

Р.К. Потапова Речь: коммуникация, информация, кибернетика. М., Радио и связь. 1997. 527 с.

Распознавание устной речи, ввод текста в компьютер с помощью воспринимающих речь устройств, автоматическое чтение текста и другие сложные современные задачи сейчас становятся в центре внимания как ученых-теоретиков, так и практиков, работающих на ниве компьютеризации и информатизации современного общества. Появление новых основополагающих работ в этой области чрезвычайно важно как для собственно научных исследований, так и для преподавания соответствующих дисциплин в вузах.

Рецензируемая монография задумана как учебное пособие для студентов по специальности «Автоматизированные системы обработки информации и управления», «Линг-

вистика». Однако эта монография интересна также и специалистам, работающим в данной области.

Книга состоит из одиннадцати глав, образующих три части, имеется словарь терминов прикладной лингвистики. Во введении говорится, что обычно в работах, посвященных проблемам речи, генетический аспект ограничивается описанием органов речеобразования и слуха, при этом данные, как правило, отражают статику речеобразования. Задача заключается в том, чтобы переместить центр тяжести описания со статики на динамику. Речевая коммуникация понимается как процесс взаимного обмена сообщениями между динамическими системами. Автор обосновывает целе-

сообразность введения новой отрасли науки и техники — лингвокибернетики — и ее раздела — речевой кибернетики. Базовым понятием лингвокибернетики является естественный языковой код при условии полиарусного рассмотрения его функционирования. В книге рассматриваются новые поколения автоматизированных распознающих речь систем.

Часть I «Речевая коммуникация» открывается первой главой «Речевая коммуникация как объект эволюционного подхода в речеведении». Автор указывает на относительно поздно возникшие свойства языка — кодирование и декодирование устной речи. Акустический речевой сигнал, по мнению Р.К. Потаповой, не делится на сегменты, соответствующие буквам: характеристики отдельных звуков при речепроизводстве «сплюсываются» в пределах слога. Процесс фонетического кодирования приводит к временной компрессии речи: акустические «ключи» к начальным и конечным согласным содержатся в переходах от звука к звуку. При восприятии мозг декодирует акустический сигнал с помощью каких-либо дискретных единиц (слов, слогов и пр.). Еще одно свойство языка — синтаксическое кодирование и декодирование. Синтаксис в данном случае образует сходство с фонетикой: несколько простых предложений глубинной структуры преобразуются в одно сложное на основе трансформационных правил. Наиболее ранний фактор эволюции речи — познавательные (когнитивные) способности. Высказываемое иногда мнение о произвольности набора звуков, используемых в языке, ошибочно. Этот набор ограничен, и в нем имеется базисная часть, которая присутствует почти во всех языках.

Глава вторая, «Специфика речевой артикуляции в процессе речеобразования», содержит анализ слитной речи в коммуникации, описание общего подхода к временному структурированию речевого высказывания, эффекта коартикуляции в речеобразовании. Одной из классических проблем при экспериментальном изучении речи является выбор максимально информативных параметров, подлежащих измерению. Такие параметры должны соответствовать теоретически значимым факторам для нормальной и патологической речи. Они также должны быть воспроизводимы как для одного говорящего, так и для разных. Для того, чтобы акустический сигнал мог быть не только воспринят, но и структурирован при передаче лингвистической информации, необходимо контролировать и координировать работу органов артикуляции. В

процессе речеобразования артикуляторные жесты и позиции разных фонетических сегментов накладываются друг на друга. Степень наложения во многом определяется скоростью процесса говорения. Изменение темпа речи можно использовать для выявления и предсказания особенностей временного структурирования речевого высказывания, что особенно важно для решения различных коммуникативных задач прикладного характера. Другие главы этой части: глава 3 — «Основные аспекты процесса речевосприятия и понимания при коммуникации» и глава 4 — «Соотношение звучащей речи и языковой структуры в акте коммуникации».

Часть II. «Речевая информация» содержит три главы. Глава пятая, «Фонологическая морфологическая информация, содержащаяся в речевом сообщении», посвящена факторам, влияющим на речевой сигнал. Автор указывает, что ингерентная вариативность речевого сигнала может быть обусловлена двумя причинами, зависящими от типов текстов и диктора. Каждый диктор имеет индивидуальный темп, громкость и высоту тона высказывания. Общее эмоциональное состояние диктора также оказывает влияние на речевой сигнал. Под ингерентными факторами, зависящими от текста, понимают фактор коммуникативного типа высказывания, а также факторы смысловой нагрузки, синтаксического членения высказывания и пр. В плане артикуляции минимальным линейно неделимым отрезком речи является слог. В шестой главе, «Синтактико-семантическая и прагматическая информация в речевом высказывании», рассматриваются типы взаимодействий информативных языковых единиц внутри и на стыке синтагм, комплексная языковая информация и прагматические аспекты информационной загруженности в коммуникации. Седьмая глава — «Акустико-фонетическая информация и методы ее извлечения из речевого сигнала».

