IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

Study on Speech Synthesis by Sparse Modelling in Noisy Environment: A Review

Gagandeep Kaur[1], Mr. Naveen Sharma[2]

Computer Science & Engineering Dept, ICL Group of Colleges, Ambala, India[1,2]

Abstract: The estimation of unseen noise is the major challenge in speech enhancement algorithm in adverse environments. It is difficult to understand the speech signal under presence of noise from background areas. The human speech and hearing system is inherently sensitive to interfering noise. The use of speech enhancement algorithm removes or reduces the presence of noise. The aim of the noise reduction algorithms is to estimate the clean speech signal from the noisy recordings in order to improve the quality and intelligibility of the enhanced signal. Due to this, it presents a review on method for speech enhancement using mask estimation iteratively. In this work, it will provide the concept of optimization of cost function by iterative method. This will help for reducing the noise from signal. All simulations will be implemented in MATLAB.

Keywords: Speech Enhancement, Speech Processing, Noise Filtering, Sparse Representation etc.

I. INTRODUCTION

Hearing-aid systems are helpful for hearing-impaired people for hearing speech and understating speech. At present, most hearing-aid systems provide hearing-impaired people to hear speech by using sufficient amplification from their hearing aids. In a noisy place, hearing aids could amplify the noise as well as the desired speech signal. Furthermore, feedback associated with high output hearing aids distorts the frequency response of the hearing aid and sometimes causes oscillation. Because conventional hearing-aid systems are lack of noise reduce function, the hearing-impaired people will hear but not understand under noise environments.

It is well known that speech enhancement techniques can extract the speech signal in noise environment. It is reported that every enhancement of 1dB in noisy speech can result in a 10% promotion in intelligibility of speech. So, when the speech enhancement technique is used for a pre-treatment of noisy speech signal prior to speech recognition, it will improve the comfort of hearing-aid in noisy environment and enhance the intelligibility of speech, assisting people with hearing impairment to better gain information for verbal communication even without barrier. Therefore, speech enhancement technique for enhancing performance of hearing-aid systems is desirable.

Speech is the main carrier of human conversation, and speech communication is one of the fastest-growing communication business. With the development of speech signal processing technology, the evaluation of speech quality increases in importance. The speech quality evaluation has made great achievements in speech coding, speech recognition, speech synthesis, but the research in speech enhancement is not mature. The change of speech quality caused by speech coding essentially differs from by speech enhancement, therefore, the speech quality evaluation system in speech coding field cannot be directly applied to speech enhancement.

Speech quality evaluation measures are classified into subjective and objective methods. Subjective measures best fit human feelings, and can better reflect speech quality, but they are subjected to various test conditions, which influences the reliability of results. Speech signals from the uncontrolled environment may contain degradation components along with required speech components. The degradation components include background noise, speech from other speakers etc. Speech signal degraded by additive noise, this make the listening task difficult for a direct listener and gives poor performance in automatic speech processing tasks like speech recognition speaker identification, hearing aids, speech coders etc. The degraded speech therefore needs to be processed for the enhancement of speech components. The aim of speech enhancement is to improve the quality and intelligibility of degraded speech signal. Improving quality and intelligibility of speech signals, reduces listener's fatigue. Quality can be measured in terms of signal distortion but intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Perceptual quality and intelligibility are two measures of speech signals.

The desired speech signal passes through a convolutive acoustic channel before reaching the microphone, where it is combined with sound from other acoustic sources in the environment and it is transduced into the electronic domain.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

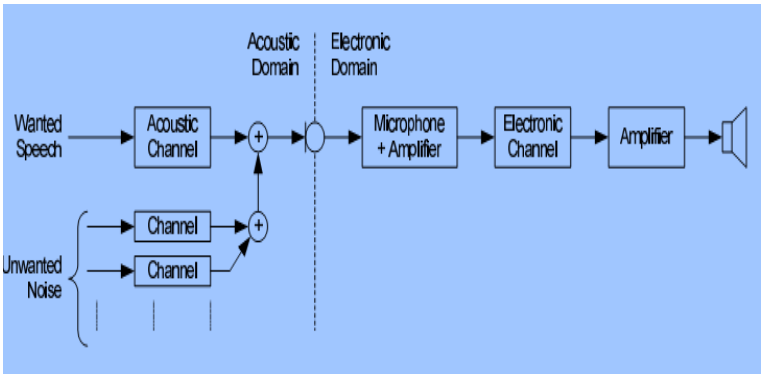

Figure 1: General Speech Recording Chain [2]

The speech signal can become degraded by further additive noise as well as by possible non-linear distortion within the electronic domain.

