ЭМПИРИЧЕСКОЕ ПРЕДСКАЗАНИЕ И РАСПОЗНАВАНИЕ ОБРАЗОВ
(Вычислительные системы)
1976 год Выпуск 67

УДК 621.396.275.6:681.3

МЕТОДЫ РАСПОЗНАВАНИЯ И СИНТЕЗА РЕЧИ С ОГРАНИЧЕННЫМ СЛОВАРЕМ

С.В.Голубцов

Решение проблемы автоматического распознавания и синтеза речи, создание устройств, моделирующих процессы восприятия и образования человеческой речи, обеспечат качественно новый подход к организации обмена информацией между человеком и машиной во многих областях автоматического управления и связи. Однако полного решения этой проблемы до настоящего времени еще не найдено. Известны методы и технические решения ряда частных аспектов указанной проблемы, наиболее важным из которых является распознавание и синтез речи с ограниченным словарем. В то же время изучение условий речевого общения между человеком и машиной показывает, что, как правило, необходима строгая формализация информации, поскольку это позволяет уменьшить вероятность ошибки оператора и повысить скорость ввода данных.

Изложенные обстоятельства привели к широкому развитию исследований методов распознавания и синтеза речи с ограниченным словарем как в нашей стране, так и за рубежом. В настоящей работе предпринята попытка отразить некоторые характерные и перспективные, с точки зрения автора, направления развития этих исследований. Работу не следует рассматривать как обзор, поскольку упоминание о тех или иных исследованиях имеет цель скорее проиллюстрировать имеющиеся тенденции развития, а не охватить все существующие направления, число которых растет год от года. Для более полного ознакомления можно рекомендовать работы [3,5,6,9,14,16,29], содержащие аналитические обзоры по ряду направлений.

Комплекс устройств, обеспечивающих ввод и вывод информации в форме устной речи, в дальнейшем изложении мы будем называть подсистемой речевой связи.

В соответствии с двумя основными функциями подсистемы речевой связи (приемом и выдачей информации в речевой форме) ниже будут рассмотрены два направления исследований – распознавание и синтез речи с ограниченным словарем.

Распознавание ограниченного набора речевых команд. В настоящее время можно считать общепринятым использование иерархического принципа распознавания речи. С одной стороны, это обусловлено тем, что на иерархическом принципе функционирует система речевосприятия человека [10]. С другой стороны, применение его для распознавания речи приводит к более простым и гибким техническим решениям. Различные авторы выделяют разное число ступеней иерархии или этапов распознавания. Например, в [10] выделяется три основных этапа, а в [72] таких этапов насчитывается пять:

а) выделение первичных параметров речи;

б) детектирование микросегментов;

в) преобразование последовательности микросегмента в фонемы;

г) распознавание на уровне слов;

д) семантическая интерпретация и коррекция результатов распознавания.

Для реализации алгоритма распознавания в техническом устройстве последняя классификация представляется наиболее удобной и полной. В той или иной форме она проявляется в большинстве работ, хотя в зависимости от целей автора отдельные этапы могут отсутствовать. Наиболее четко и последовательно перечисленные этапы анализа проявляются в работах [5,29,32,36, 42, 47]. Рассмотрим кратко особенности анализа на отдельных этапах.

Первичное описание речевого сигнала имеет задачу сократить объем анализируемой информации, поскольку непосредственный анализ речевого сигнала, преобразованного в цифровую форму, требует переработки значительных объемов информации (42 кбит/сек согласно [2]). В то же время объем полезной информации, определяемый фонетическим составом речи, равен всего 120 бит/сек. Задача сокращения объема речевого сигнала при

сохранении полезной информации решается в системах компрессии речи – так называемых вокодерах, различающихся по способу описания речевого сигнала. К настоящему времени разработано большое число способов описания, используемых в полосном, формантном, гармоническом вокодерах и их разновидностях [1,2,3]. Действие всех перечисленных систем основано на сохранении в описании формы огибающей текущего энергетического спектра речи, измеряемого с частотой квантования 25–100 гц, что позволяет сократить объем информации до (1–5 кбит/сек).

Несмотря на успехи вокодерной техники, в значительной части работ по распознаванию ограниченного словаря опыт построения вокодерных систем используется слабо. Например, очень часто в работах по распознаванию применяется описание на основе подсчета числа переходов речевой волной нулевого значения, усредненное за время порядка 10 мсек [5,8,17,18,31,38,42–45,63]. В то же время в вокодерах от подобного описания давно отказались как от малоинформативного. Это чувствуют и специалисты по распознаванию, пытаясь усовершенствовать описание, например, путем использования разложения предельно-ограниченной речевой волны по функциям Уолша [52].

Другим видом описания, часто используемым в работах по распознаванию, является полосный анализ спектра речи. При этом виде анализа большое значение играет подробность расфильтровки. По опыту разработки вокодеров для обеспечения речи удовлетворительной разборчивости необходимо анализировать сигнал в полосе частот порядка 5–7 кгц с использованием не менее 10–15 каналов.

Многие авторы, надеясь на уменьшение требований к объему полезной информации благодаря ограниченному словарю, удовлетворяются значительно меньшим количеством каналов – до 4–5 [7,26, 29,33,34,36,37,58] . Это приводит к значительным потерям информации, которые на последующих этапах анализа могут быть скомпенсированы лишь отчасти.

С другой стороны, чрезмерное увеличение подробности расфильтровки также вредно, поскольку это бесполезно загружает узлы последующего анализа, не давая прироста достоверности распознавания. С целью сравнения различных систем параметров с точки зрения их эффективности при распознавании в работе [53] приводятся результаты распознавания 10 однозначных цифр, произнесенных по-японски 100 раз одним диктором-мужчиной. В качестве первичного описания использовалось:

а) полосный анализ спектра с расфильтровкой на 96, 49 и 25 каналов;

б) автокорреляционная функция речевой волны;

в) кепстральное разложение речевого сигнала;

г) коэффициент линейного предсказания мгновенных значений речевой волны.

При распознавании применялся алгоритм распознавания, близкий к использованному в работе [33]. Сравнение показало, что наилучшие результаты (достоверность распознавания близка к 100%) получились при использовании полосного и кепстрального описаний. При этом различие в точности для расфильтровки на 96 и 25 каналов оказалось незначительным.

Для стабилизации результатов распознавания при выделении первичных признаков в ряде работ принимаются меры, направленные на ослабление влияния таких дестабилизирующих факторов, как изменение громкости речи, смены диктора, помещения, где находится микрофон, и др.

Одним из преимуществ использования признаков, основанных на клиппированной речи, является меньшая зависимость их от уровня благодаря предварительному предельному ограничению сигнала. В устройствах, где используются фильтровые и другие способы анализа, появляется необходимость применения усилителя с автоматической регулировкой усиления [73] или специальных мер по фиксации положения микрофона относительно рта говорящего. Для стабилизации результатов распознавания относительно меняющегося уровня речи могут также приниматься меры при построении решающих правил на последующих уровнях (например, использование решающих функций, не зависящих от уровня).

Уменьшение влияния индивидуальных особенностей дикторов, помещения и микрофона на этапе выделения признаков достигается автоматической подстройкой частотной характеристики [73], а также использованием информации о частоте основного тона диктора. В простейшем случае изменение частоты основного тона служит признаком для смены спектральных эталонов при распознавании [44]. Однако наиболее перспективным, хотя и достаточно сложным представляется использование спектрального анализатора, синхронизированного с частотой основного тона.

Для реализации устройства выделения признаков используются два способа. В ряде работ [7,26,33,34,35,37,7I] речевой сигнал, полученный с микрофона, сразу же преобразуется в цифровую форму и выделение признаков осуществляется по программе в ЭЦВМ. Это позволяет исключить из устройства аналоговые схемы, но значительно загружает ЭЦВМ, поскольку, как отмечалось, скорость поступления речевой информации составляет несколько десятков килобит в сек. Для сокращения времени анализа обычно применяют алгоритм быстрого преобразования Фурье, однако даже в этом случае не удается проводить анализ в реальном времени,а большие затраты машинного времени ограничивают точность анализа 4-5 полосами частот. Именно этим объясняется отмеченная выше недостаточная точность спектрального анализа в ряде работ.

