АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И СИМВОЛЬНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ
(Вычислительные системы)
1991 год Выпуск 141

УДК 621.391

# КОМБИНИРОВАННЫЙ АЛГОРИТМ РАСПОЗНАВАНИЯ ФРАЗ СЛИТНОЙ РЕЧИ

В.М.Величко

## В в е д е н и е

В статье рассматриваются вопросы построения алгоритма распознавания ограниченного набора фраз проблемно-ориентированного языка в системе понимания речи.

Проблемно-ориентированным языком системы понимания речи в рассматриваемом приложении является язык упражнений диспетчера гражданской авиации при обучении на цифровом диспетчерском тренажере. К этому языку предъявляются весьма жесткие требования по фразеологии, т.е по лексике, семантике и прагматике, определяемые нормативными документами [1] и конкретной моделируе - мой воздушной обстановкой, что подробно описано в [2].

Для решения подобной задачи обычно используются методы распознавания слитной и квазислитной речи.

Под распознаванием квазислитной речи сегодня понимают два существенно отличающихся процесса: а) распознавание ограниченного набора слитно произнесенных фраз (так называемого "фра - заря"); б) распознавание фраз, составленных из слов ограниченного словаря и произносимых пословно, с четко выраженной паузой между словами (эта задача называется также задачей распознавания дискретной речи). Обе задачи в каком-то смысле сводятся к задаче распознавания изолированных команд. Однако имеются значительные отличия.

Для задачи "а" пользователь получает иллюзию общения на естественном (как правило, проблемно-ориентированном) языке, что существенно повышает комфортность работы с автоматическим устройством и, таким образом, улучшает эргономические характеристики системы речевого диалога. С точки зрения распознавания такая задача легче, так как длинные фразы содержат больше отличий друг от друга как по продолжительности сигнала, так и по параметрам первичного описания, что способствует повышению надежности и увеличению быстродействия системы. Хорошим примером такой системы является HARPY, разработанная в 1976 году под руководством Редди (университет Карнеги-Меллона, США) в рамках крупного проекта ARPA и рекламируемая обычно как одна из первых систем распознавания слитной речи с высокими показателями по надежности. Такие системы позволяют реализовать осмысленный диалог с автоматическим устройством с учетом семантики и прагматики, но они требуют жесткого сценария, чтобы не вызвать чрезмерного, а иногда и неконтролируемого роста объема фразаря. Методика построения и использования такой системы распознавания квазислитной речи весьма привлекательна при условии жесткой регламентации речевого диалога и возможности тщательной подготовки к его осуществлению (например, в рамках ограниченного упражнения на тренажере для диспетчера управления воздушным движением).

Для задачи "б" характерно пословное произнесение осмысленных фраз. Каждый этап распознавания сводится к распознаванию изолированной команды из ограниченного набора.Достоинствами по сравнению с задачей "а" являются: гораздо более широкий набор возможных фраз, ограниченный лишь синтактико-семантико-прагматическими правилами; существенное ограничение за счет этих правил подсловаря, распознаваемого при выборе очередного слова-претендента, и,как следствие - значительное повышение надежности и быстродействия каждого шага (правда, при значительном об-

щем замедлении процесса). Очевидный недостаток - необходимость неестественного произнесения фраз пословно, что быстро создает у пользователя чувство дискомфорта. Лучшей системой такого сорта в настоящее время, по-видимому, является система распозна - вания квазислитной речи фирмы IBM, позволяющая с высокой надежностью работать со словарем в 20000 слов на материале деловой переписки фирмы (развитие системы [3]).

Задача распознавания слитной речи на базе ограниченного словаря проблемно-ориентированного языка сосредоточила сегодня наиболее крупные исследовательские силы, вызвала к жизни новые подходы. Она сходна с задачей "а" распознавания квазислитной речи по естественности и удобству произнесения, с задачей "б" по использованию лингвистической информации и базового словаря слов проблемно-ориентированного языка, но существенно сложнее обеих задач в реализации из-за отсутствия акустических границ между словами во фразах и явления коартикуляции и изменения фонетического качества звуков в слитном произнесении по сравнению с изолированно произнесенными эталонами. Несмотря на широко объявленные экспериментальные результаты и даже наличие серийно выпускаемых устройств, она далека от разрешения в плане достаточно надежного и удобного применения, а также по потребляемым вычислительным ресурсам. Лучшим объявленным результатом является, по-видимому, работа фирмы BBN (США), проведенная под руководством Махула [10].

