Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.934

Соколович
Максим Геннадьевич

Алгоритмы распознавания диктора для речевых интерфейсов интеллектуальных систем

АВТОРЕФЕРАТ
на соискание степени магистра наук
по специальности 7-06-0611-03 – Искусственный интеллект

Научный руководитель
Захарьев Вадим Анатольевич
кандидат технических наук, доцент

Минск 2026

Краткое введение

Речевой интерфейс является одним из наиболее естественных способов взаимодействия человека с интеллектуальной системой. Его применение особенно важно в условиях, когда пользователь не может постоянно обращаться к клавиатуре, экрану или другому ручному средству управления. Однако для надежного интеллектуального интерфейса недостаточно определить только содержание произнесенной фразы. Система должна также учитывать, кто произнес команду, поскольку ошибка в задаче распознавания диктора может привести к выполнению действия от имени другого пользователя.

Актуальность темы определяется необходимостью разработки алгоритмов, которые позволяют совместно анализировать содержание речевой команды и индивидуальные акустические признаки голоса. Речевой сигнал является нестационарным, зависит от темпа произнесения, микрофона, шума, канала записи и индивидуальных особенностей диктора. Поэтому задача распознавания диктора для речевых интерфейсов требует не только выбора информативных признаков, но и построения устойчивых эталонов, корректного экспериментального протокола и проверки работоспособности методов на открытых речевых данных.

Общая характеристика работы

Цель и задачи проводимых исследований

Целью диссертационной работы является разработка и экспериментальное исследование алгоритмов распознавания диктора для речевых интерфейсов интеллектуальных систем при совместном учете задачи распознавания речевых команд.

Для достижения поставленной цели необходимо решить следующие задачи:

– проанализировать область применения речевых интерфейсов интеллектуальных систем и определить требования к алгоритмам распознавания диктора;

– исследовать методы параметризации речевых сигналов, включая спектральные, кепстральные и статистические признаки;

– разработать алгоритмическую схему формирования эталонов диктора и речевой команды на основе классических методов цифровой обработки речи;

– сформировать воспроизводимый экспериментальный протокол с контролем разбиения данных, отсутствия утечек и корректным расчетом метрик качества;

– провести экспериментальную оценку эффективности предложенных методов и алгоритмов на открытых русскоязычных речевых данных;

– исследовать применение компактной сверточной нейронной сети к log-Mel-представлениям речи и сопоставить ее с классическими эталонными методами.

Объектом исследования являются речевые сигналы пользователей интеллектуальных систем.

Предметом исследования являются методы и алгоритмы распознавания диктора и речевых команд.

Связь работы с приоритетными направлениями исследований и запросами реального сектора экономики

Тема диссертации соответствует приоритетному направлению «Цифровые информационно-коммуникационные и междисциплинарные технологии, основанные на них производства» согласно перечню приоритетных направлений научной, научно-технической и инновационной деятельности на 2021–2025 годы, утвержденному Указом Президента Республики Беларусь от 07.05.2020 № 156. Актуальность темы сохраняется и для периода 2026–2030 годов, поскольку Указом Президента Республики Беларусь от 01.04.2025 № 135 среди приоритетов научной, научно-технической и инновационной деятельности определены цифровые технологии и искусственный интеллект.

Диссертационное исследование выполнено на базе учреждения образования «Белорусский государственный университет информатики и радиоэлектроники». Полученные результаты ориентированы на разработку речевых интерфейсов интеллектуальных систем, персонализированных сервисов, систем голосового доступа и программных комплексов, в которых требуется одновременно определить содержание команды и принадлежность голоса допустимому пользователю.

Личный вклад

Диссертационное исследование является квалификационной научной работой, выполненной соискателем самостоятельно на основе изучения отечественной и иностранной научно-технической литературы, анализа

открытых речевых корпусов и проведения вычислительных экспериментов. Личный вклад заключается в:

– систематизации современных подходов к построению речевых интерфейсов и распознаванию диктора;

– разработке математической схемы формирования эталонов речевых сигналов на основе MFCC/log-Mel-признаков, DTW, PCA и полиномов Чебышева;

– подготовке воспроизводимого manifest-описания данных и программного контура экспериментов на Python;

– проведении экспериментальной оценки классических методов распознавания диктора и речевых команд;

– реализации и анализе компактной CNN-модели для распознавания диктора по log-Mel-представлениям;

– подготовке таблиц, графиков и выводов, основанных на сохраненных экспериментальных артефактах.

Основные выводы, теоретические положения и практические разработки принадлежат автору диссертации и составляют содержание данной работы.

