На правах рукописи

УДК 519.95

МАЗУРЕНКО Иван Леонидович

АВТОМАТНЫЕ МЕТОДЫ РАСПОЗНАВАНИЯ РЕЧИ

01.01.09 - дискретная математика и математическая кибернетика

Д И С С Е Р Т А Ц И Я

на соискание ученой степени

кандидата физико-математических наук

Научный руководитель:

доктор физико-математических наук, профессор БАБИН Д.Н.

Москва-2001

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Образное восприятие - одно из свойств мозга, позволяющее правильно воспринимать информацию о внешнем мире. При этом происходит классификация воспринимаемых ощущений, т.е. разделение их на группы похожих образов. Образы обладают тем свойством, что ознакомление с некоторым конечным числом их реализаций дает возможность узнавать все остальные представители образа. Более того, разные люди, обучающиеся на различном материале наблюдений, одинаково распознают одни и те же объекты.

Процессу распознавания образов всегда предшествует процесс обучения, во время которого мы знакомимся с некоторым числом примеров, зная наперед об их принадлежности к каким-то образам. У человека процесс обучения заканчивается успешно. Следовательно, образы - это объективные характеристики внешнего мира, а не произвольные наборы изображений. Эту объективную характеристику образов и стремятся выявлять с помощью математических методов. Создаются аппаратные и программные распознающие системы, на вход которых подается информация о предъявляемых объектах, а на выходе отображается информация о классах, к которым отнесены распознаваемые объекты.

Две фундаментальные гипотезы, лежащие в основе работы большинства систем автоматического распознавания речи (АРР), заключаются в следующем ([1, 6]):

1) информация в речевом сигнале переносится в виде изменений во времени его амплитудного спектра

2) речь - это сложный иерархически организованный сигнал, так что более простые образы одного уровня однозначно определенными правилами объединены в более сложные образы следующего уровня (аллофоны, фонемы, дифоны, трифоны, слоги, слова, предложения и т.п.)

Все это позволяет эффективно использовать автоматные методы[34] для описания речевых сигналов.

При этом решаются задачи синтеза порождающих образ автоматов (детерминированных и вероятностных), распознавания объектов, оценки сложности этих алгоритмов. Заметим, что участвующие в задании объектов массивы чисел хотя и конечные, но необозримые, и не представляется возможным хранение достаточного запаса примеров этих объектов. К примеру, одна секунда речевой записи занимает около 10000 байт, а сами примеры этих записей, как бы мы их не накапливали, не только не повторяются и не обладают близостью в 10000-мерном пространстве, но даже имеют принципиально разную длину.

Практически все известные методы распознавания речи обладают рядом основных общих свойств:

- для распознавания используется метод сравнения с эталонами;
- сигнал может быть представлен либо в виде непрерывной функции времени, либо в виде слова в некотором конечном алфавите;
- для сокращения объема вычислений используются методы динамического программирования.

Методы распознавания речи можно условно разделить на две большие группы — непараметрические — с использованием непараметрических мер близости к эталонам (к ним можно отнести методы на основе формальных грамматик и методы на основе метрик на множестве речевых сигналов) и параметрические (вероятностные - на основе скрытых марковских моделей, нейросетевые).

Первые разработки в области создания устройств автоматического распознавания речи можно отнести к концу 40-х годов прошлого столетия. Первые устройства ([2, 3, 4]) были аналоговыми и использовали пороговую логику. Исследователи сразу же столкнулись с тем, что распознавание речи — достаточно сложная задача в силу неравномерности

растяжения речевых сигналов во времени и сложной связи сигнала с фонемной структурой произносимой речи. В силу этого первые системы распознавания речи не обладали высокой надежностью и были узкоспециализированными.

Ситуация резко изменилась к началу 70-х годов в результате послевоенного развития электроники и вычислительной техники и появления лингвистической теории речи, представляющей устную речь как производную фонетической транскрипции произносимого текста. К числу первых проектов такого рода можно отнести проект Агентства перспективных исследовательских работ ARPA ([5]). В этих системах сигнал разбивался на сегменты, которые обозначались фонемоподобными символами, а затем получившийся код сигнала распознавался с помощью формальных грамматик. Узким местом такого подхода оказалось то, что сегментация и локальное распознавание является задачей, трудно поддающейся точному автоматическому решению, несмотря на то, что человек с этой задачей вполне справляется ([7]).

Следующим этапом в развитии теории распознавания речи стало развитие непараметрических систем, основанных на мерах близости на множестве речевых сигналов как функций времени. Революционный вклад в развитие этого направления оказал подход Винцюка (1960-е г.г., [8]), предложившего использовать новый для того времени метод динамического программирования (Беллман, 1950-е г.г. [9]) для быстрого вычисления меры близости между двумя функциями, задающими изменение во времени параметров речевых сигналов. Подход Винцюка, развитый Итакурой ([10]) и другими исследователями, позволил сократить время вычисления значений функции близости к эталонам с экспоненциального (от длины сигнала) до квадратичного. В силу того, что основной спецификой метода являлось нелинейное искажение временной оси одной из сравниваемых функций, метод получил название «динамической

деформации времени» (ДДВ).

Метод ДДВ обладает рядом достоинств и недостатков. К очевидным достоинствам метода относятся простота его реализации и обучения -для работы метода достаточно в качестве эталона команды использовать одно из ее произнесений. Основным недостатком метода является сложность вычисления меры близости (пропорционально квадрату длины сигнала) и большой объем памяти, необходимый для хранения эталонов команд (пропорционально длине сигнала и количеству команд в словаре).

Ряд исследователей (Бейкер — CMU — система «Драгон» [11] и, независимо, Йелинек — IBM [12], 1970-е годы) для распознавания речи применили теорию скрытых марковских моделей (СММ, скрытых марковских процессов, вероятностных функций цепей Маркова), созданную Баумом и коллегами в конце 60-х - начале 70-х г.г ([13]). Скрытые марковские процессы представляют из себя дважды стохастические процессы — марковские цепи ([14]) по переходам между состояниями и множества стационарных процессов в каждом состоянии цепи. Основы теории СММ были опубликованы в ряде научных журналов ([13,19]), однако получили развитие среди разработчиков систем распознавания речи лишь после выхода серии обзоров, посвященных популярному изложению теории СММ ([15-18]). Для обучения моделей и вычисления функции близости к эталону (здесь — вероятности наблюдения слова на выходе СММ) был также применен метод динамического программирования (алгоритмы прямого-обратного хода [19], Баума-Уэлча, или ЕМ-алгоритм [20], Виттерби [21-22]).

Достоинствами метода СММ являются достаточно быстрый способ вычисления значений функции расстояния (вероятности) и существенно меньший объем памяти, необходимый для хранения эталонов (пропорционально количеству фонем, Трифонов и т.п. в языке), а основны-

ми недостатками — достаточно большая сложность его реализации, а также необходимость использования больших фонетически сбалансированных речевых корпусов (баз данных) для обучения параметров СММ. Последнее обстоятельство привело к тому, что вплоть до начала 90-х годов методы ДДВ были более эффективными и точными при решении практических задач, чем методы СММ.

В настоящее время распознавание речи на основе СММ переживает период бурного роста. Десятки крупных коммерческих компаний (IBM, Dragon, L&H, Philips, Microsoft, Intel...) создали и активно развивают коммерческие системы распознавания речи. Эксперты в области компьютерных технологий называют распознавание речи одной из важнейших задач XXI века.

Основными характеристиками современных систем АРР являются следующие:

—словари размером в десятки тысяч слов;

—распознавание слитной речи;

—возможность работы как с предварительной настройкой на голос диктора, так и без настройки;

—надежность работы 95-98%.

СММ, возникшие как обобщение цепей Маркова, тесно связаны с понятием вероятностного автомата. Вероятностные автоматы, впервые введенные в общей форме Дж Карлайлом (1963, [23]) и, независимо от него, Р.Бухараевым (1964, [24]) и П.Штарке (1965, [25]), представляют из себя в практическом плане устройства с конечной памятью, перерабатывающие информацию с входных каналов в выходные, переходы и выходы которых происходят на основе вероятностных законов ([26]). Скрытые марковские модели являются частным случаем вероятностных автоматов, а именно, вероятностными автоматами без входа. СММ, используемые в системах распознавания речи, обладают дополнительно тем свой-

ством, что на каждом такте работы автомата переход осуществляется в состояние с тем же или большим номером. Такие модели, предложенные впервые Бакисом ([27, 12]), называются лево-правыми (left-right), или моделями Бакиса. Автор предложил при изложении на автоматном языке называть соответствующие этим моделям вероятностные автоматы монотонными.

Распознавание речи с помощью автономных монотонных вероятностных автоматов сводится к следующему. Каждой речевой команде сопоставляется эталон в виде вероятностного автомата. Вычисляется вероятность наблюдения распознаваемого кодового слова на выходе каждого из таких автоматов, и выбирается автомат, максимизирующий эту вероятность. Соответствующая найденному автомату команда считается результатом распознавания.

Метод СММ на уровне распознавания отдельных команд (т.е. без учета контекстной зависимости слов внутри предложения на естественном языке) дает надежность распознавания, близкую к надежности распознавания речевых команд человеком. Это позволяет сделать вывод о том, что общая задача распознавания отдельно произносимых команд решается методом СММ почти с максимально возможной точностью. Тем не менее, в практических приложениях возникает необходимость построения систем распознавания речи с ограниченным словарем, которые должны обладать сравнительно более высокой надежностью работы.

Традиционно эта задача решается методом подбора словаря команд. Подбор словаря может осуществляться путем проведения экспериментов с распознаванием команд из различных словарей и сравнения результатов этих экспериментов, что является «дорогим» решением. Поэтому представляет интерес задача автоматического подбора словаря команд без проведения экспериментов для обеспечения требуемой надежности распознавания.

Для решения этой задачи автор предложил использовать функцию качества словаря команд, связанную со средней вероятностью ошибки при распознавании и вычисляемую непосредственно по эталонам команд без проведения экспериментов с помощью введеннной метрики на множестве монотонных вероятностных автоматов.

В статистике традиционно задача нахождения расстояния между случайными величинами решается на основе статистических мер близости типа расстояний Махаланобиса, Кульбака-Лейбнера, Бхатачария и др. ([30]). Однако, данные метрики не имеют эффективного обобщения на случай вероятностных автоматов. В настоящей работе предлагается новая адекватная метрика, которая вычисляется эффективно.

Целью настоящей работы является построение формальной математической модели для распознавания речи с целью исследования вопроса оценки качества работы различных автоматных алгоритмов распознавания речи для различных словарей команд.

В работе имеются следующие достижения:

1. Предложена формальная математическая модель для задачи распознавания речи, в которой выделены множества (алфавиты букв, звуков, сигналов и кодов) и действующие на них функции (транскрибирования, произнесения, описания, кодирования и распознавания). Данная модель охватывает все случаи распознавания речевых сигналов в произвольных внешних условиях. В эту модель укладывается, в частности, задача распознавания речи при наличии дополнительных неакустических источников речевой информации в условиях шумов.

2. Сформулирована постановка задачи распознавания речи в данной модели и предложена функция для оценки качества словаря команд для фиксированного алгоритма распознавания.

3. Предложен универсальный детерминированно-автоматный алго-

ритм распознавания в рамках введенной модели, для которого найдена нижняя оценка функции качества словаря команд.

4. Известный метод распознавания речи на основе скрытых марковских моделей переведен на язык вероятностных автоматов.

5. Введено новое понятие автономного монотонного вероятностного автомата. На множестве таких автоматов введены операции декартового и скалярного произведений автоматов и на их основе построена метрика, эффективно вычислимая по автоматам.

6. Введено понятие качества словаря команд для вероятностного случая, получена точная формула для вычисления качества произвольного словаря команд через метрику на множестве монотонных автономных вероятностных автоматов.

В первой главе вводится формальное определение речевой модели как математического объекта. Основой ее является понятие **звукового сигнала,** под которым понимается функция *s :* R —» R, непрерывная на некотором отрезке *[to,t{* Є R и принимающая вне этого отрезка нулевые значения. Переходя к дискретному времени, **дискретным звуковым сигналом** называется любое отображение *s'* : Z —*ї* К , такое что 3zo,^i Є Z : *s'(z)* = 0 *\/z $ [zQ,zI]*. Дискретные звуковые сигналы получаются из непрерывных с помощью операции **дискретизации:** *s'(z) = s(z • 5),* где *5* называется **периодом дискретизации.** Для простоты далее будем опускать штрих при обозначении звукового сигнала.

Носителем сигнала *s :* Z -> R, *s ф* 0 назовем отрезок [inf{z|s(z) *ф* 0},sup{z|s(2:) *ф* 0}], вне которого он принимает нулевые значения. Границы носителя назовем **началом** и **концом** сигнала *s* соответственно. Определим операцию **конкатенации** *S1S2* сигналов *s\,S2* с непересекающимися носителями как *(siS2)(z) = max(si(z),S2(z)).*

Далее дается понятие речевой модели, отдельно для случая непрерыв-

ного и для случая дискретного времени.

Понятие речевой модели в дискретном случае формулируется в определении **1.15.** Упрощенно это определение выглядит так:

Речевой моделью называется восьмерка (Б, *B,* 5, С, *d,* Г, П, г), где

- *B* — конечный алфавит звуков;
- *B* Є *B** — конечный словарь команд;
- *S* — множество дискретных звуковых сигналов — произнесений команд;
- *C* — конечный алфавит — кодовая книга описаний сигналов;
- П : *B* —*У* 2^s - функция произнесения (здесь 2^s - множество всех подмножеств множества 5); П(/5) — множество всех произнесений команды *(З;*
- *d : S* —>■ $(R^M)^*$ — функция описания речевых сигналов, сопоставляющая речевому сигналу *s* матрицу *Mxn,* где *M* - некоторая константа (размерность пространства описаний), а *n* линейно зависит от длины речевого сигнала. Каждый столбец матрицы описания задает точку в пространстве описаний R^M*;*

и

- Г : К —> *C* — кодирующее отображение, сопоставляющее каждому столбцу матрицы описания элемент кодовой книги *С.* Г естественно обобщается на слова как Г : $(Ш^м)^*$ —*У С*;*
- *г* : *С** —> 2^B — функция распознавания, определяемая как

$$г(_7) = \{/З Є В|_7 Є Г(гі(П(/З)))\}.$$

Обозначим через П(&о) множество сигналов, не являющихся произнесениями никаких звуков:

$$n(\&o) = s\backslash U^n W'$$
$$ЪеВ$$

Будем считать, что функция произнесения П удовлетворяет следующим свойствам:

1. *У/З Є В* П(/3)^0;
2. V/Зь /¾ Є Б А ^ /¾ =» ІОД) П 11(/¾) = 0;
3. V/З Є £ Vs € П(/3) s *ф* 0;
4. *УР Є В :/З* = Ьі6$_2$...6„ и п > 1 Vs Є П(/3)

   3s$_0$l, Si, S12, S2, •.., S$_n$-lnS$_n$S$_n$0 Є 5: S = S$_0$lSiSi2S$_2$...S$_n$-.i$_n$S$_n$S$_{n0}$, где

   Si Є 11(¾) для *г* = 1,2,..., n,

   s$_i$j Є 11(6,-) U 1%) U П(Ьо) для *i , j* = 1,2,..., *n,*

   SJO, s$_0$j Є П(_) U П(6») U П(6$_0$) для г = 1,2,..., п,

   _ — специальный символ алфавита (символ паузы)

   и все сигналы Sj, **Sij** имеют непересекающие носители.

Как правило, функция описания *d* строится в виде композиции оконного отображения, вырезающего с постоянным шагом *L* из сигнала временное окно в общем случае переменной длины, и функции локального описания сигнала, сопоставляющей каждому окну сигнала точку в пространстве описаний R^M.

Образом *V(b)* С *Ш.*^{м} фонемы *b Є B* как в непрерывном, так и в дискретном случае называется совокупность всех точек описания всех ее произнесений. Для дискретного случая

$$V(b) = \{_P Є\ Ш^м \backslash 3y \in d(H(b)) : y = ||yi|y2|...|p|.,.|y,,-i|y,,||\},$$

где ||yiI2/2I"-I2/п|| обозначает матрицу, состоящую из столбцов yi, y_2,..., y_n. В **утверждении** 1.6. показано, что для любых допустимых значений периода дискретизации *S* и шага описания *L* дискретный и непрерывный образы одной и той же фонемы совпадают. Это, по сути, означает эквивалентность понятий образа фонемы в непрерывном и дискретном случае, что позволяет, ничего не теряя в точности решения задачи, использовать для распознавания речевую модель с дискретным временем.

При этом можно считать, что мы рассматриваем описания сигналов с шагом 1.

Задача распознавания речи во введенной речевой модели состоит в восстановлении функции распознавания г. Содержательно, распознавание речи сводится к тому, чтобы по заданному кодовому слову *у Є С*,* про которое известно, что оно является кодом описания произнесения одной из команд словаря *В,* найти в словаре эту команду.

Задачу распознавания можно решать методом сравнения с эталонами. Пусть каждой команде /¾ *Є В* сопоставлен эталон е* - элемент некоторого множества *Е = {е*\, Є2,..., едт}- Содержательно эталон команды является компактным способом представления множества кодовых слов всех ее произнесений.

Пусть также на декартовом произведении *С* X Е* введена функция **расстояния** *р* : *С** X *Е* —> *R* между кодовыми словами и эталонами. Алгоритм *(Е,р)* распознавания речи методом сравнения с эталонами восстанавливает функцию распознавания г с помощью формулы *Ц_{Е,Р)}{I)* = $\text{Argmm}_{/3e}\text{B}(\text{p}(7,e_i),z = 1,2,..., N)$. Если *г ф* $\text{г}_{(\text{E})/э)}$, будем говорить, что имеют место ошибки распознавания.

Качество *q(E,p) (В)* словаря команд *В* для данного алгоритма распознавания *(Е, р)* естественно определить, как 1 — 2рош.> где рош. - среднее число ошибок распознавания команд из *В* алгоритмом *(Е,р).*

Важно заметить, что каждой команде /3 соответствует множество кодовых слов описаний всех ее произнесений Г(с?(П(/3))), которое является конечным подъязыком языка *С** всех кодовых слов. Автоматный подход к распознаванию речи состоит в том, чтобы представить язык кодовых слов каждой команды некоторым автоматом. В случае конечно-детерминированных автоматов мы погружаем язык кодовых слов команды в некоторый известный нам бесконечный регулярный язык, а в случае вероятностных автоматов каждому слову языка кодовых слов при-

писываем некоторое действительное число — его вероятность, что дает нам возможность рассматривать случай пересекающихся языков кодовых слов.

Во второй главе эталоны команд задаюся конечно-детерминированными автоматами. Построен алгоритм распознавания речи на основе конечных детерминированных автоматов, где множество произнесений команды представляется некоторым бесконечным регулярным языком, приведена конструкция автоматического порождения указанных регулярных язкыков, введена функция качества словаря команд и получена нижня оценка значения функции качества в детерминированно-автоматном случае.

Предложен следующий способ построения автоматов.

Пусть пространство описаний сигналов R^M разбито на m непересекающихся подмножеств

$$\mathbf{K^M} = \underset{i=l,m}{[\mathbf{J}} \quad \mathbf{Q},$$

C = {ci, C2,..., c_m}. Сопоставим каждому элементу *c* разбиения *C* символ *CI* и выберем алфавит *C* = {ci, C2, •••, $c_т$} в качестве кодовой книги нашей речевой модели.

Кодирующее отображение будем строить на основе разбиения *C.* А именно, для каждой точки *p* Є R^M определим Г(р) как Г(р) = с Є *C : p* Є с. Предположим, что при распознавании мы не умеем вычислять *Г(р)* для всех точек *p* Є R^M. Пусть, тем не менее, для каждой пары элементов разбиения *C* мы умеем "отличать их друг от друга". Более строго,

Отделяющей функцией для двух множеств с;, *dj* Є *C* назовем фун-цию $ф_{CiЧ}$: $Ш^м$ -» *C* U {?}, такую что Vр Є $Ш^м$

0° $ф_{cc}(p)$ = с;

1° $\phi_{CiCi}(p) = ipc_{jCi}\{p)$ (симметричность);

3° p Є $CI = ï\ \phi_{Ci}cM$ Є fa,?};

С помощью функции ϕ можно построить функцию локального распознавания кода описания сигнала:

Функцией локального распознавания Ф будем называть функцию Ф : $E^м$ —>• 2^{e7}, определенную для каждой точки р Є R^M следующим образом: Ф(р) = {с Є С : Vc' $\epsilon C,c'\ \phi c$ ϕ_{cc}> Є {с, ?}}.

Множество значений функции Ф порождает регулярный язык "близких "гипотез распознавания кода описания сигнала, описываемый регулярным выражением $\Gamma(y)$ — I?(Ф(I/I))Л(Ф(ï/2)) • • .$R(\hat{}(y_n))$, где матрица y = ||yi|y2I ■ • • |Уп||> ^{a} регулярное выражение R определено как $R(\{ai, a_2, \bullet\bullet, a_m\}) = (ai|a_2|... \backslash a_m)$.

Обозначим через $\{\%)$ регулярный язык, описываемый регулярным выражением $И$.

Справедливо **Утверждение 2.2.** о том, что множество Ф(р) "близ-ких"гипотез локального распознавания содержит всегда "правильную "гипотезу Г(р):

$$Vp£R^M \qquad \Gamma(p) Є \Phi(p).$$

Как следствие получаем, что для описания $y = d(s)$ любого сигнала s справедливо $\Gamma(y)$ Є (Г(у))-

Говорим, что c Є C *и.* c' Є C **отделимы** с помощью отделяющей функции ϕ, если $\backslash/p$ Є c U c' $\phi(p)$ Є {с, c'}.

Окрестностью $0(C)$ некоторого подалфавита C C C кодовой книги C будем называть множество всех букв с из C, не являющихся отделимыми хотя бы от одной буквы алфавита C.

Обозначим через $c(Ь)$ множество $c\{Ь) = \{c' Є C : c' \Gamma\backslash T>(Ь)\ \phi$ 0}.

Свойство 4 функции произнесения П позволяет нам ввести эталоны

команд в виде регулярных языков, порождаемых регулярными выражениями специального вида, которые строятся непосредственно по текстам команд.

А именно, **эталонным кодом** команды *(З* = 6i&2 • • • *К* назовем регулярное выражение

f (/З) = Д(0(с(_) U с(6о) U фi)))Д(О(с(Ь0))Д(О(с(Ьi) U *с{Ь0)* и с(Ьз)))Л(0(с(Ьа))) • • ■

... Д(0(с($Ь_п$_i)))Д(0(с(Ь,,_i) U с(Ьо) U с(Ь,,)))Д(0(ф,,)))Д(0(с($Ь_п$) U с(Ьо) U *с(b_h))).

Код r(d(s)) любого произнесения *s* Є П(/3) команды /3 лежит в эталоне этой команды - регулярном языке (Г (/3)).

Справедлива более сильная **теорема 2.1:**

В речевой модели, в которой функция произнесения обладает свойствами 1-4, для любой звуковой команды /3 = *bfa.. .$Б_п$* и описания ***у*** = d(s) ее произвольного произнесения *s* Є П(/3) имеет место вложе-ние (Г*(у))* С (Г(/3)).

Далее во второй главе показано, что если звуки объединить в классы эквивалентности с помощью отношения отделимости, и в качестве разбиения *С* пространства описаний рассмотреть образы этих класов при отображении *d* о Г, то каждый язык гипотез распознавания (Г(у)) будет состоять ровно из одного элемента - кодового слова Г (у).

Теперь автоматный алгоритм *(Е, р)* распознавания речи в модели *(В, В, S,* С, «i, Г, П, г) можно задать множеством эталонов *Е* = {ei, ег,..., e_m] *ЄI = (Г(/3I)),* и функцией расстояния *р,* определяемой для кодового слова 7 = Г(ф)),вЄП(А)как:

$$\text{I} \quad 0, \quad \text{если } 7 \in$$

$$\in I \text{ I} \quad 1, \quad \text{иначе}$$

Результат распознавания кодового слова *у* алгоритмом *(Е, р)* есть множество команд Г($я_{>/7}$)(7) = *{Р* Є I% Є (f (/3))}.