Все это привело к широкому использованию на этапе выделения признаков аналоговых анализаторов спектра, согласованных с ЭЦВМ через аналого-цифровой преобразователь. По-видимому, при современном состоянии вычислительной техники это оптимальное решение [3I,45]. Переход на полностью цифровую обработку речевого сигнала будет целесообразен после обработки технических решений цифрового моделирования вокодерных систем.Сейчас,по сути дела, решение этой задачи еще только начинается [2,II].

Следующим этапом распознавания является классификация отрезков речи на микросегменты (субфонемные последовательности). Речь представляет из себя непрерывный во времени поток звуков (реализаций фонем), связанных между собой переходными участками. В этом процессе могут быть с определенной долей условности выделены квазистационарные части звуков, обладающие более или менее определенным спектром, и переходные участки,основным признаком которых является характер изменения спектра во времени. Этот же признак является определяющим и для так называемых динамичных звуков речи (взрывные, дрожащие звуки). Микросегменты отражают указанные особенности отдельных участков речи, в связи с чем на этом этапе анализа большинство звуков речи выражено еще в неявной форме.

Состав алфавита для классификации микросегментов в различных работах значительно варьируется в зависимости от объема словаря, возможностей использованной системы признаков и процедуры распознавания. В этой связи интересно заметить, что даже сравнительно небольшой по объему алфавит микросегментов позволяет обеспечить распознавание довольно больших словарей. В работе [29] показано, что если алфавит микросегментов состоит всего из 7 элементов, представляющих собой основные группы фонем (гласные и сонорные, дрожащие, глухие взрывные, звонкие взрывные, глухие фрикативные, звонкие фрикативные, аффрикаты), то возможно разделение до 300 слов при очень незначительном (в 5-I0% случаев) предварительном подборе слов по их фонетическому составу при составлении словаря.

В ряде работ при выборе алфавита микросегментов не производится четкой привязки их к фонемному составу речи. Например, в работах [5,8,I7,I8,3I] алфавит сегментов представляет собой набор дифференциальных признаков, таких как: "звонкость", "глухость", "гулкость", "шумность", "гласность", "назальность" и т.д. В то же время большинство исследователей предпочитают придерживаться фонемного (или фонемно-группового) алфавита уже на этапе микросегментов [I5,24,26,27,29,33-37,40,4I,73].

Поскольку длительность микросегментов обычно выбирается постоянной (I0-20 мсек) и не связанной с временным положением границ между звуками в слове, задача следующего - третьего - этапа анализа состоит в преобразовании последовательности микросегментов в последовательность фонем и их вариантов.Заметим, что этот этап необходим и целесообразен только в случае распознавания большого, в пределе неограниченного, словаря.В связи с этим в большинстве работ он не выражен.Преобразование последовательности микросегментов в фонемный (фонемно-групповой, фонемно-вариантный) алфавит представляет собой в общем случае преобразование временной последовательности микросегментов по определенным логическим правилам. Правила преобразований основаны на том,что информация о фонемах в речевом сигнале рассредоточена по времени и для принятия фонемных решений необходимо учитывать не только структуру сигнала в текущий момент, но и предысторию, включающую в себя информацию о переходных участках к соседним звукам [5].

Четвертый этап - распознавание на уровне слов - является обязательным практически во всех работах.

Задача распознавания слов по последовательности распознанных фонем могла бы быть сведена к тривиальному сравнению фонетического состава произнесенного слова с фонетической транскрипцией слов словаря. Однако этому препятствует следующее:

а) в настоящее время отсутствует алгоритм распознавания фонем с достоверностью, обеспечивающей однозначное сопоставление распознанной и эталонной фонемных последовательностей;

б) дикторы произносят одно и то же слово неодинаково, что приводит к различиям в их фонетическом составе. Причины этого связаны с различной степенью редукции, диалектными особенностями и др.

Все это приводит к тому, что однозначное сопоставление распознанных и эталонных фонемных последовательностей невозможно. Именно поэтому в большинстве работ этап распознавания фонем не моделируется, а распознавание слов производится непосредственно по последовательности микросегментов. Однако здесь возникают свои трудности, связанные с отсутствием четких границ между звуками в распознаваемом слове и большой вариативностью последовательностей. Особенности произнесения приводят к тому, что для одного и того же слова может меняться количество звуков, их длительность, а также длительность слова в целом. Это вызывает рассогласование во времени между распознаваемой и эталонной последовательностями микросегментов. Для устранения рассогласования в алгоритм обычно вводится так называемое временное нормирование, для выполнения которого необходимо с высокой надежностью выделять границы отдельных слов. Последнее требование приводит к тому, что допускается произнесение только изолированных слов, разделенных паузами, поскольку надежно работающих алгоритмов членения непрерывной речи на слова неизвестно.

В настоящее время в отдельных работах делаются лишь первые попытки в этом направлении [21,28,44,55,62].

В случае, когда слова произносятся изолированно, положение границ слова во времени определить сравнительно несложно по появлению сигнала, превышающего уровень отсечки паузы. В настоящее время для выравнивания длительности анализируемого слова и его элементов с эталонными известно несколько алгоритмов.

Наиболее простым является алгоритм линейного временного нормирования [15, 51, 65]. Он предусматривает пропорциональное изменение длительности всех элементов нормируемого слова до

тех пор, пока длительность всего слова не станет равной длительности эталона. Недостаток этого метода связан с тем, что в естественной речи при изменении длительности слова длительности его отдельных элементов меняются по-разному. Более всего подвержены изменению длительности стационарных частей нединамичных звуков, менее – длительности переходных участков и элементы динамичных. Вследствие этого после пропорционального изменения длительности элементов их границы зачастую не совпадают с эталонными и сравнение происходит с ошибкой.

Для устранения рассогласования в некоторых работах предусмотрено сравнение состава последовательности микросегментов без учета длительностей сравниваемых участков [27,36]. Это позволяет сделать алгоритм нечувствительным к изменению длительности квазистационарных участков нединамичных звуков. В то же время такой метод приводит к потере некоторой доли полезной информации, так как длительность сегмента содержит признаки ударности гласных, позволяет отделять фрикативные от взрывных и др.

Наиболее полное решение задачи дает так называемое нелинейное временное нормирование [2,7,33–35,53]. Смысл его сводится к тому, что определяются характерные временные точки внутри распознаваемого слова и затем производится нелинейная трансформация временной оси, обеспечивающая совпадение этих точек у реализации и эталона. Нелинейное нормирование дает хорошие результаты, однако требует значительных машинных затрат. Поэтому иногда применяют вначале линейное нормирование (для всех слов словаря с разделением их на подгруппы), а затем нелинейное нормирование для разделения слов внутри подгруппы [33].

Для привязки временных участков при нелинейном нормировании требуется тем или иным способом внутри слова определить местоположение характерных точек. Такими точками могут быть переходы "пауза/сигнал", "тон/шум" и некоторые другие. Наиболее полным решением является использование в качестве характерных точек сигналов сегментации, т.е. сигналов, определяющих временное положение границ между звуками в слове.

Таким образом, возникает задача автономной сегментации речи на звуки. Решение этой задачи важно для распознавания как ограниченного, так и неограниченного словаря. Исследование методов сегментации проводится в ряде работ [29,46,47,67]. В качестве признака границы между звуками обычно используется про-

явление динамики тех или иных характеристик речи. В частности, в работе [29] применен следующий критерий:

$$\max_{t} \sum_{i=1}^{n} |x_i(t) - x_i(t+\tau)|,$$

где $x_i(t), x_i(t+\tau)$ – отсчеты уровней в каналах n – полосного анализатора спектра в два момента времени, отстоящие друг от друга на время τ; τ = 10–40 мксек.

При сегментации возможно появление двух родов ошибок – пропуск границы и появление дополнительных границ (ложная тревога). Менее желательным является пропуск границы. Надежность выделения границы составляет обычно 85–95%. При этом вероятность ложной тревоги в некоторых работах доходит до 100% и более, т.е. число ложных границ вдвое превышает истинное [29].

