

Recent advances in automatic speech recognition (ASR) systems have brought many benefits in man-machine speech interaction. Despite good performance in controlled conditions, relatively high sensitivity in adverse conditions is the main reason for poor use in real-life scenarios [].

Speech technologies are intended for commonly used mode of speech, i.e., normally phonated speech (neutral speech). Other speech modalities include shouted speech, louder speech, soft speech and whisper. The parameters for distinction of these 5 modes are sound pressure level (SPL), sentence duration and silence percentage, frame energy distribution and spectral tilt [].

Whispered speech is pronounced as an alternative to neutral speech for a number of reasons. ……

Recent studies demonstrated performance gain in whispered speech recognition using data augmentation techniques [1,2,3], as well as voice conversion [4].

[1] Gudepu, Prithvi R. R., Gowtham P. Vadisetti, Abhishek Niranjan, Kinnera Saranu, Raghava Sarma, M. Ali Basha Shaik and Periyasamy Paramasivam. “Whisper Augmented End-to-End/Hybrid Speech Recognition System - CycleGAN Approach.” *INTERSPEECH* (2020).

[2] Data Augmentation for Whispered Speech Recognition

[3] Data Augmentation for ASR using CycleGAN-VC

[4] Marius Cotescu, Thomas Drugman, Goeric Huybrechts, Jaime Lorenzo-Trueba, Alexis

Moinet, Voice Conversion for Whispered Speech Synthesis, IEEE Signal Processing Letters, 2020.

The goal of research study presented in this paper is present initial results in real-time speaker independent recognition of bimodal produced speech (isolated words in neutral speech and whisper) for Serbian. The ASR system for training and testing is based on CMU Sphinx platform [].

The remainder of this paper is organized in the following manner. In Section 2 short overview of hidden Markov Models is given. Section 3 presents ASR system, speech database and experiment setup for testing. In Section 4 we give results of experiments as well as its discussion, whereas concluding remarks and direction for future studies are stated in Section 5.