А Н А Л И З Т Е К С Т О В И С И Г Н А Л О В
(Вычислительные системы)
1987 год Выпуск 123

УДК 621.391

СИСТЕМА РАСПОЗНАВАНИЯ СЛОВ В ПОТОКЕ СЛИТНОЙ РЕЧИ

Б.Д.Наумов

В в е д е н и е

Одной из задач распознавания слуховых образов является создание высокоэффективных по времени и надежных по качеству распознавания средств для поиска опорных (ключевых) слов в слитной речи, состоящей из слов произвольного словаря. Определенных успехов в решении этой задачи добились некоторые исследовательские группы в Японии, США и ряде стран Западной Европы [I] (см. § 5 настоящей статьи).

Разработки систем распознавания и смысловой интерпретации слитной речи, включающих в себя процедуры обнаружения и верификации ключевых слов в непрерывном речевом потоке, ведутся в ВЦ АН СССР [2] и в ИК АН УССР [3,4]. Судить об эффективности используемых в этих системах алгоритмов распознавания слов в слитной речи трудно, поскольку в публикациях не приведены результаты соответствующих экспериментов. Такая ситуация, по-видимому, объясняется тем, что существующий в настоящее время уровень серийной вычислительной техники не обеспечивает функционирования предложенных авторами алгоритмов выделения слов в реальном масштабе времени, а применение специализированных аппаратурных средств требует, по мнению самих разработчиков (см., например, [4, с.244]), решения ряда сложных научно-технических проблем.

Предлагается алгоритм распознавания слов в потоке слитной речи, позволяющий с приемлемыми аппаратурными затратами создавать системы выделения слов с объемом словаря до нескольких сотен слов, работающие в реальном масштабе времени. Роль основной аппаратурной поддержки при этом отводится специализированному ДП-процессору, использующему процедуру динамического программирования (ДП)

для принятия решения на уровне подсловарей. Прототипом ДП-процессора является спецпроцессор [5], реализующий алгоритм распознавания изолированных слов (§I). Приведенный алгоритм в рассматриваемой системе модифицируется путем введения ограничений на форму оптимального пути (§2) и изменения начальных условий и процедуры принятия решения (§3). Для реализации модифицированного таким образом алгоритма в реальном масштабе времени предложена архитектура микропроцессорной системы выделения слов (§4). Экспериментальные исследования алгоритма проведены на системе с одним ДП-процессором (§5). В дальнейшем для повышения надежности распознавания системы предлагается ввести адаптивную подстройку алгоритма вычисления локального расстояния к уровню речевого сигнала (§6).

§I. Схема динамического программирования с фиксированными конечными точками

Алгоритм распознавания изолированных слов, реализованный с помощью спецпроцессора [5], основывается на установлении степени сходства первичного описания входного речевого сигнала с эталонными представлениями слов. Для учета возможных изменений темпа речи в нем используется метод динамической временной нормализации с фиксированными конечными точками, основной принцип которого можно трактовать как нелинейную деформацию временной шкалы контрольной реализации относительно временной шкалы эталонной реализации с целью минимизации расхождения между сравниваемыми реализациями.

Предполагается, что сравниваемые реализации представлены последовательностями векторов, описывающих речевой сигнал на заданном интервале времени (сегменте). Эталонные реализации $A^n = a^n_1, a^n_2, \ldots, a^n_i, \ldots, a^n_{I^n}$, $n = 1, \ldots, N$, хранятся в памяти, а контрольная реализация $B = b_{j_н}, b_{j_н+1}, \ldots, b_j, \ldots, b_{j_к}$, характеризующая некоторый фрагмент анализируемой речевой последовательности, подается на вход системы.

На рис.I процедура динамической временной нормализации изображена в (i,j)-плоскости, где горизонтальная ось соответствует контрольной реализации, а вертикальная - эталонной. Цель временной нормализации - нахождение пути на этой плоскости с минимальным полным (интегральным) расстоянием между сравниваемыми реализациями. В методе динамической временной нормализации с фиксированными конечными точками все возможные пути начинаются в точке $(1, j_н)$ и заканчиваются в точке $(I^n, j_к)$.

