

На правах рукописи

Ду Цзяньмин

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ
МЕТОДОВ ЗВУКОВОГО ПОИСКА В БАЗАХ
ДАнных НА ОСНОВЕ ФОНЕТИЧЕСКОГО
КОДИРОВАНИЯ И ИХ ИСПОЛЬЗОВАНИЕ
ДЛЯ УСКОРЕНИЯ РАСПОЗНАВАНИЯ РЕЧИ**

Специальность 05.13.11

Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных
сетей (технические науки)

Автореферат диссертации на соискание ученой степени
кандидата технических наук

Москва — 2019

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. В настоящее время технология электронного распознавания речи является одной из наиболее важных частей взаимодействия системы «человек – компьютер», статус устойчивого направления ведения научных исследований закрепился за ней еще с момента зарождения. В то же время с развитием информационных технологий распознавание речи находит все более широкие области применения. Можно без преувеличения сказать, что на сегодняшний день распознавание речи надолго закрепилось в повседневной жизни общества.

Процесс распознавания речи представляет собой поиск точных совпадений имеющихся в базе данных записей с предоставленным пользователем образцом, обработанным посредством серии вычислительных операций преобразования речевых сигналов.

По мере развития технологий в системах распознавания начали применяться методы цифровой обработки сигналов, статистического и вероятностного моделирования, такие как скрытые марковские модели (СММ) и искусственные нейронные сети (ИНС). В последние годы представлены значительные достижения в исследованиях алгоритма распознавания с использованием обоих приведенных методов.

В то время как величина скорости распознавания с усовершенствованием алгоритма растет, система распознавания усложняется, особенно в случае больших словарей, когда как нейронная сеть, так и скрытые марковские модели сталкиваются с проблемами относительно трудоемких и медленных вычислений, например, при обработке речи на русском языке, который обладает большими лексиконом и гибкостью.

Поэтому улучшение систем анализа русского языка становится основным направлением изысканий. Но большая часть исследований направлена на увеличение скорости распознавания путем использования распределенных вычислений и увеличения вычислительных мощностей на основе аппаратного ускорения.

Реальный вклад существующих на сегодняшний день систем распознавания речи в улучшение процессов поиска в базах данных созвучных слов все еще незначителен и не может быть признан полностью воплощенным и тщательно разработанным. Фактически немалая

часть временных сложностей распознавания речи приходится на поиск слова в словаре. Ускорение поиска слов по их звучанию видится актуальным и не представлено в достаточной степени среди известных научных публикаций.

Рассматриваемые в работе алгоритмы фонетического кодирования и методы поиска слов предполагают возможность становления другого решения для этой задачи. Алгоритмы фонетического кодирования представляют собой последовательность действий для поиска слов по их звучанию. Они могут служить базой построения улучшенных методов поиску слов в словаре.

Цели и задачи. Цель диссертационной работы заключается в разработке и исследовании методов звукового поиска в базах данных на основе фонетического кодирования для ускорения распознавания речи на большом словаре.

Для достижения данной цели были поставлены и решены следующие задачи:

- 1) разработка фонетических алгоритмов для ускорения поиска слов в базах данных;
- 2) разработка метода фонетического кодирования для последовательности фонем языка;
- 3) реализация распознавания речи при использовании разработанного метода поиска слов в базах данных на основе алгоритма фонетического кодирования и определение ее эффективности.

Объект и предмет исследования. Объектом исследования является поиск слов в базах данных. Предмет исследования – метод звукового поиска слов в базах данных при распознавании речи.

Методы исследования. Методы акустического моделирования, распознавания речи, статистической обработки результатов экспериментов, методы поиска данных.

Научная новизна. Разработаны и исследованы следующие методы:

- 1) предложено использование фонетических алгоритмов для поиска слов в базах данных, отличающиеся тем, что кодированию подвергается не последовательность букв слова, а последовательность его фонем;

2) разработан алгоритм фонетического кодирования для последовательности фонем русского языка, позволявший находить близкие по произношению слова;

3) разработан и опробован эффективный метод поиска слов в базах данных на основе предложенного алгоритма фонетического кодирования.

