УДК 004.02

АВТОМАТИЧЕСКОЕ ЧТЕНИЕ РЕЧИ ПО ГУБАМ ДЛЯ ЛЮДЕЙ С НАРУШЕНИЕМ СЛУХА

Иванько Д.В. (аспирант)
Научный руководитель – д.т.н., проф. А.А. Карпов
Университет ИТМО

В настоящее время автоматическому чтению речи по губам, также известному как визуальное распознавание речи (англ. visual speech recognition, VSR), уделяется большое внимание в связи с его потенциальным использованием в приложениях человеко-машинного взаимодействия, распознавания языка жестов, аудиовизуального распознавания речи, биометрии и биомедицины.

Люди с нарушениями слуха или глухие используют язык жестов в качестве основного способа общения. Язык жестов - это структурированная форма жестов рук, включающая движения и знаки, которая используется в качестве системы связи. Язык жестов предполагает использование разных частей тела - не только движений рук и пальцев. Глухие люди также используют движения губ как часть языка жестов, хотя они не могут слышать акустическую речь. Более того, движения губ у людей с нарушениями слуха обычно характеризуются гиперартикуляцией. Этот факт потенциально дает возможность лучше распознать визуальную речь такой группы людей, поскольку движения губ становятся более выраженными. Хотя значительное количество исследований было посвящено теме визуального декодирования речи, проблема чтения по губам для людей с нарушениями слуха остается открытой проблемой в данной области.

Одним из основных препятствий для современных исследований в области визуального распознавания речи является отсутствие подходящих баз данных. В отличие от большого количества существующих аудиоречевых корпусов, очень мало баз данных общедоступно для визуальных или аудиовизуальных систем распознавания речи. Большинство из них включают ограниченное количество дикторов и небольшой словарный запас. Баз данных на русском языке жестов, пригодных для обучения VSR, практически не существует. По этой причине в 2018 году в СПИИРАН была зарегистрирована База данных русского языка жестов. В состав базы данных входят записи 13 дикторов-носителей русского языка жестов. Каждый диктор продемонстрировал 164 фразы по 5 раз. Общее количество записей в корпусе составляет 10660.

В общем случае, визуальное распознавание речи - это процесс преобразования последовательностей изображений области рта в текст. Традиционный подход основан на виземах (базовых визуальных речевых единицах), поскольку акустические особенности некоторых фонем не четко различимы по месту их артикуляции. В нашем исследовании мы использовали расширенный список из 20 русских визем. Каждая практическая система чтения по губам обязательно включает 4 основных этапа обработки, таких как получение изображения, локализация области интереса (ROI), извлечение признаков и распознавание речи.

В данной работе представляется система чтения речи по губам для людей с нарушениями слуха. Результаты этого исследования впоследствии могут быть использованы для создания вспомогательных технологий для глухих или слабослышащих людей.

Благодарности.

Данное исследование проводится при поддержке фонда РФФИ (проект № 18-37-00306) и Правительства РФ (Грант № 08-08).