



ОТКРЫТАЯ НАУКА
издательство

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.942

ВОСПРИЯТИЕ ЧЕЛОВЕЧЕСКОГО ГОЛОСА ПРИ ПОМОЩИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

¹Некрасов Т.Д., ²Комбаров В.Д., Лозница С.Ю.

ФГБОУ ВО "САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ГРАЖДАНСКОЙ АВИАЦИИ ИМЕНИ ГЛАВНОГО МАРШАЛА АВИАЦИИ А.А. НОВИКОВА", Санкт-Петербург, Россия (196210, город Санкт-Петербург, ул. Пилотов, д.38), e-mail: ¹Kvakolka885@gmail.com, ²vlad54295@gmail.com

В статье рассматриваются принципы и технические возможности применения методов Фурье-анализа для обработки звуковых сигналов в ИИ. В современном мире технологии искусственного интеллекта (ИИ) играют всё более важную роль в нашей жизни. Они используются в различных областях, от медицины до развлечений, и становятся неотъемлемой частью нашего общества. Одной из ключевых задач ИИ является восприятие и анализ человеческого голоса. В этой работе мы рассмотрим основные аспекты восприятия человеческим голосом искусственным интеллектом, а также его применение в различных сферах деятельности.

Ключевые слова: ИИ, метод, анализ, звук, преобразование, дискретизация, голос.

PERCEPTION OF THE HUMAN VOICE WITH THE HELP OF ARTIFICIAL INTELLIGENCE

¹Nekrasov T.D., ²Kombarov V.D., Loznitsa S.Yu.

"ST. PETERSBURG STATE UNIVERSITY OF CIVIL AVIATION NAMED AFTER AIR CHIEF MARSHAL A.A. NOVIKOV", St. Petersburg, Russia (196210, St. Petersburg, ул. Pilotov, д.38), e-mail: ¹Kvakolka885@gmail.com, ²vlad54295@gmail.com

The article discusses the principles and technical possibilities of using Fourier analysis methods for processing audio signals in AI. In the modern world, artificial intelligence (AI) technologies are playing an increasingly important role in our lives. They are used in various fields, from medicine to entertainment, and are becoming an integral part of our society. One of the key tasks of AI is the perception and analysis of the human voice. In this paper, we will look at the main aspects of the perception of the human voice by artificial intelligence, as well as its application in various fields of activity.

Keywords: AI, method, analysis, sound, transformation, sampling, voice.

Актуальность темы обусловлена растущим интересом к использованию технологий ИИ для улучшения качества жизни людей. Понимание того, как ИИ воспринимает человеческий голос, может привести к созданию более эффективных систем распознавания речи, улучшению коммуникации между человеком и машиной, а также разработке новых методов диагностики и лечения речевых нарушений.

Цель данной работы — изучить основные принципы и методы восприятия человеческим голосом искусственным интеллектом. Для достижения этой цели будут решены следующие задачи:

- Рассмотреть физиологические и психологические аспекты восприятия звука человеком;
- Изучить основные характеристики человеческого голоса, важные для восприятия;
- Описать алгоритмы и методы машинного обучения для распознавания речи;
- Проанализировать применение методов Фурье-анализа для обработки звуковых сигналов в ИИ;
- Привести примеры использования ИИ для анализа человеческого голоса;
- Провести экспериментальное исследование восприятия человеческого голоса ИИ.

Перейдем к рассмотрению физиологических и психологических аспектов восприятия звука человеком.

Звук — это колебательное движение частиц среды, которое распространяется в пространстве и воспринимается органами слуха человека. Восприятие звука включает в себя несколько этапов: преобразование звуковой волны в электрические сигналы, передачу этих сигналов в мозг и их обработку.

Человеческий слух способен воспринимать звуки в диапазоне от 20 Гц до 20 кГц. Этот диапазон охватывает большинство звуков, которые встречаются в повседневной жизни. Звуковые волны, попадающие в ухо, вызывают колебания барабанной перепонки, которые затем передаются на слуховые косточки и улитку внутреннего уха. Улитка преобразует звуковые волны в электрические импульсы, которые передаются в мозг через слуховой нерв.