Практическая значимость работы. Применение разработанного метода поиска слов на основе фонетического кодирования позволяет улучшить количественные и качественные характеристики современных систем распознавания слитной речи на большом словаре.

Положения, выносимые на защиту:

1) использование разработанных фонетических алгоритмов для ускорения поиска слов в базах данных;

2) алгоритм фонетического кодирования для последовательности фонем русского языка;

3) метод поиска слов в базах данных на основе алгоритма фонетического кодирования.

Апробация работы. Основные результаты работы были изложены в докладах и получили положительную оценку на следующих конференциях и научных семинарах:

1) научный семинар кафедры «Информационные системы и телекоммуникации» МГТУ им. Н. Э. Баумана (Москва, 2017, 2018);

2) научный семинар Института проблем управления им. В. А. Трапезникова РАН (Москва, 2018);

3) международная научная конференция «Распределенные компьютерные и телекоммуникационные сети (DCCN)» (Москва, 2018);

4) молодежная научно-техническая конференция «Студенческая весна» МГТУ им. Н. Э. Баумана (Москва, 2014, 2016).

Публикации. Результаты диссертационной работы отражены в 6 научных статьях, в том числе 3 публикациях в изданиях из перечня ВАК РФ.

Структура и объем работы. Диссертационная работа состоит из введения, трех глав, заключения и списка литературы. Объем работы составляет 112 печатных страниц, включающих 12 рисунков и 25 таблиц. Библиография содержит 102 наименования.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснованы важность и актуальность темы диссертации, сформулированы цель работы, а также основные задачи, которые необходимо решить для ее достижения, охарактеризованы научная новизна и практическая ценность работы, кратко изложены основные теоретические и практические результаты работы.

В первой главе рассмотрено понятие алгоритмов фонетического кодирования. Описана теоретическая часть работы с самым известным алгоритмом фонетического кодирования SoundEx. Основным принцип кодирования в алгоритме состоит в том, что близкие по звучанию буквы кодируются одной и той же цифрой. С появлением и развитием компьютерных технологий появилось множество других алгоритмов фонетического кодирования, в том числе и для различных естественных языков. Описаны другие производные алгоритмы: NYSIS, Daich-Mokotoff SoundEx и Metaphone. Алгоритмы фонетического кодирования включают в себя не только алгоритмы для сравнения слов, но и алгоритмы определения расстояния между словами при поиске по звучанию: расстояние Левенштейна, расстояние на основе N-грамм и расстояние Джаро. Рассмотрено, для чего они применяются.

Во второй главе рассмотрена теория алгоритма поиска слов в базах данных для ускорения распознавания речи на основе фонетического кодирования.

В первой части главы представлен обзор системы распознавания речи.

Известны многие программные инструменты для исследования поиска слов при распознавании речи. Для проведения экспериментов использовалась информационная система распознавания речи CMUSphinx, которая имеет открытые исходные тексты программ, доступную акустическую и языковую модели русского языка и большой объем словаря (540 тысяч словоформ).

Во второй части главы рассмотрена типичная архитектура систем распознавания речи, показанная на Рис. 1. Входной звуковой сигнал преобразуется в последовательность акустических векторов. Для процесса декодирования слов требуется использование акустической модели, языковой модели и словаря произношений.

