International Journal of Speech Technology
https://doi.org/10.1007/s10772-020-09674-2

Fundamentals, present and future perspectives of speech enhancement

Nabanita Das[1] · Sayan Chakraborty[2] · Jyotismita Chaki[3] · Neelamadhab Padhy[1] · Nilanjan Dey[4]

Received: 7 December 2019 / Accepted: 10 January 2020

Abstract
Speech enhancement has substantial interest in the utilization of speaker identification, video-conference, speech transmission through communication channels, speech-based biometric system, mobile phones, hearing aids, microphones, voice conversion etc. Pattern mining methods have a vital step in the growth of speech enhancement schemes. To design a successful speech enhancement system consideration to the background noise processing is needed. A substantial number of methods from traditional techniques and machine learning have been utilized to process and remove the additive noise from a speech signal. With the advancement of machine learning and deep learning, classification of speech has become more significant. Methods of speech enhancement consist of different stages, such as feature extraction of the input speech signal, feature selection, feature selection followed by classification. Deep learning techniques are also an emerging field in the classification domain, which is discussed in this review. The intention of this paper is to provide a state-of-the-art summary and present approaches for using the widely used machine learning and deep learning methods to detect the challenges along with future research directions of speech enhancement systems.

Keywords Speech degradation · Backgroundnoise · Traditional speech enhancement techniques · Machine learning · Speech enhancement applications · Clustering · Classification · Deep learning

1 Introduction

Enhancement of speech is an essential requirement in the domain of speech signal processing. Speech quality refers to a speech (Paliwal 2003) sample being (i) clear, (ii) pleasant, (iii) compatible with other speech processing methods. The main idea behind speech enhancement is removal of background (Giacobello et al. 2005) noise. Echo suppression is another key area which needs to be addressed during speech (Faúndez-Zanuy et al. 2002) enhancement. Speech recorded in natural environment includes background noises as well as echo. Although, echoless speech is captured in a special anechoic room. It sounds dry and dull to a human ear. Echo suppression is also required for speech samples collected from big halls or house. The speech may pick some echo up if the distance between the speaker and microphone is large.

Speech enhancement is also required for telephonic conversation. In such cases, speech enhancement demands real-time processing (Krishnamoorthy and Mahadeva Prasanna 2008), in order to get quality sound from the speaker. Current telephone networks speech is band-limited from 300 to 3400 Hz. With the development of recent trends, the

✉ Nabanita Das
nabanita.das2008@gmail.com

Sayan Chakraborty
sayan.cb@gmail.com

Jyotismita Chaki
jyotismita.c@gmail.com

Neelamadhab Padhy
dr.neelamadhab@giet.edu

Nilanjan Dey
neelanjan.dey@gmail.com

[1] School of Computer Engineering (SOCE), GIET University, Gunupur, India

[2] Department of CSE, Ideal Institute of Engineering, Kalyani, India

[3] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

[4] Department of IT, Techno International New Town (Formerly known as Techno India College of Technology), Kolkata, India

Published online: 22 January 2020

band-limit may get increased up to 7500 Hz or even higher. People will be able to have a telephonic conversation then, even though they will be farther from the telephone network. In these cases, speech enhancement will be needed as well, as the concept is quite similar as of speaker and microphone standing in a long distance, and then possibility of background noise and echo will be higher than expected.

Background noise suppression (Vijayan et al. 2018; Christiansen et al. 2007) is needed for cases where people are having a telephone conversation in a street or noisy environment (Santos et al. 2019). Background noise suppression (Deshmukh and Espy-Wilson 2007) can be seen when the pilot of airplane sends speech signals from the cockpit to the ground or cabin.Enhancement of speech can also be seen for hearing-aids. For hearing-aid devices, a significant amount of work has been done on speech enhancement. The application of speech enhancement goes beyond (Mustière et al. 2010) real life examples, as it can be observed during criminal investigation also, as the speech enhancement plays a vital role in removing background noise from speech sample, and classifying or identifying the target speech. Speech enhancement also can be seen in some speech recognition system, as the system tends to enhance the quality of the speech prior to feature extraction, selection and classification (Baumgarten et al. 2013). Overall, it can be said, speech enhancement has a wide-range of applications, and it's necessary for almost all the devices or systems or methods related to speech signals.

There are two main conceptual requirements for assessing a speech enhancement system's performance. The enhanced speech signal quality analyses in that signal its distorted nature, clarity, and residual noise level. Quality is a subjective assessment of how happy the listener is with the improved speech signal. The second standard examines the improved signal intelligibility. This is an objective assessment that offers the percentage of terms which listeners (Sen et al. 2019a, b, c) will correctly identify. There's no need to make sense of the words in this study. There is no connection between the two performance measures. Poor intelligibility and good quality and may be properties of a speech signal, and vice versa. Many speech enhancement systems increase signal quality to the detriment of increasing their intelligibility. By carefully listening to that speech signal, listeners will typically gain more details of the noisy speech signal (Santosh et al. 2019) than from the enhanced signal. This is apparent from the theorem of information theory on data processing. Nonetheless, listeners experience tiredness during lengthy listening sessions, which results in the fragmented message being less intelligible. If such situations occur, the improved signal intelligibility may be higher than the noisy signal. Usually the listener would need less effort to decode parts of the enhanced speech signal that conform to segments of the noisy signal's high signal-to-noise ratio.

The rest of this paper is organized as follows. In Sect. 2, we briefly introduced speech signal characteristics and their requirements. In Sect. 3, we briefly described speech channel artifacts and various sources of noise which are present in the speech, in Sects. 4, 5 and 6 we review methods in following aspects: speech enhancement techniques, applications of speech enhancement and challenges respectively. Lastly, in Sect. 7 conclusions and future directions of speech enhancement systems are deliberated.

2 Speech signal characteristics and requirements

Speech signal requires speech processing. Speech processing is the analysis of speech signals and signal processing processes. In a digital form, the signals are normally interpreted, and speech processing can be viewed as a specific case of digital signal processing, adapted to speech signals. Speech recognition aspects involve collecting, handling, preserving, transmitting and distributing speech signals. The input is named as recognition of speech and the result is termed as synthesis of speech. In early, 2002, Faúndez-Zanuy et al. presented an overview of non-linear (Faúndez-Zanuy et al. 2002) speech process. In this work they showed the applications of various non-linear speech processing techniques. In 2003, Paliwal discussed the aspects of speech processing in different phases. In this paper, he explored human speech (Paliwal 2003) perceptions with respect to the phase information, and tried to anticipate the usefulness of it. This work involved short-time phase spectrum in human speech perception and automated speech recognition. Later in 2008, Giacobello et al., introduced sparse linear predictors for speech processing. In this work, they introduced two new classes of linear prediction (Giacobello et al. 2005) scheme. The paper also stated that the algorithms proposed may not be restricted to only speech processing as it can be used for several linear prediction problems too. The research work on speech signal processing later broke into several domains as specific domains wanted to focus on specific characteristics of speech signals. Figure 1 shows the classification of speech signal characteristics and requirements. Based on the characteristics of speech signals, there are two categories of speech signal features.

2.1 Temporal features

Temporal features (time domain features) that are easy to extract and easy to interpret, such as: short time zero crossing rate, autocorrelationand short time energy.

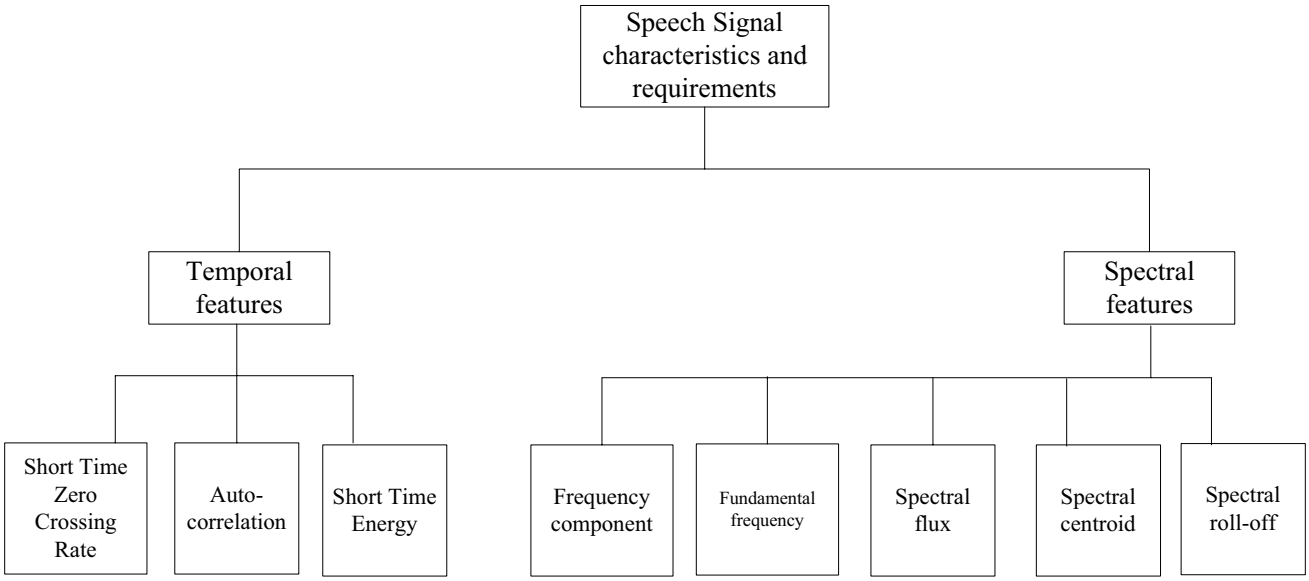

Fig. 1 Overview of speech signal characteristics

2.1.1 Short time zero crossing rate

The most popular feature of the speech signal is short time zero crossing. It is defined by point where a graph of speech signal shifts the positive sign to the negative sign. The zero crossing is specified as number of times in a speech signal alters the amplitude of the sign wave. Zero crossing rates (Bachu et al. 2010) are utilized to classify speech and end point identification of unvoiced and voiced. Zero crossing value is large when silence or unvoiced speech occurs in speech signal. It is calculated by measuring the number of times the amplitude of speech signals in a given time interval passes thru a value of zero. Short time zero crossing rates will be calculated by Eq. (1).

$$Z_m = \sum_{n=-\infty}^{\infty} |\mathrm{sgn}[x(n)] - \mathrm{sgn}[x(n-1)]| W(m-n) \quad (1)$$

where Z_m is an energy (Jalil et al. 2013) at sample m of the x signal, n is the number of speech signal frames and W is the window.

2.1.2 Autocorrelation

Autocorrelation is a method of system analysis and is also referred to as sequential correlation.It is calculated by the time series correlation (Ando 2013) compared and the identification of resemblances between past and future speech values. It is utilized to track the speech signals' repetition or periodicity. This is calculated by using Eq. (2).

$$A_m(p) = \sum_{n=0}^{M-1-p} [x(m+n)W(n)]\left[x(m+n+p)W(n+p)\right] \quad (2)$$

where A_m is an energy at sample m of the x signal, n is the number of speech (Ando 2013) signal frames and W is the window.