Поскольку в большинстве работ распознавание ограниченного набора слов является конечной целью исследования, интересно сопоставить достоверность распознавания, получаемую в различных режимах анализа. Такое сопоставление проведено в рaоте [14]. Из него следует, что, несмотря на различие алгоритмов и условий эксперимента, можно определить зависимость получаемой достоверности распознавания от объема словаря N.

Полученные в работе кривые приведены на рис.1. Кривая 1 соответствует режиму распознавания, предусматривающему подстройку под голос говорящего, а кривая 2 иллюстрирует случай работы без подстройки. Из кривых следует, что с ростом словаря достоверность распознавания слов η (при $N \geq 10$) падает для случая подстройки под голос диктора на 1–1,5%, а для случая без подстройки – на 5–6% при удвоении объема словаря. Значение достоверности η = 90%-кривая, полученная в режиме без подстройки под голос диктора, достигает в области N = 60 слов, а при подстройке под голос такое значение будет, по-видимому, при $N \approx 500$. Введение подстройки под голос диктора дает существенный выигрыш достоверности, приводя в то же время к усложнению распознающего устройства и необходимости обучения перед распознаванием. Поэтому достаточно часто встречаются оба подхода. В работах [5,8,15,17,29,38,42–44,65] распознающее устройство настраивается на некоторого "среднего" диктора и работает на различных голосах по единому комплексу эталонов. Понижение достоверности при этом авторы часто стараются скомпенсировать мерами на последующих ступенях распознавания. В работах [7,24, 33–

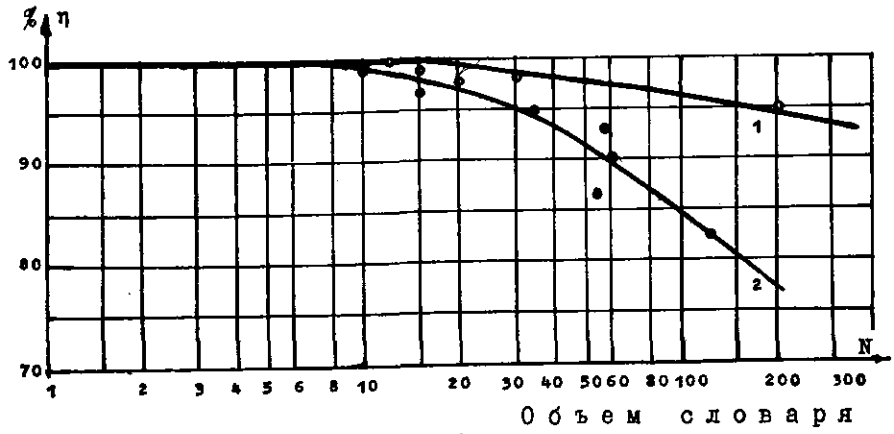

Рис. 1

36,39,40,51,60,64] исследование ведется или с использованием голоса единственного диктора, или с подстройкой под голос при работе на различных голосах. Полученные результаты подтверждают сделанный выше вывод об эффективности адаптации. Например, в работе [60] эталоны японских цифр 0–9 вырабатываются путем нескольких произнесений их диктором. Достоверность распознавания при использовании "своих" эталонов на 30 мужских голосах составила 99,8%, 30 женских – 98,6%, 7 детских – 97,6%.

Интересно отметить, что в некоторых работах [51, 60, 64] процесс адаптивного получения эталонов автоматизирован и входит в программу работы распознающего устройства. Благодаря этому удается до минимума сократить трудоемкость и время обучения.

Появилось несколько работ, в которых задача адаптации поставлена наоборот – не строить адаптирующийся под диктора автомат, а заставить диктора подстраиваться под его особенности. Были поставлены специальные опыты, в которых диктору обеспечивался оперативный контроль за результатами распознавания его речи и разрешалось повторное произнесение неправильно распознанных слов для исправления ошибки изменением манеры произношения. Полученные предварительные результаты пока разноречивы: в работе [56] делается вывод, что диктор почти не способен адаптироваться под автомат, а в [63] показано, что вероятность

правильных ответов заметно растет по мере тренировки. Следует отметить, что в работах использованы распознающие устройства невысокой эффективности, что, возможно, повлияло на результат.

Для большинства практических случаев вводимая информация составляет не отдельные слова, а целые фразы, иногда весьма большой продолжительности. Поэтому последний, пятый, этап анализа из числа перечисленных выше является необходимым. Распознавание фразы, в принципе, может быть представлено как пословное сравнение последовательности слов, ее составляющих с последовательностью, распознанной на предшествующих этапах анализа. Вероятность правильного распознавания фразы η_Φ в целом при этом будет определяться величиной $\eta_\Phi = \eta^M$, где η – средняя достоверность распознавания слов словаря, а M – количество слов во фразе. Поскольку $\eta \leq 1$, $\eta_\Phi \leq \eta$, т.е. при таком способе неизбежна потеря достоверности распознавания по сравнению с предшествующими этапами.

Для повышения достоверности распознавания фраз привлекаются грамматические и семантические связи, позволяющие выявлять и устранять ошибки, допущенные ранее, и таким способом не только не понижать, но и существенно повышать достоверность распознавания вводимых команд. При использовании семантических связей важную роль начинает играть совокупность понятий, с которыми может оперировать диктор при общении с автоматом.

Следует подчеркнуть, что на уровне распознавания фраз уже нельзя ограничиваться только вопросами ввода речевой информации, необходимо ввод и вывод рассматривать в комплексе, поскольку наиболее эффективна организация общения оператора и машины в форме диалога. Наиболее сложно организовать такое общение, не привязываясь к определенной теме или проблеме [23]. Первые опыты в этом направлении (без устного ввода/вывода речи) описаны в [22]. Авторами разработаны две программы, позволяющие вести диалог с ЭВМ на медицинские темы (программа "Доктор") и на вольную тематику (программа "Элиза"). Отмечается, что для организации работы программ должна быть заранее выработана схема-сценарий, построенная в форме ветвящегося графа-дерева. В начале графа стоят общие, широкие темы, а по мере продвижения по ветвям дерева затрагиваются частные вопросы. В случае, если эти вопросы в ходе беседы исчерпаны, возможен возврат к одной из общих тем и повторное прохождение дерева по другому пути.

Схема [22] носит в значительной мере демонстрационный характер. Для практического применения более рационально использование проблемно-ориентированных языков, организуемых с помощью порождающих грамматик. В частности, в [43] рассмотрена грамматика, образуемая упорядоченной четверкой $\{ V_T, V_N, S, P \}$, состоящей из V_T – основного тематического словаря; V_N – вспомогательного словаря; S – начально-конечных символов; P – конечного множества правил грамматики вида $A \to B$.

Применение этого принципа для организации общения оператора и ЭВМ [8,17,32,42-44,69] позволяет разбивать весь словарь на ряд групп слов (подсловарей), определяемых структурой вводимой информации. Это в несколько раз сокращает число распознаваемых в данный момент слов, поскольку большая часть словаря при заданном сценарии на отдельных этапах диалога невозможна. Последнее позволяет, в свою очередь, значительно повысить достоверность распознавания путем запрета тех или иных распознаваемых слов. Например, в [44] при общем словаре 57 слов такой метод позволил на каждом этапе диалога с ЭВМ БЭСМ-6 распознавать сменный словарь объемом от I до 20 слов. Благодаря этому при распознавании I848 фраз, произнесенных 5 мужскими и 2 женскими голосами, встретилось всего 0,5% ошибок и 2,3% ответов "не знаю", в то время как средняя достоверность распознавания всех слов составляет около 90%.

Кроме семантической информации, для повышения эффективности распознавания в ряде работ привлекаются правила грамматики [20,32,69]. В работе [20] рассматривается задача распознавания фраз достаточно разнообразного содержания (ограниченного составом словаря), но со стандартным построением: определение-подлежащее-сказуемое-дополнение. Кроме того, весь словарь разбивается на ряд классов:

а) основной субъект;

б) дополнительные предметы и субъекты (выражение принадлежности или качества);

в) действие;

г) состояние.