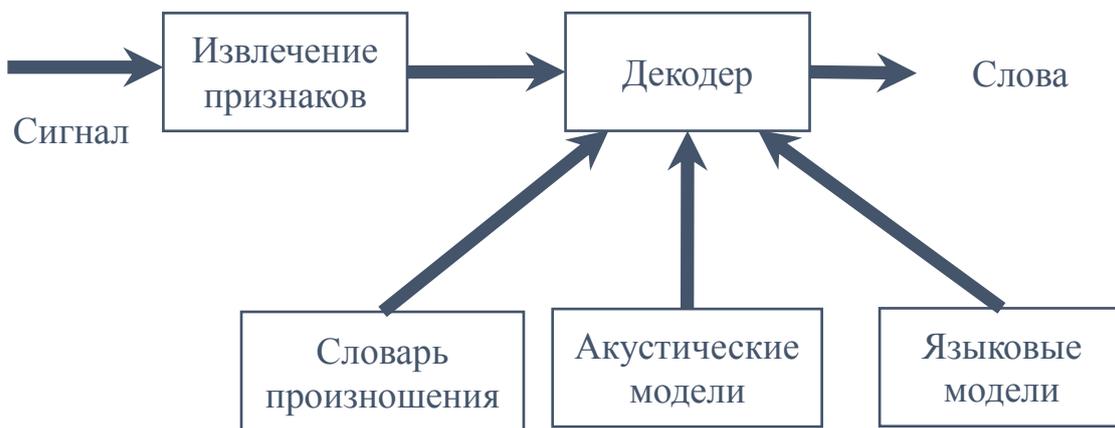


Рис. 1.

Основные компоненты систем распознавания речи

Словарь произношений, или фонетический словарь, реализуемый в виде базы данных, содержит отображение слов и их вариантов произношения на фонемы. Декодер просматривает последовательности фонем из базы данных слов и выбирает наиболее близкое слово или слова.

Одной из тенденций в распознавании речи является рост словаря, приводящий к увеличению времени декодирования. Фактически, большая часть времени распознавания речи тратится на звуковой поиск слов в базах данных. Зависимость времени распознавания от объема словаря показана в Рис. 2.

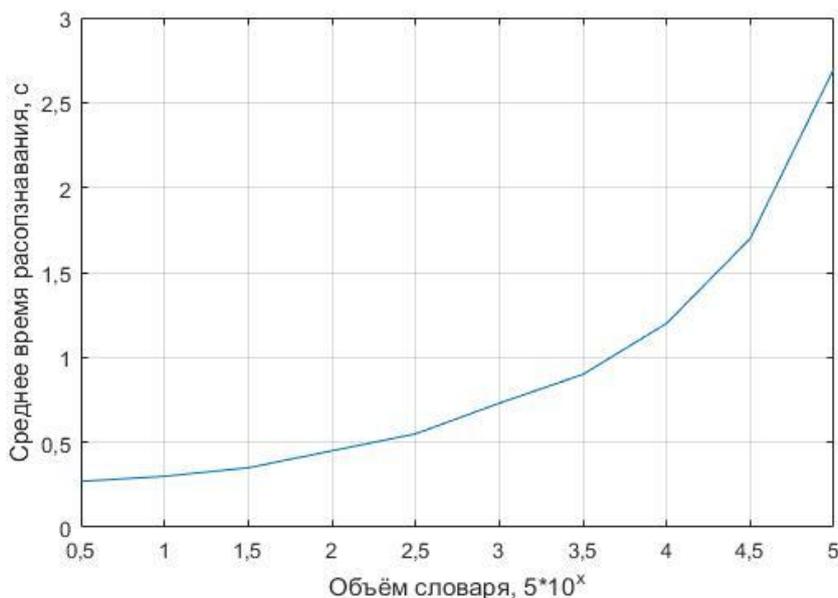


Рис. 2.

Среднее время распознавания отдельного слова с разными объемами словаря

В третьей части главы предложен метод ускорения поиска слов, основанный на идее фонетического кодирования. Суть подхода состоит в уменьшении объема словаря, используемого в процессе распознавания путем построения фонетического кода для распознаваемого слова и поиска в базах данных только тех слов, которые имеют близкие фонетические коды. Основные компоненты системы распознавания речи с помощью этого метода показаны на Рис. 3.

Из рисунка видно, что перед декодером добавлен модуль поиска слов на основе фонетического кодирования. В этом модуле признаки речи сначала анализируются и потом фонетически кодируются. Затем по результатам кодирования с помощью поиска слов на основе фонетического кодирования получаются «слова близкого произношения» для распознавания в декодере.