Speech enhancement has been studied because of its many applications, such as voice communication, voiced –control systems, and the transmitted speech signals. It is a noise suppression technology which has important significance for solving the problem of noise pollution, improving the quality of voice communications, improving speech intelligibility and speech recognition rates, etc.. The objective of speech enhancement is to restore the original signal from noisy observations corrupted by various noises [1]. Speech enhancement techniques have been developed for a single microphone and multiple microphones.

The paper is ordered as follows. In section II, it represents related work with proposed system. In Section III, It defines the description of speech level estimation system. Section IV describes the problem definition of system. Finally, conclusion is explained in Section V.

II. RELATED WORK

Xia Yousheng et. al. [1] proposed a novel multi-channel speech enhancement method by combining the wiener filtering and subspace filtering with a convex combinational coefficient. Because of using both the advantage in noise reduction of the subspace speech enhancement technology and the stable characteristic of the Wiener filtering technology, the proposed multi-channel speech enhancement method had a better performance in robustly removing colored noise from noisy speech signals. Simulation examples confirmed that under different colored noise, the proposed multi-channel speech enhancement method can obtain better speech recovery results than the traditional subspace multi-channel speech enhancement method.

Shoko Araki et. al. [2] investigated a multi-channel de-noising auto-encoder (DAE)-based speech enhancement approach. In recent years, deep neural network (DNN)-based monaural speech enhancement and robust automatic speech recognition (ASR) approaches have attracted much attention due to their high performance. Although multichannel speech enhancement usually outperforms single channel approaches, there has been little research on the use of multi-channel processing in the context of DAE. In this paper, they explored the use of several multi-channel features as DAE input to confirm whether multi-channel information can improve performance.

Zheng Gong et. al. [3] developed two embedded hearing aid systems with noise reduction, respectively using Kalman filter and Wiener filter techniques. Next, they gave a comparative study on the two speech enhancement-based hearing aid systems by testing subjective auditory in noise environment. The comparative result showed that the hearing aid system based on the Kalman filter-based speech enhancement can increase the rate of speech recognition and the hearing comfort of hearing impaired persons in a noisy environment, compared with the hearing aid system based on the Wiener filter-based speech enhancement.

Feng Deng et. al. [4] proposed a sparse hidden Markov model (HMM) based single-channel speech enhancement method that models the speech and noise gains accurately in non-stationary noise environments. Autoregressive models were employed to describe the speech and noise in a unified framework and the speech and noise gains are modelled as random processes with memory. The likelihood criterion for finding the model parameters is augmented with a regularization term resulting in a sparse autoregressive HMM (SARHMM) system that encourages sparsity in the speech- and noise- modelling. In the SARHMM only a small number of HMM states contribute significantly to the model of each particular observed speech segment.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

Pejman Mowlaee et. al. [5] presented a harmonic phase estimation method relying on fundamental frequency and signal-to-noise ratio (SNR) information estimated from noisy speech. The proposed method relies on SNR-based time-frequency smoothing of the unwrapped phase obtained from the decomposition of the noisy phase. To incorporate the uncertainty in the estimated phase due to unreliable voicing decision and SNR estimate, they proposed a binary hypothesis test assuming speech-present and speech-absent classes representing high and low SNRs. The effectiveness of the proposed phase estimation method is evaluated for both phase-only enhancements of noisy speech and in combination with an amplitude-only enhancement scheme.

Swati Pawar et. al. [6] presented an algorithm for improving speech intelligibility. Various speech enhancement algorithms were developed but only some of them can be used for real time hearing aid applications. This proposed algorithm can be used for practical hearing prosthetic devices. Implementation of the binary masking algorithm uses a bank of band-pass filters to perform mapping of signals. Also, classification is performed with a signal-to-noise (SNR) estimate and a comparator. This includes spatial filtering method, classification of signals such as original and noisy signal. After this based on SNR threshold level signals are recombined to obtain reduced noise level in speech signal.

