УДК 0004.934

МЕТОДЫ ГОЛОСОВОЙ БИОМЕТРИИ НА КОРОТКИХ ДЛИТЕЛЬНОСТЯХ В ТЕЛЕФОННОМ КАНАЛЕ

Волкова М.В. (Университет ИТМО), **Гусев А.Е. (Университет ИТМО)**
Научный руководитель – к.т.н., Новоселов С.А.
(Университет ИТМО)

В данной работе рассматривается задача верификации диктора на коротких произнесениях в телефонном канале. Анализируется применение глубоких нейронных сетей на базе архитектур TDNN и ResNet для построения голосовых моделей.

Несмотря на широкое распространение голосовых биометрических систем, верификация на коротких длительностях все еще остается трудной задачей, поскольку надежность таких систем значительно ухудшается с сокращением длительности произнесений. Однако верификация на коротких длительностях востребована в таких системах с короткими произнесениями, как интерактивные голосовые меню (IVR - Interactive Voice Response), что к тому же подразумевает телефонный канал передачи данных.

Задачи голосовой биометрии последние несколько лет успешно решаются с применением глубоких нейронных сетей. Наиболее популярным решением стали x-вектора на основе TDNN. В данной работе мы рассматриваем несколько архитектур на базе TDNN и ResNet для построения голосовых моделей и анализируем влияние адаптации под целевые условия IVR. В качестве входных признаков мы используем 23-мерные mfcc и 64-мерные fbanks.

Существует два основных сценария верификации на коротких длительностях. К первому относится случай, когда эталонная модель диктора строится по длинному аудиофайлу, но верификация происходит с коротким фрагментом речи (long - short). Второй сценарий подразумевает короткие длительности для построения как эталонной, так и тестовой модели голоса диктора (short - short) и является более трудным. В нашей работе тестирование проводилось в сценариях long-short и short-short на русскоязычной базе IVR, собранной в различных шумовых условиях, а также на базах Telecom и NIST SRE. Мы показали, что ResNet34, обученная на 64fbanks дает наилучшее качество в сценарии IVR.