

Shadowing Synthesised Speech - Analysis of Phonetic Convergence

The fact that the shape of an utterance can be connected to various aspects such as physiology, speaker dialect, idiolect and conversational settings has been demonstrated in recent phonetic studies. It was presented that talkers converge to interlocutor in phonetic and acoustic aspects which is considered to be a property of natural dialogue [1]. The fact that speakers can converge with interacting partners by subconscious manipulation of attributes such as accent, speaking rate, intensity, utterance duration and frequency of pauses has been studied and discussed in many articles. However, the topic of convergence to the synthesised speech remains unclear [2].

An experiment of phonetic convergence in humans when shadowing synthetic speech in the Polish language was conducted. The objective was to determine the level of phonetic convergence when shadowing synthesised speech in Polish. For the study, four sets of sentences with specific linguistic phenomena were prepared. These were the following: (1) orthographic word-medial letters *ę*, *ą* before vowels other than fricative, (2) the letter *ń* in various positions, (3) realisation of *em(n)*, *om(n)* word-initially in loanwords, (4) combinations of letters *trz*, *strz*. The whole procedure of the experiment includes several stages: (1) baseline production – reading, (2) visual task, (3) shadowing task, (4) postproduction. The stimuli for the shadowing task was synthetic. The analysis of the level of convergence is based on the comparison of segmental and suprasegmental characteristics of the utterance. The suprasegmental characteristics were measured with the My-Voice Analysis Python library [3]. The following features were extracted: number of syllables, number of pauses, rate of speech, articulation rate, original duration, speaking duration, F0 (mean, median, standard deviation, minimal and maximal value).

In order to compare the level of the convergence to synthesised speech to the level of convergence to the natural speech two speakers were recorded as the model, one male and one female. The procedure of this part of the experiment was the same as when shadowing the synthesised speech.

- [1] J. S. Pardo, “Phonetic convergence in shadowed speech: A comparison of perceptual and acoustic measures,” Lyon, 2013.
- [2] I. Gessinger, E. Raveh, S. Le Maguer, B. Möbius and I. Steiner, “Shadowing Synthesized Speech — Segmental Analysis of Phonetic Convergence,” in *Shadowing Synthesized Speech — Segmental Analysis of Phonetic Convergence*, Stockholm, 2017.
- [3] “GitHub,” 2019. [Online]. Available: <https://github.com/Shahabks/my-voice-analysis>. [Accessed 19 April 2019].