

МИНОБРНАУКИ РОССИИ
Федеральное государственное
бюджетное учреждение науки
Санкт-Петербургский институт
информатики и автоматизации
Российской академии наук
(СПИИРАН)

14 линия, 39, Санкт-Петербург, 199178
Телефон: (812) 328-33-11, факс: (812) 328-44-50
E-mail: spiiiran@iias.spb.su,
http://www.spiiiran.nw.ru
ОКПО 04683303, ОГРН 1027800514411
ИНН/КПП 7801003920/780101001

№ 060-11-5-11/769 от 06.04.2020

УТВЕРЖДАЮ
Директор СПИИРАН

д.т.н., проф.



А.Л. Ронжин

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертационную работу Полиева Александра Владимировича, на тему «Разработка алгоритмов для распознавания команд речевого интерфейса кабины пилота», представленную к защите на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (авиационная и ракетно-космическая техника)

Рациональная и надёжная организация человеко-машинного взаимодействия представляет собой важную и актуальную задачу. На сегодняшний день взаимодействие человека с компьютерными системами через управление речевыми командами является бесконтактным, естественным и интуитивным интерфейсом. Активно развиваются голосовые интерфейсы в интеллектуальных системах управления различными техническими устройствами и подвижными объектами. Для повышения уровня безопасности полёта необходимо нивелировать нагрузку пилота от задач, отвлекающих его от выполнения основных функций. В последнее время в разных странах активно ведутся работы по развитию речевого интерфейса управления бортовым оборудованием летательных аппаратов. Поэтому тема диссертационного исследования Полиева А.В. является актуальной.

Важными требованиями к речевому интерфейсу кабины пилота являются: хорошее качество распознавания, поскольку ошибки могут повлиять на безопасность полета; устойчивость и робастность системы распознавания речи к акустическим шумам, уровень которых в современных самолетах остается весьма высоким; а также компактность, автономность и высокое быстроедействие автоматической системы распознавания.

Отдел документационного
обеспечения МАИ

Вх. №

« 08 06 2020 »

В диссертационной работе Полиева Александра Владимировича 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 гражданских самолётов.

Отличительной особенностью работы является то, что автор рассматривает распознавание заранее определенного набора команд в виде управляющих слов-команд и фраз. Этот выбор является рациональным, поскольку использование свободной речи в системе управления самолетом даже при условии правильного распознавания может породить проблему неоднозначной интерпретации слов пилота, что усложняет процесс управления.

Соискатель при выполнении диссертационной работы сосредоточился на решении таких фундаментальных задач, как предобработка входящего речевого сигнала путём выделения фонетически однородных частей фраз, улучшение качества эталонов с помощью метода главных компонент и численной оптимизации, использование в системе распознавания нескольких эталонов и байесовского оценивания, использование свёрточных нейронных сетей, обучаемых на выборке малого объёма как при отсутствии, так и при воздействии аддитивных акустических шумов.

При проведении исследования автор решил актуальную научную задачу, заключающуюся в разработке совокупности алгоритмов, обеспечивающих повышение вероятности правильного распознавания команд речевого интерфейса кабины пилота, в том числе в условиях аддитивных акустических шумов.

В диссертационной работе сформулированы следующие основные научные положения, выносимые на защиту:

1. Алгоритм разбиения слов на фонетически однородные части на основе модифицированного метода динамического программирования.
2. Алгоритм оптимизации эталонов с использованием разложения в линейную комбинацию главных компонент.
3. Алгоритм оптимизации размерности параметрических портретов с предложенным выделением значимой информации с использованием полиномов Чебышёва.
4. Алгоритм распознавания с использованием нескольких эталонов с последовательным оцениванием с расчетом апостериорных байесовских вероятностей.
5. Алгоритм распознавания команд свёрточными нейронными сетями с обучением на выборках малого размера.

Научная новизна работы определяется применением оригинальных алгоритмов распознавания команд речевого интерфейса кабины пилота. Автор впервые разработал алгоритмы разбиения речевых команд на фонетиче-

ски однородные части на основе модифицированного метода динамического программирования, оптимизации эталонов с использованием разложения в линейную комбинацию главных компонент, распознавания с использованием нескольких эталонов с последовательным оцениванием и расчетом апостериорных байесовских вероятностей, распознавания команд свёрточными нейронными сетями с обучением на выборках малого размера. Указанные методы позволяют увеличить процент правильного распознавания изолированных команд речевого интерфейса кабины пилота.

