На правах рукописи

Полиев Александр Владимирович

Разработка алгоритмов для распознавания команд речевого интерфейса кабины пилота

Специальность 05.13.01 —
«Системный анализ, управление и обработка информации»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Москва — 2020

Работа выполнена на кафедре управляющих и информационных систем Московского физико-технического института (национального исследовательского университета).

Научный руководитель: **Корсун Олег Николаевич**
доктор технических наук, профессор,
профессор кафедры «Проектирование и сертификация авиационной техники» Института №1 «Авиационная техника» МАИ

Официальные оппоненты: **Никульчев Евгений Витальевич,**
доктор технических наук, профессор,
профессор кафедры управления и моделирования систем ФГБОУ ВО «МИРЭА — Российский технологический университет»

Чучупал Владимир Яковлевич,
кандидат физико-математических наук,
ведущий научный сотрудник ФГУ «ФИЦ «Информатика и управление» РАН»

Ведущая организация: ФГБУН Санкт-Петербургский институт информатики и автоматизации РАН

Защита состоится «21» мая 2020 года в 14:00 часов на заседании диссертационного совета Д 212.125.12 в ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)» по адресу: 125993, Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО «Московский авиационный институт (национальный исследовательский университет)» по адресу: 125993, Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4, а также на сайте института по адресу https://mai.ru/events/defence/index.php?ELEMENT_ID=110619.

Отзывы, заверенные печатью, просим направлять по адресу: 125993, Москва, А-80, ГСП-3, Волоколамское шоссе, д. 4, Ученый совет МАИ.

Автореферат разослан «__» _____ 2020 года.
Телефон для справок: +7 499 158-43-55.

Ученый секретарь
диссертационного совета Д 212.125.12,
кандидат технических наук, доцент  Старков А. В.

Общая характеристика работы

Актуальность темы исследования. На сегодняшний день взаимодействие человека с компьютерными системами через управление речевыми командами является одним из самых удобных и перспективных форматов.

В современных системах применяются 3 основные группы методов распознавания речи. Первая группа — это скрытые марковские модели. В них входная речь рассматривается как последовательность фонем с определёнными вероятностями перехода. Распознавание производится через поиск наиболее вероятной последовательности фонем для данного входного сигнала. Вторая — это методы, основанные на сравнении с эталоном. Для каждого слова из словаря некоторым образом составляется эталон. При распознавании выбирается то слово, эталон которого наиболее близок к входному сигналу. Третья группа методов основана на искусственных нейронных сетях. Суть методов состоит в нахождении такой решающей функции, которая по входному сигналу может определить его принадлежность к определённому классу. Искусственные нейронные сети построены по принципу организации биологических нейронных сетей и хорошо справляются с широким спектром задач.

В данной работе решается задача повышения вероятности правильных распознаваний и снижения влияния акустических шумов путём разработки и совершенствования алгоритмов распознавания команд речевого интерфейса пилота для управления бортовым оборудованием современных самолётов. По сравнению с обычной задачей распознания речи к речевому интерфейсу кабины пилота предъявляются следующие требования:

- распознавание ограниченного словаря из слов или фраз;
- компактность, автономность, высокое быстродействие;
- хорошее качество распознавания в условиях сильного шума.

С учётом этих требований широко используемые скрытые марковские модели не подходят из-за низкого качества распознавания в условиях шума. Остальные две группы методов в настоящий момент не обеспечивают необходимой надёжности распознавания. По этой причине тема настоящей работы, направленной на совершенствование методов распознавания речевых команд с помощью сравнения с эталоном и с использованием нейронных сетей, является актуальной.

Исследования, выполненные в рамках данной работы, направлены на решение таких практически значимых и актуальных задач, как предобработка входящего сигнала путём выделения однородных частей, улучшение качества эталонов с помощью выделения в них главных компонент, использование систем распознавания из нескольких эталонов, использование свёрточных нейронных сетей, обучаемых на выборке малого объёма.

Степень разработанности темы исследования. Первая попытка конструирования системы автоматического распознавания речи была сделана в 1952 году в работе K.H. Davis, а уже в 1970-е годы исследования F. Itakura, H. Sakoe, S. Chiba, В.М. Величко, Н.Г. Загоруйко в области распознавания речи достигли значительных успехов. В 1980-х годах научные работы в области распознавания речи перешли к моделированию статистическими методами на основе скрытых марковских моделей (Hidden Markov Models, HMM). Работы J.K. Baker были одними из первых, в которых для решения задачи распознавания речи были применены HMM. С 1990-х годов распознавание речи несколько усовершенствовалось. Словарь распознаваемых слов вырос до нескольких десятков тысяч. Использование быстрых методов декодирования позволило производить распознавание в реальном времени. В современных дикторозависимых системах, распознающих отдельные слова ошибки составляют около 1.5–2.5 %. И около 5 % ошибок в независимых от диктора системах, которые распознают слитную речь.

При этом в настоящее время очень мало работ, которые исследуют методы распознавания отдельных команд с небольшим объёмом словаря. Также недостаточно изучены возможности обучения на выборке небольшого размера в методах распознавания, основанных на нейронных сетях глубокого обучения.

Объект и предмет исследования. В работе в качестве объекта исследования рассматриваются речевые команды, а предметами исследования являются методы и алгоритмы распознавания речевых команд.

Целью исследования является повышение вероятности правильных распознаваний и снижение влияния акустических шумов, путём разработки алгоритмического обеспечения для распознавания команд речевого интерфейса кабины пилота в виде отдельных слов и фраз. За рамками работы остались выбор оптимального состава команд и их интерпретация.

Для достижения поставленной цели решаются следующие научно-технические **задачи**:

- системный анализ статистических свойств речевой информации и её обработка для решения задачи управления и принятия решений при распознавании речевых команд;
- разработка алгоритмов предварительного разбиения записей на однородные части в целях оптимизации процесса распознавания;
- разработка алгоритмов исключения шума и выделения наиболее значимых компонент в эталоне;
- исследование статистических закономерностей верного и неверного распознавания речевой информации и их использование для уменьшения количества ошибок;
- разработка алгоритмов использования нескольких эталонов одного слова для улучшения качества распознавания;

– исследование современных типов и архитектур искусственных нейронных сетей глубокого обучения для применения в задаче распознавания речевых команд.

Методология и методы исследования. Основными методами исследования, используемыми в работе, являются: анализ данных, цифровая обработка сигналов, теория вероятностей, математическая статистика, численная оптимизация, проектирование программных средств.

Научная новизна работы заключается в разработке совокупности алгоритмов, обеспечивающих повышение вероятности правильных распознаваний команд речевого интерфейса кабины пилота:

– алгоритм разбиения речевых команд на фонетически однородные части на основе модифицированного метода динамического программирования;
– алгоритм оптимизации эталонов на основе метода, в котором искомый эталон формируется как линейная комбинация главных компонент, оптимизирующая предложенный критерий качества;
– алгоритм оптимизации размерности параметрических портретов с предложенным выделением значимой информации с использованием полиномов Чебышёва;
– алгоритм распознавания команд по нескольким эталонам, отличающийся применением предварительного оценивания с использованием байесовского подхода и метода комитетов;
– алгоритм распознавания команд нейронными сетями глубокого обучения, отличающийся обучением на выборках малого размера.

