

ОБЗОР СИНТЕЗА ЭМОЦИОНАЛЬНОЙ РЕЧИ

А. Калиев

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, г. Санкт-Петербург)

Научный руководитель – доктор технических наук, Ю.Н. Матвеев

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, г. Санкт-Петербург)

Работа проводится в рамках темы НИР № 616029 «Разработка и исследование методов и алгоритмов распознавания эмоционального и психофизического состояния человека по многомодальным данным»

Целью этой работы является предоставление детального обзора развития эмоционального синтеза речи. Данная статья рассматривает методики синтеза речи, подходы анализа и оценки параметров эмоциональной речи, а также методы включения параметров эмоциональной речи в синтез речи.

Как программное приложение или интегральная часть операционной системы, синтез речи появилась вначале 1980-х в компьютерах, таких как Apple Macintosh и Commodore Amiga[1]. До середины 1980-х исследованием в этой области могли позволить себе только крупные лаборатории и компаний, но появление относительно дешевых и мощных компьютеров сподвигнуло распространению исследования по многим университетам и лабораториям. В дальнейшем с увеличением компьютерной памяти и вычислительной мощности, исследователи искали пути улучшения качества синтеза речи, что породило множества методов конкатенаций речи[2].

Несмотря на огромные усилия многих великолепных ученых имитировать физические процессы, качество синтеза речи оставалась неестественным и неприятным для человеческого слуха. Имитировать физические процессы пытались с помощью генерации артикулярной модели вокального тракта, затем путем синтеза модели из спектральной временной изменчивости и из временных свойств речи, а позже с помощью конкатенации речевых элементов [3]. Одна из причин не успешности синтеза речи с помощью конкатенации стало то, что элементы, использовавшийся для склеивания были записаны в лабораториях, где речь была специально записано в просодический нейтральном тоне. Хотя, речевые элементы содержали соответствующие спектральные характеристики для заданной звуковой последовательности, они не могли достаточно правильно моделировать различные динамические артикуляторные характеристики этой последовательности в разных контекстах[4].

Литература

1. David Suendermann, Harald Höge, Alan Black. Challenges in Speech Synthesis, In book: Speech Technology, pp.19-32, July 2010
2. Alan W Black, Kevin A. Lenzo. Building Synthetic Voices. For FestVox 2.7 Edition. 24th December, 2014
3. Taylor, P. Text-to-Speech Synthesis. Cambridge University Press. 2008
4. N. Campbell and A. W. Black. Prosody and the selection of source units for concatenative synthesis. In J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, editors, Progress in Speech Synthesis. Springer Verlag, 1995.