Таким образом, использование подхода "б" не устраивает нас из эргономических соображений, слитная речь в настоящее время не распознается с требуемой надежностью, а использование распознавания изолированных фраз (подход "а") не проходит из-за чрезмерных затрат вычислительных ресурсов, потому что несколько сотен фраз, позволяющие реализовать полезное упражне - ние, не помещаются в оперативную память ПЭВМ, не говоря уже о

программе и вспомогательных массивах, а время распознавания на стандартной отечественной технике неприемлемо.

Поэтому при разработке системы была применена комбинация методов распознавания изолированных команд и слитной речи. В качестве эталонов были применены изолированные слова, входящие в состав фраз упражнения. Эталоны фраз компилировались из эталонов составляющих фразы слов как их прямая последовательность. Такой прием позволяет сократить память, требующуюся для хранения эталонов, до вполне приемлемых размеров.

Рассмотрим последовательно характеристики наборов фраз и их специфику, собственно алгоритм распознавания, схему обучения, экспериментальные результаты и направления продолжения работ.

## 1. Характеристики распознаваемых наборов фраз

В качестве распознаваемого материала использовались ограниченные наборы слитно произносимых фраз - от нескольких десятков до нескольких сотен в зависимости от упражнения для диспетчерского тренажера. Количество фраз в рамках одного упражнения ограничено, а главное, строго фиксировано по составу. Система заранее знает весь набор фраз, возможных для произнесения. Специфика задачи позволяет сделать такое ограничение, так как при обучении диспетчера отклонение от требований фразеологических стандартов не разрешается и должно исправляться в процессе занятий (например, после отказа системы распознавать неразрешенную фразу). Специфика фразеологии упражнений состоит в том, что словарный состав реальных упражнений очень ограничен. Например, анализ фрагментов упражнений, предложенных для апробации системы и содержащих требуемое количество фраз, показал, что словарный состав, как правило, не превышает сотни слов. Так, словарь упражнения для зоны посадки составил 60 слов (словосо-

четаний) при количестве фраз 141; словари упражнений для зоны круга - 72 слова при 260 фразах и 90 слов при 273 фразах [2]; словарь упражнения для зоны подхода - 67 слов при 596(!) фразах; словарь упражнения для зоны районного центра - 63 слова при 375 фразах.

Разнообразие возможных продолжений фразы после окончания очередного слова обычно оценивается средним коэффициентом ветвления (хотя есть более адекватная оценка в виде коэффициента перплексии). Для проанализированных упражнений этот коэффициент лежит в пределах от 1 до 20, максимально может увеличиться до 40, а в среднем составляет меньше 10. Средняя длина фразы - около 3-4 слов или словосочетаний. Среди словосочетаний могут быть довольно длинные, например,"эшелон перехода 1200 по давлению 755" (9 слов). Такие словосочетания целесообразно рассматривать как единое целое, если в них нет меняющихся фрагментов (например, числовых). Диапазон длин фраз - от 1 до 13 слов или словосочетаний.

## 2. Алгоритм распознавания

Структура фразаря упражнения позволяет использовать тот факт, что многие фразы содержат одинаковые участки, которые могут распознаваться одновременно для нескольких фраз без дублирования вычислений. Например, упражнение для зоны диспетчера круга среди 273 фраз содержит 118 фраз-позывных. Различных первых слов среди них всего 6. Это значит, что в начале команды, при обращении к распознаванию позывного борта, возможно лишь 6 вариантов слов вместо нескольких десятков (от 30 до 90) фраз при стандартном подходе. Налицо экономия вычислительных ресурсов не только по памяти из-за замены эталонов фраз последовательностями эталонов слов (приблизительно с 22 Кбайт до 3 Кбайт для рассматриваемого упражнения), но и по быстродействию - соответственно тоже на порядок.

Такой подход имеет добавочные преимущества. Как известно, ошибки распознавания при использовании метода динамического программирования часто возникают из-за того, что случайный шум при оценке меры расстояния на одинаковых словах (фразах) может превысить разницу в расстояниях на отличающихся участках раз - ных слов (фраз). При использовании одинаковых эталонов слов в разных фразах они дадут в точности одинаковый шум, и вся раз - ница в расстояниях между фразами будет образована за счет фо - нетически различающихся слов. Этот фактор приводит к повышению надежности распознавания, особенно в длинных фразах, отличаю - щихся лишь одним словом (например, "86 121" - "86 123").