Опубликование результатов диссертации

По материалам выполненных исследований опубликованы 2 научные статьи. Обе публикации прошли рецензирование и отражают результаты исследования методов анализа спектральных признаков речевых сигналов и устойчивости алгоритмов распознавания к акустическим искажениям.

Апробирование результатов диссертации

Основные результаты диссертационной работы докладывались и обсуждались на I Республиканской конференции молодых ученых в рамках работы «Оценка устойчивости CNN-классификатора голосовых команд к акустическим искажениям на основе анализа спектральных признаков».

Результаты работы могут быть использованы при разработке прототипов голосового доступа, персонализированных речевых команд и модулей дикторской идентификации в интеллектуальных информационных системах.

Краткое содержание работы

В первой главе выполнен обзор подходов к формированию речевых интерфейсов интеллектуальных систем. Рассмотрены области применения речевого управления, особенности речевого сигнала как нестационарного акустического процесса, методы параметризации речи, включая MFCC, log-Mel, LPC и спектральные представления. Также проанализированы основные методы автоматического распознавания речи и диктора: сравнение с эталоном, скрытые марковские модели, статистические модели, i-vector/x-vector-представления и нейросетевые подходы.

Во второй главе сформирована математическая база исследования. Описаны алгоритмы предварительной обработки речевого сигнала, построения признакового пространства, формирования эталонов диктора и речевой команды, динамического выравнивания временных последовательностей, снижения размерности методом главных компонент, аппроксимации траекторий полиномами Чебышева, байесовского объединения решений и комитетов классификаторов. Отдельное внимание уделено совместной постановке задач «кто говорит» и «что сказано».

В третьей главе приведена экспериментальная оценка эффективности предложенных методов и алгоритмов. Описана методика проведения экспериментов, состав открытых русскоязычных данных, правила разбиения на train, dev и test, контроль отсутствия утечки данных, метрики accuracy, macro-F1, FAR, FRR и EER. Выполнено сравнение классических методов, анализ влияния числа обучающих записей, исследование шумовой устойчивости и проверка ограниченного proxy-режима совместного распознавания диктора и команды.

В четвертой главе рассмотрен нейросетевой подход к распознаванию диктора на основе log-Mel-представлений и компактной сверточной нейронной сети. Описаны подготовка входных признаков, архитектура модели, принципы обучения и регуляризации, а также сопоставление CNN с классическими эталонными алгоритмами. Показано, что на малом корпусе интерпретируемые классические методы сохраняют преимущество, тогда как CNN требует большего объема обучающих данных и дополнительной настройки.

Заключение

Основные результаты работы заключаются в следующем:

– на основе анализа области применения речевых интерфейсов показано, что для интеллектуальных систем важна совместная постановка задач распознавания содержания команды и принадлежности голоса конкретному диктору;

– разработана математическая схема формирования эталонов речевых сигналов, включающая предварительную обработку, вычисление спектральных признаков, нормализацию, динамическое выравнивание, снижение размерности и полиномиальную аппроксимацию;

– сформирован воспроизводимый экспериментальный протокол на открытых русскоязычных данных с использованием анонимных идентификаторов дикторов, фиксированных разбиений train/dev/test и контроля отсутствия утечки данных;

– экспериментальная оценка показала, что на подготовленном поднаборе Common Voice RU v22 наилучшее качество closed-set распознавания диктора среди классических методов обеспечивают эталоны на основе полиномов Чебышева с accuracy 0.966102;

– установлено, что командный эксперимент на ограниченном словаре имеет диагностический характер из-за малого объема данных и неоднородности классов, поэтому его результаты используются для оценки работоспособности программного контура, а не как финальная промышленная метрика;

– исследована компактная CNN-модель на log-Mel-представлениях, которая на малом корпусе достигла accuracy 0.677966 и уступила лучшему классическому методу, что подтверждает необходимость большего объема обучающих данных для нейросетевого подхода;

– подготовлены программные скрипты, CSV-результаты, таблицы и графики, обеспечивающие воспроизводимость экспериментальной части диссертационной работы.

Список опубликованных работ

1–А. Соколович, М. Г. Оценка устойчивости CNN-классификатора голосовых команд к акустическим искажениям на основе анализа спектральных признаков / М. Г. Соколович // Материалы I Республиканской конференции молодых ученых. – Минск, 2026.

2–А. Соколович, М. Г. Алгоритмы распознавания диктора для речевых интерфейсов интеллектуальных систем / М. Г. Соколович. – 2026.