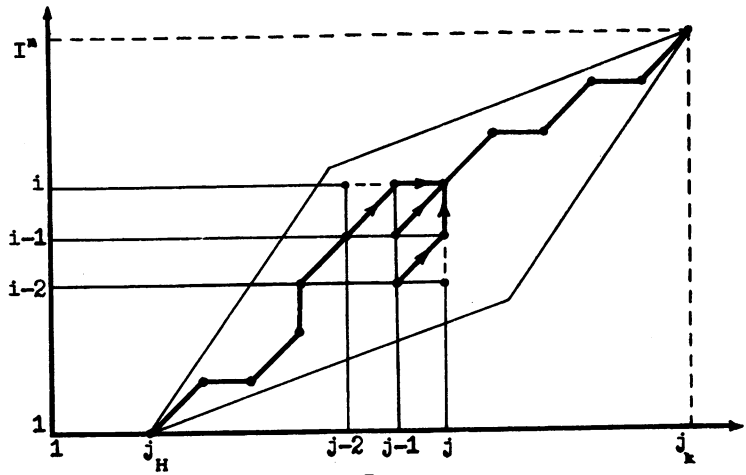

Рис.I

Для каждой точки (i,j)-плоскости вычисляется локальное расстояние d(i,j) между сегментами a_i и b_j. Интегральное расстояние $D(A^n,B)$ между реализациями A^n и B - сумма локальных расстояний между сегментами сравниваемых реализаций вдоль оптимального пути, определяется по формуле:

$$D(A^n,B) = \min_{U(j)} \sum_{j=j_н}^{j_к} (d(U(j),j)\cdot g(j,U(j),\dot{U}(j))),$$

где U(j) - функция нелинейной нормализации по времени, $\dot{U}(j)$ - производная функции U(j) по времени, g - весовая функция.

Текущее интегральное расстояние D(i,j) вычисляется в каждой точке плоскости нормализации. Для строго монотонно возрастающих функций U(j) при $g \equiv 1$ схема вычислений D(i,j) реализуется следующим образом:

$$D(i,j) = \min\{D(i-1,j), D(i-1,j-1), D(i,j-1)\} + d(i,j).$$

Начальные условия имеют вид:

$$D(i,j) = \begin{cases} \infty, & i=0,\ j \geq j_н; \\ 0, & i=0,\ j = j_н - 1; \\ \infty, & i>0,\ j = j_н - 1. \end{cases}$$

Интегральное расстояние $D(A^n,B) = D(I^n, j_k)$ следует рассматривать как оценку расхождения между реализациями A^n и B. В качестве критерия для принятия решения используется нормализованное расстояние $D(A^n,B)/(I^n + j_k - j_н)$. Для каждой эталонной реализации вычисляется нормализованное интегральное расстояние до контрольной реализации, после чего определяется эталон с наименьшим значением расстояния.

§2. Введение ограничений на форму оптимального пути

При вычислении интегрального расстояния в описанном выше алгоритме форма оптимального пути ограничена только требованием монотонности функции $U(j)$. К сожалению, для получения корректной оценки различия между сравниваемыми словами этого ограничения недостаточно [6]. В реальном процессе классификации речевых сигналов желательно учитывать только такие локальные деформации временной шкалы, которые фактически отвечают за передачу информации, содержащейся в речи. Формализовать их трудно, поскольку степень временной изменчивости, которую можно считать допустимой, неизвестна. Так, например, некоторые фонемы в различных реализациях одного и того же слова могут значительно отличаться по длительности.

Исследование различных вариантов введения ограничений на форму оптимального пути проводилось во многих работах по распознаванию речи [7-10,16,17]. Большинство предлагаемых ограничений, направленных на сокращение объема вычислений, не гарантируют получение глобально-оптимального решения, что нежелательно для наших целей. Из известных методов, сохраняющих возможность получения глобально-оптимального решения, особый интерес вызывает метод [16], в котором используются как локальные, так и глобальные ограничения, налагаемые на функцию $U(j)$ функцией g. Глобальные ограничения задают максимальный коэффициент сжатия или растяжения временной шкалы контрольной реализации и имеют следующий вид:

$$g(j,U(j),\dot{U}(j)) = \begin{cases} \infty, & \dot{U}(j) < 1/2; \\ 1, & 1/2 \le \dot{U}(j) \le 2; \\ \infty, & \dot{U}(j) > 2. \end{cases}$$

Локальные ограничения вводятся для того, чтобы локальные изменения временной шкалы имели тот же диапазон, что и глобальные. В частности, запрещаются переходы из точек $(i,j-2)$ и $(i-2,j)$ в

точку (i,j) (см. [IO]). В литературе алгоритм с подобными ограничениями известен как алгоритм динамического программирования со степенью деформации, равной двум. На рис.I локальные ограничения, налагаемые на функцию U(j), интерпретируются стрелками, показывающими допустимые пределы нелинейных деформаций при изменении темпа речи. Внутренняя область параллелограмма, границы которого заданы глобальными ограничениями, является областью существования функции U(j).