Справедливо **утверждение 2.4** о том, что в условиях модели для любого кодового слова 7 имеет место вложение г(7) С *Г($Е_{,Р}$)(I)*-

Функцию качества словаря для введенного детерминированно-авто-матного алгоритма распознавания *(E, p)* определим следующим образом. Если словарь *B* состоит из одного слова /3, то положим, что его качество равно 1 (естественно предположить, что в словаре из одной команды распознаватель не делает ошибок). Для каждого словаря *B — {/3,(3'},* состоящего из двух команд, определим функцию качества как

(а а'\ о 7ЄГ(<i(П(/3)иП(/3'))) |гдаоз)ипоз')))|

Пусть *S* — некоторое конечное множество, / : *S* х *S* —> *R.* Обозначим через *@s(f)* среднее значение функции / на *S* х *S:*

$$^{W5}\wedge I \sim\sim \quad \frac{|5|-(15|-1)}{2}$$

Теперь для произвольного словаря команд *B* определим его качество как среднее качество по всем парам слов из *B:*

$$Я(E,_P)\{B) = e_B\{q(_{E>P}))$$ Введем расстояние р(е, *e')*

между регулярными языками е и e' как

{ О, если е П e' ф 0 1, иначе Справедлива **Теорема 2.2** о нижней оценке качества словаря команд при распознавании детерминированно-автоматным алгоритмом *(E,p):*

$$<?(Я,Р)(£)>Є_В(/З).$$

В третьей главе эталоны задаются в виде автономных монотонных вероятностных автоматов.

Этот подход не является новым. В современных системах распознавания речи распознавание производится с помощью метода скрытых марковских моделей [18]. В основе этого метода лежит алгоритм, который можно перевести на язык вероятностных автоматов.

Формально, **автономный вероятностный автомат** — это тройка (С, Q,z/), где *С* - алфавит выходных символов; *Q* - конечный алфавит состояний автомата, *v* - функция, определенная на множестве состояний автомата *Q* и принимающая в качестве своих значений вероятностные меры на множестве *С* х *Q*, такая, что она разлагается в произведение *v* = 7Г • *P,* где 7г действует на множестве *Q* и имеет значениями вероятностные меры на множестве *Q,* а *P* действует на том же множестве *Q* и имеет значениями вероятностные меры на множестве *С* [26]. Содержательно *v{c,q'/q)* интерпретируется как условная вероятность перейти в состояние *q'* и выдать символ *с* при условии, что в предыдущий момент времени автомат находился в состоянии *q.* Функцию *и* можно задать двумя матрицами - квадратной ***m*** х ш-матрицей 7г = ||7rjj||: 7г^ = *7r(qj/qi),i,j* = l, m и прямоугольной ***m*** X п-матрицей *P* = ||^Ъ||: *Pік = P(ci/qi),i* = 1, *m, I* = 1, *к, к* = *C*\. Матрицы 7Г и *P* являются стохастическими (по строкам). Далее для вероятностных автоматов вида (С, Q, 7г • *P)* будем использовать обозначение *(С, Q,* 7г, *P).*

В каждый момент времени состояние автомата характеризуется вектором распределения вероятностей вида (pi,P2»Pз, • • • >Рт), где ***m*** = *Q*\, *Pi* — вероятность находиться в состоянии $, *2_] Pi* — 1- Будем называть

г=1,m

вероятностный автомат **инициальным,** если заданы распределения u^0 и v^F для его начального и финального состояний соответственно (обозначение (С, Q, 7Г, Р, u^0, v^F))-

Можно рассмотреть обобщение понятия вероятностного автомата, положив, что матрицы 7г и *P* не являются стохастическими, но состоят из неотрицательных элементов и сумма элементов в каждой их строке не превосходит 1 (такие матрицы будем называть **слабостохастическими).** Далее автоматы со слабостахостическими матрицами 7г и *P* также будем называть вероятностными.

Функционирование инициального автомата определяется следующим

образом. Начальное состояние определяется распределением z/°. На каждом шаге, находясь в некотором состоянии *q,* автомат сначала выдает некоторую букву с алфавита С, исходя из распределения вероятностей *P(q),* а затем переходит в следующее состояние, исходя из распределения вероятностей 7г(д). Поскольку матрицы 7г и Р не являются в общем случае стохастическими, будем считать, что на каждом шаге существует вероятность того, что автомат не выдал никакой буквы или не осуществил перехода. Если на некотором шаге автомат перешел в состояние *qi,* для которого _, $^{ni}J^{=}$ ^> считаем, что автомат завершил работу. В

этом случае буквы, которые автомат начиная с начального состояния последовательно выдавал на выход в процессе своей работы, образуют некоторое слово 7 над алфавитом *С.* Говорим в таком случае, что автомат "выдал"слово 7-

Если некоторое слово 7 = $c\backslash C2...c_n$ выдано автоматом (С, Q, 7г, Р, г/°, u^F), **вероятностью** этого слова назовем величину

$$p_A(_1) = u^G P(c_1)P(c_2)\text{-}...\text{^}P(c_n)(u^F)^T_:$$

где ^{m} означает транспонирование матрицы, а для каждого Q Є *С* P(Q) -квадратная *m* X *m* матрица и *P{c{)ц = -кц • Рц* — вероятность того, что находясь в состоянии $, автомат выдал букву с/ и перешел в состояние

Чу

Инициальный автономный вероятностный автомат *Л* = (С, Q, 7г, Р, z/°, *v* \Q\ = *mn* будем называть **монотонным,** если выполнены условия:

1. *iTij* = 0 при *i* > *j;*

2. i/° = (l,0,0,..., 0); 3.^ =

(0, 0, ..., 0, 1);

Теорема 3.1. утверждает, что для любого монотонного автономного

вероятностного автомата *Л* = (С, Q, 7г, *P, и*^{0}, *и*^{F}), для которого выполен-но условие *ттц* < 1 для всех *i* = l,m, события, связанные с выдачей автоматом различных слов из С*, являются несовместными.

Это позволяет сопоставить каждому автономному монотонному вероятностному автомату ***Л*** **стохастическую (слабостохастическую) словарную функцию** рд : *С** -» [0,1], вычисляющую вероятность выдачи слов из *С** автоматом *Л* и обладающую свойством

где сумма берется по всем словам 7г из С*, занумерованным в лексикографическом порядке. Каждую такую словарную функцию можно задать бесконечным вектором

$$(\text{РЬР2}, \ldots, \text{Р п}, \text{- - -}) \in h\ C\ h,$$

где *pi = рлЬI), h = {{pi,P2,* • ■ • *,Pn,* ■ • •) : *Yl* W < °°^'

$$\text{а /2} = \{(\text{pi,P2,} \bullet \bullet \bullet \text{,}Pn\text{,} \blacksquare \blacksquare \bullet) : \wedge 2\ \text{W}^{2} < °°\text{Ь}$$

Методу скрытых марковских моделей соответствует в нашей формализации алгоритм *(Е, р)* распознавания речи, в котором эталоны *Е* = {ei, Є2,..., едг} команд {/Зi, /¾) • • • > А/v} задаются в виде некоторых автономных монотонных вероятностных автоматов, синтезированных по примерам кодовых слов одним из известных методов[19-22], а расстояние между эталонами и кодовыми словами вводится как р(е^, 7) = 1 ~*PeXl)*-

Содержательно метод сравнения с эталонами в вероятностно-автоматном случае сводится к поиску команды, на которой достигается максимум вероятности выдачи распознаваемого кодового слова соответствующим этой команде эталонным вероятностным автоматом.

Качество словаря команд для алгоритма *(Е, р)* будем также вводить следующим образом. Для *В* вида *{/З}* положим *Я(Е,р){{@})* = 1- Д$^{\text{ля}}$ словаря из двух команд определим его качество как единица минус удвоен-

ная вероятность ошибки распознавания в этом словаре при условии, что распознаваемые кодовые слова выдаются автоматами, соответствующими каждой из команд, и автоматы включаются по очереди с одинаковой вероятностью 1/2. Для словаря, состоящего из более чем двух команд, определим его качество как среднее качество всех его подсловарей из двух команд.

Далее в третьей главе решается задача оценки качества словарей команд для алгоритма *(E, p).* С этой целью на множестве эталонов команд (автономных монотонных вероятностных автоматов) вводится метрика.

Если на множестве стохастических словарных функций метрика каким-либо образом введена, расстоянием между автономными монотонными вероятностными автоматами определяем как расстояние между соответствующими им словарными функциями.

Определения 3.8, 3.9, 3.11 задают три способа введения метрики на множестве стохастических словарных функций:

- *PI(PI,P2)* = $\sum_{7i\in C^*}$ IPi(Тг) -Р2ЫЗ;
- *P2(ръР2)* = i - ^2 $^{min}_{нес^*}$*(piM,P2bi));*
- *Р3[РЪР2)* = -j=*Ti* -------- ||Ї ~ 11 --- ||Ї >

где || • || — норма в *її.*

Все эти три функции расстояния обладают свойствами метрики. Показано также, что на множестве стохастических словарных функций первые две метрики совпадают с точностью до постоянного множителя.

Связь метрики р2 и вероятности ошибки при распознавании иллюстрирует Лемма 3.4., утверждающая, что $p_2(p_{e1},Pe_2)$ = 1 - 2p_0ш., где p_0ш. — вероятность ошибки распознавания в словаре {/31,/¾}.

Это позволяет выразить функцию качества словаря команд через расстояние между соответствующими командам эталонными вероятностны-

ми автоматами. **Теорема** 3.2 утверждает, что

$$Q(E,P)(B) = Q_B(P2)\text{-}$$

Далее доказывается, что метрика /¾ эффективно вычисляется по автоматам.

Для этого в определении 3.12 вводится понятие **декартова произведения автономных автоматов:**

Декартовым произведением $A = Лi$ x Ai автономных вероятностных

и

автоматов $A\backslash$ = (С, Qi, яi, $Pu\ ^{v}\backslash > ^{v}\backslash$)

$A2$ = (С, фг? я"2> Лb ^2°) $V2^{F}$) называем вероятностный автомат

^ = ($c^{f}, g_2 x g_2, 7r, 2, i/°, i/^{F}>$,

где 7Г-ШХ m-матрица, ***m*** = *rriimz, mi* = \Qi\, *i* = 1,2,

$^{P}(hj)k = (Pl)ik * (^{P}2)jk'i$

Аи = to^{0}). - (¾¾; *Лu)* = ("А ■ ("А-

Приведенной матрицей переходов 7Г вероятностного автомата (С, Q, тг, Л ^°, ^{F}) назовем m x m-матрицу, (г, ^)-й элемент которой определен как

$$1=1$$

Теорема 3-3 показывает связь между понятием декартова произведения *Л = Лi* x *Ai* автоматов *Лi* и *Ai* и скалярным произведением соответствующих этим автоматам словарных функций в \^\

(здесь *Е* - единичная матрица).

Используя известную связь между нормой и скалярным произведением (||v|| = v^V^))? получаем, как следствие, формулу для вычисления расстояния рз между автономными монотонными вероятностными авто-

матами *Лi = {C, Q_h* тг$_ь$ *P_h* z/Д *u/)* и *Л_2* = (С, *Q_2*, тт$_2$, Р$_2$, ^2°, ^2^F):

Рз(Л,Л) \ ^((E-П.гГ^1),^^-^)-^{1}),

m_2

(здесь *mi* = |Qi|, m_2 = IQ2I, тгix2 — приведенная матрица переходов автомата *Лi хЛ_2,* TTixi — приведенная матрица переходов автомата Лi х*Лi,* ^'2x2¯ приведенная матрица переходов автомата *Л_2 х Л_2)*

Окончательно, качество словаря команд *В* с помощью метрики рз определяется как

Приложения.

На основе изложенных в работе подходов написаны прикладные программы распознавания речи. Система распознавания речи с ограниченным словарем ([a2]) работает в операционной системе Windows и позволяет в реальном времени распознавать речевые команды из словаря объемом до 100 команд. Система позволяет вводить команды на любом естественном языке, поскольку для обучения эталонов используется поэлементный метод [33]. В качестве моделей эталонов система позволяет одинаково эффективно работать как с методом ДДВ, так и с методом СММ. Система использует алгоритм распознавания речи, основанный на конечно-автоматном подходе, изложенном во второй главе. В качестве алфавита классов звуков выбран алфавит {Л, О, /, _, *X}* ((см. приложение 2, [al]), где класс *X* включает шипящие согласные, _ — символы паузы и глухие смычки, а классы *A,0:1* соответствуют низкочастотным, среднечастотным и высокочастотным гласным звукам соответственно.

Конструктор систем распознавания речи в условиях сильных акустических шумов ([a4]) также работает с раздельно произносимыми командами и фразами из ограниченного словаря. В качестве устройства ввода

для конструктора могут подключаться как обычные микрофоны, так и системы из различных датчиков: акустических, фотодатчиков слежения за губами говорящего, датчиков дыхания и т.п. В отличие от описанной выше системы распознавания, конструктор позволяет динамически строить функции описания сигналов на основе сигналов с датчиков с помощью богатого семейства унарных и бинарных операций. Конструктор позволяет строить также устойчиво работающие модули выделения границ речевых сообщений на фоне шума, что повышает в сотни раз надежность распознавания речи в шуме по сравнению с традиционно используемыми методами распознавания. Работа конструктора основана на описанной в первой главе формальной речевой модели.

Система автоматического подравнивания произношения является дополнением к мультимедийным системам обучения иностранным языкам. После настройки на голос ученика электронный учитель позволяет определить количество, место и тип ошибок при произнесении учеником фраз на иностранном языке. При построении системы использовалась оригинальная методика построения модели «правильного произношения» для ученика в виде монотонного вероятностного автомата.

Метрика на множестве автономных вероятностных автоматов, введенная в третьей главе настоящей работы, была эффективно использована при решении задачи оптимального выбора фонетического алфавита при разработке системы распознавания русской речи в рамках гранта с фирмой Intel Corp., США ([35]). С помощью метрики *p$* была построена матрица попарных расстояний между фонемами русского языка (см. приложение 3), представленных в виде автономных вероятностных автоматов с 4 состояниями, которые были синтезированы на основе русской речевой базы данных. На основе проведенного эксперимента удалось показать, что алфавит из 150 фонемных символов для русского языка можно сократить без потенциальной потери точности при распознавании до

120 символов, что может значительно увеличить эффективность системы распознавания речи на русском языке.

Введенная в настоящей работе метрика на множестве автономных монотонных вероятностных автоматов имеет широкие возможности практического применения. Она может быть использована в тех областях, где в настоящее время требуется уметь эффективно вычислять расстояние между марковскими автоматами: при решении задач подбора фонетического алфавита ([36]) и словаря команд ([38]), распознавания и идентификации диктора, конверсии и синтеза речи ([37] - здесь метрика используется для решения задачи интерполяции голоса диктора), при программировании стратегических компьютерных игр ([40]) и моделировании стратегий поведения человека в реальных условиях ([39]). Во всех этих практических примерах в настоящее время используются статистические меры близости типа расстояний Махаланобиса, Кульбака-Лейбнера, метод Монте-Карло и им подобные. Поскольку эти меры близости зачастую не обладают всеми свойствами метрики (не выполняется неравенство треугольника), введение метрики на множестве скрытых марковских моделей может дать более эффективные результаты при решении данных практических задач.

Автор выражает благодарность своему научному руководителю доктору физико-математических наук, профессору Бабину Дмитрию Николаевичу за постановку задачи и помощь в работе над диссертацией.

ГЛАВА 1. ФОРМАЛЬНАЯ МОДЕЛЬ РЕЧИ

§1.1. Речь как математический объект. Основные определения

В данном разделе будет введена формальная модель речи для естественного языка, на котором произносятся слова. Без ограничения общности можно считать, что рассматривается русский язык, хотя построенная модель инвариантна относительно выбора языка.

Пусть A - некоторый конечный алфавит *букв* языка. Пусть *A* -фиксированное подмножество множества A* слов в алфавите A. Назовем множество *A словарем команд,* а элементы множества *A - командами.*

Пусть B - некоторый алфавит, который будем в дальнейшем называть *алфавитом звуков языка,* или *фонетическим алфавитом* (фонетический алфавит для русского языка приведен в приложении 1). Будем считать, что алфавит звуков содержит некоторый специальный символ -символ паузы _. Пусть *Ъ* - фиксированное подмножество B* слов в алфавите B, такое что BcB. Назовем множество *Ъ фонетическим словарем,* или *словарем звуковых команд,* а элементы этого множества - *фонетическими словами {звуковыми командами).*

Через |(3| будем обозначать длину слова |3 (количество букв в нем), а через *R*, N и Z - множества действительных, натуральных и целых чисел соответственно.

Определение 1.1. *Звуковым сигналом* назовем функцию s: R->R, непрерывную на некотором отрезке [to, ti]czR и принимающую вне этого отрезка нулевые значения (см. рис 1.1).

Область определения звукового сигнала будем называть *временем,* а точки на ней — *отсчетами {моментами времени).*

Началом to(s) звукового сигнала s назовем момент времени $t_0(s)=\inf\{t: s(t)*0\}$, а концом tj(s) сигнала s - момент времени $ti(s)=\sup\{t: s(t)*0\}$. *Длиной* |s| сигнала s будем называть величину $|s|=ti(s)-t_0(s)$.

Через S обозначим множество всех звуковых сигналов.

С физической точки зрения каждый звуковой сигнал s задает изменение величины звукового давления на отрезке времени [to(s), ti(s)].

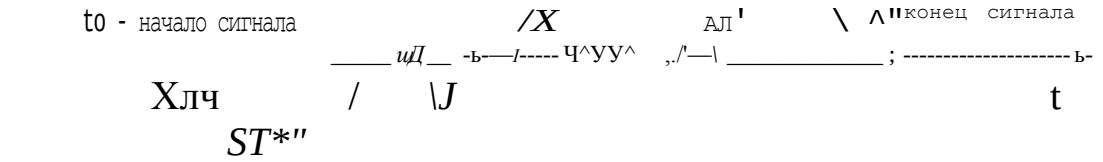


Рисунок 1.1. Звуковой сигнал

На множестве S введем отношение эквивалентности.

Определение 1.2. Сигналы si и S2 будем называть *эквивалентными,* если при всех teR выполнено $s_2(t)=Si(t+k)$ для некоторого фиксированного кєЕ.

Эквивалентность сигналов si и $\$_2$ будем обозначать как $Si\sim s_2$.

Через $S[_{UjV}]$ будем обозначать *ограничение сигнала s на отрезок* [u,v], т.е. звуковой сигнал, определенный следующим образом: S[u,v](t)=s(t) Vte[u,v] S[u,v](t)=0 Vte[u,v]

Напомним, что *преобразованием Фурье* Φ_8 функции s:R—>R называется функция O_s:R->C, $<D_s(f)= \int_{-00}^{+CO} s\{t)e^{\sim 2Kift} dt$ [29].

Определение 1.3. *Речевой моделью* назовем десятку

<A,ДЗ,В,8,т,П,Т,ц,г>, где

- A, B - конечные алфавиты букв и звуков соответственно;
- т4<zA* - конечный словарь команд;
- ScB* - конечный словарь звуковых команд;
- S - множество звуковых сигналов;
- т: A*—»B* - конечно-автоматное отображение *транскрибированим,*
- П: Б->2 - отображение *произнесения,*

обладающее следующими свойствами:

1°УРєВП((3)^0,

2° Рь Мб, MP_2 => П(Р,)пП(Р2)=0;

3° VPeB: |P|>1 Vsen(P) Φ_8(i>0 при |fj>F_{max}, где F_{max} - действительная константа (на практике F_{max}=7000 Гц), которую мы будем называть *максимальной частотой* речевых сигналов

4°УрєВ seII(P),Si~s=>Sien(P)

5° 3 константы T_{min},T_{max}: 0<T_{min}<T_{max}: VPeS Vsell(p), VteR ЗТе[T_{min},T_{max}], такое что звуковой сигнал $s_{[t}$.T/2,t+T/2] удовлетворяет свойству $Os_{[t}$.T/2,t+T/2](f)¯0 при |fl>F_{max}. 6° VPeB sell(P) => s^O (т.е. s - не тождественный 0)

- T: SxJR-»R, - функция, задающая размер *окна анализа,* T_s(t) - окно анализа сигнала s в момент времени t, определяемая как

$T_s(t)$=min{TG[T_{min},T_{max}]: $Os_{[t}$_T/2,t+T/2](i)=0 при |fl>F_{max}}.

Отрезок $[t-T_s(t)/2,t+T_s(t)/2]$ будем называть *окном анализа.*

- г: S—>2^S - *распознающая функция,* определяемая как r(s)={Pe'B: зєП(Р)};
- JJ,:SxN—»R - *отображение фонемной разметки* звуковых сигналов, такое что VpeB, $p=bib_2...b_m$, Vsell(p) выполнено 0=u(s,l)<u(s,2) <.. .<|a,(s,2m+2)=|s|, u(s,n)=0 при n>2m+2 и справедливы свойства:

1° Si~s => |u(si,n) = |u(s,n) VneN.

2° Vj=l,2,...,m Vt€[|ii(s,2j), |i(s,2j+l)] $s_{[t\cdot T(}s)/w(s)/2]$en(bj)

3° Vj=l,2,...,m+1 V tG[fj,(s,2j-l), (i(s,2j)], выполнено либо $s_№)/2\text{^}+T(s)/2]$en(b_H)un(bj), либо $s_{[t\cdot T}(s)/2,t+T(s)/2]\text{^}n(b)$ при всех ЬєВ (здесь bo = b_m+i =_ - символ паузы)

4° Если P=b - фонетическое слово из одной буквы, а sell(P) - произнесение этого слова, то fi(s,l)=u(s,2)=0, л(s,3)=|U(s,4)=|s|.

Будем называть звуковой сигнал s *произнесением* фонетического слова рєВ, если БєЩр).

Звуковые сигналы, являющиеся произнесением какого-то фонетического слова, будем называть *речевыми сигналами.* Далее, если противное не оговорено особо, символом S будем обозначать множество всех речевых сигналов.

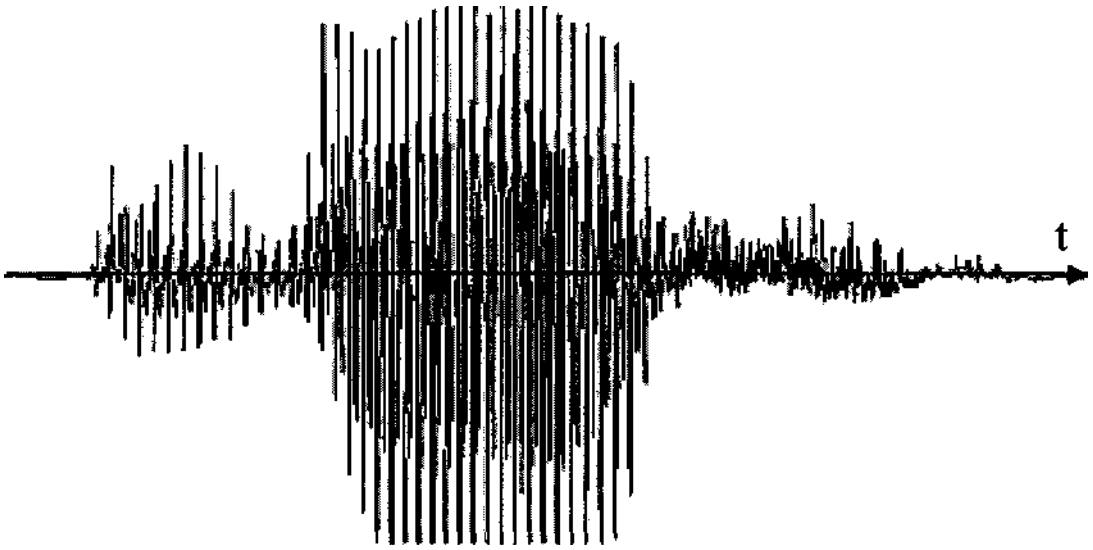


Рисунок 1.2 Пример речевого сигнала, являющегося произнесением слова «мыс»

Для каждого речевого сигнала s, являющегося произнесением некоторого фонетического слова (ЗеЗ, будем говорить, что временные отрезки вида [|Li(s,2i)$_5$|Li(s$_?$2i+l)] *соответствуют произнесению і-й фонемы* фонетического слова (3, а интервалы вида (|i(s,2i+l),|j,(s$_5$2i+2)) *соответствуют переходу от і-й к і+1-й фонеме* фонетического слова р (интервал (|LI(S,1)$_?$|LL(S,2)) будем называть *переходом от паузы к первой фонеме слова,* а интервал (|Li(s$_5$2|P|+l)$_?$jLi(s,2|p|+l)) - *переходом от последней фонемы к паузе).* Величину |Li(s,2i+l)-|Li(s,2i) будем называть *длительностью произнесения* і-й фонемы в речевом сигнале s, а величину |Li(s$_?$2i+2)-p,(s,2i+l) — *длительностью перехода* от і-й фонемы к і+1-й.