2.1.3 Short time energy

Short-time energy is a fundamental and important function of speech processing. Energy is characterized as signal strength in terms of time is naturally different in terms of energy. Short time (Jalil et al. 2013) analysis is utilized to assess the speech signal. In general, voicing, invoicing, noise regions and silence are the speech signals. Analyzing the energy-based speech signal will have higher pressure to classify certain regions of a speech signal. Voiced parts of a speech signal will have higher short-time energy and will be small when speech is not voiced. If speech signal is silent, it is too weak. Short time energy (Jalil et al. 2013) is the opposite mechanism of zero crossing rates. The Eq. (3) is used to measure it.

$$E_m = \sum_{n=-\infty}^{\infty} (x(n)W(m-n))^{-1} \quad (3)$$

where E_m is an energy at sample m of the x signal, n is the number of speech signal frames and W is the window.

From the discussion it is quite clear (Kesarkar 2003) that temporal feature analysis is easy to apply and has less hassle

in terms of computation. Although temporal analysis (Kesarkar 2003) is only capable of extracting simple features like power, energy, period etc. from speech signal. It lacks determination simple speech parameters (Santos et al. 2019) like power, energy and periodicity of speech. Also it is incapable of fault diagnostic (Gupta et al. 2019a, b) or analysis of such system in case of speech enhancement. Spectral feature analysis (Kesarkar 2003) delivers valuable information about different aspect of the speech such as vocal tract and pitch. Although this kind of analysis has proven to be more intensive, computationally. In addition, the relative equations for temporal features are presented in Table 1.

2.2 Spectral features

Spectral features (frequency-based features), which are obtained by transforming the time-based signal into the frequency domain using the Fourier Transform, like: frequency components, fundamental frequency, spectral flux, spectral centroid and spectral roll-off. Such features can be utilized to define the sounds, rhythm, pitch, and melody.

2.2.1 Frequency component

A speech signal has frequency components that fail at high frequencies. In speech signals, the energy of the high-frequency components is generally low. They may therefore not be able to carry sufficient energy (Kulkarni and Bairagi 2018) to yield useful features. Preemphasis is utilized to boost the energy of components with high frequency. Suppose the analog speech signal is digitized into d discrete samples, referred to as $s(1), s(2), \ldots, s(d)$. The preemphasized signal, $\hat{s}(d)$, of the input signal $s(d)$ is represented by Eq. (4).

$$\widehat{s}(d) = s(d) - \alpha s(d-1) \tag{4}$$

where $0 < \alpha < 1$ and d is the number of samples.

2.2.2 Fundamental frequency

The fundamental frequency of speech signals is produced by quasi-periodic oscillations in vocal folds induced by airflow from the lungs. The fundamental frequency corresponds to speed of oscillations and is therefore a measurement of the physical phenomenon. In speech, the range of the fundamental frequencies (Kulkarni and Bairagi 2018) are roughly in $F_0 \in 80\ \text{Hz} \ldots 400\ \text{Hz}$.

Table 1 Temporal feature formulations

Spectral feature	Equation
Short time zero crossing rate	$Z_m = \sum_{n=-\infty}^{\infty} \lvert \text{sgn}[x(n)] - \text{sgn}[x(n-1)] \rvert W(m-n)$
Autocorrelation	$A_m(p) = \sum_{n=0}^{M-1-p} [x(m+n)W(n)][x(m+n+p)W(n+p)]$
Short time energy	$E_m = \sum_{n=-\infty}^{\infty} (x(n)W(m-n))^{-1}$

2.2.3 Spectral flux

The spectral fluxis calculated by the signal's power spectrum change and is calculated by comparing the power spectrum. To separate the speech signal (Al-Shoshan 2006) from the music form this is the most important feature. It is also known as a square difference among two uniform magnitudes of successive spectral distribution, representing the subsequent signal frames. This is computed by the Eq. (5).

$$SpectralFlux_p = \sum_{m=1}^{N/1} \left[\left| x_p(m) \right| - \left| x_p(m-1) \right| \right]^{-1} \tag{5}$$

where $x(m)$ is the coefficient of discrete Fourier transform of p-th frame of length N.

2.2.4 Spectral centroid

This is correlated with speech brightness measurements(Al-Shoshan 2006) that characterizes the spectrum. A spectrum-specific (Kulkarni and Bairagi 2018) measure is the 'centre of gravity' utilized by Fast Fourier Transform frequency- and information on the magnitude. It is computed by the weighted average frequency of amplitudes divided by the amplitude number. The spectral centroid is computed by the Eq. (6).

$$SpectralCentroid = \frac{\sum_{n=0}^{M-1} c(n) * n}{\sum_{n=0}^{M-1} x(n)} \tag{6}$$

where $c(n)$ signify the center frequency (Kulkarni and Bairagi 2018) of the n bin with length M and $x(n)$ is an amplitude spectrum of discrete Fourier transform of the n bin or weighted frequency value.

2.2.5 Spectral roll-off

This is a measure if the skewness of the speech spectralshape. It is utilized to differentiate voiced from speech (Al-Shoshan 2006) and music that is not spoken (Unvoiced speech has a high energy content in the high frequency range of the spectrum). This can be represented by using Eq. (7).

$$SpectralRollOff = 0.85 \times \sum_{n=0}^{M-1} c(n) \tag{7}$$

where $c(n)$ signify the center frequency of the n bin.

Some of the research work included both temporal and spectral features of speech signals. Christiansen et al., presented a theoretical approach for speech processing using spectro-temporal processing.In this work, they combined both of the frequency domains: spectral and temporal domain in order to increase the quality of low-frequency speech signals. In early 2008, Krishnamoorthy et al., analyzed temporal and spectral processing (Krishnamoorthy and Mahadeva Prasanna 2008) of degraded speech signals. In this work, they included the degraded speech signal and in order to enhance them, they explored both temporal and spectral domain approach. In 2010, Vijayan et al. discussed speech signals (Vijayan et al. 2010) for temporal alignment. In this paper, they analysed not only speech signals, but also the singing signals and their characteristics to identify the common features of those signals in temporal alignment. The formulations related to various spectral features are shown in Table 2.

3 Speech channel artifacts and sources of noise

Speech enhancement always tends to stumble upon the noise or multiple problems with sources or channel or path of the transmission.These problems are the reason behind the introduction of speech enhancement technique. Such challenges and problems of speech enhancement are discussed below and shown in Fig. 2.

3.1 Speech degradation

Prior to reaching the listener, speech can get completely corrupted by noise at any stage. The speech degradation has mainly these three following ways.

Table 2 Spectral feature formulations

Spectral feature	Equation
Frequency component	$\hat{s}(d) = s(d) - \alpha s(d-1)$
Fundamental frequency	$F_0 \in 80\ \text{Hz} \ldots 400\ \text{Hz}$
Spectral flux	$SpectralFlux_p = \sum_{m=1}^{N/1} \left[\lvert x_p(m) \rvert - \lvert x_p(m-1) \rvert \right]^{-1}$
Spectral centroid	$SpectralCentroid = \frac{\sum_{n=0}^{M-1} c(n) * n}{\sum_{n=0}^{M-1} x(n)}$
Spectral roll-off	$SpectralRollOff = 0.85 \times \sum_{n=0}^{M-1} c(n)$

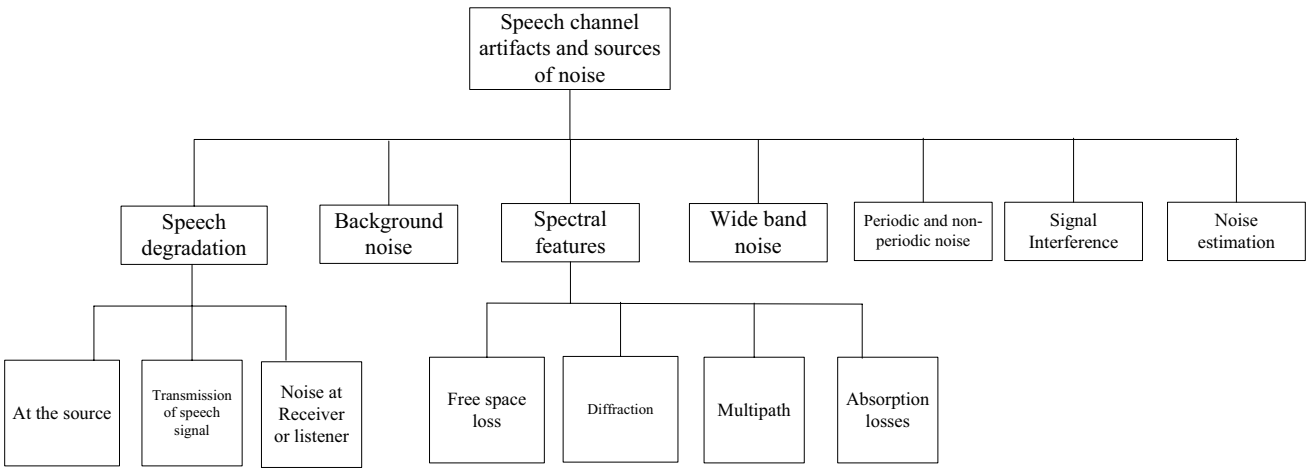

Fig. 2 An overview of speech channel artifacts

3.1.1 At the source

In this case, the source of the speech is itself in a noisy environment. A lot of background noise may get added that can't be avoided. Such cases may include examples such as: talking inside a noisy cockpit, or talking inside a vehicle while other vehicles are also adding noises (Sen et al. 2019a, b, c), or the noise can come from nature or environment also, as animal or rain sound may get overlapped with speech. This kind of noise can also occur due to room reverberation. The noise gets overlapped with the speech signal at the source in this case.

3.1.2 Transmission of speech signal

Transmission of signals happens through channels. Channels are nothing but the medium as it can be wired or wireless. While talking on the phone, it's wired channel. In this case, noise may get mixed with the speech during transmission. The main culprit of such noises is the channels itself. As the channel's weird behaviour due to some reason can cause such speech degradation (Sen et al. 2019a, b, c) during transmission. This can happen while converting speech into data or vice-versa, as the noise can be added then also.

3.1.3 Noise at receiver or listener

It's not always necessary to have a noiseless receiver, as the noise can get added at the listener's end also. But such cases are very rare as researchers tend to be careful during speech enhancement (Sen et al. 2019a, b, c) as it can affect large chunk of database. Although from a human perspective, listener's fatigue may get increased that lead to quality of speech getting low quality.

Hence, it can be observed that speech degradation is possible at any stage of speech signal. The important aspect is to identify that particular stage, and take necessary steps in order to execute speech enhancement successfully.

3.2 Background noise

Background noise or more specifically the environmental noises, plays an integral part in degradation of speech. Most of the speech enhancement techniques face challenges of background noises as the speech quality becomes poor with an effective background noise. Even the communication devices such as mobile (Foth et al. 2013) phones or recording devices like CCTV also suffer due to background noise, as researchers try to use such samples as databases. Signal to noise ratio (SNR) is widely popular technique to identify such background noise in a speech sample. Previously lot of work has been done in order to reduce the background noise. Background noise was identified by researchers as the main culprit of low-quality sound; hence a lot of work has revolved around to address this problem. Manohar and Rao et al., discussed a speech enhancement (Manohar and Rao 2006) technique, in 2005, that dealt with the background noise of the environment. This work used voice activity detector (VAD) for noise estimation. This work also involved STSA or short-time spectral attenuation to estimate the boundaries of VAD. In 2010, Shrawankar and Thakare proposed a noise estimation (Shrawankar and Thakare 2010) and removal mechanism that would work in adverse condition of the environment. Their work mainly focused on VAD for noise estimation and used MMSE for noise removal. A more recent advance of removal of background noise includes the work of Atmaja et al. as they proposed a speech enhancement (Atmaja et al. 2016) technique for the voice recording that used smartphone as the capturing device. This work included MMSE, STSA and NMF (Non-negative matrix factorization) in order to remove the background noise during smart phone devices' capturing audio.