Психологические аспекты восприятия звука включают в себя такие понятия, как высота тона, громкость, тембр и длительность. Высота тона определяется частотой звуковых колебаний, громкость — амплитудой этих колебаний, тембр — наличием обертонов, а длительность — продолжительностью звука.

Основные характеристики человеческого голоса, важные для восприятия

Основными характеристиками человеческого голоса, которые важны для восприятия, являются частота, амплитуда, спектр и тембр. Частота определяет высоту тона голоса, амплитуда — его громкость, спектр — распределение частот в звуке, а тембр — его окраску.

Частота измеряется в герцах (Гц) и определяет количество колебаний звуковой волны в секунду. Чем больше частота, тем выше тон голоса. Амплитуда измеряется в децибелах (дБ) и определяет громкость звука. Спектр представляет собой график распределения частот в звуке. Тембр определяется наличием обертонов — дополнительных частот, которые придают голосу его уникальную окраску.

Для восприятия человеческого голоса важны такие характеристики, как чёткость произношения, интонация, ритм и темп речи. Чёткость произношения определяет ясность и понятность речи, интонация — эмоциональную окраску голоса, ритм — чередование ударных и безударных слогов, а темп речи — скорость произнесения слов.

Рассмотрим вопрос восприятие человеческого голоса с помощью искусственного интеллекта.

Алгоритмы машинного обучения позволяют ИИ распознавать и анализировать человеческий голос[2]. Существует несколько основных подходов к распознаванию речи:

- Распознавание по шаблонам: сравнение входного сигнала с заранее записанными шаблонами;

- Скрытые марковские модели: использование статистических моделей для описания последовательностей событий;
- Нейронные сети: обучение на больших объёмах данных для выявления закономерностей и зависимостей.

Нейронные сети являются наиболее перспективным подходом к распознаванию речи. Они способны обучаться на больших объёмах данных и выявлять сложные закономерности, что позволяет им достигать высокой точности распознавания.

Часто для обработки звуковых сигналов в ИИ используется метод Фурье-анализа. Методы Фурье-анализа позволяют преобразовать звуковые сигналы из временной области в частотную. Это позволяет выделить основные частоты в сигнале и определить его спектральный состав [3].

Применение методов Фурье-анализа в ИИ позволяет улучшить качество распознавания речи за счёт выделения ключевых характеристик голоса. Например, можно использовать методы Фурье-анализа для определения высоты тона голоса, его громкости и тембра.

Суть метода Фурье заключается в следующем [1-3]:

1. Функция $f(x)$, определённая на некотором интервале $[a, b]$, представляется в виде суммы бесконечного числа синусов и косинусов с различными частотами и амплитудами. Это представление называется рядом Фурье.

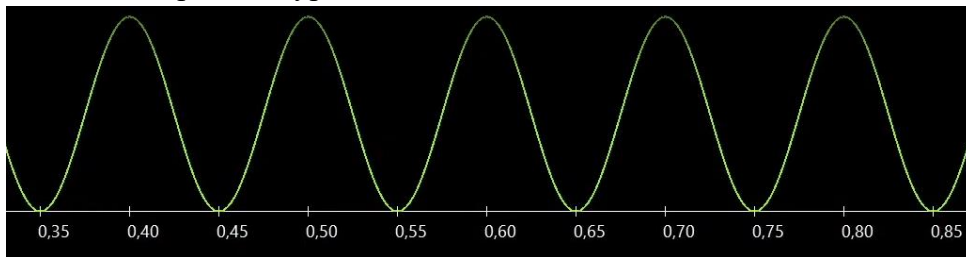


Рисунок 1 – Ряд Фурье.

2. Коэффициенты ряда Фурье вычисляются через интегралы от функции $f(x)$. Они зависят от амплитуды и фазы каждой синусоиды в разложении.

