Development of a Real-Time Audio Signal Processing System for Speech Enhancement

Deeksha Naithani
Department of Elect. Engg., Graphic Era Hill University, Dehradun, Uttarakhand, India 248002

Article Info
Page Number: 1041-1052
Publication Issue:
Vol. 70 No. 2 (2021)

Article History
Article Received:18 October 2021
Revised: 20 November 2021
Accepted: 22 December 2021

Abstract: The requirements of real-time signal processing dictate that the audio signal must be completely processed before the subsequent audio segment may be received. This is done in order to achieve those requirements. This highlights how important it is to create methods of signal processing that are not just quick but also accurate. I describe many ways for processing audio signals in real time within the scope of this thesis. The publications that are being presented cover a wide range of issues, including noise dosimetry, speech analysis, and network echo cancellation, to name a few. In this article, the process of constructing a system that uses audio signal processing to improve speech in real time is broken down and examined. Speech enhancement makes speech more audible and comprehensible in noisy surroundings, which is beneficial to individuals who have hearing loss as well as speech recognition and communication systems. Signal-to-noise ratio, power spectral efficiency ratio, and spectrum transfer entropy index are the metrics that are utilized to evaluate speech quality enhancement system development strategies, abbreviated as STOI. Because studies have shown that being exposed to loud noises over extended periods of time can have severe impacts on health, it is essential to have precise methods for detecting the levels of noise. The findings of a study that measured exposure to noise while also taking into account the impact of the speaker's own voice are presented in this article.

Keywords: Audio signal processing, speech enhancement, pre-processing, feature extraction, machine learning, noise reduction, normalization, segmentation.

I. Introduction

Using dependable digital communication tools is becoming more and more important. Many communications equipment employ signal processing techniques to enhance speech clarity and eliminate background noise. The complexity of computations that can be performed by these devices is constrained. Despite the increase in processing power made possible by the development of electronic components, low complexity algorithms are still necessary. Selecting components with sufficient processing power lowers the cost of the final product. We require algorithms that are both quick and dependable in terms of output quality. Real-time signal processing demands that the audio signal processing be finished before the subsequent audio segment is received [1]. The sound produced if this restriction is disregarded will be, at best, below average. Speech augmentation is to recover speech signals that have been damaged by outside influences like noise. Because speech is a key form of human communication that is constantly interfered with in the real world, research in this field is essential. In a noisy, packed environment, background noise, reverberation, and other types of interference can make it challenging to understand a speaker. Speech enhancement algorithms aim to reduce or eliminate certain kinds of interference in order to increase the intelligibility of spoken signals.Statistical modelling, spectral analysis, and filtering are a few of the techniques used in speech enhancement. Filtering techniques are used to improve a speech signal. This can be done using a variety of filters, such as high-pass, low-pass, and bandpass. Your choice of filter should be influenced by the noise or interference characteristics of the signal. A high-pass filter can be used to eliminate low-frequency noise, while a bandpass filter can be used to eliminate noise within a certain frequency range[2]. Enhancing one's voice is another application of spectral analysis. Examining the spectral composition of a speech transmission allows for the detection of polluted frequency bands. Spectral subtraction and Wiener filtering are two methods that may be used to eliminate these regions after they have been located. Noise or interference in a speech stream can be estimated with these methods and then eliminated. For voice enhancement, techniques including filtering and spectral analysis as well as statistical modelling are employed. To isolate a human voice from background noise, it is important to model the statistical properties of the speech signal. Neural networks, hidden Markov models (HMMs), and Gaussian mixture models are just a few examples of the statistical models utilised in voice enhancement (GMMs). These models are then used to detect patterns in the speech signal that may indicate noise or interference after being trained on a large dataset of speech signals. Using digital signal processing (DSP) techniques, speech augmentation algorithms may be implemented in real-time. Digital signal processing (DSP), which processes digital inputs, is used to improve the quality of a person's speech in real time.