Словарь системы распознавания речи описывает произношение слова в виде его транскрипции. Большинство систем распознавания речи имеют функции прямого распознавания последовательности фонем из сигнала речи с помощью только акустической модели и простого выделения фонем.

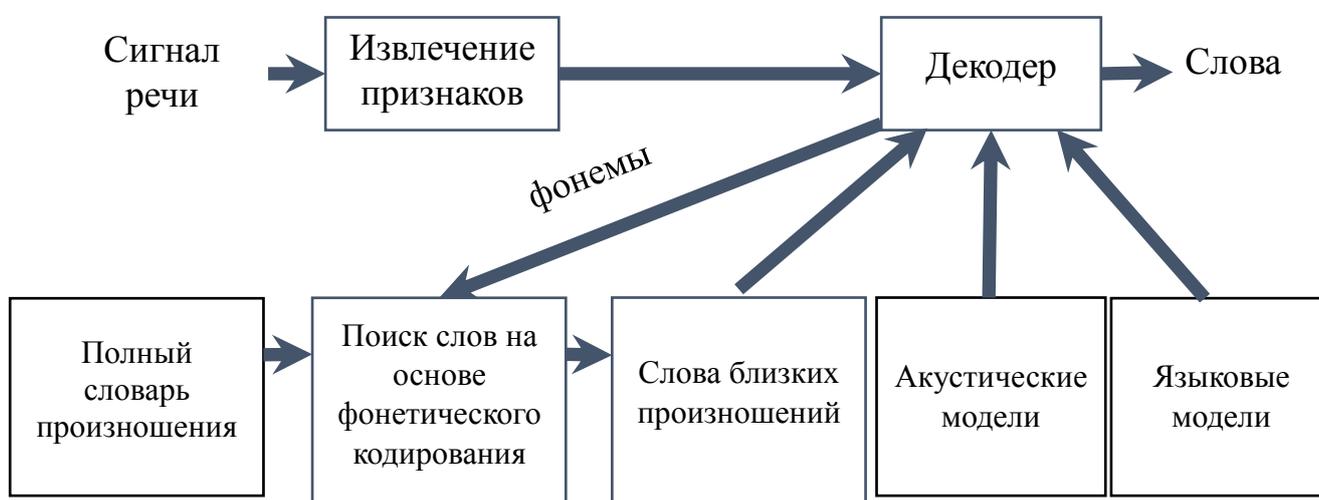


Рис. 3.

Основные компоненты системы распознавания речи с помощью поиска слова на основе фонетического кодирования

Примеры слов из словаря, их транскрипции и последовательностей фонем показаны в Таблице 1. Для кодирования

знаков транскрипций и фонем русского языка использованы обозначения, заимствованные из моделей русского языка в CMUSphinx.

Таблица 1.

Примеры слов, их транскрипции и фонемы

Слово	Транскрипция	Фонемы
дедушка	dj e1 d u0 sh k a0	bj dj e0 d u0 sh kj a0
ресница	rj e0 s nj i1 c a0	e0 z n1 dj z a1 m
поплакать	p o0 p l a1 k a0 tj	k a0 p l a1 g j a0

В процессе анализа результатов сопоставления последовательностей фонем и транскрипций слов русского языка были найдены три фонетических правила, которые могут использоваться для формирования алгоритма фонетического кодирования для последовательности фонем русского языка:

- 1) длины последовательности фонем и длины транскрипции соответствующего слова близки по значению;
- 2) число гласных в последовательности фонем и число гласных в транскрипции соответствующего слова близки;
- 3) влияние первого согласного на результат распознавания меньше по сравнению с последующими согласными.

По этой причине для фонетического кодирования слов были выбраны длина последовательности фонем, количество гласных и первый согласный в слове.

В Таблице 2 показаны диапазоны длин транскрипций слов при различных длинах распознаваемой последовательности фонем, соответствующих этому слову в реальных условиях распознавания речи.