Nasser Mohammadiha et. Al. [7] proposed a novel speech enhancement method that was based on a Bayesian formulation of NMF (BNMF). To circumvent the mismatch problem between the training and testing stages, they proposed two solutions. First, they used an HMM in combination with BNMF (BNMF-HMM) to derive a minimum mean square error (MMSE) estimator for the speech signal with no information about the underlying noise type. Second, they suggested a scheme to learn the required noise BNMF model online, which was then used to develop an unsupervised speech enhancement system. Extensive experiments are carried out to investigate the performance of the proposed methods under different conditions.

Anuradha R. Fukane et. Al. [8] reported a performance evaluation of Spectral subtraction algorithm and its modified versions for Hearing aids in different environments such as restaurant, train and Car environments. Clean speech signals were corrupted by background noise respectively multi-talker babble noise, train noise, and car engine noise at four different signal-to-noise ratio levels -2dB, 0dB, 5dB,10dB. Subjective and objective type evaluation of enhanced speech signals were carried out. The evaluation of intelligibility and quality of enhanced speech was reported for hearing Aids.

III. DESCRIPTION OF SPEECH LEVEL ESTIMATION

Speech sounds can be broadly divided into two categories: voiced and unvoiced. Voiced sounds are produced when the vocal folds are vibrating, producing a quasi-periodic signal, while unvoiced sounds are articulated without vibration of the vocal folds. Speech consists of a sequence of vowels and consonants together with brief silences between phonemes and words [1]. Vowels are created by a voiced sound without any constriction in the vocal tract.

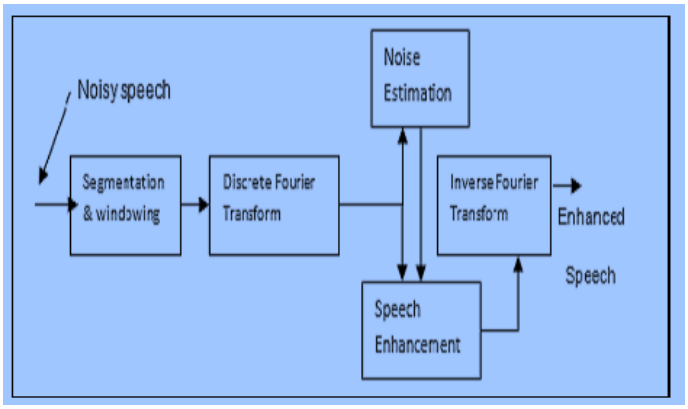

Figure 2: General Block Diagram of Speech Enhancement

Noise, in contrast to speech, can originate from any kind of source and have any spectral and temporal characteristics. There are, however, some common assumptions made about the noise when approaching the speech enhancement problem:
(i) The power spectrum of noise is more stationary than that of speech, and
(ii) Speech and noise are statistically independent.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

The background noise is considered additive and uncorrelated to the speech signal. Let s(n), w(n) and x(n) represent the speech signal, noise signal and noisy speech signal respectively.

$$x(n)=s(n)+w(n)$$

Many speech enhancement techniques require an estimation of the noise power spectrum, or, equivalently, the SNR at each time-frequency bin. The accuracy of the noise estimation technique has a major impact on both the quality and intelligibility performance of the processed speech. The parameters which determine the rejection/acceptance of a time-frequency bin vary according to different binary mask definitions. The original goal of the binary mask estimation was to identify the regions where the SNR was higher than 0dB.

The active level of a speech signal is defined to be its average power during intervals when speech is present. The measurement of a signal's active level is an essential component in any application where the input speech power needs to be normalized, such as in non-intrusive metrics for quality assessment [10]. It is also important whenever a pre-trained speech model is combined with an estimated noise model as in the parallel model combination technique or to determine the SNR of an input signal. For binary mask estimation, the speech active level can be used to make the process independent of the initial speech level.

In the binary mask approach to speech enhancement, a binary-valued gain mask is applied to the speech in the time-frequency domain and the signal is then transformed back into the time-domain. This procedure is similar to that used in conventional approaches such as spectral subtraction or MMSE estimators except that, in the latter cases, a continuously variable gain function is applied. The principal advantage of the binary mask approach over other state-of-the-art algorithms operating in the time frequency domain is that the problem of enhancement is changed from one of gain estimation to one of classification.