Рисунок 1.3. Речевая модель на примере команды "Пе'тя"

Содержательно функционирование речевой модели происходит следующим образом: команда аєД транскрибируется при помощи отображения т в звуковую команду реЪ, Р=т(ос), затем р "произносится" с помощью отображения П, в результате чего получается звуковой сигнал s, seli(P), как одно из возможных произнесений этого фонетического слова. Далее с помощью распознающей функции г можно восстановить, какое фонетическое слово было произнесено, а с помощью функции JU выполнить фонетическую разметку этого произнесения.

Функционирование речевой модели является формализацией процесса речеобразования для отдельного человека - носителя естественного языка. Считая, что каждый носитель языка произносит слова неизменным

образом (у него не меняется тип голоса, акцент, словарный запас и т.п.), можно считать, что для каждого отдельно взятого человека и фиксированного словаря команд речевая модель единственна. Поэтому каж-

дое из отображений т, П, г фиксировано для данной речевой модели и, в силу связи с реальными процессами произнесения слов и звуков и восприятия речи, обладает рядом описанных выше свойств, содержательный смысл которых состоит в следующем.

Отображение транскрибирования т задается правилами "озвучивания" отдельных слов естественного языка, поэтому для каждого типа акцента, диалекта и т.п. отображение т задано единственным образом. Для каждого естественного языка правила, задающие отображение т, известны. Пример таких правил для русского языка приведен в приложении 1.

Отображение произнесения П задает различные варианты произнесения отдельных звуков, сочетаний звуков, фонетических слов данным носителем языка. Поэтому отображение П единственно и обладает рядом естественных свойств.

Свойства 1°-6° содержательно означают, что у любого фонетического слова естественного языка есть хотя бы одно произнесение (1°); один и тот же речевой сигнал не может быть произнесением различных фонетических слов (2°); речевые сигналы содержат в себе только колебания с частотами меньше некоторой частоты F_{max} (3°); если какой-то речевой сигнал является произнесением некоторого слова, то все эквивалентные ему речевые сигналы также являются произнесениями того же слова (4°); для любого речевого сигнала в любой момент времени существует окрестность этого момента времени, имеющая длину в заданных пределах, которая содержит в себе только колебания с частотами меньше частоты F_{max} (4°); произнесение слова не может быть нулевым сигналом (5°).

Распознающая функция г моделирует процесс восприятия речи человеком — она восстанавливает фонемную транскрипцию слова по любому его произнесению.

Функция размера окна анализа T определяется как длина минимального окна, такого что сигнал, ограниченный на него, имеет ограниченный спектр.

Наконец, функция фонемной разметки (i позволяет в речевом сигнале выделить отрезки времени, соответствующие как произнесению отдельных фонем, так и переходу между фонемами. Эквивалентные сигналы имеют одинаковую фонемную разметку.

Зафиксируем речевую модель *(A,A$_j$B,B,S,x,Yl,T,[i,r).*

Определение 1.4. Функцию 0:SxR—»S, определенную как

0(s,t)(i)= $s(t + г), X \epsilon$ Ш $T_s(t)$

, будем называть *оконным отображени*-0, *иначе*

ем.

Определение 1.5. Отображение d:SxR.-»R^M, где d(s, t)=d'(0(s,t)), где d':S->R^M, - некоторая функция, будем называть *функцией описания* сигнала, а отображение d' - *функцией описания окон,* если при любом фиксированном s функция d(s,t) непрерывна по t на отрезке [t_0(s), ti(s)].

При фиксированном s функцию времени d_s: [t_0(s), tj(s)]->R^M, где d_s(t)=d(s,t), будем называть *описанием звукового сигнала* s. Согласно определению 1.5, d_s(t) - непрерывная M-мерная действительнозначная функция времени.

Определение 1.6. Функцию описания d=d'^{0}0 будем называть *нормальной,* если отображение d' инъекивно.

Описание речевого сигнала с помощью нормальной функции описания также будем называть *нормальным*.

Содержательно это означает, что нормальными мы будем называть функции описания, для которых по описанию речевого сигнала в данный момент времени однозначно восстанавливается звуковой сигнал на окне анализа.

Приведем пример функции описания звуковых сигналов, основанной на операторе преобразования Фурье.

Пример 1.1. В качестве примера функции описания возьмем отображение, сопоставляющее окну анализа вектор значений преобразования Фурье на равноотстоящих значениях частот $7_M F_0$, $^2/_M F_{0)}$... $^M/_M F_0$:
cГт = (<X>s(V_M % $Ф_8$ ($^2/_M$FO), ... Ф, (% F_0)), где F_0>F_{max}.

В §1.2 будет показано, что введенная таким образом функция описания для достаточно больших M является нормальной (утверждение 1.4).

Пусть s - некоторый речевой сигнал. Пусть d_s - описание этого речевого сигнала, т.е. функция зависимости M-разрядного вектора действительных чисел от времени. Рассмотрим все возможные произнесения всех фонем.

Определение 1.7. *Образом V(b) фонемы* ЬєВ для данной функции описания d будем называть множество D(b)={d_s(t) | БЄП(Ь), te[t_0(s), t,(s)]}ci^M

Утверяедение 1.1. Образы различных фонем для нормальной функции описания не пересекаются. Доказательство.

Пусть d - нормальная функция описания сигналов. Предположим противное. Тогда существуют две различные фонемы bi и b2, два сигнала Si и S2, являющиеся произнесениями этих фонем, и ДВа МОМеНТа Времени ti И t2, Такие ЧТО tie[to(Si)$_9$ ti(Si)], t_2G[t_0(S2), ti(S2)] и d(Sbti)=d($s_{2?}t_2$). Отсюда, учитывая инъективность отображения d', получаем: *$з_{t}$(u+x)=Б2($г_2$+x)* ViG[-T(t)/2, T(t)/2], где T(t)=min(T_{Sl}(t), Ts_2(t)), пусть, для определенности, T(t)=Tsi(t). По определению для любого речевого сигнала s и момента времени t T_s(t)=min{T: T>T_{min}, <E>($s_{[t}$-T/2,t+T/2])(f)=0 при £>F_{max}} => для сигнала s_2 Ts_2(t)<T(t) => Ts_2(t)=Tsi(t).

Однако по свойству 2° функции ^ каждый из сигналов Si(t), i=l,2, ограниченный на отрезок $\left|\ \frac{2}{2}\ \right|$, является произнесением соответствующей фонемы bi- Поскольку эти сигналы совпадают, по свойству 2° функции П должны совпадать и фонемы bi и $Ь_2$. Мы пришли к противоречию.

Утверждение доказано.

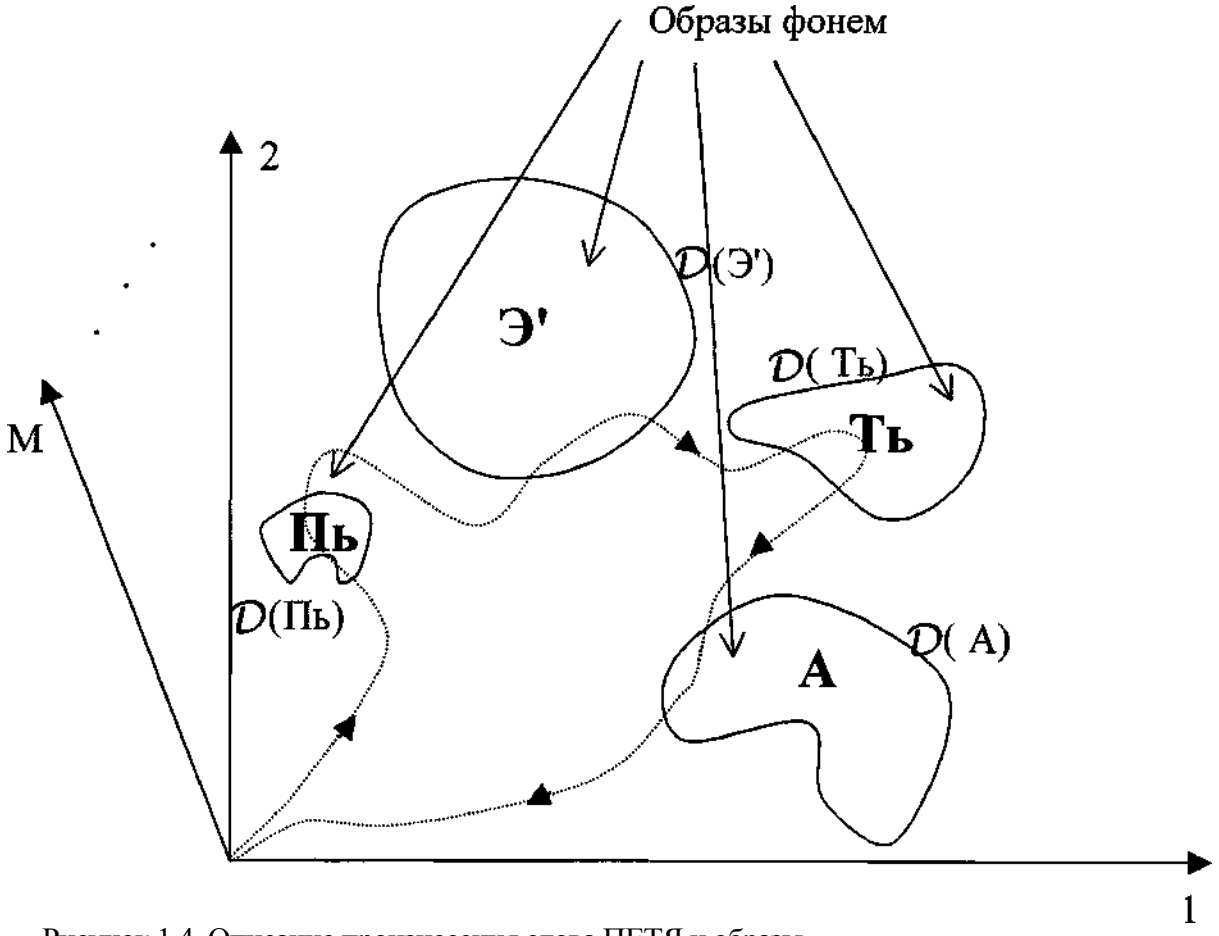


Рисунок 1.4. Описание произнесения слова ПЕТЯ и образы фонем из транскрипции этого слова для нормального оператора описания сигналов.

§1.2. Дискретизация

Перейдем теперь от непрерывного времени к дискретному. Зафиксируем некоторую действительную величину 8>0, которую назовем периодом дискретизации. Выделим на оси времени точки {nS+ф, neZ}, где ф - некоторая действительная константа (начальный сдвиг), т.е. точки, расположенные на расстоянии 5 друг от друга. Известная из теории связи теорема Котельникова [32] утверждает, что для того, чтобы по точкам, равноотстоящим друг от друга на величину 5, можно было однозначно восстановить сигнал s(t), необходимо, чтобы было выполнено условие <D_s(f)=0 при |fl>l/28. Если 8<l/(2F_{max}), то l/25>F_{max} => по свойству 3° функции произнесения П O_s(f)=0 при |f|>l/28.

Следовательно, любой речевой сигнал может быть восстановлен единственным образом по последовательности равноотстоящих отсчетов (з(п8+ф), neZ}, если l/S>2F_{max}, т.е. 8<l/(2F_{max}).

Замечание 1.1. Согласно теореме Котельникова[32], для восстановления звукового сигнала s по значениям s'(n) этого сигнала в точках {п8+ф, neZ} используют формулу:

$$\text{«O-I/OO^r^g^ (i)-} \quad \sin\frac{\pi(\text{л-п8-ф})}{6}$$

Определение 1.8. *Дискретным звуковым сигналом* будем называть любое отображение s': Z->R, такое что 3z_0, ъ\ GZ : s'(z)=0 Vzg[z_0, zj].

Понятия начала t_0(s'), конца ti(s') и длины |s'| дискретного звукового сигнала определяются аналогично соответствующим понятиям для непрерывных звуковых сигналов.

Обозначим через S' множество дискретных звуковых сигналов, а через R^+ - множество неотрицательных действительных чисел.

Определение 1.9. *(8, ф)-дискретизацией* назовем отображение

$Д5_{,ф}$: $SxR^+xR—>S'$, сопоставляющее звуковому сигналу s дискретный звуковой сигнал $S5_{,9}=A_{§!9}(s)$: \fneZ $S5_{,9}(n)=s(n8+9)$, где $8GR^+$ - период дискретизации, (peR- *начальный сдвиг* дискретизации.

Дискретным речевым сигналом (для фиксированных 8 и ф) будем называть любую (8,ф)-дискретизацию любого речевого сигнала. Будем считать, что период дискретизации 8 выбран так, что он гарантирует однозначность обратного восстановления непрерывного сигнала из дискретного (это возможно по свойству 3° функции произнесения). При этом по дискретному речевому сигналу s' можно восстановить непрерывный сигнал s при заданных 8 и ф по формуле (1), т.е. существует отображение A'^ф, обратное к $A_{§,ф}$. Отображения $A_{§,ф}$ и $Д''^{1}$^ при ф=0 будем обозначать через A§ и A" § соответственно.

Таким образом, для каждого фиксированного $8<l/(2F_{max})$ и фиксированного сдвига ф между множеством речевых сигналов S и *множеством дискретных речевых сигналов* (будем его обозначать через $Sg_{>9}$) существует взаимнооднозначное соответствие, задаваемое отображением

A5,ср.

Утверяедение 1.2. Для любых 9, ф $Ss_{>9}=S5,e-$

Доказательство.

Пусть s'eS^. Тогда 3seS: s-$Ag_{>9}(s)$, такой что VkeZ $s(k8H^)=s'(k)$. Рассмотрим сигнал У : *s(t) = s(t-Q+q>)* VteR. Очевидно, что звуковой сигнал S"~s, поэтому он также является речевым сигналом. Рассмотрим

s"=A_5,e(?). VkeZ s"(k)=?(k5+0)=s(k5-e+9+e)=s(k5+9)=s'(k) => s'=s" => s' є Ss;e $^{=>}$ S§$_{i9}$cS5_5e.

Аналогично доказывается, что Sg^cSg^. Значит, S§;(p=S8,e.

Утверждение доказано.

Следовательно, в обозначении множества дискретных речевых сигналов второй параметр - величину сдвига - можно опускать и обозначать это множество через Ss-

Далее, если обратное не оговорено особо, будем для упрощения опускать штрих в обозначении дискретных звуковых сигналов.

Через S[$_{nm}$], n,m6Z, будем обозначать *ограничение дискретного сигнала s на отрезок* [n,m], т.е. дискретный звуковой сигнал, определенный следующим образом:

S[n,m](i)=s(i) Vi=n,n+1,...,m

S[n,m](i)¯0 Vi<n и Vi>m

Введенную нами в §1.1 речевую модель (определение 1.3) можно переопределить для дискретного времени. Далее, если противное не оговорено особо, будем рассматривать везде дискретизацию с нулевым сдвигом (ф=0).

Итак, пусть для данного естественного языка заданы некоторая речевая модель (A,4,B,B,S,x,n,T,fi,r) и период дискретизации 5<l/(2F_{max}).

Определение 1.10. *Речевой моделью с дискретным временем* назовем десятку <A,A,B,13»S5,T,n5,T,|Js,r8}, где

- A, B - конечные алфавиты букв и звуков соответственно;
- ДcA* - словарь команд;

• SaB* - словарь звуковых команд;

• Sg - множество дискретных речевых сигналов;

• т:A*-^B* - отображение транскрибирования;

• *И§:3^>2* , П5=А5°П - отображение произнесения;

• T: SgxZ^N, $Ts_5(n)=[Ts(n5)/5]$ - функция, задающая размер *окна анализа.*

• rgiSs—»в, rg^roAs"[1] -*распознающая функция;*

• (o^SgxN-^Z - *отображение фонемной разметки* звуковых сигналов, такое что V(3G3, (3=bib$_2$...b$_m$, VSSGIT^P) выполнено 0=M,$_6$(S$_6$, 1)<|Li5(s5?2) <...<|i_6(s$_{55}$2m+2)=|s_s|, |i_6(s_5,n)=0 при n>2m+2.

Отображения (j, и (Lis связаны соотношением Vi= 1,2,..., 2m+2 j$Li_5(s_5,i)$=[(|u(A_5"[1]$(s)_5$ i)/5].

Свойства функции произнесения $П_5$ и отображения фонемной разметки Ц5 для дискретного случая легко получаются из соответствующих свойств для непрерывного случая.

Свойства функции П§:

1°V(3GB $П_5(p)$*0,

2° $p_ь$ P_2eB$_э$ Pi*p_2 => $П_5$(Pi)п$П_5(P_2)$=0 3°,5° для

дискретного случая теряют смысл 4° Vpe8

$s_5Gn_5(P)_5$ s'$_6$~s_s => 8'$_6$є$П_6$(P) 6°VpeSs$_6Gn_5$(P)=>s^0

Свойства функции \x§\

2° Vj=l..m Vn=($i_6(s_5,2j)$+1, j$i_5(s_5,2j)$+l,... |i_5(s_5?2j+l)-l

S8[n-[T(s)/2],n+[T(s)/2]]^{є} П5(Ьц)

3° Vj=l..m+1 Vn=|i§($s_{5?}$2j-1), n$($s_5$,2j-l)+l,... |$LL_6$(s6,2j), выполнено, что либо $S5_{[n\cdot[T(s)}$/2],n_+[T(s)/2]]en_5(bj.i)un_5(bj), либо VbeB S5[n.[T(s)/2];n+[$T(_S)$/2]]^n_5(b) (здесь $b_0 = b_{m+}$i =_ - символ паузы) 4° Если p=Ь - звуковая команда из одной буквы, а sgellg(p) -произнесение этой команды, то |Ag(sg, 1)=1115(85,2)=0, (a_6($s_{5?}$3)=|A6(s5,4)=js_5|.

Как и для непрерывных речевых сигналов, для каждого дискретного речевого сигнала sg, являющегося произнесением некоторой звуковой команды (Зє73, будем говорить, что отрезки натурального ряда вида [|Hg(sg_9 2i),|i(ss, 2I+1)] *соответствуют произнесению і-й фонемы* звуковой команды р, а отрезки натурального ряда вида *[[іь($ь,* 2i+l)+l,|Hg(sg, 2i+2)-l] *соответствуют переходу от і-й к і+1-й фонеме* звуковой команды р. Величину p,g(sg,2i+l)-|Ag (sg,2i)+l будем называть *длительностью произнесения* і-й фонемы в дискретном речевом сигнале Sg, а величину (i5(ss,2i+2)-|Li5 (sg,2i+l)-l — *длительностью перехода* от і-й фонемы к і+1-и.

Определение 1.11. Отображение *d§:* SsxN^R^M будем называть *функцией описания дискретных речевых сигналов,* если Vs5eS5,VieZ выполнено dg(s6,i) = d(Ag" (sg),i5), где d:SxR-^R^M - некоторая функция описания непрерывных речевых сигналов.

Содержательно это определение означает, что для описания дискретного речевого сигнала мы сначала восстанавливаем из него непрерывный речевой сигнал, а затем применяем какую-либо функцию описания к непрерывному сигналу.

Замечание 1.2. Как в непрерывном, так и в дискретном случае можно считать, что область значений функции описания является многомерным комплексным пространством. Для приведения комплексного случая к действительному нужно действительные и мнимые части значений функции описания рассматривать как отдельные координаты многомерного действительного пространства.

Поскольку по определению любая функция описания непрерывных речевых сигналов представляется в виде d=d'°0_5 то Vs§GS5,ViGZ выполнено

de(S5,i) = ^(0(^^(85),16)).

Определение 1.12. Функцию 0§:S_5xZ^Ss, определенную как

$0_5(s_6,z)(z')=$ $s_s(z + z')$, $z^{\}}$- *ТЛ*)* **ЗД**

, будем называть дискретным

Оэ *иначе*

оконным отображением.

Утверждение 1.3. VsgeSgjVieZ 0(A8"^{i}(s_s)$_5$i8) = A^Os^i)).

Доказательство

Возьмем любое teR, такое что t=kS, к є Ъ.

Если t€[-$r_{A\_}$, (i5)/2, $Г_{д\_}$, (iS)/2]$_?$ то по определению 1.12 имеем

$0(A_5''^1(s_5)_3i5)$ (t) = $A5^{-1}(s_5)$ (t+i8) = s_5(k+i), в то время как АДОб^бД)) (t) = $0_5(s_5,i)$ (k) = s_5(i+k), откуда $0(A_5''^1(s§),i8)(t) = A_5'^1(0_5(s5_5i))(t)$ для t из этого отрезка.

Если же tgf-T^ $_}$(i8)/2, $Г_{д\_}$,(i5)/2], то по определению 1.4 оконного отображения О имеем $0(A_S''^{1}(ss),iS)$ (t) = 0, в то время как $As^{-1}(0_5(s5,i))$ (t) = Os(s_5,i) (k) = 0 по определению 1.12.

Звуковой сигнал A_8~ ($O_s(s_5,i)$) удовлетворяет свойству $Ф_д$-«,(0,(,* ,-))(0=0 при $|f|>F_{max}$ по определению функции As"[1], а звуковой сигнал $0(A5^{-1}(ss),i5)$ удовлетворяет этому свойству в силу свойства 5° функции произнесения. По теореме Котельникова два сигнала с ограниченным спектром, совпадающие на множестве точек t=k5, к є Ж, совпадают всюду.

Утверждение доказано.

Следствие 1.1. Любая функция описания $d_§$ дискретных речевых сигналов может быть представлена в виде d_s=dVO§, где

d'§: S-»R^M - дискретная функция описания окон;

0_8: дискретное оконное отображение.

Доказательство

По утверждению 1.3 Vs_8eS_8,VieZ $d_s(s_5,i) = d(A_8\text{-}^1(s_5),i5) = d'(0(A_5'^1(s_5)_5i5))=d'(A_8''^1(0_5(s_8,i)))$, откуда d_8 = d'oAg'^Os. Обозначим d'_8=d'°A5" . Отображения d'_8 и 0_8 обладают, очевидно, искомыми свойствами.

Следствие доказано.

Содержательно следствие 1.1 означает, что для описания дискретного сигнала нет необходимости восстанавливать непрерывный сигнал из дискретного и применять непрерывную функцию описания. Существует дискретный аналог функции описания и дискретный аналог оконного отображения, которые позволяют определить функцию описания для дискретного случая аналогично непрерывному случаю как композицию оконного отображения и функции описания окон.

Посмотрим теперь, какая дискретная функция описания соответствует функции описания для непрерывного случая, описанной в примере 1.1.

Дискретным преобразованием Фурье *Ф* дискретного звукового сигнала Ss называется функция $Ф^{д}8_{8}(ї)= \sum_{«=-00}^{+00} 2>Ж^{2}$^ ([32]).

Утверяедение 1.4. Пусть функция описания непрерывных звуковых сигналов d порождена функцией описания окон d'? заданной в виде *d\s)=(OsC/M¥o)9* <I>S(²/MFO), ••., $O_s(^{M}/_{M}F_0)$)? где F_0 - некоторая постоянная частота из отрезка *[¥max,1/2ь]*-

Тогда для соответствующей функции d_s описания дискретных речевых сигналов функция описания окон $d'_{§}$ имеет вид d'8(s6H*V/MF_0>5, ФЧ(²/мF_0)-5, ..., $Ф^{д}8_5(^{м}/_{M}F_0)$-5), где $Ф^{д}$ - оператор дискретного преобразования Фурье.

Доказательство.

Пусть s: R-»R - некоторый речевой сигнал, s_5=A5(s) - его 8-дискретизация.

В приведенных ниже выкладках для простоты опущены нижние индексы s и s§ в обозначении преобразования Фурье.