Диапазоны определены так, что каждый диапазон и соответствующая ему характеристика (длина транскрипции) удовлетворяли следующему соотношению,

$$P(L \in R) \geq P_0,$$

где L – число распознаваемых фонем, R – диапазон длин транскрипций слова. P_0 – минимальная вероятность правильного распознавания (в работе принята равной 0,95).

Таблица 2.

Длина транскрипций

Число фонем в распознаваемой последовательности	Длина транскрипции слова в словаре
1	[1, 4]
2	[1, 4]
3	[1, 5]
4	[2, 7]
5	[2, 8]
6	[3, 9]
7	[3, 10]
8	[4, 11]
9	[5, 13]
10	[6, 15]
11	[8, 17]
12	[9, 18]
12+	[9, 18+]

Например, если последовательность фонем имеет длину 8, то соответствующее слово может иметь длину транскрипции от 4 до 11 фонем.

В Таблице 3 показываються диапазоны для числа гласных в транскрипции распознаваемого слова.

Как было указано ранее, первый согласный является определяющим в распознавании слова. Однако из-за ошибок преобразования векторов признаков в фонемы распознавание первого согласного слова сопровождается тремя видами ошибок:

1) первый согласный в транскрипции не является первым согласным в последовательности фонем, перед ним появляются другие независимые согласные;

2) первый согласный отсутствует в последовательности фонем;

3) первый согласный трансформировался в согласный другой фонетической группы, где в одной группе объединены сходные по звучанию фонемы.

Таблица 3.

Число гласных в транскрипции

Число гласных в последовательности фонем	Число гласных в транскрипции
1	[0, 2]
2	[1, 3]
3	[2, 4]
4	[2, 5]
5	[3, 6]
6	[4, 8]
7	[5, 9]
8	[6, 9]
9	[7, 10]
9+	[8, 10+]

Первая ошибка присутствует в слове «дедушка» (Таблица 1), где в последовательности фонем «bj d e0 dj u0 sh kj a0» первой фонемой является согласный «bj». Пример второй ошибки виден в слове «ресница» с транскрипцией «rj e0 s nj il c a0» и соответствующей ей последовательности фонем «e0 z nldj z a1 m». Третья ошибка возникла в слове «поплакать»: согласный «р» был распознан как «к».

Для учета ошибок второго и третьего типа сгруппируем согласные фонемы в группы в соответствии с Таблицей 4.

Для получения Таблицы 4 использованы экспериментальные результаты классификации фонем на основе смеси гауссовых моделей и скрытых марковских моделей. Согласные объединялись в одну группу при выполнении следующего неравенства,

$$P(c \in \mathbf{C}_N | r_c \in \mathbf{R}_N) \geq P_0,$$

где \mathbf{C}_N – множество первых согласных транскрипции в группе N , r_c – результат распознавания согласного c , \mathbf{R}_N – множество первых согласных последовательности фонем в группе N , P_0 – минимальная вероятность правильного распознавания (в работе принята равной 0,95).

Например, из Таблицы 4 видно, что если первый согласный звук выражен фонемой «р», то в транскрипции слова первый согласный может быть выражен фонемами «b» «bj» «p» «pj» «k» «kj» «t» и «tj».

Таким образом, для фонетического кодирования слов русского языка имеет смысл использовать код, состоящий из трех цифр:

- 1) длина транскрипции;
- 2) количество гласных;
- 3) группа первого согласного.

Таблица 4.

Группы первых согласных слова

Группа	Первый согласный последовательности фонем	Первый согласный транскрипции
1	z, zj, s, sj, c, ch, sch, zh, sh	t, tj, z, zj, s, sj, c, ch, sch, zh, sh
2	p, pj	b, bj, p, pj, k, kj, t, tj
3	d, dj	d, dj, k, kj, z, zj, zh
4	b, bj	d, dj, b, bj, p, pj
5	k, kj	p, pj, b, bj, k, kj, g, gj
6	g, gj	k, kj, g, gj
7	m, mj, n, nj	m, mj, n, nj
8	t, tj	z, zj, zh, d, dj, k, kj, t, tj
9	первая фонема – гласный звук	z, zj, zh, p, pj, b, bj, p, pj, k, kj, g, gj, m, mj, n, nj, t, tj

С учетом того, что часто возникает риск неправильного определения согласной фонемы, для кодирования последовательности фонем будем использовать код, состоящий из четырех цифр:

- 1) длина последовательности фонем;
- 2) количество гласных;
- 3) группа первого согласного;
- 4) группа второго согласного.