На предшествующих этапах распознавания авторами используется алгоритм, не позволяющий однозначно определять слово. Это делается в конце анализа путем определения соответствия слова

и его функции в предложении с проверкой допустимости этого слова с точки зрения разрешенных семантических связей.

При анализе речи на уровне фраз большую роль начинают играть суперсегментные, просодические параметры такие, как частота основного тона, темп речи, изменение среднего уровня. Совокупность этих параметров обусловливает появление различных видов интонации (повествование, вопрос, приказ и т.д.), а также словесного и логического ударений. Естественно, что просодические характеристики высказывания могут быть использованы при распознавании фраз. С другой стороны, просодика наиболее информативна в слитной речи, когда слова не разделяются искусственно паузами. В работе [21] предлагается перед распознаванием членить слитную речь на слова по положению словесного ударения, которое определяется распределением относительных длительностей звуков. В [54] по комплексу просодических параметров перед распознаванием производится классификация гласных по схеме:

Затем полученные данные используются для членения фраз на слова и их распознавание. В работе [68] предлагается построение двух параллельных систем: распознавание слов по сегментным признакам и распознавание синтаксической структуры фраз по суперсегментным признакам. На уровне распознавания фраз обе системы должны взаимодействовать.

Первые результаты экспериментальной проверки метода распознавания слитной речи с ограниченным проблемно-ориентированным словарем описаны в [44]. Для членения фраз используются ключевые слова как включаемые специально в текст, так и входящие в него органически (например, цифры, названия и др.). Тематически опыты были привязаны к задаче автоматизации диспетчерской службы аэрофлота. Проверка алгоритма распознавания на двух коротких фразах: "Слушай, прошел Серпухов" и "Слушай, время 15.23" на 10 голосах (около 750 реализаций) показало, что достоверность превышает 95%.

Важной стороной работ по распознаванию речи является изучение эффективности речевого ввода в различных областях его применения. Дело в том, что в ряде случаев целесообразность речевого ввода оказывается неочевидной [14,74], поскольку для ввода информации существуют и другие, широко используемые и более простые в техническом отношении способы, например клавишные устройства и др.

С другой стороны, по данным инженерной психологии, для человека-оператора речь является одной из наиболее высокоскоростных и удобных форм обмена информацией. Поэтому для выяснения положительных и отрицательных сторон речевого ввода/вывода информации необходимо накопление опыта эксплуатации подобных устройств.

По данным работ [5,14,64,75], устройства для распознавания ограниченного словаря могут быть эффективно использованы при общении оператора и ЭВМ, в системах информационно-справочной службы, в области телефонной связи.

Для оценки эффективности речевой связи с ЭВМ в США была разработана специальная программа исследований "SPECOMICOM" [75].

Перспективность использования устройств речевого ввода обусловлена следующими его преимуществами:

а) речь слышна в темноте без прямой видимости объекта управления;

б) речь обходит препятствия;

в) связь с объектом осуществляется без физического контакта оператора с ним, освобождая ему руки и зрение;

г) речь—наиболее удобный и естественный способ общения, ее использование наряду с другими видами ввода информации повышает пропускную способность оператора.

Однако до настоящего времени достаточно обоснованных доказательств преимуществ и недостатков речевого ввода перед другими способами ввода еще не получено.

Вывод информации в речевой форме. Для вывода информации в форме устной речи в настоящее время известно несколько методов, отличающихся как устройством аппаратуры, синтезирующей речь, так и способом управления ею. Общим

для задачи речевого вывода является наличие двух основных узлов - узла, воспроизводящего речь по данным, введенным заранее в его память, и узла управления, определяющего последовательность вывода речевых элементов в соответствии с содержанием высказывания.

Способы речевого вывода по способу хранения данных о речи разделяются на аналоговый и дискретный, а по способу управления - на синтез методом компиляции и синтез по правилам. Синтез методом компиляции предполагает создание речи путем воспроизведения речевых элементов, записанных заранее с голоса диктора последовательно во времени в соответствии с содержанием высказывания, зачастую без специальных мер по стыковке соседних элементов. Синтез по правилам предусматривает создание речевого сигнала из элементов, которые генерируются по специальным правилам в синтезаторе. Правила содержат сведения о спектрально-временном составе речевых элементов и способе их сочленения в слитную речь.

Для управления синтезом по методу компиляции достаточно знать последовательность речевых элементов, составляющих речь, и их адреса в памяти ЭВМ. При управлении синтезатором по правилам в памяти устройства необходимо иметь комплекс этих правил для всех элементов речи и состав последовательности, соответствующей требуемому содержанию речевого ответа.

В настоящее время речевой вывод начинает все шире использоваться практически. По данным различных публикаций, в 1973 г. за рубежом работало в промышленности около 500 таких устройств, а к 1975 г. ожидается увеличение их числа до 10 тысяч. При этом большая часть существующих устройств (около 80%) построена на аналоговом принципе.

Аналоговый принцип построения устройств речевого вывода предусматривает предварительную запись мгновенных значений речевой волны на тот или иной носитель с последующим воспроизведением их в заданной последовательности. В качестве носителя используются магнитные ленты, диски, барабаны, а также фотозапись, имеющая преимущества по сроку службы. Речевой сигнал обычно записывается в аналоговом виде, однако в некоторых случаях он преобразуется в дискретную форму [57]. Этот случай мы будем относить также к аналоговому, поскольку ему присущи все основные особенности этого принципа.

Преимущества аналогового принципа состоят в следующем:

а) речь претерпевает минимальные преобразования, благодаря чему обеспечивается высокая естественность и разборчивость воспроизводимых слов;

б) аппаратура для записи, воспроизведения и хранения речевой информации достаточно проста.

Недостатки аналогового принципа построения узла речевого ответа:

а) речевые элементы практически не могут быть меньше слова, поскольку речь, составленная, например, из предварительно записанных звуков, звучит неразборчиво вследствие отсутствия переходов между ними. Организация простых стандартных переходов между звуками дает недостаточный эффект-сигнал получается низкого качества (разборчивость слов W = 75% [25], что соответствует речи, не приемлемой для связи [4]). Из этого следует, что по аналоговому принципу не может быть реализован речевой вывод с неограниченным словарем, поскольку для этого необходимо запасти в памяти устройства около 130 тыс. слов, а с учетом требуемых словоформ это число возрастает до 500-700 тыс. [9], что практически нереализуемо;

б) вследствие того, что речевой сигнал не подвергается почти никаким преобразованиям, в него очень трудно ввести какие-либо изменения, например, управление интонацией, создание логического ударения и др.;

в) объем информации для запоминания речевого сигнала относительно велик (порядка 20 кбит/слово).

Перечисленные особенности позволяют с помощью аналогового принципа синтезировать речь только методом компиляции.

Дискретный принцип синтеза речи предусматривает предварительное преобразование речевого сигнала в дискретную форму с одновременной компрессией объема информации, что позволяет более эффективно загрузить память устройства речевого вывода. В процессе вывода речевой информации производится восстановление объема речевого сигнала и преобразование его в аналоговую форму, пригодную для восприятия речи человеком. При сокращении объема речевого сигнала анализируется его спектр, а при воспроизведении речь создается путем управления параметрами синтезатора аналогично тому, как это делается в параметрических вокодерах.

Поскольку дискретный принцип предусматривает запоминание не мгновенных значений речевого сигнала, а параметров, при его реализации открываются широкие возможности для изменения характеристик речи. Это имеет большое значение в первую очередь при создании плавных переходов между речевыми элементами, что, в свою очередь, позволяет использовать в качестве последних не только слова и фразы, но и более мелкие элементы, вплоть до фонем. Использование фонем в качестве речевых элементов при синтезе речи крайне привлекательно, так как позволяет синтезировать речь с ограниченным словарем при очень небольшом (порядка 50) алфавите речевых элементов. Кроме того, при использовании дискретных принципов синтеза упрощается вопрос имитации различных суперсегментных явлений (интонации, ударения и др.).