Теоретическая и практическая значимость. Полученная в результате работы совокупность алгоритмов повышает точность распознавания речевых команд при различных уровнях шума, в том числе с учётом случая статически неустойчивого самолёта. Результаты работы могут быть применены в учебном процессе и в ходе разработки алгоритмического обеспечения речевого интерфейса пилота для таких задач, как отображение информации, выбор частоты радиооборудования, прокладка маршрута, управление системой опознавания и датчиками, запрос запаса топлива.

Основные положения, выносимые на защиту:

1) Алгоритм разбиения речевых команд на фонетически однородные части, отличающийся от существующих применением модифицированного метода динамического программирования.
2) Алгоритм оптимизации эталонов, отличающийся от существующих тем, что искомый эталон формируется как линейная комбинация главных компонент, оптимизирующая заданный критерий качества.
3) Алгоритм оптимизации размерности параметрических портретов, отличающийся выделением наиболее значимых составляющих с использованием полиномов Чебышёва.

4) Алгоритм распознавания команд по нескольким эталонам, отличающийся применением последовательного оценивания с расчётом апостериорных байесовских вероятностей.
5) Алгоритм распознавания команд нейронными сетями глубокого обучения, отличающийся от существующих обучением на выборке малого размера.

Достоверность результатов обеспечивается корректным применением математической статистики, методов идентификации и анализа данных, подтверждением полученных теоретических результатов с помощью экспериментов на различных наборах входных данных с несколькими уровнями шума, а также сравнением с известными результатами, ранее полученными другими авторами.

Публикации. По теме диссертации автором опубликовано 4 научных работы [1–4]: 3 из них в изданиях из списка, рекомендованного ВАК РФ [1–3], и 2 из них в изданиях, индексируемых в международных системах цитирования Scopus и Web of Science [1;4].

Апробация работы. Основные результаты исследования докладывались на следующих конференциях: Всероссийская научно-техническая конференция «XII Научные чтения по авиации, посвящённые памяти Н.Е. Жуковского» (Москва, 2015) [5], Восьмой Международный Аэрокосмический Конгресс IAC'15 (Москва, 2015) [6], Всероссийская научно-техническая конференция «XIII Научные чтения по авиации, посвящённые памяти Н.Е. Жуковского» (Москва, 2016) [7], Юбилейная Всероссийская научно-техническая конференция «Авиационные системы в XXI веке» (Москва, 2016) [8; 9], Вторая Международная научно-практическая конференция «Эрго-2016: Человеческий фактор в сложных технических системах и средах» (Санкт-Петербург, 2016) [10], международный семинар Workshop on Contemporary Materials and Technologies in the Aviation Industry — CMTAI (Москва, 2016) [11], Всероссийская научно-техническая конференция «Навигация, наведение и управление летательными аппаратами» (Москва, 2017) [12], Девятый Международный Аэрокосмический Конгресс IAC'18 (Москва, 2018) [13], Всероссийская научно-техническая конференция «Моделирование авиационных систем» (Москва, 2018) [14].

Объем и структура работы. Диссертация состоит из введения, четырёх разделов и заключения. Полный объём диссертации составляет 152 страницы, включая 26 рисунков и 54 таблицы. Список литературы содержит 94 наименования.

Содержание работы

Во **введении** обоснована актуальность работы, сформулированы цель и задача исследования, научная новизна и практическая значимость полученных результатов.

Первый раздел работы посвящён обзору современных методов распознавания речи, алгоритмов параметризации речевых сигналов и формирования эталонов. Также описаны основные математические методы, используемые в разработанных алгоритмах распознавания речевых команд. Анализ главных методов автоматического распознавания речи показывает их преимущества и недостатки, а также описывает сами методы, их вычислительную сложность и области применимости. Изложено подробное описание алгоритма получения параметрических портретов эталона и метода их использования для получения эталонов.

Среди математических методов, используемых в работе, присутствует метод главных компонент, который используется для улучшения качества используемых эталонов, и метод динамического программирования, используемый при разделении входного сигнала на однородные части. Также рассмотрены методы подстройки эталонов по длительности и полиномиальной аппроксимации по Чебышёву.

Второй раздел посвящён описанию всех предложенных алгоритмов, связанных с распознаванием путём сравнения с эталоном.

Вначале проводится изучение статистических свойств речевых команд и их параметрических портретов. Ключевое значение имеет исследование закона распределения и обоснование того, что во многих случаях это распределение для исследуемых параметров является нормальным. Такая проверка необходима, так как подавляющее большинство применяемых алгоритмов использует гипотезу о нормальности.

Рассмотрим M реализаций слова во временной области: $\tilde{x}_k(t)$, $k = 1, 2, \ldots, M$. Для каждой реализации можно получить параметрический портрет $X_k(i, j)$, представляющий собой матрицу, в которой строки $i = 1, 2, \ldots, N_t$ соответствуют делению слова на N_t интервалов по времени, а столбцы $j = 1, 2, \ldots, N_f$ соответствуют частотным компонентам для каждого временного интервала.

Полная энергия слова в форме дискретного сигнала принимает следующий вид: $E(x) = \sum_{i=1}^{N} x_i^2$. Энергия сигнала в полосе частот от f_0 до f_1 определяется как $E(f_0, f_1) = \int_{f_0}^{f_1} S(t)df = \int_{f_0}^{f_1} |X(f)|^2 df$. В качестве средней частоты принимается такая частота $f_{\text{ср}}$, что энергия составляющих сигнала с частотами в диапазоне $f \in [0, f_{\text{ср}}]$ равняется энергии составляющих сигнала с частотами $f \in [f_{\text{ср}}, +\infty)$, что для дискретного сигнала эквивалентно условию $\sum_{j=1}^{j_{\text{ср}}} \widehat{S}_x(f_j) \approx \sum_{j_{\text{ср}}+1}^{N/2} \widehat{S}_x(f_j)$.

Каждое слово записывается с разной амплитудой прежде всего из-за флуктуаций громкости произношения, изменения расстояния и ориентации губ диктора относительно микрофона. Это означает, что уже во временной области каждая реализация имеет индивидуальный коэффициент усиления c_k. В работе показано, что для исключения влияния c_k все M слов следует привести к единому масштабу по амплитуде. Коэффициент коррекции по амплитуде для каждого слова находится по формуле

$b_k = \sqrt{E_{mean}/E_k}$, где E_{mean} — средняя энергия сигнала по всем словам, а E_k — энергия слова k. В итоге получается сигнал, скорректированный по амплитуде, для которого должны выполняться условия нормальности.

Далее предлагается подход к автоматическому разделению слова на фонетически однородные части, при котором границы частей определяются в результате решения задачи многопараметрической оптимизации. В естественной речи длительность произношения заданного слова, как и длительность каждого звука в слове, не является постоянной величиной. Ручное выделение однородных частей в слове позволяет улучшить результаты распознавания слов через их сравнение с эталоном. Поэтому данный алгоритм может применяться в любой процедуре распознавания, в которой присутствует сравнение с эталоном. Но для эффективного алгоритма распознавания слов необходимо реализовать автоматический алгоритм разбиения слов на однородные части.

Фонетический состав слов естественных языков представляет собой некоторый код, предназначенный для передачи информации. Как известно, количество содержащейся в сообщении информации обратно пропорционально числу возможных вариантов данного сообщения. Применительно к звукам слов естественных языков этот результат можно интерпретировать в том смысле, что количество содержащейся в каждом звуке информации тем больше, чем сильнее он отличается от других звуков в слове и, прежде всего, от соседних звуков. Тогда определение границ между частями слова можно свести к математической задаче на поиск экстремума.