Например, слова «дедушка» (Таблица 1) с транскрипцией «dj e0 d u0 sh kj a0» будет закодировано как 733, а его последовательность фонем «bj dj e0 d u0 sh kj a0» – как 8343.

Так как коды слов в словаре и коды последовательности фонем различаются, необходимо определить правило перекодирования,

согласно которому для каждого кода последовательности фонем будут определяться коды транскрипций слов, которые могут быть распознаны из этой последовательности фонем.

Код последовательности фонем состоит из четырех частей: из длины L , числа гласных V , кода группы первого согласного последовательности фонем $C1$ и кода второго согласного последовательности фонем $C2$. Множество слов R соответствующих заданному коду последовательностей фонем выражается формулой,

$$R = R_L \cap R_V \cap (R_{C1} \cup R_{C2}),$$

где R_L – множество слов с длинами транскрипции из диапазона, куда попадает число фонем L (см. Таблицу 2), R_V – множество слов с числом гласных в транскрипции из диапазона, куда попадает число гласных V (см. Таблицу 3), R_{C1} (R_{C2}) – множество слов с первыми (вторыми) согласными из групп $C1$ ($C2$).

Мощность множества R во много раз меньше мощности всего множества слов.

На Рис. 4 пояснен процесс перекодирования последовательности фонем с кодом 8343, который получен в результате произношения слова «дедушка» (см. Таблицу 1) в коды транскрипций слов, которые могут быть распознаны из этой последовательности фонем, а в Таблице 5 приведены результаты перекодирования.



Рис. 4.
Процесс перекодирования

Из Таблицы 5 видно, что код 8343 соответствует 48 кодам транскрипций. Заметим, что распознаваемое слово с кодом транскрипции 733 включено в таблицу.

Таблица 5.

Результаты перекодирования кода 8384

424	524	624	724	824	924	1024	1124
434	534	634	734	834	934	1034	1134
444	544	644	744	844	944	1044	1144
423	523	623	723	823	923	1023	1123
433	533	633	733	833	933	1033	1133
443	543	643	743	843	943	1043	1143

Из примера видно, что в конечном итоге просмотру подлежит не весь исходный словарь, а только та его часть, которая содержит слова с кодами транскрипций из найденного множества.

В четвертой части главы описан процесс поиска слов в базах данных на основе фонетического кодирования.

Перед запуском процесса распознавания речи нужно создать таблицу базы данных, содержащую слова и их транскрипции, а также фонетический код, полученный по описанному в главе 3 алгоритму, который будет являться индексом этой таблицы.

Создается также таблица перекодирования, которая используется для преобразования кода последовательности фонем в индексы близких по произношению слов из первой таблицы. Процесс звукового поиска слов в базе данных показан на Рис. 5.

