

ОТЗЫВ

официального оппонента на диссертационную работу Сагацияна Максима Владимировича на тему «Разработка и исследование коллективных нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов», представленную на соискание учёной степени кандидата технических наук по специальности 05.12.04 Радиотехника, в том числе системы и устройства телевидения

1. Актуальность проведённых исследований

В настоящее время растёт важность массового внедрения новых интерфейсов взаимодействия человека с техническими системами, поскольку традиционные интерфейсы во многом уже достигли своего совершенства, а вместе с ними и своих пределов. При традиционно высокой значимости информации, поступающей к нам через органы зрения, и ее высокой доле среди всей сенсорной информации, данный канал восприятия человека становится в значительной степени перегружен. И первоочередной альтернативой здесь видится коммуникация именно по акустическому каналу. Преимущества речевого диалога перед традиционными средствами общения исследованы достаточно давно и многократно описаны в литературе. Привлекательность речевого общения, достижения в области распознавания речи, а также сопутствующее им развитие и удешевление электронно-вычислительной и микропроцессорной техники в последнее время привели к тому, что сфера внедрения систем распознавания речи существенно расширяется, захватывая различные отрасли производственной, административной, и даже бытовой деятельности.

Учитывая вышесказанное, можно заключить, что поставленная в диссертации задача разработки и исследования коллективных нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов является актуальной.

2. Характеристика содержания работы

Во введении обосновывается актуальность темы исследования, определяются предмет, цель и задачи исследования, методы исследования, научная новизна, практическая ценность работы, достоверность полученных результатов, приводятся сведения об апробации работы и внедрении результатов исследования, личном вкладе соискателя, указываются положения, выносимые на защиту.

В первой главе проведён обзор существующих систем распознавания речи. Предпринята попытка спроектировать систему распознавания речевых сигналов, максимально приближенную к человеческой слуховой системе, так как человеческие возможности воспринимать и обрабатывать речевые сигналы превосходят имеющиеся на сегодняшний момент технические возможности их распознавания.

Показана необходимость использования представления речевого сигнала в виде мел-частотных кепстральных коэффициентов. Данное представление приближено к восприятию речевых сигналов слуховым анализатором человеческого мозга. Также выбран инструмент распознавания – искусственные нейронные сети, так как данный метод в настоящее время является перспективным и в определенной степени приближен к принципу работы человеческой нервной системы.

Во второй главе разработаны и исследованы нейросетевой, коллективный нейросетевой и модифицированный коллективный нейросетевой алгоритмы дикторонезависимого распознавания речевых сигналов.

Для анализа данных алгоритмов созданы две речевые базы «КРИПТОН-01» и «КРИПТОН-02». База «КРИПТОН-01» содержит 10 видов речевых сигналов, а «КРИПТОН-02» – 102 вида.

При разработке коллективного нейросетевого алгоритма предложено объединить отдельные нейросетевые распознаватели в единую систему по

принципам коллективного равноправного голосования методом bagging. Проведены исследования по определению наилучших параметров для нейросетевого bagging-коллектива дикторонезависимого распознавания русскоязычных речевых сигналов.

Для увеличения технических возможностей распознавания речевых сигналов коллективного нейросетевого алгоритма предложена модификация bagging-алгоритма. Она позволяет увеличить размер словаря без потери качества дикторонезависимого распознавания речевых сигналов.

В третьей главе исследованы разработанные коллективные нейросетевые алгоритмы с заданными наилучшими параметрами с разными алгоритмами обучения. Рассмотрено три коллективных нейросетевых алгоритма: bagging-коллектив 12-слойных перцептронов на основе обучения Левенберга-Марквардта; bagging-коллектив 12-слойных сетей Эльмана на основе обучения GDX и bagging-коллектив 12-слойных перцептронов на основе обучения SCG.

Для оценки работы алгоритмов обучения нейронных сетей выбраны критерии среднего квадрата ошибки и вероятности распознавания речевых сигналов. Лучшие результаты распознавания 10 речевых сигналов показал 12-слойный перцептрон на основе обучения SCG, а худшие – 12-слойный перцептрон на основе обучения Левенберга-Марквардта.