Придадим этим общим рассуждениям математический смысл. Фонетически однородной частью, границы которой подлежат определению, назовём часть, содержащую 2 или более элементарных интервалов. Такая часть соответствует прежде всего звуку, в отдельных случаях — слогу. Выразим границы частей через номера интервалов a_i, которые могут принимать значения $1 \le a_i \le N_t$, $i = \overline{0, L}$. Тогда для частей $k = \overline{1, L}$ границы задаются следующим образом: $k = 1 : [a_0; a_1], k = 2 : [a_1 + 1; a_2], \ldots, k = L : [a_{L-1} + 1; a_L]$, где $a_0 = 1$, $a_L = N_t$ и $a_1, a_2, \ldots, a_{L-1}$ — граничные интервалы частей.

Итак, в качестве принципа разбиения на части выберем однородность части и отличие его от соседних частей. В терминах коэффициентов корреляции между элементарными интервалами это можно представить в виде следующих условий:

- $\max_{a_0,\ldots,a_L} \sum_{k=1}^{L} \frac{1}{A_k} \sum_{i=a_{k-1}+1}^{a_k} \sum_{j=i+1}^{a_k} r_{ij}$ - интервалы, входящие в одну часть, должны иметь высокие коэффициенты корреляции;
- $\min_{a_0,\ldots,a_L} \sum_{k=1}^{L} \frac{1}{A_k - 1} \sum_{i=a_{k-1}+1}^{a_k} \sum_{j=i+1}^{a_k} (r_{ij} - \widehat{M}_k)^2$ - дисперсия взаимных коэффициентов корреляции между интервалами, входящими в одну часть, должна быть мала;

– $\min_{a_0,\ldots,a_L} \sum_{k=1}^{L-1} \frac{1}{A_{k,k+1}} \sum_{i=a_{k-1}+1}^{a_k} \sum_{j=a_k+1}^{a_{k+1}} r_{ij}$ - интервалы, входящие в одну часть, должны иметь малые коэффициенты корреляции с интервалами, входящими в соседние части;

где $A_k = \frac{1}{2}(a_k - a_{k-1})(a_k - a_{k-1} - 1)$, $A_{k,k+1} = (a_k - a_{k-1})(a_{k+1} - a_k)$, $\widehat{M}_k = \frac{1}{A_k} \sum_{i=a_{k-1}+1}^{a_k} \sum_{j=i+1}^{a_k} r_{ij}$. В общем случае результаты оптимизации разбиения по критериям могут не совпадать, хотя допустимо рассчитывать на близость результатов.

При решении оптимизационной задачи методом перебора задаются число частей и начальные значения пограничных точек или узлов. Шаг приращения принимается равным одному элементарному интервалу. Но получается, что использование полного перебора практически возможно только при небольших приращениях индексов от начальных значений. Для более точного разбиения слов на части потребуются очень большое время для работы алгоритма, поэтому целесообразно использовать методы динамического программирования.

Рассмотрим модифицированную схему динамического программирования, показанную на рисунке 1, для критерия, зависящего от двух соседних частей. В данном случае, в силу взаимосвязи между соседними частями на каждом этапе, необходимо рассматривать 2 части, то есть выполнять перебор значений 3 узлов, а не 2, как в стандартной схеме.

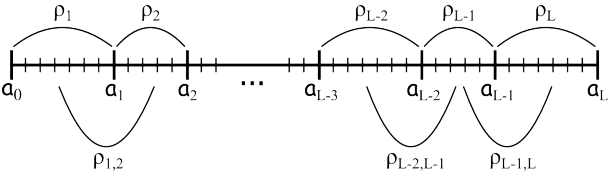

Рисунок 1 — Модифицированная схема динамического программирования

На первом этапе рассматриваем 2 крайних правых части. Последний узел a_L, как и ранее, считаем фиксированным. Задаём приращения двум другим узлам и перебираем все сочетания значений a^i_{L-1}, $i = \overline{-m, m}$, a^j_{L-2}, $j = \overline{-m, m}$, то есть $(2m+1)^2$ комбинаций. Рассматриваемые 2 узла полностью определяют 2 крайних правых части, что позволяет вычислить значение суммы критериев 1 и 2 для каждой комбинации. Для этого для всех индексов i, j вычисляем оценки средних коэффициентов корреляции внутри каждой части ρ^i_L, $\rho^{j,i}_{L-1}$, а также оценку среднего коэффициента корреляции между частями $\rho^{j,i}_{L-1,L}$. Тогда для 2 рассматриваемых частей $[a_{L-1}+1; a_L]$, и $[a_{L-2}+1; a_{L-1}]$ суммарные оценки средних коэффициентов корреляции (далее — суммарные оценки) вычисляются по формуле $P^{j,i}_{L-1,L} = \rho^i_L + \rho^{j,i}_{L-1} + \rho^{j,i}_{L-1,L}$. Всего получаем $(2m+1)^2$ значений суммарных оценок, которые приписываем каждой из комбинаций значений узлов a^i_{L-1} и a^j_{L-2}.

На втором этапе вводим в рассмотрение узел a_{L-3}, которому также придаём $2m+1$ значений a^l_{L-3}, $l=\overline{-m,m}$. Новый узел позволяет вычислить коэффициенты ρ_{L-2} и $\rho_{L-2,L-1}$. Эти коэффициенты зависят также от значений границ предыдущей части, то есть от значений a^i_{L-1}, a^j_{L-2}. Используя эти значения и перебирая $(2m+1)^3$ комбинаций по переменным l, j, i, вычисляем суммарные оценки $P^{l,j,i}_{L-2,L-1}=\rho^{l,j}_{L-2}+\rho^{l,j,i}_{L-2,L-1}+P^{j,i}_{L-1,L}$. Выделяем значения, соответствующие узлам a_{L-3} и a_{L-2}, то есть индексам l, j. Каждой комбинации этих коэффициентов соответствуют $2m+1$ значений узла a_{L-1}, то есть индекса i. Возьмём по индексу i максимум. В итоге получим $(2m+1)^2$ значений суммарного коэффициента, соответствующих узлам a_{L-3} и a_{L-2}: $P^{l,j}_{L-2,L-1}=\max_i\{P^{l,j,i}_{L-2,L-1}\}$. Итого, результатом второго этапа являются $(2m+1)^2$ комбинаций коэффициентов a_{L-3} и a_{L-2}, каждой из которых присвоено значение суммарной оценки, оптимальное по всем возможным положениям узла a_{L-1}.

Действуя аналогично на предпоследнем этапе для узлов a_1 и a_2, получаем $(2m+1)^2$ значений, каждому из которых присвоено значение суммарной оценки, оптимальное по всем возможным положениям узлов $a_3,\dots,a_{L-1}$: $P^{l,j}_{1,2}=\max_i\{P^{l,j,i}_{1,2}\}$. На последнем этапе добавляем узел a_0 и находим оптимальный вариант: $P_{1,2}=\max_{l,j}\{\rho^l_1+\rho^{l,j}_{1,2}+P^{l,j}_{1,2}\}$.