Согласно формуле обратного преобразования Фурье для непрерывного времени ([29]) при t=5n имеем

$$s_6(n)=s(5n)=-^{-}\ h(f)e^{i2}{}^{\wedge}df = {}^{\wedge}- \sum_{(2*-l)—}^{(2к+\)^-} /Ф(/У^{2}*Ч =$$

(поменяем местами интеграл и сумму, сделав в интегралах замену)

$$\frac{1}{2n}\ c- \quad 2\partial \quad i \quad +co \quad i$$
$$U \quad к=-со$$

С другой стороны, по формуле обратного преобразования Фурье для дискретного времени ([32])

$$s_5(n)=f- /Ф-С/У^{2}*\wedge$$

171 \
~2S

Приравнивая два выражения для $s_§(n)$, получим соотношение

1 +00 1

0 *=-«> *0*

Поскольку в нашем случае рассматриваются значения / из промежутка (О, F_0], $F_{max}<F_0<l/28$ и по свойству 3° отображения П Ф(їі)=0 при $|f|>F_{max}$, то в сумме (*) только одно слагаемое отлично от нуля (это слагаемое соответствует случаю к=0, т.к. для k>0 $f+k/8>0+2F_{max}=2F_{max}$, а для к<0 $f+k/5<Fo-2F_0=-F_0<-F_{max}$)

=> $Ф^{Д}(/) = \frac{}{\partial} Ф(/)$, или, с индексами s и s§, $Ф^{Д}з§(:Е)=Ф_§(г)/8$.

Последнее равенство справедливо для любых звуковых сигналов s и $s_s=A5(s)$, для которых $Ф_5(г)=0$ при $|fl>F_{max}$, в частности, для удовлетворяющих этому свойству по свойству 5° отображения П звуковых сигналов С>5(S5) И 0(s).

Следовательно, выполнено равенство О*0g(S5)(f)8 = Фоф®, откуда dg(s5,i) = d^{f}(0(s,i8)), что завершает доказательство утверждения.

Утверждение доказано.

Определение 1.13. Функцию описания дискретных речевых сигналов будем называть *нормальной,* если нормальной является соответствующая ей функция описания непрерывных речевых сигналов.

Утверждение 1.5. Функция описания непрерывного речевого сиг-нала, порожденная функцией описания окон d'(s)=(<D_s($/_{2М}$5)> $Ф_8$(*hus),* ■■-, $Ф_5(^{м}/_{2}мб)$) (здесь $F_0^{=1}/_2$§) и соответствующая ей дискретная функция опи-

сания, порожденная функцией описания окон ^^^(Ф^зС^мб)^, $Ф^{Д}85(^{2}/2M5)'5$, ..., $Ф^{Д}85(^{M}/2Mб)'§)_{?}$ являются нормальными функциями опи-сания при $M>T_{max}/8$.

Доказательство.

Рассмотрим описание произвольного речевого сигнала s:R-^R с помощью функции описания d=d'oO в произвольной точке to- Из свойства 4° функции произнесения П следует, что можно не ограничивая общности считать, что to=5-io, для некоторого іоєЖ (действительно, если это не так, то вместо сигнала s можно рассмотреть эквивалентный ему речевой сигнал, смещенный по оси времени на t_0-[t_0/5]8).

Покажем, что отображение d^f инъективно.

Обозначим через *s′* речевой сигнал s после применения оконного преобразования O: $s'(t) = O(s_9 t_0)(t)_9$ а через s'_s речевой сигнал sg после применения оконного преобразования Og: $s'_s(n) = O_s(s_s j_0)(n)$.

Тогда $Ф8<^{т}/_{2M}б)/5 = Ф^{д}8._5(^{т}/2M5) - \frac{-2mn\ ----}{2M}$ £*>)*

$$\frac{^{T}S;(fr)}{П=\ T_s*Oo)}\ t\ 4(*K"^{2J}"$$

$)e \quad ^{LM}$ - (поскольку $M>T_{max}/8$, откуда $M>Ts_s(io)/2$)

$$5»)e \qquad ™ = Y,s^f_s(n-M)e$$

$$e\sim 5>;(и\text{-}лое^{\frac{n\sim -M}{2M}}{}_{-2mn\ ----}{}^{n=0}{}_{\partial?yu}{}_{1У\pm\ J\&}$$

Применяя формулу обратного дискретного преобразования Фурье ([32]), получим:

$$-mm \quad 1 \quad \backslash Ч \quad \kappa \quad i2лк(n+M) \quad \kappa \quad ч \quad 2M$$

к=M,M+1,...2M-1 положим $Ф^{*}\#\left(\frac{A:}{2MS}\right) = Ф^{*}Ч\left(\frac{2M—1—A:}{2M8}\right)$ (из свойств преобразования Фурье для действительного сигнала, [32]).

Возвращаясь снова к непрерывному сигналу, $s'(t) = O(s,t_0)(t)$ =

$O(AV(\wedge)_5 t_0)(O$ = (по доказанному ранее утверждению) =

$A\text{-}j(O_d(s)_9 i_0)(t) = A\ V\ (s'_s)(O$. Итак, $s'(t) = A''^{1}\ (s'_s)(t)_9$ в то время как

$$■ \quad 1 \quad {}^{2}\wedge \quad Л \quad Ия к(n+M) \quad Л2M$$

Таким образом, мы показали, что сигнал s' однозначно восстанавливается по значению d'(<v'), т.е. что отображение d^f - инъективно. Следовательно, порожденная этим отображением функция описания нормальна.

Утверждение доказано.

Пусть X-прямоугольная mxn матрица, X^i, i=l,2,..., п - ее столбцы. Будем в таком случае писать, что X=||X |X |...|X ||.

Матрицу $d_5(s_6)$ = $||d5(s5,to(s5))|d_5(s_6,to(s5)+l)|...|d_6(s5_3ti(s_s))||$ размерности Mx|s5| будем называть *описанием дискретного речевого сигнала* S5.

С целью экономии памяти в практических приложениях удобнее вместо матрицы dg(s5) описания речевого сигнала s§ использовать матрицу, составленную из некоторого подмножества множества столбцов матрицы $d_6(s_5)$.

Более строго, *описанием дискретного речевого сигнала s$ с шагом L*, LeN, будем называть матрицу размерности Mx[|s5|/L], столбцы кото-

рой соответствуют описанию сигнала в моменты времени, кратные L, т.е. являются подмножеством множества столбцов матрицы ds(ss):

$$d5,L(S6)^{-}\|d_5(s_5,\ \underset{VL}{\ }\ \frac{\Phi б)}{L}+1\quad Z)|d_5(s_s,\ \frac{\Phi б)}{L}+2\quad \blacksquare Z)|...|d_6(s5,\ \frac{\Phi^{*})}{L}\quad L)\backslash\backslash$$

Очевидно, что $d_5(s_5) = d_5,i(s_5)$.

Пусть sg - некоторый дискретный речевой сигнал. Пусть $d_{5jL}(s5)$ -описание этого речевого сигнала с шагом L. Столбцы этой матрицы, рассматриваемые как координаты точек пространства R^M, задают некоторое множество точек в этом пространстве.

Рассмотрим теперь все возможные произнесения всех фонетических слов для данного естественного языка.

Определение 1.14. *Дискретным образом* £\ь(Ь) *фонемы* ЬєВ для данной функции описания *d§* и шага описания L будем называть следующее подмножество пространства R^M:

$t>_{5;}L(b)=\{d_5(S5)(z)\ I\ 8_6 Є П_5(Ь),\ ZG[t_0(s_5),\ ^(¾)]\}$.

Содержательно данное определение означает, что дискретный образ фонемы состоит из тех и только тех точек, соответствующих дискретным описаниям с шагом L всех произнесений этой фонемы, которые задаются дискретными моментами времени, кратными шагу L.

Утверждение 1.6. Для любых допустимых значений периода дискретизации 5 и шага описания L дискретный и непрерывный образы одной и той же фонемы совпадают.

Доказательство.

Очевидно, что каждый дискретный образ фонемы вложен в соответствующий непрерывный, т.е. Ds,L(b)cD(b). Это следует из того, что

для моментов времени вида t=Ln5 соответствующая индексу Ln точка описания дискретного сигнала принадлежит описанию соответствующего непрерывного сигнала.

Обратное утверждение следует из свойства 4° функции произнесения П. Пусть b - произвольная фонема, *хеТХЪ).* Тогда существует непрерывный звуковой сигнал SGil(b) и момент времени t': d(s, t')=x. Рассмотрим сигнал s': s'(t)=s(t-t') VteR. Очевидно, что s'~s => Б'єЩЬ) (свойство 4° функции П) и fi(s,t)=fi(s',t) Vte[t_0(s), tj(s)] (свойство 1° функции **И)-**

d(s',0)=d(s,t')=x. Если обозначить через s'_8 дискретный звуковой сигнал s'g^As^'), то по определению d_s имеем $d_s(s'_s,0)$=d(s',0)=x. Момент времени 0 имеет вид k8L, keZ (здесь k=0) => по определению "D^iXb) имеемхєОбьФ) => ДЬ)е0_5ь(Ь).

Итак, 0_5,ь(Ь)сДЬ) и D(b)cD_5L(b), следовательно, *TXb)=$V_{b,L}$(b).*

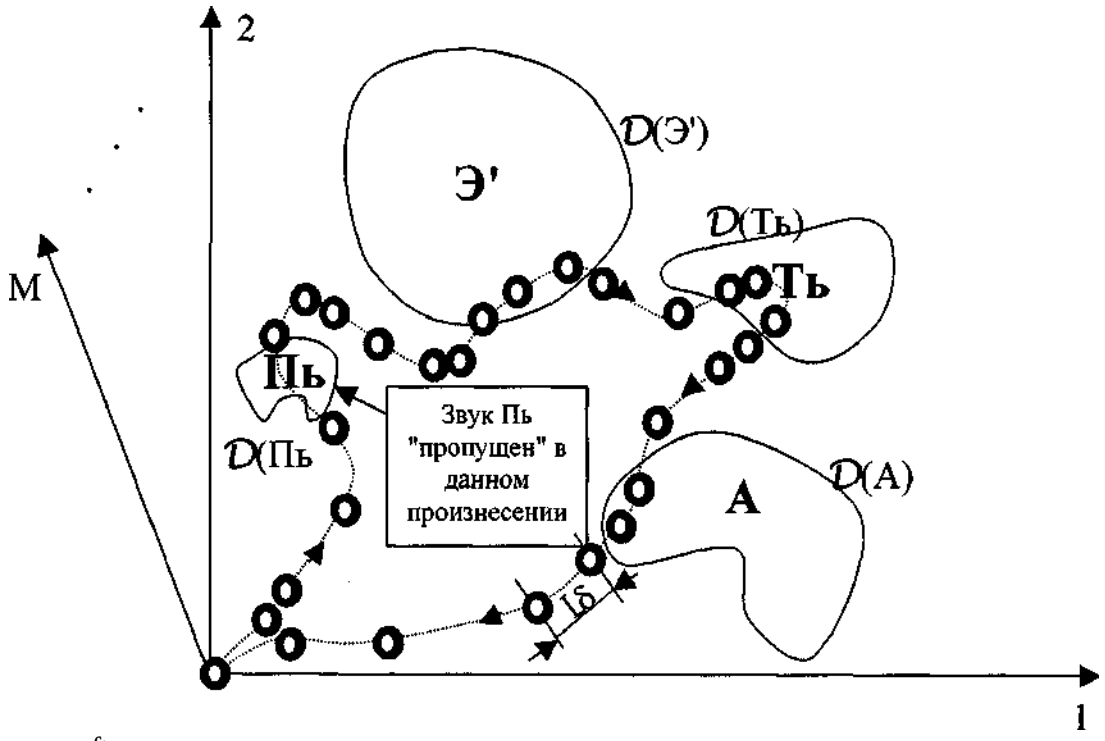
Утверждение доказано.

Содержательный смысл утверждения 1.6 следующий: как бы сильно мы не "прореживали" сигнал при его описании, образы фонем будут в точности теми же, поскольку мы имеем свободу при выборе начальной точки дискретизации и прореживания.

Поскольку 1ХЬ)=0§,ь(Ь) V допустимых 5 и L, индексы 5 и L можно опустить и в дальнейшем и в дискретном случае обозначать образы фонемы b через *ТХЪ).*

При дискретизации речевого сигнала при достаточно большом значении шага описания L может случиться, что множество точек описания сигнала, соответствующих дискретному описанию сигнала с шагом L, не будут пересекаться с каким-либо дискретным образом фонемы, несмотря на то, что соответствующая этому множеству фонема (символ алфавита

В) входит в произнесенное фонемное слово. Будем говорить в таком случае, что в описании данного (дискретного) произнесения данного фонемного слова указанная фонема *"пропущена"* (см. рисунок 1.5).



ft - точки, соответствующие дискретному описанию сигнала

Рисунок 1.5. Описание с шагом L дискретного речевого сигнала, являющегося произнесением слова «Пе'тя»

§1.3 Формализация задачи распознавания речи. Функция качества словаря команд.

В предыдущем параграфе мы ввели понятие дискретного речевого сигнала и показали, что если период дискретизации 5 выбран правильно, множества дискретных и непрерывных речевых сигналов находятся во взаимнооднозначном соответствии.

Понятие речевой модели было сформулировано как для случая непрерывного (определение 1.3), так и для случая дискретного времени (определение 1.10). Было показано, что операторы описания в непрерывном и дискретном случае устроены одинаково (следствие 1.1).

Мы показали также, что несмотря на то, что дискретное описание речевого сигнала берется с некоторым временным шагом, образы фонем и в дискретном, и в непрерывном случае в точности одни и те же (утверждение 1.6).

Таким образом, не ограничивая общности, в дальнейшем можно считать, что мы имеем дело с дискретными речевыми сигналами и рассматриваем их описания с шагом 1. Для упрощения записи индекс 8 при обозначении элементов речевой модели с дискретным временем будем в дальнейшем опускать.

Поскольку распознающая функция г восстанавливает по речевому сигналу речевые команды из "8, а не тексты команд из *A,* на практике используются такие словари команд, транскрипции которых попарно различны (т.е. в словаре нету т.н. "омонимов" - одинаково звучащих слов). В таком случае без ограничения общности можно считать, что команды -это слова в алфавите В, и исключить из нашей речевой модели алфавит букв А и отображение транскрибирования т.

На практике системы распознавания речи работают не со звуковыми сигналами, а с их описаниями как с исходным материалом для обучения и распознавания речевых образов. Если через (R^M) обозначить множество прямоугольных матриц с М строками и произвольным количеством столбцов с элементами из R (иначе говоря, множество слов над континуальным алфавитом R^M), то описание любого речевого сигнала является элементом множества (R^M)*.

Часто бывает удобным разбить пространство R^M на конечное число непересекающихся подмножеств: $R^M = \bigcup_{i=\overline{1,2,...Jc}} Jc_i$, $C=\{c_i, i = 1,2,...Д\}$, которым можно сопоставить символы некоторого конечного алфавита $C=\{c_1,c_2,...,c_k\}$ (алфавит С называется обычно *кодовой книгой).* В этом случае каждая матрица из (R^M) может быть закодирована словом в алфавите С, если каждому столбцу этой матрицы сопоставить символ из С, соответствующий элементу разбиения С, содержащему точку в R^M, координаты которой задает этот столбец.

Таким образом, понятие речевой модели, с учетом включения в модель функции описания, кодовой книги С и кодирующего отображения Г, может быть, не ограничивая возможностей ее применимости, сформулировано следующим образом:

Определение 1.15. *Речевой моделью* называется совокупность 11 элементов <B,B,S,(R^M)*,C,n,T,|i,d,r,r>, где

- B - конечный алфавит звуков;
- B - словарь команд, BcB*;
- S - множество дискретных речевых сигналов;
- (R^M)* - множество прямоугольных матриц над R с M строками;
- C - кодовая книга (конечный алфавит);

• Г: $R^{,M}$ -»С - кодирующее отображение, которое естественно обобщается на слова как Г: (R^M) ->С*, если сопоставить каждой матрице у£(R^M)\ у=||y^1|y^2|...|y^n|| слово Г(у)= Цу^Цу^.-Цу^{11}) - последовательный *код* у.

• d: SxN->R^M - функция описания речевых сигналов, имеющая вид d=d'°0, где d': S-»R^M - функция описания окон, а О: SxZ—»S - оконное отображение; d также обобщается на сигналы как d: S-KR^M/ , если сопоставить сигналу s его описание d_s=d(s)=||d(s,l)|d(s,2)|...|d(s,n)||;

• *И:Ъ-^2* - функция произнесения, обладающая свойствами:

l°VJ3eB П((3)*0,

2° $p_ь$ p_2єЗ, Pi^p_2 => П(Р,)пП(p_2)=0

4°VPeB sen^s^s^SjenCP)

6°VpGBsGn(P)=>s^0

• Т: SxZ-»N, - функция, задающая размер *окна анализа,* такая что T_{min}<T_s(n)<T_{max}, где T_{min},T_{max}єМ - некоторые натуральные числа, T_{min}<T_{max} (здесь T_s(n) - окно анализа сигнала s в момент времени п)

• ц: SxN-»Z - *отображение фонемной разметки* звуковых сигналов, такое что УРєВ, Р=bib_2...b_m, VSGII(P) выполнено 0=^i(s,l)<p,(s,2) <...<|a(s,2m+2)=|s|, u(s,n)=0 при n>2m+2 и справедливы свойства: 1° Si~s => |Li(si,n) = |i(s,n) VneN. 2° Vj=l..m Vn=^(s,2j)+1, Ks,2j)+1, ..., |a(s,2j+l)-l

S[n-T(s)/2,n+T(s)/2]$^{\epsilon}$ n(bj)

3° Vj=l..m+1 Vn=ju(s,2j-1), p,(s,2j-1)+1,... }J.(s,2j), выполнено, что либо $s_{[n.}$|T(sy2],n+iT(sy2]]eII(bj.i)uri(bj), либо VbeB

S[n-[T(s)/2],n+[T(s)/2]] ¢- П(Ь)

(здесь b_0 = b_{m+}i =_ - символ паузы)

4° Если (3=b - звуковая команда из одной буквы, а sell(P), то |i(s,l)=|i(s,2)=l,|i(s,3)=|i(s,4)=|s|. • г: Г-»2^{B} *-распознающая функция,* определенная как г(у)={Рє"В: уєГ(ci(П(Р)))}.

Замечание 1.2. Если образы фонемных множеств при отображении Г не пересекаются, то \/уеГ(ё(ПСВ))) |г(у)|=1, т.е. речевая команда однозначно восстанавливается по коду описания любого ее произнесения.

Замечание 1.3. Необходимым условием для того, чтобы \/уєГ(е(П(В))) выполнялось |г(у)|=1, является условие нормальности функции описания d.

Под *задачей распознавания речи* в речевой модели (B,B,S,(R^{M})^C,n,T,|u,d,r,r) будем понимать задачу восстановления функции г на основе конечного числа примеров произнесения каждой из звуковых команд из *Ъ.*

Обычно задача распознавания речи решается с помощью *метода сравнения с эталонами.* Опишем этот метод более строго.

Итак, пусть словарь команд В=(Рь p_2,.-- P_N}- Сопоставим каждой команде pj из *Ъ эталон* е; - элемент некоторого абстрактного множества E={$e_b e_2$,...,e_N}.

Содержательно эталон е; команды Р; является компактным способом представления множества кодовых слов всех ее произнесений и

строится обычно на основе набора примеров ее произнесения - некоторого НепуСТОГО МНОЖеСТВа КОДОВЫХ СЛОВ {уц, УЦ, ..., Уі,п(і)}сГ(П(Зі)).

На декартовом произведении С*хЕ введем функцию *расстояния* р: С*хЕ—»Ж между кодовыми словами и эталонами. Распознавание речи методом сравнения с эталонами производится по формуле Г(Е,р)(у)= *Argmm(p{y,e_i)*). Если Г5а-($_{E)}$р), будем говорить, что имеют место

ошибки распознавания. Алгоритм распознавания речи, основанный на сравнении кодовых слов с эталонами из Е с помощью расстояния р, будем обозначать через (Е, р).

Содержательно метод сравнения с эталонами позволяет находить эталоны, ближайшие к распознаваемому коду.

Качество q(E,p)CB) словаря команд В для данного алгоритма распознавания (Е,р) естественно определить, как 1-2$р_{ош}$., где $р_{ош}$. -среднее число ошибок распознавания команд из В алгоритмом (Е,р) (более строгое определение функции качества будет дано для конкретных примеров автоматных алгоритмов распознавания речи в следующих главах).

Важно заметить, что каждой команде р соответствует множество кодовых слов описаний всех ее произнесений Г(с1(П(р))), которое являет-

ся конечным подъязыком языка С всех кодовых слов. Автоматный подход к распознаванию речи состоит в том, чтобы представить язык кодовых слов каждой команды некоторым автоматом. В случае конечно-детерминированных автоматов (глава 2) мы погружаем язык кодовых слов команды в некоторый известный нам бесконечный регулярный язык, а в случае вероятностных автоматов (глава 3) - в множество всех слов В* над алфавитом В, приписав при этом каждому слову языка кодо-

вых слов действительное число - его вероятность. Последний подход дает нам возможность рассматривать случай пересекающихся языков кодовых слов.

ГЛАВА 2. ДЕТЕРМИНИРОВАННЫЕ АВТОМАТНЫЕ МОДЕЛИ §2.1

Решение задачи локального распознавания речи.

Пусть задана речевая модель (B,B,S,(R^M)*,C,II,T,|j,,d,r,r} и решается задача распознавания речи в этой модели. Предположим, что разбиение *С* и соответствующая ей кодовая книга С таковы, что существует алгоритм распознавания речи (Е,р), восстанавливающий г(у) без ошибок при условии, что для каждого рєК,^{м} мы умеем вычислять Г(р) (далее в этой главе будет показано, что такое *С* существует). К сожалению, на практике вычисление Г(р) для таких С оказывается сложной задачей. Существование "идеального" алгоритма (Е,р) распознавания кодовых слов сводит, таким образом, задачу распознавания речи к задаче восстановления кодирующего отображения Г. Поскольку Г вычисляется отдельно для каждого элемента описания сигнала, задачу восстановления Г назовем *задачей локального распознавания речи.*

Будем говорить, что задача локального распознавания речи решена, если каждой точке реR^M сопоставлено некоторое множество ^{x}F(p)cC, такое что Г(р)є$^{x\cdot}$ї$'$(р).

Для конструктивного задания функции локального распознавания *F предположим, что задана функция \|/, которая умеет "отличать" элементы разбиения *С* попарно. Перейдем к формальным определениям.

Определение 2.1. Функцию \|/: С х С х R^M-»Cu{?} будем называть *отделяющей,* если для VpeR (0°) i|/$_{cc}$(p>c

(1°) Vl/$_c$icj(p)=V|/cjci(p) (СИММЄТРИЧН0СТЬ)

(2^0)4/cm(pMci,$_{Cj}$,?}.

(3°) рєCi^м/сіезСрМсі,?}

Смысл функции i|/$_{cic}$j следующий - она выдает ответ на вопрос, "какому из множеств c_t, *cj* принадлежит точка р при условии, что она лежит в объединении c_t^*jCjT*. Ответы интерпретируются следующим образом: \\/$_c$icj(p)=Ci => РЄ C_t , \\/cicj(p)=Cj =>pЄ^c/, '?' - НеИЗВеСТНО.

Определение 2.2. *Функцией локального распознавания* ¥ будем называть функцию *Ч*:* R^M->2 (здесь 2^e - множество всех подмножеств алфавита С), определенную для каждой точки рeR^M следующим образом:

^(p)={сєС: Vc'eC, c'#c => ц/$_{cc}$(p)є{с,?}}.

Смысл функции ¥ такой: сєЧ^р), если гипотезу "рє *с*" невозможно опровергнуть при помощи отделяющей функции *f.*

Определение 2.3. Будем говорить, что функция *W отделяет* множества c_i и *Cj*, *Щ$_y$* (иначе говоря, *множества* c_i и *Cj отделимы с помощью функции ¥),* если для любой точки рeR^M выполнены условия:

(1°) рєс, =>Cj^(p)

(2°) *рес j =>a€¥(p)*

Можно дать графовую интерпретацию вычисления функций \|/ и ¥ для данной точки р из множества R^M. Итак, пусть точка р фиксирована. Построим ориентированный граф, вершины которого будут помечены символами алфавита С. Для каждой пары (c_{i5}Cj):

если \\/$_C$i$_C$j(p)=Cj, то проводим направленное ребро из вершины, помеченной символом С; в вершину, помеченную символом C_J;

если v|/$_C$icj(p)=Cb то проводим направленное ребро из вершины, помеченной символом Cj в вершину, помеченную символом CJ; если \|/$_{ciC}$j(p)-?', никаких ребер не проводим.

Утверяедение 2.1. cє^p) <=> из вершины, помеченной символом c, не исходит ни одного ребра.