Утверждается [IO], что степень деформации, равная двум, является оптимальной для слов японской речи, поскольку в экспериментах для нее получена самая высокая надежность распознавания. Проверка справедливости этого утверждения для слов русской речи была выполнена нами для значений степени деформации h, лежащих в диапазоне от двух до семи. Для сравнения в таблице также приводятся результаты распознавания, полученные для полного алгоритма ДП ($\max_{1 \le n \le N} I^n < h$, h = 100).

Эксперименты по распознаванию изолированных слов проводились на тестовом материале, составленном одним диктором из "фонетически богатых" слов (I2 реализаций по I00 слов) и акустически подобных слов (20 реализаций по 67 слов). Полученные реализации записывались на магнитную ленту и при распознавании предъявлялись поочередно в качестве эталонной реализации. Остальные реализации данного словаря использовались как контрольные. Локальное расстояние в алгоритме ДП вычислялось по формуле:

$$d(i,j) = \sum_{k=1}^{p} |x_k^{(i)} - x_k^{(j)}|,$$

где p - размерность пространства признаков; $x_k^{(i)}$ - значение k-го признака для i-го сегмента эталонной реализации; $x_k^{(j)}$ - значение k-го признака для j-го сегмента контрольной реализации. Длительность сегмента равнялась I7 мс. В качестве признаков использовались интенсивности речевого сигнала в шести спектральных полосах, нормированные по общей интенсивности на данном сегменте.

Результаты экспериментов (см.таблицу) позволяют предположить, что для слов русской речи оптимальное значение допустимой степени равно трем (а не двум, как утверждается в [8]).

Таблица

Зависимость надежности распознавания от степени деформации

Объем словаря	Степень деформации h						
	2	3	4	5	6	7	I00
67 слов	87,6	89,4	88,8	86,9	86,5	86,I	85,7
I00 слов	97,9	98,I	98,0	97,7	97,6	97,4	97,2

§3. Алгоритм распознавания слов в потоке слитной речи

Эффективность использования метода динамической временной нормализации с фиксированными конечными точками во многом зависит от надежности выделения граничных точек сравниваемых реализаций [II]. При получении эталонов проблема выделения границ решается сравнительно просто, поскольку процесс обучения (в отличие от распознавания) не ограничен рамками реального времени и обычно происходит в режиме диалога человека с ЭВМ. При определении граничных точек контрольной реализации в слитной речи часто бывает довольно сложно найти точные границы между словами. Поэтому при распознавании слитной речи, на наш взгляд, более предпочтительны алгоритмы, не требующие предварительного членения анализируемого речевого потока на слова. Именно такому условию удовлетворяет алгоритм распознавания слов в слитной речи, основанный на идее метода динамической временной нормализации без фиксации конечных точек [I2], при ограничениях на форму оптимального пути, обоснованных в предыдущем разделе.

Метод динамической временной нормализации без фиксации конечных точек (называемый также методом динамической временной нормализации с "плавающими" концами) отличается от приведенного ранее алгоритма ДП начальными условиями ($D(0,j)=0$ для всех j) и "скользящей" во времени процедурой принятия решения.

Динамическую временную нормализацию с "плавающими" концами можно интерпретировать как процедуру параллельного сравнения эталона с множеством контрольных реализаций $\Omega(j_н)$. Последние получаются из некоторой исходной контрольной реализации $B(j_н,j)$ с начальной точкой $j_н$ и правым концом, ограниченным текущим значением j, путем сдвига по оси времени начальной точки в прямом на -

правлении соответственно на I,2 и более сегментов. Так как начальные точки соседних контрольных реализаций отличаются при этом всего на один сегмент, то траектории оптимальных путей, связанных с этими реализациями, могут пересекаться друг с другом, что приводит к отсечке оптимального пути для контрольной реализации, предшествующей во времени. Другими словами, в процессе нормализации из контрольной реализации автоматически выделяется фрагмент, самый близкий по интегральному расстоянию к сравниваемому эталону.

