

Московский государственный технический университет имени Н. Э. Баумана
(национальный исследовательский университет)

Научная исследовательская работа

«Классификация методов обнаружения образцов голоса, синтезированных с помощью нейронных сетей»

Студент: Ахмад Халид Каримзай ИУ7и-74Б

Научный руководитель: А.С. Кострицкий

Москва, 2023 г.

Цель и задачи работы

Цель - В рамках данной научной-исследовательской работы, было рассмотрено следующие цели и задачи:

Задачи:

- Синтезирование аудио: Понятие и Типы;
- Характеристики и особенности аудиоматериала для изучения;
- Понятие и схема работы системы обнаружения синтетического звука;
- Классификация и обзор синтезированного звука.

Понятие синтезированного голоса

Под термином "Синтезирование голоса" обычно понимается любой аудио-сигнал, важные характеристики которого были изменены при помощи технологий нейронных сетей, сохраняя при этом воспринимаемую естественность.

- преобразование текста в речь;
- преобразование голоса;
- подделка эмоций;
- подделка сцен;
- частично подделка.

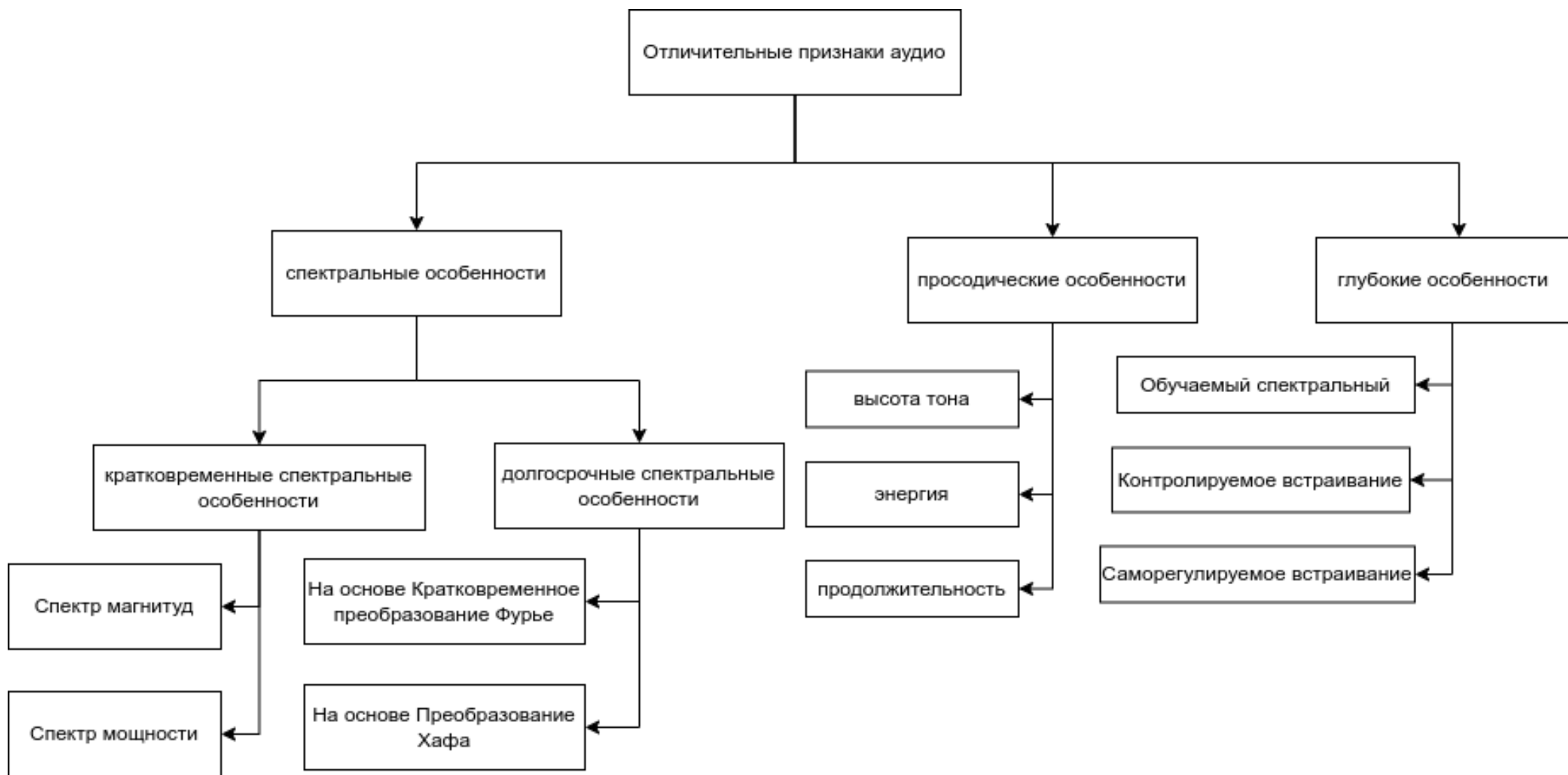
<i>Поддельный тип</i>	<i>Поддельная черта</i>	<i>Поддельная продолжительность</i>	<i>С помощью продолжительность нейронной сети</i>
Преобразование текста в речь	Личность спикера	полностью	да
Преобразование голоса	Личность спикера	полностью	да
Подделка эмоций	эмоция спикера	полностью	да
Подделка сцен	Акустическая сцена	полностью	да
Частично подделка	Речевое содержание	частично	да

Признаки аудио для изучения

Извлечение признаков представляет собой ключевой модуль классификатора аудио-дипфейков. Основной целью этого процесса является изучение характерных особенностей путем выделения акустических артефактов из речевых сигналов, которые могут свидетельствовать о наличии поддельных атак. Большое количество исследований подчеркнуло важность определения полезных признаков для эффективного обнаружения дипфейков.

