World Academy of Science, Engineering and Technology
International Journal of Cognitive and Language Sciences
Vol:18, No:07, 2024

Text-to-Speech in Azerbaijani Language via Transfer Learning in a Low Resource Environment

Authors : Dzhavidan Zeinalov, Bugra Sen, Firangiz Aslanova

Abstract : Most text-to-speech models cannot operate well in low-resource languages and require a great amount of high-quality training data to be considered good enough. Yet, with the improvements made in ASR systems, it is now much easier than ever to collect data for the design of custom text-to-speech models. In this work, our work on using the ASR model to collect data to build a viable text-to-speech system for one of the leading financial institutions of Azerbaijan will be outlined. NVIDIA's implementation of the Tacotron 2 model was utilized along with the HiFiGAN vocoder. As for the training, the model was first trained with high-quality audio data collected from the Internet, then fine-tuned on the bank's single speaker call center data. The results were then evaluated by 50 different listeners and got a mean opinion score of 4.17, displaying that our method is indeed viable. With this, we have successfully designed the first text-to-speech model in Azerbaijani and publicly shared 12 hours of audiobook data for everyone to use.

Keywords : Azerbaijani language, HiFiGAN, Tacotron 2, text-to-speech, transfer learning, whisper
Conference Title : ICSST 2024 : International Conference on Speech Science and Technology
Conference Location : New York, United States
Conference Dates : July 11-12, 2024