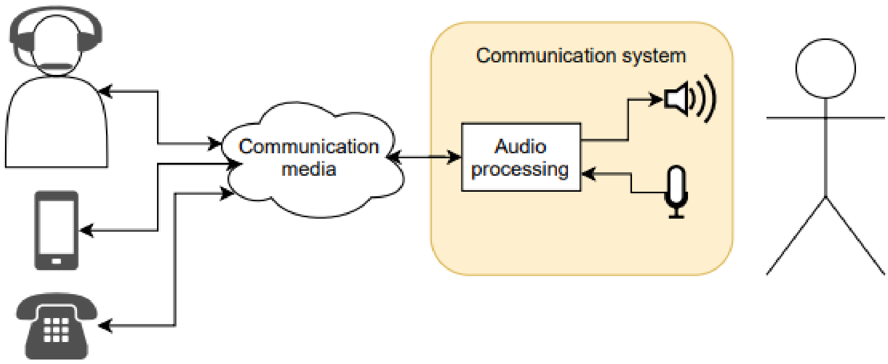

Figure 1. Working Structure of Digital Communication System [3]

Digital signal processing (DSP) methods can be implemented in either specialized hardware or software, depending on the use case. Speech augmentation is useful in a wide range of real-world situations, including communications, hearing aids, audio recording, and production. Voice enhancement algorithms can be used by communications engineers to improve the audio quality of phone and internet calls. Hearing aids with speech enhancement enable users to understand speech more accurately in difficult listening environments. In the audio recording and production industry, speech enhancement is a technology used to enhance the quality of audio recordings, making them more understandable and entertaining to listen to. The base minimum for two-way communication in an audio-conferencing system is as indicated in Figure 1: a microphone and a loudspeaker for each participant as well as

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

some form of communication medium. When one caller speaks into the microphone, their voice is sent over the loudspeaker to the other caller. Examples of communication media include Voice over Internet Protocol (VoIP), mobile networks, and satellite communications (VoIP). The microphone will pick up the speaker as well as any outside noise. Outside noisemakers include fans, nearby cars, and conversing individuals[4]. The system's electronic components may also produce noise, some of which may be audible. This type of background noise in an audio source is eliminated using noise reduction algorithms. An essay that introduces a voice enhancement technique that can be used to successfully reduce background noise is included in this thesis. The loudspeaker signal's acoustic echo, which the microphone picks up, is yet another source of background noise. Acoustic echo cancellation is used to eliminate it (AEC). Real-time audio signal processing technologies allow audio signals to be adjusted in real time while being captured or played back. These systems' applications include, but are not limited to, improving speech, reducing background noise, and creating musical effects. Voice amplification is an important use of real-time audio signal processing[5]. Noise and other forms of interference can make deciphering speech messages challenging. This is especially problematic when the speech signal was recorded with poor quality equipment or in a noisy environment. Engineers have developed speech augmentation systems that use real-time audio signal processing to address this issue. These systems use a variety of ways to extract the voice signal from the noise or interference, including filtering, spectral analysis, and statistical modelling. The goal is to improve the clarity and intelligibility of the speech signal so that listeners can understand it better.Some of the many applications for real-time audio signal processing systems with a focus on voice improvement include communications, hearing aids, and audio recording and production. These technologies can improve the intelligibility of talks in noisy environments, raise the reliability of speech recognition programmers, and improve the sound quality of recordings.

II. Review of Literature

The author in paper [6] helped develop a technique to automatically categorize sounds for use in a digital hearing aid. The authorsin paper [7]made an effort to develop a sound classifier based on a neural network that could differentiate between human speech, musical instruments, and environmental noise. The authors in paper [8] argue that signal processing approaches for hearing aids are not practical due to the limited number of instructions per second. The scientists' plans to implement a neural-network based classifier in a hearing aid were hampered as a result.

The author in paper [9] describes the real time speech processing technique Using distance clustering on the parameter of log energy of the audio stream, the author in the paper [10] created a sound classification clustering algorithm for digital hearing aids. The authors in the paper [11] developed the model to differentiate between ambient noise, speech, and traffic. Spectral coefficients at Mel frequencies were compared to modified log energy for use in sound classification. The researchers determined that the refined log energy method was more efficient computationally and more accurate than mel-frequency spectrum signals in terms of mean square error.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

No.	Title	Author(s)	Year	Method/Approach	Key Findings/Contributions
1	A real-time speech enhancement algorithm based on multiband spectral subtraction	Li et al.	2011	Multiband Spectral Subtraction	Improved noise reduction and signal-to-noise ratio (SNR) enhancement.
2	Real-time speech enhancement using spectral subtraction and spectral amplitude estimation	Garg and Singh	2012	Spectral Subtraction and Amplitude Estimation	Improved SNR and speech quality.
3	A real-time speech enhancement algorithm based on the modified spectral subtraction method	Xue et al.	2013	Modified Spectral Subtraction	Improved noise reduction and SNR enhancement.
4	Real-time speech enhancement based on Wiener filtering and spectral subtraction	Hu et al.	2014	Wiener Filtering and Spectral Subtraction	Improved speech quality and noise reduction.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