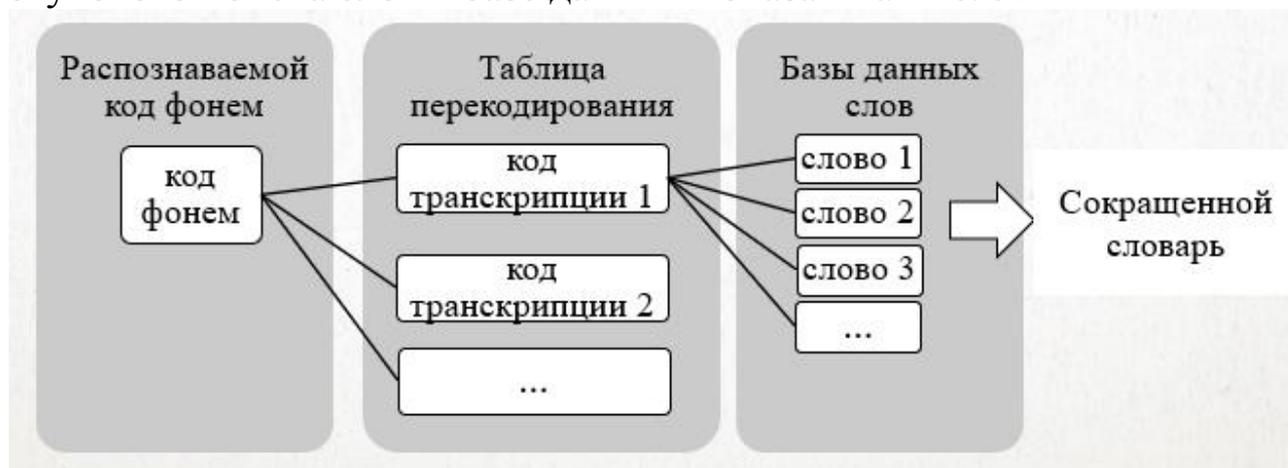


Рис. 5.

Процесс поиска слов в базах данных на основе фонетического кодирования

В последней части главы описаны способы применения разработанного метода звукового поиска на основе фонетического кодирования.

При распознавании речи (см. Рис. 2) в процессе декодирования обрабатывает большое число гипотез для каждого фрагмента аудиосигнала, который может иметь в своем составе различные слова с различным произношением.

В распространенном и используемом при таком декодировании методе динамического декодирования на основе лексического дерева, управляемого историей (англ. History Conditioned Lexical Tree, HCLT) используется словарь произношений, скомпилированный в виде дерева лексических префиксов. Лист дерева представляет собой устоявшуюся в естественном языке форму одного или нескольких слов.

Когда декодером достигнут один узел (префикс слова), он сохраняет состояние и продолжает поиск далее. В случае неудачи производится откат назад и поиск продолжается. Этот процесс продолжается до окончания распознаваемой последовательности фонем (см. Рис. 6). Возникающая при этом рекурсия приводит к очень большой сложности реализации метода и к большому времени поиска.

Здесь имеется возможность использования разработанного метода поиска слов на основе фонетического кодирования. Когда декодер сформирует гипотезу слова, метод может на основе имеющейся последовательности фонем выбрать возможные наиболее вероятные ветви дерева, что уменьшает число проверяемых гипотез.

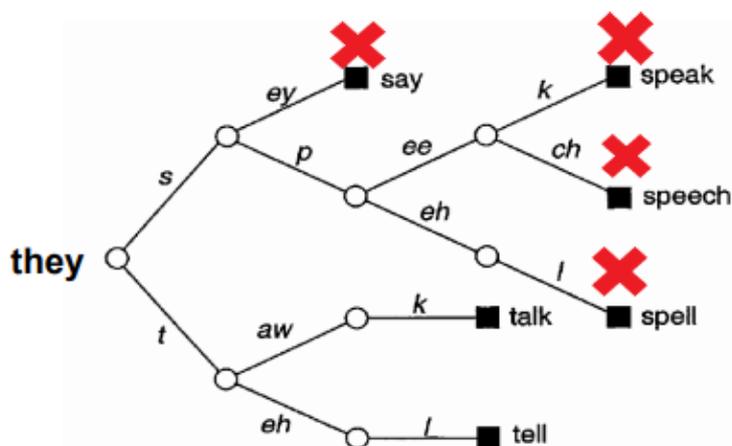


Рис. 6.

Пример обрезки на основе фонетического кодирования

Также разработанный метод может быть применен при распознавании речи на основе грамматик. В этом случае используется некоторая формальная грамматика, которая задает языковую модель (см. Рис. 7), определяющую устойчивые сочетания слов в языке. В этом случае декодер может проверять гипотезы о возможности следующей группы слов из грамматики путем вычисления и сравнения фонетических кодов.

