УДК 004.934.2

ИДЕНТИФИКАЦИЯ МЕТОК РАЗГОВОРНОГО ЯЗЫКА В СЛОЖНЫХ АКУСТИЧЕСКИХ УСЛОВИЯХ ПРИ ИСПОЛЬЗОВАНИИ ЭНХАНСЕРОВ РЕЧЕВЫХ СИГНАЛОВ

Пелсе А. В.[1]
Научный руководитель – канд. техн. наук, доцент Волохов В. А.[2]
[1]Университет ИТМО, [2]ООО «ЦРТ-инновации»
shashka_aya@mail.ru

Введение

Исследование направлено на изучение влияния реверберации и методов её устранения на точность идентификации разговорного языка. Актуальность работы обусловлена необходимостью повышения надёжности алгоритмов автоматического распознавания языка в реальных условиях, характеризующихся наличием шумов, эха и реверберации. Эти факторы существенно снижают качество выполнения многих задач обработки звучащей речи, таких как идентификация говорящего [1] и идентификация языка, что подчёркивает важность разработки эффективных методов предварительной обработки сигналов.

Основная часть

Настоящее исследование проводилось с использованием тестового набора базы VoxLingua107 [2], предназначенной для оценки качества идентификации разговорного языка. Для моделирования реальных акустических условий использовалась специализированная библиотека PyRoomAcoustics [3], позволяющая имитировать распространение звуковых волн в виртуальных помещениях различного типа (больших залах, офисных комнатах и др.).

Процедура удаления реверберации осуществлялась с применением диффузионной модели, основанной на использовании мостов Шрёдингера. Этот подход демонстрирует высокую эффективность в задачах обработки речи благодаря своей способности сохранять структуру исходного сигнала даже в сложных акустических ситуациях [4, 5].

В рамках исследования рассмотрены два энхансера: стандартный энхансер, реализованный внутри фреймворка Nvidia NeMo, и его аналог, обученный на другом тренировочном наборе данных – корпусе VoxCeleb1 train вместо WSJ0 train.

Проведённые эксперименты позволили установить следующие результаты: без наложения искусственных акустических эффектов точность идентификации разговорного языка достигает 90,39%. Наложение искусственной реверберации резко снижает точность до 72,56%, демонстрируя значительное негативное воздействие акустики помещения. Применение деревербератора, обученного на тренировочном множестве VoxCeleb1 train, повышает точность до уровня примерно 83,78%, показывая существенный прирост эффективности. Стандартный деревербератор из Nvidia NeMo обеспечивает улучшение до уровня около 79,55%.

Выводы

Полученные результаты подтверждают, что использование методов предварительной обработки аудиосигналов играет критически важную роль в улучшении точности идентификации разговорного языка в сложных акустических условиях. Наиболее эффективным оказался энхансер деревербератор, обученный на датасете VoxCeleb1 train, обеспечивающий повышение точности на 11,22% по сравнению с идентификацией метки языка на данных, искажённых реверберацией. Менее эффективным, но тоже полезным

оказался стандартный девербератор из Nvidia NeMo (+6,99%). Таким образом, внедрение рассмотренных алгоритмов улучшения качества речевых сигналов позволяет повысить надёжность решений в области автоматического распознавания разговорного языка в практических приложениях.

Литература

1. Jawarkar N. P. Speaker Identification in Noisy Environment / Naresh P. Jawarkar // International Journal of Current Engineering and Scientific Research (IJCESR). – 2017. – Т. 4, № 7. – С. 37–43. – ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697. – Доступно по адресу**:** https://troindia.in/journal/ijcesr/vol4iss7part5/37-43.pdf (дата обращения: 25.01.2026).
2. Valk J., Alumae T. VOXLINGUA107: a dataset for spoken language recognition / Tallinn University of Technology, Estonia. – Доступно по адресу: https://arxiv.org/pdf/2011.12998 (дата обращения: 25.01.2026).
3. Scheibler, R. Pyroomacoustics: A Python package for audio room simulations and array processing algorithms / R. Scheibler, E. Bezzam, I. Dokmanić // 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – Calgary, CA, 2018. – P. 351–355. – DOI: 10.1109/ICASSP.2018.8461310.
4. Nasretdinov R., Korostik R., Jukic A. Robust Speech Recognition with Schrodinger Bridge-Based Speech Enhancement // arXiv preprint arXiv:2505.04237v1. – 2025. – 7 May. – Доступно по адресу: https://arxiv.org/pdf/2505.04237 (дата обращения: 10.03.2026).
5. Wang S., Liu S., Harper A., Kendrick P., Salzmann M., Cernak M. Diffusion-based Speech Enhancement with Schrodinger Bridge and Symmetric Noise Schedule // arXiv:2409.05116v2 [eess.AS]. – 13 сентября 2024 г. – Режим доступа: https://arxiv.org/pdf/2409.05116 (дата обращения: 27.02.2026).