**1. Титульный слайд**

Уважаемые члены Ученого Совета! Вашему вниманию представляется работа на тему: **Разработка алгоритмов для распознавания команд речевого интерфейса кабины пилота.**

**2. Состояние проблемы и основные трудности в предметной области**

Перспективным способом взаимодействия человека и компьютерной системы является речевое управление. Сейчас используются 3 подхода к распознаванию речи:

Первый: скрытые марковские модели. Речь рассматривается как последовательность фонем с некоторыми вероятностями перехода. Распознавание проводится через поиск наиболее вероятной последовательности фонем.

Второй: сравнение с эталоном. Для каждого слова из словаря составляется эталон. При распознавании выбирается то слово, эталон которого наиболее похож на входной сигнал.

Третий: искусственные нейронные сети. В них находится решающая функция, которая по входному сигналу определяет его принадлежность к некоторому классу.

Проведенные раннее исследования показали, что скрытые марковские модели в условиях шума показывают около 50% ошибок, поэтому в данной работе будут использоваться 2 оставшихся подхода: эталоны и нейронные сети.

Используемые в самолётах системы распознают от 100 до 300 команд и показывают менее 10% ошибок.

Фирма Google использует технологии голосового ввода и показывает снижение ошибок с 23% в 13 году до 5% в 17 году.

В качестве сравнения, точность распознавания речи человеком зависит от словаря: от 0.1% для записей цифр и до 4% для телефонных разговоров.

**3. Цель исследования и практическая значимость**

Сложность бортового оборудования возрастает, поэтому необходимы дополнительные независимые каналы связи пилота и бортовой системы.

Цель исследования заключается в повышении вероятности правильных распознаваний и снижении влияния шумов путём разработки алгоритмов распознавания команд речевого интерфейса кабины пилота.

Еще не созданы нормативы, задающие процент ошибок для подобных речевых систем, поэтому в работе ставилась цель выйти на уровень распознавания речи человеком.

Алгоритмы должны обладать автономностью, быстродействием, обеспечивать высокую вероятность правильных распознаваний и устойчивость к шуму.

Ошибки могут привести к разным последствиям, но в работе принимается гипотеза, что ошибки первого и второго рода нежелательны в одинаковой степени. Поэтому применяется критерий максимума апостериорной вероятности, который даёт максимальную вероятность правильных распознаваний.

Речевой интерфейс используются для задач, некритичных для безопасности полета, например, отображение информации и управление датчиками. Также на первом этапе возможно ручное подтверждение команд.

**4. Научная новизна**

Научная новизна заключается в разработке совокупности алгоритмов, обеспечивающих повышение вероятности правильных распознаваний:

– алгоритм разбиения команд на однородные части; отличается применением модифицированного метода динамического программирования;

– алгоритм оптимизации эталонов; где искомый эталон формируется как линейная комбинация главных компонент, оптимизирующая предложенный критерий качества;

– алгоритм оптимизации размерности параметрических портретов; отличается выделением значимых составляющих полиномами Чебышёва;

– алгоритм распознавания команд по нескольким эталонам; отличается применением последовательного оценивания с расчётом апостериорных байесовских вероятностей;

– алгоритм распознавания команд свёрточными нейронными сетями; отличается обучением на выборках малого размера.

**5. Положения, выносимые на защиту**

Положениями, выносимыми на защиту, являются озвученные на прошлом слайде алгоритмы.

**6. Статьи и конференции**

Результаты опубликованы в 4 статьях, из них 3 в журналах ВАК и 2 в журналах из Scopus и Web of Science. А также апробировались на 9 конференциях.

**7. Тестовая база речевых данных**

На основе анализа аналогичных исследований и с учетом доступных возможностей авторами создана тестовая база, состоящая из 3 наборов: 20 слов для 9 дикторов, 3 слова для 13 дикторов и 11 фраз для 7 дикторов. Несмотря на небольшой объём, выборка достаточно представительна и в дикторонезависимых экспериментах получен качественный результат. Также используется запись шума из кабины пилота, что позволяет задавать отношение сигнал/шум.