Управление синтезом при дискретной реализации речевого вывода информации может осуществляться как методом компиляции, так и по правилам. Вне зависимости от способа управления дискретный принцип речевого вывода требует применения специфического узла – синтезатора речи. В принципе, для этих целей может быть использован любой тип параметрического синтезатора из числа используемых в параметрических вокодерах [1,2,3].

Дискретный синтез методом компиляции наиболее удобно выполнять на основе технических решений, применяемых в вокодерной технике, например, с использованием полосных гармонических или формантных анализаторов и синтезаторов речи. Подобное устройство предусматривает анализ речевого сигнала и преобразование получаемых функционалов – параметров в дискретную форму в аналого-цифровом преобразователе (АЦП). Перед началом вывода информации в речевой форме диктор читает слова (или фразы), входящие в словарь. Параметры речи записываются в память ЭВМ, при этом каждому слову присваивается свой адрес.

При работе в режиме речевого вывода информации в соответствии с требуемым содержанием по этим адресам из памяти ЭВМ извлекаются значения параметров нужных слов и после необходимой стыковки подаются на управление синтезатором соответствующего типа через преобразователь цифра–аналог. Синтезатор преобразует последовательность значений параметров в слитную речь.

Дискретный синтез по правилам, в отличие от метода компиляции, не требует предварительного заполнения памяти значениями параметров естественной речи с помощью анализатора. Этот метод синтеза основан на использовании в качестве исходных достаточно мелких речевых единиц, например фонем. В память ЭВМ заносится комплекс правил, определяющих спектр каждого такого элемента, и способ сочленения его с другими элементами при синтезе слитной речи (например, длительности переходных участков в спектре сочетающихся звуков [16]). При осуществлении синтеза задается последовательность элементов, из которых состоит речь, и ЭВМ вычисляет значения параметров в соответствии с заложенными в ее память правилами. После преобразования в цифро–аналоговом преобразователе параметры управляют синтезатором одного из перечисленных выше типов.

Использование метода синтеза по правилам допускает применение синтезаторов, существенно отличающихся от используемых в вокодерах. Это, в свою очередь, позволяет повысить качество речи, устранив некоторые искажения, характерные для параметрических вокодеров, и упростить управление синтезатором.

В этой связи следует в первую очередь отметить синтезаторы, полностью или частично моделирующие артикуляционный аппарат человека (так называемый электрический аналог речевого тракта [9,59,80]). Основные особенности их состоят в следующем:

а) формирование речевого спектра производится в последовательно включенных частотно–избирательных управляемых цепях. Это позволяет избавиться от интерференционных искажений, присущих параллельным схемам синтезаторов (например, полосным);

б) формирование сигналов управления осуществляется на основе задания положения и характера взаимодействия отдельных органов речевого аппарата человека (через так называемые "цели движения" [9,59]) и последующего пересчета в сигналы, управляющие синтезатором. Такой метод позволяет хранить в памяти не все спектрально–временное описание фонем и переходов, а траектории изменения положения артикулирующих органов вычислять на основе данных об их динамических характеристиках.

Наибольшее приближение к естественному прообразу достигается при использовании пневмомеханических синтезаторов речи [1]. Эти синтезаторы предусматривают моделирование речевого тракта в механической модели, что дает преимущества в части качества речи при весьма простых технических решениях.

Поскольку при выводе речевой информации необходимо обычно составлять более или менее протяженные фразы, для обеспечения

естественности и выразительности речи большое значение приобретают суперсегментные, просодические характеристики, которые должны быть определены одновременно в процессе синтеза и наложены на получаемый речевой сигнал. В последние годы в этом направлении ведутся интенсивные исследования и получены первые положительные результаты [9,48-50,6I,76-78].

Задача в большинстве работ рассматривается в аспекте автоматического преобразования печатного текста в звучащую речь. С точки зрения проблематики речевого общения человека и машины, кроме этапов считывания печатных знаков и преобразования их в фонетическую транскрипцию, все вопросы, рассматриваемые в этих работах, актуальны. В частности, большое значение просодики выводимых фраз имеют формализованные правила для

а) имитации основных видов интонации с помощью типовых контуров изменения основного тона, энергии, длительности элементов (в том числе и пауз);

б) имитации словесного и логического ударений.

При формировании правил просодики слова во фразе разбиваются на содержательные и служебные ; выводимая информация оформляется в синтагмы, разделяемые паузами. Тип интонации содержится в выводимой информации. Текущие значения основного тона и энергии задаются в виде линейно-ломаной линии (путем линейной интерполяции характерных отсчетов просодических параметров).

Эффективность работы различных синтезаторов можно сопоставить по таким характеристикам, как качество речи (в частности, ее естественность и разборчивость), объем памяти ЭВМ на I слово и связанный с этим объем словаря.

Для оценки разборчивости речи в настоящее время в СССР разработана достаточно строгая методика, определяемая ГОСТ-7I53-68 и состоящая в оценке относительно числа S правильно принятых бессмысленных слогов бригадой слушателей (аудиторов). За рубежом также используются подобные оценки, однако сопоставление результатов испытаний можно делать только приблизительно. В ряде случаев как в отечественных, так и зарубежных работах оценивается разборчивость не слогов, а слов W.

Для оценки естественности известен метод мнений [4], состоящий в определении числа высказываний экспертов-аудиторов в

пользу той или иной записи речи из числа предъявленных на экспертизу. В число предъявляемых должна входить "эталонная" речь, качество которой известно заранее.

Разборчивость речи, получаемой методом компиляции, определяется, в основном, теми преобразованиями, которым она была подвергнута в процессе записи ее в память ЭВМ и обратного воспроизведения. В частности, для аналоговых методов компиляции могут быть получены сигналы, практически ничем не отличающиеся от исходной естественной речи.

Для дискретных методов компиляции разборчивость восстановленной речи определяется качеством вокодерных преобразований и в общем соответствует разборчивости вокодерной речи, имеющей для хорошо отработанных систем значения S = 85-90% [2,3], что согласно [4] соответствует классу отличного качества и словесной разборчивости W = 98-99%.

Естественность компилированной речи определяется рядом факторов:

а) искажениями, вносимыми системой преобразования речевых сигналов;

б) качеством стыковки речевых элементов в процессе синтеза;

в) степенью отработки просодических характеристик, накладываемых на речь в процессе синтеза.

Как общую тенденцию можно заметить, что качество речи тем лучше, чем из более крупных единиц она составляется [9].

Речь, синтезированная по правилам, к сожалению, обычно не испытывается на разборчивость с той же строгостью, как вокодерная речь. Из работы [77] известно, что разборчивость односложных английских слов, синтезированных по правилам, составляет 80%, а фраз - 90%.

Отечественный метод синтеза речи по правилам, описанный в [I6], обеспечивает словесную разборчивость W = (96,5±I)% при разборчивости слогов S = (7I,4±I,3)%, что соответствует классу речи хорошего качества.

Естественность речи, синтезированной по правилам, обычно ниже естественности компилированной речи и зависит от тщательности отработки правил синтеза, степени их формализации и особенностей применяемого синтезатора.

Сопоставим различные методы синтеза по объему памяти, необходимому для воспроизведения одного слова. В целях удобства сопоставления как для аналогового, так и для дискретного методов оценку будем вести в пересчете на объем бинарного речевого сигнала. При подсчете будем предполагать, что средняя длительность синтезируемых слов составляет 0,5 сек.

Аналоговый метод требует запоминания мгновенных значений речевого сигнала, что требует, как отмечалось, затрат около 40 кбит/сек или 20 кбит в пересчете на I слово. Кроме запоминания мгновенных значений речи, метод компиляции на уровне слов, обычно используемой здесь, требует некоторых затрат памяти на программу, обеспечивающую выбор слов в заданной последовательности из запасенного словаря. Сугубо ориентировочно для достаточно большого словаря эти затраты можно оценить цифрой I0 кбит.

