов по тем или иным правилам.

Это совершенно очевидно, если речь идет о методах одиночного стимула, когда самой инструкцией испытуемому задается требование отнести каждый из предъявленных сиг-

налов к одному из заранее заданных классов (идентифицировать сигнал).

Менее очевиден случай парного сравнения двух сигналов, но и его можно свести к классификации, если допустить, что стимулом является вся пара, а оцениваемым признаком – разность величин исследуемого параметра в сигналах, образующих данную пару. Тогда при классификации «стимул», в котором эта разность не превосходит заданного порога, будет вызывать ответы «одинаковые», в противном случае – ответы «разные». Если задана трехальтернативная система оценок, то при отрицательной разности будут возникать ответы «меньше», при положительной – «больше», а в остальных случаях – ответы «равны».

Даже при численном субъективном шкалировании испытуемым приходится классифицировать предъявляемые сигналы, соотнося их с единицами заданной шкалы, тем более, что в большинстве случаев им предписывается соотносить параметры сигналов с ограниченным диапазоном целочисленных «делений» шкалы.

Как следует из анализа современных моделей восприятия речи (см. гл. I), процедуры классификации составляют основу многих из них (особенно на стадии «обучения» конкретному языку). В гл. II отмечалось, что способность к самообучению является основой системы восприятия речи и должна учитываться при создании соответствующей функциональной модели. Естественно возникает необходимость оценить, в какой степени такая способность используется в ходе психоакустического эксперимента и как она может повлиять на его результаты. Для этого попытаемся проанализировать некоторые из наиболее широко используемых психоакустических методик и оценить, в какой мере полученные с их помощью количественные характеристики системы восприятия могут быть использованы при построении функциональной модели восприятия речи.

## МЕТОД ИДЕНТИФИКАЦИИ И ВОЗМОЖНОСТИ ИССЛЕДОВАНИЯ ПЕРЦЕПТИВНЫХ ЭТАЛОНОВ

Как уже говорилось выше (см. гл. II), появление речевого акустического сигнала на входе функциональной модели системы восприятия речи человеком приводит в действие две подсистемы: подсистему преобразования исходного акустического сигнала во внутреннее (субъективное) представление в виде

непрерывно изменяющихся субъективных параметров и многоуровневую подсистему принятия решений, для которой эти параметры являются исходной информацией.

Одним из первых этапов принятия решений, если не самым первым, должна быть фонетическая интерпретация внутреннего (субъективного) представления речевого сигнала (пока неизвестным числом непрерывно изменяющихся параметров), подразумевающая переход к более «сжатому» описанию сигнала в виде последовательности дискретных категорий (единиц, классов). По-видимому, количество и конкретный вид этих категорий должны определяться потребностями уровней лингвистического или, точнее, собственно языкового анализа сигнала, и в данный момент невозможно дать сколько-нибудь полный перечень таких категорий.

Однако, каковы бы ни были эти дискретные категории и какие бы параметры сигнала ни использовались для их описания, процедура перехода от непрерывного параметрического описания сигнала к дискретному (или от предварительной сегментации, обеспечиваемой психоакустическим модулем, к «лингвистической») обязательно должна включать этап сопоставления текущих значений параметров с некоторыми «эталонными» их значениями, характеризующими каждую дискретную категорию. Эталоны должны вырабатываться в процессе обучения языку и могут быть разными у носителей разных языков.

Предположим, например, что эталонная область гласных представлена кругом (эллипсом) на плоскости  $F_1$ – $F_2$  (это не слишком большое упрощение, так как большинство гласных уверенно идентифицируется по положению первых двух формант). Если на вход системы фонетической интерпретации с таким эталоном поступает множество реализаций разных гласных, то с учетом возможных ошибок субъективного измерения параметров сигналов («частот» формант) вероятность отнесения этих сигналов к одной из категорий будет представлена «колоколом», вершина которого совпадает с центром круга-эталона (эллипса). Если же физические параметры используемых в психоакустическом опыте сигналов располагаются на линии в плоскости  $F_1$ – $F_2$  (наиболее часто встречающийся вариант построения экспериментального набора), распределение вероятностей должно быть представлено соответствующим сечением «колокола» (см. рис. 4.2).

Но такое же понятие – «эталон» (стандарт) – присутствует в описании большинства психофизических экспериментов в ка-

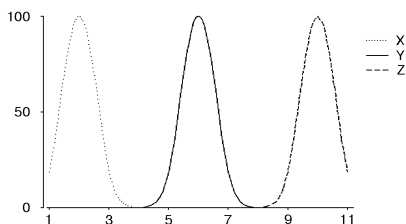


Рис. 4.2. Гипотетические «кривые идентификации» трех условных единиц языка (X, Y, Z), полученные методом свободного выбора. Распределения неопределенных ответов (не-ответов) на рисунке не показаны.

По оси абсцисс – номер «стимула» в предполагаемом экспериментальном наборе; по оси ординат – относительная частота «ответов», проценты.

проводимого методом одиночных постоянных стимулов (точнее, предполагаемая модель поведения испытуемого в таком эксперименте). Этот метод и называют обычно идентификацией, или классификацией.

В качестве стимулов, как правило, используется набор сигналов с заданными числовыми значениями изучаемого параметра. Эти значения перекрывают некоторый ограниченный диапазон и остаются неизменными в течение всего эксперимента. Число градаций исследуемого параметра обычно не превышает 10–15, и каждый сигнал с заданным значением параметра повторяется в наборе несколько раз. В результате в каждом эксперименте испытуемому предъявляются 100–200 сигналов. Сигналы предъявляются поодиночке, и для каждого из них испытуемый должен дать оценку путем отнесения его к одному из дискретных классов. Количество и конкретные типы классов могут быть либо заданы экспериментатором (принудительный выбор), либо выбираются самим испытуемым на основании его внутреннего представления о характере предъявляемых сигналов (применительно к речевым сигналам это можно рассматривать как своего рода диктант).

честве его цели и результата. Естественно, возникает потребность проанализировать, в какой степени полученные в таком эксперименте данные об эталонах могут быть использованы в функциональной модели восприятия речи. Точнее, нас должно интересовать, как следует построить психофизический (психоакустический) эксперимент, чтобы его результаты можно было использовать в действующей модели.

Наибольшее сходство с гипотетической процедурой фонетической интерпретации обнаруживает процедура психофизического эксперимента,

По результатам эксперимента для каждого сигнала подсчитывается количество ответов каждого типа и строятся графики распределения количества ответов (или относительной частоты появления каждого типа ответов) в зависимости от величины исследуемого параметра. Получаются так называемые кривые идентификации (рис. 4.3).

Экспериментальные данные аппроксимируют аналитической функцией (чаще всего – функцией нормального распределения вероятностей) или некоторой вероятностной моделью, описывающей предполагаемое поведение испытуемого, и параметры стимулов, соответствующие абсциссам точек пересечения кривых, принимают в качестве «точек субъективного равенства». Применительно к речевым и речеподобным сигналам эти параметры используют в качестве «фонемной границы» (разумеется, в тех случаях, когда исследуется восприятие сегментных признаков сигнала).

Очевидно, что получающиеся в подобных экспериментах кривые идентификации существенно отличаются от кривых, о которых мы говорили выше (рис. 4.2). Соответственно отличаются и процедуры определения положения «эталона» и «фонемной границы».

Попытаемся понять природу этих различий и оценить степень соотнесенности «фонемных границ» с гипотетическими эталонами, используемыми человеком при фонемной интерпретации естественной речи.

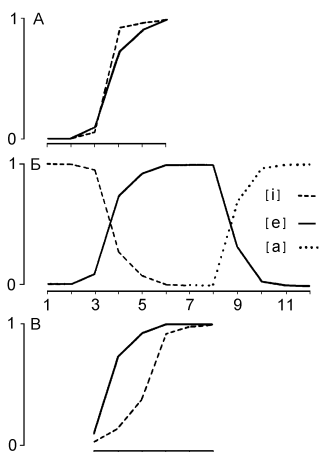


Рис. 4.3. Результаты опытов по идентификации синтезированных гласных. А и В – результаты идентификации гласного [е] в опытах с «усеченными» наборами при двухальтернативном вынужденном выборе (штриховые кривые) в сравнении с аналогичными данными, полученными для полного набора при трехальтернативном выборе (сплошные кривые); Б – результаты опытов с полным набором при трехальтернативном выборе. По оси абсцисс – номер гласного в наборе из 12 сигналов, упорядоченном по частотам первой и второй формант; по оси ординат – относительные частоты ответов.

Возможны, по крайней мере, два фактора: один обусловлен особенностями организации конкретного эксперимента, другой определяется возможной стратегией поведения испытуемого в опыте.

Могло случиться так, что исследователь выбрал для эксперимента диапазон изменения параметров сигналов, который не перекрывает полностью все предполагаемые эталонные эллипсы в пространстве признаков. Тогда стимулы, соответствующие концам выбранного диапазона, могли оказаться в центре эталонных эллипсов, а в результате мы не обнаружили ожидаемого снижения частоты (вероятности) появления соответствующих оценок (см. данные для [i] и [a] на рис. 4.3,Б). Смущает лишь то обстоятельство, что подобные кривые идентификации получают-ся во всех психофизических экспериментах, проводимых с использованием такой методики, и маловероятно, что экспериментаторы всегда выбирают параметры стимулов столь неудачно.

Вместе с тем в естественной речи заведомо будут представлены сигналы (участки), в равной степени не относящиеся ни к одному из имеющихся у слушателя эталонов языковых единиц. Это обусловлено и неизбежными акустическими шумами, и сильной редукцией безударных гласных (слогов), если речь идет о русском языке, и шумами собственно системы восприятия. При сложившейся у носителя языка системе эталонов эти сигналы останутся неинтерпретированными, и при необходимости давать им оценку в опыте испытуемый должен будет оценивать их как не-стимулы, т.е. как не относящиеся ни к одному из известных ему классов. Применительно к тому опыту, результаты которого приведены на рис. 4.3,Б, испытуемый должен был бы на часть стимулов отвечать «не [i], не [e] и не [a]», т.е. должен был бы использовать 4 класса ответов. Когда же использование не-класса испытуемому запрещено (принудительный выбор), все попадающие в этот класс стимулы он вынужден относить к одному из разрешенных инструкцией классов, пользуясь некоторыми правилами. Важно отметить, что правила «выбирает» (причем подсознательно) сам испытуемый, и экспериментатору остается только строить предположения об их конкретном содержании.

Если таким правилом является простое гадание, то предполагаемые не-ответы будут в равных долях заменены ответами, заданными инструкцией, и предполагаемые кривые идентификации, сохранив до некоторой степени «колоколообразную»

форму, примут вид, представленный на рис. 4.4. Очевидно, что здесь мы практически не приближаемся к результатам, получаемым в психоакустическом эксперименте.

Другая возможность приблизить результаты работы модели к результатам эксперимента заключается в изменении стратегии приписывания оценки не-стимулам: оценивается субъективное расстояние до каждого из используемых в опыте этапов, и ответ соответствует тому из них, расстояние до которого окажется минимальным (Nearey, 1990). При равенстве расстояний решение принимается гаданием. При этих условиях стимул 1 (рис. 4.2) почти всегда будет отнесен к классу «X», и в сумме такой ответ будет получен в 100% случаев, а стимул 11 – к классу «Z» с тем же результатом. С учетом естественной вариативности субъективных расстояний модельные «кривые идентификации» могут стать похожими на те, которые получаются в эксперименте (ср. рис. 4.3,Б). Другими словами, испытуемый будет оценивать сигналы по их похожести на тот или иной гласный.

Способность испытуемых использовать во время опыта разные процедуры оценки одного и того же стимула можно проиллюстрировать данными о восприятии носителями чешского языка (100 испытуемых) набора из 200 синтезированных гласных (Janota, 1967). Их просили оценить каждый стимул в терминах пяти гласных фонем (а, о, и, е, и) и при этом отмечать те из них, которые, по мнению испытуемых, соответствуют гласным чешского языка. С обоими заданиями испытуемые справились без особых затруднений. Не идентифицированными оказались лишь 0,43% стимулов, но только 22% стимулов были оценены как «чешские» гласные. Таким образом, обнаруживается способность слушателя идентифицировать как некоторые языковые единицы даже те сигналы, которые в естественной речевой среде данного носителя языка не встречаются (правда, в рабо-

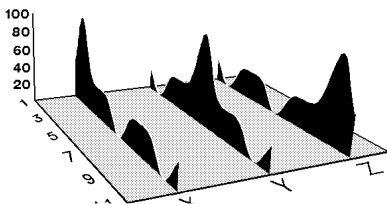


Рис. 4.4. Гипотетические «кривые идентификации» трех условных единиц языка (X, Y, Z), полученные методом вынужденного выбора, (оценки неопределенных сигналов выбираются гаданием с равной вероятностью появления каждой из трех допустимых оценок). По вертикали – вероятность отнесения стимулов (с 1-го по 11-й) к каждому из трех классов, проценты.

те нет сведений, в какой степени испытуемые владели другими языками).

Очевидно, в рамках такой гипотетической модели поведения испытуемого в экспериментальной ситуации «фонемные границы» должны соответствовать стимулам, равноудаленным от всех эталонов в пространстве исследуемых признаков. Очевидно также, что положение в этом пространстве самих эталонов по результатам таких экспериментов установить невозможно. Отметим еще раз, что в системе «фонемных границ», как правило, не находится места ни области «неопределенных» сигналов (не похожих ни на одну из известных испытуемому дискретных единиц или похожих на множество таких единиц), ни очерчивающим эту область границам. В то же время в естественной речи такие сигналы (участки речевого потока) встречаются обязательно, о чем говорилось выше.

Следует добавить, что, по данным многих исследований, «точка субъективного равенства» в пространстве физических параметров сигнала сдвигается при изменении выбранного исследователем диапазона варьирования этих параметров в конкретном опыте, но ее положение остается неизменным, будучи выражено в терминах номеров стимулов (разумеется, при условии, что номера присваиваются в порядке возрастания/убывания исследуемого параметра). Эффект этот наблюдается как в опытах с достаточно простыми сигналами (см. рис. 4.10), так и в опытах с речевыми сигналами (Krause, 1982).

Таким образом, на результаты опытов оказывает влияние изменение «контекста», в котором предъявляются те или иные сигналы: физически один и тот же стимул (например, груз заданного веса или синтезированный гласный) оценивается испытуемым по-разному в зависимости от того, в каком наборе он предъявляется.

Примером такого влияния являются, в частности, результаты исследования, проведенного Кузьминым и Лисенко (1971). В опытах использовались два набора (сигналы с 1-го по 6-й и с 3-го по 8-й) синтезированных гласных, взятых из полного набора в 12 гласных. Спектральные характеристики (частоты формант) гласных были выбраны так, что 1-й сигнал соответствовал гласному [i], 6-й и 7-й – гласному [e], а 12-й – гласному [a]. Остальные сигналы были промежуточными между [i]–[e] и [e]–[a] (подробное описание сигналов см. в работе: Chistovich e. a., 1966).

Ранее сигналы этого полного набора были использованы в опытах по идентификации с вынужденным выбором, где испытуемые должны были оценивать их в терминах трех гласных русского языка ([i]–[e]–[a]) (Chistovich e. a., 1966; Голузина, 1971). Полученные в этих исследованиях распределения оценок приведены на рис. 4.3,Б.

При изучении контекстуальных влияний каждый из частных наборов (назовем их «1–6» и «3–8» соответственно) предъявлялся группе испытуемых, которые должны были ответить, какой гласный ([i] или [e]) они слышат. На рис. 4.3,А и 4.3,В приведены кривые идентификации (частота ответов «е») для этих наборов. Там же для этих сигналов даны частоты ответов «е», полученные в опытах с полным набором из 12 сигналов при трехальтернативном выборе, т.е. соответствующие участки кривых идентификации с рис. 4.3,Б. Отчетливо прослеживается зависимость результата идентификации каждого стимула от того, в каком наборе он предъявлялся испытуемым: например, стимул «4» оценен как «е» с относительной частотой 0,75 в полном наборе, 0,934 – в наборе «1–6» и 0,129 – в наборе «3–8».

Очевидным результатом таких контекстуальных влияний является сдвиг получаемых в опыте «фонемных границ»: положение границы в пространстве параметров сигнала, помимо свойств сенсорной системы, определяется еще и структурой предъявляемого в опыте набора сигналов. О каких же эталонах тут может идти речь?!

Явление это называют «выработкой уровня адаптации» (Helson, 1948). Его образную иллюстрацию дали в своей статье Бродбент и Ладефогед (Broadbent, Ladefoged, 1960) на примере оценки испытуемыми грузов разной тяжести: «...для профессионального штангиста груз весом в 100 фунтов покажется "средним", а груз в 50 фунтов он оценит как "легкий". Клерк может считать "средним" только груз весом 20 фунтов и потому груз в 50 фунтов оценит как "тяжелый". В условиях эксперимента субъективную шкалу ответов можно сместить, предъявляя испытуемым последовательность грузов: если за грузом в 50 фунтов последует набор более тяжелых грузов, весьма вероятно, что первый будет оцениваться как "легкий". Стимул, на который в таком наборе будет дан нейтральный ответ, называется "уровнем адаптации" ('adaptation level') на данный момент».

Из приведенного описания следует, что до начала эксперимента с поднятием грузов каждый испытуемый имеет свой

собственный эталон «легкого» или «тяжелого» груза, а под влиянием условий эксперимента происходит изменение этого эталона так, что в конце концов он оказывается примерно одинаковым для разных испытуемых и отражает особенности предъявленного в опыте набора стимулов (грузов): как правило, положение эталона соответствует середине диапазона изменения данного параметра стимула в предъявленном наборе. Об этом последнем обстоятельстве Бродбент и Ладефогед не говорят, но оно известно каждому, кто использует методы психофизики.

Что же заставляет испытуемого изменять свой внутренний эталон применительно к условиям конкретного эксперимента? Ведь обычная инструкция «оценивать грузы как легкие или тяжелые» допускает возможность того, что все предъявленные в опыте грузы будут отнесены к одному классу (для одного испытуемого к «легким», для другого – к «тяжелым»). Однако такого результата не наблюдал ни один исследователь. Видимо, испытуемый дополняет данную ему инструкцию своим собственным правилом: «в опыте обязательно должны появляться и легкие, и тяжелые грузы». Следуя этому правилу, он изменяет свой внутренний эталон, чтобы иметь возможность одну часть грузов оценивать как «легкие», а другую – как «тяжелые».

Но, изменяя внутренний эталон в условиях конкретного эксперимента, испытуемый должен иметь возможность вовремя остановиться, чтобы не впасть в другую крайность, когда внутренний эталон снова окажется вне диапазона параметров экспериментального набора, но уже с противоположной стороны. Необходима так называемая обратная связь. Что может использовать испытуемый в качестве источника (сигнала) такой обратной связи?

Если испытуемый действительно дополняет данную ему инструкцию правилом «в наборе должны быть и легкие, и тяжелые грузы», то обнаружить непригодность своего внутреннего привычного эталона для работы с предлагаемым набором внешних сигналов (грузов) он может только путем проверки на соответствие данному правилу тех оценок, которые он вырабатывает при восприятии каждого сигнала. Обнаружив, что система оценок, образующаяся в результате сопоставления внешних сигналов с его привычным эталоном, не соответствует такому правилу, испытуемый изменяет эталон до тех пор, пока новая система оценок не станет подчиняться указанному правилу. Так как в опыте испытуемый фиксирует свои оценки путем отнесения каждого стимула к одной из заданных в инструкции категорий,

для экспериментатора процесс изменения эталона (переобучения) будет проявляться в динамическом изменении характера ответов по ходу опыта. В конечном счете все сводится к тому, что разные категории ответов должны появляться одинаковое число раз (Parducci, 1965; Treisman, Williams, 1984), а при двух категориях каждый тип ответа должен появляться с вероятностью (частотой) 0,5.

В качестве примера на рис. 4.5 приведены такие вероятности (частоты), вычисленные по данным Аллана и Кристофферсона (Allan, Kristofferson, 1974) для нескольких испытуемых и при разных комбинациях диапазона изменения исследуемого в опыте параметра (длительности световой вспышки). В опытах ставилась задача оценить как «долгие» или «краткие» вспышки света разной длительности (от 70 до 1020 мс). Из рис. 4.5 следует, что частоты действительно близки к 0,5 для всех испытуемых и независимо от абсолютных длительностей сигналов. Подтверждением универсальности правила могут служить результаты любого опыта, использующего метод идентификации.

Она подтверждается и результатами описанного выше эксперимента по исследованию контекстуальных влияний при восприятии синтезированных гласных (Кузьмин, Лисенко, 1971). Для набора «1–6» ответы «е» испытуемые давали с частотой 0,493, для набора «3–8» – 0,567. Но если бы стимулы набора «3–8» оценивались бы в сравнении с тем эталоном, который был принят при оценке стимулов набора «1–6», частота ответов «е» должна была бы быть 0,826. Таким образом, контекстуальные влияния проявляются в сдвиге внутреннего эталона, приводящем к выравниванию вероятностей появления в ответах испытуемых заданных инструкцией оценок. В качестве сигнала обратной связи может использоваться апостериорная вероятность (частота) появления этих оценок, а точнее, отклонение

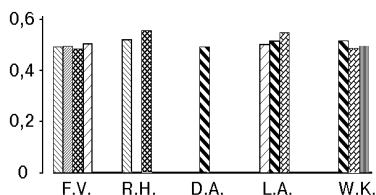


Рис. 4.5. Относительная частота (вероятность) появления ответа «долгий» (по оси ординат) у разных испытуемых в опытах по идентификации длительности вспышек света.