Попутно решается задача существенного сокращения времени обучения для нового диктора, которая теперь сводится к однора- зовому произнесению нескольких десятков слов-эталонов вместо нескольких сотен фраз. При среднем темпе обучения порядка 15-20 слов в минуту это для приведенных словарей потребует 3-7 ми- нут от каждого диктора при более высоком качестве обучения (подстройки под диктора).

Однако основным преимуществом, предопределившим выбор данного метода, была возможность в полной мере использовать лингвистические ограничения задачи. Отмеченное выше преимуще - ство комбинирования схем распознавания слитной и дискретной ре- чи состоит в возможности охватить весь диапазон разрешенных фраз заданного проблемно-ориентированного языка путем задания лингвистических ограничений. Лингвистические ограничения, по нашему мнению, должны удовлетворять следующим требованиям:

- все разрешенные фразы проблемно-ориентированного языка не должны противоречить формальным ограничениям;

- количество фраз, не разрешенных в проблемно-ориентиро - ванном языке, но разрешаемых формальными лингвистическими ог - раничениями, должно быть минимальным;

- ограничения должны формулироваться экономным способом с точки зрения: а) требуемой для их записи памяти; б) времени их использования в процессе распознавания.

Примером идеальной схемы лингвистических ограничений может служить схема для распознавания слитно произносимых чисел в диапазоне от 0 до 999. Ее практическая значимость не вызывает сомнений, так как позволяет фактически произнести любые числа в любом контексте (после добавления слов "тысяча", "миллион", "запятая" и т.д. на соответствующих местах и в соответствующих синтаксических формах). Схема предполагает деление всех слов, отражающих числа в указанном диапазоне, на пять фиксированных групп (0, единицы, десятки, сотни, ...надцать, т.е. 10-19) [6]. Для слов каждой группы задается перечень групп слов, которые могут за ними следовать. Схема разрешает все фразы, составляющие числа от 0 до 999, и не допускает ни одной фразы вне этого перечня. Она предельно экономна.

Однако в большинстве практически ориентированных случаев формальная запись лингвистических ограничений приводит к созданию довольно громоздкой грамматики, допускающей наряду с полным учетом разрешенных фраз существование большого числа дополнительных фраз. Так, в схеме проблемно-ориентированного языка диспетчера для зоны посадки [5], ориентированной на все возможные случаи и достаточно экономичной по памяти и быстродействию, возможны продолжения, допускаемые формальной схемой и запрещенные в реальной ситуации. В более сложных случаях ограничения играют еще более слабую роль, разрешают большее количество запрещенных для произнесения фраз и тем самым снижают надежность распознавания и быстродействия системы. В развитых системах распознавания слитной речи учет грамматических ограничений занимает до 95% общего времени распознавания, как, например, в прообразе системы Махула [10], выполненной в рамках вышеупомянутого проекта ARPA.

Полный перечень фраз снимает вопрос о грамматических ограничениях: сам перечень удовлетворяет первым двум сформулированным требованиям. При не слишком большом объеме фразаря (ориентировочно до 500-1000 фраз) этот способ достаточно экономичен. Дополнительным преимуществом служит возможность использовать стандартные приемы распознавания изолированных команд с известным эталоном (длина команды и ее составляющих) и возможностями отсечек по длительности контрольной реализации и ее акустическому сходству с эталоном [4].

При описании алгоритма будем исходить из изложенного метода конкатенации заранее записанных слов-эталонов в качестве эталонов допустимых в данный момент речевого диалога фраз контрольной реализации. Вопрос формирования слов-эталонов рассматривается в следующем разделе. Скажем лишь, что они хранятся в памяти ПЭВМ как последовательность номеров таксонов, записанных в кодовой книге при обучении системы распознавания [6].

В отличие от довольно громоздкой в программной реализации, хотя и быстрой и экономичной схемы с адаптивным коридором [4], была выбрана простая схема динамического программирования [7], позволяющая в полной мере использовать априорные сведения о фразе-эталоне и ее составляющих, а именно о расположении коридора вокруг главной диагонали матрицы расстояний, устанавливаемого по длине фразы, и о длительностях слов-эталонов в составе фразы для перехода от предыдущего слова к последующему.