Достоверность научных результатов обеспечивается корректным применением методов анализа данных и идентификации, математической статистики, подтверждением полученных теоретических результатов с помощью экспериментов на различных наборах входных данных с несколькими уровнями шумов, а также сравнением с некоторыми известными результатами, ранее полученными другими авторами. Результаты опубликованы в 3 научных работах в изданиях из перечня, рекомендованного ВАК при Минобрнауки России, и неоднократно докладывались на научно-практических конференциях всероссийского и международного уровня.

Обоснованность полученных в диссертации выводов, результатов и рекомендаций, обеспечивается корректным выбором допущений и ограничений при формировании исходных данных и моделировании, применением апробированного математического аппарата и корректностью постановки частных научных задач.

Теоретическая и практическая значимость работы состоит в том, что разработанные методы и алгоритмы могут использоваться для создания программного обеспечения речевого интерфейса кабины пилота для выполнения голосовых команд, некритичных для безопасности полета, таких как выбор частоты радиооборудования, управление системой опознавания и датчиками, запрос запаса топлива самолета.

Результаты диссертационной работы могут быть внедрены в авиастроительных концернах («Сухой», «Ил», «МиГ» и др.), а также в учебном процессе в профильных российских ВУЗах (МАИ, СПбГУ ГА, МГТУ ГА и др.).

Диссертационная работа содержит 4 главы, в которых изложены предложенные автором модели, методы и результаты их реализации.

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

Первая глава посвящена 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, представлено описание самих методов, их вычислительной сложности и области применения. Подробно описан алгоритм получения параметрических портретов эталона речевой команды.

Вторая глава посвящена описанию предложенных алгоритмов распознавания речи, основанных на сравнении с эталоном. Изучаются статистические свойства речевых команд и их параметрических портретов. Проведено исследование закона распределения и обоснование того, что во многих случаях это распределение для исследуемых параметров является нормальным.

Далее предлагается подход к автоматическому разделению изолированных речевых команд-слов на фонетически однородные части, при котором границы частей определяются в результате решения задачи многопараметрической оптимизации. Подробно описывается теоретическая часть алгоритма разбиения слов на однородные части с помощью модифицированного метода динамического программирования.

После этого описывается алгоритм формирования эталонов на основе метода главных компонент. В нём оптимальный эталон формируется путём разложения усреднённого эталона на главные компоненты и дальнейшей оптимизацией коэффициентов разложения на обучающей выборке с помощью метода покоординатного спуска.

Затем рассматривается разработка алгоритма формирования эталонов на основе полиномов Чебышёва. Предполагается, что использование полиномов Чебышёва помогает выделить только информативную часть используемых параметрических портретов.

В конце главы рассмотрена разработка алгоритмов распознавания команд с использованием записей речи нескольких дикторов на основе формулы Байеса и метода комитетов. Показано, что для обеспечения дикторнезависимости распознавания следует увеличивать разнообразие речевого материала в обучающей базе, например, за счёт применения нескольких эталонов, сформированных по записям разных дикторов. В алгоритме на основе формулы Байеса 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. Второй алгоритм основан на методе комитетов и заключается в независимом распознавании команд разными эталонами, после чего частные оценки, полученные по всем эталонам, суммируются и формируют итоговую оценку, на основе которой определяется результат распознавания речи.

Третья глава посвящена экспериментальному оцениванию характеристик распознавания предложенных алгоритмов. В начале главы описана тестовая база речевых данных. После этого представлена проверка гипотез о

нормальности распределения отклонения элементов параметрических портретов слов от эталона, а также показаны результаты расчётов длительности слов, их энергии и частоты.

Далее описываются результаты экспериментов по разделению слов на однородные части. Приведены примеры, показывающие хорошее совпадение результатов ручного и автоматического разбиения слов.

Затем приводятся результаты эксперимента для получения оптимального эталона через метод главных компонент на примере записей слов с различным акустическим уровнем шума в наушниках. Приведенные результаты показывают уменьшение качества неправильно распознанных речевых команд в 4 раза.

После этого описываются результаты экспериментов, связанных с аппроксимацией полиномами Чебышёва. Показано, что наиболее оптимальным в плане уменьшения размера параметрического портрета является одновременное сжатие по частотным полосам и по временным интервалам.

В конце показаны результаты распознавания речи алгоритмами на основе формулы Байеса и метода комитетов. Оба метода показали уменьшение процента неправильных распознаваний в 1,5-2 раза, а при использовании дополнительной подстройки по времени процент ошибок уменьшился в 2,7 раза от начального результата. Отмечено, 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.

В четвёртой главе приведено описание разработки алгоритмов автоматического распознавания речевых команд на основе свёрточных нейронных сетей глубокого обучения.