3.3 Channel loss/path loss

A channel refers to any medium for communication. In general case, while capturing the speech, the air acts as the channelling device to fetch the signal data into the receiver end of the recording device. If the medium is congested with lot of nano-particles that may affect the signal. This is known as the channel loss, which is also another challenge for speech enhancement. The same can happen for wired communication also. During the study, the following has been observed as the causes that lead to channel loss.

3.3.1 Free space loss

If the coverage area increases the channel for signal to transmit increases also. The signal may get lost in the process, as the free space doesn't provide the boundary to limit the signal to transmit through a proper path. The signal has to cover a wider area becomes a headache. An example of such case may include radio communication signals breaking when sphere is increased.

3.3.2 Diffraction

If there is an object in the path of the transmission then diffraction of signal may occur. The signal may get diffracted (Flamand et al. 2016) around the object, but losses occur. This loss gets increased when the area of such object is more rounded. An example of such case can be seen during radio transmission, as it works better around sharp edges.

3.3.3 Multipath

Signals get reflected and reaches the receiver while it's inside a real terrestrial environment. Although such movement (Treichler and Agee 1983) of signal may affect itself and losses may occur. A mobile phone receiver is one good example of this.

3.3.4 Absorption losses

If the channel is not fully transparent or semi-transparent, then absorption loss happens. The non-transparent or semi-transparent channel may include atmosphere moist, glass, polluted air etc. as the particles inside these things tends to absorb a small portion of the speech signal, that leads to signal's quality getting decreased.

The Channel or path loss is not one of the major concerns, but it affects the speech signal which is the bottom-line. Hence, this should be addressed during speech enhancement.

3.4 Wide band noise

Speech enhancement also faces a massive task in the term of wideband as wideband refers to the incident of message's bandwidth going beyond the coherence bandwidth (or maximum bandwidth allowed) of the channel. A significant amount of work has been done in order to eradicate this problem during speech enhancement. Early in 2007, Deshmukh et al., introduced MPO or Modified Phase-Opponency (Deshmukh and Espy-Wilson 2007) technique. The prime objective of this proposal was to enhance the speech by removing the additive noise. This model was also able to extinguish between the wideband noise and narrowband noise. In 2010, Mustière et al., proposed a novel technique (Mustière et al. 2010) that can adapt the bandwidth extension dynamically so that speech enhancement doesn't face the issue of wideband noise. They used SNR to measure the changes of noise amount following the extension of bandwidth. An interesting piece of work done by Zhang and Dong in 2012, reported that it was possible to separate uncorrelated (Zhang et al. 2012) wide-band sound sources with the help of a hierarchical approach. In this work, they included the wideband noises that overlap each other, uncorrelated, in space and frequency domains.

3.5 Periodic and non-periodic noise

Speech or sounds are basically signals. The signal that fails to repeat itself after a specific period is called non-periodic signals. On the contrary, if a signal manages to repeat itself, then it will be called as periodic signal. This phenomenon is called as the periodicity of signal.This is an integral property of speech signal, as it plays a vital role in order to achieve fundamental frequency and pitch of the voice in speech. Periodicity of a signal can be enhanced which ultimately leads to speech enhancement. In 2009, speech enhancement using periodic noise was done by Sasaoka et al. In this work, they overcame the issues created by the presence of periodic (Sasaoka et al. 2009) signal in a speech, using noise estimation filter. Later, Yoshizawa et al., used DFT for noise reduction (Yoshizawa et al. 2011) of periodic signals. They compared the results with the results of applying Non-harmonic analysis, as both were applied on periodic signal. In, 2015, Chen and Hohmann used periodicity degree (Chen and Hohmann 2015) In order to estimate signal to noise ratio (SNR). In 2016, Huang et al. proposed a periodicity enhancement for noisy speech. The enhancement was done on linear prediction (Huang et al. 2015) residual of speech.

3.6 Signal interference

Beam forming with microphone arrays has been commonly utilized in a variety of applications to improve speech signals of interest and eliminate noise and interference. Beaming usually means that the speech source of concern and the source of interference are incidental to the array from various directions in order to make it work. Due to signal interference, speech quality suffers badly. Hence, a lot of work has been done in order to remove this problem. In a work proposed by Arslan and Hansen, the problem of cross-talk interference (Arslan and Hansen 1997) was addressed. This work showed a substantial improves in terms of quality and intelligibility (Yu et al. 2009) of speech signal. Cohen and Berdugo proposed a speech enhancement technique that used beam to reference ratio.In this work they tried to eradicate the problem of acquiring (Khosravy et al. 2010) a noiseless speech while being in the car environment, as car horns or sounds (Chawla 2011) work as beam to interfere the main signal i.e. speech signal. Leng et al., later in 2010 tried to enhance the speech (Leng et al. 2018) quality by studying the beamforming (Khosravy et al. 2015) of the microphone. The work studied the source of speech (Gupta et al. 2019a, b) and interference source, and used the beamforming as an array from same or different direction.

3.7 Noise estimation

Noise estimation refers to the process of calculating the amount of noise or unnecessary data is present in the speech sample.This is an important part of speech enhancement, as Noise estimation helps to determine whether the obtained result is acceptable or not. The most commonly used noise estimation technique is known as the SNR or Signal to Noise ratio. SNR is basically, the ratio of a signal's power and noise's power. It's defined as shown in Eq. (6).

$$SNR = {}^{P_{Signal}}/_{P_{Noise}} \quad (6)$$

As previously stated, in order to enhance the speech, detecting the noise is necessary. Hence, a lot of research work has been done in this domain.In the early 2006, Scalart and Filho proposed a speech enhancement technique (Scalart and Vieira-Filho 2006) that was based on priori SNR estimation. This method was mainly introduced in order to enhance the speech taken by a single microphone device in a noisy environment. Later, Plapous et al. enhanced the Signal-to-noise ratio.In this work, they proposed a two-step noise reduction technique (TSNR) (Plapous et al. 2006) that addressed the problem of SNR depending on the previous frame's spectrum estimation. Recently, in 2016, Lee et al., used a little bit of different approach (Lee and Lee 2016) during noise estimation. In this work, they included a minimum-statistics method (MS) that used a non-linear function and an SAP or Speech Absence probability, in order to estimate the noise.

4 Speech enhancement techniques

The main task of speech enhancement is to remove the additive noise from a speech signal. Basically, speech intelligibility can be regarded as an aspect of quality, on the other hand, high-quality speech always gives good intelligibility.So, no speech enhancement systems can improve both speech quality and intelligibility. Over the past decades to enhance speech quality and intelligibility many methods have been placed, among which time-domain and frequency-domain algorithms are two prominent categories. The speech enhancement approaches basically have three main steps (Kalamani et al. 2014) in the methodology such as (i) feature extraction, (ii) feature selection, (iii) classification or clustering.

These techniques basically use classifiers in order to classify the speech signals, then remove noise and finally reconstruct the speech signals without noise. In this way, many researchers have proposed various algorithms that dealt with the noisy environment for recorded speech samples. Some of the significant techniques are discussed below.

This section mainly focused on various speech enhancement approaches include spectral subtraction (Boll 1979), Wiener filtering (Wang et al. 2010), statistical model-based methods (Ephraim and Malah 1984), non-negative matrix factorization (Fakhri et al. 2018) etc. In time-domain algorithms, the enhancement filter is implemented to diminish the additive noise corrupting the speech without introducing noticeable distortion in the enhanced speech output where as in the case of frequency-domain algorithms normally noisy speech is changed into the frequency-domain through the Discrete Fourier Transform (DFT), and then the clean speech spectrum is obtained based on the observed noisy spectrum. At last it again transformed back to the time-domain by the inverse DFT. Mostly the frequency-domain algorithms increase the amplitude where as in time-domain algorithms increase both amplitude and phase at the same time implicitly.

4.1 Feature extraction

Extraction of features is the method of extracting various features from the speech signal, like pitch, power and vocal tract configuration. The method of transforming these features into signal parameters via differentiation and concatenation processes is parameter transformation.

In 2009, Yan et al., combined speech enhancement with discriminative feature extraction (Yan et al. 2009) for robust speaker recognition. In this work, they presented a new strategy which used discriminative feature extraction to overcome the acoustics mismatch between training and testing data in the noisy environment. Later in 2010 Even et al. addressed the need of feature extraction (Even et al. 2010) in blind signal, during speech enhancement for samples collected from noisy environment. In this work, blind estimation of diffuse background noise for hands-free speech interface was used for feature extraction. This was also called as Blind Signal Extraction (BSE). The work also focused on Blind signal separation and its significance over BSE. Sharma et al. in 2015 studied various robust feature extraction techniques for speech recognition and enhancement. In this work, they presented a survey of feature extraction techniques for speech recognition and enhancement system. The feature extraction techniques involved, Linear Prediction based features (LPC), linear prediction cepstral coefficient (LPCC), Partial Auto-correlation coefficient (PARCOR) etc. Recently, in 2016, Purushotham and Suresh, introduced a novel feature extraction technique to enhance the speech signal for mobile communication. In this work, they worked on a method that extracted features from speech signal (Purushotham and Suresh 2016) in order to enhance it in real-time. The extracted signal was reconstructed after removing the noise from it, making it an enhanced speech signal.

4.1.1 Time-domain methods

This section mainly focused on various time domain methods of speech enhancement. Spectral subtraction (Boll 1979) is one of the old time-domain approaches utilized for improving the degradation of speech due to additive stationary background noise. A "spectral subtraction" based on preprocessing noise suppression system is used in order to analyze the magnitude frequency spectrum that is lying under the clean speech. It is done by reducing the noise magnitude

spectrum from noisy speech spectrum for improvement in quality and intelligibility. Wiener filtering is another spectral subtraction technique for improving the speech signal, which can recover the noisy signal from the original speech signal by minimalizing the mean square error (MSE) amongst the original signal and the assessed signal. Thresholding in the wavelet domain and decomposition of wavelet influenced a Wavelet de-noising method is proposed in Johnstone and Silverman (1997) for removing noise. This method decomposes signal containing noise into several sub-bands, and the noise reduction is achieved by either soft or hard thresholding. Drawback of this technique is its capability to modify some valuable original speech components. Depending on the adaptation of the sample to sample filter impulse response, an Adaptive Wiener filtering technique is presented in Abd El-Fattah et al. (2013). It is observed that this method gives best results than other speech enhancement techniques, in case of both high and low SNR values. Also the filters perform decently in both AWGN and coloured noise cases. Pandey et al. introduced a convolutional neural network (CNN) based speech enhancement technique (Pandey et al. 2019). For distinguish voiced and unvoiced speech and silence and also to look further intofeatures of noise a Kalman-Filtering Speech Enhancement Method is illustrated in Goh et al. (1999). A new speech enhancement technique in combination with generative adversarial networks (GAN) (Pascual et al. 2019) is proposed to regenerate corrupted signals and convert them into a clearversion, that depends on the raw signal. A Kalman filter (KF) algorithm based on deep neural network (DNN) is implemented in Yu et al. (2019). It is observed that DNN is used to access the significant parameters in the KF, namely, the linear prediction coefficients (LPCs). This algorithm allows for a more precise and reliable calculation of LPCs from noisy speech. The DNN-KF approach beats current KF-based speech amplification approaches with respect to speech performance and intelligibility.