Латинскими буквами обозначены испытуемые; колонки разной штриховки – данные для разных «стандартных» длительностей в экспериментальных наборах.

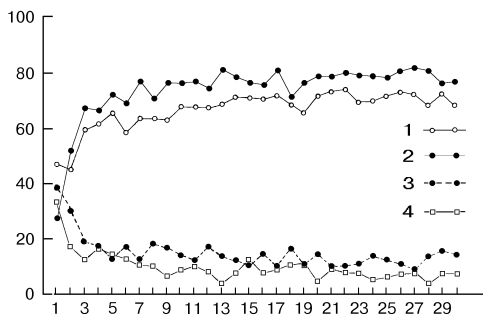


Рис. 4.6. Динамика выработки знаний об априорных вероятностях появления стимулов разных классов в опытах по предсказанию пространственного положения источника света (по: Gardner, 1957).

1 — при появлении источника «слева» в 70% процентах случаев (выбор из двух вариантов — «слева» и «справа»); 2 — при появлении источника «слева» в 70% случаев; 3 — «в середине» в 20% случаев и 4 — «справа» в 10% (выбор из трех вариантов).

По оси абсцисс — номера последовательно предъявляемых групп из 15 стимулов каждая; по оси ординат — процент правильно предсказанных положений.

этой частоты от априорной вероятности (0,5 при двух категориях, 0,333 при трех и т.д.).

Способность человека узнавать в результате обучения апостериорные вероятности появления конкретных сигналов в экспериментальных наборах и соответственно этому регулировать частоту появления тех или иных ответов доказана экспериментально (см., например, Gardner, 1957). На рис. 4.6 хорошо видна динамика этого процесса. Причем очень важно отметить, что частота ответов, с которой начинается процесс переобучения, соответствует равновероятному появлению в

опыте разных классов сигналов.

Такая же способность обнаруживается и при исследовании чувствительности сенсорной системы человека методами теории статистических решений (Аткинсон и др., 1969; Леонов, 1977; Забродин, 1977): в среднем по всему опыту частота появления ответа «сигнал» (обычно задается две возможности — «шум» и «сигнал») хорошо совпадает с заданной исследователем частотой появления сигнала (на фоне шума) в экспериментальном наборе стимулов. Расхождение не превышает 10%.

Таким образом, поведение испытуемого в психофизическом эксперименте (в данном случае с использованием метода одиночных стимулов) оказывается направленным к оптимальному решению поставленной в эксперименте задачи: разбить предъявленный набор сигналов на заданные инструкцией клас-

сы. При этом испытуемый в состоянии решить поставленную задачу, хотя условия эксперимента не всегда совпадают с привычными для него условиями естественного поведения.

Количественно это выражается в том, что при равном числе предъявлений каждого из стимулов экспериментального набора, равномерно распределенных в пространстве исследуемого параметра, и в предположении близкой к линейной зависимости между физическими величинами параметра и их субъективными представлениями «точка субъективного равенства» должна совпадать с серединой диапазона изменения этого параметра в данном эксперименте. Если в опыте используется несколько наборов стимулов с различающимися значениями параметров, зависимость между оценкой «точки субъективного равенства» и величиной параметра, соответствующей середине исследуемого диапазона, оказывается близкой к линейной. Так, при восприятии длительности коротких световых вспышек (Allan, Kristofferson, 1974) «точка субъективного равенства» практически совпадает с серединой диапазона длительностей стимулов в экспериментальном наборе (рис. 4.7) (отклонение не превышает 2,5%). И это при отсутствии в наборах стимулов с соответствующей длительностью.

Отклонения от данного правила наблюдаются при неравномерном расположении

стимулов в исследуемом диапазоне (Parducci, 1965; Krause, 1982) или при неодинаковом числе предъявлений каждого из стимулов (Parducci, 1965). При этом степень отклонения определяется конкретным распределением частот предъявления стимулов в данном наборе (Treisman, Williams, 1984), но при выполнении условия равновероятности появления ответов разных типов.

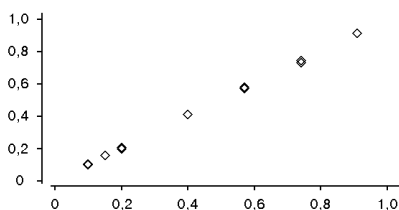


Рис. 4.7. Зависимость положения «точки субъективного равенства» на шкале длительностей от параметров стимулов в экспериментальном наборе.

По оси абсцисс – длительность световой вспышки, соответствующая середине диапазона изменения длительностей стимулов в экспериментальном наборе, с; по оси ординат – длительность стимула, соответствующая «точке субъективного равенства», с.

Должны они наблюдаться и при условии, что мы имеем дело с нелинейными субъективными шкалами для конкретных физических параметров сигналов.

Поскольку экспериментальная процедура определения «фонемной границы» по сути ничем не отличается от процедуры, используемой в методе одиночных стимулов, следует ожидать, что положение «фонемной границы» в пространстве физических параметров сигнала в основном будет определяться структурой и параметрами конкретного экспериментального набора, а не теми внутренними критериями системы восприятия, которыми она пользуется в условиях естественного речевого общения. И этим объясняются контекстуальные влияния, о которых шла речь выше.

Тогда «фонемная граница» для данного параметра сигнала может быть обнаружена в любой точке физического пространства при соответствующем подборе экспериментальных сигналов. Очевидно, что последнее справедливо лишь до тех пор, пока количественные изменения параметров сигнала не приводят к качественным изменениям его оценок испытуемыми, а именно это наиболее вероятно, когда исследуются речевые или речеподобные сигналы.

Так, к примеру, изменяя частоту основного тона гласного и давая инструкцию оценивать сигнал в терминах «высокий/низкий», мы вполне можем обнаружить границу в любом произвольном месте допустимого диапазона частот основного тона в зависимости от диапазона ее изменения в конкретном эксперименте. Но если испытуемому дать задание оценивать те же сигналы в терминах «детский/женский/мужской голос», положение границы должно зависеть не только от параметров конкретного набора, но и от имеющихся у испытуемого сведений об ограниченности диапазона частот основного тона для детского, женского и мужского голосов.

В еще большей степени это должно относиться к «фонемным границам» в пространстве такого параметра, как десинхронизация шума и голоса (см. с. 177). Здесь количественная оценка степени десинхронизации в конце концов превращается в качественную оценку способа образования согласного, а диапазон величин десинхронизации, при котором может возникнуть этот переход, достаточно ограничен и обуславливается особенностями артикуляции в конкретном языке. И действительно, в опытах Брэди и Дарвина (Brady, Darwin, 1978) было установлено, что величина «фонемной границы» для признака

«способ образования смычных согласных» ([d]–[t]), измеренная в пространстве параметра «десинхронизация шума и голоса», не связана линейной зависимостью с характеристиками стимулов в экспериментальном наборе (рис. 4.8).

Суммируя сказанное, следует отметить, что метод одиночных стимулов (идентификации) в его классическом варианте не позволяет исследовать систему эталонов, используемую при восприятии речи. Необходимо существенно модифицировать экспериментальную процедуру.

Потребуется

изменить структуру экспериментального набора сигналов: он должен быть построен так, чтобы исключить или, по крайней мере, минимизировать возможность выработки «уровня адаптации» ('adaptation level') и тем самым уменьшить контекстуальные влияния. Один из возможных способов – создать у испытуемого впечатление неповторяемости стимулов в ходе эксперимента, например использовать большой набор разных слов и задавать разные значения исследуемого параметра в разных словах.

Можно попытаться использовать предполагаемую нелинейность субъективных шкал и исследовать эталонные области как зону изменения положения «фонемных границ» при целенаправленных контекстуальных изменениях. Еще одна возможность заключается в оценке качественных изменений ответов испытуемых при определенных изменениях контекста (структуры экспериментального набора).

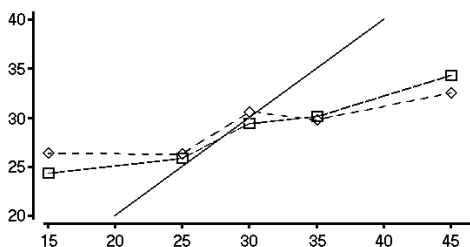


Рис. 4.8. Зависимость положения «фонемной границы» для способа образования смычных согласных от параметров стимулов в экспериментальном наборе.

Квадраты и ромбы – данные опытов с разным порядком предъявления широкого и узких диапазонов изменения величины десинхронизации; прямая линия – предполагаемая на основании опытов по идентификации длительностей зависимость положения «фонемной границы» от параметров набора стимулов.

По оси абсцисс: величина временной десинхронизации шума и голоса, соответствующая середине диапазона изменения этого параметра в экспериментальном наборе, мс; по оси ординат: положение «фонемной границы» на шкале десинхронизации, мс.

## ЗАКОН ВЕБЕРА: ФАКТ ИЛИ АРТЕФАКТ?

В предыдущих разделах речь шла о том, что любая из предлагаемых моделей восприятия речи человеком на определенном этапе обработки сигнала предполагает переход от непрерывного признакового описания к последовательности дискретных «символов». На этом этапе конкретное содержание понятия «символ» может быть разным в зависимости от класса обрабатываемого сигнала (речевой/неречевой) и типа предлагаемой модели. Для речевого сигнала это должна быть последовательность символов, представляющих собой некоторые единицы того языка, носителями которого являются говорящий и слушающий. Такая последовательность формируется в результате сопоставления текущего признакового описания, создаваемого по одинаковым для всех звуковых сигналов правилам, с внутренними субъективными эталонами соответствующих единиц языка.

Хотя о номенклатуре символов (и эталонов) еще нельзя сказать ничего сколько-нибудь определенного, совершенно очевидно, что из-за естественной вариативности сигнала для каждого эталона должна быть задана «зона нечувствительности» (порог), в пределах которой несоответствие реальной величины признака и эталона еще не приводит к выбору другого символа в качестве описания данного участка сигнала.

Так как наличие такой зоны позволяет классифицировать все участки речевого сигнала на, скажем, «Х-символы» и «не Х-символы», ее очевидной аналогией в психоакустике будет дифференциальный (разностный) порог. Естественно поэтому выяснить, как и в какой степени количественные оценки величины этого порога, получаемые в психофизическом эксперименте, могут быть использованы в реально работающих функциональных моделях восприятия речи. Это тем более важно, что, как следует из результатов множества экспериментов, дифференциальные пороги не являются чем-то постоянным. Следовательно, возникает дополнительная задача: учесть в модели эти возможные изменения, если они имеют к ней какое-либо отношение.

Один из основоположников современной экспериментальной психологии, Фехнер, полагал, что субъективные ощущения связаны с физическими параметрами соответствующих явлений внешнего мира логарифмической зависимостью (Fechner, 1860, цит. по: Бардин, 1976). Им же высказано пред-

положение, что едва заметное приращение ощущения, осознаваемое и оцениваемое человеком, есть величина постоянная.

Очевидно, что едва заметное приращение ощущения и соответствует порогу (зоне нечувствительности), так как превышение этого порога приводит к смене ощущений, т.е. вместо одного символа сигналу будет приписан другой. То же самое справедливо и для дифференциального (разностного) порога, ибо два внешних явления, различающиеся на эту величину, вызовут разные оценки наблюдателя, т.е. разные ощущения.

Отсюда вытекает утверждение о постоянстве относительного дифференциального порога (при измерении его в физических единицах), получившее название «закон Фехнера-Вебера». Многочисленные экспериментальные исследования, кажется, подтверждают в целом справедливость этого утверждения, хотя обнаруживаются и отклонения от этого правила, для которых исследователи пытаются найти разумное объяснение (с большим или меньшим успехом).

Само по себе существование такого правила, или закона, не может встретить у создателей функциональной модели восприятия речи каких-либо возражений. При том неперменном, однако, условии, что оно задано в виде универсального алгоритма с известными количественными характеристиками, т.е. может быть реализовано в модели в виде формальной процедуры преобразования сигнала. На первый взгляд, исчерпывающим решением этой задачи является использование (как это предполагал Фехнер) логарифмической шкалы при переходе от физических параметров сигнала к параметрам его внутреннего

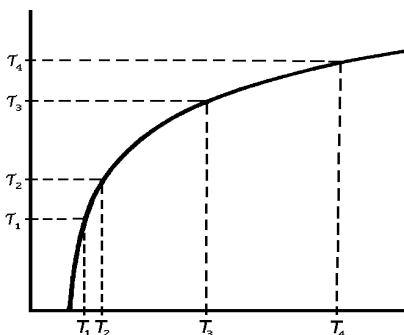


Рис. 4.9. Схема образования постоянного относительного дифференциального порога (по длительности) в предположении существования логарифмической шкалы субъективных длительностей.

Постоянной разности субъективных длительностей соответствует постоянное отношение физических.

По горизонтали — физическая длительность стимула; по вертикали — субъективная длительность.

субъективного представления. Схематическое изображение процесса такого преобразования (рис. 4.9) демонстрирует на примере длительности, как может быть связана постоянная разностная чувствительность в субъективном пространстве с изменяющейся чувствительностью в пространстве физических признаков сигнала.

Если такое преобразование является внутренним свойством сенсорной системы, то логарифмический характер шкалы, описывающей переход от физических параметров сигналов к их субъективным параметрам, должен так или иначе проявляться в результатах любых психофизических исследований. Кроме того, для нормального функционирования сенсорной системы (и ее модели) необходимо, чтобы количественные характеристики этого преобразования были достаточно стабильны для всех параметров сигнала. В крайнем случае, можно согласиться с фактом существования изменений в этих характеристиках, но при условии, что известны правила, по которым они происходят. И вот здесь-то и возникают проблемы.

В многочисленных работах с использованием методов прямого численного шкалирования субъективных ощущений исследователи получают самые разнообразные виды шкал, вплоть до линейных. При этом оказывается, что параметры получаемой шкалы могут зависеть от структуры предъявляемого в опыте набора стимулов (Stevens, Galanter, 1957) или заданного инструкцией диапазона возможных ответов испытуемого (Ross, DiLollo, 1971). Иначе говоря, оказывается, что логарифмическое преобразование не является обязательной процедурой получения внутреннего представления сигнала.

Далее. Величины относительных дифференциальных порогов для разных параметров акустического сигнала (длительности, громкости, высоты, положения спектрального максимума и т.д.) оказываются существенно разными. С этим можно было бы смириться, заложив различия в масштабные коэффициенты тех преобразователей, которые превращают физические параметры внешнего сигнала в их внутреннее субъективное представление в сенсорной (слуховой) системе.<sup>1</sup> Но даже для

---

<sup>1</sup> Заметим, что в сенсорной системе на уровне ощущения физическая природа явления, вызвавшего это ощущение, уже никак не проявляется; материальным носителем здесь являются возбуждение и торможение определенных участков нейронной сети. Следовательно, можно было бы ожидать, что дифференциальные пороги на данном уровне (с учетом логарифмического преобразования по Фехнеру) окажутся

Таблица 4.1. Результаты опытов по субъективному сравнению длительностей синтезированных гласных: относительные частоты ответов «второй в паре долгий»

Длит. 1-го гласн.	Длительность 2-го гласного в паре									
	50	63	79	100	125	159	200	250	316	400
50	0,465	0,395	0,480	0,530	0,735	0,830	0,920	0,960	0,990	0,980
63	0,335	0,305	0,465	0,715	0,735	0,910	0,995	0,990	0,999	0,990
79	0,105	0,300	0,455	0,375	0,630	0,820	0,960	0,970	0,980	0,990
100	0,060	0,255	0,115	0,390	0,505	0,780	0,935	0,980	0,990	0,999
125	0,075	0,085	0,140	0,420	0,380	0,615	0,925	0,980	0,990	0,999
159	0,050	0,095	0,110	0,270	0,155	0,420	0,735	0,880	0,980	0,975
200	0,075	0,125	0,100	0,140	0,160	0,350	0,560	0,855	0,950	0,995
250	0,030	0,005	0,030	0,010	0,060	0,180	0,305	0,605	0,940	0,990
316	0,000	0,000	0,030	0,010	0,020	0,030	0,110	0,380	0,555	0,865
400	0,010	0,010	0,000	0,010	0,020	0,020	0,035	0,120	0,275	0,505

одного и того же параметра сигнала (например, длительности) относительные дифференциальные пороги, полученные разными исследователями, оказываются неодинаковыми (см. Чистович и др., 1976), чего никакими масштабными коэффициентами учесть невозможно, пока не выяснена природа различий.

Анализ экспериментальных процедур, использованных разными исследователями при определении величины относительного дифференциального порога, позволяет предположить, что его величина (по крайней мере, для длительности звуковых сигналов) определяется не только свойствами сенсорной системы, но и структурой набора экспериментальных сигналов, который предъявляется испытуемому. Особенно наглядно это проявилось в результатах опытов по различению длительностей синтезированных гласноподобных звуков, приведенных в табл. 4.1, где каждая строка соответствует длительности первого гласного в паре.

Если допустить, что относительный дифференциальный порог в 10 – 20% является раз и навсегда заданным свойством сенсорной системы, а абсолютная величина порога определяется

---

постоянными вне зависимости от физической природы вызвавшего их явления внешнего мира. Но установить это путем непосредственного измерения исследователю не в состоянии, и ему приходится оценивать величины субъективных порогов через соответствующие значения физического параметра сигнала, что приводит к утрате «масштабных коэффициентов».

длительностями двух сравниваемых сигналов в паре (как это и должно быть при наличии логарифмического преобразователя на входе субъективного измерителя параметров входного сигнала), то для пар сигналов с длительностями 50–63, 50–79 и 50–100 мс следует ожидать различия с относительной частотой выше 0,75. Но, как видно из табл. 4.1, такая различимость достигается при гораздо большей разнице в длительностях сигналов пары. Видимо, в этом случае величина дифференциального порога («зоны нечувствительности») устанавливалась по каким-то иным правилам. Расчеты показывают (Венцов, 1983), что она едина для всего набора стимулов данного опыта и составляет около 80 мс. Из этого можно заключить, что получаемые в опыте величины порогов отражают одновременно как свойства сенсорной системы, так и особенности воспринимаемого набора сигналов.

Было высказано предположение (Ventsov, 1987), подкрепленное некоторыми экспериментальными данными, что ширина «зоны нечувствительности» зависит от диапазона изменения исследуемого параметра сигнала в конкретном опыте. Во время опыта происходит как бы «переобучение» испытуемого: он «подстраивается» к заданному набору стимулов.

На такое предположение натолкнуло то обстоятельство, что данные, полученные в психофизическом эксперименте с любыми стимулами методом парного сравнения, образуют как бы единую генеральную совокупность (рис. 4.10), будучи нанесены на график в зависимости от номера стимула в упорядоченном экспериментальном наборе.

В соответствии с принятыми в психофизике методиками обработки и интерпретации подобных экспериментальных результатов для данных, приведенных на рис. 4.10, абсолютная величина дифференциального порога окажется примерно одина-

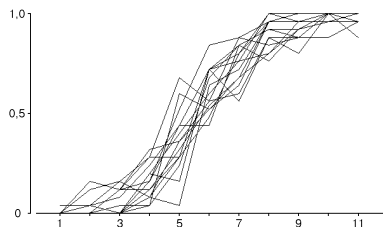


Рис. 4.10. Зависимость относительной частоты ответов «второй больше» при сравнении пары длительностей от длительности переменного сигнала в паре. Данные для 15 экспериментальных наборов.

По оси абсцисс - номер стимула в экспериментальном наборе, организованном в порядке возрастания разницы длительностей переменного и стандартного сигналов; по оси ординат - относительная частота появления соответствующих ответов.

В соответствии с принятыми в психофизике методиками обработки и интерпретации подобных экспериментальных результатов для данных, приведенных на рис. 4.10, абсолютная величина дифференциального порога окажется примерно одина-

ковой для всех использованных в опыте наборов стимулов, если вычислять ее в долях номера стимула, и пропорциональной диапазону изменения исследуемого параметра в данном экспериментальном наборе, если измерять ее в соответствующих физических величинах. Поскольку величину диапазона исследователь выбирает заранее, планируя эксперимент, результат исследования – величина дифференциального порога – как бы задается изначально, еще до проведения самого эксперимента.

Возможно, исходя из этого предположения, удастся объяснить феномен появления эффекта, получившего в конце концов свое формальное описание в виде закона Вебера: такой зависимости дифференциального порога от величины стандартного сигнала, используемого в опыте в качестве базы для сравнения с ним остальных сигналов экспериментального набора, при которой сохраняется неизменным относительный дифференциальный порог.

Представим себе, как исследователь мог бы планировать эксперимент, поставив себе задачу исследовать дифференциальную чувствительность системы к изменениям конкретного параметра сигнала. По-видимому, прежде всего он выберет те параметры стандартного сигнала, для которых он намеревается получить оценку дифференциальной чувствительности. Затем ему надо будет выбрать для каждого экспериментального набора, содержащего только один из стандартов, диапазон изменения этого параметра в переменном сигнале, сравниваемом со стандартом. Совершенно естественно, что, выбрав подходящую величину изменений параметра для некоторого характерного значения стандарта, исследователь постарается сохранить выбранное соотношение величины диапазона изменений и величины стандарта для всех других значений последнего, т.е. примет для всех наборов стимулов постоянный относительный диапазон изменения исследуемого параметра. Исходя из этого, следует ожидать и постоянства относительного дифференциального порога как следствие специфического выбора параметров сигналов.

