

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

УДК _____

Жук

Иван Николаевич

Система распознавания речевых команд на основе нейронной сети для
«умного дома»

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-40.80.01 Элементы и устройства вычислительной
техники и систем управления

Научный руководитель

Азаров Илья Сергеевич

к.т.н., доцент

Минск 2015

ВВЕДЕНИЕ

Когда Кристофер Лэтем Шоулз разработал планировку клавиатуры QWERTY в 1860-ых (часто предполагалось, что данная планировка будет уменьшать производительность машинистов), немногие могли себе представить, что данная инновация станет непревзойденным методом ввода в 20-м веке. В начале 21-го века (так называемом веке информации) данная технология ввода до сих пор лидирует, несмотря на многие, объективно лучшие, разработанные методы ввода. Бесспорно, пришло время рассмотреть альтернативы, в частности самый естественный метод человеческой коммуникации — разговорный язык.

Разговорный язык не только естественный, но во многих случаях и более быстрый метод, нежели печатный или с помощью мыши. Так же он индифферентен ко времени и расположению, тогда как клавиатуры и мышь могут быть неудобными. В частности, в мире в растущем количестве встраиваемой техники, так называемый «умные дома» могут стать первым массовым внедрением систем голосового взаимодействия.

К тому же, большинство лежащих в основе технологий уже существует — различная бытовая техника, электронные устройства, мультимедиа системы, сенсоры и другие приборы достаточно умны для объединения в сеть. Домашние беспроводные сети очень быстрые и широко распространены. Системы синтеза речи могут синтезировать практически естественную речь. Так же широко распространены технологии записи и воспроизведения звукового сигнала. Современные компьютеры высоконадежные, относительно дешевые и встроены практически во всю бытовую технику в качестве встроенных систем. Однако камнем преткновения реализации подобных систем для умного дома является автоматическое распознавание речи (АРР) и понимание естественных языковых оборотов.

В данной работе исследуются АРР системы использующие в своей основе рекуррентные нейронные сети, как наиболее естественный способ представления длинных последовательностей, с имеющимися временными зависимостями. Так же представлен анализ определенных операционных условий работы системы, в частности определенной структуры команды, и ограниченного словаря.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Голосовые интерфейсы распознавания речи обретают все большую популярность, как в академических, так и в коммерческих отраслях. Это связано с последними достижениями в данной области, которые можно воочию увидеть в нашей повседневной жизни: голосовые интерфейсы в компьютерах, роботах, сотовых телефонах, в интернет-браузерах, и даже в бытовой технике. Однако проблема до сих пор остается нерешенной.

Исследовательская часть настоящей диссертации посвящена распознаванию и классификации голоса и речи в бытовых условиях повседневной жизнедеятельности.

Целью и задачами работы являются исследование особого вида нейронных сетей — сетей с обратными связями (рекуррентных) в контексте голосового интерфейса в условиях умного дома, изучение различных архитектурных решений организации нейронных связей, исследование и реализация алгоритмов обучения подобных сетей, обучение сети на аудио-данных из соответствующих условий и сравнение результатов с альтернативным подходами.

Практическая ценность полученных результатов работы заключается в следующем:

- предложенная архитектура распознавателя, может успешно справляться с бытовыми односложными командами в условиях зашумлённости
- выявлены наиболее ценные с точки зрения классификации подходы и методики, что может быть полезным при обучении других голосовых интерфейсов с русским языком.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 51-ой научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2015).

Список опубликованных работ

По теме диссертации опубликовано 1 печатная работ – статья в сборнике материалов 51-ой научной конференции аспирантов, магистрантов и студентов БГУИР.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе описывается исследуемая проблема и поставленные задачи, а так же производится обзор основных подходов и методик решения этих задач. Описывается специфика проблемы в контексте бытовой жизнедеятельности и актуальность данной проблемы..

Во второй главе рассматриваются основные подходы к решению задач распознавания и классификации — скрытые Марковские модели и различные виды нейронных сетей. Аргументируется выбор рекуррентных нейронных сетей для рассматриваемой проблемы.

В третьей главе рассматривается архитектура предлагаемого распознавателя на основе нейронных сетей с обратными связями, а так же приводятся различные алгоритмы обучения данных сетей.

В четвертой главе устанавливаются условия экспериментов, оговариваются методы предобработки звука. После чего приводятся экспериментальные результаты обучения на реальных аудио-данных для ключевого события, которое сигнализирует о начале команды и для самих команд, производятся выводы из результатов обучения.

В заключении приведены общие выводы и результаты по всей диссертации вместе с дальнейшими научными направлениями, которые представляют определенный интерес для автора и которые могут представлять интерес для научного сообщества в целом.

ЗАКЛЮЧЕНИЕ

В данной работе были рассмотрены основные подходы к распознаванию и пониманию речи и связанные с ними алгоритмы и модели. В частности были рассмотрены математические модели нейронных сетей и их прикладные аспекты. Особенным видом нейронных сетей, являются рекуррентные нейронные сети, которые являются центром данной работы.

Нейронные сети с обратными связями хорошо зарекомендовали себя в различных областях науки, в том числе и в обработке и распознавании речи. Полученные мною результаты подтверждают актуальность их прикладной эффективности. С помощью предложенной мною нейронной сети с обратными связями, обученной алгоритмом Левенберга-Марквардта удалось достичь результата выявления ключевого события в 90%, и средней точности классификации в 87.5%. Данные результаты позволяют сделать вывод о возможности практического применения и внедрения подобных систем.

На данном этапе развития, технологии речевого управления развиты достаточно, что бы сравниться в классическими подходами ввода информации в такой области как управление умным домом. Однако недостаточно развиты для полной без компромиссной замены клавиатуры и мыши.