

Persuasive synthetic speech: voice perception and user behaviour

Mateusz Dubiel¹, Pilar Oplustil Gallegos², Martin Halvey³, Simon King²

¹University of Luxembourg, ²University of Edinburgh, ³University of Strathclyde

corresponding author: mateusz.dubiel@uni.lu,

The following abstract is based on *Persuasive Synthetic Speech: Voice Perception and User Behaviour* (Dubiel et al., 2020) research paper that was published at the 2nd Conference on Conversational User Interfaces (CUI'20). When the paper was published the first author was affiliated with the University of Strathclyde.

Abstract (max 300 words)

Previous research indicates that synthetic speech can be as persuasive as human speech (Stern et al., 1999). However, there is a lack of empirical validation on interactive goal-oriented tasks. In our two-stage study (Dubiel et al., 2020), which comprised of online listening test and interactive evaluation, we compared participants' perception of the persuasiveness of synthetic voices created from speech in a debating style (*IBM Debater* dataset (Mirkin et al., 2018)) vs. speech created from audio-books (*Libri TTS* dataset (Zen et al., 2019)). The goal of the first stage was to select the single most persuasive synthetic voice per dataset, thus providing a persuasive voice for the second stage, and a strong baseline voice for comparison. In the second stage (interactive evaluation), participants undertook a series of search tasks, interacting with a Conversational Agent (CA) to achieve the goal of selecting a flight. In each task, the CA attempted to persuade participants to change their original flight selection by providing counterarguments. We evaluated the persuasiveness of the CA via: (1) questionnaires adapted from Stern et al. (1999), covering perception of the message, voice, and personal qualities of the speaker; (2) the number of times a participant followed the recommendation of the CA. We found that participants who interacted with the CA using the voice created from the debating style speech rated it as significantly more truthful and more involved (qualities of persuasive speakers) than the CA using the audio-book-based voice. However, there was no difference in how frequently each group followed the CA's recommendations during interactive goal-oriented tasks. While our findings are preliminary and further experiments in different domains would be required to assess their validity, our proposed experiment is an important step towards more ecologically-valid evaluation of text-to-speech quality for systems designed to support decision making in goal-oriented tasks.

[Abstract Word count: 299]

References

- Dubiel, M., Halvey, M., Gallegos, P. O., & King, S. (2020). Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (pp. 1-9). <https://doi.org/10.1145/3405755.3406120>
- Mirkin, S., Jacovi, M., Lavee, T., Kuo, H-K., Thomas, S., Sager, L., Kotlerman, L., Venezian, E. and Slonim, N. (2018). *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* <https://aclanthology.org/L18-1037>

Stern, S. E., Mullennix, J. W., Dyson, C. L., & Wilson, S. J. (1999). The persuasiveness of synthetic speech versus human speech. *Human Factors*, 41(4), 588-595.
<https://doi.org/10.1518%2F001872099779656680>

Zen, H., Dang, V., Clark, R., Zhang, Y., Weiss, R. J., Jia, Y., ... & Wu, Y. (2019). LibriTTS: A corpus derived from LibriSpeech for text-to-speech. *arXiv preprint arXiv:1904.02882*.