Таким образом, общие затраты памяти C_a при выводе речевой информации аналоговым методом можно оценить как $C_a = 10+20\,N$ (кбит), где N - объем словаря.

Дискретный метод синтеза, использующий компиляцию на уровне слов, требует запоминания вокодерных параметров речи, для которых характерна скорость передачи, лежащая в пределах 0,6-5,0 кбит/сек [I,2,3,5]. Это означает, что для запоминания одного слова необходимо 0,3-2,5 кбит. Однако опыт построения вокодеров показывает, что речь, передаваемая на скорости порядка 0,6 кбит/сек, обладает весьма низким качеством, в связи с чем скорости ниже I,0-I,5 кбит/сек для речевого вывода использовать нецелесообразно. Для оценки наиболее приемлемой скорости, по-видимому, будет некоторая средняя скорость передачи информации, равная, скажем, 3,2 кбит/сек. Следовательно, объем памяти для I слова равен I,6 кбит.

Для организации вывода в этом методе потребуется программа примерно того же объема, что и в предыдущем. На основании сказанного можно записать $C_g = 10+1,6\,N$ (кбит), где C_g -объем памяти, а N - объем словаря.

Приведем в заключение ориентировочную оценку объема памяти для синтеза речи по правилам на фонемном уровне. Этот метод характерен тем, что в памяти ЭВМ должны быть заложены спектрально-временные характеристики эталонов всех звуков речи и сравнительно сложная программа вычисления текущих значений функционала, управляющих синтезатором. Для метода синтеза, описанного в [I6], этот начальный объем памяти может быть оценен в 70 кбит, независимо от объема словаря. Для других алгоритмов синтеза речи по правилам, например, алгоритма, примененного в работе [9], эта величина, по-видимому, будет несколько меньше.

В то же время эталон слова при синтезе речи по правилам определяется только последовательностью звуков, его составляющих. Если известны правила синтеза звуков, эталон слова потребует очень небольшого объема информации. В самом деле, при числе фонетических элементов алфавита не более 2^7 (что вполне достаточно для синтеза) и 6-7 звуках в слове (в среднем), количество информации, определяющей слово, составит всего 42-56 бит. Таким образом, объем памяти для синтеза слов по правилам C_n составит $C_n = 70+0,05\,N$ (кбит).

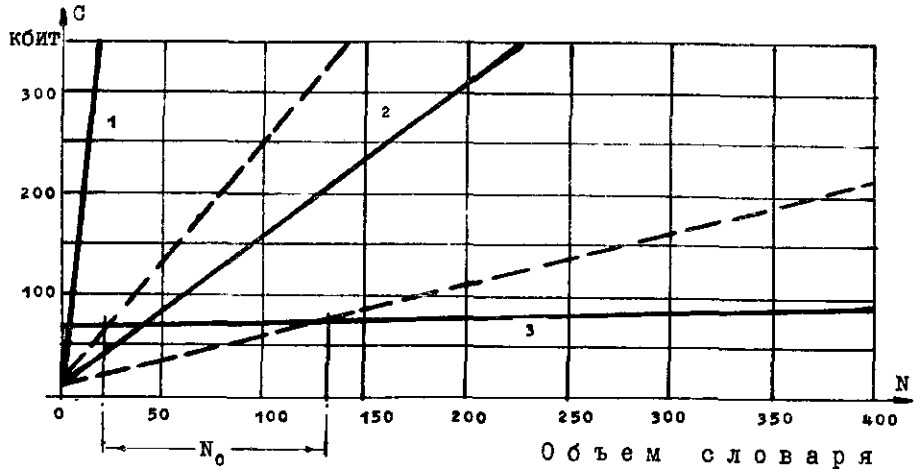

Рис.2. Оценка объема памяти для различных видов синтеза: I - аналоговый синтез; 2 - дискретный синтез на уровне слов; 3 - дискретный синтез на уровне фонем.

Сопоставление приведенных оценок объема памяти C в зависимости от объема словаря N приведено на рис.2. Из него видно, что по этому показателю аналоговый метод оказывается неконку -

рентоспособным с другими методами для любых N, а синтез по правилам становится выгоднее дискретного синтеза методом компиляции при числе слов N_0, лежащем в пределах $20 \leq N_0 \leq 130$ для различной скорости работы синтезатора (1,2–5,0 кбит/сек).

Следует подчеркнуть, что в связи с неустановившимися еще в настоящее время методами синтеза речи по правилам приведенные оценки могут значительно измениться. Более вероятно, что это изменение будет происходить в сторону уменьшения величины N_0.

Возможная структура подсистемы речевой связи. Анализ работ по распознаванию и синтезу речи позволяет сформулировать предложения по структуре подсистемы речевой связи, которая могла бы быть реализована уже в ближайшее время.

Предлагаемый вариант структуры подсистемы речевой связи имеет следующие особенности, обусловленные возможностями существующих алгоритмов распознавания и синтеза речи:

а) подсистема речевой связи должна разрабатываться как единый комплекс, позволяющий, в принципе, организовать диалог оператора и машины;

б) подсистема должна обеспечивать оперативную смену словарей в узлах распознавания и синтеза речи;

в) узел распознавания должен обеспечивать подстройку эталонов под голос говорящего;

г) при вводе речевой информации слова-команды должны разделяться паузами. Вывод речевой информации возможен в форме слитной речи;

д) при реализации подсистемы речевой связи первичный анализ и синтез речи должен осуществляться в специализированных узлах, работающих совместно с базовой ЭЦВМ, снабженной подпрограммами распознавания и синтеза речи.

Вариант укрупненной блок-схемы подсистемы речевой связи, учитывающий перечисленные требования, приведен на рис.3.

Речевой сигнал с микрофона М поступает на полосный анализатор речевого спектра. Кроме полосных каналов, анализатор на входе имеет схемы автоматической регулировки усиления и выделения основного тона речи. Последняя необходима для интонирования синтезированной речи при работе в режиме речевого вывода.

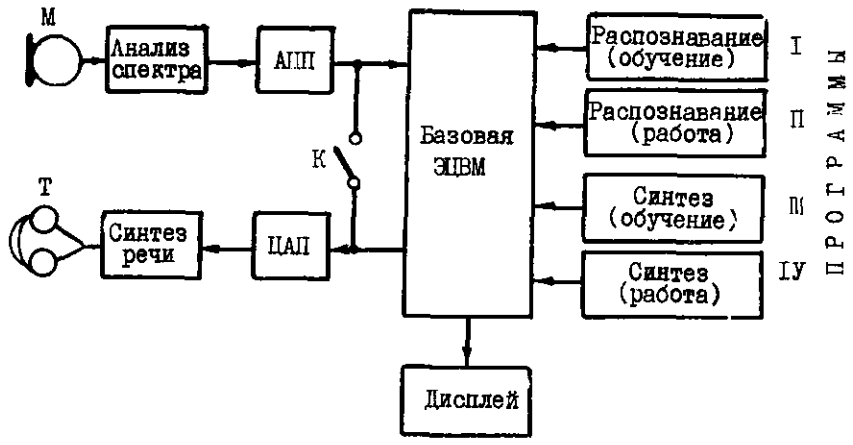

Рис. 3

Спектральные параметры и сигнал основного тона преобразуются в цифровую форму аналого-цифровым преобразователем АЦП и поступают в ЭЦВМ. Точность преобразования АЦП при линейной шкале квантования может быть ограничена 6 разрядами бинарного кода, а при логарифмической шкале – 4 разрядами для частоты опроса 50–100 гц.

В память ЭЦВМ заложены программы обработки речевого сигнала. Первая программа предназначена для получения эталонов при распознавании и используется в случае смены оператора или состава словаря. Эта программа реализует алгоритм, предусматривающий автоматическое формирование правил распознавания слова по значениям его параметров, поступающих из анализатора спектра через АЦП в ЭВМ. Для формирования эталонов слов оператор перед началом работы должен несколько раз произнести каждое слово своего рабочего словаря. Сформированные эталоны заносятся в долговременную память ЭЦВМ, откуда вызываются в случае повторной работы данного оператора с данным словарем.

