

ОТЗЫВ

Официального оппонента на диссертацию Полиева Александра Владимировича "Разработка алгоритмов для распознавания команд речевого интерфейса кабины пилота", представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – системный анализ, управление и обработка информации (авиационная и ракетно-космическая техника)

Диссертация Полиева А.В. выполнена на *актуальную* тему, связанную с исследованием методов распознавания речи, которые сохраняют работоспособность в условиях помех и искажений. Такие условия характерны для использования систем распознавания в различных роботизированных комплексах, в частности, в кабинах боевых самолетов. Таким образом, выбор направления диссертационного исследования представляется актуальным и имеет научное и практическое значение.

Основная цель диссертационного исследования Полиева А.В. заключалась в разработке алгоритмов и программ распознавания ограниченного набора команд речевого интерфейса пилота самолета с учетом искажений, вносимых в сигнал акустическими помехами и ломбард-эффектом.

Научная новизна работы заключается в: разработке новых алгоритмов распознавания команд речевого интерфейса кабины пилота:
– сегментации речевых команд на акустически однородные части на основе модифицированного метода динамического программирования;
– формированием эталонов с использованием метода главных компонент;
– понижения размерности эталонов с использованием полиномов Чебышёва;
– распознавания команд по нескольким эталонам с использованием байесовского подхода и метода комитетов;
– распознавания команд с помощью глубокой сверточной нейросети

Диссертация состоит из введения, четырех глав, заключения и списка использованных источников.

Во введении обоснована актуальность темы диссертации, определена цель работы, задачи и методы исследования.

В первой главе проведен обзор наиболее перспективных, по мнению автора, методов, которые могут быть использованы при создании речевого интерфейса для управления оборудованием современных самолетов.

Во второй главе посвящена разработке новых алгоритмов формирования эталонов для автоматического распознавания речевых команд на основе модифицированного метода динамического программирования, метода главных компонент, использования полиномов Чебышёва. Также приведены алгоритмы распознавания по нескольким эталонам и дикторам на основе формулы Байеса и метода комитетов.

Третья глава содержит описание и результаты численных экспериментов по оценке показателей эффективности распознавания команд речевого интерфейса пилота при использовании предложенных во второй главе алгоритмов.

В четвертой главе описаны результаты экспериментальных исследований алгоритмов автоматического распознавания речевых команд с использованием глубоких сверточных нейронных сетей.

В заключении сформулированы основные результаты работы и делаются выводы об эффективности разработанных методов и алгоритмов и целесообразности их использования в системах речевого управления бортовым оборудованием.

Обоснованность результатов диссертации. Теоретические рассуждения, приведенные в диссертации выполнены достаточно корректно, алгоритмы доведены до программной реализации и проверены на собранном автором материале. Положительные результаты экспериментов свидетельствуют об обоснованности выводов, сделанных в диссертации.

По тексту диссертации имеются следующие **замечания**.

1. В обзорной главе нет сведений о гибридных и сквозных архитектурах систем распознавания речи, которые на сегодняшний день определяют мировой научно-технический уровень в этой области.
2. Характеристики корпуса данных для обучения и тестирования исследуемых методов (словари из 3 и 20 слов, 11 фраз и количества дикторов, в экспериментах - 7) представляются недостаточными, поэтому заключения об эффективности предлагаемых методов имеют предварительный характер.
3. Архитектура использованной в диссертации сверточной нейросети зависит от размерности входа, которая, в свою очередь, определяется эвристически и различается в зависимости от состава словаря. Таким образом при модификациях словаря нужно изменять архитектуру сети, что неудобно.
4. Предлагаемые в диссертации методы не являются методами распознавания речи в реальном времени поскольку входом является весь речевой образ, от начала до конца. Поэтому возникает вопрос относительно границ их применимости в системах речевого управления.
5. Используемые в экспериментах раздела 4.1 нейросети некорректно определены: сеть, использованная для получения данных в таблицах 26 - 29 именована как сеть с одним скрытым слоем, на самом деле это сеть без скрытых слоев. Сеть, использованная для расчета таблиц 30-32 содержит всего один (а не два, как в тексте) скрытых слоя.
6. В разделе 4.4. не приведено определение величины отношения сигнал-шум (ОСШ или SNR), которым пользуется автор. Необоснован выбор фиксированного значения ОСШ = +4 для тестов: при исследовании методов распознавания речи в шумах диапазон ОСШ выбирается так, что включаются отрицательные значения, например от -10 дБ до +15дБ.
7. Из текста непонятно почему в таблицах 30,31 и 44 для выбора оптимального количества итераций при обучении минимальные значения ошибок соответствует другим соотношениям параметров (большему числу итераций), чем фактически выбранные для дальнейшего использования.