После этого рассматривается алгоритм формирования эталонов на основе метода главных компонент. Оптимальный эталон формируется путём разложения усреднённого эталона на главные компоненты и дальнейшей оптимизацией коэффициентов разложения на обучающей выборке с помощью метода покоординатного спуска. Полученный в результате оптимизации эталон может быть применён в любых алгоритмах, которые используют сравнение с эталоном.

Пусть имеется M параметрических портретов различных реализаций одного слова $X=\{x_{ij}(k)\}$, $k=1,2,\dots,M$; $i=1,2,\dots,N_t$; $j=1,2,\dots,N_f$. Преобразуем для каждого k матричный портрет в одномерный массив с числом элементов $i=1,2,\dots,P$, $P=N_fN_t$ и объединим эти M векторов в матрицу размерности $P\times M$:

$$X=\begin{bmatrix}x_1 & x_2 & \dots & x_M\end{bmatrix}=\begin{bmatrix}x_{11} & x_{12} & \dots & x_{1M}\\ x_{21} & x_{22} & \dots & x_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ x_{P1} & x_{P2} & \dots & x_{PM}\end{bmatrix}. \qquad (1)$$

Матрица корреляционных моментов равна $K_x=X^TX$ и является симметричной. Для неё можно вычислить M собственных чисел $\lambda_1,\lambda_2,\dots,\lambda_M$ (упорядочены по убыванию) и соответствующие им собственные векторы $l_1,l_2,\dots,l_M$. Первая главная компонента a_1 определяется как линейная комбинация исходных векторов $x_1,x_2,\dots,x_M$, взятых с коэффициентами, равными элементам собственного вектора $l_1^T=[l_{11}l_{21}\dots l_{M1}]$.

Аналогично вычисляются остальные главные компоненты. Смысл применения главных компонент состоит в том, что поведение системы определяется в основном несколькими первыми главными компонентами a_j, $j = 1, 2, \ldots, M'$, $M' < M$. Это позволяет уменьшить размерность задачи и рассматривать M' главных компонент вместо M исходных векторов.

Далее, представим эталон как линейную комбинацию M' главных компонент и постоянной составляющей с некоторыми коэффициентами: $E_{syn} = k_0 a_0 + k_1 a_1 + \cdots + k_{M'} a_{M'}$. Задача поиска оптимального эталона сводится к подбору таких коэффициентов $k_0, \ldots, k_{M'}$, которые будут удовлетворять критерию: значение Z-коэффициента корреляции $\to \max$. Для случая распознавания 3 слов суммарный критерий описывается формулой $F = \Delta Z_1^{low} + \Delta Z_2^{low} + \Delta Z_3^{low}$, где $\Delta Z_i^{low} = \min(\Delta Z_{ij}, \Delta Z_{ik})$, $i \neq j$, $i \neq k$, $j \neq k$, а ΔZ_{ij} — это разница Z-преобразований Фишера коэффициентов корреляций i-го эталона с распознаваемым i-м словом и j-го эталона с распознаваемым i-м словом.

Также можно дополнительно штрафовать за неправильное распознавание слова, что эквивалентно штрафу за значения Z_i^{low} меньше нуля: $F = \Delta Z_1^{*low} + \Delta Z_2^{*low} + \Delta Z_3^{*low}$, где α — это некоторое положительное число и

$$\Delta Z_i^{*low} = \begin{cases} \Delta Z_i^{low}, & \Delta Z_i^{low} \geq 0, \\ \Delta Z_i^{low} - \alpha(\Delta Z_i^{low})^2, & \Delta Z_i^{low} < 0. \end{cases} \quad (2)$$

Подбор коэффициентов на каждой итерации производится для каждого слова по очереди, при этом коэффициенты разложения двух других слов остаются неизменными.

Затем рассматривается разработка алгоритма формирования эталонов на основе полиномов Чебышёва. Пусть у нас есть параметрический портрет X, который содержит N_f частотных полос и N_t временных интервалов. Данный параметрический портрет помимо самого речевого сигнала содержит ещё и неинформативные сигналы, обусловленные особенностями речи определённого диктора и шумами. Использование полиномов Чебышёва поможет выделить только информативную часть, решив при этом сразу несколько задач. Во-первых, позволит уменьшить размерность параметрического портрета без существенной потери информативности, что упростит его хранение и ускорит обработку. Во-вторых, выделение только самого речевого сигнала может повысить качество распознавания. Сжатие можно производить по частотным полосам, по временным интервалам и по обоим измерениям одновременно. Также возможно использование полиномов Чебышёва не только для записей слов, но и для эталонов.

В конце рассмотрена разработка алгоритмов распознавания команд с использованием нескольких дикторов на основе формулы Байеса и метода комитетов. Для обеспечения дикторонезависимости распознавания

следует увеличивать разнообразие речевого материала в обучающей базе, например, за счёт применения нескольких эталонов, сформированных по записям разных дикторов. В данных алгоритмах для улучшения результатов распознавания могут быть применены подстройка по времени и оптимизированные эталоны, а также для ускорения работы может быть использовано сжатие используемых параметрических портретов.

Первый алгоритм использует формулу Байеса. В этом случае на основе обучающей выборки формируются априорные условные вероятности возможных вариантов распознавания слов, используемые для расчёта апостериорных вероятностей, что позволяет улучшить оценки и, как следствие, качество распознавания при невозможности выбора состава команд.

Пусть имеются гипотезы $H_1, \ldots, H_M$, соответствующие полной группе несовместных событий с априорными вероятностями $P(H_1), \ldots, P(H_M)$. Пусть в результате распознавания по одному эталону произошло событие A_k, то есть принята гипотеза H_k. Тогда по формуле Байеса условная апостериорная вероятность каждой гипотезы при условии, что произошло событие A_k, равна $P(H_i|A_k) = P(H_i)P(A_k|H_i)/\sum_{j=1}^{M} P(H_j)P(A_k|H_j)$, $i = 1, \ldots, M$. Для использования данной формулы необходимо определить априорные вероятности гипотез $P(H_1), \ldots, P(H_M)$ и получить для всех событий A_k оценки априорных вероятностей события $P(A_k|H_i)$, $k = 1, \ldots, M$. Изначально все априорные вероятности можно принимать равновероятными $P(H_i) = 1/M$ или использовать апостериорные вероятности, полученные на предыдущем этапе для случая многоэтапной процедуры распознавания.

Необходимо также получить оценки априорных условных вероятностей $P(A_k|H_i)$. Для этого необходимо использовать обучающую выборку: $P(A_k|H_i) = e_{ki}/E$, где e_{ki} — число событий A_k, при условии, что верна гипотеза H_i. Для случая L эталонов в формуле Байеса необходимо изменить только количество индексов: $P(H_i|A_{k_1 \ldots k_L}) = P(H_i)P(A_{k_1 \ldots k_L}|H_i)/\sum_{j=1}^{M} P(H_j)P(A_{k_1 \ldots k_L}|H_j)$.

Также дополнительная возможность для улучшения распознавания заключается в использовании значений меры близости Z как показателей качества распознавания. Введём характеристику качества распознавания $\Delta Z = Z_{\max} - Z_{\max -1}$, где $Z_{\max}$ — максимальное значением меры близости Z между параметрическими портретами распознаваемого слова и всеми M субэталонами, а $Z_{\max -1}$ — значение, ближайшее к максимальному. Можно принять, что качество распознавания прямо пропорционально значению показателя ΔZ. Тогда формула для одного эталона с учётом качества распознавания будет следующей:

$$P(H_i|A_k^{\Delta Z}) = \frac{P(H_i)P(A_k|H_i)P(\Delta Z|H_i)}{\sum_{j=1}^{M} P(H_j)P(A_k|H_j)P(\Delta Z|H_j)}. \qquad (3)$$

Для случая нескольких эталонов формула определяется аналогично.