Доказательство.

ce¥(p) <=> Ус'єС, с'^с выполнено: либо i|/$_{cc}${p)=?, либо \j/$_{cc}$(p)=b <=> Увершины, помеченной символом с'є С, с'^с выполнено: либо вершины, помеченные символами с и с', не соединены ребром, либо в графе есть ребро, направленное от вершины, помеченной символом с' к вершине, помеченной символом с.

Утверждение доказано.

Следствие 2.1. Множество T(p) совпадает с множеством меток вершин графа, из которых не исходит ни одного ребра.

Пример 2.1. Пусть в речевой модели <B,B,S,(R^M)*,C,n,T,(i,d,r,r> функция описания d - нормальная, а разбиение С пространства R^M на подмножества имеет вид R^M= $\underset{<=0,2..|Д|}{\backslash}$*JV(b$_t$),* где £;.=t)(bi) при i=l..|B| - образы фонем из B={bi,b$_2$,...,b$_{|B|}$}, а c_0 =0(bo)=R^M\ $\underset{<'=1.2..|Д|}{[]}$^(Ь,) - *«межфонемное пространство».* В этом случае можно считать, что C=Bu{bo}. Как будет показано ниже, для этого примера существует детерминированно-автоматный алгоритм распознавания речи, восстанавливающий речевые команды по кодовым словам описаний их произнесений.

Пример 2.2. Пусть задана речевая модель из примера 2.1, B={П,Р,И,М,Е} (алфавит для слова ПРИМЕР), C=Bub_0 и функция \|/ в точке р задана следующей таблицей:

	П	Р	И	М	Е
П	П	9	И	М	Е
р	?	Р	И	М	?
И	И	И	И	?	И
М	М	М	?	М	?
Е	Е	?	И	?	Е

Тогда граф, соответствующий функции ц/ в приведенной выше графической интерпретации, будет выглядеть так (вершины, соответствующие ¥(р), обведены дважды):

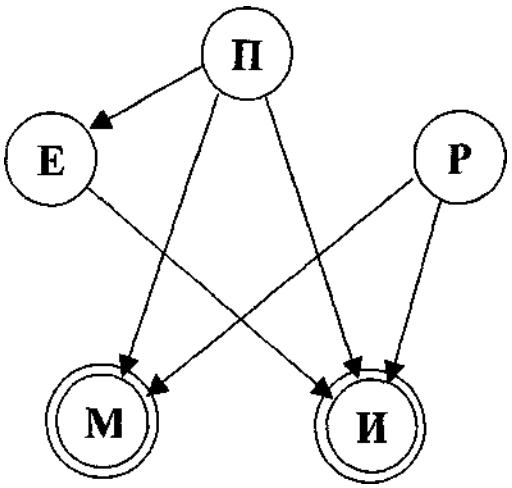


Рисунок 2.1. Графовая интерпретация функции ¥.

В данном примере ^{x}Р(р)={М,И}.

Утверждение 2.2. УрєЕ^{м}Г(р)єТ(р)

Доказательство.

Если *рес$_0$,* то Г(р)=c_0 => по определению функции *ці* Vc'eC, c'^Co выполнено УсоС'Є{со,?} => по определению ¥ c_0єТ(р) => Г(р)є^{x}Р(р).

Если же pg c_0, то Эс: Г(р)=с и рє c => по определению функции \\i Vc'eC, cVc выполнено i|/$_{CC}$'Є{с,?} :=> по определению 4^{y} сє^(р) ■=>

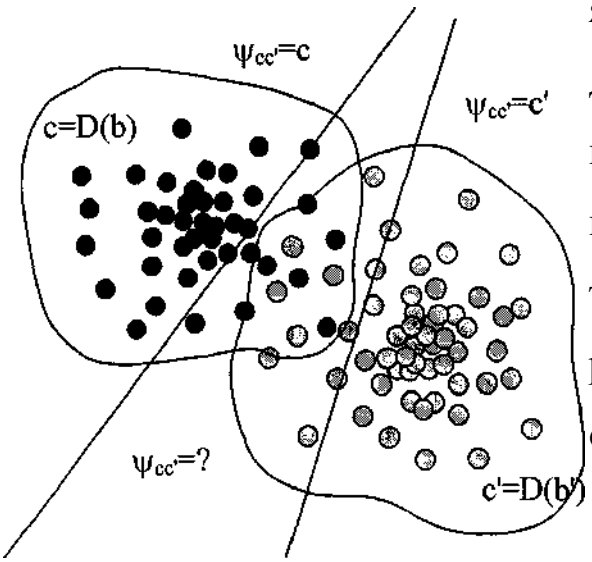
Г(р)є^(р).

Утверждение доказано.

Содержательно утверждение 2.2 означает, что множество "неопро-вергнутых гипотез локального распознавания " Т(р) обязательно содержит "правильную гипотезу" Г(р).

На практике отделяющая функция \|/$_{CC}$' может быть введена с помощью построения поверхности (например, состоящей из фрагментов гиперплоскостей), разделяющей образы элементов разбиения, в нашем примере - двух фонем *V{h)* и 0(Ь').

Если такая поверхность на основе обучающей выборки может быть построена, данная пара образов фонем отделяется с помощью функции \j/, являющейся характеристической для данной поверхности (если положить \|/$_{cc}${р)=с, если точка р и множество 0(b) лежат с одной и той же стороны относительно разделяющей поверхности, а иначе определить, что \|/$_{cc}$(р)=с').



2.2. Построение отделяющей функции *ц,* для случая М=2

В противном случае, если Рисунок

^{так}У^{ю} разделяющую поверхность построить не удается (например, когда Ц)(Ь)пО(Ь')^0), зачастую можно построить две непересекающиеся разделяющие поверхности, которые разбивают все множест-

во *ТХЪ)uТХЪ')* на три подмножества. Точки первого подмножества гарантированно принадлежат *ТХЪ),* второго подмножества - объединению D(b)u'D(b^l) и третьего подмножества - множеству **D(b').**

§2.2 Модель распознавания последовательного кода команды с помощью детерминированных автоматов.

С помощью функции локального распознавания *Ч** невозможно однозначно восстановить код произнесенной команды, поскольку для каждого момента времени i функция *W* дает вместо однозначного ответа распознавания (кода Г(у') точки y^1 описания у сигнала в этот момент времени) множество гипотез распознавания, которое содержит в силу доказанного выше утверждения "наилучшую" гипотезу Г(у'), но может содержать и другие буквы кодовой книги С.

Множество гипотез локального распознавания для каждого момента времени порождает некоторый язык слов в алфавите С возможных кодов сигнала, если рассматривать сигнал в целом.

Введем некоторые обозначения.

Пусть дано произвольное конечное множество A={$a_ьa2,...a_т$}. Обозначим через R(A) регулярное выражение ([34]), описывающее регулярный язык однобуквенных слов в алфавите А: R(A)=(ai|a_2|...|a_m), где | означает символ «или».

Пусть **а** - регулярное выражение. Обозначим через (а) регулярный язык, которое оно описывает.

Пусть s - речевой сигнал, у=||y^1|y^2|.. .|y^n|| є(К.^{м})* - его описание. Рассмотрим язык "гипотез локального распознавания" матрицы у:

{y| y=cic$_2$...c$_n$, CiЄ^y^1)}. Этот язык, очевидно, описывается регулярным выражением f(y)=R(^{v}F(y^1))R(^{v}F(y^2))...R(^{v}F(y^n)), причем Г(у)є<Г(у)>.

Определение 2.4. Пусть C'cC - подалфавит кодовой книги. Окрестностью О(С') алфавита С назовем следующий подалфавит С:

0(C')={сєС: Зс'єС: с и с' не являются отделимыми с помощью функции Т}.

Лемма 2.1 Для любой точки рєК,^{м} ^(р)сО(Г(р)).

Доказательство.

Пусть Г(р)=с => по утверждению 2.2 сє^(р). Возьмем любую другую букву *с'еЧ(р),* cVc. Множества с' и с не являются отделимыми с помощью функции $Ч^1$, так как противное противоречило бы определению отделимости. Следовательно, с'єО(с)=0(Г(р)).

Лемма доказана.

Объединив утверждение 2.2 и доказанную лемму, получим, что имеет место вложение: Г(р)є^(р)сО(Г(р)).

Обозначим через *сф)* множество элементов разбиения С, пересекающихся с образом D(b) фонемы Ь:

$$c(b)=\{ceC:\ crYD(b)*0\}.$$

Через с(Ь) будем, соответственно, обозначать множество букв кодовой книги, соответствующих элементам разбиения из *c(b).* Очевидно, что для любого звука ЬєВ r(D(b))c:c(b).

Покажем, что множество (f (y)) гипотез локального распознавания описаний *y* произнесений любой речевой команды можно описать регулярным выражением, которое строится непосредственно по тексту этой команды.

Определение 2.5. Пусть рєВ - некоторая команда, p=bjb$_2$...b$_m$. *Эталонным кодом* Г ((3) команды |3 будем называть регулярное выражение f(P)=R(O(cO^c(b$_1$)uc$_0$}))*R(O(c(b$_1$)))*R(O(c(b$_1$)uc(b$_2$)uc$_0$))* R(O(c(b$_2$)))*...R(O(c(b$_m$)))*R(O(c(b$_m$)uc(Juc$_0$))* (здесь * - символ итерации).

Теорема 2.1. Пусть задана речевая модель <B,B,S,(R^M)*,C,n,T,|a,d,r,r>, отделяющая функция у и функция локального распознавания ¥. Тогда для любой звуковой команды (3=bib$_2$...b$_m$ выполнено (f(y))c:(f ((3)), где y=d(s) - описание произвольного произнесения sen(P) команды р.

Доказательство.

Поскольку Г(у) є (Г (у))

Пусть НIуУi.-.IУЧИад и м(s)=Ms,1), |i(s,2),..., fi(s,2m+2)) - фонемная разметка сигнала s.

По свойству 2° отображения фонемной разметки Vj=l..m Vk=fi(s,2j)+1, Ks,2j)+1, ..., |u(s,2j+l)-l s$_{[k·T(s)/2;k+T(s)/2]}$6n(bj) => d(s,k)e^bj)cc(bj) => r(d(s,k))ec(bj) и по лемме 2.1 ¥(d(s,k))cO(r(d(s,k)))cO(c(bj))r> <Г(Р(8,^,2i)+1)И(8,ц(8,2i)+2)|...И(8,ц(8,2i+1)-1||))^(0(c(Ь^))*).

По свойству 3° отображения фонемной разметки Vj=l..m+1 Vk=p,(s,2j-1), u(s,2j-l)+l,... |i(s,2j), выполнено, что либо

S[k-[T(s)/2],k+[T(s)/2]]en(bj.i)un(bj), либо не существует звука ЬєВ:

S[k-[T(s)/2],k_+[T(s)/2]]en(b) ^ d(s,k)e^b_j)uD($b_{j\cdot 1}$)u'D(bo)cc(b_j)uc(bj.i)uc_0 =>

r(d(s,k))Gc(bj)uc(bj.i)uc(b_0) и по лемме 2.1

^(d(s,k))cO(r(d(s,k)))cO(c(b_j)uc($b_{j\cdot 1}$)uc(b_0))=>

<f(||d(s,Ks,2j-l))|d(s,^(s,2j-l)+l)|...|d(s^(s,2j))||)c

e<R(0({c(^),c(6,$_{yi}$),Co}))*>.

Следовательно, (f(d(s))) = (f(||d(s,l)|d(s,2)|...|d(s,n)||)> = =

<f(||d(s,rtM))|d^ d(s,n<s,3))|...|M<s,2m+2)||)> = = <f

(||dCs,Ks,l))|d<:s,H<s,l)+l)|. - -|d(s,n(s,2))||)f (||dCs,M<s,2>4-l)|.. .|d(s,n(s,3)-

l)||)...f(||d(s,Ks,2m+l))|...|^(s$_5$2m+2)||))c

c (R(O(cO^c(b_1)uc_0))*R(O(c(b_1)))$^{!!!}$R(O(c(b_1) uc(b_2)uc_0))*

R(O(c(b_2)))*...R(O(c(b_n)))*R(0(c(b_n)ucO^c_0))*> = f ((3), что и требовалось доказать.

Теорема доказана.

Теорема 2.1 позволяет нам решить задачу локального распознавания речи в общем случае. Однако в связи с тем, что отделяющая функция \|/ для большинства пар фонем устроена плохо - фонемы плохо отделимы друг от друга для используемых функций описания сигналов, - практическое применение этой теоремы ограничено. Дело в том, что для различных пар команд (3 и {3' языки Г (р) и Г (р') часто пересекаются, что не позволяет использовать теорему 2.1. для решения задачи распознавания речи. Выход заключается в том, чтобы объединить некоторые фонемы в классы фонем и вместо разбиения пространства описаний сигналов на образы фонем использовать укрупнение этого разбиения - разбиение пространства BL^M на образы фонемных классов.

§2.3 Модель и алгоритм распознавания кода команды для случая классов звуков. Нижняя оценка качества словаря команд.

Пусть теперь 6=6^6^...^6^{1}^ - разбиение множества фонем на непересекающиеся классы, Б'сБ, 6^0 Vi=l..k. Рассмотрим разбиение

R^{M} на подмножества (ДБ^{1}), i=l..k}uP(B^{0}), где c_t =Я(В>= *{]Щ)* - Обра-$_{бєв'}$ зы классов фонем, а через c_0 =О(B^{0}) обозначено межфонемное пространство R^{M}\ *JV(B')* (откуда ДВ°)=Д$Ь_0$)). Сопоставим каждому элементу

с I этого разбиения букву c_4 кодовой книги С={CI, i=0..k}. В этом случае справедливо следующее утверждение:

Утверждение 2.3. Если для всех Cj,CjeC функция *F отделяет q и c_{J5} то

1)УрєЕ^{м}¥(p)={Г(p)}

2)VpESVyEd(n(b))(f(y))={r(y)}

Доказательство.

1) По лемме 2.1 имеет место вложение Г(р)є^{x}ї$^{/}$(p)c:0(Г(p)). Но по условию для кодовой книги С 0(с)=с для любой буквы с є С. Значит, ¥(p)={Г(p)}.

2) f(y)=Rmy^{1}))Rmy^{2}))...RWy^{n}))=(В силу 1) =

(Г(у^{1}))(Г(у^{2}))...(Г(у^{п})) => (Г(у)> = {Г(у)}

Утверждение доказано.

Содержательный смысл этого утверждения следующий.

Если разбиение множества фонем на фонемные классы произведено так, что удается отделить друг от друга с помощью функции Ч^{7} образы этих классов фонем, то

1) Локальное распознавание каждой точки пространства описаний дает только один вариант распознавания - код этой точки
2) Код любого произнесения восстанавливается с помощью функции локального распознавания *Ч** однозначно

Замечание. Для случая, когда разбиение задано с помощью классов звуков, эталонный код Г((3) любой команды (ЗєЗ, p=bib$_2$...b$_m$ будет равен f(P)=R(cQuc(b$_1$)uco)*R(c(b$_1$))*R(c(b$_1$)uc(b$_2$)uc$_0$)* R(c(b$_2$))*...R(c(b$_m$))*R(c(b$_m$)uc(J)uco)* (здесь * - символ итерации).

Предложим теперь механизм разбиения множества фонем на фонемные классы, позволяющего применять утверждение 2.3.

Пусть заданы речевая модель, функция отделимости \|/ и функция локального распознавания ¥. Введем на буквах кодовой книги С бинарное *отношение отделимости,* обобщив определение отделимости, приведенное ранее. Определение отделимости букв с и е ' (обозначение с~с') будем вводить индуктивно.

Определение 2.6.

1. Если сие' отделимы с помощью *Ш* в смысле определения 2.5, то с~с'.
2. Если с~с', с'~с'', то с~с''.

Очевидно, что отношение отделимости ~ является отношением эквивалентности. Классы эквивалентности, на которые это отношение разбивает элементы кодовой книги из примера 2.1, являются примером удачного разбиения множества фонем на фонемные классы, позволяющего использовать образы этих классов для построения алгоритма распознавания речи.

Пусть словарь команд В речевой модели <B,B,S,(R^M) ,С,П,Т,ц,с1,Г,г) имеет вид В={(3ь p_2v Pn}- Пусть заданы функции отделимости \|/ и локального распознавания *Ч**. Пусть кодовая книга С и соответствующее ей разбиение *С* заданы на основе образов классов эквивалентности введенного отношения отделимости ~, а кодирующее отображение Г определено с помощью ¥ на основе утверждения 2.3 (пункт 1). Определим алгоритм распознавания речи, сопоставив каждой команде Pi из В эталон

Єі=(г(РО) **и** введя на декартовом произведении С*хЕ, Е={ei,e_2,...,ек}

функцию расстояния $P(.У>^{e})$ - ' О, *если у Є е*
1, *иначе*

Справедливо

Утверждение 2.4. Для алгоритма распознавания речи (Е,р) для любого кодового слова уєГ(сі(П(В))) имеет место вложение г(у)сГ($_{E,P}$)(у)

Доказательство.

Пусть рєг(у). Тогда по определению г уеГ(ё(П((3))). По утверждению 2.3 уе(f (у))с(f (р))=е => по определению р р(у,е)=0 => Р є *Argmm(p(/,e_i))* =г$_{(E)P)}$(у). Итак, мы доказали, что г(у)сг$_{(E\ p)}$(у).

ДєЯ

Утверждение доказано.

Содержательный смысл доказанного утверждения состоит в том, что построенный на основе разбиения звуков на классы детерминирован-но-автоматный алгоритм распознавания речи каждый раз включает в результат распознавания все "правильные ответы".

Утверждение 2.5. Пусть задан введенный выше алгоритм распознавания речи (Е,р). Если словарь команд В нашей речевой модели та-

ков, что для любой пары команд р^єВ соответствующие им эталоны ej и ej не пересекаются, то г=Г(E,p).

Доказательство.

По утверждению 2.5 для любого у г(у)сг$_{(E>p)}$(у). Но если эталоны всех команд словаря попарно не пересекаются, то для любого у множество $\Gamma_{(E,p)}$(у)=Лг£1ШП(/?(>,e;))

МВ

по построению функции расстояния состоит ровно из одного слова. Поскольку г(у) не пусто (оно содержит Р: уєГ(с!(П(р)))), получаем, что г(у)=г$_{(E!P)}$(у). Утверждение доказано.

Утверждение 2.5 имеет следующее практическое значение. Если на этапе построения системы распознавания речи мы имеем возможность выбирать команды произвольно, то можно подобрать команды, эталоны которых не пересекаются. В этом случае алгоритм распознавания будет работать идеально.

Возникает вопрос, как, не проводя экспериментов, сравнить среднее число ошибок алгоритма (E,p) для двух различных словарей команд, каждый из которых содержит пары пересекающихся эталонов. С этой целью введем понятие качества словаря команд.

Качество словаря команд будем определять индуктивно. Положим, что качество словаря, состоящее из одной команды, равно 1. Для словарей из двух команд определим качество как:

Определение 2.7. Качеством словаря команд В={р,Р'} для введенного алгоритма распознавания речи (E,p) назовем величину

гда^ипот)

Содержательно качество $q(E,_P)(\{P>P'\})$ словаря {(3,p'} связано со средним числом ошибок при распознавании алгоритмом (E,p).

Пусть 5 - некоторое конечное множество, *f.* 5x5—»R. Обозначим через ©.(/) среднее значение функции/на парах различных элементов из *SxS:*

$$©_s(/) = \frac{£/(<r,a')}{S|(|S|-1)}$$

2 Определение 2.8. Качеством словаря команд *Ь,* состоящего из более чем двух команд, назовем величину $q(E,p)C8)^{-}0_8(q(E,p))$.

Лемма 2.2. Для любого словаря команд *Ь* $0<q_{(EiP)}(S)<l$.

Доказательство.

Для каждого у $Г(E,_P)(У)$ СОСТОИТ либо из одной команды (и в этом случае $Г(_{E;P})(у)= г(у))$, либо из двух команд. Следовательно, по определению 2.7 качество $q(E_{lP})(\{P,P'\})$ любого словаря из двух команд {(3,(3'} лежит на отрезке [0,1]. При усреднении по всем парам команд из словаря *Ъ* неравенство $0<q_{(E},p)(S)<l$ сохранится.

Лемма доказана.

Определим расстояние *p(e,e')* между множествами е и е' как

$$p(e,e') = \begin{matrix} ГО, & если\ ес\backslash ёФ0 \\ 1, & иначе \end{matrix}$$

Справедлива теорема о нижней оценке качества словаря команд для детерминированно-автоматного алгоритма (E,p):

Теорема 2.2. Для введенного алгоритма распознавания речи (E,p) для любого словаря команд *Ъ* справедлива оценка:

$$q_{(E,p)}(S) > 0_6(£).$$

Доказательство.

Случай словаря из одной команды для задачи распознавания речи лишен смысла.

Если словарь состоит из двух команд, то по определению

$$n \quad ([R\ /?'П - ? \qquad \underline{гбГ(<*(П(i)иП(Я))}$$
$$ЯшШР*)-* \quad |_{Гда(у}£)иП(Я))Г$$

Если эталоны команд **р** и р' пересекаются, то 0 (***р***)=0, откуда, используя лемму 2.2, получаем требуемую оценку для этого случая.

Если же эталоны команд Р и $р^1$ не пересекаются, то 0 (*Р)=1.*

Однако в этом случае для каждого *у* $г_{(E,}$р)(у) состоит ровно из одной команды, откуда и $q_{(EjP}$)(3)=1.

Усреднение по всем парам команд из словаря не изменит этой оценки, откуда следует справедливость теоремы для словарей с произвольным числом команд.

Теорема доказана.

ГЛАВА 3. ВЕРОЯТНОСТНЫЕ АВТОМАТНЫЕ МОДЕЛИ.

§3.1. Понятие монотонного автономного вероятностного автомата. Модель распознавания речи с помощью вероятностных автоматов.

Наиболее распространенным математическим объектом, использующимся на практике при построении алгоритмов распознавания речи, является вероятностный автомат. Обусловлено это тем, что применение конечных детерминированных автоматов, как это показано в предыдущей главе, позволяет решать задачу распознавания речи только для специально подобранных словарей команд. При создании эффективных алгоритмов распознавания речи в общем случае используются вероятностные автоматы.

Вероятностный автомат хорош тем, что он может применяться в качестве "эквивалента" множества кодов описаний всех речевых сигналов, являющихся произнесениями какого-либо слова словаря или звука речи. Вероятностный автомат в нашем случае используется в качестве генератора этих кодов.

Обучение вероятностных автоматов происходит по правилам, так что коды описаний произнесений «родных» команд порождаются ими с большой вероятностью, и эта вероятность выше вероятности порождения этих кодов «чужими» автоматами. Задача обучения параметров вероятностных автоматов находится за рамками настоящего исследования.

Пусть (B,B,S,(R^M) ,C,П,T,|и,(i,Г,г> - речевая модель, задан словарь команд S={pi, {Зг,... Pn} И решается задача распознавания кодов описаний произнесений этих команд, т.е. кодовых слов *yeT(d(nCB)))*.

Определение 3.1. *Вероятностным автоматом* называют четверку (B, C, Q, v>, где

B - конечный алфавит входных символов;

C - конечный алфавит выходных символов; Q - алфавит (в нашем случае - конечный) состояний автомата, v - функция, определенная на декартовом произведении BxQ и принимающая в качестве своих значений вероятностные меры на декартовом произведении BxQ ([26, 34]).

В содержательной интерпретации вероятностного автомата как преобразователя информации числа v(b, q)(c, q') є [0,1] означают условную вероятность перехода автомата в состояние q' и выдачи символа c, при условии, что автомат находился в состоянии q и получил на вход символ Ь. Будем обозначать поэтому v(b,q)(c,q') := v(c,q'/b,q). Числа v(c,q'/b,q) в силу этого обладают свойствами 0<v(c,q'/b,q) <1 и ^]v(c,g76,#) = l Vb, q.

Вероятностный автомат функционирует подобно детерминированному автомату. Находясь с некоторой вероятностью в некотором начальном состоянии q, автомат работает по тактам и в каждый такт времени получает на вход букву b алфавита B, выдает на выход букву c и переходит в состояние q' с вероятностью v(c,q'/b,q), и т.д.

Инициальным вероятностным автоматом называется пятерка (B, C, Q, v, v_0), где vo - вероятностная мера на множестве состояний Q, задающая распределение для начального состояния автомата.

Нам будет интересно рассмотреть частный случай вероятностных автоматов - вероятностный автомат Мура без входа. Именно такие автоматы, обладающие дополнительно свойством монотонности (о нем будет рассказано ниже), используются при построении алгоритмов распознавания речи.

