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.

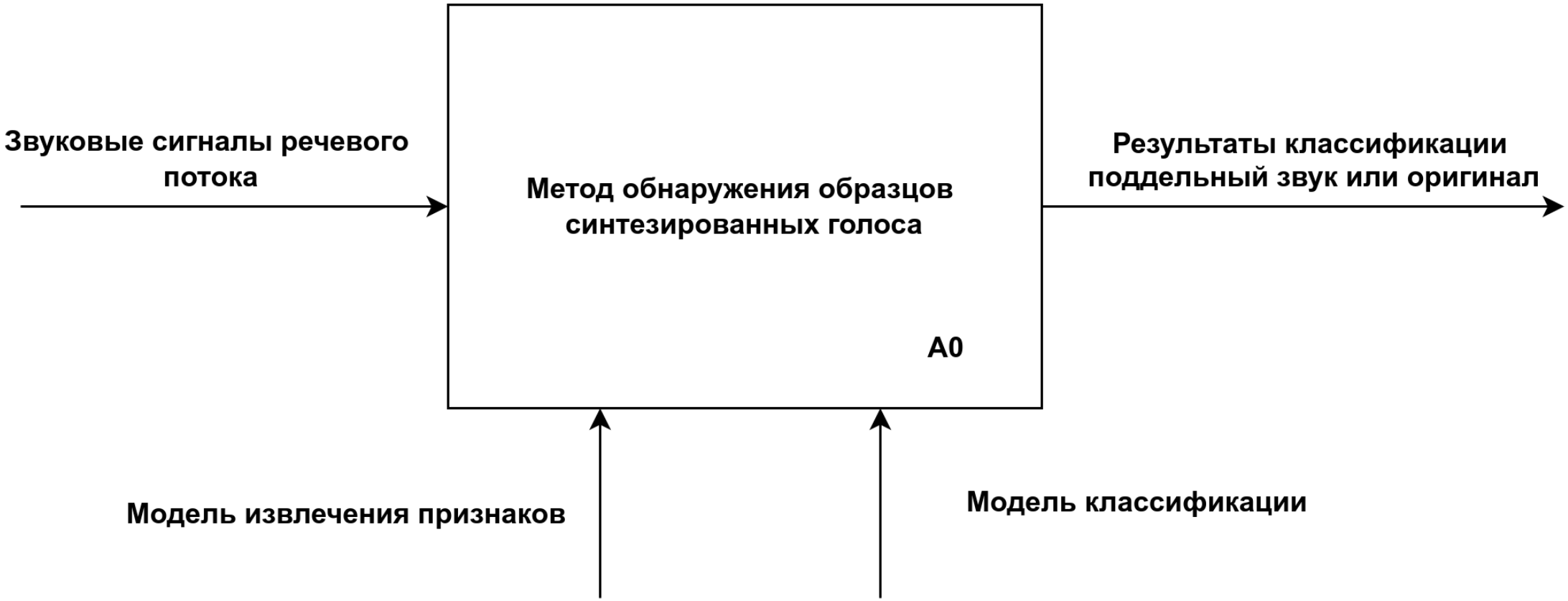
# Цель и задачи работы

**Цель** - В рамках данной научной-исследовательской работы, было расмотрено 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.

# Постановка задачи



# Анализ предметной области

Под термином ”Синтезирование голоса” обычно понимается любой аудио-сигнал, важные характеристики которого были изменены при помощи технологий нейронных сетей, сохраняя при этом воспринимаемую естественность, в проведенные исследования в основном выделяли пять видов дипфейкового звука:

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

*Классификации аудиодипфейков по способу генерации*

| *Поддельный тип* | *Поддельная черта* | *Поддельная продолжительность* | *С помощью нейронной сети* |
|---|---|---|---|
| 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 |

# Классификации аудио по признаками

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

# Система обнаружения поддельного звука

1. Сквозная система - в этом варианте система обнаружения поддельного звука, получает на вход речевой поток;
2. Комбинированная система - в этом варианте система обнаружения поддельного звука, состоит из двух модулей:
    - Модуль извлечения признаков;
    - Модуль классификации.

Существующие методы:

1. Метод с гениализацией признаков;
2. Метод с использованием интегрированного спектрально-временного подхода;
3. Метод с использованием трансферного обучения.

# Критерии сравнения методов

1. Равная частота ошибок (EER):

$$P_{\text{ложный}}(\theta) = \frac{\text{количество фальшивых голосов с партитурой} > \theta}{\text{полное количество фальшивых голосов}}$$

$$P_{\text{пропущенный}}(\theta) = \frac{\text{количество настоящих голосов со счетом} \leq \theta}{\text{полное количество истинных голосов}},$$

$$ERR = P_{\text{ложный}}(\theta) = P_{\text{пропущенный}}(\theta),$$

2. Функция затрат на тандемное обнаружение (mint – DCF):

$$mint - DCF = \min_{\theta}\{C_0 + C_1 P_{\text{пропущенный}}(\theta) + C_2 P_{\text{ложный}}(\theta)\},$$

# Классификация причисленных методов обнаружения синтетического звука

1. K1 - Точность обнаружение поддельной речи, для этой цели рассматриваем оценка ошибки $\mathrm{ERR}$ относителя корпус данных ASVSpoof, данное значение настолько меньше, настолько выше точность работы метода;
2. K2 - Устойчивость к различным типам поддельной речи, для этой цели рассматриваем оценку функция затрат на тандемное обнаружение (min t-DCF) относителя корпус данных ASVSpoof \cite{yamagishi2021asvspoof}, данное значение настолько меньше, настолько выше точность работы метода относительно различных видов синтезирования звука;
3. K3 - Принимает ли на вход аудиосигнал;
4. K4 - Требует ли обучение модель классификации.

*Классификация приведенных методов обнаружения синтетического звука*

| *Метод* | *К1* | *К2* | *К3* | *К4* |
|---|---|---|---|---|
| Метод с генерализацией признаков | 4.07% | 0.102 | Нет | Да |
| Метод с использованием трансферного обучения | 8.09% | 0.2116 | Нет | Да |
| Метод с использованием интегрированного спектрально-временного подхода | 0.83% | 0.0275 | Да | Да |

# Заключение

В рамках данной работы было проведено изучение системы обнаружения аудиодипфейков, анализ предметной области, рассмотрение признаков аудио для изучения и обучения, а также проведена классификация и обзор методов обнаружения аудиодипфейков.
В итоге можно описать структуру работы системы обнаружения синтетического звука следующим образом:

1. На вход поступает аудиозапись.
2. Модель извлечения признаков предварительно обрабатывает запись.
3. В некоторых методах модели используют признаки для обучения, а в некоторых других просто принимают звуковую речь.
4. Модель классификации использует признаки или саму звуковую речь для обучения и распознавания.