4.1.2 Frequency-domain methods

In this section numerous noise reduction methods that are performed in the frequency domain is described. Basically, a frequency response is calculated per each block, and a multiplication in the frequency domain is applied to this frequency response. Frequency domain speech scrambling's one of the most significant real-time (Ahmed and Ikram 2003) implementation is reported in Ahmed and Akram's work.This work consisted of frequency-Inversion, frequency Hoping Inversion and rolling code Inversion. A comprehensive study of an efficient frequency domain optimal linear estimator (Hu and Loizou 2004a, b) is demonstrated that involves the masking properties of human's hearing system which makes residual noise modulation inaudible. Frequency domain adaptive line enhancer influenced significantly in the introduction of new single-channel speech enhancement system (Nakanishi et al. 2006). After implementing this method, the optimal setting of frequency domain decorrelation parameters was proposed and accustomed according to the existence of speech and the dominancy of the speech elements in frequency domain. By adding the original noisy speech, the speech distortion of speech is first compensated and after that the noise is subtracted by a post-filter. A feature extraction technique using frequency domain (Li 2008) linear prediction (FDLP), that was built upon theory ofmodelling temporal envelopes of the speech signal is addressed in Thomas et al. (2008). For enhancing the intelligibility of speech signal, a new method is proposed by amplifying the location of extracted transient components in frequency domain. Using this method, pre-processing of tonal components (Rezvani and Kahaei 2015) are no longer needed. Later, wavelet-threshold multitaper spectra is implemented for frequency-domain speech improvement techniques as a substitution of conventional fast Fourier transform (FFT) magnitude spectra (Parchami et al. 2016). This method gives good result of the minimum mean-square error log-spectral amplitude estimator (MMSE-LSA), mainly when wavelet-thresholdedmultitaper spectra were applied instead of the FFT spectra. All the above conventional methods enhance signal-to-noise ratio (SNR), along with creation of speech distortions using frequency domain methods. Table 3 presents presents an overview of feature extraction based methods which includes a both time domain methods and frequency domain methods.

4.2 Feature selection

In order to enhance the accuracy in classification system of speech enhancement, feature selection is used. It is done by selecting the most uncorrelated features.

In 2011, Mporas et al. discussed dynamic selection (Mporas et al. 2011) method for speech enhancement and speech recognition that can be used for speech signals recorded in motorcycle environment. In this work, they presented a speech pre-processing scheme (SPPS) that used GMM clusters based feature selection method to improve speech signals in real-time. In 2014, Kalamani et al. discussed feature selection algorithms (Kalamani et al. 2014) for automatic speech recognition. In this work, they explored various feature selection algorithms that used Melfrequency Cepstral Coefficient (MFCC) as feature extraction method. This paper used meta-heuristics such as Particle Swarm Optimization, Genetic Algorithm, Ant Colony Optimization as feature selectors. They also used Support Vector Machine, Neural Networks and Fuzzy Rough set for feature selection. Recently, in 2017, Manolov et al. explored various feature selection techniques that are used for speech classification.

Table 3 An overview of feature extraction based methods

Technique	Field of usage	Works
Discriminative feature extraction	To overcome the acoustics mismatch between training and testing data in the noisy environment	Yan et al. (2009) and Biem et al. (1993)
Blind signal extraction (BSE)	Blind estimation of diffuse background noise for hands-free speech interface	Even et al. (2010) and Cichocki and Thawonmas (2000)
Linear prediction based features (LPC),	A survey of feature extraction techniques for speech recognition and enhancement system. The feature extraction techniques involved, Linear Prediction based features (LPC)	Sharma et al. (2015) and Marchi et al. (2014)
Mobile communication based Feature extraction technique	To enhance the speech signal for mobile communication	Purushotham and Suresh (2016) and Hong Kook and Cox (2000)
Spectral subtraction	Utilized for improving the degradation of speech due to additive stationary background noise	Boll (1979) and Islam et al. (2014)
Wavelet de-noising method	Removing noise	Johnstone and Silverman (1997) and Baishya and Kumar (2018)
Adaptive Wiener filtering	This method gives best results than other speech enhancement techniques	Marwa et al. (2014) and Yelwande et al. (2017)
CNN based speech enhancement technique	Distinguish voiced and unvoiced speech and silence and also to look further into features	Pandey et al. (2019) and Bhat et al. (2019)
Generative adversarial networks (gan)	To regenerate corrupted signals and convert them into a clear version, that depends on the raw signal	Pascual et al. (2019) and Donahue et al. (2018)
Kalman filter (KF) algorithm based on deep neural network (DNN)	To access the significant parameters in the KF, namely, the linear prediction coefficients (LPCs)	Wan and van der Merwe (2001) and Yu et al. (2019)
Optimal linear estimator	Involves the masking properties of human's hearing system	Ephraim and Malah (1983) and Hu and Loizou (2004a, b)
Wavelet-threshold multitaper spectra	Frequency-domain speech improvement techniques as a substitution of conventional fast Fourier transform (FFT) magnitude spectra	Hu and Loizou (2004a, b) and Parchami et al. (2016)
Single-channel speech enhancement system	Optimal setting of frequency domain decorrelation parameters was proposed	Virag (1999) and Nakanishi et al. (2006)

In this work, they presented a method for emotion classification (Manolov et al. 2017) of speech signals in order to identify emotional rate of the speaker. They used mutual information maximization (MIM) feature for scoring criterion of emotional rate. They used wrapper and embedded methods that were classifier dependent, and filter method that was classifier independent. They used this method on EMO-DB. A summary of various feature selection methods is given below in Table 4.

4.3 Classification

Classification is the issue of assessing which of a group of categories (sub-populations) belongs to a new inference,

Table 4 An overview of feature selection methods in speech enhancement

Technique	Field of usage	Works
Dynamic selection method	Speech pre-processing scheme (spps) that used gmm clusters based feature selection method to improve speech signals in real-time	Koniaris et al. (2010) and Mporas et al. (2011)
Automatic speech recognition	Explored various feature selection algorithms that used melfrequency cepstral coefficient (mfcc) as feature extraction method	Kalamani et al. (2014) and Sahu and Ganesh (2015)
Emotion classification	Applied on speech signals in order to identify emotional rate of the speaker	Manolov et al. (2017) and Deshpande et al. (2017)

based on a training set of data comprising findings (or instances) whose membership in the category is known. In speech enhancement classification has generally been applied on the noises in order to reduce it and enhance the speech sample.

Earlier in 2013, Jiang et al. proposed a close talk speech enhancement (Jiang et al. 2013) technique based on classification. In this work, they used binary classification in order to enhance the quality of close talk speech that is recorded in noisy environment. In 2016 Saki and Khetarnavaz proposed a classification technique (Saki and Kehtarnavaz 2016) for noise that would lead to speech enhancement in hearing aid devices. In this work they used voice activity detector as a feature extractor and random forest classifier for classification. Recently, in 2018, Fakhri et al. introduced another speech enhancement (Fakhri et al. 2018) technique that used classification of noisy signals. In this work they used supervised non-negative matrix factorization or NMF, in order to enhance the speech samples. They specifically focused on decomposing noisy observation into multiple parts and then enhanced signal is reconstructed by combining less-corrupted ones. They used Support Vector Machine for classification in this work.

Deep learning (also known as hierarchical learning or deep structured learning) is part of a wider family of artificial neural networks-based machine learning techniques. Deep learning architectures such as deep belief networks, deep neural networks, convolutional neural networks and recurrent neural networks have been applied to fields such as speech recognition, computer vision, natural language processing, social network filtering, audio recognition, machine translation, bioinformatics, medical image analysis, drug design, board game programs and material inspection where they have produced outcomes equivalent to and in some cases greater to human experts.

4.3.1 Deep learning based speech enhancement

In 2014, Xu et al. studied the possibility of speech enhancement using deep neural network (DNN) (Xu et al. 2014). In this work they presented a regression-based speech enhancement framework using deep neural networks (DNNs). They used a multiple-layer deep architecture that trained 100 h of simulated speech data. In 2017, Tu and Zhang worked on speech enhancement (Tu and Zhang 2017) that used deep neural networks with skip connections. In this paper, they focused on fed forward network-based architecture that added skip connections between network inputs and outputs. They used DNN to measure the ideal ratio mask. Later in, April 2018, Karjol et al. discussed a speech enhancement technique that would be possible by using multiple deep neural (Karjol et al. 2018) networks. In this system, they applied multiple DNN to (i) estimate clean speech spectrum, (ii) calculate the weighted average, (iii) use that average to train multiple DNN jointly. The objective function was set to be mean square log error between target spectrum and estimated spectrum. This work showed an improvement of 0.07% while comparing it with single DNN based system, in terms of signal to noise ratio. The discussion highlighted on most signinficant deep learning based speech enhancement techniques, which is also reflected on Table 5.

4.4 Clustering

Clustering is the process of splitting the data points or population into a number of groups so that data points within the same groups are more similar to other data points within the same group and different from data points in other groups. Clustering algorithms are built on similar time–frequency bins being grouped together. These include, in general, methods to computational auditory scene interpretation, which depend on psychoacoustic signals, and methods focused on spectral clustering. Regression and classification algorithms, on contrary, are utilized to determine the origin of the target class or to identify the type of origin that controls each time frequency bin.

In 2005, Srinonchat improved the clustering algorithm that was used in speech coding without degrading the quality of the speech. In this paper, a new clustering technique (Srinonchat 2005) is proposed in order to enhance the performance of the cluster process to provide more accuracy. It also endeavours to use as few cluster centres as possible to represent the data group. Two clustering methods are studied here and applied to the speech signal. In August, 2013, Nakatani et al. proposed a speech enhancement technique that used spatial clustering approach (SCA) (Nakatani et al.

Table 5 An overview of deep learning based speech enhancement methods

Technique	Field of usage	Works
Close talk speech enhancement	Binary classification in order to enhance the quality of close talk speech that is recorded in noisy environment	Jiang et al. (2013) and Matheja et al. (2013)
Voice activity detector	Used as a feature extractor and random forest classifier for classification	Zhao et al. (2014) and Saki and Kehtarnavaz (2016)
Supervised Non-negative matrix factorization or NMF	To enhance the speech samples	Fakhri et al. (2018) and Barman and Lee (2008)

2013). In this paper, they used SCA along with factorial model approach that is based on two different features of signals, spatial and spectral features. In 2018, Nesbitt et al. proposed a speech segment clustering (Nesbitt et al. 2018) technique for real-time speech enhancement. In this work, they proposed a system that can resolve the problem of time complexity in exemplar-based speech enhancement. They also used this system to be used in real-time. To summarize the above discussed clustering techniques and their field of usage, Table 6 is presented below that focuses on key clustering techniques that has been used for speech enhancement.

