

## ЧУВСТВИТЕЛЬНОСТЬ СОВРЕМЕННЫХ АЛГОРИТМОВ ВЕРИФИКАЦИИ ДИКТОРА К СТАРЕНИЮ ГОЛОСА: БАЗЫ ДАННЫХ ДЛЯ ТЕСТИРОВАНИЯ И ОБУЧЕНИЯ МОДЕЛЕЙ

Явшиц Е. В.<sup>1</sup>

Научный руководитель — Корневская А. М.<sup>1</sup>

<sup>1</sup>Университет ИТМО

[kate\\_yavshitz@mail.ru](mailto:kate_yavshitz@mail.ru)

### Введение

В современной биометрии голос человека является одним из ключевых признаков для верификации и идентификации. При разработке и валидации таких систем критическую роль играет качество и репрезентативность данных, на которых они обучаются и тестируются. Стандартные открытые базы (VoxCeleb, LibriSpeech), как правило, не содержат размеченной информации о возрасте дикторов и датах записи, что делает их малоприспособленными для исследования долговременной внутридикторской изменчивости.

В то же время, анализ физиологической литературы и клинических исследований показывает, что голос человека претерпевает естественные возрастные изменения — пресбицию [1]. Это создает противоречие: системы верификации, обучаемые на статичных данных, эксплуатируются в долгосрочных сценариях, при этом голос пользователя неизбежно меняется. Для изучения этого эффекта необходимы специализированные базы данных, содержащие записи одних и тех же дикторов с интервалами в несколько лет. Существующие зарубежные базы данных (например, VoxCeleb с доразметкой от сторонних исследователей) и проприетарный корпус ИТМО не закрывают полностью потребность в репрезентативных данных, что и определяет актуальность данного исследования, построенного на комбинации двух уникальных наборов данных.

### Основная часть.

В качестве решения проблемы дефицита релевантных данных предлагается не только поиск новых размеченных корпусов, но и методология тестирования, основанная на совместном использовании двух баз данных, обладающих взаимодополняющими характеристиками. Проприетарная база данных «Voice Biometrics Age» (НИУ ИТМО). Ее ключевая особенность — наличие для каждого диктора записей, сделанных с контролируемыми временными промежутками (от 1 до 5 лет). База сбалансирована по полу (50/50) и содержит, преимущественно, записи на русском языке. Это позволяет исследовать внутридикторскую изменчивость в «чистом» виде, минимизируя влияние языковых и канальных факторов. Также предполагается использование публичной базы данных VoxCeleb2 [2] с уточненными метаданными. Поскольку «сырой» VoxCeleb2 не содержит информации о возрасте, в работе использована открытая сторонняя разметка, позволившая определить возраст дикторов и годы записи для тестового подмножества (36 дикторов, 945 файлов). Это дает возможность сопоставить результаты с мировыми исследованиями и оценить влияние «реалистичного» шума и разнообразия каналов, присущих видеохостингам.

Суть экспериментального подхода заключается в анализе динамики скоринговых оценок системы верификации (модель биометрической верификации представлена архитектурой ResNet34) в зависимости от увеличения временного интервала между эталонной и тестовой записью внутри каждой из баз. Новизна метода состоит в кросс-валидации результатов на двух качественно разных наборах данных: «чистом» (лабораторные условия) и «зашумленном» (реалистичные условия интернета).

Экспериментальные результаты показали, что на обеих базах данных наблюдается устойчивая деградация качества верификации с ростом временного разрыва. При этом на более сложном и вариативном наборе VoxCeleb2 эффект падения точности (рост равновероятной ошибки) выражен значительно сильнее, чем на проприетарной базе. Это позволяет сделать важный вывод о том, что старение голоса и внешние шумовые факторы действуют синергетически, многократно усиливая негативный эффект.

### **Выводы**

Проведенное исследование, базирующееся на двух специализированных базах данных, позволяет сделать следующие выводы. Подтверждена критическая важность репрезентативных данных: Экспериментально доказано, что использование только «чистых» краткосрочных корпусов для валидации систем верификации некорректно. Для долгосрочных сценариев необходимы базы данных с временной разметкой. Выявлена специфика доступных датасетов: проприетарная база «Voice Biometrics Age» пригодна для изолированной оценки влияния фактора времени, в то время как VoxCeleb2 с доразметкой позволяет оценить устойчивость систем в условиях, приближенных к реальной эксплуатации. При этом выявлено, что даже молодые голоса (преимущественно представленные в базах) демонстрируют измеримую изменчивость за 3-5 лет. Обоснована необходимость пополнения баз: Деградация точности верификации на существующих данных указывает на необходимость создания новых, более крупных открытых корпусов, включающих записи пожилых дикторов и охватывающих более широкие временные промежутки (10+ лет).

### **Литература:**

1 Румянцева, В.В. Акустические показатели голоса в разные возрастные периоды у взрослых и их прогнозирование / В. В. Румянцева, О. С. Бестолкова // Российская оториноларингология. — 2015. — №1 (74). — URL: <https://cyberleninka.ru/article/n/akusticheskie-pokazateli-golosa-v-raznye-vozzrastnye-periody-u-vzroslyh-i-ih-prognozirovanie> (дата обращения: 22.01.2026).

2 Chung, J. S. Voxceleb2: Deep speaker recognition / J. S. Chung, A. Nagrani, and A. Zisserman // Interspeech. — 2018. — P. 1086-1090.