Именно это мы наблюдаем во всех психофизических экспериментах, и все эти эффекты и привели к формулированию и широкому применению «закона Вебера».

Очевидно, что каждое отступление от правила «сохранять в опыте неизменной относительную величину диапазона изменения исследуемого параметра» должно приводить к отклонению

Таблица 4.2. Результаты опытов по субъективному сравнению длительностей временных интервалов между короткими щелчками

Номер набора	«Стандарт», мс	Диапазон изм., мс	Ср. квадр. откл., мс	
			Исп. D.G.	Исп. T.W.
1	50	30	5,86	7,6
2	100	50	8,58	11,34
3	200	80	11,0	21,66
4	400	100	16,0	26,15
5	600	200	31,35	39,21
6	800	300	41,25	60,75
7	1000	200	58,17	57,9
8	1200	250	74,8	64,1
9	1400	300	87,03	75,95
10	1600	400	100,0	86,17
11	1800	400	88,2	116,8
12	2000	400	108,3	140,6
13	2400	1000	180,65	124,2
14	2800	1500	260,87	309,3
15	3200	2000	423,9	558,5

от «закона Вебера», независимо от причины, вызвавшей это отступление. Не являются ли подобные отклонения причиной того, что дифференциальные пороги в опытах разных исследователей оказываются разными, а при малых значениях стандарта наблюдается отступление от «закона Вебера» – увеличение относительного дифференциального порога?

Чтобы проверить, в какой мере подобная «философия» согласуется с экспериментальными данными, воспользуемся результатами, полученными Гетти (Getty, 1975). Его исследование было предпринято специально с целью установить, насколько выполняется «закон Вебера» для дифференциальных порогов по длительности. Исследовалась способность испытуемых различать длительности интервалов между двумя щелчками. Параметры стимулов и основные результаты опытов приведены в табл. 4.2. В каждом опыте использовалась одна из 15 стандартных длительностей, с каждой из которых сравнивались 11 изменяемых в опыте длительностей. Опыт проводился методом двухальтернативного вынужденного выбора. Результаты каждого опыта аппроксимировались интегральной кривой нормального распределения, по которой определялись положение «точки субъективного равенства» и величина среднеквадратического отклонения ( $s$ ) (в соответствии с принятой в психофизике методикой величина дифференциального порога определяется как  $0,675s$ ).

На рис. 4.11, А приведены величины относительного среднеквадратического отклонения для каждого из 15 опытов, соответствующего определенному значению стандартной длительности. Из рисунка следует, что говорить о постоянстве относительного дифференциального порога, отличающегося от приведенных величин среднеквадратических отклонений только масштабным коэффициентом, можно лишь применительно к средней части диапазона стандартных длительностей: при малых и больших стандартах наблюдается существенное увеличение относительного среднеквадратического отклонения.

Соответственно при этих значениях стандарта наблюдается резкое возрастание ошибки аппроксимации экспериментальных данных «законом Вебера» (см. рис. 4.14, А), что с сожалением констатировал и сам автор исследования, результаты которого мы здесь используем.

Интересно отметить, что график зависимости относительной величины диапазона изменения длительности от номера опыта примерно воспроизводит характер изменения относительного среднеквадратического отклонения (рис. 4.11, Б). Это вполне согласуется с высказанным выше предположением о связи величины дифференциального порога (среднеквадратического отклонения) с диапазоном изменения параметров стимулов в экспериментальном наборе.

На рис. 4.12 величины стандартных отклонений, полученные при аппроксимации результатов работы Гетти (Getty, 1975), отложены в зависимости от величины диапазона изменения

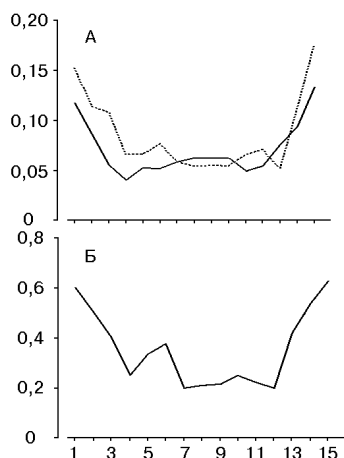


Рис. 4.11. Сопоставление Веберовского отношения (для двух испытуемых) с относительным диапазоном изменения длительности переменного сигнала.

По оси абсцисс – номер экспериментального набора (в упорядоченной по величине стандартного стимула последовательности); по оси ординат – Веберовское отношение (А) и относительный диапазон (Б).

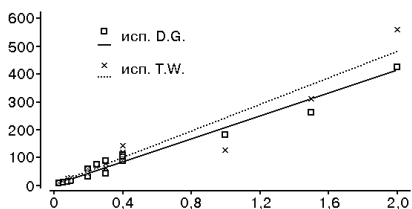


Рис. 4.12. Зависимость между среднеквадратическим отклонением в уравнении интегральной кривой нормального распределения, аппроксимирующей экспериментальные данные, и диапазоном варьирования длительности переменного сигнала в соответствующих экспериментальных наборах. Данные двух испытуемых. Прямыми линиями нанесены результаты линейной аппроксимации экспериментальных зависимостей.

По оси абсцисс — диапазон изменения длительности, с; по оси ординат — среднеквадратическое отклонение, мс.

дурта (см. рис. 4.14,Б), и этим она в лучшую сторону отличается от «закона Вебера». «Модель I» предполагает вполне определенную стратегию поведения испытуемого в психофизическом эксперименте. Прежде всего стимулом для него является не каждый из сравниваемых сигналов в отдельности, а вся пара, оцениваемым же признаком пары — разница параметров образующих ее сигналов (в нашем случае — длительностей). В зависимости от диапазона изменения длительности переменного сигнала в данном экспериментальном наборе испытуемый так настраивает дифференциальную чувствительность слуховой системы, чтобы во время опыта разные внутренние оценки пары («меньше», «равны», «больше») возникали одинаково часто.<sup>2</sup> Как видим, в такой модели игнорируется процесс субъективного

длительности тестового сигнала в соответствующем опыте. Очевиден большой разброс результатов, но тем не менее вполне отчетливо проявляется линейный характер зависимости. Две прямые, прохо-

дящие через начало координат (при регрессионном анализе величина смещения прямых статистически не отличается от нуля), аппроксимируют (назовем эту аппроксимацию «модель I») полученные зависимости стандартного отклонения от величины диапазона изменения длительности стимула в опыте.

При такой модели ошибки примерно одинаковы при всех длительностях стан-

<sup>2</sup> Метод вынужденного выбора не дает возможности обнаружить это явление, так как ответы «равны» испытуемому не разрешены и ему приходится «гадать». Как следствие этого ответы «больше» и «меньше» оказываются примерно равновероятными (заметим, при любой возможной вероятности предполагаемых ответов «равны»).

измерения длительности сигнала и получения ее внутреннего представления. Предполагается, что сразу и непосредственно может быть получена субъективная оценка разности длительностей двух сигналов, хотя они предъявляются последовательно во времени и для получения искомой разности придется запоминать длительность первого до появления оценки длительности второго. Далее из модели следует, что испытуемый способен подстраиваться под любой сколь угодно малый диапазон изменения параметра, т.е. достигать практически бесконечной чувствительности, причем независимо от абсолютной величины длительностей сравниваемых сигналов. Однако уже имеющийся эмпирический опыт многих исследователей противоречит последнему выводу: очевидно, что разница длительностей, скажем, в 5 мс не будет восприниматься одинаково при абсолютных длительностях стандарта в 1 с и 50 мс.

Возникает необходимость каким-то образом привести к единой метрике результаты, полученные в опытах с разными длительностями стандартного сигнала. Поэтому мы попытались проанализировать характер зависимости относительного среднеквадратического отклонения от величины относительного диапазона, т.е. после нормирования соответствующих абсолютных величин по длительности стандартного сигнала.

На рис. 4.13 приведена такая зависимость для одного из испытуемых по материалам, взятым из работы Гетти (Getty, 1975). В первом приближении она может быть описана выражением (назовем это «модель II»):

$$\begin{aligned}\sigma' &= \alpha && \text{при } \delta T \leq \delta T_k \\ \sigma' &= \alpha - \beta(\delta T - \delta T_k) && \text{при } \delta T \geq \delta T_k,\end{aligned}$$

где  $\sigma' = \sigma/T_c$  – относительное среднеквадратическое отклонение;  
 $\delta T = \Delta T/T_c$  – относительный диапазон изменения длительности

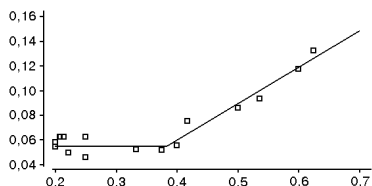


Рис. 4.13. Пример кусочно-линейной аппроксимации экспериментальной зависимости Веберовского отношения от величины относительного диапазона изменения параметров стимулов в экспериментальном наборе.

По оси абсцисс – относительный диапазон; по оси ординат – Веберовское отношение (величина относительного дифференциального порога).

стимулов в опыте;  $\delta T_k$  – «критическая» величина относительного диапазона.

На основании экспериментальных данных для двух испытуемых были вычислены следующие величины коэффициентов, входящих в указанное выражение:

	$\delta T_k$	$\alpha$	$\beta$
исп. D.G.	0,384	0,05472	0,297
исп. T.W.	0,350	0,06230	0,345

Соответствующие ошибки аппроксимации экспериментальных данных приведены на рис. 4.14, В. Отчетливо видно, что «модель II» описывает экспериментальные результаты гораздо лучше, чем «закон Вебера»: удовлетворительное совпадение экспериментальных и расчетных величин относительного средне-квадратического отклонения наблюдается для всех значений стандарта, т.е. для всего опыта в целом.

Поскольку минимизация ошибки является пока единственным критерием качества аппроксимации (и адекватности той или иной модели), интересно сравнить по этому признаку три описанных выше метода аппроксимации («модели»). Рассчитанные на основании рассмотренных моделей ошибки аппроксимации экспериментальных данных, полученных Гетти (Getty, 1975) применительно к относительной дифференциальной чувствительности, для двух испытуемых приведены ниже:

закон Вебера	0,0256	0,0405
«модель I»	0,0162	0,0178
«модель II»	0,0067	0,0127.

Как видим, «модель II» дает наилучшую аппроксимацию экспериментальных результатов. Какие же субъективные механизмы могут лежать в основе этой «модели»?

Напомним, прежде всего, гипотезу Крилмана (Creelman, 1962) о механизме субъективного измерения длительностей. Он предположил, что при измерении длительности используется внутренний тактовый генератор (возможно, пуассоновского типа) и длительность сигнала преобразуется в число импульсов, выдаваемых этим генератором за время действия сигнала. Предположим теперь, что испытуемый обладает способностью так перестраивать частоту генератора, чтобы средней длительности сигнала в предъявленном ему экспериментальном наборе («точка субъективного равенства») всегда соответствовало постоянное число импульсов, независимо от абсолютной величины средней длительности. Основанием для такого допущения служит способность испытуемых вырабатывать 'adaptation level'

применительно к условиям конкретного психофизического эксперимента (см. предыдущий раздел). Если теперь «зону нечувствительности» системы сравнения (дифференциальный порог) задать как некоторое фиксированное число импульсов, то в пространстве физических длительностей мы получим эффект постоянства относительного дифференциального порога.

Если же в экспериментальном наборе сигналов диапазон изменения длительности одного из сравниваемых сигналов становится настолько большим, что доля оценок «равны» оказывается значительно меньше доли оценок «меньше» и «больше», испытуемый расширяет «зону нечувствительности» (повышает порог) до тех пор, пока не получит примерно одинаковые частоты появления всех трех типов оценок. На способность человека изменять дифференциальную чувствительность указывают, в частности, результаты, полученные Фернбергером (Fernberger, 1931). В пространстве же физических длительностей мы обнаружим увеличение относительного дифференциального порога, когда относительный диапазон изменения длительности превысит некоторую критическую величину.

Таким образом, анализ экспериментальных данных, полученных Гетти (Getty, 1975) в опытах по различению длительностей, позволяет сделать следующие выводы:

1) относительный дифференциальный порог действительно может зависеть от диапазона изменения величины исследуемого

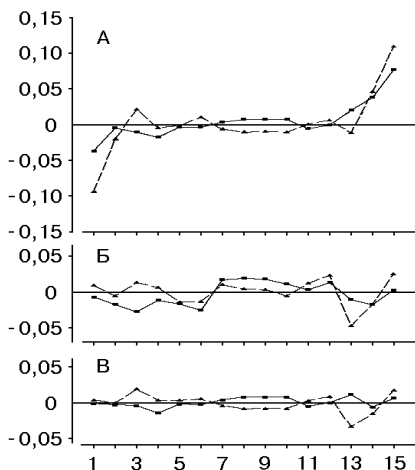


Рис. 4.14. Разница между относительными разностными порогами, полученными в эксперименте и рассчитанными в соответствии с тремя гипотетическими моделями. Данные двух испытуемых.

А – по закону Вебера; Б – согласно «модели I»; В – согласно «модели II».

По оси абсцисс – номер экспериментального набора; по оси ординат – разница между экспериментальными и расчетными величинами относительного разностного порога (относительная ошибка аппроксимации).

параметра сигнала, выбранного исследователем при организации конкретного эксперимента;

2) в то же время этот порог не может быть меньше некоторой «критической» величины, которая для данного испытуемого есть величина постоянная, т.е. следует «закону Вебера»; поскольку мы анализировали данные, касающиеся только восприятия длительности акустических сигналов, не исключено, что для других параметров численное значение и природа этой величины могут оказаться иными в зависимости от свойств соответствующего субъективного измерителя;

3) в количественном отношении зависимость оказывается разной у разных испытуемых, что может быть объяснено различиями в их способности изменять чувствительность своей слуховой системы.

Возможно, в этом и заключается причина того, что разные исследователи получали в своих опытах значительно различающиеся оценки величины относительного дифференциального порога: они работали с разными испытуемыми и выбирали неодинаковые относительные диапазоны изменения длительности в экспериментальных сигналах.

Напрашивается вывод, что не существует универсального правила, определяющего величину дифференциального порога и реализованного путем применения шкалы какого-либо вида в блоке преобразования физических параметров внешнего сигнала в их субъективный эквивалент. Более вероятно, что получаемая в результате обработки экспериментальных данных величина дифференциального порога является следствием применения испытуемым каких-то универсальных правил принятия решений в заданной экспериментальной ситуации, которая определяется и структурой набора предъявляемых сигналов, и заданной исследователем инструкцией, и внутренними установками самого испытуемого.

Очевидно, что создатель работоспособной функциональной модели восприятия речи может использовать данные о дифференциальной чувствительности слуховой системы, полученные в психоакустических экспериментах, с известной осторожностью и только будучи твердо уверенным в том, что условия проведения этих экспериментов совпадали с условиями восприятия естественного речевого сигнала.

## ИЗБИРАТЕЛЬНАЯ АДАПТАЦИЯ: ДЕТЕКТОРЫ ПРИЗНАКОВ ИЛИ ПРАВИЛА ПРИНЯТИЯ РЕШЕНИЙ

Выше говорилось о том, что при восприятии речи переход от непрерывного признакового описания речевого потока к дискретному описанию в терминах некоторых единиц языка осуществляется в результате сопоставления текущих значений параметров сигнала с соответствующими каждой дискретной единице эталонными областями. Для этого у каждого носителя языка в процессе обучения формируются система правил и необходимые «устройства» их реализации. Естественно, что они являются предметом неубывающего интереса исследователей и разработчиков функциональной модели восприятия речи.

Высказывается предположение, что «характерные фонологические признаки человеческой речи декодируются нейросенсорными рецептивными полями – "детекторами признаков", врожденно структурированными для выделения многочисленных различительных признаков акустического потока и соответствующего реагирования на эти признаки» (Abbs, Sussman, 1971, р. 23). По замыслу исследователей, такие детекторы должны срабатывать (выдавать выходной сигнал) всякий раз, когда во входном сигнале возникает комбинация признаков (качественная и количественная), на которую настроен данный детектор. В сущности, детекторы представляют собой устройства, реализующие идею сопоставления непрерывного признакового описания речевого сигнала с эталонами определенных дискретных событий, могущих возникнуть в сигнале. В качестве аргумента приводятся данные о существовании в нервной системе животных специализированных нейронов, избирательно реагирующих на сложные акустические сигналы, биологически важные для данного вида, а также о специализированных рецептивных полях зрительной системы.

Идея детекторов выглядит очень привлекательно и вполне правдоподобно. Однако следует заметить, что до настоящего времени не существует сколько-нибудь устоявшегося представления о том, что же собой представляют признаки, которые должны выделяться детекторами. Подобная неопределенность приводит к тому, что иногда высказываются сомнения в возможности и необходимости самого существования детекторов в системе восприятия речи (Diehl, 1981). И это несмотря на то, что такой чисто речевой признак, как глухость/звонкость со-

гласного, во многих языках определяется фактом и знаком (опережением или запаздыванием) временной десинхронизации моментов появления шума взрыва и колебаний голосовых связок и не может быть получен иначе как путем использования специального детектора – неважно, будет ли это набор правил или специализированное устройство. При этом практически не обсуждается проблема исследования этих детекторов – их структуры и конкретных признаков, ими выделяемых, – у человека: по понятным причинам мы лишены возможности исследовать реакции отдельных нейронов человеческого мозга методами нейрофизиологии, поэтому очень важно оценить возможности исследования детекторов методами психофизики.

В связи с этим в середине 70-х и в начале 80-х годов много внимания было уделено опытам по так называемой «избирательной адаптации». В них испытуемые сначала идентифицировали набор стимулов, подобранных для изучения некоторого признака, а затем опыт повторяли, но перед каждым стимулом или группой стимулов испытуемому многократно предъявляли элемент набора, на который испытуемые дают максимальное число ответов заданного класса. При интерпретации результатов таких опытов исходят из допущения, что длительное предъявление одного сигнала должно вызывать «утомление» (адаптацию) детектора исследуемого признака. Если сигнал, использованный в качестве адаптирующего, действительно вызывает изменения в результатах опыта, считается, что существует детектор, настроенный на признаки этого сигнала.

К сожалению, алгоритмы работы предполагаемых детекторов при этом не описываются. Отсутствует также анализ того, какими количественными характеристиками должен обладать детектор, чтобы результаты работы модели, созданной на его базе, были сопоставимы с данными психоакустических исследований. В связи с этим возникает вопрос, действительно ли все эффекты, наблюдаемые в опытах по избирательной адаптации, являются следствием «утомления» соответствующих детекторов? Остается также непонятным, почему адаптация («утомление») происходит только во время эксперимента и не наблюдается в условиях естественного речевого общения, когда многие детекторы будут работать непрерывно, а значит, должны «утомляться».

Едва ли не единственная работа, в которой дано достаточно подробное описание одного типа детекторов – детекторов с

избирательными характеристиками, – принадлежит Эймасу и Корбиту (Eimas, Corbit, 1973, p. 101):

«(а) Существуют детекторы, избирательно реагирующие на десинхронизацию шума и голоса (об этом термине см. ниже. – А.В., В.К.), причем наибольшая чувствительность соответствует величинам, характеризующим различие по глухости/звонкости при речеобразовании (Lisker, Abramson, 1964). (b) При некоторых задержках возбуждаются два детектора, однако, при прочих равных условиях, следующих уровней обработки достигает только наиболее мощный из двух выходных сигналов. (c) Фонемная граница приходится на величину задержки, при которой оба детектора возбуждаются одинаково. (d) После адаптации чувствительность детекторов уменьшается, т.е. ослабляется их выходной сигнал. Допустим, что выходной сигнал уменьшается в равной степени для всех величин задержки, к которым чувствителен данный детектор. Отсюда следует, что в результате избирательной адаптации фонемная граница сдвигается вдоль шкалы величин задержки как следствие смещения точки равновесия ответов двух детекторов. (e) Предположение, что верхние уровни обработки не различают выходные сигналы детектора, вызванные двумя разными величинами задержки, позволит объяснить появление максимумов различимости стимулов».

Описанный детектор работает с признаком, определяющим, по мнению многих исследователей, классификацию начальных смычных согласных по способу образования – глухой/звонкий – в большинстве языков мира. Этим признаком является величина временной десинхронизации моментов образования шума взрыва и включения голосовых связок (голоса). В англоязычной литературе принято называть этот признак *voice onset time* (букв. «время начала голоса»). Мы будем называть его «десинхронизацией шума и голоса» (ДШГ). При анализе речевых сигналов и синтезе стимулов в психоакустических исследованиях отрицательные величины этого параметра принято приписывать опережающему включению голоса, а положительные – запаздывающему. Здесь надо отметить, что пока не существует никаких формальных правил обнаружения этих событий в естественном речевом сигнале при его обработке в слуховой системе.

Попытаемся создать на основе этого общего описания основных принципов функционирования детекторов признаков «работающее» устройство и сравнить предполагаемые результаты

его работы с данными, полученными в психоакустическом эксперименте.

При этом без нарушения общности рассуждений можно ограничиться рассмотрением процедуры выделения признака только для одного класса согласных, например губных. Кстати, при создании общей модели придется искать ответ на вопрос, должны ли существовать разные детекторы «способа образования» (по крайней мере, различающиеся количественными характеристиками) для согласных разного места образования или это должен быть единый для всех смычных согласных детектор. Кроме того, понадобится еще и детектор «способа образования» для щелевых согласных: ведь среди них тоже есть глухие и звонкие.

