УДК 004.934

ОБЗОР КОРПУСОВ ДЛЯ ДЕТЕКТИРОВАНИЯ ЭКСТРАЛИНГВИСТИЧЕСКИХ СОБЫТИЙ

Ролинский С.О (Университет ИТМО), **Двойникова А.А.** (Санкт-Петербургский Федеральный исследовательский центр Российской академии наук)
Научный руководитель – д. т. н., доцент Карпов А. А.
(Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Университет ИТМО)

В работе описывается актуальность и значимость анализа акустических событий, а также практическая значимость результатов анализа в различных областях науки. В ходе исследований проводится аналитический обзор существующих корпусов, содержащих разметку звуковых экстралингвистических явлений, включающих воспроизводимые человеком невербальные звуки, такие как смех, плачь, всхлипы, вздохи, кашель и пр.

Введение. Экстралингвистическое акустическое событие – звуковое явление, имеющее определенную (как правило, короткую) длительность, время возникновения и окончания. Данное исследование направлено на детектирование событий, создаваемых человеком, таких, как кашель, чихание, смех, вздохи, выдохи, всхлипы. Каждое из этих событий несет в себе определенную информацию о состоянии диктора, которую можно использовать для дальнейшего анализа речи и определения состояния человека как физиологического, так и эмоционального. Наиболее актуальные задачи, где находит применение детектирование экстралингвистических событий – распознавание эмоций (при помощи смеха или плача), мониторинг здоровья говорящего (кашель, чихание, хрипы), определение психологической стабильности (короткие/длинные вдохи и выдохи) и другие. В современном мире, особенно в условиях частой заболеваемости и повышенной значимости эмоций, задача детектирования акустических экстралингвистических событий человека является актуальной и значимой.

Основная часть. Существуют различные корпусы, в которых присутствует разметка акустических событий разного рода, относящиеся к звукам человека. Рассмотрим основные из них:

- Belfast storytelling dataset [McKeown G. et al. The Belfast storytelling database: A spontaneous social interaction database with laughter focused annotation //2015 International Conference on Affective Computing and Intelligent Interaction (ACII). – IEEE, 2015. – С. 166-172.]. Набор данных, содержит в себе разговор по группам из 4 человек. Разговор проходит на различные позитивные темы и юмористические истории, в котором участники могут как активно коммуницировать между собой, так и являться пассивными слушателями одного из участников. Разметка аудиоданных описываемого корпуса содержит в себе временные обозначения начала и конца историй, а также маркеры событий смеха. Данные также размечены на множество различных подвидов смеха (восторг, злорадство, облегчение и т.д.).
- AMI (Augmented Multi-party Interaction) [McCowan I. et al. The AMI meeting corpus //Proceedings of the 5th international conference on methods and techniques in behavioral research. – 2005. – Т. 88. – С. 100.]. Аудиовизуальный корпус содержит в себе данные участников деловых встреч, проходящих как по сценарию, так и в естественной среде. Набор размечен неологизмами, звукоимитирующими явлениями и отметками для прерываний и заиканий. Также в разметках присутствуют маркеры акустических событий, таких как смех и кашель/хрип.
- ASVP-ESD (Audio, Speech, and Vision Processing Lab Emotional Sound database) [https://www.kaggle.com/dejolilandry/asvpesdspeech-nonspeech-emotional-utterances]. – Корпус состоит из аудиоданных из фильмов, шоу, роликов на YouTube, выбранных

вручную так, чтобы содержать явления смеха, глубоких вздохов, всхлипов и других акустических событий. Всего выделяется 13 различных эмоций (скука, счастье ,злоба, страх, восторг, отвращение, удивление, боль и другие), не все из которых содержат акустические события, также присутствует деление на язык, возрастную группу и эмоциональную нагрузку.

- MMLI (Multimodal Multiperson Corpus of Laughter in Interaction) [Niewiadomski R. et al. MMLI: Multimodal multiperson corpus of laughter in interaction //International Workshop on Human Behavior Understanding. – Springer, Cham, 2013. – С. 184-195.]. Корпус аудиовизуальных данных содержит рассказы веселых и смешных ситуаций, направленных на получение естественного смеха. Участников просили смотреть смешные видео, играть в словесные игры и рассказывать скороговорки. Почти все участники были друзьями, что позволяло участникам чувствовать себя комфортно, и таким образом создавались максимально естественные условия для выражения натурального смеха.
- Real and Synthetic Audio-Cough Events [Monge-Álvarez J. et al. Robust detection of audio-cough events using local hu moments //IEEE journal of biomedical and health informatics. – 2018. – Т. 23. – №. 1. – С. 184-196.]. Содержит речевые высказывания с кашлем. Она состоит из двух частей – первая содержит имитированный кашель, полученный синтетическим путем (наложением обыкновенных аудио речи на звуки кашля из библиотеки звуков), а вторая – естественный кашель пациентов.

Помимо этого, для аугментации данных исследователями часто создаются собственные наборы данных с помощью объединения речевых аудиоданных и звуков, взятых из соответствующих библиотек (например, FreeSound).

С использованием вышеперечисленных наборов данных планируется создать систему, которая будет выделять и определять в этих данных такие экстралингвистические события как кашель, смеха, всхлипы, хрипы.

Выводы. В рамках работы был проведен сравнительный обзор существующего информационного обеспечения (корпусов) экстралингвистических звуков говорящих людей. Из представленного обзора можно сделать вывод, что существует несколько корпусов для детектирования кашля, чихания, смеха и пр. С использованием описанных наборов данных акустических событий в речи человека и современных методов машинного обучения можно разработать автоматическую систему детектирования таких экстралингвистических событий как кашель, смех, вздохи, чихание и т.д. Данная система будет актуальной и полезной для сферы речевых технологий и человеко-машинного взаимодействия.

Исследование выполнено при поддержке Совета по грантам Президента РФ (грант № НШ-17.2022.1.6)