{Поверните | Поверни} {направо | направо}

{Идите | Иди} {вперед | назад}

Рис. 7.

Пример грамматических правил

В этом случае ускоряется поиск наиболее близкого слова в каждой грамматической группе или отклоняется поиск в текущей ветви грамматики по причине отсутствия слов, выраженных заданной последовательностью фонем.

В третьей главе освещено экспериментальное исследование.

Метод звукового поиска слов в базах данных на основе фонетического кодирования внедрен в существующую систему распознавания речи. Для проведения экспериментов использовалась информационная система распознавания речи CMUSphinx с речевым материалом длительностью восемь часов, содержащим семь тысяч наиболее употребительных слов русского языка, которые произнесли четыре диктора: двое мужчин и две женщины.

Речевой материал был разделен на две части. Первая часть из пяти тысяч слов использовалась для разработки правил формирования фонетических кодов слов русского языка. Вторая часть, состоящая из оставшихся двух тысяч слов, была использована для анализа эффективности модифицированной системы распознавания речи и оценки правильности распознавания слов.

Результаты экспериментального исследования приведены в Таблице 6.

Из Таблицы 6 видно, что модифицированная информационная система, работает быстрее исходной более чем в два раза. При этом точность поиска отдельных слов (правильность их распознавания) также увеличивается с 56,1 до 61,7 процентов.

Результаты экспериментального исследования

Стадия распознавания речи, характеристика	Исходная информационная система, с		Модифицированная информационная система, с	
	Доверительный интервал	Среднее время	Доверительный интервал	Среднее время
Получение фонем	-	-	[0,039, 0,049]	0,044
Извлечение словаря	-	-	[0,34, 0,51]	0,42
Процесс распознавания	-	-	[0,73, 1,09]	0,91
Полное время	[2,34, 3,46]	2,90	[1,21, 1,55]	1,38
Правильность распознавания	56,1 %		61,7 %	

В заключении приведены основные результаты работы, которые заключаются в следующем:

1. Предложено использование фонетических алгоритмов для кодирования последовательности фонем при поиске слов в базах данных

2. Разработан эффективный алгоритм фонетического кодирования для последовательности фонем русского языка, позволяющий находить близкие по произношению слова.

3. Разработан метод поиска слов в базах данных на основе алгоритма фонетического кодирования.

4. Определена эффективность полученных результатов.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Выхованец В.С., Ду Цзяньмин, Сакулин С.А. Обзор алгоритмов фонетического кодирования // Управление большими системами. Выпуск 73. М.: ИПУ РАН, 2018. С.67-94 (1,75/0,88 п.л.).

2. Ду Цзяньмин. Метод сокращенного перебора слов при распознавании на основе фонетического кодирования // Динамика сложных систем. 2018. №3. С. 79-83 (0,31 п.л.).

3. Ван Л., Петросян О.Г., Ду Цзяньмин. Распознавание лиц на основе дерева коэффициентов для трехмасштабного вейвлет преобразования // Проблемы информационной безопасности. Компьютерные системы. 2018. №3. С. 126-136 (0,68/0,30 п.л.).

4. Ду Цзяньмин, Выхованец В.С. A method of reducing the search for words in speech recognition based on phonetic coding algorithm // Материалы XXI международной научной конференции «Distributed computer and communication networks» (DCCN 2018, Москва) М. 2018. С.21-30 (0,63/0,32 п.л.).

5. Ду Цзяньмин. Выборочное распознавание фонем с помощью смеси Гауссовых распределений // Молодежный научно-технический вестник. М.: Академия инженерных наук им. А.М. Прохорова. 2016. № 6. С.25 (0,56 п.л.).

6. Ду Цзяньмин. Отложенная слоговая сегментация при распознавании речи // Молодежный научно-технический вестник. М.: Академия инженерных наук им. А.М. Прохорова. 2014. № 3. С.29 (0,56 п.л.).