Так как целью является повышение вероятности правильных распознаваний, за рамками работы остались выбор оптимального состава команд и их интерпретация.

**8. Получение параметрического портрета из звукового сигнала**

Обычно речь представлена в виде оцифрованного сигнала. Но для работы нужен более удобный формат и одним из общепринятых представлений является параметрический портрет. Это 2-мерная матрица: столбцы соответствуют временным интервалам, строки – частотным полосам, а значения описывают интенсивность сигнала.

Далее идёт алгоритм получения параметрического портрета. Сначала усиливаются высокочастотные компоненты. Затем проводится временное квантование, применяется одно из спектральных окон, вычисляется и логарифмируется модуль быстрого преобразования Фурье.

Затем составляется эталон как среднее нескольких параметрических портретов. Минус такого подхода – это смешение звуков из-за их разной длительности у разных людей.

**9. Пример – входной сигнал и параметрический портрет 35 х 48**

На слайде показан пример входного аудио сигнала; параметрический портрет с хорошим разрешением; часть используемой во всех алгоритмах матрицы; и визуализация этой матрицы.

Результаты прошлых исследований показывают, что 35 частотных полос и 48 временных интервалов оптимальны для распознавания через сравнение с эталоном. Также, в качестве меры близости оптимальным является использование Z-преобразования Фишера от среднего коэффициента корреляции соответствующих столбцов параметрических портретов.

**10. Разбиение на фонетически однородные части – функционалы**

Ручное выделение однородных частей решает проблему смешения звуков в эталоне и улучшает результаты распознавания, поэтому предлагается алгоритм автоматического разбиения. Слова являются результатом эволюции, поэтому можно предположить их максимальную распознаваемость. Это значит, что положение границ даёт максимальное различие соседних частей и сходство внутри одной части.

Это сводится к задаче поиска экстремума по следующим критериям:

1-й: интервалы одной части имеют максимальные коэффициенты корреляции.

2-й: дисперсия коэффициентов корреляции интервалов одной части минимальна.

3-й: интервалы одной части имеют минимальные коэффициенты корреляции с интервалами соседних частей.

Итоговым функционалом является нормированная комбинация предложенных критериев. Полный перебор границ возможен только с малым числом частей, поэтому целесообразны методы динамического программирования.

**11. Стандартная схема динамического программирования**

Применение целочисленного варианта динамического программирования возможно, если критерий вычисляется отдельно на каждой части. Этому условию удовлетворяют первый и второй критерии.

Стандартный алгоритм реализуется следующим образом. Начиная с конца слова, каждой границе задаются приращения. Затем для каждого приращения вычисляется функционал и определяются оптимальные положения всех границ правее текущей. На последнем шаге выбираются оптимальные границы для всего слова.

**12. Модифицированная схема динамического программирования**

Предложенная модифицированная схема нужна для третьего критерия, зависящего от двух соседних частей. Отличием является одновременное задание приращений для 2 соседних границ. Для каждой пары приращений вычисляется функционал и определяются оптимальные положения всех границ правее текущих.

**13. Разбиение на фонетически однородные части – примеры**

Далее описываются результаты. Эталоном было визуальное ручное разбиение. Рассматривалось разбиение на 3–7 частей, каждая от 2 до 25 интервалов по 10 мс.

Один из примеров — это слово «тысяча» на нижнем рисунке. В нём выделяются 6 хорошо различимых звуков, по одному на каждую букву. Для этого и других слов получено достаточно хорошее совпадение эталонных и найденных границ.

**14. Формирование эталона с помощью метода главных компонент – теория**

Далее предлагается способ улучшения качества распознавания через подстройку эталонов. Оптимальный эталон формируется путём разложения усреднённого эталона на главные компоненты, полученные из эталонов нескольких дикторов. Затем подбираются коэффициенты разложения, максимизирующие критерий F.

