Чистович Л.А.

Физиология речи

Восприятие речи человеком

Москва
«Книга по Требованию»

УДК 03
ББК 92
Ч-68

Чистович Л.А.
Ч-68 Физиология речи: Восприятие речи человеком / Чистович Л.А. – М.: Книга по Требованию, 2012. – 386 с.

ISBN 978-5-458-35915-3

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

ISBN 978-5-458-35915-3

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

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

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

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

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

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

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

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

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

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

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

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

ПОЯСНЕНИЯ К ТРАНСКРИПЦИИ

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

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

При цитировании работ зарубежных авторов сохранялась символика оригинала.

РАЗДЕЛ I

РЕЧЕВОЙ СИГНАЛ И ПРОБЛЕМЫ ЕГО ОПИСАНИЯ

Глава 1

СВОЙСТВА РЕЧЕВОГО СИГНАЛА И НЕКОТОРЫЕ ВОПРОСЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

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

Определим исходный речевой сигнал как функцию $p(t)$, представляющую изменения звукового давления происходящие в течение того времени, когда говорящий произносит фразу. Функция $p(t)$ может быть выражена электрическим напряжением или другой физической величиной, может быть зарегистрирована тем или иным способом. Ясно, что во всех случаях $p(t)$, определенная на отрезке фразы, полностью характеризует последнюю. При таком определении речь представляется множеством функций $p(t)$, соответствующих разным фразам, произнесенным множеством дикторов.

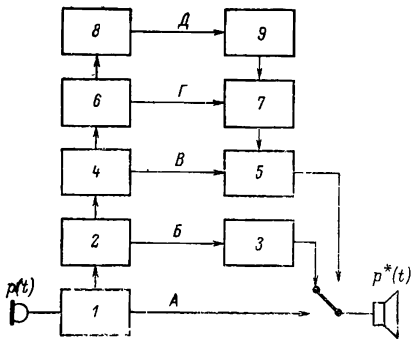

Рис. 1.1. Соотношение основных типов преобразований речевого сигнала в технических системах.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.1. ЭЛЕМЕНТЫ ТЕОРИИ РЕЧЕОБРАЗОВАНИЯ

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

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

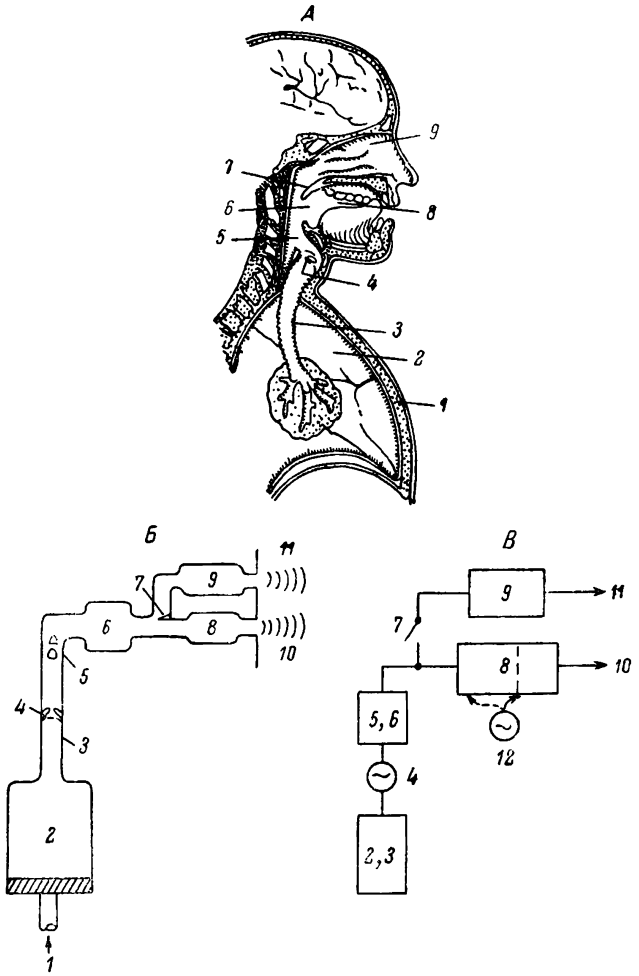

Рис. 1.2. Схема речеобразующего аппарата.

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

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

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

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

1.1.1. КЛАССИФИКАЦИЯ ЗВУКОВ РЕЧИ

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

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

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

3) артикуляторные жесты выполняются последовательно один за другим.

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

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