Распознавание слов путем сравнения с эталоном обычно сводится к выделению из контрольной реализации участков речевой последовательности по мере сходства со словами из заданного словаря. Поэтому для принятия решения остается определить эталон с минимальным расстоянием от контрольной реализации $B(j_H,j)$ в момент времени j, выбранный таким образом, чтобы в интервал $[j_H,j]$ попала реализация слова из заданного словаря. С этой целью проводится последовательное вычисление интегрального расстояния между каждым из N эталонов и контрольной реализацией для всех j, начиная с произвольно выбранного момента j_H. На каждом шаге этой процедуры определяется эталон с вычисленным минимальным расстоянием, нормированным на длину эталона I^n, т.е. находится значение n, удовлетворяющее условию:

$$D(A^n,B(j_H,j)) = \min_{1 \le r \le N} (D(A^r,B(j_H,j))/I^n).$$

Введем следующие пороги: P1 - по абсолютному значению расстояния; P2 - по глубине просмотра вправо; P3 - по минимальной длине слова.

Если на каком-либо шаге j' расстояние между доминирующим эталоном и анализируемой реализацией окажется ниже порога P1, то этот эталон выбирается в качестве основного претендента на распознавание.

Следуя в прямом направлении от точки j', определим первый локальный минимум функции $D^*(j) = D(A^n,B(j_H,j))/I^n$, где $n = n(j)$ - номер доминирующего эталона на j-м шаге. Предположим, что этот минимум достигается в точке l_1 (рис.2). Если

$$D^*(l_1) = \min_{l_1 \le j \le l_1+P2} D^*(j),$$

то процедура вычисления расстояния на данном этапе принятия решения прекращается. В противном случае находим точку l_2, в которой достигается следующий локальный минимум функции $D^*(j)$, и вновь проверяем, является ли этот минимум глобальным на интервале $[l_2, l_2 + P2]$. Окончанию процесса поиска нужного минимума на k-м шаге соответствует принятие решения в точке $j_k = l_k$.

При переходе к распознаванию следующего слова естественно выбрать новую начальную точку интервала анализа $j'_н$, равную $l_k + 1$. При этом следует иметь в виду, что в окрестности точки принятия решения, а именно на интервале $[j_k, j_k + P2]$, могут существовать точки, соответствующие неоптимальным путям на плоскости нормализации, в которых возможно ложное срабатывание системы. Для предотвращения этого предлагается ограничить слева интервал принятия решения значением $j_н + P3$.

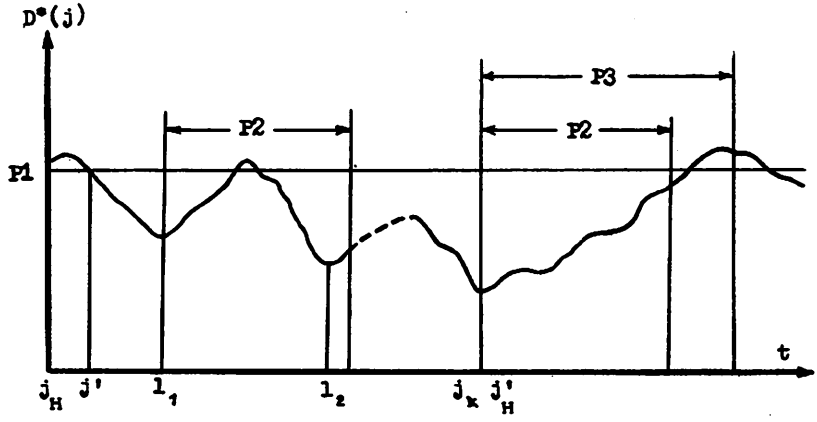

Рис. 2

В результате условие обнаружения эталонного слова в непрерывном потоке речи приобретает следующий вид:

$$D(A^a, B(j_н, j_k))/I^a = \min_{\substack{1 \le r \le N \\ j_н + P3 \le j \le j_k + P2}} D(A^r, B(j_н, j_k))/I^r < P1 .$$

§4. Архитектура системы распознавания слов в потоке слитной речи

Как отмечалось в начале статьи, система, предназначенная для решения задачи выделения слов из слитной речи в реальном масштабе времени, должна обладать вычислительной мощностью, превышающей возможности отдельных серийно выпускаемых в настоящее время процессоров. Поэтому для реализации системы распознавания слов в потоке слитной речи было решено использовать мультипроцессорную архитектуру с магистральной организацией (рис.3).

