УДК 004.89

БЕЗЭТАЛОННЫЙ МЕТОД ОБЪЕКТИВНОЙ ОЦЕНКИ РАЗБОРЧИВОСТИ РЕЧИ НА ОСНОВЕ НЕЙРОСЕТИ

Зиганшин Г.М. (ИТМО)
Научный руководитель – кандидат технических наук Столбов М.Б. (ИТМО)

Введение. Разборчивость речи (РР) – способность слушателя расшифровать содержимое речи: слоги, слова, фразы. Искажение речевого сигнала (РС), приводящее к ухудшению разборчивости, может быть вызвано различными факторами: шумом, реверберацией, потерей данных при кодировании и передаче аудиосигналов, искажениями в электронных устройствах.

Оценка РР важна во многих областях: в системах связи; в системах распознавания и синтеза речи; в решении задач идентификации дикторов; в медицине (системах протезирования и восстановления слуха), в оценке алгоритмов компенсации шумов и многих других.

Методы оценки РР можно разделить на два больших класса: субъективные и объективные. Объективные методы основаны на оценке характеристик звуковых сигналов и могут быть реализованы техническими средствами. Этот класс методов можно также разделить на две группы: требующие наличия эталонного сигнала (эталонные) и способные делать оценку по самому оцениваемому сигналу (безэталонные). Вторая группа методов представляет больший интерес, так как имеет широкий спектр применения, например, в системах реального времени, когда доступ к эталонному сигналу затруднен или невозможен.

Основная часть. В настоящий момент актуальной задачей является разработка методов, способных делать оценку РР только по данным из оцениваемого сигнала. Применение нейронных сетей (НС) позволяет исследовать новые подходы к решению этой задачи. Целью работы является разработка безэталонного метода объективной оценки разборчивости речи на основе нейросетей.

Основной идеей предложенного безэталонного метода оценки РР является обучение нейросети для предсказания выбранной метрики РР. В качестве метрики РР использовалась объективная оценка разборчивости STOI (Short-Time Objective Intelligibility) [1]. Значение метрики STOI равное 1 соответствует 100% разборчивости. В качестве входных данных для нейросети выбраны мел-кепстральные коэффициенты, вычисляемые на кадрах речи зашумленной фонограммы. Выбрана архитектура нейросети, представляющая собой комбинацию сверточных и полносвязных слоев.

Для обучения и тестирования нейросети был составлен набор фонограмм, включающий в себя как чистые, так и зашумленные РС. Зашумленные сигналы были получены из исходных РС наложением различных шумов: музыки, белого шума, шума транспорта, разговорного шума. Зашумление проводилось в диапазоне от -6 дБ до 12 дБ. В качестве исходных данных использованы аудиозаписи на английском, французском, японском, итальянском языках и записи шумов из базы данных «Supplement 23 to the P series of ITU-T Recommendation: coded-speech database» [2]. Частота дискретизации во всех аудио равна 16 кГц.

Полученный набор данных был разделен на обучающие и тестовые выборки. На этапе обучения для оценки STOI использовалны пары фонограмм эталонных и зашумленных сигналов.

Результаты. В результате обучения НС была достигнута корреляция между истинными и предсказанными значениями STOI. Среднее значение (в диапазоне различных значений STOI) среднеквадратичного отклонения предсказанной оценки от истинного значения составило приблизительно 0,1-0,15.

Таким образом, предложенный метод может быть положен в основу безэталонной оценки РР зашумленных РС. Дальнейшим направлением работы является исследование границ работоспособности предложенного метода.

Список использованных источников.

1. Cees H. Taal. A short-time objective intelligibility measure for time-frequency weighted noisy speech / Cees H. Taal. Richard C. Hendriks, Richard Heusdens, Jesper Jensen – Текст: электронный // IEEE International Conference on Acoustics, Speech and Signal Processing : материалы конференции, 14-19 марта 2010 г. / IEEE Xplore: сайт. – Дата публикации: 28 июня 2010 – URL: https://ieeexplore.ieee.org/document/5495701 (дата обращения: 20.11.2023).
2. ITU-T P.Sup23. Supplement 23 to ITU-T P-Series recommendations: ITU-T coded-speech database : рекомендации : издание официальное : одобрен в соответствии с процедурой Резолюции WTSC 27 февраля 1998 г. № 5 / подготовлено ITU-T study group 12 (1997–2000 гг.) – Текст: электронный / ITU-T publications – 1998 – URL: https://www.itu.int/rec/T-REC-P.Sup23-199802-I/en (дата обращения 20.11.2023).