Определение 3.2. *Вероятностный автомат Мура без входа* (далее *-автономный вероятностный автомат)* - это тройка (C, Q, v), где C - алфа-

вит выходных символов; Q - конечный алфавит состояний автомата, v -функция, определенная на множестве состояний автомата Q и принимающая в качестве своих значений вероятностные меры на множестве CxQ, такая, что она разлагается в произведение v=7i-P, где *n* действует на множестве Q и имеет значениями вероятностные меры на множестве Q, а P действует на том же множестве Q и имеет значениями вероятностные меры на множестве C ([26]).

Содержательно v(c,qVq) интерпретируется как условная вероятность перехода в состояние q' и выдачи символа *c* при условии нахождения в предыдущий момент времени в состоянии q. Условие представимости вероятностной меры v в виде произведения вероятностных мер тг и P означает содержательно, что два события - переход в следующее состояние и выдача некоторого символа на выход автомата - несовместны.

Прямоугольную матрицу будем называть *стохастической,* если все ее элементы неотрицательны и сумма элементов которой в каждой строке равна 1.

Как в общем случае, так и для автономных автоматов, функция v полностью определяется некоторой конечной системой матриц.

Так, для автономного автомата функция v задается стохастической матрицей 7Г размерности mxm (где |Q|=m) вероятностей переходов и стохастической матрицей P размерности mxk (где |C|=к), i-я строка которой задает распределение вероятностей выходных символов для состояния qj. Матрицы тс и P являются стохастическими.

Определение автономного вероятностного автомата можно обобщить на случай, когда матрицы *n* и P не являются стохастическими, но обладают

тем свойством, что сумма элементов в каждой их строке не превосходит 1 (будем такие матрицы называть *слабо-стохастическими).* Добавив дополнительное (поглощающее) состояние и дополнительную (пустую) букву, можно доопределить данный автомат естественным образом до автономного вероятностного автомата со стохастическими матрицами переходов $к^1$ и выходов $P^!$ следующим образом:

ж =

7Г	*л*
0	**1**

, *P′* =

P	*P*
0	**1**

(вектор-столбец *к* в матрице тс' состоит из чисел, дополняющих каждую из строчек матрицы % до 1, а вектор-столбец *P*, соответственно, дополняет до 1 распределение выходных букв в каждом состоянии, т.е. строчки матрицы Р)

Благодаря возможности такого обобщения многие утверждения, справедливые для автономных вероятностных автоматов, верны и для автоматов со слабо-стохастическими матрицами. Если это не оговорено особо, мы будем и такие автоматы называть вероятностными.

Будем считать, что выделено некоторое подмножество состояний автомата, которые будем называть финальными. Множество финальных состояний автомата однозначно задается вектором-столбцом ju_F, в котором в i-й позиции стоит единица тогда и только тогда, когда состояние qi - финальное, а иначе там стоит 0.

Будем считать, что вероятностный автомат, в котором заданы финальные состояния, функционирует таким образом, что при попадании в финальное состояние он прекращает свою работу, или, что эквивалентно, в финальном состоянии автомата (вне зависимости от наличия у автомата входа и того, какой символ поступил на вход, если у автомата вход есть) вероятность перехода в себя и в любое другое состояние равна 0. Вероятностные автоматы, в которых выделены финальные состояния, будем обозначать, включая вектор финальных состояний в число объектов, определяющих автомат, на-

пример, (B, C, Q, v, v_0, v_F). Заметим, что матрица вероятностей переходов тс в таком случае уже не будет стохастической, т.к. финальным состояниям в ней будут соответствовать нулевые строки.

Для заданного автономного инициального автомата *A=(C,* Q, тс, P, Vo, VF) можно рассматривать вероятность р(у) того, что автомат перешел в финальное состояние, выдав при этом на выход слово у є С*. Известно ([26]), что р(у) можно посчитать с помощью матричной формулы

р(у) = $ц_0$М(у)ць где М(у)=М(с!$С_2$.. .$с_п$)= М(сОМ($с_2$). • -М($с_п$) (1)

(здесь через М(с/) обозначена матрица вероятностей переходов автомата и одновременной выдачи символа С/, *M(ci)y=miyPц).*

Более подробно эта формула выглядит так:

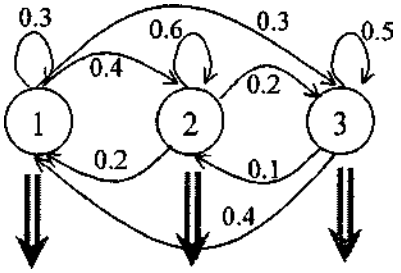
Р(У)= *ЦР(Г\Ю =* ^(чО^КА^К, •••••W^,

В дальнейшем мы будем рассматривать только автономные вероятностные автоматы, поэтому слово автономный для простоты будем иногда опускать.

Пример 3.1.

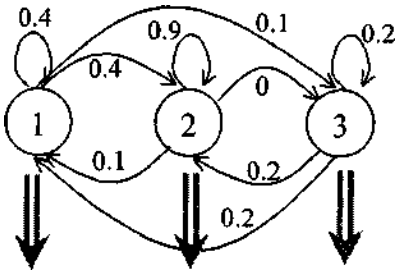
Пусть вероятностные автоматы *А* = {{a,Ь,c},{ 1,2,3}, $тс_ь$ Pi} и

A_2 = {{a,Ь,c},{ 1,2,3}, $тс_2$, Рг} представлены диаграммами:



a	0,2	a	0,1	a	0,8
b	0.5	b	0.9	b	0.1
c	0.3	c	0	c	0.1

Автомат т4]



a	0,3	a	0,7	a	0
b	0	b	0.1	b	1
c	0	c	0.1	c	0

Автомат $Д_2$

У автомата *A* матрицы переходов и выходов являются стохастическими:

$$7Ui = \begin{vmatrix} 0,2 & 0,5 & 0,3 \\ \mathbf{0Д} & 0,9 & 0 \\ 0,8 & \mathbf{0Д} & \mathbf{0Д} \end{vmatrix}$$

$$\begin{vmatrix} 0,3 & 0,3 \\ 0,2 & 0,2 \\ 0,4 \quad \text{ОД} & 0,5 \end{vmatrix}, Pi = \begin{vmatrix} \\ \\ \\ \end{vmatrix}$$

У автомата *Aj*, напротив, ни матрица переходов, ни матрица выходов стохастическими не являются:

$$\text{^2} \begin{vmatrix} 0,3 & 0 & 0 \\ 0,7 & \mathbf{0Д} & \mathbf{0Д} \\ 0 & 1 & 0 \end{vmatrix}$$

$$\begin{vmatrix} 0,3 & 0,4 & \mathbf{0Д} \\ \mathbf{0Д} & 0,9 & 0 \\ 0,2 & 0,2 & 0,2 \end{vmatrix}, P_2 = \begin{vmatrix} \\ \\ \\ \end{vmatrix}$$

Матрицы Mj(a), Mj(b), Mj(c), i=l,2, участвующие в формуле (1), для автоматов *A* и *Ai* имеют следующий вид:

$$Mi(a) = \begin{vmatrix} 0,06 & 0,08 & 0,06 \\ 0,02 & 0,06 & 0,02 \\ 0,32 & & 0,08 \\ 0,4 & & \end{vmatrix}, M,(Ь) = \begin{vmatrix} 0,15 & 0,2 & 0,15 \\ 0,18 & 0,54 & 0,18 \\ 0,04 & & 0,01 \\ 0,05 & & \end{vmatrix}, M,(c) = \begin{vmatrix} 0,09 & 0,12 \\ 0,09 & \\ 0 & 0 \\ 0\ 0,04 & 0,01 \end{vmatrix}$$

0,05

$$M_2(a)=\begin{Vmatrix} 0,09 & 0,12 \\ 0,03 & 0,07 \\ 0,63 & 0\ 0 \\ 0 & 0 \end{Vmatrix}, M_2(Ь)=\begin{Vmatrix} 0 & 0 \\ 0,01 & 0,09 \\ 0,2 & 0,2 \end{Vmatrix}, M_2(c)=\begin{Vmatrix} 0 & 0 \\ & 0 \\ 0,01 & 0,09 \\ & 0 \\ 0 & 0 \\ & 0 \end{Vmatrix}$$

Каждому автономному вероятностному автомату естественным образом ставится в соответствие *словарная функция* р: С*-»[0,1], сопоставляющая каждому слову у из С* вероятность р(у). Эту функцию можно рассматривать как счетное множество пар вида (у,р(у)), где пары следуют согласно лексикографическому порядку, заданному на словах у.

Определение 3.3. Два инициальных вероятностных автомата называются *эквивалентными,* если тождественно равны их словарные функции.

Определение 3.4. Для автономного инициального автомата

т4={С,Q,тс,Р,Vo,v_F} квадратную матрицу *я,* где *Пц* = $n^{к}$^P_a, будем называть

приведенной матрицей переходов (здесь к=|С| - мощность выходного алфавита).

(ij)-ii элемент приведенной матрицы переходов имеет смысл вероятности того, что автомат перешел из i-го состояния в j-ое и подал какую-то букву на выход.

Лемма 3.1. Для любого автономного вероятностного автомата r4=<C,Q,7i,P,Vo,v_F) выполнено равенство $\wedge p(r) = V_0 K'' V_F > {}^{г}Д^{e}$ *я* " приведенная

$_{ГеС':|y|=n}$

матрица переходов автомата *А.*

Доказательство.

$$2>(r) = \sum_{y\epsilon C:\backslash_r\backslash=n} \sum_{r=c_h c_h \ldots c_{jn}} 2\, v_0(fti)^{*>} (c_л\, к_{i(Д}\, (c_л\, x_2, \bullet \ldots \bullet p_{la}\, \{_{Cjn}\, к_л$$

$$L\text{—}i \qquad I\text{—}i \quad UVJ/, / \quad /, V \quad y, \bullet \quad /i/_2 \quad '2^{V} \quad \wedge 2^{y} \quad '2'3 \quad 'я^4 \quad Уп^7 \quad Vn$$

$$Z^{v}o(Qi\ )Яц^{\Pi}ïI^{в} — 'Яi\ I \qquad У\wedge\ (c1\ )P_t\ (c,\ )\text{-}\ldots\text{-}£\ (c,\ )$$

$$r^{sC}JI^{c}J2\text{-}^{c}J,$$

$$Z \qquad vofe,\ )¾¾ * - ■ \wedge_{л+1}\ Z\wedge,\ (<v,\ )Z\wedge_2\ (^{c}л\ )\text{-}IX\ (^{c}л)$$

$$^{v}F\ (<II_{й+}]H$$

$$4=?,y7/2\text{-}<7,Л_{+}i$$
$$M?/_{„+1})=l$$

Лемма доказана.

Следствие 3.1. Для любого автономного вероятностного автомата A={C,Q,7C,P,Vo,v_F) со стохастической матрицей P выполнено $\sum_{y\epsilon C^*:\backslash\Lambda=n} \wedge p(y) = v_b n^n v_F$.

Доказательство

Для этого автомата матрицы *n* и *n* совпадают, откуда и следует искомое равенство.

Лемма доказана.

Лемма 3.2. Если R - квадратная матрица (такая что матрица E-R - невырожденная) и матричный ряд $\sum_{/7=1}^{+co}]Гд^{и}$ сходится (т.е. сходящимися последовательностями являются все соответствующие элементы $\prod_{/7=1}^{N}$ матриц $\wedge R''$,

N= 1,2,...), то сумма этого ряда равна $(E- R)''^{1} - E$.

Доказательство.

$$N \qquad N \qquad N+l \qquad XT\_L1$$

$$\sum_{/i=0}^{N} J\,Л'' = (E-R^{N+l})(E-RY^{1} \prod_{/1=1}^{л=1} E \qquad iN+b$$

$$=> £V$$

$$=(E-R^{N+l})(E-Ry^{l}-E$$

$$«=0$$

При стремлении N к бесконечности $i?^{N+1}$ из условия сходимости ряда

$\sum_{n=1}^{+\infty}$ J\R" стремится к нулевой матрице, откуда получаем искомое

тождество. Лемма доказана.

Определение 3.5. *t-м косым рядом* размерности m называется матрица I_t размерности mxm, у которой на t-й диагонали стоят 1, а все остальные элементы равны 0, т.е. It cy)1 <=> $\}^{=t+}$ЪIt(ij)= 0 <=> j^i+t

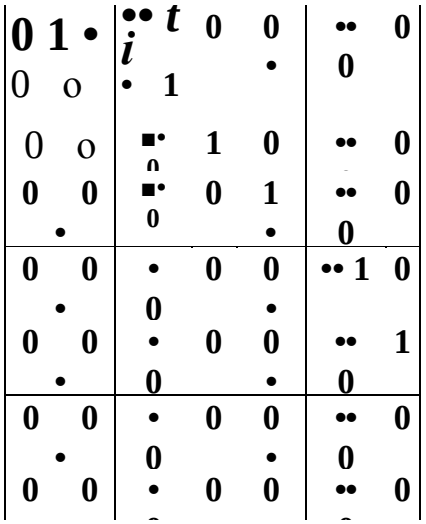


Рисунок 3.1. t-йкосойряд.

Заметим, что Io - единичная матрица, а I_t при t>m-l - нулевая матрица. Кроме того, 1-й косой ряд обладает свойством: $Ii^{t=}\mathbf{I_t}$.

Определение 3.6. *Жордановой клеткой* J(X) размерности m называется квадратная mxm матрица J(A,) = A£+Ii.

Из определения жордановой клетки следует, что на ее главной диагонали стоят числа *X,* на диагонали над главной - единицы, а все остальные элементы равны 0.

Лемма 3.3. Для *X: 0<к<1* матричный ряд $\sum_{и=0}^{+\infty} J(X)''$ сходится.

Доказательство.

В следующих выкладках равенство матричных рядов понимается в том смысле, что они либо оба сходятся к одной и той же матрице, или оба расходятся.

$$\sum_{=}^{+00} \text{Ёля)-} = \sum_{}^{+00} \text{5К}^{+}\text{'.)''} = \sum_{и}^{+\circ\circ} \text{Щя'/г-}^{1} \text{—} = \sum_{}^{+00} \text{XI}^{I}\text{,-}^{\text{-}}$$

to to to to /!(«-/)! to to /!(л-/)!

+со +00 i ic-n-i +Ю +00 *(il. 1 г* +» m-1 i'; _i_ Л-Ч• m-1 +00 /• . L\P

Таким образом, все элементы на k-й диагонали суммы данного матрич-

ного ряда равны $\sum_{}^{+\circ\circ} \frac{a + kv}{i} = \frac{X(i +}{}$ *k)(i* + £ -1)...(/ + Г). Ряд, стоящий справа,

при 0<А,<1 является сходящимся, поэтому сходится и данный в условии матричный ряд.

Лемма доказана.

Перейдем теперь к рассмотрению частного случая автономных вероятностных автоматов. А именно:

Определение 3.7. Инициальный автономный вероятностный автомат 74=(C,Q,7r,P,Vo,v_F) будем называть *монотонным,* если выполнены условия:

1) T^j = 0 при i>j

2) v_0=(1,0,0,...,0)

3) v_F^T=(0,0,. -.,0,1)

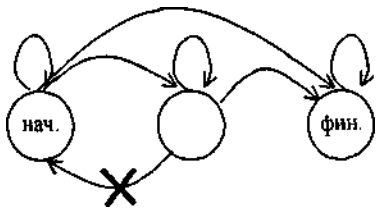


Рисунок 3.2. Монотонный автомат.

Теорема 3.1. Если для монотонного автомата *4=(C,Q,7I,P,VO,VF) при

ВСеХ I 7ljj< 1, то ^ $p\{y)$ < 1.
$y \in C$

Доказательство.

Заметим, что]Г/?00 $= \sum_{n=\Gamma y.|y|=n}^{+00} P(Y)$ и если РЯД в правой части равенства

сходится, то сходится и ряд слева.

Зафиксируем некоторое натуральное N.

$\pounds 2>0) = 1><\sum_{n=1}^{N} Ч = ^o !>" \Gamma^{N} \quad \backslash^{V}F$ (по лемме 3.1).

$\sum_{\Lambda=1 \underline{y:|y|-n}}$ $\sum_{n=1}$ $\backslash n=\backslash$)

Устремим N->со и докажем сходимость ряда, стоящего в скобках. Для этого представим матрицу *ft* в Жордановой форме; *it* =*QJQ'*\ где Q - некоторая невырожденная матрица, а J - Жорданова форма матрицы *И.*

$$\pounds^{*"} = fj\&or') = 1_L Q^{J"} Q \sim^{X} = \text{fifi}^{\wedge}V^{1} \qquad (2)$$

Собственные числа матрицы *it* являются решениями уравнения *n*-Щ=0. Однако матрица *И* - верхнетреугольная, поэтому |тг-А£| =

(й\\-X)(n 22-^)- --(^пЯт-Г^Х"^)- СлеДОВаТеЛЬНО { $\pounds_{1b}$ #22, ---^m-lm-b 0} - Соб-

ственные числа матрицы *n.* Жорданова форма J является блочно-диагональной матрицей, поэтому J^n=diag(J(A;)^n, *](Я2)*^{n},... J($\wedge_s$)^n), где *{Я], Л2,... Xs)* - различные собственные числа из множества *{it u, it*22, .--^т-іт-ь 0}- По лемме 3.3 ряд, составленный из степеней Жордановых клеток, сходится, поэтому сходится и ряд из степеней Жордановой формы матрицы *it.*

По лемме 3.2 получаем $\sum_{1=1\ У'-|y|=n}^{(+\ 00} \pounds/Ю0 = JU_0 \wedge \sum_{\backslash л=1}^{Л} m\varepsilon'' JU_F = //_0((\pounds\text{-тг})^{1}\text{-}E)/Л_P =$

$((\pounds - тг)"^{1} - \#)_{,,} = ((E - i)^{-1})_{,m}$, т.е. искомая сумма равна правому верхнему угловому элементу матрицы $(E - n)\sim^{x}$. Докажем, что он не превосходит 1.

Пусть X = *(E - fty^x*. Тогда из верхнетреугольности матрицы *ft* следует, что и матрица X имеет также верхнетреугольный вид. Причем из условия (E-£)X=E следует равенство (E-;г)x=(0,0,...,0,1)^T, где x - последний столбец матрицы X, откуда:

1) (1-я-ц)Xн = 1 => $X_и$ = ———, в частности, X_{mm}=l,
1 - 7V_n

2) (l-£$_m$-lm-l)X_m-i_m + (-^$_m$-lm)l=0, ОТКуда $X_{m·lm}$ = *ч_ы . Из УСЛОВИЯ
*■ ^{n}*m*-*m*-\

^m-lm-1+^m-lm ^ 1 СЛЄДУЄТ, ЧТО *ft$_{mAm}$* < l-#$_m$-lm-b СЛЄДОВаТЄЛЬНО, X_m_i_m<l.

m) (l-^n)Xi$_m$ + *{-ft 12* X_{2m}) + (-^13 X_{3m}) +...+*{-ft$_ы$* X_{mm})=0, откуда (1-^ll)Xi$_m$ = *ft \2* Xгт +^13 Xз$_г$ +...+ *ft \$_m$* X_{mm} < *ft \$_2$* + #13+...+ *ft \$_m$* < \-*ft* \\, ОТ-куда Xi$_m$<l, что и требовалось доказать.

Теорема доказана.

Следствие 3.2. Для любого монотонного автомата A=(C,Q,7r,P,v_0,v_F), такого, что Vi 7iн<1, а матрицы P и л (кроме последней строчки) - стохастические, выполнено]|Г *P(y)* =! •
y

Доказательство.

Доказательство для этого частного случая автоматов отличается от доказательства доказанной выше теоремы тем, что в этом случае *ft -n* - стохастическая матрица по всем строчкам, кроме последней.

Поэтому если X = *(E-л)*$^{-1}$, то из условия (E-яг)x=(0,0,...,0,1)^T, где x -последний столбец матрицы X, получаем:

1) (1-¾)¾ = 1 => $X_и$ = ------ , в частности, X_{mm}=l,
1-я-,

2) (l-^.im-i)X_m.i_m + (-^.im)l=0. В нашем случае л^-im-i+^i-im = 1, отку да следует, что яьi-im = I-я^-іт-ь следовательно, $X_{m·lm}$=l.

m) (1-яїі)$X_{1т}$ + (-яІ2 X_{2m}) + (-яІЗ X_{3m}) +...+*(-щ$_m$* X_{mm})=0, откуда (1-

^ll)Xi_m = ЛГ12 $X_{2т}$ + Яіз $X_{3т}$ +…+ *Щ$_m$* $X_{тт}$ = #J2 + *Щз* + ■••+ Яіт = 1-7Гц, ОТКуДД

$Xi_т$=1, что и требовалось доказать.

Следствие доказано.

Доказанное следствие имеет следующий смысл: события "монотонный автомат завершил работу и выдал слово у" для разных у несовместны.

Из доказательства теоремы 3.1 следует, что для монотонного автомата т4=(С,Q,7С,Р,|LIO,|LIF) сумма матричного ряда]Г£" равна матрице *(E-Sty*^{1}, кото-

$_{я=1}$

рую будем называть *предельной приведенной матрицей переходов.*

(у)-й элемент предельной приведенной матрицы переходов - это вероятность того, что автомат перешел из i-го состояния в j-ое за конечное число шагов.

Введем теперь вероятностно-автоматный *алгоритм распознавания речи.* Пусть одним из известных методов ([18-22]) на основе набора примеров-кодовых слов каждой команде (3; словаря команд В сопоставлен вероятностный автомат *А*\. Назовем эталоном команды Р; стохастическую словарную функцию Єi=p_Ai- Расстояние между кодовыми словами и эталонами команд определим как р(у,Єi)=1-p_Ai(у).

Алгоритм распознавания речи (Е, р) называется традиционно методом скрытых марковских процессов. Результат $г_{(E;\ p)}$(у) распознавания кодового слова у этим методом - это команда р, максимизирующая вероятность РАІ(У) наблюдения слова у на выходе вероятностных автоматов Д.

§3.2 Метрика ρi на множестве стохастических словарных функций.

Словарные функции, удовлетворяющие условию £/>00 = 1, будем называть *стохастическими.* Множество стохастических словарных функций будем обозначать через /s. Если £/>00<1, словарные функции будем называть *слабостохастическими.* Множество слабостохастических словарных функций будем обозначать через /s'.

Если на множестве стохастических (слабостохастических) словарных функций введена метрика, то она порождает метрику на вероятностных автоматах, где *расстоянием между монотонными автоматами* будем называть расстояние между соответствующими им стохастическими словарными функциями.

Каждую стохастическую словарную функцию {(у, р(у)} можно представить в виде вектора р=(р(уi), р($у_2$), ..., р(уО,...)» ГДе слова $у_ь$ $у_2$, ..., у;,,... занумерованы в лексикографическом порядке. Будем обозначать множество таких стохастических векторов также символом /s (или 4' для слабостохастических словарных функций).

Метрику на множестве стохастических (слабостохастических) словарных функций можно ввести различными способами.

Поскольку £/700=1, то рє/ь где /i={(ХьХ$_2$,...$Х_п$,...): £|х,,|<оо}

([29]). Введем на множестве 4 метрику из пространства *1*\.

Определение 3.8. Расстоянием ρi между стохастическими (слабо-стохастическими) словарными функциями Рi=$^{=}$ {(у,Рi(у)),уєС } и Р2=$^{=}$ {(у,Р2(у)),уєС } назовем функцию

Рі(РьР2)=£іАО)-лО)і
у

Утверяедение 3.1. рі - метрика в /$_s$ и в /$_s$'.

Доказательство.

Проверим корректность определения. Ряд сходится, т.к. рі(рь рг)

у *у* *її*

Выполнимость свойств данной метрики из пространства *1*\ проверяется просто:

1) Симметричность очевидна
2) рі(р, р)=0 - очевидно
3) Неравенство треугольника: $\wedge\backslash P\backslash(y)-p_г\{y)\backslash-$
у

$(T,\backslash МГ)-МГ)\backslash + ВР1(Г)-P_2(Г)\backslash)=$
у *у*

$\wedge\backslash \boldsymbol{МГ)-P_2(Г)\backslash-\backslash МГ)-МГ)\backslash-\backslash МГ)-P_2(Г)\backslash\wedge}$
у

Утверждение доказано.

§3.3 Метрика p2 на множестве стохастических словарных функций. Функция качества словаря команд.