Итак, в соответствии с пунктом (а), детектор обладает избирательной чувствительностью к некоторому диапазону величин ДШГ, при этом наибольшая чувствительность соответствует тем участкам диапазона, которые характеризуют различие звуков речи по признаку «глухости/звонкости» в речеобразовании.

Ниже приведены данные о величинах ДШГ для трех категорий губных смычных согласных (звонкого – [b], глухого – [p] и глухого аспирированного (придыхательного) – [p<sup>h</sup>]), полученные усреднением соответствующих величин по всем языкам, сведения о которых имеются в работе Лискера и Абрамсона (Lisker, Abramson, 1964):

		[b]	[p]	[p <sup>h</sup> ]
в изолированных словах	-80	12	76	
во фразах	-58	9	56	
относит. сокращение		0,725	0,750	0,738

При этом отдельно усреднены данные для изолированно произнесенных слов и для фраз. Вычислено также относительное сокращение длительности задержки во фразах по сравнению с изолированным произнесением.

Видно, что интересующие нас интервалы укорачиваются во фразах по сравнению с изолированным произнесением, причем относительное укорочение примерно одинаково для согласных с разным способом образования, что, вероятнее всего, связано с известным общим увеличением темпа артикуляции во фразах. Из этого следует, что нормальная работа детекторов, настроенных на характерные для разных способов образования величины ДШГ, возможна только при условии автоматического нормирования измеренных величин задержек по темпу речи. Со стороны внешнего наблюдателя такая норми-

ровка будет выглядеть как изменение диапазона величин ДШГ, на которые реагирует детектор, т.е. как изменение вида избирательных характеристик. Все это отражено на рис. 4.15, где схематически приведены предполагаемые характеристики трех детекторов.

Из всех параметров амплитудной характеристики детектора можно считать известным только положение ее вершины (максимума избирательности) на шкале ДШГ. Нам неизвестны ни форма этих характеристик, ни величина выходного сигнала детектора, соответствующая его максимальной чувствительности, ни диапазон величин ДШГ, перекрываемый этой характеристикой. Поэтому для начала мы исходили из того, что все детекторы имеют одинаковые характеристики избирательности и что можно ограничиться выбором треугольной формы этих характеристик (рис. 4.15), как это сделал Айнсворт (Ainsworth, 1977).

Согласно пункту (b), детекторы сами по себе могут иметь «перекрывающиеся» амплитудные характеристики, т.е. на некоторые величины ДШГ могут реагировать одновременно два детектора (рис. 4.15), однако на следующий уровень обработки попадают только наибольшие выходные сигналы детекторов. Поскольку термин «наибольшие» точно не определен, правильнее будет говорить, что на следующий уровень обработки попадают ответы детекторов, превосходящие некоторую критическую величину, т.е. ввести понятие порога. Так как на характер реакции системы в целом влияние оказывает только то, что передается на следующий уровень обработки, нам интересны только те величины ответов детекторов (и соответствующие им величины ДШГ), которые превышают этот условный порог. В результате, если работу детектора оценивать по данному кри-

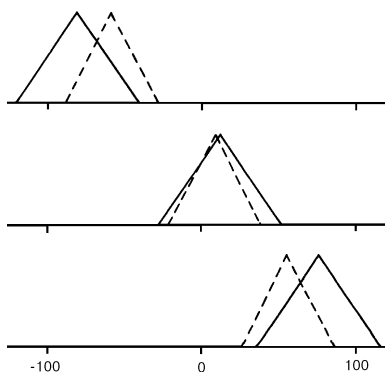


Рис. 4.15. Гипотетические избирательные характеристики детекторов «звонкости», «глухости» и «придыхательности» (сверху вниз). Сплошные линии – при изолированном произнесении, штриховые – при произнесении во фразе. По горизонтали – величина ДШГ (отрицательные значения соответствуют опережающему включению голоса), мс.

терию, его характеристика будет иметь вид, представленный штриховыми линиями на рис. 4.16. Введение порога очевидным образом сокращает область величин ДШГ, на которые в конечном счете реагирует система. Степень такого сокращения определяется величиной «порога»: чем он выше, тем степень сокращения больше.

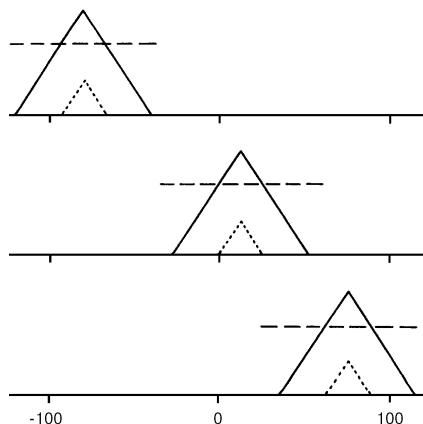


Рис. 4.16. Гипотетическая трансформация характеристик детекторов, приведенных на рис. 4.15 с учетом передачи на следующий уровень обработки только «наибольших» выходных сигналов детектора.

Штриховыми линиями отмечена величина предполагаемого «порога», пунктирными – амплитудные характеристики детектора после порогового устройства.

Остальные обозначения те же, что на рис. 4.15.

тудные характеристики детекторов, чтобы одновременно выполнялись два условия: положение проекций их вершин на шкалу ДШГ должно соответствовать величинам ДШГ для согласных разного способа образования, полученным в результате измерений в акустическом сигнале (см. с. 178), а проекции точек пересечения характеристик разных детекторов – «фонемным границам» для соответствующих звуков, полученным в психоаку-

Из пункта (с) следует, что при прочих равных условиях «фонемная граница» приходится на значение ДШГ, при котором одинаково возбуждаются два детектора одновременно. Это допущение противоречит тому, что было сказано в пункте (b): если даже на более высоких уровнях обработки сохраняется возможность одновременного возбуждения двух детекторов, предполагаемый порог должен быть установлен достаточно низко, так, чтобы амплитудные характеристики детекторов обязательно имели перекрывающиеся области. Но тогда на следующий уровень обработки станут попадать не только «наибольшие» ответы детекторов. Так как понятие «наибольший» не было строго определено, допустим все же, что эти ответы образуют перекрывающиеся зоны на шкале ДШГ.

Посмотрим, какую форму должны иметь ампли-

стическом эксперименте. Примем, что максимальные выходные сигналы одинаковы для всех детекторов, а характеристики по-прежнему имеют треугольную форму. Воспользуемся данными о положении «фонемных границ» для губных согласных разного способа образования, полученными Лискером и Абрамсоном (Lisker, Abramson, 1970) на носителях тайского языка в результате анализа соответствующих кривых идентификации (рис. 4.17). Здесь границе между [b] и [p] соответствует величина ДШГ, равная – 20 мс, а границе между [p] и [p<sup>h</sup>] – величина, равная 40 мс.

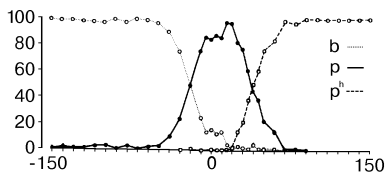


Рис. 4.17. Результаты идентификации стимулов с разными величинами ДШГ носителями тайского языка (по: Lisker, Abramson, 1970). По оси абсцисс – величина ДШГ, мс; по оси ординат – относительная частота ответов «звонкий» (b), «глухой» (p) и «придыхательный» (p<sup>h</sup>), проценты.

На рис. 4.18 приведены характеристики всех трех детекторов, удовлетворяющие указанным условиям. За основу при построении принята характеристика детектора «глухости» (Дг), поскольку она обязательно должна охватывать область значений ДШГ, включающую в себя обе «фонемные границы».

Хотя форма характеристики данного детектора была выбрана совершенно произвольно (это касается величины максимума и ширины перекрываемой ею области на шкале ДШГ), приведенные выше условия построения остальных характеристик достаточно жестко регламентируют их окончательные параметры.

Как видно из рис. 4.18, детекторы имеют принципиально

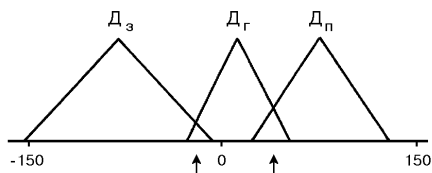


Рис. 4.18. Реконструкция избирательных характеристик гипотетических детекторов способа образования согласных на основании данных психоакустических экспериментов и измерений в речевом сигнале.

Дз – амплитудная характеристика детектора «звонкости», Дг – детектора «глухости» и Дп – «придыхательности».

По горизонтали – величина ДШГ, мс.

разные характеристики, хотя мы по-прежнему не знаем их точных количественных параметров и у нас нет никакой возможности найти их (и это, видимо, – главное), так как все

доступные прямому измерению величины уже учтены при построении характеристик, приведенных на рис. 4.18.

Ситуация окажется еще более сложной, если при реконструкции характеристик детекторов мы учтем еще одно допущение авторов приведенного выше описания: величина выходного сигнала детекторов после порогового устройства не зависит от величины ДШГ, а определяется только фактом попадания

последней в область чувствительности данного детектора, ср. пункт (е) выше. Без этого казалось невозможным объяснить результаты опытов по субъективному различению стимулов, отличающихся физической величиной ДШГ (Eimas, Corbit, 1973).

Какими могут быть окончательные характеристики детекторов, описанных Эймасом и Корбитом, с учетом этого последнего допущения? Рисунки 4.19-4.21 иллюстрируют три возможных варианта.

На рис. 4.19 показан уже рассмотренный выше вариант, когда на предыдущем уровне детекторы имеют перекрывающиеся избирательные характеристики (см. рисунок 4.18). Здесь характеристики тоже перекрывающиеся, но теперь «фонемная граница» представлена не точкой на шкале ДШГ, а областью на этой же

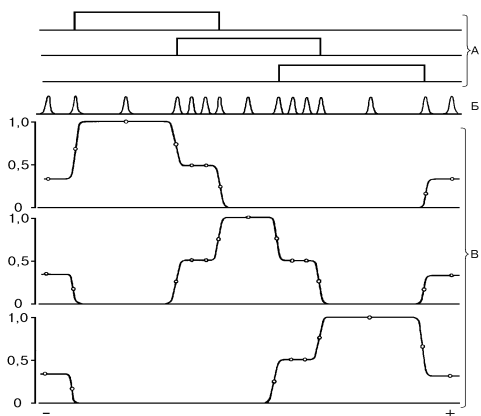


Рис. 4.19. Схематическое представление взаимного расположения на шкале ДШГ перекрывающихся характеристик детекторов способа образования смычных согласных при условии, что на высшие уровни обработки передаются только «наибольшие» ответы с предыдущего уровня.

А – амплитудные характеристики детекторов «звонкости», «глухости» и «придыхательности» (сверху вниз); Б – схематическое изображение распределений субъективных оценок физических величин ДШГ; В – «кривые идентификации» (зависимость вероятности появления ответов «звонкий», «глухой» и «придыхательный»), получающиеся при работе соответствующих детекторов (сверху вниз).

По горизонтали – величина ДШГ (знаками «-» и «+» отмечены крайние области диапазона возможных изменений).

шкале. Рис. 4.20 отражает предельный случай варианта с перекрывающимися характеристиками: характеристики примыкают друг к другу. «Фонемная граница» может быть определена как точка на шкале ДШГ, в которой наблюдается примыкание соседних характеристик. Наконец, на рис. 4.21 приведены характеристики детекторов в случае, когда на верхний уровень передаются только наибольшие ответы с предшествующего уровня, а характеристики детекторов разнесены по шкале ДШГ, и для определения места «фонемной границы» необходимы какие-то дополнительные правила.

Теперь попытаемся оценить, какие результаты должны получиться в психоакустическом эксперименте при использовании в системе восприятия речи и в ее функциональной модели детекторов с указанными характеристиками (рис. 4.19-4.21). Для этого договоримся о следующем: 1) когда отвечают два детектора одновременно, система принимает решение гаданием, выбирая один из двух возможных ответов в соответствии с типами возбужденных детекторов; 2) когда не отвечает ни один из детекторов, система принимает решение гаданием, выбирая один из нескольких возможных ответов, число которых либо определяется системой языка, носителем которого является испытуемый, либо задается экспериментатором в инструкции; применительно к системе на рис. 4.19-4.21 это число равно 3; 3) при многократном предъявлении в психоакустическом опыте одного и того же стимула (с заданной физической величиной ДШГ) на вход детектора будут попадать разные субъективные

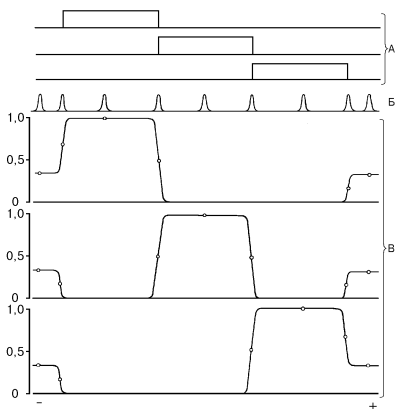


Рис. 4.20. Схематическое представление взаимного расположения на шкале ДШГ примыкающих характеристик детекторов способа образования смычных согласных при условии, что на высшие уровни обработки передаются только «наибольшие» ответы с предыдущего уровня. Обозначения те же, что на рис. 4.19.

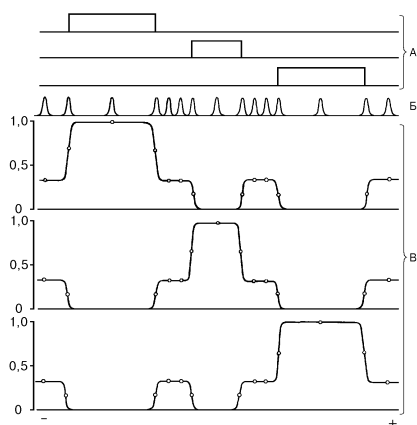


Рис. 4.21. Схематическое представление взаимного расположения на шкале ДШГ разнесенных характеристик детекторов способа образования смычных согласных при условии, что на высшие уровни обработки передаются только «наибольшие» ответы с предыдущего уровня. Обозначения те же, что на рис. 4.19.

величины ДШГ, разброс которых из-за ошибок измерения будем считать отвечающим закону нормального распределения.

С учетом этих замечаний были рассчитаны вероятности появления ответов «звонкий», «глухой» и «придыхательный» для нескольких стимулов с разными физическими величинами ДШГ и построены соответствующие графики зависимости этих вероятностей от величины ДШГ для всех трех комбинаций характеристик детекторов. Результирующие «кривые идентификации» приведены на рис. 4.19-4.21.

При сопоставлении полученных графиков с кривыми идентификации, построенными Лискером и Абрамсоном (Lisker, Abramson, 1970) по результатам психоакустического эксперимента (см.

рис. 4.17), обнаруживается по крайней мере одно принципиальное отличие. Оно касается тех областей в левой и правой частях шкалы ДШГ, где оказываются равновероятными все три типа ответов: «звонкий», «глухой» и «придыхательный». Ничего подобного в реальных психоакустических экспериментах не наблюдается. Из этого мы можем сделать вывод, что использование в функциональной модели восприятия речи детекторов с избирательными характеристиками не позволит удовлетворительно воспроизвести поведение человека в условиях психоакустического эксперимента.

Отметим здесь еще одно обстоятельство. Чтобы совместить предполагаемые результаты деятельности описываемых детекторов с данными, получаемыми в реальном эксперименте, авторам пришлось ввести в алгоритм детекторов несколько

взаимоисключающих правил. При этом, с одной стороны, нарушилась логическая стройность всего алгоритма, а, с другой стороны, сохранилась неясность относительно способов проверки справедливости предлагаемых правил. Поэтому не станем здесь анализировать варианты возможного поведения описанных выше детекторов в условиях избирательной адаптации, а отметим лишь, что для детекторов с примыкающими характеристиками (рис. 4.20) общий вид зависимости вероятностей ответов всех трех типов от величины ДШГ совпадал бы с экспериментально полученными кривыми идентификации, если бы удалось исключить равновероятное появление трех типов ответов на краях исследуемого диапазона величин ДШГ. Очевидное решение проблемы состоит в таком изменении формы характеристик детекторов «звонкости» и «придыхательности», чтобы первый реагировал на любые величины ДШГ, меньшие некоторой отрицательной величины, а второй – на любые величины, превосходящие некоторую положительную величину. Эти новые детекторы перестают быть избирательными в том смысле, как их описали Эймас и Корбит (Eimas, Corbit, 1973), а становятся детекторами релейного типа, к рассмотрению которых мы и обратимся.

Если решающим признаком при определении способа образования смычного согласного действительно является временная десинхронизация шума взрыва и голоса, оцениваемая величиной ДШГ, то слуховой системе достаточно было бы некоторого определенного опережения включения голоса, чтобы оценить соответствующий согласный как звонкий. Сколько бы мы затем ни увеличивали степень опережения в эксперименте, для системы восприятия это не имело бы значения, так как некоторое необходимое и достаточное опережение уже достигнуто. При речеобразовании же величина опережения определяется аэродинамическими процессами в речевом тракте и инерционностью артикулирующих органов и, как показывают измерения в акустическом сигнале (Lisker, Abramson, 1964), может быть достаточно большой.

Предположим, что величина ДШГ вырабатывается (до поступления на соответствующий детектор) линейной измерительной системой с неограниченным динамическим диапазоном. Когда шум взрыва и голос возникают одновременно (величина ДШГ равна нулю), система идентифицирует согласный как глухой. Однако даже самые совершенные технические системы всегда имеют «зону нечувствительности», в пределах которой

изменения внешнего сигнала никак не регистрируются. Тем более это должно быть справедливо для биологических «измерительных» систем: их основу составляют нейроны, обладающие «порогом», т.е. некоторым ненулевым значением входного сигнала, еще не вызывающим реакции нейрона. Поэтому вполне резонно допустить существование области величин ДШГ по обе стороны «нуля», в пределах которой система еще никак не

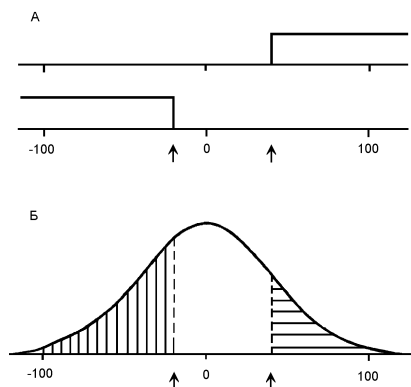


Рис. 4.22. Характеристики детекторов релейного типа (А) и распределение субъективных величин ДШГ для стимула с нулевой физической ДШГ (Б).

Стрелками отмечены границы области «нулевой десинхронизации» («фонемные границы»). Заштрихованные площади – вероятность возбуждения каждого из детекторов.

По горизонтали – величина ДШГ, мс.

согласные, все величины ДШГ, превышающие левую (отрицательную) границу области «нулевой десинхронизации», будут вызывать оценки «глухой» и так далее. Собственно говоря, границы этой области «нулевой десинхронизации» (отрицательная левая и положительная правая) и должны быть определены как «фонемные границы» между звонкими и глухими, глухими и придыхательными соответственно.

реагирует на возможную десинхронизацию взрыва и голоса («считает» ее нулевой), а соответствующий согласный оценивает как глухой. Величины ДШГ, выходящие за правую (положительную) границу этой области, будут вызывать оценки согласного как придыхательного, величины, выходящие за левую (отрицательную) границу, – оценки согласного как звонкого. Таким образом, речь может идти о существовании в системе двух детекторов: один реагирует на положительные величины ДШГ, превышающие некоторую пороговую величину, другой – на отрицательные величины ДШГ, меньшие соответствующей пороговой величины (рис. 4.22, А).

Так будет работать система в языках, в которых различают три способа образования согласных. Если же в языке отсутствуют придыхательные

Если бы при анализе и оценке стимулов испытуемые не допускали никаких ошибок и стимул с заданной величиной ДШГ оценивался бы ими совершенно одинаково при всех его предъявлениях, положение «фонемной границы» на шкале ДШГ можно было бы определить как величину ДШГ, при которой резко изменяется ответ испытуемого: например, вместо ответа «звонкий» испытуемый начинает давать ответ «глухой». Подбором шага изменения ДШГ в экспериментальных сигналах можно было бы установить положение «фонемной границы» с любой требуемой точностью. Реально работа испытуемого в психофизическом эксперименте, по-видимому, всегда сопровождается ошибками, ибо получить постоянную оценку одного и того же стимула при многократном его предъявлении практически невозможно. Очевидно, из-за ошибок измерения сигнал с заданной физической величиной ДШГ представляется на шкале субъективных ДШГ областью, в пределах которой величины ДШГ можно считать распределенными нормально (о чем уже говорилось выше).

Соответственно на каждый экспериментальный сигнал может быть получено несколько разных ответов, частота появления которых определяется параметрами распределения субъективных величин ДШГ для этого сигнала, а также взаимным положением этого распределения и области «нулевой десинхронизации» (рис. 4.22,Б). На рисунке приведено распределение субъективных ДШГ, которое гипотетически должно получиться при многократном предъявлении стимула, у которого физическая величина ДШГ равна нулю. Из рис. 4.22,Б следует, что для этого стимула частота ответов «звонкий» будет определяться частью площади под кривой нормального распределения с вертикальной штриховкой, частота ответов «глухой» — площадью без штриховки, а «придыхательный» — площадью с горизонтальной штриховкой. Использование в опыте достаточно большого набора сигналов, равномерно расположенных по шкале физических величин ДШГ, позволяет для каждого из них получить частоты появления этих разных ответов. В результате будут построены графики зависимости этих частот от величины ДШГ (рис. 4.23), обнаруживающие хорошее качественное совпадение с кривыми идентификации, полученными в реальном психоакустическом эксперименте (см. рис. 4.17).