Часть III — «Речевая кибернетика». Во введении говорится, что имеющиеся в настоящее время способы микросегментации речи (сегментации на субзвуки, звуки, слоги) можно классифицировать следующим образом: 1) использование степени стабильности во времени каких-либо акустических параметров речевого сигнала, таких как концентрация энергии в частотном спектре; 2) наложение акустических меток на речевой сигнал через регулярно повторяющиеся короткие интервалы; 3) сравнение выборки речевого сигнала в коротких

временных окнах при регулярных интервалах с выборками из фонем-прототипов. Различаются контекстно-зависимые и контекстно-независимые методы сегментации. Для решения проблемы сегментации звучащей речи большое значение имеет обращение к слогу.

В главе 8 «Единицы сегментно-звукового уровня в системах автоматического распознавания речи» рассматриваются два подхода к автоматическому распознаванию речи. Ранее полагали, что изучение методов построения простых распознающих систем приведет естественным путем к более сложным системам. В 50-е годы были построены системы с обнадеживающими характеристиками для распознавания гласных и распознавания цифр. Однако эти методы и результаты не могли быть распространены и экстраполированы на более сложные системы. Стало ясно, что для создания стратегии распознавания должны быть привлечены лингвистические и контекстуальные данные. По мнению некоторых исследователей, бесперспективно строить систему, понимающую речь, по принципу сравнения с заложенным в память машины множеством эталонных конструкций. Предлагается использовать механизм «человеческого» типа. Распознающая система в этом случае распознает не отдельные слова и предложения, а извлекает значения из высказываний. Такая система функционирует как система блоков. Базисным блоком является акустико-фонетический процессор, который дает дискретное описание речевой волны; блок грамматического анализатора координирует свою работу с синтаксическим и семантическим компонентами.

Глава 9, «Анализ просодических характеристик в системах автоматического распознавания речи», содержит детальный анализ связи просодических характеристик с семантикой. Наиболее эффективным просодическим параметром является частота основного тона. Она имеет специфические контуры, связанные с прагматическим значением или намерением говорящего. Каждый из просодических параметров извлекается в результате довольно сложного анализа. Используемые источники знаний, в которых применяется просодическая информация, зависят от возможности точного извлечения просодических признаков из речевого сигнала. Сегментация каждого неизвестного речевого высказывания начинается с задания для каждого речевого фрейма вектора признаков. Для каждого класса находится произведение соответствующего весового вектора и вектора

признаков, что дает в результате три скалярные величины, которые и представляют каждый из речевых классов.

В главе 10, «Преобразование текст-речь в автоматизированных системах» рассматриваются вопросы развития систем, синтезирующих речь, акустические свойства фонетических сегментов, используемые в синтезе речи, некоторые подходы к синтезу по правилам, высшие языковые уровни в синтезе речи, преимущества и недостатки различных методов синтеза речи. В современных усовершенствованных моделях голосового тракта большое внимание уделяется включению в модель звеньев, определяющих потери, зависящие от частоты, учету движения стенки полости рта при низких частотах, а также более точному моделированию меняющегося во времени сопротивления нагрузки полости рта. Еще один подход к моделированию вибрации голосовых связок привел к созданию трехмерной структуры, состоящей из множества соединенных между собой масс. Эти модели позволяют получить более сложные виды вибрации и даже имитировать некоторые виды отклонений. Тем не менее, ни в одной из физиологических моделей не найдено удовлетворительного решения, которое позволило бы смоделировать то, что происходит, когда голосовые связки смыкаются на средней линии и несколько деформируются во время поглощения энергии при столкновениях.

В настоящее время нет еще сложных артикуляторных моделей, в которых точные акустические аспекты смогли бы дать естественное звучание.

Относительно моделей синтезаторов речи можно сформулировать четыре основных вывода: 1) современные формантные синтезаторы самых различных конфигураций способны имитировать, и весьма успешно, многие мужские голоса; 2) некоторые упрощения, наблюдаемые в формантных синтезаторах, приводят к неудовлетворительным результатам при имитации гласных звуков высокого тона, произносимых с придыханием, которые часто встречаются рядом с глухими согласными в речи женщин и детей; 3) анализ по методу линейного предсказания через синтез является весьма действенным способом имитации высказываний с высокой точностью воспроизведения, однако его использование для синтеза текста в общем случае является ограниченным; 4) по всей видимости, артикуляторная модель будет окончательным решением проблемы синтеза разборчивой речи с помощью машин, однако вычислительные

затраты и отсутствие данных, необходимых для синтеза по правилам, не дают возможности достаточно широко реализовать эту модель в настоящее время.

Предложение нельзя синтезировать выстраивая линейную последовательность морфем или слов. Для того, чтобы предложение звучало разборчиво и более-менее естественно, необходимо правильно определить длительность звучания, интонацию и варианты аллофонов. Просодическая информация также помогает слушателю разделить акустический поток на слова и фразы.