5 Applications of speech enhancement techniques

Highly efficient and natural mode for communication is speech. Speech enhancement is a technique which works on the noisy speech signal. In a wide range of application domains different methods of speech enhancement is implemented. This section mainly focused on various application domains of speech enhancement techniques namely: speaker recognition, video-conference, speech transmission through communication channel, speech-based biometric system, mobile phones, hearing aids, microphones, voice conversion.

5.1 Speaker recognition

Speaker recognition is the process of identifying the speakers from their spoken words, which carry the information about their identities. Speech enhancement improves the quality of automatic speech recognition. For automatic speaker identification (SID), four different speech enhancement algorithms namely spectral subtraction (SS) (Berouti et al. 1979), statistical modelbased (MMSE) (Ephraim and Malah 1984), subspace (pKLT) and Wiener filtering (Scalart and Vieira-Filho 1996) (SS, MMSE, WIN, pKLT) are used at pre-processing stage to suppress background noise (Jabloun and Champagne 2003). To improve the single channel speaker identification system, the spectral subtraction procedure is implemented and for multi-channel speaker identification system, the enhancement has been done by using adaptive noise cancellation and delay-and-sum speech beamforming (Ortega-Garcia and Gonzalez-Rodriguez 1996). The study of quantile-based spectral subtraction (Bai and Wan 2003) followed by adaptive wavelet denoising (Fu and Wan 2003) approach is implemented which offers best speaker identification accuracy at low SNRs (SNR $\leq$ 10 dB) is described in El-Solh et al. (2008). The speech enhancement techniques are used to reduce the acoustical noise in the speech signal before the speaker recognizes (El-Solh et al. 2008). A variational bayesian algorithm for joint speech enhancement and speaker identification (Sadjadi and Hansen 2010) is addressed in Maina and Walsh (2011). A detailed study of an efficient speech cryptosystem technique and a proposed speaker identification system using cancelable features for speaker identification systems is described in Priyanka (2017) and Soliman et al. (2017).

5.2 Video-conference

Another objective of speech enhancement is to improve speech quality at the time of video conference when video is shot in noisy environment. Video need to be enhanced for the voice of a speaker seen in the video. For this purpose, a spectral subtraction approach through maximum likelihood estimate (MLE) is implemented (Zhou et al. 2008) which simulates probability distribution of useful signals and decrease the noise maximum. To enhance speech communication systems such as video conference an adaptive beamforming method based on post multistage Wiener filter is proposed in Wang et al. (2010) to overpower the coherent and incoherent noise signal more effectively. Detailed study of conference management and speech enhancement technique for video conference is addressed in Prabhu et al. (2012).It is observed that DNN-based system that incorporates audio and visual information for speech enhancement gives a feature-level-fusion method in multimodality research (Hou et al. 2016).For separating the voice of a visible speaker from background noise an end-to-end neural network model is implemented (Gabbay et al. 2018). Video indexing and retrieval of audiovisual data are very important for many applications like education (Bureš et al. 2016), surveillance, media production etc.

Table 6 An overview of clustering methods in speech enhancement

Technique	Field of usage	Works
Clustering	To enhance the performance of the cluster process to provide more accuracy	Li et al. (2014) and Srinonchat (2005)
Spatial clustering	Sca along with factorial model approach that is based on two different features of signals, spatial and spectral features	Nakatani et al. (2013) and Brandstein and Griebel (2000)
Speech segment clustering	For real-time speech enhancement	Kamper et al., (2014) and Nesbitt et al. (2018)

5.3 Speech transmission through communication channel

Speech enhancement has given a good impact in speech transmission through communication channels. Implementation of digitized speech transmission followed by existing VHF FM repeater gives an operational flexibility in the design and use of a radio system (Petrovie 1985). It is observed that data transmission or secure communications through a GSM (Global System for Mobile communications) speech dedicated channel is perfectly handled by digital modulations. At a low bit rate speech transmission through the underwater acoustic communication channel is addressed in Goalic et al. (2005) through which real time transmission (Chmayssani et al. 2008; Zhang et al. 2009) of information (images, speech, text and data) can be accomplished in harsh channels. Performance evaluation of the WiMAX (Worldwide Interoperability for Microwave Access) system under the different wireless channels in real time image and speech transmission is reported in Sedani et al. (2012). In Gao and Zhao (2013), an error control transmission scheme based on UEP (unequal error protection) is applied to the speech source to reconstruct a good quality of speech signals.

5.4 Speech-based biometric system

Speech biometric is the one of the most important applications of speech enhancement techniques. Biometric system is used to communicate with the authorized person and network. It is a powerful way of security (Yamin and Sen 2018). Biometric speaker (Davis and Mermelstein 1980; Reynolds et al. 2000) recognition, reduces the susceptibility of wireless communication and can help to keep the sensitive information. A detailed discussion of an effective feature extraction technique followed by two different speaker verification technique in noisy environments is described in Thulasimani (2012). In Kopparapu (2009), a speech biometric system is introduced which allows the speaker recognition for key-less access to vehicles. In Shukla et al. (2010), by using speech features a neuro-fuzzy based simulation model has been developed which recognizes the speaker along with their gender and mental status. In Shen et al. (2010), the MFCC (Mel-frequency Cepstral coefficients) features are used for speech-based speaker recognition and the GMM-UBM (universal background model) framework is used for speaker verification. Automatic speaker verification using whispered speech is reported in Sarria-Paja et al. (2015).

5.5 Mobile phones

Speech enhancement plays a vital role in the field of digital communication. Many signal processing applications like mobile communication is enhanced.For use in a DSR (distributed speech recognition) system the study of speech enhancement methods with an auditory model-based front end is described in Flynn and Jones (2008).The speech quality of mobile phones is improved by using of speech enhancement techniques. In Lee et al. (2013), on the basis of imparity of two-microphone signals a speech absence probability estimation approach followed by an MMSE estimator is implemented which clean the speech under a noisy environment. To enhance the quality of the speech signal for mobile phones in the noisy environments, simple time domain approach with gain adjustment and the frequency domain approach using psychoacoustic is addressed in Premananda and Uma (2013). An improved speech enhancement method based on "Wavelets" is introduced which provides a good speech quality for mobile phones (Nabi et al. 2016).

5.6 Hearing aids

Speech plays a crucial role in the human-physical (Dey et al. 2018) world's communication. For smooth functioning, it is important for a person to hear different audio signals. Man who lacks his ability to listen the voice, certain electronic equipment are utilized to recover this capacity, which detects sound signals from the physical world and magnifies them to the appropriate level, such that the man is clearly heard by using this device. Speech enhancement plays an important role in digital hearing aids. Generalized maximum posteriori spectral amplitude (GMAPA) speech enhancement algorithm (Lai et al. 2013) is illustrated which aim is to enhance signal-to-noise (SNR) level of received speech signals and increase the speech intelligibility for hearing-loss individuals. In Panahi and Kehtarnavaz (2016), a speech processing pipeline method is introduced to improve speech signals while various types of environmental noises are present in case of hearing aid applications. Using these method significant enhancements in speech is noticed for both users i.e. normal and impaired hearing (Ang et al. 2016). In Modhave et al. (2016a, b), a "Matrix wiener filter" technique for speech enhancement is implemented in hearing aid devices which offers better result than single and multi channel wiener filter. Various spectral subtraction algorithms were discussed in Fukane and Sahare (2011) which are suitable for hearing aids in different noisy Environments. Extraction of pure speech by reducing the estimated noise from a corrupted speech, multichannel wiener filter algorithm is introduced for the application of hearing aids (HA), reported in Modhave et al. (2016a, b). Use of voice activity detector (VAD) is done in order to distinguish the input sound status as unvoiced or voiced, pure noise or speech plus noise for hearing aids are addressed in Saki and Kehtarnavaz (2016a). Detailed study of three speech enhancement (SE) algorithms (Low SNR and ICA based one microphone

speech enhancement, Spectral-Coherence on two microphone speech enhancement) for hearing aid are illustrated in Panahi et al. (2017).

5.7 Microphones

For hands free communication, microphone array-based speech enhancement techniques (Beh et al. 2006) have gain so much attention. In Vu et al. (2010), a cost effective, speech enhancement method is introduced that was based on independent component analysis (ICA) comparison study. The comparative study was done between co-located microphone arrays that contain microphones and traditional uniform linear arrays. The traditional uniform linear arrays formed from omnidirectional microphones was addressed in Shujau et al. (2010). Combination of GSC (Generalized sidelobe canceller) algorithm and spectral subtraction speech enhancement algorithm is illustrated and it efficiently defeats the impact of the residual background noise and enhances speech-noise ratio and intelligibility of speech (Yu and Su 2015). A concave microphone array for VR/AR headsets is presented by Ma et al. (2017). This method was indrouced to increase the speech signals for different financial application scenarios (Ma et al. 2017). This work also included beamforming and ICA methods. A computational auditory scene analysis (CASA) based speech enhancement technique (Jiang and Liu 2017) is proposed. This system was done on flexible dual microphone setting which emphasize on the speech enhancement between the matched and unmatched training and test conditions. In Zhang et al. (2017), a new noise suppression algorithm was introduced by combining beamforming technique and multiband spectral subtraction based on microphone array is implemented by subtracting both background as well as musical noise from speech sample. The study of several approaches to selecting a reference microphone for each speaker by distributed microphone array processing as a frontend of meeting recognition is found in Araki et al. (2018).

5.8 Voice conversion

The methodology of voice conversion (VC) to alter speech acoustics is used to retain the linguistic material transfer of non-/para-linguistic data to any form. VC plays a crucial role in enhancing speech expression, like speech-aid, silent speech interaction, and voice alteration. A VC/EVC-basedAL-to-Speech system is established and various types of alaryngeal speech, such as esophageal speech (ES), electrolaryngeal speech (EL), and body-conducted silent electrolaryngeal speech (silent EL) are discussed in Doi et al. (2011). Implementing this method important improvement on each type of alaryngeal speech is observed. A brief study on trajectory-based conversion method and the capability of effectively reproducing natural speech parameter trajectories utterance by utterance and highlights several techniques is discussed in Toda (2014). Using non-parallel corpora an enhanced variational auto-encoder (EVAE) which can transform the speech have a good target orientation and a well voice quality for voice conversion, is described in Chaudhari and Dhonde (2015). A spectralmapping using Gaussian Mixture Model (GMM) and auditory masking is illustrated (Malathi et al. 2018) and the significant enhancement of electrolaryngeal speech compared to previous enhancement methods is clearly observed. In Kobayashi and Toda (2018), a statistical voice conversion (VC) technique based on Gaussian mixture models (GMMs) has been implemented to electrolarynx (EL) speech enhancement and it is observed that for both objective and subjective evaluation scores this method performs significantly better than the conventional method.

6 Challenges and new perspectives

Different state of the art presented in this study clearly depicts the significance of speech enhancement. It is very difficult job for hearing impaired person as well as normal hearing person to recognize speech signals in the presence of background noise (Dhanj and Eng 2001).Therefore, the challenges in this field are prominent because of the presence of several types of noise in speech signal. Different types of noise and different techniques for noise reduction are discussed here. The spectral subtraction approach is commonly utilized for the methodology of speech improvement, but it suffers from noise which arises due to the difference amongst the expected noise and true noise. Several researchers have contributed to the improvement in the spectral subtraction approach for better noise cancellation efficiency. Moreover, in respect to quality major traditional speech enhancement algorithms gives good outcome but not for intelligibility. Based on four different algorithms namely minimum mean square error (MMSE), spectral subtraction (SS), Wiener Type (WT) and ideal binary mask (IBM) implementation-quality and intelligibility are enhanced. According to the survey, it is observed in the field of artificial neural network (ANNs) in speech processing is the simplicity of the MCP (McCulloch & Pitts) model that form the foundation of the modern neural computing (Zhang and Zhang 1999). Due to presence of noises many speech enhancement techniques suffer from degradation. To overcome this problem adaptive beamforming techniques are introduced.To extract features for speech enhancement in non-stationary noisy environment depends on the speaker characteristics and temporal dynamics, a Corpus based speech enhancement method is implemented. Voice recognition in speech enhancement still has a lot of problems. The challenges in this field are (i)

poor adaptability of the system, (ii) improved recognition model is required improved and (iii) insufficient application extension.