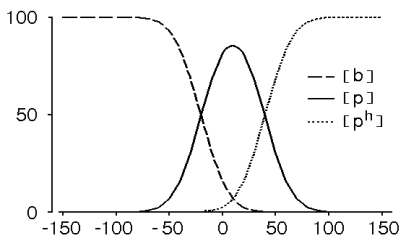


Рис. 4.23. Расчетные «кривые идентификации», при использовании в системе идентификации способа образования смычного согласного детекторов, характеристики которых приведены на рис. 4.22.

Среднеквадратическая ошибка измерения субъективной ДШГ принята равной 20 мс.

По оси абсцисс — ДШГ, мс; по оси ординат — относительная частота появления соответствующих ответов.

Частота ответов «звонкий», равная 0,5, будет получена для стимула, у которого среднее значение субъективной ДШГ совпадает с левой (отрицательной) границей области «нулевой десинхронизации». В языках, где существуют три класса согласных, различающихся по способу образования, для этого стимула с частотой 0,5 будут появляться и суммарные ответы «не звонкий», т.е. «глухой» плюс «придыхательный». Ответы же «придыхательный» станут появляться с частотой 0,5, когда величина ДШГ совпадет с правой (положительной) границей этой области. Таким образом, в опытах, где используются три и бо-

лее категорий ответов, за «фонемную границу» следовало бы принимать те величины ДШГ, при которых с частотой 0,5 появляются ответы, соответствующие обоим концам шкалы, а не проекции точек пересечения кривых идентификации.

Подобное «устройство» легко может быть реализовано в слуховой системе, так как его характеристики практически полностью соответствуют возможностям нейронных схем: 1) нейрон обладает порогом срабатывания и «молчит» при любом входном сигнале, не превышающем этого порога, т.е. не дает выходного сигнала; величиной этого порога может определяться положение «фонемной границы» на шкале ДШГ; 2) нейрон реагирует на входные сигналы только одного знака; поэтому устройство, отвечающее и на положительные, и на отрицательные величины ДШГ, может быть реализовано только с помощью двух разных нейронов; при этом каждый нейрон-детектор будет обладать своим порогом, и величины порогов могут быть разными; 3) после каждого ответа нейрона (после появления на его выходе сигнала) происходит кратковременное повышение порога срабатывания; если ответы следуют достаточно часто и за время между двумя последовательными ответами порог «не

возвращается» к исходному уровню, постепенно устанавливается новая — «динамическая» — величина порога: нейрон адаптируется, но эта величина порога всегда больше, чем у неадаптированного нейрона.

Из этого следует, что область «нулевой десинхронизации» может образоваться благодаря работе двух нейронов и отражает такое их состояние, когда входной сигнал (величина ДШГ) не превосходит порогового уровня ни в одном из них. Ответы «звонкий» будет давать один нейрон, ответы «придыхательный» — другой, а ответы «глухой» будут соответствовать состоянию системы, когда ни один из нейронов не реагирует на данный входной сигнал.

Так как положение «фонемных границ» в системе с детекторами такого рода определяется величинами порогов соответствующих нейронов, логично предположить, что положения однотипных «границ» должны совпадать у носителей разных языков, поскольку определяются они свойствами нейронов, а последние вряд ли могут существенно различаться только потому, что люди говорят на разных языках.

При «утомлении» нейрона-детектора происходит повышение порога его чувствительности, и этим ограничивается диапазон параметров входного сигнала (в данном случае величины ДШГ), на которые этот детектор дает ответы. Таким образом, подобный детектор реализует сразу две модели, предложенные Коулом и Купером (Cole, Cooper, 1977) для объяснения эффектов избирательной адаптации: «утомления» (fatigue) и «перенастройки» (retuning) детекторов. Поскольку повышение порога происходит в самом детекторе, а вариации субъективных параметров сигналов принято относить на счет системы измерения этих параметров, а не выделения признаков, следствием избирательной адаптации должен быть параллельный сдвиг кривых идентификации вдоль оси параметров стимулов относительно их исходного положения (без адаптации).

Теперь посмотрим, что произойдет, если перед очередным тестовым сигналом на такое нейронное устройство многократно подавать адаптирующий сигнал. Начнем с ситуации, когда в системе исследуемого языка присутствуют три типа согласных, различающихся по способу образования.

Если адаптером является сигнал, соответствующий сто процентно звонкому согласному, то отвечать на него и адаптироваться будет только нейрон, реагирующий на отрицательные величины ДШГ. Результатом будет повышение порога этого

нейрона, т.е. сдвиг «фонемной границы», разделяющей «звонкость/глухость», в направлении более отрицательных величин ДШГ (в направлении адаптера). При таком адаптере никакие ошибки в измерении его ДШГ не приведут к возбуждению второго нейрона (его входные сигналы должны иметь другой знак). Следовательно, этот нейрон не будет подвержен адаптации и, значит, никакого сдвига «фонемной границы» для «глухости/придыхательности» возникнуть не должно. Если бы некоторые реализации адаптирующего сигнала все-таки вызвали ответы этого нейрона, сдвиг границы мог бы произойти только в сторону уменьшения вероятности появления ответов «придыхательный», т.е. в данном случае в направлении от адаптера.

Если в качестве адаптера используется придыхательный согласный, картина будет обратная.

При использовании в качестве адаптера стопроцентно глухого согласного адаптация может привести к сдвигу «фонемных границ» только в направлении от адаптера, т.е. в сторону уменьшения вероятности появления как ответов «звонкий», так и ответов «придыхательный», независимо от того, будут ли адаптироваться сразу оба нейрона или только один из них.

Таблица 4.3. Положение «фонемных границ» для губных смычных согласных тайского языка (по: Donald, 1976)

Тип	Испыт. границы	До адаптации	После адаптации стимулом:		
			[b]	[p]	[p <sup>h</sup> ]
[b]–[p]	1	-25	-33	-24	-23
	2	-19	-21	-17	-18
	3	-18	-32	-8	-25
	4	-23	-34	-22	–
	5	-24	-33	-21	-28
[p]–[p <sup>h</sup> ]	1	27	25	24	33
	2	22	19	–	23
	3	20	16	13	31
	4	28	27	26	–
	5	30	26	24	46

В табл. 4.3 приведены результаты, полученные Дональд (Donald, 1976) в опытах по идентификации стимулов с разными величинами ДШГ (от -80 до 70 мс) методом избирательной адаптации. В качестве «звонкого» ([b]) адаптера использовался сигнал с величиной ДШГ -80 мс, «глухого» ([p]) – 5 мс и «придыхательного» ([p<sup>h</sup>]) – 70 мс.

Результаты типичны для опытов по избирательной адаптации: границы сдвигаются в направлении адаптирующего сигнала. При этом одна и та же граница ( $[b]-[p]$  или  $[p]-[p^h]$ ) сдвигается в разных направлениях в зависимости от параметров адаптирующего сигнала. Но, как мы выяснили выше, такой результат не может получиться, если сдвиг границы является следствием «утомления» нейрона-детектора и вызванного этим повышения порога его чувствительности.

Если в системе конкретного языка используются всего две категории согласных, различающиеся по способу образования (например, звонкие и глухие), то ответы одного из нейронов становятся избыточными и на следующих уровнях обработки могут игнорироваться. Система будет работать с ответами одного нейрона и оценивать согласный в зависимости от их наличия или отсутствия. Так, для носителей американского варианта английского языка «глухим» является только согласный с большой положительной величиной ДШГ, т.е. «придыхательный» в системе с тремя категориями. Все остальные согласные оцениваются как «звонкие». Следовательно, ответы нейрона, реагирующего на «звонкость», оказываются избыточными, и при дальнейшей обработке информации можно пользоваться только ответами нейрона, реагирующего на «придыхательность». При использовании в качестве адаптера «звонкого» согласного предполагаемое повышение порога у нейрона-детектора «звонкости» не должно сказываться на результатах работы системы в целом, так как ответы этого детектора ею не используются, а нейрон-детектор «придыхательности» отвечать не должен и, значит, не должен повышаться порог его чувствительности. Таким образом, при данном адаптере в результатах психоакустических экспериментов не должен обнаруживаться сдвиг «фонемной границы» и только использование в качестве адаптера «придыхательного» сигнала должно приводить к сдвигу «фонемной границы» в направлении адаптера.

В табл. 4.4 приведены фонемные границы и их сдвиги под влиянием адаптирующего сигнала, полученные Дональд (Donald, 1976) на носителях американского варианта английского языка. Видно, что адаптация стопроцентно «придыхательными» сигналами приводит к сдвигу «фонемной границы» в направлении, предсказанном на основании анализа работы гипотетического нейрона-детектора. Использование других адаптеров вызывает сдвиги границы, которые невозможно объяснить, исходя из параметров рассматриваемых детекторов.

Таблица 4.4. Положение «фонемной границы» для губных смычных согласных американского варианта английского языка (по: Donald, 1976)

Тип	Испыт. границы	До адаптац.	После адаптац. стимулом:		
			[b]	[p]	[p <sup>h</sup> ]
[b]–[p]	1	15	11	10	23
	2	15	7	13	25
	3	18	10	13	27
	4	10	-2	3	24

Из сказанного напрашиваются два вывода: либо для выделения признаков, характеризующих способ образования согласных, используются детекторы с более изощренной структурой, чем описанный выше детектор нейронного типа, либо эффект избирательной адаптации отражает какие-то более сложные процессы в системе, возможно, никак не связанные с работой собственно выделителей признаков.

Тот факт, что эффект избирательной адаптации может не быть следствием изменения параметров конкретного детектора в результате длительной непрерывной его работы, подтверждается некоторыми прямыми экспериментальными исследованиями. В частности, данными, которые были получены в опытах Пизони (Pisoni, 1980), когда он исследовал восприятие неодновременности включения двух тональных посылок и сдвиги границы для таких стимулов под влиянием адаптации речеподобными сигналами с соответственно заданными величинами ДШГ. Автор предполагал, что выделитель неодновременности появления в сигнале некоторых его составляющих должен вести себя одинаково при разной физической природе этих составляющих и в разных сигналах (речевых/неречевых). Как следствие, процесс утомления (адаптации) также не должен зависеть от того, в сигналах какого типа детектор обнаруживает неодновременность появления отдельных составляющих. Оказалось, однако, что использование в качестве адаптера сигнала, качественно отличающегося от тестовых сигналов (речеподобного для тональных посылок и наоборот), не приводит к сдвигу «фонемных границ».

Наблюдаются и обратные эффекты, когда адаптация детектора, теоретически предназначенного для выделения одного признака в сигнале, приводит к сдвигу «фонемных границ» в пространстве другого признака. Так, Савуш (Sawusch, 1977) ссылается на работы (Diehl, 1975; Ganong, 1975), в которых исполь-

зование в качестве адаптера стимулов, содержащих информацию о смычном согласном только на участке шума взрыва, приводило к сдвигу «фонемной границы» в наборе стимулов, где восприятие смычного согласного определялось только формантными переходами.

Чтобы справиться с этими противоречиями, иногда делается предположение о существовании детекторов признаков двух уровней – акустического и фонетического (точнее, видимо, фонологического) – и необъяснимые с точки зрения возможностей акустических детекторов результаты относятся на счет адаптации фонетических детекторов, собирающих данные с выходов нескольких акустических детекторов.

Как может работать «фонетический» детектор? Допустим, что речь идет о детекторе места образования согласного и на его вход может поступать информация от двух «акустических» детекторов: детектора, реагирующего на спектральные характеристики шума взрыва, и детектора формантных переходов гласного.

Если «фонетический» детектор дает ответ только при поступлении на него сигналов с обоих «акустических» детекторов, отсутствие в экспериментальных стимулах одного из признаков (шума взрыва или формантного перехода гласного) делает невозможным появление ответов «фонетического» детектора, а, значит, становится невозможной и его адаптация.

Если же детектор отвечает при поступлении сигнала уже от одного из «акустических» детекторов, сдвиги «фонемных границ» становятся возможными и при несовпадении акустических признаков адаптера и стимулов основного экспериментального набора. Необходимо лишь, чтобы каждый из этих «акустических» признаков вызывал ответы одного и того же «фонетического» детектора. Однако в этом случае специального объяснения потребует тот факт, что использование слогов СГ в качестве адаптера не вызывает сдвигов «границы» в экспериментальном наборе, составленном из слогов ГС с теми же согласными (Ades, 1974). Существуют и другие экспериментальные данные, прямо противоречащие гипотезе «фонетических детекторов» (Roberts, Summerfield, 1981).

Возможно, разрешить это противоречие позволит анализ поведения испытуемого в психофизическом эксперименте вообще и в опытах по избирательной адаптации в частности.

Используя метод «избирательной адаптации», исследователи давно обратили внимание на тот факт, что полученные ими

данные чрезвычайно похожи на результаты психоакустических опытов с использованием «опорного» ('anchoring') стимула (Simon, Studdert-Kennedy, 1978; Diehl e. a., 1980). Суть последних заключается в том, что один из стимулов представлен в экспериментальном наборе чаще остальных. При этом все стимулы предъявляются в случайном порядке, и на каждый из них испытуемый дает ответ в соответствии с данной ему инструкцией. Результатом таких опытов является сдвиг «фонемной границы» в направлении «опорного» стимула, и величина этого сдвига вполне сравнима с величинами сдвигов при избирательной адаптации. В то же время считается, что об адаптации (утомлении) детекторов в опытах с «опорным» стимулом речи быть не может.

В обоих случаях сдвиг «фонемных границ» проявляется в виде параллельного сдвига кривых идентификации вдоль оси исследуемого параметра сигнала относительно их положения при отсутствии адаптации (см., например, рис. 4.24). Здесь обнаруживается совершенно очевидное сходство со сдвигами «фонемных границ» вследствие контекстуальных влияний (см. выше). Из этого делается вывод, что эффект адаптации может сводиться к выработке новой «точки субъективного равенства», или «уровня адаптации» (adaptation level) по терминологии, принятой в зарубежной психофизической литературе.

Таким образом, делается предположение, что эффект избирательной адаптации является следствием специфической организации экспериментальной процедуры, а не адаптации (утомления) гипотетических детекторов признаков. Отмечается, что с учетом поведения испытуемого удастся объяснить направление смещения «фонемных границ» в результате избирательной адаптации, но используемые для этого модели поведения испытуемого не дают надежного количественного описания эффекта избирательной адаптации. Последнее обстоятельство часто используется в качестве аргумента в пользу сенсорной природы эффекта избирательной адаптации, т.е. утомления детекторов. Попробуем показать, что при удачном выборе модели поведения испытуемого удастся получить удовлетворительное количественное описание результатов опытов по избирательной адаптации, учитывая эффект изменения критерия под влиянием условий эксперимента (выработки нового положения «точки субъективного равенства»)

Обратимся еще раз к описанию процедуры опыта по избирательной адаптации. Опыт проводится в два этапа. На первом

этапе испытуемым предъявляют стимулы с выбранными для исследования параметрами и дают инструкцию классифицировать эти стимулы в соответствии с заданными классами. Стимулы предъявляют группами с интервалом между стимулами в 1–5 с и с паузами между группами 10 с и более. На этом этапе получают привычные «кривые идентификации», т.е. распределение частот ответов разных классов соответственно параметрам предъявленных стимулов.

На втором этапе тот же набор стимулов предъявляют группами, но в группу включают меньшее число стимулов и перед каждой группой многократно с короткими интервалами (50–350 мс) предъявляют адаптирующий сигнал. В качестве последнего используют один из стимулов набора, параметры которого обычно соответствуют одной из крайних величин диапазона изменения исследуемого в опыте признака. По инструкции испытуемый должен давать ответы только на стимулы основных групп.

Однако это последнее обстоятельство еще не означает, что испытуемый никак не оценивает, по крайней мере подсознательно, и адаптирующие сигналы тоже. Результаты опытов (Brown, 1953) показывают, что в подобной ситуации полностью исключить такую возможность не удастся, даже когда в инструкции специально указывается на необходимость не оценивать такие сигналы и экспериментатор при каждом их предъявлении напоминает об этом. Избежать подобного эффекта удастся только при условии, что адаптирующий («опорный», anchoring) сигнал заведомо не принадлежит к экспериментальному набору (является «чужим» для него).

В итоге в «субъективном» наборе стимулов (стимулов, прошедших оценку сенсорной системой) будет чаще представлен один из них. Но в опытах по идентификации, а именно к этому классу относятся все опыты с избирательной адаптацией, стратегия поведения испытуемого направлена на достижение равновероятного появления всех заданных условиями эксперимента типов ответов (см. с. 156 и сл.). Чтобы сохранить баланс при более частом появлении в экспериментальном наборе одного из стимулов, системе придется сдвинуть критерий (изменить положение «эталоны» на шкале параметров) в направлении, уменьшающем возможность появления избыточного числа оценок одного типа. Совершенно очевидно, что сдвиг всегда должен происходить в направлении адаптирующего стимула. В целом система будет стремиться к выравниванию вероятности

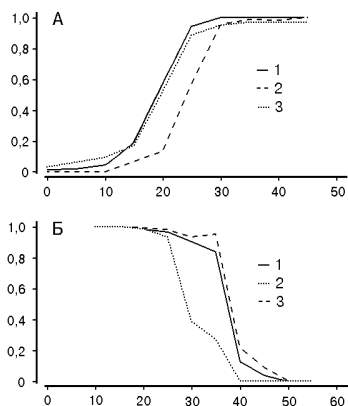


Рис. 4.24. Результаты идентификации синтетических слогов типа [ta]–[da] в обычных условиях и при избирательной адаптации (по: Dechovitz, Mandler, 1978). 1 – в обычных условиях; 2 – в условиях избирательной адаптации сигналом из основного экспериментального набора; 3 – при адаптации сигналом из другого («чужого») набора (перекрестная адаптация). По оси абсцисс – величина ДШГ, мс; по оси ординат – относительная частота ([ta]–ответов – А, [da]–ответов – Б).

формантные синтетические гласные [a] с разными длительностями формантных переходов: 15 и 70 мс (назовем их «тип-15» и «тип-70» соответственно). Из таких сигналов были образованы два экспериментальных набора стимулов с изменяющейся от стимула к стимулу величиной ДШГ (по 10 градаций с шагом 5 мс от 0 до 45 мс для сигналов типа-15 и от 10 до 55 мс – для типа-70). При работе с каждым из наборов в качестве адаптеров использовались стимулы с ДШГ, равной 25 мс, при которой тип-15 уверенно оценивался как слог [ta], а тип-70 – как [da]. Адаптация производилась серией из 32 последовательных предъявлений адаптера с интервалом между предъявлениями в 350 мс. Адап-

появления разных оценок с учетом возможных оценок адаптирующих стимулов.

Строя по результатам такого опыта «кривые идентификации» (естественно, не принимая во внимание возможные оценки адаптирующего сигнала), исследователь обнаруживает преобладание ответов, соответствующих стимулам, параметры которых располагаются на противоположном адаптеру конце шкалы. Это выглядит как сдвиг «фонемной границы» в направлении адаптирующего сигнала. Данные, приведенные в таблицах 4.3 и 4.4, свидетельствуют, что именно такие сдвиги и наблюдаются в опытах по избирательной адаптации.

Адаптация же сигналами из другого набора не приводит к смещению «фонемных границ», как это следует из рис. 4.24, где приведены кривые идентификации, полученные при использовании в качестве адаптера «чужих» сигналов (Dechovitz, Mandler, 1978). В их опытах использовались два типа сигналов, основу которых составляли трехфор-

тирующая серия предъявлялась перед группой из 6-и стимулов основного набора. Таких групп было 20 (каждый стимул предъявляли 12 раз).

Как следует из рис. 4.24, адаптация сигналом из другого набора не приводит к сдвигу «фонемной границы». Там же, где сдвиг имеет место, можно ожидать, что он вызван неосознанным включением в число оцениваемых сигналов минимум 20 дополнительных (адаптирующих) сигналов. Посмотрим, так ли это.

Предположим, что в опыте используются  $m$  разных стимулов, и каждый из них повторяется  $n$  раз. Испытуемые оценивают стимулы, относя каждый из них к одному из двух классов. Частота появления ответов одного из классов при отсутствии адаптации получится равной  $P_{ni}^1$ , где  $i = 1, \dots, m$ . Суммарное число

ответов этого типа будет  $N_n^1 = \sum_{i=1}^m n \cdot P_{ni}^1$  и должно быть примерно равно  $n \cdot m / 2$ .

В условиях адаптации один из стимулов,  $1 \leq j \leq m$ , предъявляется чаще других (первый стимул адаптирующей последовательности). Число предъявлений этого стимула на  $K$  превышает число предъявлений остальных стимулов. Для оценки стимулов используются те же классы ответов.

Частота появления того же ответа, что и в предыдущем случае, будет теперь  $P_{ai}^1$ , где  $i = 1, \dots, m$ . Суммарное число ответов этого типа в опыте с адаптацией будет:

$$N_a^1 = \sum_{i=1}^{j-1} n \cdot P_{ai}^1 + (n+K) \cdot P_{aj}^1 + \sum_{i=j+1}^m n \cdot P_{ai}^1.$$

На основании изложенных выше соображений следует ожидать выполнения условия  $N_a^1 = N_n^1$ . Отсюда:

$$K = \left\{ \frac{\sum_{i=1}^m P_{ni}^1 - \sum_{i=1}^{j-1} P_{ai}^1 - \sum_{i=j+1}^m P_{ai}^1}{P_{aj}^1} - 1 \right\} \cdot n,$$

здесь  $P_{ni}^1$  и  $P_{ai}^1$  — экспериментально полученные частоты ответов для каждого стимула в опыте без адаптации и при избирательной адаптации соответственно.