Another difficulty concerning the research of speech enhancement is that many types of noises are non-stationary. In contrast to stationary noises, the spectral properties of non-stationary ones are difficult to predict and estimate, which makes noise removal challenging. To separate speech and noise in the spectral domain, nonnegative dictionary learning has been extensively studied recently [10]. In this idea, one first trains two groups of nonnegative bases: the speech related basis and the noise related basis.

The noisy input speech spectrum is subsequently represented by the convex combination of both bases. Finally, the clean speech spectrum is estimated by the linear combination of the speech bases weighted by their coefficients. In case of an unseen noise type, no noise dictionary can be obtained beforehand. Hence, only a speech dictionary is available and leaving the noise bases to be learned on the fly during the enhancement. Another approach is to train a group of noise bases from some known noise types and then to utilize the bases to unseen noise conditions, regardless of the possible mismatch between the training and testing noises.

IV. PROBLEM DEFINITION

Speech enhancement or noise reduction has been one of the main investigated problems in the speech community for a long time. The problem arises whenever a desired speech signal is degraded by some disturbing noise. The noise can be additive or convolutive. Even this additive noise can reduce the quality and intelligibility of the speech signal considerably. Therefore, the aim of the noise reduction algorithms is to estimate the clean speech signal from the noisy recordings in order to improve the quality and intelligibility of the enhanced signal. Due to this, it proposes a method for speech enhancement using mask estimation iteratively. The main focus is to improve the cost of system.

In this work, we investigate a method without any pre-training of noise models. The only assumption about the noise is that it is different from the involved speech. Therefore, the noise estimation turns out to be finding the components which cannot be adequately represented by a well defined speech model. Given the good performance of deep learning in signal representation, a deep auto encoder (DAE) is employed for accurately modelling clean speech spectrum.

V. CONCLUSION

Speech signals can be degraded in many ways during their acquisition in noisy environments and they can also be further degraded in the electronic domain. This paper provides a review on speech enhancement method based on sparse modelling in noisy environment. This method will provide a noise reduction procedure which functions and gives low residual noise, high quality speech. The main parameters are BER, noise sigma and cost of function. For this, it will use the concept of sparse modelling and PCA vectors. The low variability of the gain function during stationary

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 9, September 2017

input signals will give an output with less tonal residual background noise, thus low noise distortion. After this, it will use the concept of ANN for minimizing the error.

REFFERENCES

[1] Xia Yousheng, Huang Jianwen, "Speech Enhancement Based on Combination of Wiener Filter and Subspace Filter", IEEE 2014.
[2] Shoko Arakit, Tomoki Hayashi, "Exploring Multi-Channel Features For Denoising-Auto-encoder-Based Speech Enhancement", IEEE 2015.
[3] Zheng Gong and Youshen Xia, "Two Speech Enhancement-Based Hearing Aid Systems and Comparative Study", IEEE International Conference on Information Science and Technology, April 24-26, 2015, China.
[4] Feng Deng, Changchun Bao, "Sparse Hidden Markov Models for Speech Enhancement in Non-Stationary Noise Environments", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 11, November 2015.
[5] Pejman Mowlaee and Josef Kulmer, "Harmonic Phase Estimation in Single-Channel Speech Enhancement Using Phase Decomposition and SNR Information", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 9, September 2015.
[6] Swati R. Pawar, Hemant kumar B. Mali, "Implementation of Binary Masking Technique for Hearing Aid Application", IEEE International Conference on Pervasive Computing, 2015.
[7] Nasser Mohammadiha, Paris Smaragdis, "Supervised and Unsupervised Speech Enhancement Using Nonnegative Matrix Factorization", IEEE Transactions on Audio, Speech, And Language Processing, Vol. 21, No. 10, October 2013.
[8] Anuradha R. Fukane, Shashikant L. Sahare, "Enhancement of Noisy Speech Signals for Hearing Aids", IEEE International Conference on Communication Systems and Network Technologies, 2011.
[9] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in Proc. ICASSP, 2002, pp. 4164–4167.
[10] K. Paliwal, K. Wjcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain,"Speech Commun., vol. 52, no. 5, pp. 450–475, 2010.
[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., vol. ASSP-33, no. 2, pp. 443–445, Apr. 1985.
[12] P.C. Loizouand S. Member, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp. 857–869, Sep. 2005.
[13] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," IEEE Trans Audio, Speech, Lang. Process., vol. 20, no. 4, pp. 1383–1393, May 2012.