Рекуррентные формулы алгоритма динамического программирования (в отличие от [7]) выбраны следующим образом:

$$S(i,j) = \min(S(i-1,j), S(i-1,j-1) + d(i,j), S(i,j-1)) + d(i,j), \quad (1)$$

где $S(i,j)$ - длина оптимального пути на матрице расстояний в точке $(i,j)$; $d(i,j)$ - расстояние между i-м сегментом эталонной фразы и j-м сегментом распознаваемой (контрольной) фразы; мет-

рика выбрана следующим образом:

$$d = \sum_{l=1}^{6} |x_l - e_l|; \qquad (2)$$

$X = (x_1,\ldots,x_6)$ - вектор признаков j-го сегмента контрольной фразы; $E = (e_1,\ldots,e_6)$ - вектор признаков i-го сегмента эталонной фразы, который берется из кодовой книги в соответствии с нумерацией этого сегмента [6]; $i = 1 \div M$, $j = 1 \div N$; M - длина эталонной фразы; N- длина контрольной фразы; S(M,N) - итоговая длина оптимального пути.

Начальные условия задаются следующим образом:

$S(i,0) = S(0,j) = 30000$ для $i = 1 \div M$, $j = 1 \div N$;

$S(0,0) = 0$.

Величина 30000 выбрана как эквивалент машинной бесконечности для 16-разрядного представления целых чисел, применяемых в ПЭВМ типа ДВК-3. Схема многократно описана, в частности, в [8] и близка к примененной в [7]. Отличие заключается в использовании меры расстояния вместо сходства, что требует минимизации функционала на матрице расстояний вместо максимизации в [7] и, следовательно, изменения схемы динамического программирования с 0-1-0 в [7] на 1-2-1, где числа обозначают коэффициенты при расстояниях (сходствах) в сумме (1) соответственно при переходе по матрице слева направо, по диагонали слева сверху - направо вниз и сверху вниз, что легко видно из формулы (1) по индексам при d.

Существенной особенностью, часто применяемой в системах распознавания, является использование коридора вдоль главной диагонали матрицы расстояний [7]. Ширина коридора выбрана фиксированной вдоль всей фразы и задается в режиме диалога с пользователем. В экспериментах обычно выбиралось значение 7 (плюс-минус 3 от главной диагонали матрицы расстояний). При вычисле-

нии элемента главной диагонали для различных фраз-эталонов используется таблица котангенсов - отношение длин фраз эталонов к длине контрольной реализации. Для ускорения вычислений в 16-разрядной машине котангенсы взяты нормированными к целочисленным значениям, чтобы использовать целочисленные операции. А чтобы избежать потери точности и не допустить переполнения, нормировка проводится каждый раз по максимальной длине фразы-эталона. Из этих соображений выбрана величина 32767 в формуле (4) как максимальное 16-разрядное целое число.

Конкретные формулы нормировки следующие:

- номер элемента строки матрицы расстояний между сегмен - тами контрольной фразы и k-й эталонной фразы для j-го сегмента распознаваемой фразы, соответствующей центру коридора вдоль главной диагонали матрицы

$$i = (j*ctg(k))/L; \qquad (3)$$

- нормировочный коэффициент

$$L = 32767/(\max(M(k),N) + 1); \qquad (4)$$

- нормированный котангенс для k-й эталонной фразы дли - ной M(k)

$$ctg(k) = L*M(k)/N; \qquad (5)$$

N - длина контрольной фразы.

На границы коридора накладываются некоторые ограничения, которые будут рассмотрены при описании алгоритма в целом.

Использование векторного квантования для представления эталонных фраз позволяет сократить вычисления за счет запоми - нания расстояний между текущим сегментом контрольной фразы и векторами из кодовой книги. При повторном появлении номера вектора из кодовой книги расстояние не вычисляется заново, а извлекается из таблицы. При запоминании табличных значений необ-

ходимо отметить те номера векторов, которые уже встречались. Чтобы не очищать массив встретившихся номеров при переходе к следующему сегменту контрольной фразы, был применен следующий прием. Первоначально массив заполняется нулями. Затем при обработке очередного сегмента элементу массива, соответствующему номеру вектора из кодовой книги, присваивается значение на единицу большее, чем для предыдущего сегмента. Таким образом, нет необходимости каждый раз обнулять массив. Достаточно сравнить содержимое ячейки массива с текущим значением присваемой величины. Обнуление массива происходит при достижении значением элементов массива предельной величины, например, 255 (что соответствует максимальному числу для 1 байта при однобайтных элементах массива). Указанный прием универсален, он позволяет экономно помечать для каких-либо целей элементы массива без затрат времени на очистку массива после каждого цикла. В описанной реализации требуется дополнительная память по 255 элемен - тов для однобайтного массива номеров векторов кодовой книги [6] и двухбайтного массива для запоминания расстояний. Схема поиска - простейшая из-за линейного расположения номеров эле - ментов.