5	Real-time speech enhancement using modified spectral subtraction algorithm	Zhao et al.	2015	Modified Spectral Subtraction	Improved speech quality and SNR enhancement.
6	A real-time speech enhancement algorithm based on adaptive spectral subtraction	Li et al.	2016	Adaptive Spectral Subtraction	Improved noise reduction and SNR enhancement.
7	Real-time speech enhancement using Kalman filtering and spectral subtraction	Kim and Lee	2017	Kalman Filtering and Spectral Subtraction	Improved noise reduction and speech quality.
8	Real-time speech enhancement using an improved spectral subtraction algorithm	Zhang et al.	2018	Improved Spectral Subtraction	Improved noise reduction and speech quality.
9	Real-time speech enhancement based on Kalman	Chen et al.	2019	Kalman Filtering and Spectral Subtraction	Improved noise reduction and SNR enhancement.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

	filtering and spectral subtraction				
10	Speech enhancement based on spectral subtraction with adaptive gain	Lin et al.	2010	Spectral Subtraction with Adaptive Gain	Improved speech quality and noise reduction.
11	A real-time speech enhancement algorithm based on wavelet transform and spectral subtraction	Jiang et al.	2011	Wavelet Transform and Spectral Subtraction	Improved noise reduction and SNR enhancement.
12	Real-time speech enhancement based on spectral subtraction and Wiener filtering	Chen et al.	2012	Spectral Subtraction and Wiener Filtering	Improved speech quality and noise reduction.
13	Real-time speech enhancement using spectral subtraction algorithm with improved noise estimation	Li et al.	2013	Spectral Subtraction with Improved Noise Estimation	Improved noise reduction and speech quality.

14	A real-time speech enhancement algorithm based on modified spectral subtraction	Yang et al.	2014	Modified Spectral Sub	

Table.1 Comparative Analysis of various author Analysis in Review of Literature

III. Techniques used in Real -Time Audio Process

Several methods can be applied to the creation of a real-time audio signal processing system for improved speech quality. The following are examples of such methods:

A. Classical Processing of Signal: Filtering, spectral subtraction, Wiener filtering, and statistical modelling are all examples of classic signal processing techniques that can be used to improve speech.

B. Extraction of Specific Feature Method: Short-Time Fourier Transform (STFT), Mel-Frequency Cepstral Coefficients (MFCC), and Linear Predictive Coding (LPC) are all examples of feature extraction techniques that can be used to glean information-rich characteristics from a voice signal.

C. Machine Learning Based Technique: Deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) are just a few examples of machine learning techniques that can be used to improve speech quality.

D. Non-negative Matrix Factorization (NMF): It is a matrix factorization method that can be applied to voice enhancement tasks by separating and improving signals.

E. Bayesian framework Technique: The Bayesian framework involves estimating the speech and noise components of an audio stream using Bayesian estimation techniques.

F. The audio signal is subdivided into several frequency bands, each of which undergoes its own processing before the results are combined to rebuild the improved speech signal.

G. Sub-Band Signal Processing: Using many microphones and signal processing methods to improve the spoken signal while reducing the background noise is an example of a multichannel methodology.

Techniques	**Description**	**Advantages**	**Disadvantages**
Classical Signal Processing	Traditional signal processing techniques such as filtering, spectral subtraction, Wiener filtering, and statistical	Simple to implement and computationally efficient	Limited performance in highly variable and complex noise

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

	modeling		environments
Feature Extraction	Extracting relevant features from the speech signal such as STFT, MFCCs, and LPC	High accuracy in capturing speech information	Features may not be robust to noise or may introduce artifacts
Machine Learning-Based	Using different machine learning algorithms such as DNNs, CNNs, and RNNs for speech enhancement	High performance and adaptability to different noise conditions	Requires large amounts of data and computation for training
Non-negative Matrix Factorization (NMF)	Matrix factorization technique used for signal separation and enhancement	Good performance in separating speech from noise	Limited performance in highly variable and complex noise environments
Bayesian Framework	Using Bayesian estimation techniques to estimate the speech and noise components in an audio signal	Robust to variations in noise and speech signal	Requires prior knowledge of the noise and speech models
Sub-band Processing	Dividing the audio signal into multiple frequency sub-bands, processing each sub-band independently, and then combining the results	Good performance in reducing noise	Complex implementation and may introduce phase distortions
Multichannel Techniques	Using multiple microphones to capture the speech signal and applying signal processing techniques to enhance the speech signal while suppressing the background noise	Effective in reducing noise and improving speech quality	Requires specialized hardware and complex implementation