In this study, different speech enhancement methods are evaluated. For speech enhancement a large performance gap between the ideal noise PSD matrix estimation is present, so there is considerable room for further research in this newly explored topic. ANN has become very popular nowadays in this study. LP-ANN model can generate better results in the field of ANNs in Speech Processing. Various adaptive beamforming techniques like LCMV based MVDR, TF-GSC, DTF-GSC, CTF-GSC is introduced which can improve application like hearing aids, teleconference and in stereo recordings in the future. Future work may consist of the speech recognition system functioning properly for strong and complex problems. A new future task of speech recognition method is to apply all incentives and appear on the face of the business and the desires of society in a friendly man–machine conversation.

7 Conclusion

Speech enhancement is a technique having objective of increasing the quality of speech signal. The speech signal is degraded due to various types of noise. In this study, different speech enhancement methods with their applications and different types of noise and its removal techniques are illustrated. Speech enhancement is very essential. The main challenge of speech enhancement is dealing with the background noise because due to the noise the speech quality becomes degraded. We have studied various types of time domain methods and frequency domain methods of speech enhancement regarding noise removal. The current work also discussed various speech enhancement techniques that used feature extraction, classification, deep learning, clustering, in order to justify the contribution of classifiers in this domain. The discussion continued with respect to various applications of such speech enhancement technique. The current work may lead to focussing on novel techniques that will help noise removal in real time; eradicating the time complexity and making the enhancement technique more efficient and robust in nature.

References

Abd El-Fattah, M. A., Dessouky, M. I., Abbas, A. M., Diab, S. M., El-Rabaie, E.-S. M., Al-Nuaimy, W., Abd El-samie, F. E. (2013). Speech enhancement with an adaptive Wiener filter. *International Journal of Speech Technology*, *17*(1), 53–64.

Ahmed, J.& Ikram, N. (2003). Frequency-domain speech scrambling/descrambling techniques implementation & evaluation on DSP. In *7th International Multi Topic Conference, 2003. INMIC 2003* (pp. 781–789).

Al-Shoshan, A. I. (2006). Speech and music classification and separation: A review. *Journal of King Saud University—WEngineering Sciences, 19*(1), 95–132.

Ando, Y. (2013). Autocorrelation-based features for speech representation. *The Journal of the Acoustical Society of America, 133*(5), 1–8.

Ang, L. M., Seng, K. P., & Heng, T. Z. (2016). Information communication assistive technologies for visually impaired people. *International Journal of Ambient Computing and Intelligence, 7*(1), 45–68.

Araki, S., Ono, N., Kinoshita, K., & Delcroix, M. (2018). Comparison of reference microphone selection algorithms for distributed microphone array based speech enhancement in meeting recognition scenarios. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 316–320).

Arslan, L. M., & Hansen, J. H. L. (1997). Speech enhancement for crosstalk interference. *IEEE Signal Processing Letters, 4*(4), 92–95.

Atmaja, B. T., Farid, M. N., & Arifianto, D. (2016). Speech enhancement on smartphone voice recording, 8th international conference on physics & its applications (ICOPIA). *Journal of Physics: Conference Series, 776*, 1–6.

Bachu, R., Kopparthi, S., Adapa, B., & Barkana, B. (2010). Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In K. Elleithy (Ed.), *Advanced techniques in computing sciences and software engineering* (pp. 279–284). Dordrecht: Springer.

Bai, H. & Wan, E.A. (2003). Two-pass quantile based noise spectrum estimation. *Center of Spoken Language Understanding, OGI School of Science & Engineering at OHSU* (pp. 12–16).

Baishya, A., & Kumar, P. (2018). Speech de-noising using wavelet based methods with focus on classification of speech into voiced, unvoiced and silence regions. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*.

Barman, P. C., & Lee, S.-Y. (2008). Nonnegative matrix factorization (NMF) based supervised feature selection and adaptation. In *Intelligent Data Engineering and Automated Learning—IDEAL 2008* (pp. 120–127).

Baumgarten, M., Mulvenna, M. D., Rooney, N., & Reid, J. (2013). Keyword-based sentiment mining using twitter. *International Journal of Ambient Computing and Intelligence, 5*(2), 56–69.

Beh, J., Baran, R. H., & Ko, H. (2006). Dual channel based speech enhancement using novelty filter for robust speech recognition in automobile environment. *IEEE Transactions on Consumer Electronics, 52*(2), 583–589.

Berouti, M., Schwartz, R. & Makhoul, J. (1979). Enhancement of speech corrupted by acoustic noise. In *Proceedings on IEEE ICASSP'79, Washington, DC, Apr. 1979* (pp. 208–211).

Bhat, G. S., Shankar, N., Reddy, C. K. A., & Panahi, I. M. S. (2019). A real-time convolutional neural network based speech enhancement for hearing impaired listeners using smartphone. *IEEE Access, 7*, 78421–78433. https://doi.org/10.1109/access.2019.2922370.

Biem, A., Katagiri, S., & Juang, B.-H. (1993). Discriminative feature extraction for speech recognition. In *Neural Networks for Signal Processing III—Proceedings of the 1993 IEEE-SP Workshop*.

Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, & Signal Processing, 27*(2), 113–120.

Brandstein, M. S., & Griebel, S. M. (2000). Nonlinear, model-based microphone array speech enhancement. In *Acoustic signal processing for telecommunication* (pp. 261–279).

Bureš, V., Tučník, P., Mikulecký, P., Mls, K., & Blecha, P. (2016). Application of ambient intelligence in educational institutions:

Visions and architectures. *International Journal of Ambient Computing Intelligence, 7*, 94–120.

Chaudhari, A., & Dhonde, S. B. (2015). A review on speech enhancement techniques. In *2015 International Conference on Pervasive Computing (ICPC)* (pp. 272–275).

Chawla, M. P. S. (2011). PCA and ICA processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison. *Applied Soft Computing, 11*(2), 2216–2226.

Chen, Z., & Hohmann, V. (2015). Online monaural speech enhancement based on periodicity analysis & a priori SNR estimation. *IEEE/ACM Transactions on Audio, Speech, & Language Processing, 23*(11), 1904–1916.

Chmayssani, T., Baudoin, G., & Hendryckx, G. (2008). Secure communications through speech dedicated channels using digital modulations. In *2008 42nd Annual IEEE International Carnahan Conference on Security Technology* (pp. 312–317).

Christiansen, T.U. Dau, T. Greenberg, S. (2007). Spectro-temporal processing of speech—An information-theoretic framework. In *Hearing—From sensory processing to perception* (pp. 59–523).

Cichocki, A., & Thawonmas, R. (2000). On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics. *Neural Processing Letters, 12*(1), 91–98.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, & Signal Processing, 28*(4), 357–366.

Deshmukh, O. D., & Espy-Wilson, C. Y. (2007). Speech enhancement using the modified phase-opponency model. *Journal of the Acoustical Society of America, 121*(6), 3886–3898.

Deshpande, G., Viraraghavan, V. S., Duggirala, M., Reddy, V. R., & Patel, S. (2017). Empirical evaluation of emotion classification accuracy for non-acted speech. In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP).*

Dey, N., Ashour, A. S., Shi, F., Fong, S. J., & Tavares, J. M. R. S. (2018). Medical cyber-physical systems: A survey. *Journal of Medical Systems, 42*(4), 1–13.

Dhanj, S. & Eng, J.P. (2001). Artificial neural networks in speech processing: Problems & challenges. In *2001 IEEE Pacific Rim Conference on Communications, Computers & signal Processing*. PACRIM (vol. 2, pp. 510–514).

Doi, H., Nakamura, K., Toda, T., Saruwatari, H., & Shikano, K. (2011). An evaluation of alaryngeal speech enhancement methods based on voice conversion techniques. In *2011 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 5136–5140).

Donahue, C., Li, B., & Prabhavalkar, R. (2018). Exploring speech enhancement with generative adversarial networks for robust speech recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp.2018.8462581

El-Solh, A. &Cuhadar, A. &Goubran, R. (2008). Evaluation of speech enhancement techniques for speaker identification in noisy environments. In *Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007)* (pp. 235–239).

Ephraim, Y., & Malah, D. (1983). Speech enhancement using optimal non-linear spectral amplitude estimation. ICASSP '83. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. https://doi.org/10.1109/icassp.1983.1171938

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions of ASSP, 32*(6), 1109–1121.

Even, J., Saruwatari H., Shikano, K., Takatani, T. (2010). Speech enhancement in presence of diffuse background noise: Why using blind signal extraction. In *2010 IEEE International Conference on Acoustics, Speech & Signal Processing* (pp. 4770–4774).

Faúndez-Zanuy, M. M., Esposito, S., Hussain, A., Schoentgen, J., Kubin, G., Kleijn, W. B., et al. (2002). Nonlinear speech processing: Overview & applications. *Control & Intelligent Systems, 30*(1), 1–9.

Fakhri, M., Poorjam, A.H., Christensen, M.G. (2018). Speech enhancement by classification of noisy signals decomposed using NMF & Wiener filtering. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 16–21).

Flamand, J., Le Bihan, N., Martin, A. V., & Manton, J. H. (2016). Low-resolution reconstruction of intensity functions on the sphere for single-particle diffraction imaging. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP).

Flynn, R., & Jones, E. (2008). Speech enhancement for distributed speech recognition in mobile devices. In *2008 Digest of Technical Papers—International Conference on Consumer Electronics* (pp. 1459–1463).

Foth, M., Schroeter, R., & Ti, J. (2013). Opportunities of public transport experience enhancements with mobile services and urban screens. *International Journal of Ambient Computing and Intelligence, 5*(1), 1–18. https://doi.org/10.4018/jaci.2013010101.

Fu, Q. & Wan, E. (2003). Perceptual wavelet adaptive denoising of speech. In *8th European Conference on Speech Communication & Technology, Euro Speech 2003, September 1–4, 2003* (pp. 577–580).

Fukane, A. R., & Sahare, S. L. (2011). Enhancement of noisy speech signals for hearing aids. In *2011 International Conference on Communication Systems & Network Technologies* (pp. 490–494).

Gabbay, A., Shamir, A. & Peleg, S. (2018). Visual speech enhancement. In *Interspeech 2018 2–6 September 2018, Hyderabad* (pp. 1–5).

Gao, D., & Zhao, X. (2013). A speech coding error control transmission scheme based on UEP for bandwidth-limited channels. In *2013 International Conference on Computational & Information Sciences* (pp. 318–321).

Giacobello, D., Christensen, M. G., Dahl, J., Jensen, S., Moonen, M. (2005). Sparse linear predictors for speech processing. In *Proceedings of the International Conference on Spoken Language Processing*, 2008 (pp. 4–7).