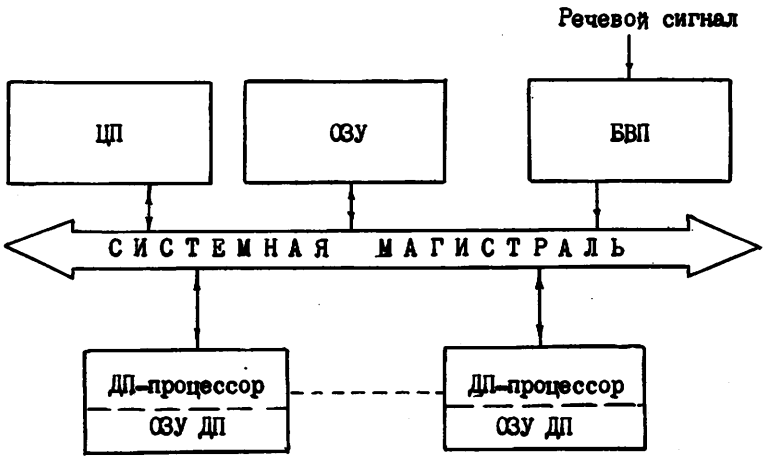

Рис. 3

Функции устройства управления в системе выполняет центральный процессор (ЦП) микро-ЭВМ "Электроника-60 М" с системной магистралью Q-bus. Помимо центрального процесса к системной магистрали (каналу ЭВМ) подключаются: системное оперативное запоминающее устройство (ОЗУ), блок выделения признаков речевого сигнала (БВП) и до восьми специализированных ДП-процессоров, реализующих алгоритм, приведенный в §3. Каждый из ДП-процессоров имеет встроенную память (ОЗУ ДП), в которой хранятся эталоны слов и элементы матриц расстояний, вычисляемых в процессе принятия решения спец-процессором.

Если в системе используется более чем один ДП-процессор, то словарь выделяемых слов разбивается на соответствующее количество подсловарей. Максимальный объем подсловаря определяется суммарной длительностью эталонных реализаций, размещающихся в отведенной для них области памяти спецпроцессора (при средней длительности реализации 0.7 с объем подсловаря равен 50).

При параллельной работе нескольких ДП-процессоров контрольная реализация одновременно сравнивается со всеми эталонами слов, хранящимися в памяти этих процессоров (длительность цикла сравнения не зависит от структуры подсловаря и всегда равна I5 мс). После окончания цикла сравнения каждый из спецпроцессоров вырабатывает номер доминирующего эталона из заданного подсловаря и соответствующее ему нормализованное интегральное расстояние. Центральный процессор из полученного таким образом набора доминирующих эталонов выбирает эталон с минимальным значением расстояния и в соответствии с предложенным алгоритмом определяет, выполняется ли на данном шаге условие обнаружения эталонного слова в потоке слитной речи.

Сегмент речевого сигнала в системе представлен четырьмя I6-разрядными словами, характеризующими интенсивность сигнала на интервале анализа длительностью не менее I6 мс в восьми спектральных полосах (на стадии отладки системы в качестве блока выделения признаков применялось устройство для первичной обработки речевого сигнала с шестью спектральными полосами [I3]).

§5. Экспериментальные результаты

Основными характеристиками системы распознавания слов в потоке слитной речи являются эффективность распознавания слов и количество ложных срабатываний за определенный промежуток времени.

Эффективность распознавания слов определяется как отношение количества правильно распознанных слов к общему количеству слов из выделенного словаря (в дальнейшем именуемых выделенными словами), предъявленных к распознаванию. Принятие решения о распознавании какого-либо выделенного слова в момент времени, отличный от момента появления этого слова в контрольной реализации, квалифицируется как ложное срабатывание или ошибка второго рода.

Эти характеристики определялись на специально подобранном словаре из 25 слов, имеющих существенные фонологические отличия. Для распознавания предъявлялись фразы различной длины (от трех до

семи слов), произносимые одним диктором при отношении сигнал/шум, приблизительно равном 20 дб. Каждая фраза содержала от одного до трех выделенных слов, представляла одну синтагму и произносилась слитно. Система признаков и метод вычисления локальных расстояний остались такими же, как и при проведении экспериментов по определению оптимальной степени деформации сравниваемых речевых образов.