Более естественной метрикой на множестве монотонных автоматов является метрика p_2. Она имеет прямую связь с тем, как человек воспринимает различные слова и звуки и оценивает их похожесть.

Определение 3.9. $p_2(Рьр2) = 1- \min_{усе'}(A(r)\wedge_2(r))-$

Утверждение 3.2. p2 является метрикой на множестве /s.

Доказательство.

Проверим корректность определения. Ряд сходится, т.к.

yeC' yeC'

Проверим выполнимость свойств метрики:

1) Симметричность очевидна
2) $p_2(p, P)=l- \underset{yeC'}{Xmin}(p(r),p(r))=l- \underset{yeC''}{]>}(г)=1-1=0$
3) Неравенство треугольника.

Необходимо доказать, что для любой тройки словарных функций рь р2, Рз выполнено условие:

$p_2(рь\ p_2)<p_2(рь\ Рз)+Р2(Рг, Рз)$, что эквивалентно

$$1- \underset{yeC'}{2тіп}(\wedge(\wedge),/?_2(/))-1+ \underset{yz.C'}{\wedge тіп}(_л(/),_л(г))-1^+ \underset{yєC*}{2^{тіп}\wedge 2}(г)'Рз(г))\wedge 0,$$

или

$$\underset{yeC'}{2^{min}}(\#(У)>Рз\,00) + тіп0_2(y), p_ъ(y)) \sim minfo(y), p_2(/)) < 1$$

Рассмотрим следующие случаи:

а) у: pi(y)<p_2(y)^рз(y), обозначим множество таких у через y_{l2}з

б) у: pi(y)<pз(y)<P2(y)? обозначим множество таких у через $y_{ш}$

в) у: p2(y)<pi(y)-Pз(y)? обозначим множество таких у через y_2n

г) у: P2(y)^pз(y)<Pi(y)> обозначим множество таких у через y_2зi

Д) У: Pз(y)<pi(y)^P2(y)$_5$ обозначим множество таких у через узi2

е) У: Pз(y)<P2(y)<Pi(y)$_3$ обозначим множество таких у через $y_3$2i

Для каждого из случаев а)-е) посчитаем, чему равно f(y)=

$min(p_J(y), p_ъ(y)) + mm(p_2(y), p_ъ(y)) - mm(p_x(y), p_2(y))$

а) f(y)=Pi(y)+P2(y)-pi(y)=P2(y) Vy£y_{123}

б) %)=Pı(у)+Pз(у)-Pı(у)=Pз(у) Vy£y_{132}

в) f(y)=Pi(y)+P2(y)-p2(y)=Pi(y) Vyey$_{21}$3

г) %)=Pз(y)+P2(y)-p2(y)=Pз(y) Vyey$_{23}$i

Д) %)=Pз(y)+Pз(y)-pi(y)=2p$_3$(y)-pi(y) Vyey$_{312}$ є)

%)=Pз(y)+Pз(y)-p2(y)=2pз(y)-p$_2$(y) Vубy_{32}i

Нам необходимо доказать, что]£/(» ^ 1 •

Пользуясь тем, что все множества ур попарно не пересекаются, преобразуем левую часть неравенства:

$$I\ /00 = E/oo + 1/0-) + E/oo^{+}\ Елг) + ''£/(/) + £/oo =$$

$$= 2>(г) + EA00 + 5>,(г) + 1A(г) + 1(2_A(г) - A(г)) +$$

$$T(^2 P3(Г) - P_2(Г))$$

$$= 2>2<Г) + 2>з(г) + 5>.00^{+} ЕЛ(Г) + Sfe(r) - A(r)) + X>з00 +$$

$$I(й00 - л00)^{+} 2>_3 00 =$$

$$2>зОО^{+} ЕЛ(Г) + 2>00 + Z(ft(T) - A(r)) + S(ft(y) - A(r)) -$$

Далее, поскольку $p_3(y)-p_2(y) < 0$ при уєу$_3$і$_2$, то JfaOO-AOO^O.

Аналогично ^$(p_3(y) - p_2(y))^{<}$®-

Так как $p_2(y)<p_3(y)$ *приуеу$_т$,то* £$p_2(y)$< £р$_3$*(У)*• Аналогич-но

£/>,(г)^ Е^зОО-
ГЄГ213 Г$^{\varepsilon}$/213

Следовательно, £/00^ 2>з(г) + 1>$_3$(г)+ Х^з(г)=
Г /^132^231^312^321 ГЄ/123 /^^213

= ^*Р3(у)* = ^*Р3(у)* =1»$^{чт0\ и}$ требовалось доказать.
УЄГізг ^Ггзі "Гзі2 ^Гз2і '-Гш ^*Гт* *У*

Утверждение доказано.

Замечание 3.1. Функция p_2 не является метрикой на множестве /$_s$', т.к. свойство 2 метрики для слабостохастических словарных функций не выполнимо.

Будем называть стохастические словарные функции рі и p_2 *«абсолютно разными»,* если выполнены условия:

1. Если рі(у)>0, то рг(у)=0
2. Если р2(у)>0, то рі(у)=0

Утверждение 3.3. На множестве стохастических словарных функций 4 метрики рі и p_2 совпадают с точностью до постоянного множителя.

Доказательство.

Р2(РьР2)=1-Е^{шіп}(А0').Л(г)) = $\%^{MF)+\frac{P.JF)}{t}}$ - Етіп($_{А}$(г),p_2(г))

Sl '$_{А}$ОО+л(?0 -тіп^О^О)) -у$\underline{I^i(г)-/?_2(У)I}$-1 Рі(РьРг)-

Утверждение доказано.

Определение 3.10. Пусть W и V - матрицы одинаковой размерности. Обозначим через *min(W, V)* матрицу той же размерности, элементы которой задаются соотношением *min(W, V)i/=min(Wij, Уф.*

Утверждение 3.4. ПуСТЬ ^^ICQbTTbPbVojVp} и

$42^{=}${C,Q_2,7r_2,P2,Vo,v_F} - монотонные вероятностные автоматы, , где

IQiMChhm, Vo=(1,0,...,0), v_F=(0,0,...,0,l). Пусть p! и p_2 - соответствующие этим автоматам стохастические словарные функции.

Тогда $p_2(A$ ь $A_2)$= 1 - $J^{min}_{V_2}$(^i 00» ^{M}2 00) $\overset{\backslash}{>}$ ^{г}Д^{е} Mj - матрицы из (1).

Доказательство.

Действительно, **p_2(Л ь *А* $_2$)=1-Е^{тіп}(А0^{г}).л0'))=Ь**

$_{yeC}$

1-X^{m}тм(^{M}Ш$_{m}$,м$_2$(у\$_m$) 1-

/єС* $_{yeC}$

2тіп(М,(/),M_2(г))$_{ии}$ =1- $Y_j mm\{M_1\{y),M_2\{Y))$

(здесь v'$_0$ = (v$_0$,0); v'$_F$ = (^,0)).

Утверждение доказано.

Поскольку ряд $\wedge iun(Pi(y),p_2(y))$ сходится, можно вычислять рас-

стояние между автоматами со сколь угодно большой точностью є, задавшись достаточно большим натуральным N, по формуле:

$9i\{A$ ь $A_2)$« $p_2^N(*41, A_2)=1-\wedge^{min}_{|/|<JV}(Pi(г),P_2(г))$• Доказанное утвержде-

ниє позволяет оценить число матричных умножений при вычислении расстояния между автоматами для векторов UQ и $\backslash i_F$ специального вида. Сформулируем это более строго в виде следствия:

Следствие 3.3. Пусть даны два монотонных автомата Ar{C,Qi,7Ti,P_b|i(b^F} и $Д_2$={C,$C>_2$,TC_2,P_2,ЦO,ЦP}, такие, что |СЫ=|$C>_2$|=т, Цо=(1Д- ..,0),JLI_F=(0,0,...A1).

Пусть задано натуральное N. Тогда для приближенного вычисления расстояния между автоматами по формуле $p_2^N(A \backslash, A_2)$=1-

$\wedge тіп(M,(>),M_2(;к))$ требуется порядка $2m^N$ умножений матриц размерно-
$\{\backslash№ \quad)_ы$

сти m x m.

Доказательство

Для расчета M;(y) требуется |y| матричных умножений. Поскольку для вычисления Mi(y) можно воспользоваться индуктивной формулой M;(y)=Mi(y')Mi(c), где y=y'c, c є C*, c є C, т.е. достаточно одного дополнительного матричного умножения, то общее число умножений матриц не

превышает $У^N 2m^к = 2(\ ^{m^{N+l}}\_j$ --------- 1)~2m .

Следствие доказано.

Заметим, что если не использовать зависимость M;(y)=M;(y')M;(c), то для приближенного вычисления расстояния между автоматами требуется

уже порядка $2Nm^{N+1}$ умножений матриц, так как общее число умножений матриц в этом случае равно $2^{\wedge}2km = 2т^{0}\frac{Nm^{N+l}-(N+\backslash)m^{N}+l}{\backslash^{m}\sim 1)} \approx 2Nm^{N+1}$

Пусть решается задача распознавании речи алгоритмом (Е,р) в словаре Ъ из двух команд В={|3ь Рг}- Эту задачу можно рассматривать как задачу *классификации([35])* слов, выдаваемых соответствующими командам вероятностными автоматами $A\backslash$ и A_2. А именно, пусть в каждый момент времени срабатывает только один автомат, автоматы равновероятны, и задача состоит в том, чтобы определить, какой из двух автоматов выдал заданное кодовое слово. Данная задача решается с некоторой ошибкой -*ошибкой классификации,* вероятностью которой будет вероятность такого события, когда слово было выдано одним автоматом, но отнесено нашим алгоритмом распознавания (Е,р) к выходу другого. Ошибку классификации можно называть *ошибкой распознавания,* если выдача кодового слова у автоматом $A\backslash$ происходит точно тогда, когда уєГ(с!(П(pi)).

Вероятность ошибки классификации будет задаваться по формуле:

$$Рош = P(\text{ТЫЕАГ}) = Pi\,\text{IГ,А})P(\text{У I А})P(A,) + P(г_{(E,р)}(y) = \text{Д}\,|y,A_г)P\{y \mid A_2)P(A_2))= г^{*?}$$

$$= \text{^ЫI}_{E,р)}(y) = 13_2\backslash\text{гЛ})\text{Пг}\backslash\text{А}) + P(г_{(Ер)}(y) = \text{^}\backslash y,A_2)P(_Г\backslash A_2))$$

Здесь Р(Д;)=1/2 - вероятность "срабатывания" i-го автомата, $P\{г_{(E,р)}\{y) = P_2\ \backslash y,A_x)$ - вероятность того, что классификатор определил, что "сработал" 2-й автомат, а на самом деле у выдано первым автоматом, $P(^{Г}(E,_{P)}\ (У) = fix \backslash Г\,Л)$- наоборот.

Предположим, что для каждого слова у нам известны вероятности pi(y) и p_2(y) появления слова у на выходе автоматов $A\backslash$ и A_2 соответственно. Справедлива

Лемма 3.4. p_2(Ль A_2)=1-2 $p_{ош}$..

Доказательство.

1-2 $p_{ош}$. = *l-2^(P(r{E,p)(y) = A |y,Ax)P{У \ A) + Щ{E,P)(У) = A |y,A2)P(У I A))*

= 1- $\underset{yeC}{If}$(rw(r) = AI *уЛ)ПУ \ A) + ЦР(*$\underset{yeC}{г_{E,p)}}$*(y) = AI У,A2)P(У \A)-*

$\underset{yeC'}{T,^{P}}$^ЕАУ>Рг\УЛ)РIУ\А)= $\underset{yeC*}{1^P}$(Рг(У)<Рг(У)\УЛ)МУ)

Но $P(p_2(y) < {}_{Px}(y) \mid y, A) = \begin{cases} 1, & \text{если } p_1(y) < p_x(y) \\ 0, & \text{иначе} \end{cases}$, отсюда

$\underset{\Gamma\epsilon C*}{Y,P}$(Pi (У)<PI(У)\ГЛ) A (У)= $\underset{\Gamma\epsilon C*:p_2(?')<Л(У)}{ф}$ S A (У) =>

$1\text{-}2p_{ош} = 1 - \underset{У*C*:p_2(y)<p_x(y)}{^} \; ^A(\Gamma) - \underset{\Gamma\epsilon C:_{Pi}(y)<p_2(y)}{t} 2>_2(\Gamma) = 1 - \underset{yeC*}{2^{min}}(A(r),A(r)) =$

=*p2(A\,A2),* что и требовалось доказать.

Лемма доказана.

Содержательно данная лемма связывает расстояние между двумя автоматами как эталонами речевых объектов (слов, звуков и т.п.) с вероятностью ошибки распознавания в словаре из этих двух объектов, т.е. с вероятностью "спутать" произносимые слова (звуки и т.п.) при условии, что обучение (синтез) эталонов команд произведено идеально.

Введем теперь *функцию качества словаря команд* для вероятностно-автоматного алгоритма (E,p) распознавания речи.

Для словаря Ъ из двух команд S={pb P2} определим Я(E,P)CB) как Я(E,P)C8)=1-2p0Ш., где рош. - вероятность ошибки распознавания в этом словаре команд, о которой говорилось выше. Для словаря с произвольным числом команд, как и раньше, положим Ч(E,P)CB)=0B(Я(E,P)).

Теорема 3.2. Пусть *Ъ={$\,$2,---* Рп} - словарь команд, *A={Ai,A2,...,An}* - соответствующие этим командам вероятностные автоматы. Тогда q(E,p)CB)=0A(p2).

Доказательство.

Для доказательства теоремы достаточно применить лемму 3.4. Для каждой пары {pь p2} команд из словаря В имеем *q(E,p)CB)=l-2poul=p2(Ai, Ai).* Усредняя по всем парам команд, получаем искомое утверждение теоремы.

Теорема доказана.

Теорема 3.2 дает нам способ точной оценки качества произвольного словаря команд для алгоритма распознавания речи методом скрытых марковских моделей. Качество определяется с помощью метрики, введенной на множестве эталонов команд.

§3.4 Метрика рз на множестве стохастических словарных функций.

Третий вариант введения метрики на множестве стохастических словарных функций состоит в следующем.

Если представить стохастическую словарную функцию {(у, р(у)} в виде стохастического вектора р=(р(їі)> РСїг), • • • > р(у0>---)^{c}4> то $\underset{\text{геС*}}{\wedge p(y)=1}$

$\Rightarrow$ $\underset{\text{уеС'}}{2>^{2}00}$^1, значит, ре/$_2$, где /$_2$={(хьх$_2$,...х$_п$,...): £х,,$^{2}_{\text{и=1}}$ $\overset{+00}{<}$°°) ([29]).

/$_2$ является евклидовым пространством со скалярным произведением (х, у)=$\overset{+00}{£}$X7,,, нормой ||х||=V(x,x) и метрикой р(х, у)=||х-у|| ([29]).

Можно ввести в /$_2$ отношение эквивалентности следующим образом: х~у, если BkeR, k^O: х=ky. Это отношение разбивает /$_2$ на классы эквивалентности, причем в каждом классе (кроме класса, состоящего из нулевого вектора) существует ровно один элемент, принадлежащий множеству £ стохастических векторов. Кроме того, каждый такой ненулевой класс эквивалентности содержит ровно один элемент из /$_2$, норма которого равна 1.

Введем расстояние между стохастическими векторами следующим образом: расстоянием между рі и р$_2$ назовем расстояние между представителями классов эквивалентности, содержащими рі и р$_2$, норма которых равна 1. Более строго:

Определение 3.11. Расстоянием р$_3$ между стохастическими функциями рі и р$_2$ назовем величину рз(рь р$_2$) $^{=}$ $\underset{\text{Л}^2\ \text{НА II}\ \text{Ил II}}{\text{-HI A}^{-}\text{Аг II}}$

Утверждение 3.5. рз - метрика на множествах /$_s$ и /$_s$'.

Доказательство.

!) рз(р, р)=0 - очевидно

2) рз(рь p_2)= Рз(Р2, Pi) - очевидно

3) Неравенство треугольника для метрики p_3 и трех произвольных словарных функций $p_ь$ p_2, рз следует из справедливости неравенства треугольника для А»Ап и Дг для стандартной
ЬI \\Рг\\ ЬI\

метрики в /$_2$.

Утверждение доказано.

Как известно, *косинусом угла* между векторами pi и p_2 евклидова пространства называется величина cosZ(p_b p_2) = ~~„ , 1 , , ,~~ ([29]).
INN1

Утверждение 3.6. рз(рь p_2) = *jl-cosZ($p_v p_2$).*

Доказательство.

$$2рз(Рь\ p_2)^2 = \underset{IIAI}{IД-Д\ \|^2} = \underset{IIA}{(А-Д,\ А''\ Д\ )} = \underset{1A|\ HAII}{1+1-2тИ\#4} = \underset{A\ A}{2-}$$

2cosZ(p_b p_2), откуда p_3($p_ь$ рг) = *jl-cos Z($P_v p_2$)*

Утверждение доказано.

Следствие 3.4.

1) Для любых двух стохастических словарных функций $p_ь$ p_2 верно, что 0<рз(рь Рг)^1

2) рз(рь Рг)^1 точно тогда, когда cosZ(p_b p_2)=0

3) рз(рь p_2)=0 точно тогда, когда pi=p_2.

Доказательство следует непосредственно из формулы p_3(рь Рг) = *j\ - cosZ($p_v p_2$).*

Определение 3.12. *Декартовым произведением* $A\backslash xA_2$ автономных вероятностных автоматов $A\backslash\sim\{C\&\backslash,П\backslash,P\backslash,У\backslash o,Yi¥\}$

и

^2=(0,02,712^2^2 0/^} назовем вероятностный автомат 74=(C,QixQ2,7C,P,v_0,(iF}, где *к* имеет размерность m=mim$_2$ (mj=|Qj|, i=l,2) и %ЛШ,„Л)=^,^{ш}*и>^{P}(У)к=Pl,ik-P2jb $V_0^{(ij)}$=V! o'-Vi_0^j, $V_F^{(ij)}$=$_{Vl}p^1$-**V,**$_F^j$.

Теорема 3.3. Пусть pi и p_2 - стохастические словарные функции, порожденные монотонными автоматами 4i={C,Qi,7tbPi,Vi_0,$v_{]F}$} и

^2=(C,Q_2,7I2,P2,v_{2O},V2F}, *A=AixA$_2$={C,Q$_l$xQ$_2$,n,?,Vo,v¥}* -декартово произведение автоматов *A* и *A_2*, k=|C|, m=|QixQ$_2$|. Тогда (рьР2>= *((E-mгy^1\$_m$*,

к

где *n\ я$_y$= я^P$_a$* -приведенная матрица переходов автомата *A*.

Доказательство.

(РьРг) = $\sum_{yeC}$EA(r)A(r)=Z $_{yeC\ \ V\ Ч\backslash}$ $_{Чг}$ 2^0^)^0^)

f

$\sum_{yeC}$ Е *ЦР^Г'ЯМIУ'ЯI)* = $\sum_{r\&\ \backslash Чi,Чi}$ *ЦР(Г^ХЧ$_2$)* = £Р00-$_{yeC\ J}$

Здесь q_x - путь из начального состояния в финальное в автомате *A*, q_2 - путь из начального состояния в финальное в автомате A_2, q_xxq_2 - путь из начального состояния в финальное в автомате *A=A\xA$_2$.*, *p(y,q)*- совместная вероятность выдачи слова у и прохождения по пусти *q*,

Р(Г,Ю=Р(У\ЮР(Ч)-

По теореме 3.1 £р00 = *((E-mгy^1\$_m$=>* (рьр$_2$)=((£-£)"%, что и тре-

$_{yeC}$

бовалось доказать.

Теорема доказана.

Содержательно теорема 3.3 означает, что скалярное произведение двух стохастических словарных функций эффективно вычисляется как вероятность дойти из начального состояния до финального в декартовом произведении любых двух автоматов, реализующих эти словарные функции. В свою очередь, вероятность дойти из начального состояния до финального вычисляется путем нахождения правого верхнего углового элемента в предельной приведенной матрице переходов *(Е -я)~^{x}*.

Следствие 3.5. Расстояние между стохастическими словарными функциями pi и p_2, реализующими автоматы *А* и *Аг,* эффективно вычисляется по формуле

Рз($_{P1}$, p_2)=Vl-cos($_{A}$,^)= 1- ~*л$^{Ш)}$ui^{m,m}:* „.а > ^{г}Д^{е} *Кг*~

приведенная матрица переходов декартова произведения A=74IXA$_2^{=:}${C,QIXQ2,71,P,^IO,)LIF} автоматов *А* и *A_2, $л_{ы}$*- приведенная матрица переходов декартова произведения ДіхЛі, *$л_{2x2}$*- приведенная матрица переходов декартова произведения *АІКАІ.*

Доказательство получается непосредственной подстановкой формулы (рьр$_2$)$^{=}$ *((Е-лу^{1}\$_{m}$в* формулу рз(р$_ь$ рг)= *Jl-cosZ(pi,p_2)* для вычисления расстояния p_3.

Главным итогом этого параграфа является существования способа эффективного введения метрики на множестве автономных монотонных вероятностных автоматов. Для вычисления метрики рз оказывается достаточным применить конечное число операций типа умножения матриц, нахождения обратной матрицы и определителя, извлечения квадратного корня, умножения и деления действительных чисел.

По аналогии с тем, как мы вводили функцию качества с помощью метрики p_2, можно дать следующее определение:

Пусть 8={pbP2v Рп} - словарь команд, $A''=\{A_ъA_2,...,A_n\}$ - соответствующие этим командам вероятностные автоматы. Тогда *качеством* этого словаря назовем величину ч'(E,p)CB)=$0_д$(рз).

§3.5 Связь между метрикой и вероятностью.

Пусть слово $y_0 \in C^*$.

Определение 3.13. *Характеристической словарной функцией слова У* назовем следующую словарную функцию:

[O, *иначе* Пусть $y=CiC_2...c_n$. Тогда монотонный вероятностный автомат, соответствующий функции pp, можно задать автоматом, приведенном на рис. 3.2:

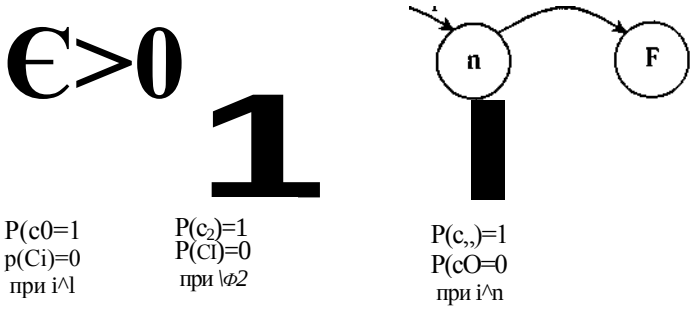


Рисунок 3.2.

Используя характеристическую словарную функцию для кодового слова, можно ввести расстояние между кодовыми словами и словарными функциями:

Определение 3.14. Пусть р - метрика на множестве $/_s$ (или /Д у $\in C^*$, р $\in /_5$(/Д *Расстоянием меэюду словом* у *и словарной функцией р* назовем величину $p(y, p) = p(p_Y, p)$.

Посмотрим, чему равно расстояние между словом и словарной функцией для введенных ранее метрик pi, p_2 и p_3.

Лемма 3.5. Пусть уєС*, ре/$_S$. Тогда

(1) p,(y,p)=2(1-p(y)),

(2) p_2(y,p)=1-p(y)

(3) РЗ(Т.Р)=ff .-„pi, -норма словарной функции р в /,

Доказательство.

(1) p,(y, p) = $_{Pl}$(p_Y, p) = ^\$P_I$(Г')-Р(Г')\ = ?,\$_{yФy}$РАГ')-Р(Г)\+I-Р(У)

2>(7') +1-P(Y) = 2>0O -P(Y)+1-P(Y) = 1-p(y)+1-p(y)=2(1-
$_{г*г}$ p(y)).
$_{г'}$

(2) p_2(y, p) = p_2(p_y, p) = 1- $_{y'eC}$^mn($_{Pг}$(/),p(/))=\- p(y)

(3) рз(y, p) = Рз(p_y, Р) = J1-FT4'

IIPyII =]ЦРг^2(г')=1, (P_Y> Р) = $_{r'}$T,Ру(Г)р(Г) = Р(У)> откуда p_3(y,p)
V f

U P(r)
V IИ

Лемма доказана.

