

ОСОБЕННОСТИ ОЦЕНКИ ЭКСПРЕССИВНОГО СИНТЕЗА РЕЧИ В ЯНДЕКС ЗАДАНИЯХ

Хайдукова Т.В. (ИТМО)

Научный руководитель – Томилов А.А. (ООО «ЦРТ-Инновации»)

Введение. Современные системы синтеза речи активно внедряются в повседневные приложения. Одной из ключевых задач остаётся обеспечение выразительности и эмоциональной окраски синтезируемой речи, что позволяет сделать опыт использования автоматизированных сервисов более естественным [1]. Вместо традиционной оценки качества синтеза, актуальной становится задача определения потребности в эмоциональной окраске речи. В данной работе предлагается методика оценки потребности в эмоциональной окраске речи с помощью пользовательских опросов, позволяющая определить значимость экспрессивного синтеза.

Основная часть. Методика оценки потребности в эмоциональной окраске речи состоит из нескольких этапов: проведения пользовательских опросов, сбора данных, статистического анализа и интерпретации результатов. Оценка предпочтений пользователей проводилась с помощью MOS тестирования [2] на платформе Яндекс Задания. В исследовании приняли участие 300 пользователей от 18 до 72 лет, все – носители русского языка. Им предлагалось оценить синтезированные фразы в двух вариантах: нейтральном и эмоциональном. Опрос включал сопоставление синтезированной речи с воспринимаемой эмоцией, оценки эмоциональной выразительности и эмоциональной уместности синтезированной речи и сравнительный выбор между двумя вариантами синтеза (А/В-тестирование). Статистический анализ ответов позволил выявить ключевые закономерности в предпочтениях пользователей и определить важность эмоциональной окраски. Анализ результатов показал следующее. Респонденты наиболее точно распознали злость в синтезированной речи (92%), наибольшая показатель полноты у синтеза с проявлением счастья – респонденты обнаружили 86% всех фактических случаев «счастья». Среднее значение выразительности речи составило 3,59 из 5, когда уместность эмоциональности составила 3,77 из 5. В 58,5% случаев пользователям нравилась «эмоциональная» версия, нейтральный вариант предпочли 13,8% участников, а оставшиеся 27,8% не увидели разницы или затруднились с выбором. Таким образом, благодаря трём разным опросам и совокупности в 100 участников на каждый блок, удалось собрать комплексную оценку. Эмоциональная окраска действительно считывается слушателями и повышает субъективное восприятие. Полученные результаты дают основания улучшать синтез и точнее настраивать интонационные признаки при злости или радости.

Выводы. Использование разработанной методики оценки экспрессивного синтеза речи позволяет значительно улучшить качество синтезированной речи.

Список использованных источников:

1. Burkhardt F., Campbell N. Emotional Speech Synthesis // The Oxford Handbook of Affective Computing / Oxford Library of Psychology – 2015. – P. 286–295
2. Streijl R. C., Winkler S., Hands D. S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives // Multimedia Systems. – 2016. – V. 22. – №. 2. – P. 213–227.