Для оценки справедливости изложенных выше соображений посмотрим, какие величины  $K$  могут быть получены на основании данных некоторых работ по избирательной адаптации (Dechovitz, Mandler, 1978; Cole, Cooper, 1977; Cooper, 1974).

Таблица 4.5. Сопоставление количества адаптирующих сигналов, вычисленного на основании экспериментальных данных, с заданным в наборе сигналов

Статья	Стимул	Адаптер	Колич. адап. групп	
			заданное	расчетное
Dechovitz, Mandler, 1978	Слог СГ, Г – тип–15	Слог СГ, Г – тип–15 Г – тип–70	20	23
			20	1
	Слог СГ, Г – тип–70	Слог СГ, Г – тип–70 Г – тип–15	20	17
			20	-4
Cooper, 1974	Слоги [be]	Слог [be]	35	25
	" [de]	" [de]	35	28
	" [ge]	" [ge]	35	29
Cole, Cooper, 1977	Слоги [ja-da]	Слог [ja]	9	12
		Слог промежут. между [ja]-[da]	9	6

Оценивая результаты расчетов (табл. 4.5), следует иметь в виду следующее: 1) расчетные формулы не учитывают тех ограничений в смещении «фонемных границ», которые обусловлены допуском на изменение положения «фонемной границы» в системе конкретного языка; 2) с одной стороны, испытуемый может учитывать в адаптирующей последовательности оценку только первого стимула, а, с другой стороны, в какой-то момент времени может вообще перестать обращать внимание на адаптирующую последовательность. Тем не менее, проведенный анализ убедительно показывает, что сдвиги «фонемных границ», наблюдаемые в опытах по избирательной адаптации, можно предсказать (и качественно, и количественно), зная параметры экспериментального набора сигналов и правила принятия решений испытуемым в психофизическом эксперименте.

Из всего сказанного выше можно сделать следующие выводы.

1. Можно придумать вполне правдоподобную схему детекторов признаков, использование которых в функциональной модели восприятия речи обеспечит достаточно хорошее совпа-

дение результатов работы модели с результатами психоакустических экспериментов, использующих метод идентификации.

2. Эффекты сдвига «фонемных границ», наблюдаемые в опытах по избирательной адаптации, не являются следствием адаптации (утомления) детекторов: утомление подобных детекторов должно приводить к сдвигам, прямо противоположным тем, которые наблюдаются в эксперименте. Сдвиги границ определяются особенностями стратегии принятия решений испытуемым в психофизическом эксперименте, а именно неосознаваемым стремлением к уравниванию частот появления разных оценок для данного набора стимулов, т.е. реализацией правила: «стимулы разных типов (и ответы на них) должны появляться в наборе одинаково часто».

3. Результаты опытов по избирательной адаптации не могут служить ни подтверждением, ни опровержением гипотезы о существовании специализированных детекторов признаков. Не годится эта методика и для исследования возможной структуры и свойств таких детекторов.

4. Большого успеха можно добиться, опираясь при построении возможных схем детекторов на свойства нейронных структур.

## ПЛАНИРОВАНИЕ ПСИХОАКУСТИЧЕСКОГО ЭКСПЕРИМЕНТА В РЕЧЕВЫХ ИССЛЕДОВАНИЯХ

Характерной особенностью, свойственной трем рассмотренным выше методам экспериментальной психофизики (психоакустики), является четко обнаруживаемая способность испытуемого к самообучению. Результатом реализации этой способности становятся «контекстуальные влияния» (в чем бы конкретно они ни проявлялись). Видимо, психофизический эксперимент в его классической постановке может быть успешно использован при изучении алгоритмов самообучения, в том числе и процессов обучения языку. Однако в поисках параметров функциональной модели восприятия звуковых сигналов (в частности, речевых) мы, как правило, экспериментально исследуем слушателя с уже сложившимся опытом восприятия таких сигналов (в частности, носителя конкретного языка). При этом больше всего нас интересует, по каким правилам он производит обработку и интерпретацию этих сигналов в естественных, а не экспериментальных условиях взаимодействия с окружающим миром.

Не исключено, что и при исследовании системы восприятия естественной речи классические методы психофизики могут оказаться полезными, но лишь в той мере, в какой механизм самообучения используется носителями конкретного языка в условиях естественного речевого общения. Применительно же к изучению «инвариантных» параметров системы восприятия речи у нормального носителя языка методы психоакустики в их классической постановке оказываются, как мы видели выше, малоэффективными, ибо дают оценки параметров системы, справедливые только для условий данного эксперимента и не более того.

Чтобы результаты, полученные в эксперименте, могли быть перенесены в модель (в дальнейшем речь будет идти только о восприятии речи и его моделировании), либо условия эксперимента должны точно соответствовать условиям естественного восприятия сигналов, либо заранее должны быть известны правила перехода от искусственных экспериментальных условий к естественным. К сожалению, на сегодняшний день мы такими правилами не располагаем. Поэтому следует так планировать эксперимент, чтобы для испытуемого условия работы с предъявляемыми в нем сигналами практически не отличались от естественных.

Выше речь шла о том, что система восприятия речи условно может быть разделена на две подсистемы (модуля): одна из них производит по единым правилам обработку любых акустических сигналов, другая использует разные алгоритмы для анализа сигналов разного происхождения и, в частности, имеет специальный режим («речевой») обработки именно речевого сигнала. Исходя из этого, при экспериментальном исследовании первой подсистемы можно было бы ограничиться неречевыми сигналами (какой бы сложности они ни были), тогда как для изучения второй нужны только речевые сигналы. Сложность, однако, заключается в том, что на сегодняшний день исследователям неизвестен ни полный набор возможных субъективных параметров акустических сигналов, используемый человеком при восприятии внешнего мира, ни даже та его часть, которая необходима и достаточна для нормального функционирования системы восприятия речи. Поэтому у экспериментаторов зачастую нет уверенности в том, что исследуемый параметр сигнала действительно будет использован в условиях естественного восприятия речи. Кроме того, психофизические (психоакустические) методы исследования за-

ранее предполагают, что работа первичной слуховой системы выделения и измерения того или иного признака может наблюдаться исследователем только через механизм принятия решений испытуемого. При этом у исследователя не может быть полной уверенности в том, что правила принятия решений одинаковы для неречевых сигналов и для естественной речи, тем более, что правила эти неизвестны, а восприятие речи предполагает существование специальных режимов работы системы принятия решений. Поэтому, исследуя систему восприятия речи методами психоакустики, предпочтительнее пользоваться в качестве стимулов именно речевыми сигналами.

Многие исследователи, используя в качестве стимулов синтезированные речевые и речеподобные сигналы, исходят из того, что получить естественные речевые сигналы с точно заданными параметрами, да еще изменяющимися в достаточно широких пределах, очень сложно. К тому же при интерпретации результатов опытов с естественными речевыми сигналами приходится принимать во внимание множество не имеющих отношения к акустико-фонетическим признакам сигнала факторов, влияющих на ответ испытуемого и не всегда поддающихся учету.

Однако результаты многих исследований показывают, что надежность восприятия носителями языка синтезированной речи даже при высоком ее качестве уступает надежности восприятия естественной речи. Кажется, экспериментальные данные (Bacri, 1987) подтверждают мнение, что синтезированная речь представляет собой самостоятельный класс сигналов, чем-то отличающихся от естественной речи. Тем более должны отличаться от естественной речи речеподобные синтезированные сигналы, а соответственно и режим обработки их слуховой системой (подсистемой принятия решений) может отличаться от режима восприятия естественной речи.

В то же время проблемы получения экспериментальных стимулов из естественных речевых сигналов практически не существует. Применение вычислительной техники позволяет не только вычленять любой нужный участок из непрерывного речевого потока и измерять его характеристики с требуемой степенью точности, но и модифицировать эти характеристики в соответствии с задачами конкретного экспериментального исследования, не нарушая существенным образом естественности звучания полученных «искусственных» сигналов.

В предыдущих разделах мы убедились, что процедура типичного психоакустического эксперимента создает благоприятные условия для самообучения испытуемого в ходе самого эксперимента. Результатом этого является выработка испытуемым правил и критериев принятия решений, позволяющих ему успешно справиться с поставленной перед ним в данном опыте задачей на базе общих стратегий решения поведенческих задач (возможно, заложенных генетически), которыми пользуется человек. Чтобы избежать подобной «подстройки» к заданным условиям эксперимента, определяемым набором стимулов и инструкцией, и попытаться получить информацию о правилах и критериях, используемых в ситуациях естественного речевого общения,<sup>3</sup> следовало бы резко ограничить в эксперименте возможность самообучения: исключить многократное повторение в наборе одних и тех же сигналов и не предъявлять набор конкретному испытуемому более одного раза. Для получения статистически достаточного объема экспериментальных данных можно увеличить, во-первых, число привлекаемых к опыту испытуемых, а, во-вторых, разнообразие сигналов в экспериментальном наборе так, чтобы многократное включение в него какого-то элемента не создавало впечатления частой его повторяемости. Например, гласный заданного качества и длительности может быть многократно использован в составе **разных** слогов и слов, образующих набор. Это позволит многократно «предъявить» его одному испытуемому, но, надо надеяться, не создаст у последнего впечатления многократной повторяемости одного и того же элемента.

По-видимому, следует отказаться и от метода «вынужденного выбора», когда экспериментатор задает испытуемому заранее ограниченный набор классов, к которым надо отнести каждый из предъявляемых в опыте стимулов. Нет сомнений, что в естественных условиях носители языка пользуются для описания сигналов (по крайней мере, их сегментного состава) достаточно ограниченным набором категорий (классов). Од-

---

<sup>3</sup> Здесь мы исходим из того, что типичный носитель языка представляет собой уже обученную систему, которая в главной своей части, отражающей особенности конкретного языка (по крайней мере, его фонетического строя), не требует переобучения. В то же время процесс восприятия естественной речи, видимо, может сопровождаться «текущим» переобучением. Принятие решения «речевой/неречевой» – пример этого. Не исключено, что существуют и другие возможности.

нако количество и характер этих категорий являются предметом поиска, т.е. заранее неизвестны экспериментатору, и нет никаких гарантий, что заданные инструкцией вынужденного выбора категории совпадут (по числу и характеристикам) с имеющимися у испытуемого. Если такого совпадения не произойдет, испытуемый будет стремиться выполнить данную ему инструкцию и станет относить предъявляемые стимулы к заданным исследователем классам, вырабатывая правила классификации в ходе опыта, зачастую вопреки тем правилам, которые он использовал бы в естественных условиях. Ясно, что последующее использование результатов такого эксперимента в модели восприятия речи становится бессмысленным.

Возможно, более надежные результаты даст метод «диктанта», когда испытуемого просят записывать с максимальной возможной точностью все, что он слышит. При такой инструкции испытуемый должен (теоретически) пользоваться категориями и классами, сложившимися у него в результате опыта речевого общения. Но, давая такую инструкцию, следует проявлять осторожность при последующей интерпретации результатов. Это связано с неопределенностью термина «слышит». Выше мы уже говорили, что слушатель не в состоянии осознавать результаты работы модуля первичной обработки акустического сигнала и что осознаваемыми становятся только решения, принимаемые фонетико-фонологическим модулем на основании полученных субъективных признаков. Поэтому чаще всего термин «слышит» следует относить именно к принятым решениям, хотя нередки случаи, когда испытуемый «слышит» то, чего не было в исходном акустическом сигнале.

По этой причине исследователь должен попытаться путем подбора экспериментального материала и детализацией инструкции уменьшить влияние системы принятия решений на ответы испытуемого.

К примеру, во избежание влияния внутреннего словаря слушателя и частотности лексем при принятии решений о сегментном составе желательно использовать в качестве стимулов слова, заведомо отсутствующие в словаре данного слушателя. При этом надо учитывать, что в процессе принятия решения носитель языка может учитывать не только наличие/отсутствие предъявляемого стимула в словаре (если этот стимул имеет размерность слова), но и вероятность существования в языке отдельных комбинаций звуков, образующих стимул.

Возможно, в подобных случаях испытуемых следует слегка «обманывать», сообщая им в инструкции, что предъявляемые стимулы взяты из слов (фраз) неизвестного им языка, и предложив записать их буквами русского алфавита с максимальным сохранением особенностей звучания услышанного стимула. Ведь умеем же мы передавать на письме особенности звуков, создаваемых животными или возникающих в окружающей нас природе от других естественных источников.

Все, о чем говорилось выше, хотя и имеет непосредственное отношение к планированию эксперимента, касается, скорее, организационной его стороны, т.е. того, как следует или не следует проводить эксперимент. Наиболее важной его стороны – концептуальной, т.е. того, что предполагается исследовать в конкретном эксперименте, мы практически не касались, хотя именно этим определяется и выбор методики эксперимента, и характер стимулов, и вид инструкции, в соответствии с которой должен будет работать испытуемый. Здесь все зависит от того, какой из модулей системы восприятия речи предполагается исследовать, а также от гипотетических алгоритмов работы этого модуля. Очевидно, что неременным условием продуктивности планируемого эксперимента является предварительное детальное описание всех гипотетических алгоритмов и предполагаемых результатов их работы.

Настоящий раздел монографии распадается на две относительно самостоятельные части. В первой выделяются результаты предпринятого теоретического и экспериментального исследования, которые авторам представляются наиболее важными. Вторая посвящена принципам становления перцептивного механизма как самоорганизующейся и самообучающейся системы.

Система восприятия речи предстает как «многоблочное» устройство; основная граница разделяет психоакустический модуль, призванный производить первичную обработку акустического сигнала, и модули высших уровней.

Психоакустический модуль максимально автономен, функционирует по типу «снизу вверх» и не подвержен влиянию фонетико-фонологического и прочих «лингвистических» (языковых) и общекогнитивных высших блоков системы.

Можно считать, согласно с утвердившейся в специальной литературе традицией, что важным компонентом психоакустического модуля являются детекторы акустических признаков. Но под такими детекторами разумно понимать лишь нейронные устройства с жестко очерченными, сравнительно простыми функциями, с генетически заданными свойствами и возможностями. Детекторы не могут быть источником окончательной информации там, где речь идет о процессе принятия решений относительно функциональных (языковых) единиц и их свойств.

Модули, «расположенные над» психоакустическим, принимают решения о функциональной структуре сигнала, основываясь на объективных показаниях психоакустического модуля (данных, однако, в виде некоторого субъективного представления) и огромного множества разнообразных факторов – от собственно языковых, относящихся ко всем компонентам языка, до фоновых знаний. Основной принцип работы здесь – «сверху вниз» (при прочих равных условиях предпочтение отдается информации высших уровней), в то же время происходит постоян-

ный обмен информацией между параллельно работающими блоками-модулями.

Восприятие речи есть приписывание сигналу языковой структуры. Соответственно разной может быть «глубина» восприятия: от опознания фонемного облика до установления полной семантической структуры высказывания (текста).

Соотношение сегментации и идентификации акустического сигнала не должно рассматриваться как постоянно заданное. Перцептивные стратегии многообразны и подвижны: если, например, объективные свойства речевого сигнала способствуют более или менее надежной сегментации, уместно ожидать, что именно этому аспекту в работающей системе будет придан более высокий функциональный вес.

Типологические особенности языков могут выдвигать на первый план те или иные перцептивные стратегии; простейший пример – высокий функциональный вес процедур членения на слова в языках с фонологически фиксированным ударением.

При изучении восприятия речи мы всегда имеем дело с деятельностью целостной системы, когда нерасчлененно представлены результаты работы всех модулей, блоков, уровней. Задача исследователя – разграничить «эффекты», относящиеся к функционированию различных компонентов системы.

Один из распространенных методических недостатков перцептивных экспериментов – прямолинейное сопоставление параметров экспериментального сигнала с ответами испытуемых без должного учета скрытых от внешнего наблюдателя внутренних процессов, которые могли привести к данному результату.

Иначе говоря, исследование интересующего нас типа невозможно без выдвижения гипотез относительно природы перцептивного механизма с последующей их верификацией. Единственно реальный путь построения действующей модели восприятия речи, базирующейся на тех или иных гипотезах, – это «прогривание» ее возможных компьютерных версий.

В литературе к настоящему времени накоплено огромное количество релевантных данных, которые позволяют уже сейчас ставить в практической плоскости вопрос о разработке действующих компьютерных моделей восприятия речи, несмотря на общепризнанную сложность задачи.

Вполне очевидно, что реально это могут быть лишь модели «поверхностного» восприятия (см. гл. II); особое внимание

естественно было бы уделить имитации упрощенного доступа к словарю по минимуму признаков (гл. III).

Во второй части заключительного раздела мы хотели бы попытаться дать предварительный ответ на следующий вопрос: какими механизмами должен владеть человек от рождения, чтобы восприятие речи стало возможным, и какие знания он должен приобрести в процессе овладения языком применительно к указанной задаче? Заметим, что когда мы говорим о врожденных механизмах, то не разграничиваем аспекты, связанные с доступностью тех или иных операций уже в ранний постнатальный период, и все стороны деятельности, становящиеся возможными в силу созревания организма. Мы не различаем также в данном случае механизмы и знания, поскольку последние, если это знания процедурного типа, с функциональной точки зрения представляют собой механизмы.

При ответе на поставленный вопрос мы будем использовать результаты нашего исследования, а также литературные данные, в том числе не фигурировавшие в основной части книги.

Решение проблемы, контуры которого мы хотим наметить, с нашей точки зрения, и должно быть естественным итогом начального этапа моделирования восприятия речи (и, шире, речевой деятельности). Ведь такая модель, как мы не раз отмечали, должна быть самообучающейся, а это прежде всего и означает, что должны быть заданы «стартовые условия», а также стратегии, позволяющие от последних перейти к некоторому «функционально зрелому» состоянию. «Стартовые условия» — это те возможности, которыми располагает человек генетически, а стратегии, по преимуществу также врожденные, обеспечивают формирование развитого языкового механизма на базе уже доступных от рождения возможностей и взаимодействия со средой.

Начать можно с того, что, как показывают некоторые наблюдения, ср. (Brown 1973) и др., уже в раннем постнатальном периоде дети по-разному реагируют на неречевые звуки и речь. Более того, дети 4-месячного возраста, как обнаружено, в окружающей их речевой среде выделяют так называемую материнскую речь (*motherese*), т.е. речь, для которой характерны сравнительно медленный темп, небольшая длительность каждого высказывания, выделенность синтагм и фраз паузами, некоторая утрированность мелодических перепадов (Fernald, Simon, 1984): в экспериментальных условиях взор младенцев

дольше задерживается на громкоговорителе, когда через него воспроизводится материнская речь, чем тогда, когда звучат произвольные речевые образцы (вне зависимости от того, используется ли в эксперименте речь матери данного ребенка), см. (Gleitman e.a. 1989, p. 153).

Иначе говоря, владение базовой оппозицией «речь/не-речь» (Касевич 1983), где в качестве первого члена оппозиции выступает речь материнская, принадлежит, вероятно, к генетически заданным «стартовым условиям». Из этого, в свою очередь, следует, что человеку от рождения присуще умение выделять из звукового фона определенные признаки организованности в акустических событиях.

Специфические характеристики материнской речи отчасти указывают на то, каковы именно параметры, которые должен обрабатывать ребенок, чтобы воспринимать речь (напомним: чтобы воспринимать речь, необходимо владеть языковой системой; однако, чтобы овладеть данной языковой системой, нужно уметь воспринимать, в той или иной степени, языковой материал соответственной речевой среды – процесс не может не быть рекурсивным). Достаточно ясно, что ребенок должен уметь определять длительность акустического сигнала, а также оценивать крутизну частотных перепадов.

Несколько сложнее обстоит дело с паузами. Как следует из наших результатов (см. гл. III), то, что с функциональной точки зрения трактуется как пауза, акустически представляет собой комплексное событие: частотно-динамические изменения определенного вида плюс перерыв фонации, причем последний может быть факультативным. Данные неопубликованной работы К.Хирш–Пасека и др., на которую ссылаются Л.Глайтман и др. (Gleitman e.a., 1989), указывают, что именно таким образом функционирует пауза уже в восприятии детей доречевого периода. Хирш–Пасек и его соавторы изучали реакцию 8-месячных детей на материнскую речь, передаваемую через громкоговоритель, в двух экспериментальных режимах: в одном на границах предложений вводилась постоянная пауза длительностью в 0,5 с, в другом та же пауза фигурировала в позициях, случайных с синтаксической точки зрения, обычно перед третьим словом от конца предложения. Было обнаружено, что дети лучше реагируют на вариант с паузами в конце предложений, что определялось по времени фиксирования взгляда на громкоговоритель. Л.Глайтман и др. адекватно, с нашей точки зрения, интерпретируют экспериментальные данные: «...раз-

личные акустические ключи одновременно присутствуют в позиции конца предложения. Результатом экспериментальной модификации [сигнала] было либо поддержание этого совпадения различных ключей (когда паузы вводились на границах предложений), либо диссоциация ключей (когда паузы вводились там, где другие [сопутствующие им] параметры обычно не представлены). Результаты свидетельствуют, что младенцы способны отождествить указанные ключи с границами предложений, стремятся к этому и ожидают, что ключи будут совпадать в некоторых временных точках...» (Gleitman e.a., 1989, p. 163).