При распознавании 3 слов, F состоит из 3 слагаемых. ∆*Zilow*— это разница корреляций распознаваемого слова с истинным эталоном и с лучшим из оставшихся. Также вводятся штрафы за неправильное распознавание, то есть когда ∆*Zilow* отрицательное.

**15. Формирование эталона с помощью метода главных компонент – результаты 1**

Результаты для записей 4 дикторов в условиях с шумом в наушниках. Видно улучшение не только для диктора из обучающей выборки (это 3-й диктор), но и для других. Средний процент ошибок уменьшился в 4 раза.

**16. Формирование эталона с помощью метода главных компонент – результаты 2**

Также протестировано малое число записей и итераций при построении оптимального эталона. Оказалось достаточно всего 1 реализации слова и 10 итераций, что заметно улучшает быстродействие.

**17. Сжатие с помощью полиномов Чебышёва – теория**

Другая проблема – это долгое время работы предыдущих алгоритмов. Для её решения предлагается алгоритм оптимизации параметрических портретов на основе полиномов Чебышёва.

Их использование поможет выделить информативную часть портрета, уменьшая размерность без существенной потери информативности, что упростит хранение и обработку портрета. Сжатие можно проводить независимо по каждому измерению.

**18. Сжатие с помощью полиномов Чебышёва – результаты**

Эксперимент проводился на портретах 20 слов 9 дикторов. Для исходных записей получено 1.6% ошибок. Полное совпадение числа ошибок достигается при очень большом числе полиномов, позволяя лишь незначительно уменьшить размеры портрета. Для полиномов до 18 степени получается 1.8% ошибок, до 14 степени — 1.9%, до 12 степени – 2.0%. Это позволяет уменьшить размер портрета в 5–10 раз.

**19. Распознавание несколькими эталонами – метод Байеса 1**

Для дикторонезависимости нужно увеличивать разнообразие обучающей базы, например, за счёт эталонов разных дикторов. Эту задачу решают предложенные алгоритмы на основе формулы Байеса и метода комитетов.

Первый алгоритм использует формулу Байеса. На обучающей выборке вычисляются априорные условные вероятности *P*(*A|H*) возможных вариантов распознавания, которые используются для расчёта апостериорных вероятностей по первой формуле. Это улучшает качество распознавания при невозможности выбора состава команд. Изначально априорные безусловные вероятности *P*(*H*) задаются равными.

**20. Распознавание несколькими эталонами – метод Байеса 2**

Также улучшение распознавания возможно при введении меры качества ∆*Z*, равной разнице максимального значения меры близости *Z* и значения, ближайшего к максимальному. В итоге получается формула расчёта апостериорной вероятности для нескольких эталонов с учётом качества распознавания.

**21. Распознавание несколькими эталонами – метод комитетов**

Второй алгоритм основан на методе комитетов. Неизвестное слово распознаётся несколькими эталонами и для каждого вычисляются коэффициенты r.

Формула исходит из эвристических соображений. Истинный вариант имеет наибольшую меру близости Z, наилучшее качество ∆Z и наивысшее место в рейтинге *p*. Баллы по всем эталонам формируют итоговую оценку, по которой определяется результат распознавания.

**22. Распознавание несколькими эталонами – результаты**

Проверка проводилась на наборе из 20 слов 8 дикторов с дикторонезависимым распознаванием. В первом столбце указан вариант распознавания: одним эталоном для сравнения, 7 эталонами с точностью до 1 слова и до группы из 2 и 3 слов. Видно, что для 7 эталонов ошибки уменьшаются в 1.5–2 раза — с 8.5 до 5.5%.

Лучший эффект достигается при распознавании с точностью до 3 слов за счёт перехода к иерархическому распознаванию: вначале быстродействующими алгоритмами выделяется малая группа, а затем проводится поиск более точными алгоритмами, например, подстройкой границ, представленной в начале. Это снижает в ошибки почти в 3 раза до 3.1%.