Для снятия характеристик использовалась мультипроцессорная система (§4), содержащая один ДП-процессор. Распознавание слов происходило в реальном масштабе времени, задержка выдачи результата не превышала 16 мс.

Эффективность распознавания слов на контрольном материале из 500 фраз составила 95-96% при 4-6 ложных срабатываниях. При введении ограничений на величину ошибок второго рода (появление не более одного ложного срабатывания) эффективность распознавания снижалась до 84-87%.

Полученные результаты следует считать предварительными. Эффективность работы системы предполагается повысить с помощью цифрового процессора для анализа речевых сигналов [14] и нового метода вычисления локального расстояния, описанного в §6.

Качество решения задачи выделения слов во многом определяется выбором тестового материала и, как следствие, эффективность распознавания может изменяться в широких пределах в зависимости от уровня шумов и степени ограничений, учитываемых в тестовом материале. Поэтому невозможно провести сколько-нибудь аргументированное сравнение различных подходов к решению задачи выделения слов при оценке их эффективности на различных тестовых материалах.

Трудность сопоставления полученных результатов с известными системами выделения слов также усугубляется тем обстоятельством, что при создании системы, описанной в §5, основное внимание уделялось увеличению объема словаря, тогда как зарубежные разработки в большей степени ориентированы на определение сравнительно небольшого количества ключевых слов в речи произвольного диктора на фоне помех. Так, например, система выделения слов, разработанная фирмой Dialog Sistems [15], обеспечивает выделение одного ключевого слова в 90-95% случаев при 4-6 ложных тревогах в 1 ч на тестовом материале, составленном из специальной подборки новостей, которую читали вслух девять человек. При испытаниях с десятью другими дикторами эффективность выделения ключевого слова составила 70% при 6 ложных тревогах в 1 ч.

В [I5] также сообщается о более сложной системе выделения слов, при испытании которой была обнаружена определенная корреляция между характеристиками распознавания и фонологической структурой выделяемых слов. Если, например, допускается не более двух ложных сигналов в процессе выделения десяти ключевых слов при анализе 24-минутной речи на фоне шумов, то правильно будет выделено лишь 38% ключевых слов. При этом 60% пропущенных ключевых слов составляют четыре слова, которые подвержены наибольшей фонологической редукции при разговорной речи.

§6. Адаптивный метод вычисления расстояния между сегментами речевого сигнала

Слабая зависимость между формализмом ДП и метрикой пространства признаков речевого сигнала дает возможность применять методику ДП к различным первичным описаниям, разрабатываемым независимо от процедуры принятия решения на уровне слов. С первичным описанием непосредственно связана лишь проблема вычисления локального расстояния между отдельными сегментами речевого сигнала.

Рассмотрим случай, когда используемые признаки в явном виде содержат информацию об уровне речевого сигнала, как это имеет место в системах с полосно-спектральным описанием, при этом интенсивность шумов постоянна. Тогда вектор признаков при изменении интенсивности произнесения отличается от эталонного вектора множителем $\alpha > 0$. Бесспорно, что этот факт следовало бы учитывать при вычислении локального расстояния.

В наиболее распространенных в настоящее время методах вычисления локального расстояния используется предварительная нормализация спектрально-полосных признаков речевого сигнала одним из следующих способов:

I) по общей интенсивности на данном сегменте;

2) по максимуму общей интенсивности на длине всей реализации.

Первый достаточно прост и надежен в реализации, но дает большую погрешность при малых уровнях речевого сигнала. Второй - частично устраняет этот недостаток, но возникает необходимость определения максимума общей интенсивности на длине всей реализации, что затрудняет применение этого способа в системах реального времени.

Способ учета изменений интенсивности речевого сигнала, свободный от указанных недостатков, предложен в [4]. Он предполагает, что за эталонным вектором a_i^n $(1 \le n \le N,\ 1 \le i \le I^n)$ с относительной интенсивностью $\alpha_{n\,i}$ следует эталонный вектор a_{i+1}^n с относительной интенсивностью в интервале $[\alpha_{n\,i} - \varepsilon,\ \alpha_{n\,i} + \varepsilon]$. При этом задается возможный диапазон изменения интенсивности $\alpha_{min} \le \alpha_{n\,i} \le \alpha_{max}$ и вводится шкала дискретных значений $\alpha_m = \alpha_{min} + m\,(\alpha_{max} - \alpha_{min})/M$, $m = 0, \ldots, M$. Затем генерируются все возможные в рамках заданной шкалы эталоны слова с различной интенсивностью произнесения. Решение задачи распознавания для этого случая приводит к двумерному варианту схемы ДП, что в $\varepsilon \cdot M^2/(\alpha_{max} - \alpha_{min})$ раз увеличивает объем вычислений по сравнению с обычным алгоритмом ДП. В результате затраты на реализацию этого способа не оправдывают полученный выигрыш в надежности распознавания [4]. Для сокращения объема вычислений нами разработан следующий метод вычисления локального расстояния между сегментами речевого сигнала.