Иначе говоря, можно считать, что оперирование паузами как сложными акустическими событиями, связанными с сегментацией речевого потока (прежде всего на предложения-высказывания), тоже принадлежит к врожденным умениям. Пауза – перерыв фонации определенной длительности – при восприятии речи используется, как полагают, также для того, чтобы реализовать переход от поверхностно-синтаксической структуры предложения к его глубинной структуре (см. гл. III, с. 116), и эта стратегия, будучи универсальной, также должна быть врожденной.

Существуют ли какие-то врожденные механизмы для выделения из звучащего текста слов? С одной стороны, имеются многочисленные свидетельства того, что ребенок достаточно рано «извлекает» из взрослой речи слова, при этом заменяя их квазиэквивалентами – чаще всего соответствующими ударными слогами, например *ko* вместо *молоко*, *raff* вместо *giraffe*, см. об этом (Slobin, 1973; Касевич, 1977; Gleitman, Wanner, 1982). С другой стороны, исследования Р.Уир (Weir, 1966) показали, что оперирование ударением едва ли можно считать врожденным умением, ибо в собственной лепетной речи детей использование ударения или тона определяется языковой средой: если американские дети произносят итерированные слоги, организованные акцентным контуром, то китайские дети того же возраста в своей лепетной речи используют тонированные однослоги. Исходя из сказанного, можно предположить, что к врожденным стратегиям принадлежит умение выделять синтагматически контрастные слоги: ударный слог всегда – по определению – обладает выделенностью по отношению к соседним безударным, а тонированный слог противопоставлен своим «соседям» как самой сферой действия тона, который в наиболее типичном случае распространяется именно на слог, так и тем обстоятельством, что

ситуация соположения слогов с разными тонами статистически преобладает. Обратимся к вопросу о реалистичности наличия врожденных стратегий, относящихся к сегментации речевого потока на минимальные функциональные элементы, в типичном случае отвечающие фонемам, и к идентификации минимальных сегментов.

Что касается последнего аспекта, то в литературе в свое время активно обсуждались результаты, согласно которым не только дети предречевого периода, но и животные – крысы и некоторые другие грызуны, обезьяны-макаки – реагируют на различия согласных по звонкости/глухости, гласных по формантной структуре и т.п., см. об этом (Kuhl, Miller, 1975; Чистович и др., 1976; Jusczyk, 1986; Kluender e.a., 1987) и другие работы. Как уже упоминалось в гл. I, на этом основании ряд авторов приходит к выводу, согласно которому дифференциальные признаки фонем – врожденные категории, врожденным является и само умение воспринимать соответствующие акустические параметры категориально.<sup>1</sup>

Ситуация представляется более сложной. Приведем данные чрезвычайно интересных экспериментов, в которых изучались реакции 2-месячных детей на открытые слоги, отличающиеся согласными (Eimas, 1975). Интенсивность реакции определялась по частоте сосательных движений; при сжатии соски «пустышки» губами ребенок слышал акустический сигнал – открытый слог с тем или иным согласным. Было обнаружено, что частота сосательных движений возрастает при первом и некотором числе последующих предъявлений, например, слога /ba/, затем уменьшается при привыкании, а потом снова возрастает при предъявлении слога /pa/. Весьма любопытно, что в описанных экспериментальных условиях японские дети точно таким же образом реагировали на различие слогов /ra/–/la/ – при том, что носители японского языка «знамениты» своей «неспособностью» различать [r] и [l].

---

<sup>1</sup> Существует и точка зрения, согласно которой те или иные умения новорожденных, в том числе относящиеся к речи, не должны рассматриваться как свидетельства их врожденности: предполагается, что процесс обучения может начаться уже в период внутриутробного развития, см. напр., (Krasnegor, 1989). Не обсуждая специально соответствующую концепцию, отметим лишь, что даже признание ее реалистичности мало что изменило бы в наших представлениях, ибо и в этом случае речь шла бы о некоторых генетически заданных стратегиях, только на сей раз пре-, а не постнатальных.

Здесь уместно вспомнить, что в период, который условно и неточно принято называть доречевым и последним этапом которого служит стадия лепетной речи, ребенок обнаруживает умение произносить все разновидности гласных и согласных, которые встречаются в языках мира, см. об этом, например, (Касевич, 1983). В дальнейшем, в зависимости от языкового окружения, одни звуковые различия сохраняются, приобретая статус фонологических, другие же, не получая подкрепления, «отмирают». Иначе говоря, имеет место взаимодействие врожденных умений, универсально – что абсолютно естественно – обслуживающих усвоение любого языка, и умений приобретаемых, становление которых определяется средой. В то же время практически очевидно, что стратегии приобретения такого рода умений в свою очередь должны быть врожденными.

По аналогии с этим, а также учитывая приведенные выше экспериментальные данные, можно предположить, что в области восприятия речи мы встречаемся с принципиально той же картиной. От рождения человек вооружен способностью различать любые гласные и согласные, возможные в языках мира, т.е. способностью оперировать любыми акустическими параметрами, которые могут характеризовать результат работы артикуляторного аппарата.<sup>2</sup> В раннем онтогенезе, однако, на эти универсальные потенции налагается влияние языковой среды, которая поддерживает одни различия и не поддерживает другие. В результате и формируются перцептивные стратегии, которые можно отнести к фонологическим. Врожденные умения, позволяющие различать любые «человеческие» гласные и согласные, конечно, не утрачиваются, они лишь меняют «уровень», а точнее, входят составной частью в набор операций психоакустического модуля, когда над ним надстраивается фонологический (ср. Werker, Tees, 1984).

Наконец, о такой же линии развития, в котором взаимодействуют врожденные механизмы и приобретенные под влиянием среды умения, можно говорить и применительно к сегментации речевого потока на минимальные звуковые сегменты. Л.Глайтман и ее соавторы пишут, что, «по-видимому, не требуется никакого специального процесса обучения (no learning

---

<sup>2</sup> Общность в этом отношении с животными, которая упоминалась выше, может объясняться тем, что между характеристиками речевых и неречевых звуков нет непроходимой пропасти; впрочем, этот вопрос заслуживает отдельного изучения, см. (Porpell e.a., 1991) и другие работы.

apparatus) для начальной сегментации акустической волны на дискретные фонемы. Сегментация обеспечивается [самой] нервной системой [человека]» (Gleitman e.a., 1989, p. 154). Если под фонемами, как это обычно принято, понимать минимальные звуковые сегменты, отвечающие фонемам, т.е. выступающие конкретными экземплярами абстрактных фонем, то приведенное суждение вряд ли справедливо. Широко известно, что физически один и тот же сегмент может получить монофонемную интерпретацию в одном языке и полифонемную – в другом. Это было бы невозможным, если бы сегментация на фонемы, каждый из которых равен ровно одной фонеме, была предопределена генетически, обусловлена непосредственно параметрами нервной системы человека.

И здесь также можно предположить, что существует врожденная способность дискретизировать акустический сигнал, выделяя сегменты, которые в принципе, потенциально могли бы быть представителями некоторых фонем. Под воздействием языковой среды в дальнейшем вырабатывается умение: относительно одних сегментов принимать решение, согласно которому они выступают представителями соответствующих фонем, относительно других – решение об объединении с другим сегментом или сегментами «в рамках» одной фонемы (а возможно, и решение о том, что данный сегмент есть последовательность двух или более фонем).

Иначе говоря, механизмы врожденной «автоматической» сегментации (для действия которых, кстати, как и для механизмов такой же идентификации, речевой сигнал может содержать недостаточную информацию) по мере адаптации ребенка к его языковой среде тоже оказываются ограниченными сферой психоакустического модуля; на основании данных, поставляемых этим модулем, принимаются фонологические решения, культурно – в широком смысле – обусловленные.

Мы затронули лишь некоторые аспекты, относящиеся к принципам функционирования самообучающейся системы, способной к восприятию речи (в стороне остались все вопросы семантики и грамматики, применительно к которым тоже необходимо исследовать соотношение врожденного и приобретаемого). В целом можно согласиться с Н.Хомским, который говорит: «Мы должны принять, что существуют некоторые долигвистические понятия, с помощью которых возможно извлекать из [окружающего] мира его элементы – скажем, такие-то значения и такие-то звуки» (Chomsky, 1982, p. 119).

- Антипова А.М.** Ритмическая система английской речи. М., 1984.
- Аткинсон Р., Бауэр Г., Кронорс Э.** Введение в математическую теорию обучения / Под ред. О.К.Тихомирова. М., 1969.
- Бардин К.В.** Проблема порогов чувствительности и психофизические методы. М., 1976.
- Бахтин М.М.** Эстетика словесного творчества. М., 1979.
- Веккер Л.М.** Психические процессы. Т. 1. Л., 1974.
- Венцов А.В.** Чем определяется разностная субъективная чувствительность по длительности // Доклады X Всесоюзн. акустич. конференции (Москва, 1983): Секц. X. М., 1983.
- Венцов А.В., Надпорожская Е.В., Солнушкин С.Д., Чихман В.Н.** Организация интерактивного режима обработки речевых сигналов // Сенсорные системы. 1990. Т. 4. N 3.
- Галунов В.И., Люблинская В.В., Чистович Л.А.** О моторной теории восприятия звуковых сигналов // Вопросы бионики / Отв. ред. М.Г.Гаазе-Рапопорт. М., 1967.
- Галунов В.И., Родионов В.Д.** Моделирование процессов передачи информации в звуковом диапазоне. Л., 1988.
- Ганзен В.А.** Восприятие целостных объектов. Л., 1974.
- Гибсон Дж.** Экологический подход к зрительному восприятию. М., 1988.
- Голузина А.Г.** Шкалирование субъективных расстояний между синтетическими гласными // Анализ речевых сигналов человеком / Под ред. А.В.Венцова и др. Л., 1971.
- Гордина М.В., Быстров И.С.** Фонетический строй вьетнамского языка. Л., 1984.

- Забродин Ю.М.** Введение в общую теорию сенсорной чувствительности. (Адаптивные особенности сенсорных процессов в задачах исследования порогов чувствительности) // Психофизические исследования / Под ред. Б.Ф.Ломова и Ю.М.Забродина. М., 1977.
- Зиндер Л.Р., Касевич В.Б.** Фонема и ее место в системе языка и речевой деятельности // Вопр. языкознания. 1989. N 6.
- Кант И.** Критика чистого разума // Кант И. Соч.: В 6 т. Т. 3. М., 1964.
- Касевич В.Б.** О критериях адекватности модели распознавания речи // Автоматич. распознавание слуховых образов: Тезисы докладов VIII Всесоюзн. семинара. Львов, 1974. Ч. I.
- Касевич В.Б.** Элементы общей лингвистики. М., 1977.
- Касевич В.Б.** Фонологические проблемы общего и восточного языкознания. М., 1983.
- Касевич В.Б.** Морфонология. Л., 1986. .
- Касевич В.Б.** [Рец. на кн.:] Антипова А.М. Ритмическая система английской речи // Вопр. языкознания. 1988. N 1.
- Касевич В.Б., Рыбин В.В., Шабельникова Е.М.** Словесная разборчивость, тип языка и стратегии восприятия речи // Проблемы инженерной психологии. Вып. 3. Специальные проблемы инженерной психологии / Отв. ред. В.Ф.Ломов и др. Л., 1984. Ч. I.
- Касевич В.Б., Рыбин В.В., Шабельникова Е.М.** Перцептивное квантование китайского текста (в сопоставлении с японским и русским) // Актуальные вопросы китайского языкознания: Материалы V Всесоюзной конференции. М., 1990а.
- Касевич В.Б., Рыбин В.В., Шабельникова Е.М.** О стратегиях сегментации текста (на материале китайского, японского и русского языков) // Востоковедение, 18. Л., 1993.
- Касевич В.Б., Шабельникова Е.М., Рыбин В.В.** Ударение и тон в языке и речевой деятельности. Л., 1990б.
- Клаас Ю.А.** О членении на слова предложений, лишенных лексической информации // Механизмы речеобразования и

восприятия сложных звуков / Под ред. А.В.Венцова и др. М.;Л., 1966.

**Кривнова О.Ф.** Прагматическая функция интонационного членения // Проблемы фонетики и фонологии: Материалы Всесоюзного совещания. М., 1986.

**Кривнова О.Ф.** Интонационное членение как средство управления процедурой смыслового распознавания // Экспериментальная фонетика. М., 1989.

**Кузьмин Ю.И., Лисенко Д.М.** Контекстуальные влияния в экспериментах по определению фонемных границ // Анализ речевых сигналов человеком / Под ред. А.В.Венцова и др. Л., 1971.

**Леонов Ю.П.** Теория статистических решений и психофизика. М., 1977.

**Лисенко Д.М.** О членении человеком непрерывного потока речи на слова // Механизмы речеобразования и восприятия сложных звуков / Под ред. А.В.Венцова и др. М.;Л., 1966.

**Люблинская В.В.** Воспроизведение простых контуров изменения частоты основного тона звуков // Анализ речевых сигналов человеком / Под ред. А.В.Венцова и др. Л., 1971.

**Мурзин Л.Н., Штерн А.С.** Текст и его восприятие. Свердловск, 1991.

**Полани М.** Личностное знание: На пути к посткритической философии. М., 1985.

**Психологический словарь.** М., 1983.

**Румянцев М.К.** Тон и интонация в китайском языке. М., 1972.

**Румянцев М.К.** Машинное моделирование речевых единиц. М., 1991.

**Трубецкой Н.С.** Основы фонологии. М., 1960.

**Финк Л.М.** Сигналы, помехи, ошибки...: Заметки о некоторых неожиданностях, парадоксах и заблуждениях в теории связи. М., 1978.

- Чистович Л.А., Венцов А.В. и др.** Физиология речи. Восприятие речи человеком: Руководство по физиологии. Л., 1976.
- Чистович Л.А., Венцов А.В. и др.** Слуховые уровни восприятия речи. Функциональное моделирование // Акустика речи и слуха / Под ред. Л.А.Чистович. Л., 1986.
- Чистович Л.А., Кожевников В.А. и др.** Речь. Артикуляция и восприятие. М.;Л., 1965.
- Чистович Л.А., Шупляков В.С.** Слуховое измерение первой форманты // Анализ речевых сигналов человеком / Под ред. А.В.Венцова и др. Л., 1971.
- Штейнфельдт Э.А.** Частотный словарь современного русского литературного языка. Таллинн, 1963.
- Штерн А.С.** Перцептивный аспект речевой деятельности. СПб., 1992.
- Abbs J.H., Sussman H.M.** Neurophysiological feature detectors and speech perception: A discussion of theoretical implications // J. Speech and Hear. Res. 1971. Vol. 14. N 1.
- Ades A.** How phonetic is selective adaptation? Experiments on syllable position and vowel environment // Percept. and Psychophys. 1974. Vol. 16. N 1.
- Ainsworth W.A.** Mechanisms of selective feature adaptation // Percept. and Psychophys. 1977. Vol. 21. N 4.
- Allan L.G., Kristofferson A.B.** Judgements about the duration of brief stimuli // Percept. and Psychophys. 1974. Vol. 15. N. 3.
- Allen G.D., Hawkins S.** Phonological rhythm: Definition and development // Child Phonology. Vol. 1. Production / Eds. G.H.Yeni-Komshian e.a. New York e.a., 1980.
- Altmann G.T.M.** Cognitive models of speech processing: An introduction // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990a.
- Altmann G.T.M.** Lexical statistics and cognitive models of speech processing // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990b.

- Arbib M.A.** Schemas, cognition and language: Toward a naturalistic account of mind // *Perspectives on Mind* / Eds. H.R.Otto and J.A.Tuedio. Dordrecht e.a., 1986.
- Bacri N.** Perceptual spaces and the identification of natural and synthetic sentences // *Proc. of the XIth Intern. Congr. of Phonetic Sci.*, August 1-7, 1987, Tallinn. 1987. Vol. 2.
- Bailly G.** Integration of rhythmic and syntactic constraints in a model of generation of French prosody // *Speech Communication*. 1989. Vol. 8. N 2.
- Bard E.G.** Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others // *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Becker C.A.** Semantic context effects in visual word recognition: An analysis of semantic strategies // *Memory and Cognition*. 1980. Vol. 8. N 4.
- Bell C.G. e. a.** Reduction of speech spectra by analysis-by-synthesis techniques // *J. Acoust. Soc. Amer.* 1961. Vol. 33. N 12.
- Bever T.G., Lackner J.R., Kirk R.** The underlying structures of sentences are the primary units of immediate speech processing // *Percept. and Psychophys.* 1969. Vol. 5. N 4.
- Boomer D.S., Dittman A.T.** Hesitation pauses and juncture pauses in speech // *Language and Speech*. 1962. Vol. 5. Pt. 4.
- Brady S.A., Darwin C.J.** A range effect in the perception of voicing // *J. Acoust. Soc. Amer.* 1978. Vol. 63. N 5.
- Broadbent D.E., Ladefoged P.** Vowel judgements and adaptation level // *Proc. Roy. Soc., B*. 1960. Vol. 151.
- Browman C.P., Goldstein L.M.** Dynamic modelling of phonetic structure // *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* / Ed. V.A.Fromkin. New York, 1985.
- Brown D.R.** Stimulus-similarity and the anchoring of subjective scales // *Amer. J. Psychol.* 1953. Vol. 66. N 2.
- Brown G.** Prosodic structure and the Given/New distinction // *Prosody: Models and Measurements* / Eds. A.Cutler and D.R.Ladd. Berlin e.a., 1983.

- Brown R.A.** First Language: The Early Stages. Harmondsworth, 1973.
- Butcher A.** Aspects of the Speech Pause: Phonetic Correlates and Communicative Functions. Kiel, 1981.
- Cairns H.S., Hsu J.R.** Effects of prior context on lexical access during sentence comprehension: A replication and reinterpretation // J. Psycholinguistic Res. 1980. Vol. 9. N 4.
- Carroll J.M., Tanenhaus M.K.** Functional clauses and sentence segmentation // J. Speech and Hear. Res. 1978. Vol. 21. N 4.
- Cazanave-Delyfer T.Th.** Phenomenes de pause // Revue de Phonetique Applique. 1988. Fasc. 87+89.
- Chapin P., Smith T., Abrahamson A.** Two factors in perceptual segmentation of speech // J. Verb. Learn. and Verb. Behav. 1972. Vol. 11. N 2.
- Chistovich L.A. e. a.** Mimicking of synthetic vowels // Speech Transm. Lab. Quart. Progr. Status Rep. 1966. Vol. 2.
- Chomsky N.** The Generative Enterprise: A Discussion with Riny Huybreds and Henk van Reimsdijk. Dordrecht, 1982.
- Chuang C.K., Hiki S., Sone T., Nimura T.** The acoustical features and perceptual cues of the 4 tones of Standard Colloquial Chinese // Proc. of the 7th Intern. Congress on Acoustics. Budapest, 1972. Vol. 3.
- Cole R.A., Cooper W.E.** Properties of friction analyzers for [j] // J. Acoust. Soc. Amer. 1977. Vol. 62. N 1.
- Cole R.A., Jakimik J.** A model of speech perception // Perception and Production of Fluent Speech / Ed. R.A.Cole. Hillsdale, New Jersey, 1980.
- Cooper W.E.** Adaptation of phonetic feature analyzers for place of articulation // J. Acoust. Soc. Amer. 1974. Vol. 56. N 2.
- Creelman C.D.** Human discrimination of auditory durations // J. Acoust. Soc. Amer. 1962. Vol. 34. N 5.
- Cruttenden A.** On the so-called grammatical function of intonation // Phonetica. 1970. Vol. 21. N 3.

- Cutler A.** Syllable omission errors and isochrony // Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler / Eds. H.W. Dechert and M. Raupach. The Hague; Paris; New York, 1980.
- Cutler A.** Auditory lexical access: Where do we start? // Lexical Representation and Process / Ed. W. Marslen-Wilson. Cambridge, Mass.; London, 1989.
- Cutler A.** Exploiting prosodic probabilities in speech segmentation // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M. Altmann. Cambridge, Mass.; London, 1990.
- Cutler A., Carter D.M.** The predominance of strong initial syllables in the English vocabulary // Computer Speech and Language. 1987. Vol. 2.
- Cutler A., Norris D.** The role of strong syllables in segmentation for lexical access // J. Exper. Psychol.: Human perception and performance. 1988. Vol. 14. N 1.
- Dechovitz D., Mandler R.** Duration-contingent effects in adaptation // Haskins Lab. Status Rep. on Speech Research. 1978. SR-55/56.
- Diehl R.L.** The effect of selective adaptation on the identification of speech sounds // Percept. and Psychophys. 1975. Vol. 17. N 1.
- Diehl R.L.** Feature detectors for speech: A critical reappraisal // Psychol. Bull. 1981. Vol. 89. N 1.
- Diehl R.L., Lang M., Parker E.M.** A further parallel between selective adaptation and contrast // J. Exper. Psychol.: Human Percept. and Perform. 1980. Vol. 6. N 1.
- Donald S.L.** The effects of selective adaptation on voicing in Thai and English // Haskins Lab. Status Rep. on Speech Research. 1976. SR-47.
- Eimas P.D.** Auditory and phonetic coding of the r - l distinction by young infants // Percept. and Psychophys. 1975. Vol. 18. N 2.
- Eimas P.D., Corbit J.D.** Selective adaptation of linguistic feature detectors // Cognitive Psychology. 1973. Vol. 4. N 1.