Второй алгоритм основан на методе комитетов и заключается в независимом распознавании команд разными эталонами. На вход системы распознавания поступает неизвестное слово, которое последовательно распознаётся при помощи L эталонов, то есть вычисляются значения скалярной меры близости Z_i^j, где j — номер эталона, $j = 1, 2, \ldots, L$, а i — номер субэталона, $i = 1, 2, \ldots, M$. Предлагается для каждого j-го эталона сформировать коэффициенты $r_i^j = Z_i^j \Delta Z^j / p_i^j$, где p_i^j — место в рейтинге субэталонов.

Формула исходит из очевидных эвристических соображений. Действительно, «правильный» субэталон, то есть соответствующий распознаваемому слову, должен иметь наибольшую меру близости Z_i^j, наибольшее удаление ΔZ^j от всех остальных субэталонов и наименьшее по порядковому номеру, то есть первое, место в рейтинге. Поэтому для «правильного» субэталона коэффициент r_i^j максимален. Далее баллы, полученные по всем эталонам, суммируются и формируют итоговую оценку, на основе которой определяется результат распознания.

Третий раздел посвящён экспериментальному оцениванию характеристик распознавания предложенных алгоритмов. В начале описывается тестовая база речевых данных. В работе используется 2 основных набора слов и 1 набор фраз. Первый набор представляет собой записи 3 слов, произнесённых 13 дикторами, каждый диктор произносит около 50 реализаций каждого слова. Второй набор данных содержит 20 слов, произнесённых 9 дикторами, при этом каждый диктор произносит каждое слово по 30 раз. Последний набор состоит из 11 фраз, произнесённых 7 дикторами, при этом каждый диктор произносит каждую фразу по 30 раз.

Также используются записи с фоновым шумом и с шумом в наушниках. Шум в наушниках реализуется следующим образом. Диктор надевает наушники, в которые подаётся шум заданной громкости, при этом диктор плохо слышит свой голос, поскольку к наушникам не подключена обратная связь. Это используется для того, чтобы в экспериментальных условиях проверить степень изменения речи диктора в зависимости от уровня шума.

После этого была проведена проверка гипотез о нормальности распределения отклонения элементов параметрических портретов слов от эталона, а также показаны результаты расчётов длительности слов, их энергии и частоты. Для проверки был использован первый набор данных, состоящий из записей 3 слов без шума. Результаты показывают, что происходит увеличение доли проверок, подтверждающих нормальность по критерию Пирсона, с 30 % для случая некорректированных по амплитуде слов до 86 % проверок для слов, корректированных по амплитуде. По результатам статистической проверки можно сделать вывод, что после выравнивания слов по амплитуде входной сигнал удовлетворяет критерию нормальности. Поэтому дальнейшее использование алгоритмов, требующих нормальности входных данных, обосновано.

Далее описываются результаты экспериментов по разделению слов на однородные части. В качестве эталона принимался результат «ручного» разбиения, при котором границы частей определялись по результатам прослушивания и визуального анализа спектрограмм записи слова. На нескольких примерах было проведено сравнение метода полного перебора и рассмотренных выше алгоритмов динамического программирования, которое показало высокую точностью совпадения результатов. Поэтому в дальнейшем использовался только метод динамического программирования, который позволяет проводить разбиение записи одного слова меньше чем за секунду. Рассматривалось разбиение слов на 3–7 частей, при этом одна часть могла содержать от 2 до 25 интервалов. Диапазон варьирования границ частей принимался от минус 5 до плюс 5 элементарных интервалов от начального положения каждой границы. Обработка каждого слова выполнялась несколько раз для разных априорных значений положения границ. При этом практически отсутствовала зависимость результатов от априорных оценок при условии, что оптимальное значение находилось в области поиска.

Один из рассмотренных примеров — это слово «тысяча». В нём можно выделить 6 хорошо различимых звуков, по одному на каждую букву слова, что хорошо видно на рисунке 2 и в таблице 1.

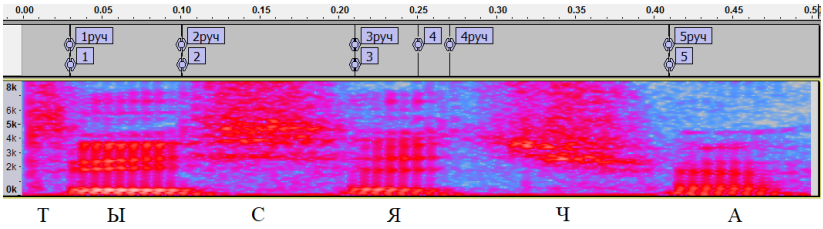

Рисунок 2 — Результаты разбиения слова «тысяча»

Таблица 1 — Границы частей для разных критериев, слово «тысяча»

Функционал	Границы частей, с					Значение функционала
	1	2	3	4	5	
J_1	0.03	0.10	0.21	0.25	0.41	0.75
J_2	0.03	0.10	0.21	0.24	0.41	0.01
J_3	0.03	0.10	0.21	0.25	0.41	0.13
J_{12}	0.03	0.10	0.21	0.24	0.41	0.74
J_{123}	0.03	0.10	0.21	0.24	0.41	0.60
Вручную	0.03	0.10	0.21	0.27	0.41	—

Аналогичные разбиения проводились и для других 5 слов. Анализ показывает достаточно хорошее совпадение границ, найденных автоматически, и эталонных границ, полученных в ручном режиме обработки. Проведённая обработка показывает также, что результаты, полученные при помощи различных критериев, близки между собой, особенно в тех случаях, когда однородные части содержат отдельные звуки, что соответствует допущениям, принятым при формировании критериев. Таким образом, проведённые эксперименты подтверждают работоспособность предложенных алгоритмов автоматического разбиения слов на фонетически однородные части.

Затем приводятся результаты эксперимента с получением оптимального эталона через метод главных компонент. Здесь тестирование алгоритма производится на примере простой задачи распознавания 3 слов, которые произносятся 10 различными дикторами в условиях без шума и 4 дикторами в условиях с шумом в наушниках (80 дБ и 90 дБ).

Вначале проводится проверка эффективности выделения главных компонент. Было показано, что первая главная компонента несёт в себе порядка 98 % имеющейся информации, первые 3 главных компоненты — 99 %, а при 6 главных компонентах доля достигает 99.5 %.

Результаты применения оптимизированного на обучающей выборке эталона, приведённые в таблице 2, показывают улучшение при распознавании реализаций не только для слов диктора, используемого в обучающей выборке, но также и для других дикторов, не входящих в обучающую выборку. Процент ошибок для 1200 записей трёх слов до оптимизации равнялся 5 %, а после оптимизации уменьшился до 1.25 %.