Вторая программа реализует алгоритм распознавания слов и фраз и используется в режиме речевого ввода. Входной информацией при работе программы являются значения параметров речи,

поступающих с анализатора спектра во время произнесения оператором той или иной команды. Результат распознавания высвечивается на экране дисплея.

Третья программа предназначена для получения эталонов синтезируемых слов. В этом режиме диктор читает перед микрофоном и словарь выводимых слов. Значения спектральных параметров речи и сигнала основного тона с узла анализатора спектра через АЦП записываются в память ЭВМ. Одновременно выполняется маркировка слов и редактирование информации (например, выбрасывание длительных пауз).

Речевой вывод осуществляется с помощью программы IУ.В соответствии с содержанием выводимой информации из памяти в требуемом порядке выбираются параметры слов, выполняется сочленение слов во фразе, введение пауз, формируется интонационный контур. Подготовленные таким образом речевые параметры через согласующийся узел и цифроаналоговый преобразователь ЦАП подаются на синтезатор речи. Синтезированная речь прослушивается через телефоны Т.

В режиме контроля аппаратуры замыкается ключ К, благодаря чему образуется тракт полосного вокодера, который может быть подвергнут речевым и другим измерениям известными методами.

Литература

I. ФЛАНАГАН Д.Л. Анализ, синтез и восприятия речи.М., "Связь", I968.

2. АКБУЛАТОВ А.Ш., БАРОНИН С.П., КУЛЯ В.И. и др. (под ред. А.А.Пирогова) Вокодерная телефония. Методы и проблемы.М., "Связь", I974.

3. САПОЖКОВ М.А. Речевой сигнал в кибернетике и связи.М., "Связьиздат", I963.

4. ПОКРОВСКИЙ Н.Б. Расчет и измерение разборчивости речи. М., "Связьиздат", I962.

5. ЦЕМЕЛЬ Г.И. Опознавание речевых сигналов. М.,"Наука", I97I.

6. ВАСИЛЬЕВ В.И. Распознающие системы. Киев, "Наукова думка", I969.

7. ЗАГОРУЙКО Н.Г. Методы распознавания и их применение.М., "Сов.радио", I972.

8. ТРУНИН-ДОНСКОЙ В.Н., ЦЕМЕЛЬ Г.И. Опознающее устройство для речевого ввода данных в вычислительную машину. Авт.свид. 25I270 кл. 42м^6 9/00, опубл. 26/УШ-69.

9. ФЛАНАГАН Д.Л. и др. Синтезированная речь для ЦВМ.-"Зарубежная радиоэлектроника", I97I, № I0, с.45-8I.

I0. БОНДАРКО Л.В. и др. Модель восприятия речи человеком. Новосибирск, "Наука", I968.

II. БАЙЛИ, АНДЕРСОН. Цифровой полосный вокодер.-"Зарубежная радиоэлектроника", I97I, № 7, с. 3-I6.

I2. ДУКЕЛЬСКИЙ Н.И. Принципы сегментации речевого потока. М.-Л., "Наука", I962.

I3. МЯКИШЕВА И.И. О некоторых особенностях восприятия и измерения разборчивости речи, синтезированной по правилам на фонемном уровне. -В кн.: Труды УI Всесоюз. семинара АРСО-УI, Таллин, I972, с. I45-I48.

I4. ГОЛУБЦОВ С.В. Задачи и перспективы распознавания речи. -Там же, с. 64-73.

I5. ОСАДЧИЙ Ю.Н. Оценка возможности распознавания ограниченного набора команд с использованием суффонемных последовательностей. -В кн.: Труды Акустич. ин-та. Вып. I2. М., I970, с. 60-63.

I6. ГОЛУБЦОВ С.В. Синтез речи. -В кн.: Труды IУ Всесоюз. семинара АРСО-УI , Киев-Канев, I968, с. I07-I29.

I7. ВЫСОЦКИЙ Г.Я., РУДНЫЙ Б.И., ТРУНИН-ДОНСКОЙ В.Н., ЦЕМЕЛЬ Г.И.Автоматическое распознавание нескольких десятков слов и фраз, произнесенных произвольным диктором.-Там же,с.I90-20I.

I8. ГРИГОРЯН А.А. и др. Выделение и ввод признаков речевого сигнала для систем речевого управления. -В кн.: Труды УI Всесоюз. семинара АРСО-УI , Таллин, I972, с. 74-76.

I9. ГУМЕЦКИЙ Р.Я. и др. Фонемное перекодирование слов речи с использованием признаков спектральной динамики. -Там же, с. 8I-84.

20. ГУМЕЦКИЙ Р.Я. и др. Алгоритм распознавания простых фраз. - Там же, с. 85-88.

2I. ХАЙРЕТДИНОВА А.Г. Распознавание слов в слитной речи с использованием информации об ударении. -Там же, с. I79-I8I.

22. ВЕЙЦБАУМ И. Понимание связного текста вычислительной машиной. -В кн.: Распознавание образов. М., "Мир",I970, с.2I4-245.

23. ХАЛЛЕ М. Что мы, собственно, делаем, когда говорим. -Там же, с. 88-I09.

24. ВИНЦЮК Т.К. Поэлементное распознавание слов устной речи. -В кн.: Распознавание образов. Киев, "Наукова думка",I969, с. 88-I03.

25. ВЕЛИЧКО В.Г. Метод фонемного синтеза речи с помощью ЭЦВМ. - Там же, с. I03-II3.

26. ВЕЛИЧКО В.М. и др. Работа по распознаванию речевых сигналов. -В кн.: Распознавание образов. М., I973, с.98-I05.

27. ВИНЦЮК Т.К. Работы Института кибернетики АН УССР по автоматическому распознаванию речевых сигналов. -Там же, с. I06-II7.

28. ГЕРАСИМОВ В.В., МАЛУШЕНКО В.К. Задача распознавания непрерывной последовательности речевых сигналов. -Там же,с.II8-I22.

29. ЕВСЕЕВ А.И. Некоторые вопросы автоматического распознавания сменных ограниченных словарей. Автореф. дисс. на соиск. учен. степени канд. техн.наук, М., I974 (Моск.энерг. ин-т).

30. МАКСИМОВ Ю.Г. и др. Перспективы использования речевого ввода исходной информации в память ЭВМ для решения некоторых оперативных задач дальнего транспорта газа. -В сб.: Распознавание образов. М., I973, с. I67-I72 (ВЦ АН СССР).

3I. ЦЕМЕЛЬ Г.И. Построение аппаратуры выделения сегментных признаков для систем речевого управления. -Там же, с.I73-I76.

32. МАККАРТИ Дж. и др. Вычислительная машина с руками,глазами и ушами. -В сб.: Интегральные роботы. Под ред. Г.Е.Поздняка, М., "Мир", I973, с.4I-60.

33. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание ограниченного набора устных команд. -В кн.: Вычислительные системы. Вып. 36. Новосибирск, I969, с.IOI-II0.

34. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Автоматическое распознавание 200 устных команд. -В кн.: Вычислительные системы.Вып.37. Новосибирск, I969, с.73-76.

35. ВЕЛИЧКО В.М. О некоторых методах автоматического распознавания речевых сигналов. Автореф. дисс. на соиск. учен. степени канд. техн. наук, Новосибирск, I97I (Ин-т математики СО АН СССР).

36. ВИНЦЮК Т.К. Методы обучения, самообучения и распознавание речи, основанные на составлении эталонных сигналов из элементарных частей. -Сб.рефератов докладов на УШ Всесоюз. акуст. конф. Том. I, М., I973, с. 9-I0.

37. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Об одном подходе к распознаванию большого словаря. -Там же, с. 53.

38. ВАСИЛЬЕВ А.В. и др. Опознавание наборов слов с использованием алгоритма группирования слов. -Там же, с. 54-55.

39. КУЛЯ В.И. Роль адаптации при распознавании речевых сигналов. -Там же, с. 57.

40. ВИНЦЮК Т.К. и др. Экспериментальная система ввода данных в ЦВМ посредством голоса. -Там же, с. 62.