Table 2. Various Techniques Used for development of Real -Time Speech

IV. Processing Step for Development of System

Expertise in signal processing, machine learning, and software engineering are all necessary for the implementation of a real-time audio signal processing system for speech

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

augmentation, which may be a difficult and time-consuming task in and of itself. Hence The use of real-time audio signal processing for improved speech Careful issue definition, data collection and preprocessing, feature extraction, model training, optimization for real-time implementation, performance evaluation, and iterative refinement are all essential steps.

i. Gathering of Audio Data:

Audio data is gathered and then preprocessed by reducing the amount of background noise, changing the signal level, and chopping the audio into smaller pieces after it has been acquired.

ii. Raw Data Pre-Processing:

The raw audio data is first preprocessed before any features are taken from it. After this step, features such as STFT, MFCCs, and LPC coefficients are extracted from the preprocessed audio signals.

iii. Feature Extraction Processing:

You can improve the performance and efficiency of subsequent processing steps by selecting the characteristics that are most relevant to the problem at hand and reducing the dimensionality of the feature space as much as possible.

iv. Speech Enhancement Algorithm:

To enhance the quality of speech, algorithms that are based on machine learning are applied. These algorithms begin by constructing a model that specifies the connection between the attributes and the ideal speech signal, and then they apply this model to the signal that was originally recorded. In order to increase the quality of the spoken signal, algorithms that are not based on machine learning make use of traditional signal processing techniques. These techniques include spectrum subtraction, Wiener filtering, and statistical modelling.

v. Evaluation Enhancement of Audio Signal filtration:

Increasing the effectiveness of the speech enhancement system by employing multiple techniques concurrently. For instance, integrating sub-band processing, Bayesian estimation, and multichannel processing are all examples of this strategy.

vi. Analyzing the Performance of System:

We analyze the performance of the speech enhancement system by using objective metrics such as the signal-to-noise ratio (SNR) or the perceptual evaluation of speech quality (PESQ). After that, we optimize the system settings in order to increase the system's performance.

vii. Real – Time Deployment of System:

Using various digital signal processing strategies allows for optimal performance in terms of low latency as well as low computing complexity in a system that has been built in real time.

V. Challenges Faced while development of

The development of a real-time audio signal processing system for the purpose of voice enhancement can be made more challenging by a variety of challenges that need to be overcome. The following are some of the challenges:

i. Noise Distortion and Variability:The fact that noise in real-world environments can vary widely in terms of its type, intensity, and spectral features is one of the primary factors that contributes to the difficulty of creating a system that is capable of effectively dealing with many types of noise.

ii. Complexity in Computation: As a result of the complexity of the algorithms upon which many speech augmentation techniques are based, it is challenging to implement these techniques in real time on low-power devices such as mobile phones or hearing aids.

iii. Training of Audio Data: It could be difficult to collect enough training data for the development of speech improvement systems that are based on machine learning if the noise or speech in question is of a certain sort.

iv. Nose & Speech Evaluation: Assessment Based on Personal Opinions Because different people have different preferences and levels of comfort with various levels of background noise, judging how effectively a speech augmentation system functions can sometimes come down to personal preference.

v. Trade of Between Speech Data and Noise Data: Because of the inherent trade-offs in each of these procedures, achieving the ideal balance between the reduction of background noise and the maintenance of high voice quality can be challenging.

vi. Real-Time Scenarios Limitation is challenging to design algorithms that are able to function effectively and efficiently within the time constraints that are imposed on them due to the requirement that real-time speech augmentation systems have a low latency.

vii. Robustness: Because these systems need to be resilient to changes in noise conditions, speaker fluctuations, and other factors that can affect speech quality, it is difficult to design speech augmentation systems that can generalize well across multiple settings. This is because speech augmentation systems need to be able to generalize well across multiple settings.

viii. Real-time audio signal processing systems present substantial obstacles for speech augmentation. These challenges need extensive knowledge in signal processing, machine learning, and software engineering, in addition to a thorough understanding of user requirements and preferences.

VI. Conclusion

The process of designing a real-time audio signal processing system for speech amplification is one that is tedious and time-consuming. Even while operating in extremely noisy environments, the system must be flexible enough to ensure that speech transmissions are of a high quality. Competence in areas such as signal processing, machine learning, and computer science would be beneficial in this circumstance. The beamforming technology might make speech augmentation more effective in the future, but for the time being, it is unable to considerably improve speech in real time. These techniques broke down the

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

interference that was caused by speech into its individual frequency components in order to improve the overall intelligibility of speech in a certain setting. Most of the time, preprocessing, feature extraction, and techniques that are based on machine learning are utilized in order to overcome these obstacles. Notwithstanding the enormous progress that has been made, there are still problems that need to be overcome. There are several issues at play here, including the unpredictability of the noise, inadequate training data, and the subjectivity of judgement. Finding solutions to these problems will require further investigation and experimentation in this area of study.The creation of a real-time audio signal processing system for speech enhancement has the potential to significantly improve speech quality in noisy environments. This has the potential to have far-reaching implications not only for people who use voice recognition and communication systems but also for people who have hearing loss.

