

Сравнение систем автоматического распознавания возраста диктора по голосу

М. В. Маркитантов, магистрант
(Университет ИТМО, г. Санкт-Петербург)

А.А. Карпов, д.т.н., профессор
(Университет ИТМО, г. Санкт-Петербург)

Научный руководитель – А. Ю. Тропченко д.т.н, профессор
(Университет ИТМО, г. Санкт-Петербург)

В повседневном общении люди используют не только вербальную (речевую, текстовую и т.д.), но и невербальную (паралингвистическую) информацию. Последняя может содержать такие характеристики диктора, как его эмоциональное состояние, возраст и пол. При отсутствии прямого контакта с клиентом (пользователем), паралингвистическая информация о нем может быть полезна для предоставления определенных товаров и услуг в Интернете.

Автоматическое распознавание возраста диктора необходимо для систем проверки и идентификации дикторов, в частности, для улучшения человеко-машинного взаимодействия, а также в работе контакт-центров, учреждений здравоохранения и для повышения эффективности целевой рекламы.

Различные исследования по извлечению акустических признаков и разработке машинного классификатора возраста диктора по голосу пока не позволяют получить удовлетворительную точность распознавания. Получение возрастной информации по речи диктора осложняется влиянием таких факторов, как, например, фоновый шум и голосовая вариативность. Таким образом, ключевая проблема в определении возраста диктора заключается в извлечении надежных признаков и разработке эффективного метода классификации.

Целью работы является сравнение систем автоматического распознавания возраста диктора по голосу.

В работе представлен аналитический обзор нескольких речевых корпусов, включая корпуса NIST SRE2008, NIST SRE 2010 и aGender, который использовался в ходе соревнования Computational Paralinguistics Challenge (ComParE) в рамках INTERSPEECH 2010, а также анализ используемых низкоуровневых признаков, которые содержат интонацию, интенсивность, мел-частотные кепстральные коэффициенты, *i*-вектора, *x*-вектора, а также различные наборы инструментов для извлечения признаков, такие как Lirosa и openSMILE.

В работе проведено сравнение существующих систем автоматического распознавания возраста диктора по голосу, включая системы, участвующие в соревновании ComParE и Яндекс SpeechKit.