В четвертой главе предпринята попытка адаптировать разработанные алгоритмы к условиям их тестирования в зашумленной среде. Предложено модифицировать алгоритмы путем введения блока шумоподавления. Выбрано три алгоритма шумоподавления: два алгоритма на основе бинарных масок и один – на основе винеровской фильтрации.

Применение блоков шумоподавления заметно повысило приспособленность коллективных и модифицированных коллективных нейросетевых алгоритмов в условиях шумов.

3. Научная новизна

Научная новизна исследований определяется разработкой:

1. Нейросетевого алгоритма bagging-коллектива на основе персептронов Розенблатта с обучением масштабируемых сопряженных градиентов (Scaled Conjugate Gradient Backpropagation, SCG) позволяющего решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов для малого словаря с вероятностью распознавания 97,1 %, что на 4,1 процентных пункта выше существующих результатов.

2. Модифицированного коллективного нейросетевого алгоритма на основе персептронов Розенблатта с обучением SCG, позволяющего решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов для большого словаря с вероятностью распознавания 95,7 %, что на 5,29 процентных пункта выше существующих результатов.

4. Практическая значимость работы

Практическая ценность проведённого исследования заключается в предложении модификации коллективного нейросетевого алгоритма, расширяющего возможности нейросетевых алгоритмов дикторонезависимого распознавания русскоязычных речевых сигналов.

5. Рекомендации по использованию результатов диссертации

Полученные в ходе разработки и исследования результаты рекомендуется использовать для модернизации различных радиотехнических устройств с речевыми интерфейсами.

6. Обоснованность и достоверность научных положений и выводов

Достоверность полученных автором результатов подтверждается разработкой и применением корректного математического аппарата нейронных сетей и их модификаций, а также хорошим согласованием результатов экспериментальных исследований с теоретическими.

7. Автореферат и публикации

Автореферат достаточно полно отражает содержание диссертации. Основные результаты диссертационной работы отражены в 17 публикациях, среди которых имеются 3 публикации в изданиях, рекомендованных ВАК.

Автореферат соответствует содержанию диссертации, а диссертация — автореферату.

8. Замечания по работе

1. Для обучения и тестирования нейросетевых алгоритмов применялись речевые базы с ограниченной размерностью словаря. Не произведено исследования разработанных нейросетевых алгоритмов при словаре более 102 видов речевых сигналов.

2. Не во всех исследованиях разработанных нейросетевых алгоритмов оценены доверительные интервалы.

3. Мало описано характеристик дикторов, взятых для экспериментов.

4. Не оценена работа нейросетевых алгоритмов в условиях шумов, кроме белого гауссовского шума.

5. Исследуется мало алгоритмов шумоподавления.

6. Не исследованы другие известные речевые базы, которые записаны с участием большого количества дикторов.

7. Разработанные алгоритмы не модернизированы для возможности распознавания слитной речи, а только отдельных слов.

Указанные недостатки, в целом, существенно не влияют на качество полученных в работе результатов.

9. Выводы

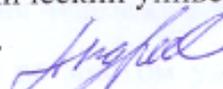
Диссертация «Разработка и исследование коллективных нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов» является законченной научно-квалификационной работой, посвященной решению

актуальной задачи создания речевых интерфейсов для различных устройств, в том числе радиотехнических. Она удовлетворяет требованиям «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям, а ее автор Сагациян Максим Владимирович заслуживает присуждения ученой степени кандидата технических наук по специальности 05.12.04 «Радиотехника, в том числе системы и устройства телевидения».

Официальный оппонент

профессор кафедры радиотехнических систем

Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Рязанский государственный радиотехнический университет»,

д-р техн. наук, доцент  Владимир Григорьевич Андреев

18.09.2015 г.

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Рязанский государственный радиотехнический университет» (ФГБОУ ВПО «РГРТУ», РГРТУ);

адрес: 390 005, г. Рязань, ул. Гагарина, д. 59/1;

тел. +7 (4912) 46-03-59, e-mail: andrejev.v.g@rsreu.ru

Подпись Владимира Григорьевича Андреева заверяю.

Учёный секретарь Учёного совета РГРТУ



В.Н. Пржегорлинский