4I. ВИНЦЮК Т.К. Распознавание некоторых классов речевых сигналов. Автореф. дисс. на соиск. учен. степени канд. техн. наук, Киев, I967 (Ин-т кибернетики АН УССР).

42. ТРУНИН-ДОНСКОЙ В.Н. Исследование вопросов речевого управления вычислительными машинами. Автореф. дисс. на соиск. учен. степени канд. техн. наук, М., I969 (ВЦ АН СССР).

43. РУДНЫЙ Б.Н. Исследование речевых сигналов и разработка методов многоуровневого распознавания в системах оперативного управления. Автореф. дисс. на соиск. учен. степени канд. техн. наук. М., I97I (ВЦ АН СССР).

44. ВЫСОЦКИЙ Г.Я. Исследование вопросов речевого управления в системах оперативного взаимодействия человека и машины. Автореф. дисс. на соиск. учен. степени канд. техн. наук. М., I973 (Моск. физ.-техн. ин-т).

45. ГРИГОРЯН А.А. Исследование динамики формантных частот гласных звуков с применением полученных признаков для опознавания набора слов. Автореф. дисс. на соиск. учен. степени канд. техн. наук. М., I972 (Ин-т пробл. перед. инф. АН СССР).

46. GOLD B. Word Recognition Computer Program. - Mass.Inst. Techn. Rept.452, 1966.

47. REDDY D.R. Segmentation of Speech Sounds. - "J.Amer.Statist.Assoc.",1966, v.40, N 2,p.307-312.

48. FUDJISAKI H., SUDO H. Synthesis by Rule of Prosodic Features of Connected Japanese". - Proc.of the 7th Int.Congr. of Acoust. Budapest, 1971, v.23C2,p.133-136.

49. HASHIMOTO S., SATIO S. Prosodic Rules for Speech Synthesis. - Proc.on the 7th Int.Congr.of Acoust. Budapest,1971, v. 23C1, p.129-132.

50. COKER C.H., UMEDA N. Toward a Theory of Stress and Prosody in American English. - Proc.of the 7th Int.Cngr. of Acoust. Budapest,1971, v.23C3, p.137-140.

51. MARVIN B., COX R.B. An adaptive Isolated Word Speech Recognition System. - "Rec.Conf.on Speech Comm. and Proc. Newton, Mass., 1972, v.C1,p.89-92.

52. CLARK M.T. and oth. Word Recognition by Means of Walsh Transforms. - Rec.Conf.on Speech Comm. and Proc. Newton,Mass., 1972, v.C3, p.97-100.

53. NAKANO Y. and oth. Evalution of Varios Parameters in Spoken Digits Recognition. - Rec.Conf.on Speech Comm.and Proc., Newton, Mass., 1972, v.C4,p.101-104.

54. HUGHES G.W. and oth. An approarch to Research on Word Spotting in Continuous Speech. - Rec.Conf.on Speech Comm. and Proc. Newton, Mass.,1972, v.C6, p.109-112.

55. MEDRESS M. A procedure for the Mashine Recognition of Speech. - Rec.Conf. on Speech Comm. and Proc. Newton,Mass.,1972, v.C7,p.113-116.

56. CARTIER M. and oth. Speaker Adaptation to an Automatic Speech Recognition System. - Rec.Conf.on Speech Comm.and Proc. Newton, Mass., 1972, v.C4, p.287-290.

57. ARAI Y. and oth. Automatic multichannel Speech Synthesizer. - Proc.of the 7th Int.Congr.on Acoust. Budapest, 1971, v.23C3.

58. REDDY D.R. and oth. Speech Recognition in a Multiprossor Environmet. - IEEE Conf.Decis. and Control, Miamy Beach , Florida, 1971.

59. RABINER L. and oth. Computer Synthesis of Speech by Concatenation of Formant-Coded Words. - BSTJ,1971,v.50, N 5, p. 1541-1558.

60. MASAKI K. and oth. Spoken Digits Mechanical Recognition System. - "Electr.Lab.Techn.Journ.,1972, v.21,N 7,p.1361-1370.

61. COKER C.H., UMEDA N. Conversion of Printed Text into Synthetic Speech. Pat.USA, kl.179, G-10.1-1/10, N 3704345, 19.03.71.

62. LEA W.A. An Approarch to Syntactic Recognition whithout Phonemics. - "IEEE Trans.Audio and Electroacoustic, 1973, v.21, N 3,p.249-258.

63. CARPENTER B.E., LAVINGTON S.H. The Influence of Human Factors on the Performance of a Real-Time Speech Recognition , System. - "J.Amar.Statist.Assoc.",1973,v.53,N 1,p.42-45.

64. GLENN J.W., HITHCOCK M.H. With a Speech Pattern Classifier, Computer Listens to its Master's Voice. - "Electronics" 1971, v.44, May, N 10, p.84-89.

65. POLS L.C.W. Real Time Recognition of Spoken Digits. - "IEEE Trans.on Comp.",1971,v.C-20, N 9, Sept., p.972-978.

66. MAKHOUL J. Speaker Adaptation in Limit Speech Recog - nition System. - "IEEE Trans.on Comp.",1971,v.C-20,N 9, Sept., p.1057-1063.

67. TAKEMOCHI J. and oth. Segmentation of Japanese Words. - "Fudjitshu Sci. and Techn.Journ., 1972,v.8,N 3,p.109-121.

68. COMPUTER given "Voice" on Bell Assemby Line. - "Bell. Labs Rec.",1972,v.50, N 3,p.98.

69. REDDY D.R. and oth. A model and a System for Machine Recognition of Speech. - "IEEE Trans on Audio and Electroaceustic",1973,v.21, N 3,p.229-238.

70. KLATT D.H., STEVENS K.N. On the Automatic Recognition on Continuous Speech: Unplications from a Spectrogram-Reading Experiment. - "IEEE Trans. on Audio and Electroacoustic",1973, v.21,N 3,p.210-217.

71. SHAFER R.W., RABINER L.R. Design and Simulation of a Speech Analysis-Synthesis System Based on Short Time Fouier Analysis. - "IEEE Trans.on Audio and Electroacoustic", 1973, v.21, N 3,p.165-174.

72. FANT G. Automatic Recognition and Speech Research.STL-QPSR 1/1970, p.32-40. Speech Transm.Lab.Rojal Inst of Techn. Stockholm, Sweden.

73. SCARR R.W.A. Normalisation and Adaptation of Speech Data for Automatic Speech Recognition. - "Intern.Journ. of Man-Mach.Studies",1970,v.2, N 1,p.41-59.

74. PIERCE J.R. Whither Speech Recognition? - "J.Amer.Statist.Assoc.",1969,v.46, N 4,pt.2, p.1049-1051.

75. LEA W.A. Towards Versatile Speech Communication with Computers. - "Intern.Journ.of Man-Mach.Studies",1970,v.2,N 2, p.107-155.

76. VENEZKY R.L. Automatic Spelling-to-Sound Conversion.- "Computat.Linguistics.Bloomington-London, Ind.Univ.Press., 1966,p.146-161.

77. RABINER L.R. A Model for Synthesing Speech by Rule. - "IEEE Trans. on Audion and Electroacoustic",1969,v.17, N 1 , p.7-13.

78. JONATAN A. Machine-to-Man Communication by Speech . Part 2. Synthesis of Prosodic Features of Speech by Rulle. - AFIPS Conf.Proc., v.32. Washington, D.C. Thompson Book Co., 1968, p.339-344.

79. CARLSON R., GRANSTROM B. Word Accent, Emphatic Stress and Syntax in a Synthesis by Rule Scheme for Swedish.STL-QPSR, 2-3/1973, Speech Transm. Lab. Rojal Inst. of Techn. Stockholm, Sweden.

80. LILJENCRANTZ. The OVE-III Speech Synthesizer.STL-QPSR, 2-3/1967,p.76-81. Speech Transm. Lab. Rojal Inst. of Techn. Stockholm, Sweden.

Поступила в ред.-изд. отд.
7 мая 1975 года