Существенным элементом алгоритма является определение момента перехода от эталона одного слова к эталону другого (следующего во фразе). Этот момент вычисляется по длине эталонного слова и соответствует его концу внутри коридора. Возможна ситуация , когда внутри коридора при его ширине, увеличивающейся с увеличением j из-за наличия текущих эталонов фраз разной длины, могут одновременно присутствовать эталоны нескольких слов (особенно при коротких словах-эталонах, например, паузе-смычке перед взрывным в начале слова). Для запоминания теку - щих границ внутри коридора и их соответствия словам-эталонам применяется таблица, обновляющаяся при изменении словарного состава внутри коридора.

Для минимизации вычислений применяется отсечка по длительности бесперспективных фраз, т.е. фраз-эталонов, выходящих по длительности за пределы N*VMIN - N*VMAX, где VMIN и VMAX - коэффициенты максимального сжатия и соответственно растяжения фразы-эталона по сравнению с контрольной фразой (обычные значения - 0,8 и 2,0). Кроме того, применяется отсечка бесперспективных продолжений, набравших меру расстояния от контрольной фразы, отличающуюся от минимального расстояния на текущей строке матрицы расстояний больше чем на величину заданного порога. Эти виды отсечек соответствуют ранее экспериментально проверенным и описанным в [4] отсечкам по длительности и по глобальной мере расстояния. Отсечка по глобальной мере расстояния используется также для сужения коридора по сравнению с вычисленным по котангенсам значением путем отбрасывания граничных элементов коридора, выходящих за пороговые значения матрицы расстояний.

Для экономии памяти запоминается переменное число элементов матрицы расстояний для каждой одновременно обрабатываемой группы фраз-эталонов в зависимости от текущей ширины коридора группы (ширина может меняться как из-за разной длины фраз-эталонов,входящих в группу,так и из-за отсечек граничных элементов коридора).Количество групп фраз-эталонов также может меняться как из-за отсечек бесперспективных продолжений, так и из-за их размножения на границах слов.Под строку матрицы выделен фиксированный,достаточно большой массив,в котором помечается начало и ширина каждого коридора в порядке рассмотрения групп фраз.Текущие значения матрицы в очередном коридоре записываются в свободный конец массива.При достижении границы массива запись переносится в начало массива,которое к этому моменту содержит уже использованные,ненужные значения,что гарантируется достаточным размером массива. По сравнению с фиксированными (следовательно, максимально возможными) значениями ширины

каждого коридора и числа групп фраз-эталонов описанный прием дает значительную экономию, так как с увеличением числа групп фраз-эталонов при разветвлении фраз уменьшается (за счет отсечек) ширина коридоров и число исходных групп фраз.

В итоге алгоритм распознавания контрольной фразы длины N выглядит следующим образом:

1. По грамматическим правилам определяется перечень фраз, которые могут быть претендентами на распознавание в данном контексте.

2. По критерию длительности определяется подмножество K фраз перечня п.1 с длинами M(k), удовлетворяющими условию N*VMIN < M(k) < N*VMAX, являющихся множеством фраз-претен - дентов.

3. Для множества фраз-претендентов определяется набор слов-лидеров, стоящих на первом месте во фразах, и для каждого из них - набор соответствующих ему фраз (группа фраз).

4. Для каждой группы фраз определяются положение и ширина коридора по формулам (3)-(5) со следующими изменениями: а) левая граница коридора определяется по самой короткой фразе группы фраз, а правая - по самой длинной; б) левая граница кори - дора не может быть левее левой границы коридора этой группы на предыдущей строке матрицы расстояний; это уточнение связано с возможным сдвигом левой границы вправо в результате отсечки по порогу расстояния. Координаты (начало и ширина) коридора записываются в информационную таблицу.