Представим сравниваемые сегменты a_i и b_j в виде точек в p-мерном пространстве признаков R^p. Пусть расстояние между этими точками вычисляется по формуле:

$$r(i,j) = \sqrt{\sum_{k=1}^{p} (x_k^{(i)} - x_k^{(j)})^2}.$$

Поставим в соответствие каждой паре точек пространства R^p параметр α_{ij}, определяемый следующим образом:

$$\alpha_{ij} = \underset{0 \le \alpha_{ij} < \infty}{\operatorname{argmin}} \sqrt{\sum_{k=1}^{p} (x_k^{(i)} - \alpha_{ij} \cdot x_k^{(j)})^2}.$$

Исключив из рассмотрения не встречающийся в реальных условиях случай, когда $x_k^{(j)} = 0$, $1 \le k \le p$, получим

$$\alpha_{ij} = \left(\sum_{k=1}^{p} x_k^{(i)} \cdot x_k^{(j)} \right) \Big/ \left(\sum_{k=1}^{p} (x_k^{(j)})^2 \right).$$

Будем считать, что определенный таким образом параметр α_{ij} характеризует относительную интенсивность векторов b_j и a_i. Если вычисленное значение α_{ij} попадает в интервал допустимых изменений интенсивности $[\alpha_{min}, \alpha_{max}]$, то локальное расстояние между сегментами речевого сигнала определяем по формуле:

$$d(i,j) = \sqrt{\sum_{k=1}^{p} (x_k^{(i)} - \alpha_{ij} \cdot x_k^{(j)})^2} =$$

$$= \sqrt{\sum_{k=1}^{p} (x_k^{(i)})^2 - (\sum_{k=1}^{p} x_k^{(i)} \cdot x_k^{(j)})^2 / (\sum_{k=1}^{p} (x_k^{(j)})^2)} . \quad (*)$$

В противном случае принимаем $d(i,j) = r(i,j)$.

Еще большего сокращения объема вычислений можно добиться, предположив допустимым такое изменение интенсивности сигнала α, при котором сегмент контрольной реализации b_j находится внутри гиперсферы с центром в точке a_i и радиусом r_0. Таким образом, если $r(i,j) \leq r_0$, то вычисляем $d(i,j)$ по формуле (*), в противном случае полагаем $d(i,j) = r(i,j)$.

Сравнительные эксперименты по распознаванию изолированных слов на словарях из 100 и 67 слов для одного диктора с применением алгоритма ДП со степенью деформации, равной трем, дали следующие результаты по надежности распознавания: первый метод нормализации - 98,3% и 89,6% соответственно, второй метод нормализации - 98,6% и 91,2%, адаптивный метод - 99,1% и 93,9%.

Полученные результаты позволяют надеяться на то, что микропроцессорная реализация предложенного метода окажется достаточно эффективным средством повышения надежности распознавания слов в потоке слитной речи.

Заключение

Одним из факторов, сдерживающих широкое применение систем автоматического распознавания речи, является чувствительность качества машинного распознавания к шумам и искажениям речевого сигнала, несущественным при восприятии человеком речи на слух. Основной процент ошибок распознавания, возникающих по этой причине, вызван: неправильным определением границ анализируемого высказывания; несоблюдением пользователем ограничений, обычно предъявляемых к структуре языка и качеству произнесения (как, например, примешивание к речи посторонних звуков и слов). Замечено, что ошибки второго рода характерны большей частью для непрофессиональных пользователей.

Сопоставляя возможности предложенной системы с этими фактами, отметим, что используемый в системе алгоритм не требует пред-

варительного определения границ слов; участки контрольной реализации, отличающиеся в известной мере от эталонов, автоматически пропускаются и непосредственно не влияют на качество распознавания системы.