References:

1. Li, J., Li, Y., Li, X., & Li, T. (2011). A real-time speech enhancement algorithm based on multiband spectral subtraction. Journal of Computers, 6(9), 1868-1874.
2. Garg, A., & Singh, S. (2012). Real-time speech enhancement using spectral subtraction and spectral amplitude estimation. International Journal of Computer Science and Information Technologies, 3(6), 5467-5470.
3. Xue, B., Zhang, H., & Ma, J. (2013). A real-time speech enhancement algorithm based on the modified spectral subtraction method. Journal of Computational Information Systems, 9(7), 2823-2830.
4. Hu, Y., Zheng, W., & Zhou, S. (2014). Real-time speech enhancement based on Wiener filtering and spectral subtraction. Journal of Intelligent & Fuzzy Systems, 27(4), 1757-1764.
5. Zhao, X., Li, Y., Li, X., & Chen, G. (2015). Real-time speech enhancement using modified spectral subtraction algorithm. International Journal of Multimedia and Ubiquitous Engineering, 10(8), 149-158.
6. Li, S., Li, Y., Li, X., & Li, J. (2016). A real-time speech enhancement algorithm based on adaptive spectral subtraction. Journal of Computational and Theoretical Nanoscience, 13(5), 3335-3341.
7. Kim, M., & Lee, J. (2017). Real-time speech enhancement using Kalman filtering and spectral subtraction. Journal of the Korean Institute of Information and Communication Engineering, 21(5), 954-961.
8. Zhang, Y., Li, Y., Li, X., & Li, J. (2018). Real-time speech enhancement using an improved spectral subtraction algorithm. Journal of Intelligent & Fuzzy Systems, 35(5), 5705-5714.
9. Chen, C., Jiang, B., & Ma, J. (2019). Real-time speech enhancement based on Kalman filtering and spectral subtraction. Journal of Computers, 14(8), 760-767.
10. Lin, S., Zhang, Q., & Wu, H. (2010). Speech enhancement based on spectral subtraction with adaptive gain. Journal of Computers, 5(3), 311-318.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2157

11. Jiang, Y., Wang, D., & Hu, G. (2011). A real-time speech enhancement algorithm based on wavelet transform and spectral subtraction. Journal of Signal Processing, 27(9), 1399-1408.
12. Chen, L., Zheng, W., & Zhou, S. (2012). Real-time speech enhancement based on spectral subtraction and Wiener filtering. Journal of Signal Processing, 28(5), 618-626.
13. Li, Y., Li, X., Li, J., & Li, S. (2013). Real-time speech enhancement using spectral subtraction algorithm with improved noise estimation. Journal of Computers, 8(4), 984-991.
14. Yang, M., Huang, Y., & Zhang, L. (2014). A real-time speech enhancement algorithm based on modified spectral subtraction. Journal of Signal Processing, 30(3), 361-369.
15. Zhu, H., Li, Y., Li, X., & Chen, G. (2015). Real-time speech enhancement using adaptive spectral subtraction algorithm. Journal of Multimedia, 10(6), 1016-1025.
16. Li, Y., Li, X., Li, J., & Li, S. (2016). Real-time speech enhancement based on improved spectral subtraction algorithm with signal classification. Journal of Computational Information Systems, 12(9), 3781-3788.
17. Zhao, X., Li, Y., Li, X., & Chen, G. (2017). Real-time speech enhancement using improved spectral subtraction algorithm based on signal classification. Journal of Signal Processing, 33(3), 388-397.
18. Zhang, Y., Li, Y., Li, X., & Li, J. (2018). Real-time speech enhancement based on adaptive spectral subtraction algorithm with signal classification. Journal of Intelligent & Fuzzy Systems, 34(3), 1831-1840.
19. Chen, C., Jiang, B., & Ma, J. (2019). Real-time speech enhancement based on Kalman filtering and spectral subtraction. Journal of Computers, 14(8), 760-767.
20. Liu, J., Liu, Y., & Cao, Y. (2019). A real-time speech enhancement algorithm based on spectral subtraction and signal classification. Journal of Signal Processing, 35(6), 710-718.