Из леммы 3.5 следует, что введенная в §3.1 при задании алгоритма распознавания (Е,р) функция р расстояния между словарными функциями и кодовыми словами совпадает в смысле определения 3.14 с метрикой

Следствие 3.6. Если нам дано расстояние от слова уєС* до стохастической функции рє/$_5$, то можно рассчитать вероятность р(у) по одной из следующих формул:

(l)p(Y)=Vi(l-pi(Y,p));

(2)p(y)=1-p_2(y,p);

(3)$P(У)=|P|\prod(1-P3^2(У,Р))$.

Метрика на вероятностных автоматах может также применяться для решения задачи **сокращения фонетического алфавита.** Коротко постановка и практический смысл этой задачи состоит в следующем. При распознавании речи вероятностные автоматы, выступающие в качестве эталонов слов словаря, обычно синтезируются путем последовательного соединения вероятностных автоматов, являющихся эталонами фонем, составляющих фонетическую транскрипцию слова.

Поскольку качество обучения (синтеза) автоматов зависит прямо пропорционально как от объема обучающей выборки, так и от количества автоматов, которые необходимо обучить, то одной из задач создания распознавателя речи является задача сокращения фонетического алфавита, имеющая важное практическое значение.

Задачу эту можно решать путем выбора в качестве начального приближения фонетического алфавита максимального алфавита, включающего в себя все возможные символы транскрипции с максимальной детализацией, а затем путем разбиения этого алфавита на классы эквивалентности и использования этих классов в качестве нового фонетического алфавита на следующем, заключительном шаге.

В один класс эквивалентности целесообразно помещать фонемы, которые близки друг к другу в смысле некоторой меры близости. Отсюда напрямую вытекает необходимость построения метрики на множестве фонем данного языка (фонетическом алфавите).

Поскольку фонемы моделируются вероятностными автоматами, возникает необходимость введения *метрики* на множестве вероятностных автоматов без входа.

Метрика рз была приведена на практике для решения задачи сокращения фонетического алфавита для русского языка. В результате

проведенного исследования было показано, что можно сократить алфавит фонем для русского языка со 143 до 132 элементов.

ПРИЛОЖЕНИЕ 1. ПРИМЕР АЛФАВИТА БУКВ И ЗВУКОВ, ПРАВИЛ ТРАНСКРИБИРОВАНИЯ ДЛЯ РУССКОГО ЯЗЫКА.

Алфавит букв русского языка: { а, б$_э$ в, г, д, е, ё, ж, з, и, й$_э$ к, л, м, н, о, п, р, с$_э$ т$_э$ у, ф, Х, ц, ч, ш, щ, ъ, ы, ь, э, ю, я,' }.

Алфавит звуков русского языка (согласно [41], [42])

Символом " * " обозначается мягкость звука слева, справа или с двух сторон (в зависимости от того, с какой стороны стоит символ " * ").

ГЛАСНЫЕ

Символом "+" обозначается гласный под ударением.

ГЛАСНЫЕ ПОД УДАРЕНИЕМ

Обозначение фонемы	Буква	Примеры	Транслитерация
а+	а	арка, буржуа	al
а*+	а	рвань	ail
*а+	а	чан	ial
	я	яма	
а+	а	часть	iail
	я	бязь	
0+	0	дом, наотмашь	ol
о*+	0	озеро, моль, наощупь	oil
*о+	0	трущоба,	iol
	е	черт, пес, ружье	
о+	0	на Печоре, йодистый	ioil
	е	песик	
У+	У	угол, жук, какаду, наука, недоумок	
у*+	У	пуля, усики, научит, баульчик	uil
*у+	У	чудо, свищу	iul
	ю	дюны, юг	
у +	У	чучело	iuil
	ю	бюстик	
и+	и	шина, цирк	yi
	ы	огурцы, пыл, сады	
и*+	и	истина, ширь, циник	yil
*и+	и	бинт	il

и+	и	линь	II1
	э	этот, пэр	el
е+	е	кафе, шест	
	э	эти, о пэре	eil
е*+	е	дельта, шесть	
	е	тесто, ель	iel
*е+	е	сельский, мель, свистеть	ieil

БЕЗУДАРНЫЕ ГЛАСНЫЕ

СИЛЬНЫЕ БЕЗУДАРНЫЕ — в 1-ом предударном слоге или после паузы в начале слова или после согласной перед паузой или последняя гласная в группах гласных.

Обозначение фонемы Буква		Примеры	Транслитерация
	о	оса. пока	а
	а	аванс, сажа	
	0	жокей	ai
	я	катя пять	ia
		атак	
4^ Q 4^		пять осин	iai
У2у*2	У	указ	u
***у2**	У	пустяк	ui
***у*2**	ю	юла	iu
	У	чудить	IIII
и	ю	тюлень, юлить	
	и	широк	У
и* *и	ы	мытарить	УI
* JJ*	и	зима	i
	я	пятак	
	и	мирить	ii
	э	экран	е
їр *пф*	э	электростанция	ei
	е	поле	ie
	е	еще	iei

СЛАБЫЕ БЕЗУДАРНЫЕ - только после согласных и
не в 1-ом предударном слоге и
не в конце слова перед паузой.

Обозначение фонемы	Буква	Примеры	Транслитерация
	а	балагур, шаловлив	ах
	ы	пасынок	
	о	богатырь, шоколад	
	е	сторожем	
ъ*	о	вымочить	ахі
	а	выкатить	
уі	У		w
у*l	У	путевой, сулема	wі
*уl	ю	полюс, вынюхать, кюрасо	іw
			іwі
*ь	а	часовой	і
	е	камнем	
ь	я	впятером	іі

В ИНОСТРАННЫХ СЛОВАХ гласные следует обозначать в соответствии с тем, как они слышатся.

СОГЛАСНЫЕ

Символом ":" обозначается длительность согласного (например, в случае сдвоенного согласного).

СОНОРНЫЕ ЗУБНЫЕ

Обозначение фонемы	Буква	Примеры	Транслитерация
Р	р	рама, пар, тигр	г
Р: р*	рр	суррогат	гг
р*:	р	рюмка, январь	гі
н	рр	территория	ггі
н:	н	нам, сын, канва	п
н*	нн	ванна, колоннада	пп
н*:	н	няня, дрянь, синька, бантик	пі
л	нн	теннис	шіі
л:	л	лавка, стол, цикл	1
л*	лл	балласт, мулла	И
л*:	л	лягу, моль, льстить	1і
	лл	капилляр	Ш

ГУБНЫЕ

Обозначение фонемы	Буква	Примеры	Транслитерация
	м	мыловар, дом, реноме	m
м			
м:	мм	грамматика	mm
м*	м	мял, семь, познакомьте	mi
м*:	мм	комментатор	mmi
б	б	бак, объект	b
б:	бб	аббат	bb
б*	б	бязь, бязь	bi
б*:	бб	баббит	bbi
п	п	пора, суп	P
	б	зуб, герб, трубка	
п:	пп	аппарат, группа	PP
п*	п	пятый, сыпь	Pi
	б	дробь, приспособьте	
п*:	пп	аппетит	ppi
	в	вуз, враг	v
	ф	Афганистан	
	в	ввоз	w
	в	вялый, вьюга, въехать, вбить	vi
в .	в	ввел	wi
Ф	ф	факт, шеф	f
	в	остров, вкус	
ф:	фф	диффузный	ff
ф*	ф	тюфяк, верфь, потрафьте	fi
	в	кровь, приготовьте, впишет	
ф*:			ffi

ЗУБНЫЕ ШУМНЫЕ

Обозначение фонемы Буква		Примеры	Транслитерация
	з	зал, призрак, язва	z
	с	сбор,сгубить	
з:	зз	беззаботный	zz
	сз	сзади	
	з	зябну, резьба, изъездить, уздечка	zi
	с	сбить	
	зз	безземельный	zzi
	с	сало, нос, стол	s
	з	глаз, смазка	
	ее	компрессор, касса, ссадина	
	с	ss	
		снесся	

	з	грызся	
	с	сяду, весь, моська	si
	з	мазь, лезть	
	ее	рассердиться	ssi
д	д	дать, подъем	d
	т	отбить	
д:	дд	аддуктор	dd
д*	д	дятел, свадьба, две	di
д[и]	дд	поддержка	ddi
	т	там, пот, отъезд	t
	д	яд, подкинуть	
	тт	атташе, оттащить, гетто	tt
	т	тяга, петь, отняли	ti
	д	сельдь, обладьте, складчина	
гр*К•	тт	аттестат, аттический, оттесывать	tti

ЗАДНЕНЕБНЫЕ

Обозначение фонемы Буква		Примеры	Транслитерация
	г	газ, когда	
	к	анекдот	
г:			
	г	гиоель	**gl** ggi
к	к	кадка, так, краб	k
	г	ногти, берег	
к:	кк	мокко	kk
к*	к	кинуть	ki
к*			kki
х	х	хата, задохся, слух	h
	г	мягко	hh
х:			
	х	хитрый	hi
	г	мягкие	hhi

ШИПЯЩИЕ И [Ц]

Обозначение фонемы Буква		Примеры	Транслитерация
ж	ж	жаль, вежливый, ружье	zh
	зж	разжать, езжу	
	еж	сжать	
	ж	ложбина	

ж: ж*	жж	жужжать	zzh
ж*: ш			zhi
			zzhi
ш:	ш	шаг, шью, шла, грош, мышь	sh
	ж	нож, рожь, ложка	
щ*	с	сшить	ssh
	3	лезший	
щ*:	щ	пощада, вещью, плющ	sch
	сч	счастье, разносчик	
	сщ	расщелина	ssch
ч	34	навязчивый	
ч:	жч	мужчина	ch
			cch
ч*'.			
ц	ч	часть, чью, чрево, врач	chi
ц:			cchi
	ц	цоколь, матрац, клецки	ts
			tts

йот

й*	я	яблоко	j

ОЗВОНЧЕННЫЕ (на стыке слов)

Обозначение фонемы	Буква	Примеры	Транслитерация
(хг)	хг	господи, белых гусей	x
(дз)	цб	отец бы	dz
(дж*)	чб	дочь бы	dxh

СЛУЖЕБНЫЕ ФОНЕМЫ

СМЫЧКИ обозначаются соответствующим согласным, заключенным в квадратные скобки:

[т], [т*], [д], [д*], [п], [п*], [б], [б*], [р], [р*], tcl ticl del did pel picl bcl bid rcl ricl
[к], [к*], [г], [г*], [ц], [дз], [ч], [ч*], [дж*], kcl kicl gel gicl tscl dzel chel dzhel
[т:], [т*:], [д:], [д*:], [п:], [п*:], [б:], [б*:], [р:],[р*:], ttcl tticl did ddicl picl ppicl bbel bbicl pel ppicl
[к:], [к*:], [г:], [г*:], [ц:], [ч:], [ч*:], [дж*] kkcl kkicl ggcl ggicl ttscl echel dzhel

ПАУЗА в начале фразы и в конце фразы - %%
ПАУЗА между словами - %
ПАУЗА вставочная (эпентетическая) - []

Транскрибирование производится по следующим правилам:
(согласно [31], [41], [42])

Переход "буква-фонема".

Осуществляет такие операции над орфографической записью, как устранение орфографических фикций, обработка особых случаев произнесения стечений согласных, устранение твердых и мягких знаков, обработка йотированных и мягких букв с соответствующей интерпретацией твердости-мягкости соседних согласных и введением йота.

Пример правил этого блока:

1) здн—»з н, лнц->н ц, стн-^с н, рдц—»р ц, и т.п. (пропуск непроизносимых согласных)

2) тся-»ц ц а, ться-»ц ц а, что->ш т о, его->е в о й т.п. (замена сочетаний согласных в окончаниях)

3) сш-»ш ш, сщ-»щ, стч-»щ, тц-»ц ц, и т.п. (ассимиляция по способу образования)

4) сть—»сь ть и т.п. (ассимиляция гомоорганных согласных по твердости-мягкости)

5) ль->ль (смягчение согласных перед мягким знаком и гласная)

6) бьи'-»бь й и' в слове воробьи и т.п. (правила вставления й в ударном и безударном слоге)

7) не-»нь е и т.п. (смягчение согласных перед йотированными гласными)

8) сс-»с, нн-»н, мм->м и т.п. (замена двух одинаковых согласных одним в некоторых позициях)

Переход "фонема-звукотип".

Включает правила, обрабатывающие случаи позиционного озвончения, оглушения согласных и редукции гласных (с учетом двух степеней редукции).

Примеры правил озвончения-оглушения:

1) б-»п в слове столб и т.п. (оглушение согласных в конце слов)

2) бс->пс в слове обстрел и т.п. (ассимиляция согласных по звонкости-глухости)

Правила редукции гласных:

Ударный гласный — наиболее полногласный, т. е. наиболее мощный по длительности и интенсивности (обозначен цифрой 3).

1-ая степень редукции — на одну ступень ниже по интенсивности и длительности, т. е. (3 - 1) -» 2.

2-ая степень редукции ещё ниже, т. е. (3 - 2) -> 1.

Формула Потебни:

1	2	3	1	1
2-ой предуд. слог	1-ый предуд. слог	ударный слог	1-ыйзаударн. слог	2-ой заударн. слог

или:

#2	2	3	1	2#
2-ой предуд. неприкр. слог	1-ый предуд. слог	ударный слог	1-ыйзаударн. слог	2-ой заударн. открыт, слог

Это означает следующее: 1-ая степень редукции наблюдается в первом предударном слоге, в абсолютном начале и в абсолютном конце слова, т. е. после или перед пробелом (знаком "#"). 2-ая степень редукции наблюдается во всех закрытых сзади согласными заударных слогах и во втором, третьем и др. прикрытых спереди согласными предударных слогах.

Например: "повтори+ть" -» 12 3

"перестро+йка" -» 12 3 2

"за+поведь" —> 3 11

"анахрони+зм" -> 2 12 3

Знак "+" после гласного означает ударность гласного. Правила замен фонем на редуцированные звукотипы для каждой степени редукции:

1-ая степень редукции

фонема	позиционно обусловленные звукотипы					
	#+Г	Г+Г	С+Г	с+Г	Г+#	с__г
0	а	а*	а*	и	а*	а
а	а	а	а*	и	а	а
э	э	э	ы*	и	э	э
ы	ы	—	ы	—	ы	ы
и	и	и	—	и	и	и
У	У2	У2	У2	У2	У2	У2

* здесь означает наличие исключений, которые задаются списками

2-ая степень редукции

фонема	позиционно обусловленные звукотипы	
	с+г	с+г
0	ъ	ь
а	ъ	ь
э	ъ	ь
ы	ъ	—
и	—	ь
У	**уi**	уi

ПРИЛОЖЕНИЕ 2. ПРИМЕР ФОНЕМНЫХ КЛАССОВ ДЛЯ РУССКОГО ЯЗЫКА

В работе [a5] использовались следующие классы фонем для русского языка:

класс _ ("паузы"): пауза,[п],[п*],[к],[л*],[т],[т*],[ч],[ц]

класс X ("шипящие"):

с,с* ,ц,ч,ш,щ,в,в* ,д,д* ,ж,з,з*,й,л,л * ,п,п* ,т,т* ,ф,ф* ,х,х*

класс A ("низкочастотные гласные"): а,а+,э,э+ **класс O**

("среднечастотные гласные"): о,о+,yl,y2,y+ **класс I**

("высокочастотные гласные"):

и,и+,*и,*и+,и*,и*+,*и*,*и*+,а*+,о*+ и т.п., л,л*,м,м*,н,н*, звонкие смычки

ПРИЛОЖЕНИЕ 3. МЕТРИКА НА МНОЖЕСТВЕ ЗВУКОВ РУССКОГО ЯЗЫКА.

Для примера взяты 10 звуков русского языка из набора, приведенного в приложении 1 (реально матрица имеет размерность 150x150)

	[п]	**М**	*а*+	У+	и	м	л*	ш	с	ж
[п]	0	14.657	143.974	134.180	136.092	116.847	138.666	96.287	48.550	114.615
М	14.657	0	144.745	139.320	136.029	117.216	139.747	99.394	48.370	116.181
а+	143.974	144.745	0	116.897	125.756	105.587	119.652	143.963	144.392	135.984
У+	134.180	139.320	116.897	0	126.364	113.008	118.581	150.204	139.529	120.719
и	136.092	136.029	125.756	126.364	0	127.115	105.292	122.289	147.063	115.253
м	116.847	117.216	105.587	113.008	127.115	0	108.844	134.464	117.647	103.921
л*	138.666	139.747	119.652	118.581	105.292	108.844	0	137.965	148.690	90.546
ш	96.287	99.394	143.963	150.204	122.289	134.464	137.965	0	100.939	110.394
с	48.550	48.370	144.392	139.529	147.063	117.647	148.690	100.939	0	109.848
ж	114.615	116.181	135.984	120.719	115.253	103.921	90.546	110.394	109.848	0

ЛИТЕРАТУРА

[I] СЕ. Левинсон. Структурные методы автоматического распознавания речи. ТИИЭР, т.73, №11, ноябрь 1985 г., с. 100-128.

[2] К.Н. Davis, R.Biddulph, and S.Balashek, "Automatic recognition of spoken digits", J.Acoust. Soc. Amer., Vol. 24, pp.637-642, 1952.

[3] P.B. Denes and M. V. Mathews, "Spoken digit recognition using time frequency pattern matching", J.acoust. Soc. Amer., vol. 32, pp. 1450-1455, 1960.

[4] H. Dudley and S.Balashek, "Automatic recognition of phonetic patterns in speech", J. Acoust. Soc. Amer., vol. 30, pp. 721-7439, 1958.

[5] D. H. Klatt, "Review of the ARPA Speech Understanding Project", J. Acoust. Soc. Amer., vol.62, no.6, pp. 1345-1366, Dec. 1977.

[6] Г. Фант. Акустическая теория речеобразования *I* Пер. под ред. В.С. Григорьева. М.: Наука, 1964.

[7] V. W. Zue and R. A. Cole, "Experiments on spectrogram reading" in Proc. ICASSP-79, pp. 116-119, 1979.

[8] Т.К. Винцюк. Распознавание слов устной речи методами динамического программирования. Кибернетика. - 1968. - №1, с. 81-88.

[9] Р. Беллман, Динамическое программирование. М.: ИЛ, 1960.

[10] F. Itakura, "Minimum prediction residual principle applied to speech recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp. 67-72, 1975.

[II] J.K.Baker, "Stochastic modeling for automatic speech understanding", in Speech Recognition, D.R. Reddy, Ed. New York: Academic Press, 1975, pp. 521-542.

[12] Ф. Джелинек [Елинек]. Распознавание непрерывной речи с помощью статистических методов. ТИИЭР, 1976, т. 64, №4, с. 131-160.

[13] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains", Ann. Math. Stat., vol. 37, pp. 1554-1563, 1966.
[14] А.А. Марков. Пример статистического исследования над текстом "Евгения Онегина", иллюстрирующий связь испытаний в цепь. Известия Академии наук, СПб., VI, т.7, 1913, №3, с. 153-162.
[15] S. E. Levinson, L.R. Rabiner, and M.M. Sondhi, "An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition". Bell Syst. Tech. J., vol 62, no.4, pp. 1035-1074, Apr. 1983.
[16] B.H. Juang, "On the hidden Markov model and dynamic time warping for speech recognition - A unified view", AT&T Tech. J., vol. 63, no.7, pp.1213-1243, Sept. 1984.
[17] L.R.Rabiner and B.H. Juang, "An introduction to the hidden Markov models", IEEE ASSP Mag., vol.3, no.l, pp.4-16, 1986.
[18] JI.P. Рабинер. "Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: обзор". ТИИЭР, т.77, №2, февраль 1989 г.
[19] L.E. Baum and J.A..Egon, "An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology", Bull. Amer. Meteorol. Soc, vol.73, pp.360-363, 1967.
[20] A. P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J.Roy. Stat. Soc, vol. 39, no.l, pp.l-38, 1077.
[21] A.J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimal decodin algorithm", IEEE Trans. Informat. Theory, vol. IT-13, pp. 260-269, Apr. 1967.
[22] Дж. Д. Форни-мл. "Алгоритм Витерби". ТИИЭР, 1973, т.61, №3, с.12-25.

[23] J.W. Carlyle. Reduced forms for stochastic sequencial machines. - J. Math. Analysis and Applic, 1963, 7, p. 167-175.
[24] Р.Г. Бухараев. Некоторые эквивалентности в теории вероятностных автоматов. - Уч. записки Казан, университета, 1964. 124, №2, с. 45-65.
[25] P.H. Starke. Theorie Stochastischen Automaten. I, II. - Elektron Informationsverarb. und Kybern., 1965, 1, №2. [26] Р.Г. Бухараев. Основы теории вероятностных автоматов. М.: "Наука", 1985.
[27] R. Bakis, "Continuous speech word recognition via senti-second acoustic states" in Proc. ASA Meeting (Washington, DC), Apr. 1976. [28] D. Kanevsky, M. Monkowsky, J. Sedivy. Large Vocabulary Speaker-Independent Continuous Speech Recognition in Russian Language. Proc. SPECOM'96, St.-Petersburg, October 28-31, 1996. [29] А. Н. Колмогоров, С. В. Фомин. Элементы теории функций и функционального анализа. М.: Наука, 1989.
[30] Дж.Бендат, А.Пирсол. Измерение и анализ случайных процессов. М.: "Мир", 1974 г., стр. 372, 342, 368.
[31] Ипатов Я.В., Коваль С.Л., Лапшина А.В., Сапожкова И.Ф. Компилятивный синтезатор речи по правилам. Параметрическое представление. Лингвистическое обеспечение.: АРСО-15, стр. 156-157.
[32] Л.Р.Рабинер, Р.В.Шафер, Цифровая обработка речевых сигналов. Пер. с английского. М., "Радио и связь", 1981 г. [33] Т.К. Винтцюк. Анализ, распознавание и интерпретация речевых сигналов. Киев, Наукова думка, 1987.
[34] В.Б. Кудрявцев, СВ. Алешин, А.С. Подколзин. Введение в теорию автоматов. М.: Наука, 1985. [35] Академическая программа Intel. http://www.intel.ru/education/Grants/grants.htm

[36] R. Singh, Bh. Raj, R. M. Stern. Structured definition of sound units by merging and splitting for improved speech recognition. 1999.
[37] T. Yoshimura, T. Masuko, K. Tokuda, T. Kobayashi, T. Kitamura. Speaker interpolation in HMM-based speech synthesis system. 1999.
[38] B.T. Tan, Y. Gu, T.Thomas. Word confusability measures for vocabulary selection in speech recognition. 1999.
[39] M. C Nechyba, Y. Xu. Stochastic similarity for validating human control strategy models. 1999.
[40] S. Siruguri. Modelling a navigation task. 1999.
[41] Н. В. Зиновьева, Л. М. Захаров, О. Ф. Кривнова, А. Ю. Фролов, И. Г. Фролова. Автоматический транскриптор. АРСО-92.
[42] Л.М. Захаров. Транскрипция текстов при синтезе и анализе русской речи. АРСО-96.

Публикации автора по теме диссертации

[al]. Мазуренко И.Л. Одна модель распознавания речи. В сб.: Компьютерные аспекты в научных исследованиях и учебном процессе. -Издательство Московского университета, Москва, 1996 г., стр.107— 112.

[а2]. Мазуренко И.Л. О сокращении перебора в словаре речевых команд в составе системы распознавания речи. В сб.: Интеллектуальные системы, т.2, вып. 1-4, Москва, 1997. Стр. 135-148.

[а3]. Мазуренко И.Л. Компьютерные системы распознавания речи. В сб.: Интеллектуальные системы, т.3. вып. 1-2 - Москва, 1998 г. Стр. 117-134.

[а4]. Мазуренко И.Л. Многоканальная система распознавания речи. В сб.: VI всероссийская конференция "Нейрокомпьютеры и их применение" Сборник докладов. Москва 16-18 февраля 2000 г. Стр. 222—225.

[а5]. Бабин Д.Н., Дудецкий В.Н., Мазуренко И.Л., Уранцев А.В. и др. Устройство для синтеза и анализа речевых сигналов. Патент N 94045004/09 (045160) решение от 20.12.96.

[аб]. Бабин Д.Н., Мазуренко И.Л., Уранцев А.В., Холоденко А.Б.Способ идентификации факта речевой активности оператора. Патент N 99103468/ 28(003418) решение от 29.12.99. МІЖ 6 В 60 К 28/06, G 10 L 5/06

В патентах [5], [6] основная заслуга автора заключается в разработке и программной реализации алгоритмов распознавания речевых сигналов.