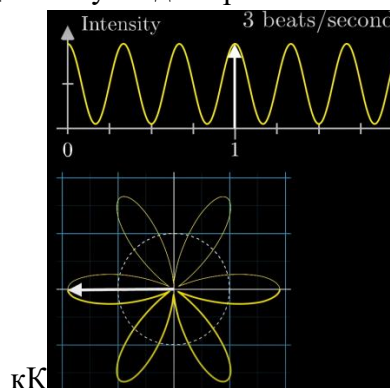


Рисунок 2 - Коэффициенты ряда Фурье.

3. После вычисления коэффициентов ряда Фурье мы получаем функцию, которая описывает зависимость интенсивности от времени. Данная функция не имеет времени в качестве входных данных, а вместо этого принимает частоту, которая и является частотой намотки. Выходные данные этой функцию представляют собой комплексное число, некоторую точку на плоскости (центр масс воображаемой намотки).

$$g(f) = \int_{+\infty}^{-\infty} g(t)e^{-2\pi ift} dt$$

4. Метод Фурье позволяет анализировать сложные функции, разлагая их на более простые синусоидальные составляющие. Это упрощает анализ и позволяет выявить закономерности в данных

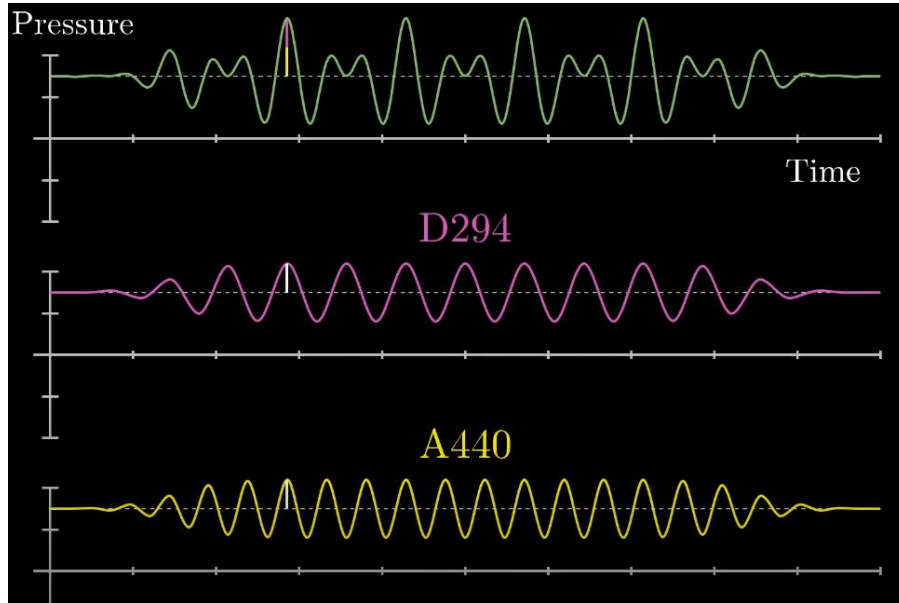


Рисунок 3 - Синусоидальные составляющие.

5. В контексте обработки сигналов и изображений метод Фурье используется для анализа частотных составляющих сигнала или изображения. Это позволяет выделить важные частоты и удалить шум из данных.

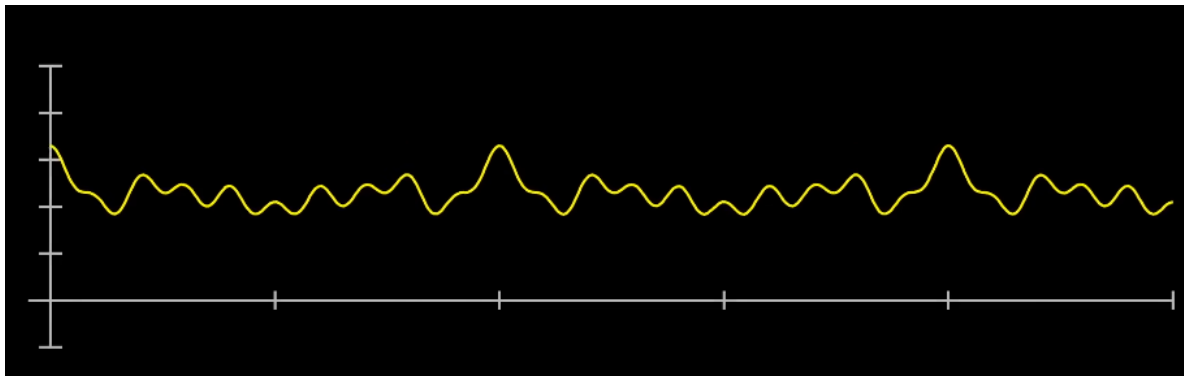


Рисунок 4 – Выделение частот и удаление шума из данных.