5. Для каждого слова-лидера группы фраз реализуется схема динамического программирования по формулам (1),(2). При вычислении очередной меры различия определяется минимальное значе - ние этой характеристики для текущего слова-лидера и для всей текущей строки матрицы расстояний. После окончания схемы динамического программирования для группы фраз производится проверка на соответствие минимального значения меры различия для

этой группы порогу отсечки, определяемому по минимальному значению матрицы расстояния на предыдущей строке (для первой строки начальное значение порога отсечки задается нулевым). В случае несоответствия группа фраз отсекается и исключается из списка фраз-претендентов. В противном случае проверяются граничные (начальные и конечные) значения строки матрицы расстояний в текущем коридоре на соответствие тому же порогу отсечки, и элементы, не соответствующие порогу, исключаются из коридора с коррекцией текущих границ коридора.

6. После достижения внутри коридора конца очередного слова в эталонной фразе производится проверка на разветвление группы фраз и при необходимости формирование новых групп фраз с новыми начальными цепочками слов-лидеров и новыми положениями коридоров для каждой группы фраз. При этом корректируется информационная таблица для коридоров.

7. После достижения конца контрольной фразы определяется минимальное расстояние S(M(k),N) для всех оставшихся K эталонных фраз-претендентов и результатом распознавания объявляется фраза, соответствующая минимальному расстоянию. При этом производится проверка на соответствие этого расстояния абсолютному пороговому значению, задаваемому заранее и определяемому экспериментально, чтобы уменьшить вероятность срабатывания системы по акустической помехе или по ошибочно произнесенной запрещенной фразе.

## 3. Схема обучения

Схема обучения включает в себя формирование таблиц грамматических ограничений и формирование эталонов слов и словосочетаний, составляющих фразы-эталоны, в виде последовательностей таксонов, характеризующих акустические признаки речевого сигнала.

Формирование таблиц грамматических ограничений подробно описано в [2]. Повторим, что вручную с помощью экспертов формируется фразарь упражнения, затем формируются количество и составы групп фраз, затем определяется возможность следования групп друг за другом в составе команды. При этом использование двух фраз из одной группы в одной команде запрещается. Все данные о перечисленных грамматических ограничениях в виде таблиц заносятся в память ПЭВМ и сохраняются в виде файлов для дальнейшего многократного использования. Эти ограничения формируют список фраз-претендентов на распознавание в рамках одного акта распознавания. В качестве примера укажем, что из 273 фраз, составляющих фразарь упражнения для зоны круга, образуются 22 группы, отражающие различные типы команд и информации. Коэффициент ветвления по группам меняется от 0 (для групп, которыми команда заканчивается) до 21 - для групп, следующих после позывного. Для фраз коэффициент ветвления соответственно меняется от 0 до 155.

Программа учета грамматических ограничений формирует также словарный состав фраз и текст обучающей последовательности, состоящей из слов и фраз. Это самый важный элемент обучения данной системы. Рассмотрим его подробно.

Исходный фразовый материал вводится с помощью стандартного текстового редактора в память ПЭВМ. В программе формирования словарного состава фразаря производится анализ пробелов текста фразаря. Текстовые блоки, находящиеся между пробелами, образуют единицы словаря, из которых строится фразарь. Использование другого разделительного знака вместо пробела позволяет использовать, кроме слов, словосочетания, в том числе устойчивые для данного фразаря, что позволяет укрупнить единицы, сократить, быть может, их общее количество и избежать ненужного дробления. Соответствующий пример приводился в первом разделе статьи. Состав фраз фразаря в единицах словаря (будем далее называть

эти единицы словами) записывается в специальный файл со своей информационной таблицей (адреса и длины фраз).

Затем автоматически проводится анализ положения слов во фразе и фонетический анализ стыков между словами. Целью такого анализа является учет изменения фонетического качества слов в составе фразы по сравнению с изолированным произнесением этих же слов вследствие изменения ударности и явления коартикуляции на границах слов, в частности, появления смычек внутри фразы перед словами, начинающимися с взрывных фонем.

Были приняты следующие упрощенные гипотезы, полученные на базе предварительных результатов по автоматизации выбора текста обучающей последовательности [9].

Во-первых, слова в конце фразы сохраняют ударение, имеющееся в изолированном произнесении. Во-вторых, слова в начале и середине фразы изменяют характер ударности и требуют для учета этого явления дополнительного эталона, вырезанного из словосочетания, где требуемое слово стоит не на последнем месте. Предполагается также, что характер изменения ударности для этих случаев одинаков, что верно лишь приближенно. В-третьих, словам, начинающимся с глухих или звонких взрывных и не занимающим первую позицию во фразе, должна предшествовать (искусственно создаваемая) смычка.