Goalic, A., Trubuil, J., Lapierre, G., Labat, J. (2005). Real time low bit rate speech transmission through underwater acoustic channel. In *Europe Oceans 2005, IEEE Xplore 03 October 2005* (pp. 319–321).

Goh, Z., Tan, K., & Tan, B. T. G. (1999). Kalman-filtering speech enhancement method based on a voiced-unvoiced speech model. *IEEE Transactions on Speech & Audio Processing, 7*(5), 510–524.

Gupta, S., Khosravy, M., Gupta, N., & Darbari, H. (2019a). In-field failure assessment of tractor hydraulic system operation via pseudospectrum of acoustic measurements. *Turkish Journal of Electrical Engineering & Computer Sciences, 27*(4), 2718–2729.

Gupta, S., Khosravy, M., Gupta, N., Darbari, H., & Patel, N. (2019b). Hydraulic system onboard monitoring and fault diagnostic in agricultural machine. *Brazilian Archives of Biology and Technology*. https://doi.org/10.1590/1678-4324-2019180363.

Hong Kook, K., & Cox, R. (2000).Bitstream-based feature extraction for wireless speech recognition. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings* (Cat.No.00CH37100).

Hou, J.C., Wang, S.S., Lai, Y.H., Lin, J.C., Tsao, Y., Chang, H.W., & Wang, H.M. (2016). Audio-visual speech enhancement using deep neural networks. In *2016 Asia-Pacific Signal & Information Processing Association Annual Summit & Conference (APSIPA)* (pp. 16–21).

Lee, H., Hu, T., Jing, H., Chang, Y., Tsao, Y., Kao, Y., & Pao, T. (2013). *Ensemble of machine learning and acoustic segment*

model techniques for speechemotion and autism spectrum disorders recognition. INTERSPEECH.

Hu, Y., & Loizou, P. C. (2004a). Incorporating a psycho acoustical model in frequency domain speech enhancement. *IEEE Signal Processing Letters, 11*(2), 270–273.

Hu, Y., & Loizou, P. C. (2004b). Speech enhancement based onwavelet thresholding the multitaper spectrum. *IEEE Transactions on Speech and Audio Processing, 12*(1), 59–67. https://doi.org/10.1109/tsa.2003.819949.

Huang, H., Lee, T., Kleijn, W. B., & Kong, Y.-Y. (2015). A method of speech periodicity enhancement using transform-domain signal decomposition. *Speech Communication, 67*, 102–112.

Islam, M. T., Shahnaz, C., & Fattah, S. A. (2014). Speech enhancement based on a modified spectral subtraction method. In *2014 IEEE 57th International Midwest Symposium on Circuits and Systems* (MWSCAS).

Jabloun, F., & Champagne, B. (2003). Incorporating the human hearing properties in the signal subspace approach for speech enhancement. *IEEE Transactions of SAP, 11*(6), 700–708.

Jalil, M., Butt, F. A., & Malik, A. (2013). Short-time energy, magnitude, zero crossing rate and autocorrelation measurement for discriminating voiced and unvoiced segments of speech signals. In *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)* (pp. 208–212).

Jiang, Y., & Liu, R. (2017). A dual microphone speech enhancement method with a smoothing parameter mask. In *2017 10th International Congress on Image & Signal Processing, BioMedical Engineering & Informatics (CISP-BMEI)* (pp. 386–391).

Jiang Y., Lu, X., Zu Y., Zhou, H. (2013). Classification-based close talk speech enhancement. In *2013 3rd International Conference on Consumer Electronics, Communications & Networks, 20–22 Nov. 2013* (pp. 192–197).

Johnstone, I. M., & Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of Royal Statistical Society, 59*(2), 319–351.

Kalamani, M., Valarmathy, S., Poonkuzhali, C., Catherine, J.N. (2014). Feature selection algorithms for automatic speech recognition. In *2014 International Conference on Computer Communication & Informatics* (pp. 2352–2356).

Kamper, H., Jansen, A., King, S., & Goldwater, S. (2014). Unsupervised lexical clustering of speech segments using fixed-dimensional acoustic embeddings. In *2014 IEEE Spoken Language Technology Workshop* (SLT). https://doi.org/10.1109/slt.2014.7078557

Karjol, P., Kumar, M.A., Ghosh, P.K. (2018). Speech enhancement using multiple deep neural networks. In *2018 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 5049–5054).

Kesarkar, M. P. (2003). *Feature extraction for speech recogntion, M.Tech*. Credit seminar report, Electronic Systems Group, EE. Dept, IIT Bombay, November, 2003.

Khosravy, M., Asharif, M. R., & Yamashita, K. (2010). A theoretical discussion on the foundation of Stone's blind source separation. *Signal, Image and Video Processing, 5*(3), 379–388.

Khosravy, M., Gupta, N., Marina, N., Asharif, M. R., Asharif, F., & Sethi, I. K. (2015). Blind components processing a novel approach to array signal processing: A research orientation. In *2015 International Conference on Intelligent Informatics and Biomedical Sciences* (ICIIBMS).

Kobayashi, K., & Toda, T. (2018). Electrolaryngeal speech enhancement with statistical voice conversion based on CLDNN. In *2018 26th European Signal Processing Conference (EUSIPCO)* (pp. 1–5).

Koniaris, C., Chatterjee, S., & Kleijn, W. B. (2010). Selecting static and dynamic features using an advanced auditory model for speech recognition. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. https://doi.org/10.1109/icassp.2010.5495648

Kopparapu, S. K. (2009). A robust speech biometric system for vehicle access. In *2009 IEEE International Conference on Vehicular Electronics & Safety (ICVES)* (pp. 174–177).

Krishnamoorthy, P., Mahadeva Prasanna, S. R. (2008). Temporal & spectral processing of degraded speech. In *16th International Conference on Advanced Computing & Communications* (pp. 9–14).

Kulkarni, N., & Bairagi, V. (2018). Use of complexity features for diagnosis of Alzheimer disease. In *EEG-Based Diagnosis of Alzheimer Disease* (pp. 47–59). https://doi.org/10.1016/b978-0-12-815392-5.00004-6

Lai, Y.-H., Su, Y.-C., Tsao, Y., & Young, S.-T.(2013). Evaluation of generalized maximum a posteriori spectral amplitude (GMAPA) speech enhancement algorithm in hearing aids. In *2013 IEEE International Symposium on Consumer Electronics (ISCE)* (pp. 245–248).

Lee, S., & Lee, G. (2016). Noise estimation and suppression using nonlinear function with A Priori speech absence probability in speech enhancement. *Journal of Sensors, 2016*, 1–7. https://doi.org/10.1155/2016/5352437.

Leng, X., Chen, J., Benesty, J., Cohen, I. (2018). On speech enhancement using microphone arrays in the presence of co-directional interference. In *2018 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 675–680).

Li, H., Mäntymäki, M., & Zhang, X. (2014). Digital services and information intelligence. *IFIP Advances in Information and Communication Technology*. https://doi.org/10.1007/978-3-662-45526-5.

Li, W. (2008). Effective post-processing for single-channel frequency-domain speech enhancement. In *2008 IEEE International Conference on Multimedia & Expo* (pp. 149–157).

Ma, R., Liu, G., Hao, Q., & Wang, C. (2017). Smart microphone array design for speech enhancement in financial VR & AR. In *2017 IEEE SENSORS* (pp. 1012–1017).

Maina, C., & Walsh, J. M. (2011). Joint speech enhancement & speaker identification using approximate bayesian inference. *IEEE Transactions on Audio, Speech, & Language Processing, 19*(6), 1517–1529.

Malathi, P., Sureshw, G. R., & Moorthi, M. (2018). Enhancement of electrolaryngeal speech using Frequency auditory masking & GMM based voice conversion. In *2018 Fourth International Conference on Advances in Electrical, Electronics, Information, Communication & Bio-Informatics (AEEICB)* (pp. 978–981).

Manohar, K., & Rao, P. (2006). Speech enhancement in nonstationary noise environments using noise properties. *Speech Communication, 48*, 96–109.

Manolov, A., Boumbarov, O., Manolova, A., Poulkov, V., Tonchev, K. (2017). Feature selection in affective speech classification. In *40th International Conference on Telecommunications & Signal Processing (TSP)* (pp. 354–359).

Marchi, E., Ferroni, G., Eyben, F., Gabrielli, L., Squartini, S., & Schuller, B. (2014). Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP).

Matheja, T., Buck, M., & Fingscheidt, T. (2013). A dynamic multi-channel speech enhancement system for distributed microphones in a car environment. *EURASIP Journal on Advances in Signal Processing, 2013*(1), 144–149. https://doi.org/10.1186/1687-6180-2013-191.

Modhave, N., Karuna, Y., &Tonde, S. (2016). Design of matrix wiener filter for noise reduction & speech enhancement in hearing aids. In *2016 IEEE International Conference on Recent Trends*

in Electronics, Information & Communication Technology (RTEICT) (pp. 843–847).

Modhave, N., Karuna, Y., & Tonde, S. (2016). Design of multichannel wiener filter for speech enhancement in hearing aids & noise reduction technique. In *2016 Online International Conference on Green Engineering & Technologies (IC-GET)* (pp. 556–559).

Mporas, I. Ganchev, T., Kocsis, O., Fakotakis, N. (2011). Dynamic selection of a speech enhancement method for robust speech recognition in moving motorcycle environment. In *2011 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 5176–5180).

Mustière, F., Bouchard M. & Bolić, M. (2010). Bandwidth extension for speech enhancement. In *CCECE* (pp. 76–84).

Nabi, W., Aloui, N., &Cherif, A. (2016). An improved speech enhancement algorithm based on wavelets for mobile communication. In *2016 2nd International Conference on Advanced Technologies for Signal & Image Processing (ATSIP)* (pp. 622–626).

Nakanishi, I., Nagata, Y., Itoh, Y., Fukui, Y. (2006). Single-channel speech enhancement based on frequency domain ALE. In *2006 IEEE International Symposium on Circuits & Systems* (pp. 389–393).

Nakatani, T., Araki, S., Yoshioka, T., Delcroix, M., & Fujimoto, M. (2013). Dominance based integration of spatial & spectral features for speech enhancement. *IEEE Transactions on Audio, Speech, & Language Processing, 21*(12), 2516–2531.

Nesbitt, D., Crookes, D., & Ji, M. (2018). Speech segment clustering for real-time exemplar-based speech enhancement. In *2018 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 5419–5423).

Ortega-Garcia, J., Gonzalez-Rodriguez, J. (1996). Overview of speech enhancement techniques for automatic speaker recognition. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96* (pp. 929–933).

Paliwal, K. K. (2003). Usefulness of phase in speech processing. In *Proceedings IPSJ Spoken Language Processing Workshop* (pp. 1–6).

Panahi, I., Kehtarnavaz, N., & Thibodeau, L. (2016). Smartphone-based noise adaptive speech enhancement for hearing aid applications. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 85–89).

Panahi, I. M., Reddy, C. K. A., & Thibodeau, L. (2017). Noise suppression & speech enhancement for hearing aid applications using smartphones. In *2017 51st Asilomar Conference on Signals, Systems, & Computers* (pp. 1890–1894).