- Elman J.L., McClelland J.L.** Exploiting the lawful variability in the speech wave // *Invariance and Variability in Speech Processes* / Eds. J.Perkell and D.Klatt. Hillsdale, New Jersey, 1986.
- Emmorey K.D., Fromkin V.A.** The mental lexicon // *Linguistics: The Cambridge Survey. Vol. 3: Language: Psychological and Biological Aspects* / Ed. F.J.Newmeyer. Cambridge e. a., 1989.
- Ewen C.J.** Segmental and suprasegmental structure// *Dependency and Non-Linear Phonology* / Ed. J.Durand. London, 1986.
- Fechner G.** Elemente der Psychophysik, Bd. I-II. Leipzig, 1860.
- Fernald A., Simon T.** Expanded intonation contours in mothers's speech to newborns // *Developmental Psychology*. 1984. Vol. 20. N 1.
- Fernberger S.** Instructions and the psychophysical limen // *Amer. J. Psychol.* 1931. Vol. 43. N 1.
- Finn A.N.** Intonational accompaniments of Japanese morphemes WA and GA // *Language and Speech*. 1984. Vol. 27. Pt. 1.
- Flores d'Arcais G.B.** Is there a memory for sentences? // *Acta Psychologica*. 1974. Vol. 38. N 1.
- Flores d'Arcais G.B.** Language perception // *Linguistics : The Cambridge Survey. Vol. 3: Language: Psychological and Biological Aspects* / Ed. F.J.Newmeyer. Cambridge e.a., 1989.
- Fodor J., Bever T.G., Garrett M.F.** The Psychology of Language. New York, 1974.
- Fonagy I.** La ripetizione creativa: Ridondanze espressiva nell'opera poetica. Bari, 1982.
- Forster K.I.** Frequency blocking and lexical access: One mental lexicon or two? // *J. Verb. Learn and Verb. Behav.* 1981. Vol. 20. N 2.
- Fox R.A., Unkefer Y.** The effect of lexical status on the perception of tone // *J. Chinese Linguistics*. 1985. Vol. 13. N 1.

- Fowler C.A.** An event approach to the study of speech perception from a direct-realist perspective // J. Phonetics. 1986. Vol. 14. N 1.
- Fowler C.A.** Calling a mirage a mirage: Direct perception of speech produced without a tongue // J. Phonetics. 1990a. Vol. 18. N 4.
- Fowler C.A.** Comments on the contribution by Pierrehumbert and Nearey // J. Phonetics. 1990b. Vol. 18. N 3.
- Fowler C.A.** [Rec. ad op.:] Cognitive models of speech processing: Psycholinguistic and computational perspectives // Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990 // Language and Speech. 1991. Vol. 34. Pt. 2.
- Frauenfelder U.H., Peeters G.** Lexical segmentation in TRACE: An exercise in simulation // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Gandour J.** Perceptual dimensions of tone: Thai // Southeast Asian Linguistic Studies. Vol. 3 / Ed. N.D.Liem. Canberra, 1979.
- Gandour J.** Tone perception in Far Eastern languages // J. Phonetics. 1983. Vol. 11. N 3.
- Ganong W.F.** An experiment on 'Phonetic adaptation' // Quart. Progr. Rep. MIT Res. Lab. Electronics. 1975. Vol. 116.
- Gardner H.** The Mind's New Science: A History of the Cognitive Revolution. New York, 1985.
- Gardner R.A.** Probability-learning with two and three choices // Amer. J. Psychol. 1957. Vol. 70. N 2.
- Getty D.J.** Discrimination of short temporal intervals: A comparison of two models // Percept. and Psychophys. 1975. Vol. 18. N 1.
- Glass J.R., Zue V.W.** Acoustic segmentation and classification // Proc. of DARPA Program Rev. San Diego, 1987.
- Gleitman L., Gleitman H., Landan B., Wanner E.** Where learning begins: Initial representations for language learning // Linguistics: The Cambridge Survey. Vol. 3: Language: Psychological and Biological Aspects / Ed. F.J.Newmeyer. Cambridge, 1989.

- Gleitman L.R., Wanner E.** Language acquisition: The state of the state of the art // Language Acquisition: The State of the Art / Eds. E.Wanner and L.R.Gleitman. Cambridge, 1982.
- Goldinger S.D., Pisoni D.B., Luce P.A.** Speech perception and spoken word recognition: Research and theory // Research on Speech Perception: Progress Report N 16 / Ed. D.Pisoni. Bloomington, Indiana, 1990.
- Goldman-Eisler F.** Pauses, clauses and sentences // Language and Speech. 1972. Vol. 15. Pt. 2.
- Goldsmith J.** Autosegmental Phonology. New York, 1979.
- Gordon B., Caramazza A.** Lexical access and frequency sensitivity: Frequency saturation and open/close class equivalence // Cognition. 1985. Vol. 21. N 2.
- Halliday M.A.K.** A Course in Spoken English: Intonation. Oxford, 1970.
- Hannan E.P., Engler L.** Juncture phenomena and the segmentation of a linguistic corpus // Language and Speech. 1967. Vol. 10. Pt. 4.
- Harrington J., Johnstone A.** The effects of word boundary ambiguity in continuous speech recognition // Proc. of XIth Intern. Congr. of Phonetic Sci. Tallinn, 1987.
- Harris M.O., Umeda N., Bourne J.** Boundary perception in fluent speech // J. Phonetics. 1981. Vol. 9. N 1.
- Hart J.'t, Collier R., Cohen A.** A perceptual study of intonation: An experimental-phonetic approach to speech melody. Cambridge e.a., 1990.
- Hebb D.O.** The Organisation of Behavior. New York, 1949.
- Helson H.** Adaptation-level as a basis for a quantitative theory of frames of reference // Psychol. Rev. 1948. Vol. 55. N 6.
- Henderson A.I., Nelms S.** Relative salience of intonation fall and pause as cues to the perceptual segmentation of speech in an unfamiliar language // J. Psycholinguistic Res. 1980. Vol. 9. N 2.
- Hood J.D., Poole J.P.** Influence of speaker and other factors affecting speech intelligibility // Audiology. 1980. Vol. 19. N 5.

- Huttenlocher D.P.** A broad phonetic classifier // ICASSP-1986. 1986.
- Huttenlocher D.P., Zue V.W.** A model of lexical access from partial phonetic information // ICASSP-1984. 1984.
- Janota P.** Personal characteristics of speech // Rozprawy Ceskoslovenske Akademie Ved / Rada Spolecenskych Ved. 1967. Rocnik 77. Sesit 1.
- Johnson C.D.** Regular disharmony in Kirghiz // Issues in Vowel Harmony / Ed. R.M.Vago. Amsterdam, 1980.
- Johnson-Laird Ph.N.** Mental Models. Cambridge, Mass., 1983.
- Johnson-Laird Ph.N.** The Computer and the Mind: An Introduction to Cognitive Science. Cambridge, Mass., 1988.
- Juszyk P.W.** A review of speech perception research // Handbook of Perception and Performance / Eds. L.Kaufman e.a. New York, 1986.
- Kasevich V.B., Slepokurova N.A., Ventsov A.V.** Strategies of segmentation in speech perception // Anthropomorphic Systems of Automatic Speech Recognition and Synthesis: Intern. Symposium. St.-Petersburg, 1993.
- Klatt D.** Speech perception: A model of acoustic-phonetic analysis and lexical access // Perception and Production of Fluent Speech / Ed. R.Cole. Hillsdale, New Jersey, 1980.
- Klatt D.** Review of Selected Models of Speech Perception // Lexical Representation and Process / Ed. W.Marslen-Wilson. Cambridge, Mass.; London. 1989.
- Kluender K. e. a.** Japanese quail can learn phonetic categories // Science. 1987. Vol. 237. N 4819.
- Krasnegor N.** Measurement of learning, sensory, and linguistic capacity early in life: a selective overview of recent research // From Reading to Neurons / Ed. A.M.Galaburda. Cambridge, Mass., 1989.
- Krause S.E.** Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults // J. Acoust. Soc. Amer. 1982. Vol. 71. N 4.

- Kreiman J.** Perception of sentence and paragraph boundaries in natural conversation // J. Phonetics. 1982. Vol. 10. N 2.
- Kuhl P.K. Miller J.D.** Speech perception by the chinchilla: Voiced-Voiceless distinction in alveolar plosive consonants // Science. 1975. Vol. 190. N 4209.
- Lamb S.M.** Linguistic and cognitive networks // Cognition: A Multiple View. New York; Washington, 1970.
- Lamb S.M.** Prolegomena to a theory of phonology // Phonological Theory: Evolution and Current Practice / Ed. V.B.Makkai. New York e.a., 1972.
- Lehiste I.** Sentence boundaries and paragraph boundaries – perceptual evidence // The Chicago Linguistic Society: The Elements: A Parasession on Linguistic Units and Levels / Eds. J.Clyne e.a. Chicago, 1979a.
- Lehiste I.** Perception of sentence and paragraph boundaries // Frontiers of Speech Communication Research / Eds. B.Lindblom and S.Ohman. New York, 1979b.
- Lehiste I., Wang W.S.-Y.** Perception of sentence and paragraph boundaries with and without semantic information // Phonologica 1976 / Eds. W.Dressler and O.Pfeiffer. Innsbruck, 1977.
- Liberman A.M. e. a.** Perception of the speech code // Psychol. Rev. 1961. Vol. 74. N 6.
- Liberman A.M., Mattingly I.G.** The motor theory of speech perception revised // Cognition. 1986. Vol. 21. N 1.
- Liberman M., Prince A.** On stress and linguistic rhythm // Linguistic Inquiry. 1977. Vol. 8. N 2.
- Lisker L., Abramson A.** A cross-language study of voicing in initial stops: acoustical measurements // Word. 1964. Vol. 20. N 3.
- Lisker L., Abramson A.** The voicing dimension: Some experiments in comparative phonetics // Proc. of the 6th Intern. Congr. of Phonetic Sci. (Prague, 1967). Prague, 1970.
- Luce P.A., Pisoni D.B., Goldinger S.D.** Similarity neighborhoods of spoken words // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.

- Marslen-Wilson W.D.** Functional parallelism in spoken word recognition // *Cognition*. 1987. Vol. 25. N 1.
- Marslen-Wilson W.D.** Activation, competition, and frequency in lexical access // *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Marslen-Wilson W.D., Welsh A.** Processing interactions during word recognition in continuous speech // *Cognit. Psychol.* 1978. Vol. 10. N 1.
- Martin J.G.** On judging pauses in spontaneous speech // *J. Verb. Learn. and Verb. Behav.* 1970. Vol. 9. N 1.
- Massaro D.W.** *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, New Jersey; London, 1987.
- McClelland J.L., Elman J.L.** Interactive processes in speech perception: The TRACE model // *Parallel Distribution Processing. Vol. 2: Psychological and Biological Models* / Eds. L.McClelland and D.E.Rumelhart. Cambridge, Mass., 1988.
- Mehler J., Dupoux E., Segui J.** Constraining models of lexical access: The onset of word recognition // *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Miller G.A., Nicely P.E.** An analysis of perceptual confusions among some English consonants // *J. Acoust. Soc. Amer.* 1955. Vol. 27. N 2.
- Miller J.D.** Auditory-perceptual approaches to phonetic perception // *J. Acoust. Soc. Amer.* 1982. Vol. 71. S112(A).
- Miller J.D., Engebretson A.M., Vemula N.R.** Vowel normalization: Differences between vowels spoken by children, women, and men // *J. Acoust. Soc. Amer.* 1980. Vol. 68. S33(A).
- Miller J.D., Jongman A.** Auditory-perceptual approaches to stop consonants // *J. Acoust. Soc. Amer.* 1987. Vol. 82. S82(A).
- Minsky M.L., Papert S.** *Perceptrons: An Introduction to Computational Geometry*. Cambridge, Mass., 1969.

- Miron M.S., Brown E.** The comprehension of rate-incremented aural coding // J. Psycholinguistic Research. 1971. Vol. 1. N 1.
- Morrish E.C.E.** The direct-realist theory of speech perception: Counter-evidence from the analysis of the speech of a glossectomee // J. Phonetics. 1990. Vol. 18. N 4.
- Morton J.** Interaction of information in word recognition // Psychol. Rev. 1969. Vol. 76. N 1.
- Morton J.** Word recognition // Psycholinguistics 2: Structures and Processes / Eds. J.Morton and J.D.Marshall. Cambridge, Mass., 1979.
- Morton J.** Disintegrating the lexicon: An information processing approach // On mental representation / Eds. J.Mehler e.a. Hillsdale, New Jersey, 1982.
- Nearey T.M.** The segment as a unit of speech perception // J. Phonetics. 1990. Vol. 18. N 3.
- Norris D.** Word recognition: Context effects without priming // Cognition. 1986. Vol. 22. N 2.
- Norris D.** A dynamic-net model of human perception // Cognitive models of speech processing: Psycholinguistic and computational perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Ohala J.J.** Against the direct realist view of speech perception // J. Phonetics. 1986. Vol. 14. N 1.
- Parallel Distributed Processing.** Vol. 1 and 2 / Eds. D.E.Rumelhart e.a. Cambridge, Mass., 1988.
- Parducci A.** Category judgement: A range-frequency model // Psychol. Rev. 1965. Vol. 72. N 6.
- Pisoni D.B.** Adaptation of the relative onset time of two-component tones // Percept. and Psychophys. 1980. Vol. 28. N. 4.
- Pisoni D.B. e.a.** Speech perception, word recognition and the structure of the lexicon // Speech Communication. 1985. Vol. 4. N 1.
- Pisoni D.B., Luce P.A.** Acoustic-phonetic representations in word recognition // Cognition. 1987. Vol. 25. N 1.

- Poppel E. e. a.** Temporal and spatial constraints for mental modality // *Frontiers in Knowledge-Based Computing* / Eds. V.P.Bhatkar and K.M.Rege. New Delhi, 1991.
- Pulgram E.** Syllable, Word, Nexus, Cursus. The Hague; Paris, 1970.
- Reich Sh.S.** Significance of pauses for speech perception // *J. Psycholinguistic Res.* 1980. Vol. 9. N 4.
- Roberts M., Summerfield Q.** Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory // *Percept. and Psychophys.* 1981. Vol. 30. N 4.
- Rosenblatt F.** Principles of Neurodynamics. New York, 1962.
- Ross J., DiLollo V.** Judgement and response in magnitude estimation // *Psychol. Rev.* 1971. Vol. 78. N 6.
- Rumelhart D.E., Hinton G.E., Williams R.J.** Learning internal representations by error propagation // *Parallel Distributed Processing. Vol. 1: Foundations* / Eds. D.E.Rumelhart e.a. Cambridge, Mass.; London, 1988.
- Sachs M.B., Young E.D., Miller M.I.** Encoding of speech features in the auditory nerve // *The Representation of Speech in the Peripheral Auditory System* / Eds. R.Carlson and B.Granstrom. Amsterdam, 1982.
- Sawusch J.R.** Peripheral and central processes in selective adaptation of place of articulation in stop consonants // *J. Acoust. Soc. Amer.* 1977. Vol. 62. N 3.
- Segui J. e.a.** The role of the syllable in speech segmentation, phoneme identification, and lexical access // *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Seidenberg M.S., Tanenhaus M.K.** Modularity and lexical access // *From Models to Modules: Studies in Cognitive Science from the McGill Workshops* / Eds. I.Gopnik and M.Gopnik. Norwood, New Jersey, 1986.
- Seneff S.** A joint synchrony/mean-rate response model of auditory speech processing // *J. Phonetics.* 1988. Vol. 16. N 1.

- Shamma S.** The acoustic features of speech sounds in a model of auditory processing: Vowels and voiceless fricatives // J. Phonetics. 1988. Vol. 16. N 1.
- Shen X.S.** The use of prosody in disambiguation in Mandarin // Phonetica. 1993. Vol. 50. N 4.
- Shillcock R.** Lexical hypothesis in continuous speech // Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives / Ed. G.T.M.Altmann. Cambridge, Mass.; London, 1990.
- Simon H.J., Studdert-Kennedy M.** Selective anchoring and adaptation of phonetic and nonphonetic continua // Hask. Lab. Status Rep. on Speech Res. 1978. SR-54.
- Slobin D.I.** Cognitive prerequisites for the development of grammar // Studies of Child Language Development / Eds. Ch.A.Ferguson and D.I.Slobin. New York, 1973.
- Slowiaczek L.M.** Effects of lexical stress in auditory word recognition // Language and Speech. 1990. Vol. 33. Pt. 1.
- Sprigg R.K.** Tonal units and tonal classification: Ranjabi, Tibetan, and Burmese // Pakha Sanjam. 1976. Vol. 8.
- Stevens K.N.** Models of phonetic recognition II: A feature-based model of speech recognition // Proc. of the Montreal Satellite Symposium on Speech Recognition (Twelfth Intern. Congr. on Acoustics) / Ed. P.Mermelstein. Montreal, 1986.
- Stevens K.N., Halle M.** Remarks on analysis by synthesis and distinctive features // Proc. of the AFCRL Symposium on Models for the Perception of Speech and Visual Form. / Ed. W.Wathen-Dunn. Cambridge, Mass., 1964.
- Stevens S.S., Galanter E.** Ratio scales and category scales for a dozen perceptual continua // J. Exper. Psychol. 1957. Vol. 54. N 6.
- Taft M., Forster K.** Lexical storage and retrieval of prefixed words // J. Verb. Learn. and Verb. Behav. 1975. Vol. 14. N 6.
- Taft M., Hamblly G.** Exploring the cohort model of spoken word recognition // Cognition. 1986. Vol. 22. N 3.

- Terken J.M.B.** The distribution of pitch accent in descriptive language as a function of informational variables // IPO Ann. Progress Report. 1980. N 15.
- Treisman M., Williams T.C.** A theory of criterion setting with an application to sequential dependencies // Psychol. Rev. 1984. Vol. 91. N 1.
- Tyler L.K., Wessels J.** Quantifying contextual contributions to word-recognition processes // Percept. and Psychophys. 1983. Vol. 34. N 5.
- Umeda N., Quinn A.M.S.** Word duration as an acoustic measure of boundary perception // J. Phonetics. 1981. Vol. 9. N 1.
- Ventsov A.** How to get parameters of the speech perception model from the results of psychoacoustic experiments? // Proc. of the XIth Intern. Congr. of Phonetic Sci., August 1-7, 1987. Tallinn, 1987. Vol. 4.
- Visual Representations of Speech Signals** / Eds. M.Cooke e.a. London, 1993.
- Weir R.H.** Some questions on the child's learning of phonology // The Genesis of Language: A Psycholinguistics Approach / Ed. D.McNeil. Cambridge, Mass., 1966.
- Werker J.S., Tees R.C.** Phonemic and phonetic factors in adult cross language speech perception // J. Acoust. Soc. Amer. 1984. Vol. 75. N 6.
- Zaltieri C.** The rhythm of the text // Semiotica. 1987. Vol. 65. N 3/4.
- Zue V.W.** Models of speech recognition III: The role of analysis by synthesis in phonetic recognition // Proc. of the Montreal Satellite Symposium on Speech Recognition (Twelfth Intern. Congr. on Acoustics) / Ed. P.Mermelstein. Montreal, 1986.

# Оглавление

---

Введение	3
Глава I. Современные модели восприятия речи: критический обзор	8
Глава II. Интегративная модель восприятия речи: возможные подходы	47
Проблема восприятия в гносеологии и психологии	—
Основные принципы представления процессов речевосприятия	52
Глава III. Взаимодействие компонентов и уровней в перцептивных процессах сегментации и идентификации языковых единиц	65
Перцептивная сегментация	—
Роль ритмических структур в сегментации и идентификации языковых единиц	97
Роль сингармонистических моделей в сегментации и идентификации	113
О пределах автономности просодических явлений в перцептивных процессах	123
Роль словаря в перцептивных процессах. Ключевые слова	134
Глава IV. Психоакустика и исследование восприятия речи	144
Психоакустика и модели восприятия речи	—
Метод идентификации и возможности исследования перцептивных эталонов	148
Закон Вебера: факт или артефакт?	162
Избирательная адаптация: детекторы признаков или правила принятия решений	175
Планирование психоакустического эксперимента в речевых исследованиях	199
Заключение	205
Литература	213

Научное издание

*Венцов Анатолий Владимирович  
Касевич Вадим Борисович*

Проблемы восприятия речи

Редактор С.Е.Хазанова  
Художественный редактор Е.И.Егорова  
Технический редактор А.В.Борщева  
Корректоры К.Я.Герловина, Т.Г.Павлова

ИБ + 4128

Лицензия ЛР + 040050 от 05.08.91 г.

---

Сдано в набор 09.12.93 г. Подписано в печать 28.02.94 г.  
Формат 60х84 1/16. Бумага офсетная. Гарнитура журнальная.  
Печать офсетная. Усл.печ.л. 13,48. Усл.кр.-опт. 13,67. Уч.-изд.л. 13,48  
Тираж 205 экз. Заказ 35.  
Издательство СПбГУ 199034, Санкт-Петербург, Университетская наб., 7/9

---

Оригинал-макет изготовлен в лаб. применения вычислительной техники в гуманитарных исследованиях (восточный факультет СПбГУ).  
Компьютерная верстка: А.В.Венцов.

---

Участок оперативной полиграфии типографии издательства СПбГУ  
199061, С.-Петербург, Средний пр., 41.