Легко заметить, что в гипотезах не учитывается в явном виде явление коартикуляции. Частично это упрощение компенсируется соответствующим выбором эталонов в слитном произнесении. Основой для полного решения проблемы может служить работа [9].

Существенной особенностью адаптивных систем распознавания речи, использующих в качестве признаков акустические характеристики сигнала без перехода к пофонемному распознаванию, является необходимость обучения для каждого нового диктора (подстройка под диктора), а также при переходе от одного блока выделения признаков к другому, как правило, путем произнесения

обучающей выборки, зависящей от языка системы распознавания. В зависимости от характеристик блока выделения акустических признаков и алгоритма определения границ слова могут быть погрешности в признаках на концах изолированно произнесенных слов и существенные отличия от произнесения тех же слов в слитном потоке. Предполагается, что эти погрешности могут быть в значительной степени скорректированы за счет автоматически формируемого дополнения к обучающей выборке.

С целью определения дополнительных эталонов, требующихся для коррекции характера ударности, проводится автоматический анализ положения каждого слова из словаря во фразе. Отмечаются слова, целиком составляющие фразу,- они в дальнейшем анализе не участвуют. Для остальных слов отмечаются начальная, конечная и средняя позиции. Одни и те же слова в разных фразах, естественно,могут встречаться в разных позициях. Слова в начальной и средней позициях помечаются для включения добавочных эталонов.

Затем производится автоматический анализ всех слов на возможные изменения за счет коартикуляции. Для этого все буквы в словах разбиваются на 6 групп в зависимости от предполагаемых изменений соответствующих им фонем.Первая группа - гласные и [р] - они,как правило, не изменяются в зависимости от окрущающих фонем и надежно выделяются на границах слов. Вторая группа - согласные, плохо выделяемые на границах слов,- это глухие и звонкие взрывные, у которых в начале слов не выделяются смычки, а в конце слов может быть пропуск взрыва и аспирации. Третья группа - согласные, не выделяемые в конце слов и перед 2-5-й группами в начале слов. Четвертая группа - звонкие согласные, перед которыми не выделяется начальная согласная из 3-й группы. Пятая группа - глухие согласные, перед которыми не выделяется начальная согласная из 3-й группы. Шестая группа - незначащие буквы - (эквивалент пробела), [ь].

Для лучшей ориентации в определении состава групп фонем использовалась вспомогательная программа параллельного вывода на экран признаков контрольной и эталонной фраз для сравнения степени редукции и переходных явлений на границах слов.

Приведем пример состава групп букв для конкретной платы и конкретного алгоритма выделения границ слова, основанного только на энергетических характеристиках акустического сигнала (инвентарь русских букв экранного редактора в порядке числовой кодировки - _,ю,а,б,ц,д,е,ф,г,х,и,й,к,л,м,н, о,п,я,р,с,т,у,ж, в,ь,ы,з,ш,э,щ,ч);

1-я группа - ю,а,е,и,о,я,р,у,ы,э;

2-я группа - б,ц,д,г,к,п,т,ч;

3-я группа отсутствует (для других плат это могут быть, например, в,ф,х);

4-я группа - л,м,н,ж,в,з;

5-я группа - ф,х,й,с,ш,щ;

6-я группа - _,ь.

В зависимости от положения (или положений) слова во фразе и групп его граничных букв формируются двух- (при необходимости трех-) словные сочетания в виде искусственных фраз, главным требованием к которым является, кроме нужного положения во фразе слов, формирующих дополнительный эталон, надежное выделение эталона из слитного словосочетания за счет контрастного сочетания букв на границах слов. Этим объясняется наличие разных групп для в общем-то однотипных по поведению звонких и глухих согласных групп 3-5. При формировании фраз учитываются все различающиеся сочетания перечисленных факторов. Таких сочетаний оказалось 11. Фразы дополняют список изолированных слов. При процедуре обучения диктор по запросу с терминала читает сначала изолированные слова, а затем фразы, составленные из них для формирования дополнительных эталонов. Согласно методике [6] из этих фраз вырезаются дополнительные эталоны и

также включаются в обучающую выборку для использования при распознавании.

Затем, как и в [6], проводится таксономия и другие стандартные для системы распознавания операции, описанные ранее.