В начале главы приведены оценки работоспособности традиционных сетей типа одно- и двухслойных перцептронов в задаче распознавания речевых команд, которые показали неудовлетворительные результаты, при этом количество ошибок стало в несколько раз больше, чем в ранее приведённых методах.

После этого представлены используемые нейронные сети глубокого обучения, для которых экспериментально подобрана оптимальная архитектура, имеющая по 2 слоя свёртки и подвыборки, за которыми идут 3 полносвязных слоя. Получены положительные результаты в дикторозависимом и дикторонезависимом (группонезависимом) режимах распознавания речи без шума и в условиях аддитивного акустического шума при использовании небольшого числа записей каждой команды в обучающей выборке. При использовании свёрточных нейронных сетей количество ошибок является за-

метно более низким, чем для других алгоритмов, единственным недостатком является длительное время обучения нейронной сети.

Автореферат, в целом, отражает основное содержание диссертации. Материалы диссертационной работы в достаточной мере опубликованы в печатных журналах и изданиях, рекомендованных перечнем ВАК, докладывались и получили положительную оценку на международных научно-технических конференциях.

Замечания по диссертационной работе:

1. Для исследований устойчивости системы к воздействию акустических шумов автор использует собственную речевую базу, к аудиозаписям которой аддитивно добавлены акустические шумы, записанные в кабине пилотов одного современного пассажирского самолета. Такой подход зависит от типа самолета и условий полета, очевидно, что для самолета другого типа, имеющего другую шумовую картину, потребуется переобучение акустических моделей системы распознавания речи.

2. В работе не показано место предложенной автором модификации динамического программирования в рамках общего подхода динамического программирования, в котором вариант автора является только одним из частных случаев, не раскрыта связь с классическим подходом динамического программирования (динамического выравнивания времени), который давно применяется для автоматического распознавания речи.

3. Не вполне ясен предложенный автором принцип разделения изолированных слов-команд на фонетически однородные части. В приведенных автором примерах в слове «больше» звук «б» не выделен в отдельную часть из-за малой длительности взрывной части звука, однако в слове «пилотаж» звук «п» выделен в отдельную часть.

4. Отсутствует экспериментальное сравнение предложенного подхода к распознаванию изолированных речевых команд с существующими подходами к автоматическому распознаванию речи.

5. В работе практически не исследуется и не упоминается так называемый «эффект Ломбарда», который важен при анализе речевых команд пилота кабины реального самолета.

Несмотря на указанные недостатки, представленная диссертационная работа является законченным научным исследованием, в котором решена актуальная научно-техническая задача разработки системы алгоритмов, обеспечивающих высокую вероятность правильного распознавания речевых команд интерфейса кабины пилота, в том числе в условиях акустических шумов, при имеющихся малых объемах обучающей выборки. Положительный эффект достигается благодаря использованию системы распознавания с малым словарем заранее определенных и фиксированных команд, а не распо-

знавания разговорной речи, однако для поставленной задачи бесконтактного управления отдельными системами бортового оборудования самолетов это ограничение является разумным и оправданным.

В целом, данные замечания не умаляют общей положительной оценки диссертации и значимости полученных научно-практических результатов.

Заключение

Личный вклад автора отражен в разработанных им новых алгоритмах и моделях. Исследование выполнено на достаточно высоком уровне, поставленные задачи решены. Содержание диссертации соответствует пунктам выбранной специальности 05.13.01 – «Системный анализ, управление и обработка информации (авиационная и ракетно-космическая техника)».

Кандидатская диссертационная работа «Разработка алгоритмов для распознавания команд речевого интерфейса кабины пилота» соответствует требованиям пункта 9 «Положения о присуждении ученых степеней», предъявляемым к диссертациям на соискание ученой степени кандидата технических наук, а ее автор, Полиев Александр Владимирович, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (авиационная и ракетно-космическая техника)».

Доклад Полиева А.В. был заслушан, а отзыв ведущей организации на диссертацию рассмотрен и одобрен на заседании Лаборатории речевых и многомодальных интерфейсов СПИИРАН 18 марта 2020 г., протокол № 1.

Г.н.с. (рук.) лаборатории речевых
и многомодальных интерфейсов
СПИИРАН,
д-р техн. наук, доц.

Карпов Алексей Анатольевич

Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН)

Адрес: 199178, г. Санкт-Петербург, 14-я линия В.О., дом 39

Телефон: +7 812 328-33-11

Email: spiiran@iias.spb.su; karpov@iias.spb.su

Официальный сайт: <http://www.spiiras.nw.ru>