Таблица 2 — Число ошибок распознавания слов с шумом в наушниках 80 дБ на обычном (1) и оптимизированном (2) эталонах и с шумом в наушниках 90 дБ на обычном (3) и оптимизированном (4) эталонах

Диктор	Слово	(1)	(2)	(3)	(4)
Б-ак	пилотаж	6	0	7	0
	масштаб	0	0	0	0
	навигация	0	8	0	1
Г-ов	пилотаж	3	0	3	0
	масштаб	0	0	0	0
	навигация	0	0	0	2
Н-ов	пилотаж	8	0	12	0
	масштаб	0	0	0	0
	навигация	0	0	0	0
Ф-ев	пилотаж	9	0	4	0
	масштаб	0	0	0	0
	навигация	0	0	8	4
Суммарные результаты		**26** $\longrightarrow$ **8**		**34** $\longrightarrow$ **7**	

Также было протестировано использование ограниченного числа реализаций слов, используемых при построении оптимального эталона, и заранее заданного количества итераций оптимизации. Эксперимент показал, что достаточно использовать только 1 реализацию слова и проводить всего 10 итераций для получения оптимального эталона, что заметно сокращает время работы программы.

После этого описываются результаты экспериментов, связанных с аппроксимацией полиномами Чебышёва. Эксперимент проводился на параметрических портретах 20 слов, произнесённых 9 дикторами. Изначальные параметрические портреты содержали 35 частотных полос и 48 временных интервалов. Наиболее оптимальным в плане уменьшения размера параметрического портрета является одновременное сжатие и по частотным полосам, и по временным интервалам.

При использовании исходных записей без сжатия было получено 1.6 % ошибок. При использовании 18 полиномов в разложении по обоим измерениям получается 1.8 %, для 14 полиномов — 1.9 % ошибок и 2.0 % ошибок для 12 полиномов. Это позволяет уменьшить число элементов в параметрическом портрете в 5–10 раз. Полное совпадение количества ошибок достигается при очень большом числе полиномов, позволяя лишь незначительно уменьшить размеры параметрического портрета.

В конце показаны результаты распознаваний алгоритмами на основе формулы Байеса и метода комитетов. Проверка алгоритмов распознавания была проведена на речевой базе, включающей 20 слов для 8 дикторов. Формирование эталона во всех экспериментах проводилось по речевому материалу диктора, который не включался в распознаваемые записи, то есть был реализован дикторонезависимый вариант распознавания.

Результаты экспериментов представлены ниже в таблице 3. В первом столбце указан вариант распознавания: по одному эталону, который формировался по записям одного диктора, с точностью до единственного слова и с точностью до группы из 2 или 3 слов. Среднее количество ошибок при распознавании одним эталоном равно 8.42 %. Для эталонов по 7 дикторам получается заметное улучшение в 1.5–2 раза — в этом случае средняя ошибка для алгоритма на основе формулы Байеса равна 5.62 %, а для алгоритма на основе метода комитетов 5.33 %.

Следует отметить, что более простой эвристический алгоритм на основе метода комитетов при тестировании показал несколько лучшие результаты, чем более сложный и математически обоснованный алгоритм, использующий формулу Байеса. Наиболее вероятная причина заключается в том, что более сложный алгоритм оказывается чувствительным к погрешностям оценок априорных условных вероятностей и, следовательно, требует увеличения объёмов обучающих выборок.

Наилучший эффект, особенно для алгоритма на основе метода комитетов, достигается при переходе к группам из 2 и 3 слов, и такая

Таблица 3 — Процент ошибок при распознавании по 7 эталонам

Вариант теста	Порядковый номер диктора								Среднее значение
	1	2	3	4	5	6	7	8	
1 эталон	5.3	9.3	15	8.0	5.0	11	7.0	6.7	**8.42**
алгоритм на основе формулы Байеса									
до 1 слова	4.8	7.3	7.3	4.8	6.5	6.3	4.7	3.2	**5.62**
до 2 слов	2.7	4.8	3.8	2.2	2.5	2.7	2.2	0.8	2.71
до 3 слов	2.5	3.5	3.0	1.8	2.2	2.2	2.0	0.7	2.23
алгоритм на основе метода комитетов									
до 1 слова	4.5	7.3	7.2	4.8	4.5	6.7	4.8	2.8	**5.33**
до 2 слов	0.7	3.3	2.8	1.3	1.7	2.7	1.0	1.3	1.85
до 3 слов	0.5	2.5	1.0	0.5	1.3	1.7	0.5	0.5	1.06
подстройка	1.8	4.5	4.3	3.2	1.9	4.8	2.3	2.2	**3.13**

возможность локализации распознаваемого слова имеет практический смысл. Это позволяет существенно сократить время обработки за счёт перехода к иерархической процедуре распознавания: вначале быстродействующими алгоритмами выделяется малая группа, а затем в рамках группы проводится поиск алгоритмами, более затратными по времени, например, подстройкой по длительности или сравнением с эталонами «чужих» слов. Результат применения такого подхода для алгоритма на основе метода комитетов снижает число ошибок до 3.13 %, что в 2.7 раз меньше изначального результата при распознавании эталоном одного диктора.

В **четвёртом разделе** приведено описание разработки алгоритмов автоматического распознавания речевых команд на основе свёрточных нейронных сетей глубокого обучения. В распознавании использовались параметрические портреты с 18 частотными полосами и 25 временными интервалами. В данном случае для уменьшения размерности параметрических портретов может быть использован метод сжатия портретов на основе полиномов Чебышёва. Также для всех портретов из обучающей выборки может быть применено разбиение на однородные части и выравнивание найденных частей относительно друг друга для каждого слова.

Вначале были проведены оценки работоспособности традиционных сетей типа одно- и двухслойных персептронов в задаче распознавания речевых команд. Они показали неудовлетворительные результаты, с количеством ошибок в несколько раз больше чем в ранее приведённых методах.

После этого были апробированы нейронные сети глубокого обучения. Для них была выбрана оптимальная архитектура сети, имеющая по 2 слоя свёртки и подвыборки, за которыми идут 3 полносвязных слоя. Между полностью связанными слоями производится регуляризация, состоящая в случайном выбрасывании определённого количества нейронов в процессе обучения. Такой приём предотвращает переобучение и повышает стабильность результатов. Структура сети показана на рисунке 3.

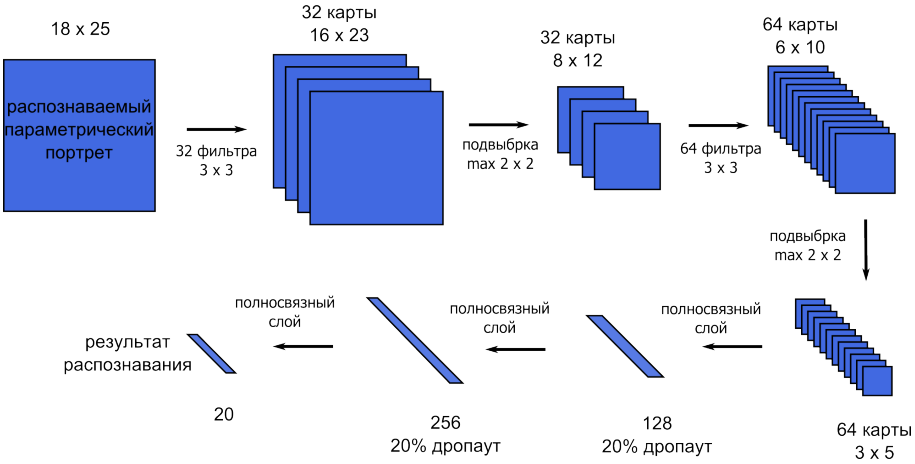

Рисунок 3 — Архитектура используемой свёрточной нейронной сети

В таблице 4 приведены суммарные результаты распознавания записей без шума при обучении на различном числе дикторов.