Результаты ограниченной экспериментальной проверки алгоритма описаны в работе [11]. По отдельности все описанные ключевые особенности алгоритма проверены экспериментально и подтвердили свою работоспособность. Это, конечно, не исключает появления в процессе комплексной проверки неучтенных аспектов. В настоящее время комплекс, использующий описанный алгоритм, устанавливается во внедряющей организации для проведения его широкомасштабных испытаний с целью определения надежности, эргономических характеристик, необходимых доработок в условиях, приближенных к реальным. Приближение заключается в том,что работать будут квалифицированные эксперты-преподаватели, знающие тренажер, но не имеющие (как и учащиеся) навыка работы с распознающими системами.

## 4. Направления дальнейших работ

Предполагается избавиться от упрощающих гипотез и формировать обучающую выборку с полным учетом изменений слов в слитном произнесении по сравнению с изолированным. Другим важным моментом является отказ от априорного формирования фразаря в явном виде. Он будет задаваться грамматикой. Таким образом,будет практически реализован алгоритм распознавания слитной речи, но на базе ограниченного фразаря с вытекающими преимуществами для системы распознавания и существенным комфортом для пользователя системы.

Автор выражает свою благодарность Л.С.Юдиной за консультации и обсуждение вопросов изменения фонетических характеристик речи при переходе от изолированного произнесения к слитно-

му. Автор также выражает свою благодарность коллективу исследователей речи НГУ и Института математики СО АН СССР, которые участвовали в обсуждении работы и реализации системы распознавания, которая включала описанный алгоритм.

## Литература

1. Правила и фразеология радиообмена при выполнении полетов и управлении воздушным движением.-М.: Воздушный транспорт, 1987.

2. ВЕЛИЧКО В.М., ГАЛЬПЕРИН Б.Т., ЧУВАКОВ В.П. Понимание речи на базе фразеологии диспетчера гражданской авиации //Методы обработки символьных последовательностей. - Новосибирск, 1989. - Вып. 132: Вычислительные системы. -С. 161-176.

3. ДЖЕЛИНЕК Ф. Разработка экспериментального устройства, распознающего раздельно произносимые слова//ТИИЭР.-1985.-Т.73, № 11. - С. 91-100.

4. ВЕЛИЧКО В.М. Минимизация вычислений в распознавании речи //Анализ символьных последовательностей. - Новосибирск, 1985. - Вып. 113: Вычислительные системы. -С. 123-132.

5. ВЕЛИЧКО В.М., ЖИДОВИНОВ А.Ф., ЗАГОРУЙКО Н.Г. и др. Система понимания слитной речи на базе ЕС ЭВМ //Автоматическое распознавание слуховых образов (АРСО-13), тезисы докладов 13-й Всесоюзной школы-семинара. - Новосибирск, 1984. -С. 272-273.

6. ВЕЛИЧКО В.М. Обучение в распознавании слитной речи //Анализ текстов и сигналов. - Новосибирск, 1987. - Вып. 123: Вычислительные системы. -С. 101-110.

7. ВЕЛИЧКО В.М., ЗАГОРУЙКО Н.Г. Распознавание 200 устных команд //Автоматическое распознавание слуховых образов. Тезисы докладов 5-го Всесоюзного семинара (АРСО-5, Сухуми, 1969 г.). Труды Акустического института. - М.: Вып. XII. - 1970.

8. SAKOE H., CHIBA S. Dinamic programming algorithm optimization for spoken word recognition //IEEE Trans. Acoust. Speech Signal Process.-1978. -Vol. ASSP-26, N 1. -P. 43-49.

9. ВЕЛИЧКО В.М., САЛОМАТИНА Н.В., ЮДИНА Л.С. Автоматизация выбора обучающей последовательности при распознавании слитной речи //Автоматическое распознавание слуховых образов: Тез. докл. 15-го Всесоюзн.семинара АРСО-15 . - Таллин, 1989.-С.265-266.

10. CUBALA F., et al. Continious Speesh Recognition Re - sults of the BYBLOS System on the DARPA 1000-Word Resource Management Database //IEEE ICASSP-88, paper s7.8.

11. Макет системы понимания речи на базе фразеологии диспетчера гражданской авиации /Величко В.М., Гальперин Б.Т., Натансон Н.Г. и др. //Автоматическое распознавание слуховых образов: Тез. докл. 16-го Всесоюзн. семинара (АРСО-16). Москва, 1991. - С. 21-23.

Поступила в ред.-изд.отд.

3 сентября 1991 года