УДК 004.934.2

Исследование нейросетевой системы шумоподавления и автоматизированной оценки произношения в аэрокосмической коммуникации

Карташев Всеволод Владимирович vsevolod.kartashev@gmail.com

ИПМ им. М.В. Келдыша РАН, Москва, Россия

Аннотация. *В работе представлена интегрированная система автоматизированной оценки произношения, устойчивая к шумовым помехам. Система основана на нейросетевых моделях и включает два ключевых этапа. Первый этап — очистка аудиосигнала от стационарных и нестационарных шумов с помощью модели автоэнкодера для минимизации искажения спектральных огибающих фонем. Второй этап — анализ очищенного сигнала для выявления ошибок произношения путем преобразования речи в текст и сопоставления звуков с эталонными фонемами. Такой комплексный подход обеспечивает высокую точность оценки речи даже в условиях сильных помех.*

Ключевые слова: *нейросетевые модели, аудиосигнал, шумовые помехи*

Введение. Современные технологии распознавания речи находят все более широкое применение, от систем обучения иностранным языкам до голосовых помощников и коммуникации в критических условиях, например, в аэрокосмической сфере. Однако эффективность таких систем напрямую зависит от двух факторов: качества входящего аудиосигнала и точности анализа произношения. Наличие фонового шума — как стационарного, так и нестационарного — значительно снижает качество распознавания речи, искажая спектральные характеристики фонем. В свою очередь, даже при чистом сигнале необходим точный механизм для оценки корректности произношения. В данном исследовании предложена комплексная система, решающая обе задачи: сначала выполняется удаление шума из аудиосигнала, а затем — детальный анализ произношения на основе очищенных данных.

Методы и материалы; результаты. Система построена на основе нескольких нейросетевых моделей, работающих последовательно.

На первом этапе для борьбы с шумовыми помехами используется нейросеть архитектуры «автоэнкодер». Модель была обучена на наборе данных, включающем сильно зашумленные аудиозаписи (в том числе из области аэрокосмической коммуникации), чистую речь и речь нескольких дикторов одновременно. Анализ показал, что шум различной природы и частотного диапазона существенно искажает спектральные огибающие фонем, что затрудняет их идентификацию. Автоэнкодер эффективно удаляет шум из аудиофрагментов, восстанавливая исходные характеристики речевого сигнала и подготавливая его для дальнейшего анализа.

На втором этапе очищенный аудиосигнал поступает в систему оценки произношения, состоящую из двух модулей. Первая нейросетевая модель преобразует аудио в текст с помощью современных алгоритмов распознавания речи. Полученный текст далее транскрибируется в последовательность фонем с использованием фонетического словаря. Второй модуль анализирует непосредственно акустические параметры звуков в очищенном сигнале и сопоставляет их с эталонными фонемами. Это позволяет с высокой точностью выявлять и подсвечивать конкретные ошибки произношения на уровне отдельных звуков, предоставляя пользователю детализированную обратную связь.

Заключение. Тестирование показало, что разработанный комплексный подход значительно повышает надежность автоматизированной оценки произношения. Предварительная очистка сигнала с помощью автоэнкодера позволяет системе эффективно работать даже с аудиозаписями низкого качества, что критически важно для применения в реальных условиях. Последующий двухэтапный анализ обеспечивает высокую точность в локализации ошибок произношения.

Таким образом, созданная система является важным шагом в развитии технологий речевого анализа. Она доказывает, что комбинация моделей для шумоподавления и оценки произношения позволяет достичь высокой точности и надежности. Дальнейшее развитие проекта предполагает расширение фонетического словаря, адаптацию системы к различным акцентам и диалектам, а также совершенствование архитектуры нейросетей для еще более качественного анализа речи.

Список источников

[1] Анастасиев С.В., Иванов А.П. *Технологии автоматического распознавания речи.* Москва, Наука, 2020.

[2] Dufour M., Lefèvre C., Toth J. et al. Phonetic recognition for language learning applications. *Speech Communication*, 2018, vol. 101, pp. 45–59. https://doi.org/10.1016/j.specom.2018.07.001

[3] Sun Y., Zhang Y., Liu J. A study of deep learning-based speech recognition systems for pronunciation assessment. *IEEE Trans. Audio, Speech, and Language Processing*, 2020, vol. 28, pp. 314–324. https://doi.org/10.1109/TASLP.2019.2945947

[4] Hinton G., Vinyals O., Dean J. Distilling the Knowledge in a Neural Network. *Advances in Neural Information Processing Systems*, 2015, vol. 28. https://doi.org/10.48550/arXiv.1503.02531

[5] Romanov P.M., Tkachenko N.V. Применение нейросетевых технологий в задачах распознавания речи: обзор современных методов. *Вестник Южного федерального университета*, 2021, № 2, с. 27–37. https://doi.org/10.22363/2313-8452-2021-2-27-37