Таблица 4 — Результаты распознаваний слов и фраз на обучающих наборах из различного количества дикторов для случая записей без шума

Процент ошибок	Количество дикторов в обучающей выборке						
	1	2	3	4	5	6	7
Слова	5.9	2.4	1.7	1.2	1.0	0.8	0.6
Фразы	18.0	11.4	6.9	5.4	4.8	4.2	—

Из результатов можно сделать вывод, что процент ошибок стабильно снижается при увеличении количества дикторов в обучающей базе. Большее число дикторов позволяет избавиться от дикторозависимости, что положительно сказывается на качестве распознавания.

На рисунке 4 представлены усреднённые результаты распознаваний слов и фраз в условиях различных уровней шума при обучении на том же дикторе. Записи каждого диктора разбивались на 2 части — на первой части обучалась свёрточная нейронная сеть, а вторая часть использовалась в качестве тестовой выборки.

По результатам видно, что при обучении на всего лишь 15 записях каждого слова получается 0.9 % ошибок для слов и 1.3 % ошибок для фраз. При добавлении шума процент ошибок незначительно возрастает. Например, при отношении сигнал/шум равному 0 дБ, процент ошибок возрастает до 1.9 % для слов и до 5.4 % для фраз.

В таблице 5 приведены суммарные результаты распознавания записей с шумом при обучении на различном числе дикторов и различном количестве наложенных шумов.

Процент ошибок в зависимости от обучающей выборки и уровня шума																
Уровень шума	Количество записей в обучающей выборке															
	1	2	3	5	7	9	11	13	15	17	19	21	23	25	27	29
слова																
без шума	29.5	13.4	6.1	3.8	1.8	1.8	1.6	1.2	0.9	0.7	0.7	0.5	0.5	0.4	0.4	1.1
SNR = 8 дБ	27.5	14.5	7.7	4.1	2.6	2.0	1.9	1.6	1.6	1.0	1.0	0.6	0.9	0.9	0.6	1.1
SNR = 6 дБ	25.6	13.8	8.3	5.1	3.2	2.5	2.1	1.8	1.7	1.4	0.9	0.8	0.9	1.1	0.4	1.1
SNR = 3 дБ	27.3	15.5	8.1	5.2	3.7	2.5	2.2	2.2	1.8	1.4	1.0	0.9	1.1	1.0	0.4	1.7
SNR = 0 дБ	32.2	16.1	10.4	5.8	4.3	3.2	2.9	2.7	1.9	1.9	1.6	1.5	1.5	1.1	0.7	2.2
SNR = -3 дБ	35.6	18.8	11.8	7.0	5.1	4.8	3.2	3.2	2.8	2.4	2.2	1.7	2.0	2.1	1.1	2.2
фразы																
без шума	22.2	15.2	9.8	7.1	5.3	4.2	4.0	2.4	1.3	1.3	1.4	0.7	0.9	0.8	0.4	1.3
SNR = 8 дБ	20.1	16.3	15.1	10.0	9.0	6.0	4.9	3.3	3.3	3.7	3.1	3.3	3.0	2.6	2.2	2.6
SNR = 6 дБ	22.1	16.6	15.6	10.6	9.4	6.1	6.1	3.7	3.1	3.9	3.3	2.6	3.3	2.9	2.2	1.3
SNR = 3 дБ	25.5	21.0	17.1	10.6	9.9	8.2	6.9	5.4	4.3	4.0	3.7	3.5	3.9	3.6	3.0	2.6
SNR = 0 дБ	27.7	23.0	19.6	13.5	11.0	9.8	8.5	6.4	5.4	5.2	4.6	4.0	3.9	3.4	3.0	2.6
SNR = -3 дБ	19.8	15.7	13.4	9.5	8.6	7.2	6.9	5.7	5.4	5.9	5.9	5.9	6.3	6.4	6.3	6.6

Рисунок 4 — Результаты распознаваний при различных уровнях шума при обучении на том же дикторе, указан процент ошибок распознаваний

Таблица 5 — Результаты распознавания по «чужому» диктору при обучении на различном числе дикторов для записей слов и фраз с шумом

—	Слова						
число дикторов	1		3		7		—
число шумов	1	7	1	3	1	3	—
процент ошибок	9.7	8.2	2.8	2.5	1.2	1.1	—
—	Фразы						
число дикторов	1			3		6	
число шумов	1	5	10	1	5	1	2
процент ошибок	22.4	21.1	20.8	11.5	10.4	8.4	7.0

Из результатов можно сделать вывод, что процент ошибок стабильно снижается при увеличении количества дикторов в обучающей базе. Также можно сделать вывод, что увеличение количества накладываемых шумов приводит к уменьшению числа ошибок, но этот эффект от увеличения обучающей базы выражен менее заметно. Большое количество шумов в обучающей выборке уменьшает эффект переобучения для определённого типа шума, что уменьшает число ошибок при распознавании записей с другим шумом.

В конце проверялось, как изменится качество распознавания команд при обучении по фразам чужого диктора с добавлением небольшого числа записей своего диктора. При этом, это будут разные реализации команд, то есть добавленные к обучающей выборке записи исключаются из тестовой выборки. Таким образом, можно учесть некоторые индивидуальные

особенности распознаваемого диктора при добавлении лишь небольшого числа фраз, что достаточно несложно реализуется на практике.

В таблице 6 приведены результаты распознавания в эксперименте для различных конфигураций.

Таблица 6 — Результаты распознавания при добавлении фраз к обучающей выборке, состоящей из записей нескольких дикторов

Число	Процент ошибок при заданном числе добавленных записей						
дикторов	0	1	2	3	5	10	15
1	18.0	8.8	7.2	6.4	5.2	3.7	1.9
2	11.4	7.1	6.1	5.2	4.8	3.6	1.6
3	6.9	5.2	4.8	4.3	3.8	3.1	1.2
4	5.4	4.6	3.7	3.7	3.2	2.7	1.1
5	4.8	4.3	4.1	3.9	3.5	2.7	1.2
6	4.2	3.9	3.8	3.7	3.2	2.9	1.2

Как видно из результатов, при добавлении всего по 1 реализации каждой из фраз «своего» диктора в обучающую выборку ошибка распознавания уменьшается больше чем в 2 раза с 18.6 до 8.8 %. Дальнейшее добавление записей снижает ошибку, но величина снижения уже не такая большая. Также, эффект от добавленных записей тем больше, чем меньше дикторов использовано в обучающей выборке.

В **заключении** приведены основные результаты работы, которые состоят в следующем:

1) Разработан автоматический алгоритм разбиения слов на однородные части, в основе которого нахождение положения границ частей производится с помощью многопараметрической оптимизации. Сформулированы критерии, реализующие принцип максимизации меры сходства фонетического материала внутри части и меры различия между соседними частями. Для численного решения задачи с высоким быстродействием предложены алгоритмы, основанные на методе динамического программирования. Эксперименты, проведённые на примерах нескольких слов русского языка, подтвердили работоспособность предложенного подхода и правомерность принятых допущений.
2) Разработан алгоритм улучшения качества эталона, основанный на выделении и оптимизации главных компонент. Эталон, полученный с помощью оптимизации коэффициентов при главных компонентах, показал значительно меньшее число ошибок при распознавании большинства записей. Общее количество ошибок для записей слов с шумом в наушниках до оптимизации равнялось 5 %, а после оптимизации уменьшилось до 1.25 %. Также был сделан

вывод о том, что для получения приемлемых результатов достаточно использовать только одну реализацию слова и проводить всего 10 итераций при получении оптимального эталона, что заметно сокращает время работы программы.