Просодию можно охарактеризовать в терминах физики через интенсивность, длительность и частоту основного тона. Эти характеристики воспринимаются как громкость, долгота и высота тона. Изменение этих просодических характеристик содержит лингвистически важную информацию.

Устройства преобразования текста в речь применяются в самых разных областях. Синтезирующие речь устройства не настолько хороши, чтобы заменить речь человека, но они могут быть эффективны, если входят в систему, предоставляющую некоторые новые виды услуг, как, например, в телефонии; при неизменной стоимости может работать большее число телефонных линий. В настоящее время в системах, использующих телефонный канал, есть ограничение: компьютер не умеет слышать так же хорошо, как и говорить. Работы в области распознавания речи несколько отстают от работ по синтезу речи, в особенности от синтеза по правилам.

Глава 11, «Новые речевые технологии в автоматизированных системах», содержит описание современного состояния проблемы и формулировку перспективных задач. При достижении определенного уровня компьютеризации современного общества возникает объективное требование создания удобного и естественного интерфейса между человеком и компьютером. Центральной задачей в области разработки речевых информационных технологий является создание систем автоматического распознавания и понимания речи и систем синтеза речи.

Для наиболее мощных современных лабораторных (экспериментальных) систем автоматического распознавания речи характерны следующие параметры:

- 1) Режим работы – распознавание слитно произнесенных фраз;
- 2) Словарь системы – одна тысяча и более слов естественного языка;
- 3) Точность распознавания слов – не ниже 95–96%; предполагается предварительная

настройка алгоритмов системы к голосу диктора по специально подобранным паролем фразам.

На смену доминировавшей в 70-х и начале 80-х годов технике распознавания через сравнение образцов с эталонами с помощью метода динамического программирования пришла техника представления речевого сигнала в виде скрытого марковского случайного процесса. Переход к использованию скрытых марковских моделей позволил более компактно представлять эталонные единицы в памяти компьютера путем задания их вероятностного распределения. Это в свою очередь дало возможность создавать многодикторские системы распознавания, оперирующие словарями свыше 1000 слов. Практически все современные системы распознавания речи, обладающие достаточно мощными характеристиками (объем словаря – 1000 и более слов, многодикторские или обладающие возможностями быстрой адаптации к диктору) выполнены на основе моделирования речевого сигнала скрытым марковским процессом. Эта техника сейчас безусловно доминирует среди остальных подходов. Можно заметить, что данный статистический подход также оказался весьма актуальным и при машинном переводе: в настоящее время ряд систем множественного (при участии нескольких языков) машинного перевода строится с использованием марковской модели статистического процесса.

Использование техники представления речевого сигнала в виде скрытого марковского процесса предполагает выбор базовых единиц (как правило, лингвистических), для которых будут строиться модели. Вследствие большого объема словаря в качестве таких единиц выступают мелкие элементы речевого потока: фонемы или аллофоны, реже – дифоны. Речевые (или фонетические) базы данных также являются новым элементом в речевых технологиях, отсутствовавших в 70-х и 80-х годах. Именно появление техники моделирования речевого сигнала скрытым марковским процессом, которая подразумевает стохастическую природу сигнала, повлекло за собой необходимость создания специальных баз данных. Основное назначение фонетических баз данных – давать информацию о распределении параметров моделей звуковых единиц, необходимую для настройки и подгонки параметров, а также для тестирования (оценки точности) работы систем распознавания.

Когда системе не представляют никакой либо контекстной, семантической, грамматической

тической или синтаксической информации, то существующие методы распознавания речи дают возможность работать со словарем порядка 200 слов. Системы же понимания речи, в которые введены синтаксические и семантические ограничения, могут успешно работать со словарем, включающим уже около 1000 слов. Однако простое увеличение вычислительных возможностей и памяти не может сделать современные методы распознавания пригодными для работы со словарем большого объема (например, свыше 1000 слов).

В «Заключении» говорится, что большие успехи предвещают современные научные исследования в области создания очувствленных роботов с адаптивными свойствами и элементами искусственного интеллекта. Так называемые интеллектуальные роботы обладают мышлением, зрением, слухом, речью, обонянием, осязанием. В настоящее время активно разрабатывается проблема

робота, способного работать во взаимодействии с человеком, сообразуясь с окружающими условиями. Эта проблема неразрывно связана с созданием искусственного интеллекта, способного обучаться на основе собственного опыта и использовать полученные знания в процессе самостоятельной работы при утере связи с оператором. Новые задачи, поставленные перед теорией искусственного интеллекта и робототехникой, прежде всего связаны с необходимостью работать с реальными объектами реального мира, а не только с абстракциями и математическими моделями.

Оценивая рецензируемую работу в целом, можно отметить, что вузовская наука получила новое весьма актуальное пособие в чрезвычайно перспективной области знания.

Ю.Н. Марчук