

ABSTRACT

Numerous approaches are used in communication, including sign language, facial expressions, gestures, postures, and many more. Speech is the most effective method of expressing one's emotions and enables one to comprehend the feelings of others in a better way. Communication is fulfilled when the listener can understand the speech without any disturbance to communication.

Communication through Speech signals is usually embedded with background noise; therefore, removing noise becomes mandatory for better understanding. The objective of speech enhancement and separation is to separate or enhance a target speech signal from a mixture of sounds generated from one or more sources. The enhancement of speech signals is quite tedious compared to other signals because of its nature which frequently changes with time. The enhanced speech signals facilitate speech recognition in human-human and human-machine interaction.

Speech recognition systems translate spoken words into text using computer algorithms. The speech signal detected by the microphone is recognized by analyzing the audio and then segmenting it into sections, digitizing it, and converting it into a machine-readable format. The recognition algorithms that convert audio into text undergo training on various speaking styles, patterns and accents. Before the recognition process, the noise embedded in the speech needs to be removed. In order to make the recognition system understand speech, speech enhancement becomes essential and mandatory. This research focuses on implementing various speech enhancement techniques for noisy speech.

The speech dataset for the speech enhancement system is taken from the University of Edinburgh, Centre for Speech Technology Research (CSTR). The clean speech is combined with various noises at varying decibels of noise, such as -10dB, -5dB, 0dB, 5dB, 10dB, and 15dB, to create the noisy speech dataset. Washing machine noise, Rainbow noise, Babble noise, Airport noise, Jet airplane noise, Street noise, Train whistle noise, and Restaurant noise are used for training and testing, and the unseen noises considered for evaluating the speech enhancement system are car noise and subway noise. Based on the Signal to Noise Ratio (SNR), segmental SNR (segSNR), Perceptual Evaluation of Speech Quality (PESQ), Short-Term Objective Intelligibility (STOI), Scale Invariant Signal-to-

Distortion Ratio (SI-SDR) and Deep Noise Suppression Mean Opinion Score (DNSMOS) the performance of the improved speech is assessed. The noisy speech is denoised using different algorithms such as Wiener filtering, traditional hybrid algorithm (combining the Wavelet Transform, Wiener filter, and Least Mean Squares (LMS) algorithm), Deep Fully Connected Neural Network (DFNN), Deep Convolutional Neural Network (Deep CNN), modified Long Short-Term Memory (modified LSTM) and modified Fully Convolutional Recurrent Network (modified FCRN).

This research work is extended by applying the algorithms for enhancing the Alaryngeal speech generated by Laryngectomy patients. When the voice box is defective, people experiencing voice loss use whispered speech as their primary method of communication. With prostheses or specialized speech therapy methods, patients who undergo partial or complete laryngectomy are usually incapable of speaking more than hoarse whispers. In order to address this major issue, machine learning and deep learning algorithms are used for the noise removal and enhancement of speech signals. Under various noise conditions, deep learning-based approaches have significantly increased the performance of speech enhancement system by improving speech intelligibility and quality.

The novelty of this work is creation of a unique resource consisting of alaryngeal speech with sentences spoken by subjects post-laryngectomy. This research contributes significantly to voice pathology detection by evaluating and comparing various speech enhancement algorithms for enhancing the speech affected by various noises. Noise removal and speech enhancement is achieved by applying various deep learning algorithms such as DFNN, Deep CNN, modified LSTM and modified FCRN. The performance evaluation is done by evaluating SNR, segSNR, PESQ, STOI, SI-SDR and DNSMOS. The comparative study on results obtained from the techniques implemented for speech enhancement is performed for normal person speech signals taken from the CSTR dataset and speech signals collected from alaryngeal patients. The result shows that the speech enhancement algorithms could help improve patients' speech with Blom Singer's non-indwelling voice prosthesis. It is obvious and evident that the noisy speech signal enhanced through the modified FCRN technique gives outstanding results as per the performance metrics analyzed. The best-performing algorithm is validated by evaluating the word error rate of the denoised speech signal. A user-friendly tool is designed as a MATLAB Application to enhance alaryngeal speech recordings in the presence of various types and levels of background noise, and the

paralinguistic features of the alaryngeal speech is compared with the normophonic speech to analyse the variation in the speech.

Keywords: Laryngectomy, Speech Enhancement, Speech Quality, Intelligibility, Alaryngeal Speech, Word Error Rate, Paralinguistic Features