Pandey, A., Wang, D. L., & Fellow, I. E. E. E. (2019). A new framework for CNN-based speech enhancement in the time domain. *IEEE Transactions on Audio, Speech, & Language Processing, 27*(7), 1179.

Parchami, M., Zhu, W. P., Champagne, B., & Plourde, E. (2016). Recent developments in speech enhancement in the short-time fourier transform domain. *IEEE Circuits & Systems Magazine, 16*(3), 45–77.

Pascual, S., Serra, J., & Bonafonte, A. (2019). Time-domain speech enhancement using generative adversarial networks. *Speech Communication, 114*, 10–21.

Petrovie, P.M. (1985). Digitized speech transmission through Vhf Fm repeaters. In *35th IEEE Vehicular Technology Conference* (pp. 205–210).

Plapous, C., Marro, C., & Scalart, P. (2006). Improved signal-to-noise ratio estimation for speech enhancement. *IEEE Transactions on Audio, Speech, & Language Processing, 14*(6), 2098–2108.

Prabhu, C., Chellappan, C., & Ramachandran, B. (2012). Conference management & speech enhancement for multiparty video conference over the MPLS Networks. *Information Technology Journal, 11*(1), 85–93.

Premananda, B. S., & Uma, B. V. (2013). Speech enhancement algorithm to reduce the effect of background noise in mobile phones. *International Journal of Wireless & Mobile Networks (IJWMN), 5*(1), 177–189.

Priyanka, S.S. (2017). A review on adaptive beamforming techniques for speech enhancement. In *International Conference on Innovations in Powerand Advanced Computing Technologies* [i-PACT2017] (pp. 1–6).

Purushotham, U,. Suresh, K. (2016). Feature extraction in enhancing speech signal for mobile communication. In *2016 1st India International Conference on Information Processing (IICIP)* (pp. 978–983).

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models. *Digital Signal Processing, 10*(1–3), 19–41.

Rezvani, M., Kahaei, M.H. (2015). Speech enhancement using transient components in frequency domain. In *2015 23rd Iranian Conference on Electrical Engineering* (pp. 164–170).

Sadjadi, S.O. & Hansen, J.H.L. (2010). Assessment of single-channel speech enhancement techniques for speaker identification under mismatched conditions. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26–30, 2010* (pp. 2138–2141).

Sahu, P. K., & Ganesh, D. S. (2015).A study on automatic speech recognition toolkits. In *2015 International Conference on Microwave, Optical and Communication Engineering (ICMOCE)*. doi:10.1109/icmoce.2015.7489768

Saki, F. & Kehtarnavaz, N. (2016). Automatic switching between noise classification & speech enhancement for hearing aid devices. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 736–740)

Santos, E., Khosravy, M., Lima, M. A., Cerqueira, A. S., Duque, C. A., & Yona, A. (2019). High accuracy power quality evaluation under a colored noisy condition by filter bank ESPRIT. *Electronics, 8*(11), 1259.

Santosh, K. C., Borra, S., Joshi, A., & Dey, N. (2019). Advances in speech, music and audio signal processing. *International Journal of Speech Technology, 22*(2), 293–296.

Sarria-Paja, M., Senoussaoui, M., & Falk, T. H. (2015). The effects of whispered speech on state-of-the-art voice based biometrics systems. In *2015 IEEE 28th Canadian Conference on Electrical & Computer Engineering (CCECE)* (pp. 1254–1259).

Sasaoka, N., Shimada, K., Sonobe, S., Itoh, Y., & Fujii, K. (2009). Speech enhancement based on adaptive filter with variable step size for wideband and periodic noise. In: *2009 52nd IEEE International Midwest Symposium on Circuits and Systems*. https://doi.org/10.1109/mwscas.2009.5236011.

Scalart, P. & Vieira-Filho, J. (1996). Speech enhancement based on a priori signal to noise estimation. In *Proceedings of IEEE ICASSP'96, Atlanta, GA, May 1996* (pp. 629–632).

Sedani, B. S., Kotak, N. A., Borisagar, K. R., & Kulkarni, G. R. (2012). Implementation & Performance analysis of efficient wireless channels in WiMAX using image & speech transmission. In *2012 International Conference on Communication Systems & Network Technologies* (pp. 630–634).

Sen, S., Dutta, A., Dey, N. (2019). Audio indexing. In *Audio processing and speech recognition. SpringerBriefs in applied sciences and technology* (pp. 1–11). Singapore: Springer

Sen, S., Dutta, A., Dey, N. (2019), Speech processing and recognition system. In *Audio processing and speech recognition. SpringerBriefs in applied sciences and technology* (pp. 13–43). Singapore: Springer.

Sen S., Dutta A., Dey, N. (2019) Audio classification. In *Audio processing and speech recognition. SpringerBriefs in applied sciences and technology* (pp. 67–93). Singapore: Springer.

Sharma, U., Maheshkar, S., Mishra, A. N. (2015). Study of robust feature extraction techniques for speech recognition system. In *2015 International Conference on Futuristic Trends on Computational Analysis & Knowledge Management (ABLAZE)* (pp. 654–659).

Shen, L., Zheng, N., Zheng, S., & Li, W. (2010). Secure mobile services by face & speech based personal authentication. In *2010 IEEE International Conference on Intelligent Computing & Intelligent Systems* (pp. 97–100).

Shrawankar, U. & Thakare, V. (2010). Noise estimation & noise removal techniques for speech recognition in adverse environment, ifip international federation for information processing 1310. In *IIP 1310, IFIP AICT 340* (pp. 336–342).

Shukla, A., Tiwari, R., & Rathore, C. P. (2010). Neuro-fuzzy-based biometric system using speech features. *International Journal of Biometrics, 2*(4), 391–406.

Shujau, M., Ritz, C. H., & Burnett, I. S. (2010). Speech enhancement via separation of sources from co-located microphone recordings. In *2010 IEEE International Conference on Acoustics, Speech & Signal Processing* (pp. 137–140).

Soliman, N. F., Mostfa, Z., El-Samie, F. E. A., & Abdalla, M. I. (2017). Performance enhancement of speaker identification systems using speech encryption & cancelable features. *International Journal of Speech Technology, 20*(9), 977–1004.

Srinonchat, J. (2005). Improvement of the clustering technique to design a codebook in speech coding. In *2005 5th International Conference on Information Communications & Signal Processing* (pp. 833–837).

Thomas, S., Ganapathy, S., & Hermansky, H. (2008). Recognition of reverberant speech using frequency domain linear prediction. *IEEE Signal Processing Letters, 15*, 681–684.

Thulasimani, L. (2012). Text dependent speech based biometric for mobile security. *International Journal of Computer Applications, 51*(17), 35–40.

Toda, T. (2014). Augmented speech production based on real-time statistical voice conversion. In *2014 IEEE Global Conference on Signal & Information Processing (GlobalSIP)* (pp. 592–597).

Treichler, J., & Agee, B. (1983). A new approach to multipath correction of constant modulus signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 31*(2), 459–472.

Tu, M. & Zhang, X. (2017). Speech enhancement based on deep neural networks with skip connections. In *2017 IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)* (pp. 5565–5570).

Vijayan, K. Xiaoxue, G. Li, H. (2018). Analysis of speech & singing signals for temporal alignment. In *Conference: Asia-Pacific Signal & Information Processing Association Annual Summit & Conference* (pp. 1–5).

Virag, N. (1999). Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transactions on Speech and Audio Processing, 7*(2), 126–137. https://doi.org/10.1109/89.748118.

Vu, N.-V., Ye, H., Whittington, J., Devlin, J., & Mason, M. (2010). Small footprint implementation of dual-microphone delay-and-sum beamforming for in-car speech enhancement. In *2010 IEEE International Conference on Acoustics, Speech & Signal Processing* (pp. 1482–1485).

Wan, E. A. and van der Merwe, R. (2001). Kalman filtering and neural networks. In *Adaptive and learning systems for signal processing, communications, and control. Wiley, 2001, ch. 7—The Unscented Kalman Filter* (pp. 221–280).

Wang, D., Fan, Z., & Li, B. (2010). An adaptive beamforming method based on post-multistage wiener filter for the speech enhancement. In *2010 2nd International Conference on Signal Processing Systems (ICSPS)* (pp. 360–362).

Xu, Y., Du, J., Li-Rong, D., & Lee, C.-H. (2014). An experimental study on speech enhancement based on deep neural networks. *IEEE Signal Processing Letters, 21*(1), 65–68.

Yamin, M., & Sen, A. A. A. (2018). Improving privacy and security of user data in location based services. *International Journal of Ambient Computing and Intelligence, 9*(1), 19–42. https://doi.org/10.4018/ijaci.2018010102.

Yan, Z., Zhenmin, T., Yanping, L. (2009). Combining speech enhancement & discriminative feature extraction for robust speaker recognition. In *2009 WRI World Congress on Computer Science & Information Engineering* (pp. 274–279).

Yelwande, A., Kansal, S., & Dixit, A. (2017). Adaptive wiener filter for speech enhancement. In *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*. doi:10.1109/icomicon.2017.8279110

Yoshizawa, T., Hirobayashi, S. & Misawa, T. (2011). Noise reduction for periodic signals using high-resolution frequency analysis. In *EURASIP Journal on Audio, Speech, and Music Processing volume, 2011*, 5 (2011) (pp. 1–19).

Yu, C., & Su, L. (2015). Speech enhancement based on the generalized sidelobe cancellation & spectral subtraction for a microphone array. In *2015 8th International Congress on Image & Signal Processing (CISP)* (pp. 1318–1323).

Yu, H., Ouyang, Z., Zhu, W.P., Champagne, B. & Ji, Y. (2019). A deep neural network based Kalman filter for time domain speech enhancement. In *2019 IEEE International Symposium on Circuits & Systems (ISCAS)* (pp. 397–403).

Yu, W., He, H., & Zhang, N. (Eds.). (2009). A probabilistic short-length linear predictability approach to blind source separation. In *23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008), Yamaguchi, Japan; Advances in Neural Networks—ISNN 2009*. Lecture Notes in Computer Science.

Zhang, E., Antoni, J., Dong, B., & Snoussi, H. (2012). Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach. *TheJournal of the Acoustical Society of America, 132*(5), 3240–3250. https://doi.org/10.1121/1.4754530.

Zhang, L., & Zhang, B. (1999). A geometrical representation of McCulloch–Pitts neural model and its applications. *IEEE Transactions on Neural Networks, 10*(4), 925–928.

Zhang, S., Shao, F., & Yu, Y. (2009). Unequal error protection of MELP compressed speech based on plotkin type LDPC code. In *2009 WRI International Conference on Communications & Mobile Computing* (pp. 166–169). https://doi.org/10.1109/cmc.2009.94.

Zhang, Q., Wang, M., & Zhang, L. (2017). A robust speech enhancement method based on microphone array. In *2017 IEEE 17th International Conference on Communication Technology (ICCT)* (pp. 1673–1678).

Zhao, Q., Yang, Y., & Li, H. (2014). A novel and efficient voice activity detector using shape features of speech wave. In *Lecture Notes in Computer Science* (pp. 375–384). https://doi.org/10.1007/978-3-319-12484-1_42

Zhou, H, Sadka, A. & Richard M. J. (2008). Speech enhancement in noisy environmets for video retrieval. In *9th International Workshop on Image Analysis for Multimedia Interactive Services. IEEE*, AUT (pp. 197–200).