На основании изложенного можно предположить, что применение данной системы распознавания слов в качестве одной из компонент речевого интерфейса связи с ЭВМ упростит диалоговое взаимодейст - вие человека с машиной, благодаря сокращению числа ошибок, вызываемых профессиональной неподготовленностью пользователей.

Литература

I. Методы автоматического распознавания речи: В 2 т. Пер. с англ. /Под ред. У.Ли. - М.: Мир, 1983.

2. ВАДОВА З.А., ВЫСОЦКИЙ Г.Я.,ПЯТКОВ В.С.,ТРУНИН-ДОНСКОЙ В.Н. Аппаратурно-программная система автоматического понимания речи //Автоматическое распознавание слуховых образов: Материалы Всесоюз. школы-семинара (АРСО-10).-Тбилиси, 1978. - С.108-110.

3. БИАТОВ К.М., ВИНЦЮК Т.К. Система смысловой интерпретации слитной речи //Автоматическое распознавание слуховых образов:Тр./ Тез.докл. 12-го Всесоюз.семинара (АРСО-12).-Киев,1982.-С.365-368.

4. ВИНЦЮК Т.К. Анализ, распознавание и интерпретация речевых сигналов. - Киев: Наукова думка, 1987. - 262 с.

5. КЕЛЬМАНОВ А.В., НАУМОВ Б.Д., ХАМИДУЛЛИН С.А. Спецпроцессор,реализующий алгоритм динамического программирования //Автоматическое распознавание образов: /Тез.докл. 12-го Всесоюз.семинара (АРСО-12).-Киев, 1982. - С.456-458.

6. ЛЕВИНСОН С.Е. Структурные методы автоматического распознавания речи //ТИИЭР. - 1985. - Т.73, № 11. - С.100-129.

7. СВИРИДЕНКО В.А. Обобщенная процедура нелинейного согласования для систем распознавания речи //Автоматическое распознавание слуховых образов: /Тез.докл. и сообщ. 14-го Всесоюз. семинара (АРСО-14). - Каунас, 1988. - С. 74-75.

8. ВИНОГРАДОВ С.В., КОСАРЕВ Ю.А. Экспериментальное исследование алгоритмов нормализации темпа речи //Там же. - С. 76-77.

9. КУЗАКОВ А.М., ЕГОРОВ А.И., ВОРОНОВ О.Г. Функциональное корректирующее окно при нелинейном сопоставлении речевых образов //Там же. - С. 78-79.

10. КАТО Я. Система распознавания связной речи фирмы NEC //Зарубежная радиоэлектроника. - 1980. -№4. - С. 107-120.

11. КЕЛЬМАНОВ А.В. Сравнительное исследование двух алгоритмов динамического программирования //Автоматическое распознавание слуховых образов: /Тез. 12-го Всесоюз. семинара (АРСО-12). -Киев, 1982. - С. 474-476.

12. СМИТ А.Р., САМБУР М.Р. Построение гипотез о словах и проверка слов для распознавания речи //Методы автоматического распознавания речи. -М., 1983. - С. 188-223.

I3. ТАРАБУНОВ И.М. Аппаратурно-программный комплекс для анализа и распознавания речевых сигналов на основе микро-ЭВМ //Автоматическое распознавание слуховых образов: /Тез.докл. и сообщ. I3-й Всесоюз. школы-семинара (АРСО-I3).- Новосибирск, I984.-С.I24-I25.

I4. НАУМОВ Б.Д. Программируемый цифровой процессор для анализа речевых сигналов //Автоматическое распознавание слуховых образов: / Тез. докл. и сообщ. I4-го Всесоюзн. семинара (АРСО-I4). - Каунас, I986. - С. 63.

I5. ЛИ У.А. Распознавание речи: прошлое, настоящее и будущее //Методы автоматического распознавания речи. -М.,I983.-С.65-I4I.

I6. ITAKURA F. Minimum Prediction Residual Principle Applied to Speech Recognition// IEEE Symposium on Speech Recognition.- Pittsburgh,1974. - P.181-185.

17. NAKAGAVA S. Speaker-Independent Phoneme Recognition in Continuous Speech by a Statistical Method and a Stochastic Dinamic Time Warping Method// Technical Report of Computer Science Department, Carnegie-Mellon University.-Pittsburg a.o.,1966.- P.34-37.

Поступила в ред.-изд.отд.
25 мая I987 года