**23. Распознавание свёрточными нейронными сетями – структура сети**

Другим подходом к расширению обучающей базы являются свёрточные нейронные сети. Была выбрана рациональная архитектура сети, имеющая по 2 слоя свёртки и подвыборки и 3 полносвязных слоя. Для снижения переобучения проводится регуляризация, состоящая в случайном выбрасывании нейронов при обучении.

В реализации использовался язык программирование Python, библиотеки TensorFlow и TFLearn и архитектура параллельных вычислений CUDA.

**24. Свёрточные нейронные сети – распознавание без шума**

В верхней таблице результаты распознавания без шума. Процент ошибок снижается при увеличении числа дикторов в обучении – это избавляет от дикторозависимости.

Далее проверялось качество распознавания при обучении по фразам чужого диктора с добавлением фраз своего диктора. Так можно учесть индивидуальные особенности, что несложно реализуется на практике.

При добавлении 1 реализации каждой фразы ошибка снижается больше чем в 2 раза с 18 до 8.8%. Дальнейшее добавление записей снижает ошибку, но не так быстро. Также, эффект от добавленных записей тем больше, чем меньше дикторов в обучающей выборке.

**25. Свёрточные нейронные сети – распознавание с шумом**

Далее идут результаты распознавания с шумом. Ошибки снижаются при увеличении числа дикторов и при увеличении числа наложенных шумов, но второй эффект менее заметен. Большое количество шумов уменьшает эффект переобучения для определённого типа шума, что снижает ошибки на записях с другим шумом.

**26. Свёрточные нейронные сети – распознавание по себе**

Здесь представлены результаты для разных уровней шума при обучении на своем дикторе. Записи каждого диктора разбивались на 2 части — первая для обучения, вторая для теста.

При обучении всего на 15 записях каждого слова получается 0.9% ошибок для слов и 1.3% ошибок для фраз. При добавлении шума ошибки незначительно увеличиваются. Например, при отношении сигнал/шум 0 дБ получается 1.9% ошибок для слов и 5.4% для фраз.

**27. Результаты работы – сравнение с эталоном**

В заключении приведены основные результаты, которые заключаются в следующем:

1. Предложен алгоритм разбиения слов на однородные части. Сформулированы критерии сходства фонетического материала внутри части и различия между соседними частями. Предложены алгоритмы оптимизации, основанные на методе динамического программирования. Проведённые эксперименты подтвердили работоспособность предложенного подхода.

2. Разработан алгоритм улучшения качества эталона, основанный на выделении и оптимизации главных компонент. Общее количество ошибок уменьшилось в 4 раза. Сделан вывод, что для получения оптимального эталона достаточно 1 реализации слова и 10 итераций, что заметно улучшает быстродействие.

3. Разработаны алгоритмы сжатия параметрических портретов с применением полиномов Чебышёва. Эксперименты показали, что при сжатии по обоим измерениям размер портрета сокращается в 5–10 раз почти без ухудшения качества распознавания.

4. Разработаны алгоритмы на основе байесовского подхода и метода комитетов, использующие нескольких эталонов. Для 7 эталонов разных дикторов, процент ошибок снижается в 1.5–2 раза, а с подстройкой границ почти в 3 раза.

**28. Результаты работы – нейронные сети**

Предложен алгоритм распознавания на основе свёрточных нейронных сетей. При обучении на наборе из 20 слов 7 дикторов ошибка равна 0.6% без шума и 1.1% с шумом. Для фраз на обучающей выборке из 6 дикторов получилось 4.2% ошибок без шума и 7.0% с шумом.

Получены положительные результаты в дикторозависимом варианте распознавания без шума и с шумом при использовании малой обучающей выборки. Также получено значительное улучшение качества распознавания при добавлении всего нескольких реализаций команд своего диктора в обучающую выборку, состоящую из записей чужих дикторов.

Таким образом, теоретическая и практическая значимость исследования заключается в том, что основные выводы и положения могут быть применимы в практической реализации речевого интерфейса кабины пилота.

**29. Доклад окончен**

Доклад закончен. Благодарю за внимание.