6. Преобразование Фурье переводит функцию из временной области в частотную. Оно показывает, какие частоты присутствуют в сигнале и с какой амплитудой они проявляются.

7. Обратное преобразование Фурье возвращает нас из частотной области обратно во временную, восстанавливая исходный сигнал.

8. Быстрое преобразование Фурье (БПФ) — это алгоритм, который позволяет вычислить коэффициенты ряда Фурье гораздо быстрее, чем при использовании прямого

подхода. БПФ широко используется в различных областях, таких как обработка сигналов, спектральный анализ, сжатие данных и т.д.

Далее рассмотрим примеры использования ИИ для анализа человеческого голоса.

Управление воздушным движением. Системы распознавания голоса могут быть использованы для управления воздушным движением, позволяя диспетчерам отдавать команды пилотам и получать от них информацию с помощью голосовых команд. Это может упростить процесс коммуникации и снизить нагрузку на диспетчеров.

Автоматизация рутинных операций. В авиации существует множество рутинных задач, которые можно автоматизировать с помощью распознавания голоса. Например, системы распознавания могут использоваться для автоматического заполнения форм, составления отчетов и других административных задач.

Обучение пилотов. Системы распознавания голоса также могут быть полезны для обучения пилотов. Они могут использоваться для оценки произношения команд, понимания акцентов и интонаций, а также для выявления ошибок в речи пилотов.

Обеспечение безопасности полётов. Системы распознавания голоса могут помочь обеспечить безопасность полётов, обнаруживая необычные или подозрительные звуки, такие как крики, стоны или другие звуки, которые могут указывать на проблемы с оборудованием или здоровьем экипажа.

Улучшение коммуникации между экипажем. Распознавание голоса может улучшить коммуникацию между членами экипажа, позволяя им быстро и точно обмениваться информацией во время полёта. Это особенно полезно в условиях стресса или ограниченной видимости.

Помощь в навигации. Системы распознавания голоса могут использоваться для помощи пилотам в навигации, предоставляя им информацию о высоте, скорости, курсе и других параметрах полёта.

Диагностика оборудования. Распознавание голоса также может быть использовано для диагностики оборудования, позволяя пилотам быстро выявлять и устранять неисправности.

Технологии искусственного интеллекта уже достигли значительных успехов в области распознавания и анализа человеческого голоса. Однако существуют определённые ограничения и проблемы, связанные с точностью и эффективностью этих технологий. Для этого мы решили изучить эту тему и найти пробелы в совершенстве распознавания голоса. Но мы также должны осознавать необходимость разработки этических принципов и норм, регулирующих использование технологий искусственного интеллекта для восприятия человеческого голоса. Необходимо учитывать возможные риски и негативные последствия, связанные с нарушением конфиденциальности, дискриминацией и другими проблемами. Важно подчеркнуть развитие данной технологии для использования ее в целях обеспечения безопасности, как и в домашних условиях, так и в рабочих.

Список литературы

1. Латыпова Н.В., Тучинский Л.И. (2011). РЯДЫ ФУРЬЕ (Ижевск).
2. А. Х. Шахмейстер(2014). Комплексные Числа (Москва).
3. К. Н. Гурьянова, У. А. Алексеева, В. В. Бояршинов (2014). Математический Анализ (Екатеринбург).

References

1. Latypova N.V., Tuchinsky L.I. (2011). *FOURIER SERIES* (Izhevsk).
 2. A. H. Shakhmeister (2014). *Complex Numbers* (Moscow).
 3. K. N. Guryanova, U. A. Alekseeva, V. V. Boyarshinov (2014). *Mathematical Analysis* (Yekaterinburg).
-