- Спектральные характеристики;
- Просодические характеристики;
- Глубокие характеристики.

Классификации аудио по признаками



Система обнаружения поддельного звука

Основной задачей системы обнаружения поддельного звука (аудио дипфейк) является процесс выявления поддельного звука в речевом потоке. В качестве входных данных используются звуковые сигналы, а на выходе представляются результаты классификации, система в целом состоит из двух частей:

1. Модуль извлечения признаков.
2. Модуль классификации.

Модуль извлечения признаков

Модуль извлечения признаков должен извлекать соответствующую информацию из речевого сигнала, отражающую артефакты, связанные с процессом преобразования или синтеза. В основном, на этом этапе в большинстве случаев извлекаются спектральные характеристики.

Самые популярные спектральные характеристики, используемые в системе обнаружения синтезированного звука, являются:

1. Кепстральные коэффициенты мел-частоты (MFCC).
2. Линейные частотные кепстральные коэффициенты (LFCC).

Модель класифікації

В системах, використовуваних для виявлення піддельного звуку, важким фактором є параметричні характеристики аудіо для вивчення, і внутрішній класифікатор грає ключову роль в глибокому розпізнаванні аудіо, внутрішні класифікатори, які часто використовуються для виявлення аудіо-дипфейків, в основному діляться на дві категорії:

1. Статистичні методи.
2. С використанням глибоких нейронних мереж.

Статистические методы обнаружения Аудио Дипфейк

Машина опорных векторов (SVM)

$$H_1 = \{x \in \mathbb{R}^n \mid f(x) > 0\}$$

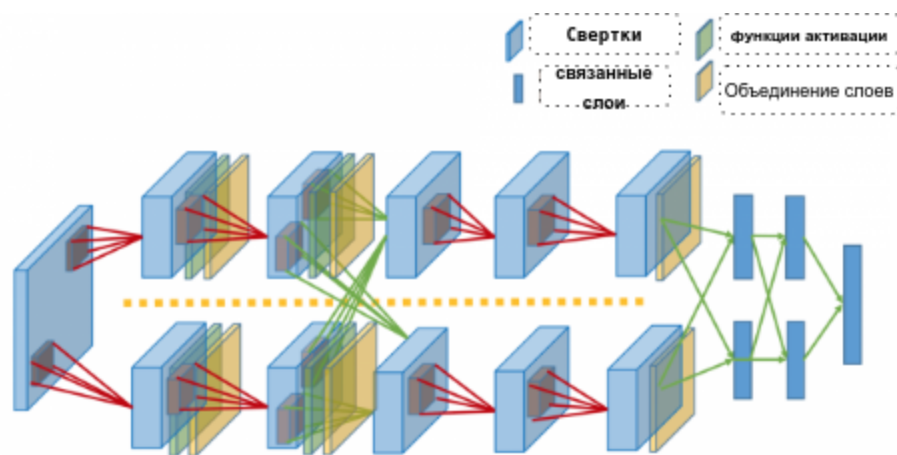
$$H_2 = \{x \in \mathbb{R}^n \mid f(x) < 0\}$$

Гауссовы модели смеси (GMM)

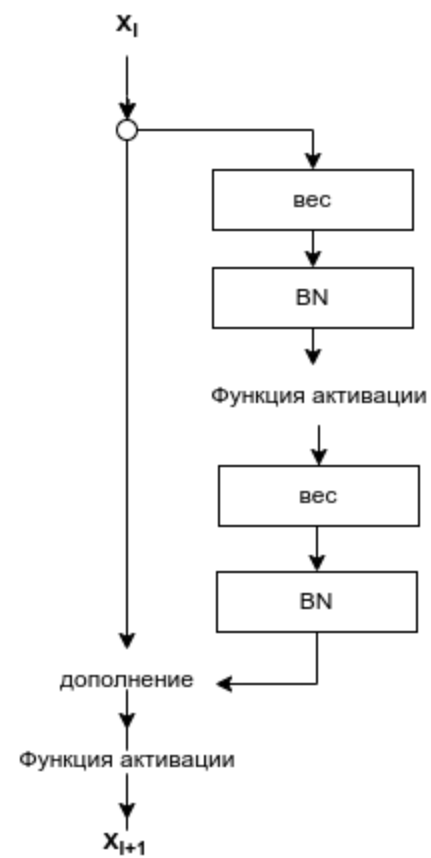
$$f_i^s(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i^s|}} \exp \left(-\frac{1}{2} (x - \mu_i^s)^T \Sigma_i^s (x - \mu_i^s) \right)$$

Методы класификации с применением глубоких нейронных сетей

Сверточная нейронная сеть (CNN)



Остаточная нейронная сеть (RNN)



Заключение

Обнаружение синтезированного звука - довольно сложная задача, и для успешного решения требуется значительное внимание. Как предложено во многих исследованиях по обнаружению синтетического звука, наилучшим вариантом классификации звука в настоящее время являются глубокие нейронные сети.

Структура работы системы обнаружения синтетического звука может быть описана следующим образом:

- На вход поступает аудиозапись.
- Модель извлечения признаков предварительно обрабатывает запись.
- Затем происходит процесс извлечения признаков.
- Модель классификации использует эти признаки для обучения и распознавания.