0.1. *Вождаева Д.А.* Методы шумоподавления речевых сигналов

В настоящий момент у ОАО «РЖД» имеется потребность в разработке системы распознавания речи для последующей разработки цифровых помощников линейного сотрудника станции. Поскольку линейные сотрудники работают на станции непосредственно с подвижным составом, система распознавания речи должна быть устойчива к шумам, возникающим при записи голоса.

В данной работе были рассмотрены методы шумоподавления речевых сигналов и протестированы на небольшом наборе шумных аудиозаписей с речью линейных сотрудников, предоставленных ОАО «РЖД». Проверка качества методов проводилась с помощью применения систем распознавания речи к очищенным от шума записям и вычисления метрик WER, CER, String Accuracy.

Первая группа методов основана на фильтрации изображений – спектрограмм преобразования Фурье аудиосигналов. Были протестированы возведение в степень нормированного изображения, билатеральный фильтр [1], двумерный вейвлет-фильтр [2] и композиция этих алгоритмов. Использовалась реализация методов фильтрации из библиотеки scikit-image Python. Исследование показало, что данные методы способны улучшить разборчивость речи и дают прирост качества ее распознавания, но, к сожалению, не применимы к подавлению шума ветра, возникающего при воздействии потока воздуха на мембрану микрофона записывающего устройства.

Помимо методов фильтрации изображений были протестированы нейронные сети, в частности, Conv-TasNet [3] – сверточная сеть, изначально созданная для задачи разделения спикеров, работающая с сигналом во временной области. Ее можно применить для шумоподавления, предположив, что в роли первого спикера выступает чистая речь, в роли второго – шум. Для обучения в качестве чистой речи использовалась часть набора записей аудиокниг из Open STT, в качестве шумов – железнодорожный шум и шум ветра из Freesound Dataset 50K и YorNoise, а также сгенерированные записи ветра [4]. В процессе обучения чистая речь и шум смешивались «на лету» со случайным отношением сигнал-шум ($SNR \in [-10, 10]$) для получения большего разнообразия тренировочных данных. Использовалась реализация данной модели на PyTorch Python. Тестирование Conv-TasNet показало эффективность этой нейронной сети для подавления железнодорожного шума и шума ветра, а также прирост качества распознавания речи.

В дальнейшем предполагается добавление в набор данных других типов шумов, обучение моделей в связке с системой распознавания речи, а также применение нейронных сетей, использующих в качестве признаков мел-спектрограммы сигналов.

Список литературы

[1] Tomasi C., Manduchi R. Bilateral filtering for gray and color images // Proc. Intern. Conf. «Sixth International Conference on Computer Vision». 1998. P. 839-846

[2] Donoho D., Johnstone I. Ideal spatial adaptation by wavelet shrinkage // Biometrika. 1994. Vol. 81. N. 3. P. 425–455

[3] Luo Y., Mesgarani N. Conv-TasNet: Surpassing Ideal Time–Frequency Magnitude Masking for Speech Separation // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2019. Vol. 27. N. 8. P. 1256–1266

[4] Mirabilii D., Habets E.A.P. Simulating Multi-Channel Wind Noise Based on the Corcos Model // 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC). 2018. P. 560-564