3) Изучены способы и разработаны алгоритмы сжатия информации о параметрическом портрете с помощью применения полиномов Чебышёва. Эксперименты показали, что сжатие может происходить как отдельно по частотам и по времени, так и по обоим измерениям одновременно. В последнем случае можно сократить место для хранения параметрического портрета в 5–10 раз практически без ухудшения качества распознавания.
4) Разработаны алгоритмы на основе формулы Байеса и метода комитетов, позволяющие заметно уменьшить количество ошибок распознавания при использовании нескольких эталонов. Первый алгоритм использует оценки априорных вероятностей, определяемые по обучающей выборке, и рассчитывает апостериорные вероятности формулы Байеса, а второй является модификацией известного метода комитетов. Выявленная в ходе тестирования возможность локализации распознаваемого слова с точностью до малой группы позволяет повышать быстродействие систем распознавания на основе иерархических процедур, в которых последовательно применяются алгоритмы распознавания разных видов. Работоспособность обоих разработанных алгоритмов подтверждается результатами тестирования. При использовании 7 эталонов, полученных по записям различных дикторов, достигается заметное снижение процента ошибок в 1.5–2 раза — средняя ошибка для алгоритма на основе формулы Байеса снизилась с 8.42 до 5.62 %, а для алгоритма на основе метода комитетов до 5.3 и до 3.13 % при использовании подстройки по времени.
5) Изучены и модифицированы алгоритмы распознавания речевых команд на основе искусственных нейронных сетей глубокого обучения. Наилучшие результаты показала архитектура нейронных сетей CNN с двумя слоями свёртки и тремя полностью связанными слоями. При обучении на словаре из 20 слов без шума на 7 дикторах средняя величина ошибки при распознавании «чужих» дикторов равна 0.6 %. При обучении в той же конфигурации, но на записях с добавленным шумом, величина ошибки достигает для «чужого» диктора 1.1 %. Эксперименты по распознаванию фраз при использовании обучающей выборки из записей 6 дикторов показали 4.2 % ошибок для случая без шума и 7.0 % для записей с шумом. Получены положительные результаты в дикторозависимом варианте распознавания без шума и в условиях шума при использовании небольшого числа записей каждой команды

в обучающей выборке. Также получено значительное улучшение качества распознавания при добавлении всего нескольких реализаций каждой из речевых команд «своего» диктора в обучающую выборку, состоящую из записей «чужих» дикторов. При использовании нейронных сетей CNN количество ошибок является заметно более низким, чем для других алгоритмов, единственный замеченный недостаток — это длительное время обучения нейронной сети.

Публикации автора по теме диссертации

Публикации в научных изданиях, входящих в перечень ВАК:

1. *Полиев, А. В.* Автоматическое выделение фонетически однородных участков в словах естественного языка на основе многопараметрической оптимизации / А. В. Полиев, О. Н. Корсун // Известия Российской академии наук. Теория и системы управления. — 2016. — № 4. — С. 115—124.
2. *Полиев, А. В.* Разработка алгоритма синтеза оптимальных эталонов на основе метода главных компонент / А. В. Полиев // Cloud of science. — 2017. — т. 4, № 4. — С. 650—661.
3. *Полиев, А. В.* Использование нескольких эталонов при распознавании речи: формула Байеса и метод комитетов / А. В. Полиев, О. Н. Корсун // Вестник компьютерных и информационных технологий. — 2018. — т. 163, № 1. — С. 14—23.

Публикации в научных изданиях, индексируемых в международных системах цитирования Scopus и Web of Science:

4. *Poliyev, A. V.* Optimal pattern synthesis for speech recognition based on principal component analysis / A. V. Poliyev, O. N. Korsun // IOP Conference Series: Materials Science and Engineering. Vol. 312. — IOP Publishing. 2018. — P. 12—14.

Публикации по теме диссертации в других научных изданиях:

5. *Полиев, А. В.* Получение оптимального эталона с помощью метода главных компонент / А. В. Полиев, О. Н. Корсун // Всероссийская научно-техническая конференция «Научные чтения по авиации, посвящённые памяти Н.Е. Жуковского». — Общество с ограниченной ответственностью «Экспериментальная мастерская НаукаСофт». 2015. — С. 455—459.
6. *Полиев, А. В.* Алгоритм разбиения слов на однородные части в интересах разработки речевого интерфейса бортового оборудования / А. В. Полиев, О. Н. Корсун // Восьмой Международный Аэрокосмический Конгресс IAC'15. — АИР. 2015. — С. 178—180.

7. *Полиев, А. В.* Разработка модифицированного алгоритма динамического программирования для разбиения слов на однородные части / А. В. Полиев, О. Н. Корсун // Всероссийская научно-техническая конференция «Научные чтения по авиации, посвящённые памяти Н.Е. Жуковского». — Общество с ограниченной ответственностью «Экспериментальная мастерская НаукаСофт». 2016. — С. 194—201.
8. *Полиев, А. В.* Определение оптимального разбиения слова на однородные участки на основе матрицы корреляционного портрета / А. В. Полиев, О. Н. Корсун // Юбилейная Всероссийская научно-техническая конференция «Авиационные системы в XXI веке». — Государственный научно-исследовательский институт авиационных систем. 2016. — С. 162.
9. *Полиев, А. В.* Определение границ однородных участков слова на основе матрицы корреляционного портрета / А. В. Полиев // Юбилейная Всероссийская научно-техническая конференция «Авиационные системы в XXI веке». — Государственный научно-исследовательский институт авиационных систем. 2017. — С. 368—375.
10. *Полиев, А. В.* Разработка метода анализа фонетически однородных частей слов естественного языка / А. В. Полиев, О. Н. Корсун // Вторая Международная научно-практическая конференция «Эрго-2016: Человеческий фактор в сложных технических системах и средах». — Межрегиональная эргономическая ассоциация. 2016. — С. 370—377.
11. *Poliyev, A. V.* The algorithm of an optimal word pattern synthesis using principal component analysis / A. V. Poliyev // Workshop on Contemporary materials and technologies in the aviation industry - CMTAI. — 2016.
12. *Полиев, А. В.* Применение формулы Байеса для распознавания слов с использованием нескольких эталонов / А. В. Полиев, О. Н. Корсун // Всероссийская научно-техническая конференция «Навигация, наведение и управление летательными аппаратами». — Издательство «Научтехлитиздат». 2017. — С. 114—116.
13. *Полиев, А. В.* Разработка алгоритма распознавания слов в условиях шума на основе сверточных нейронных сетей / А. В. Полиев, О. Н. Корсун // Девятый Международный Аэрокосмический Конгресс IAC'18. — АИР. 2018. — С. 124—126.
14. *Полиев, А. В.* Распознавание речевых команд на основе сверточных нейронных сетей / А. В. Полиев, О. Н. Корсун // III Всероссийская научно-техническая конференция «Моделирование авиационных систем». — Государственный научно-исследовательский институт авиационных систем. 2018. — С. 261.