www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 4 Issue 10 Oct 2015, Page No. 14658-14662

An Efficient Structured Sparsity Model For Reverberant Speech Separation

[1]Khamarunnisa P. [2]Sherikh K. K.

[1] M-Tech Student, Computer Science and Engineering, MES college of Engineering
Malappuram, Kerala, India
mail2khamaru@gmail.com

[2] Assistant Professor, Computer Science and Engineering, MES college of Engineering
Malappuram, Kerala, India
sherikhkk@gmail.com

Abstract: ***The growth of technology in this world is at a very high rate and hence the need to cope up with it is essential. In the field of communication, speech is one with the topmost priority. Performance of current speech recognition systems severely degrades in the presence of noise and reverberation. While rather simple and effective noise reduction techniques have been extensively applied, coping with reverberation still remains as one of the toughest problems in speech recognition and signal processing. Reverberation in speech is one of the primary factors which degrade the quality of the audio by persistence of audio in space by creating large number of echoes. Reverberation degrades the speech signal when recorded by a distant microphone and in the hands free telephonic scenarios. This reverberation corrupts the speech signal and it is difficult to carrying out communication in automatic voice recognition applications in which the voice is not properly recognized by the voice recognition applications. Reverberation in speech, caused by room reflections, is problematic especially for hands-free telephonic applications in a confined space. The problem is even severe for hearing impaired people. Therefore blind speech dereverberation is an important research area. The task is to remove reverberation from the output of a room, where the room impulse response, as well as the clean speech signal is unknown.***

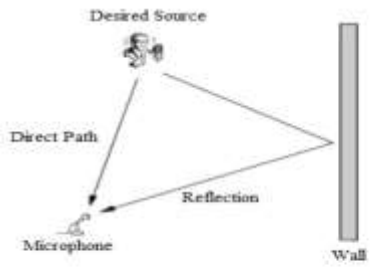
1. Introduction

As the new technologies come into existence, speech is regarded as the most important way of communication. In all forms of the communication systems that use speech, one of the main challenges of the researchers is to maintain the quality and intelligibility of the speech while the information is exchanged or transmitted from one part to another. Due to the presence of surrounding noise such as impulse noise, background noise and environmental noise the performance of communication systems in real-life application is degraded automatically. All these noises cause the distorted exchange of information during communications. The success of communication depends on the restoration of clear speech signal from the mixture of disturbances. The effect of reverberation on speech is spectrally distorted the sound distance and can also reduce intelligibility. Reverberation is described by the concept of reflections. The desired source produces wave fronts, which propagate outward from the source. The wave fronts reflect off the walls of the room and superimpose at the microphone. Due to the difference in the length of the desired source to microphone and in the amount of sound energy absorbed by the walls, each wave front arrives at the microphone with a different amplitude and phase. The term reverberation entitles the presence of delayed and attenuated copies of the source signal in the received signal. The received signal generally consists of direct sound, reverberation and reflections that arrive after the early reverberation called late reverberation.

a) Direct sound: The first sound that is received without reflection is known as the direct sound. If the source is not in line of sight of the receiver, there is no direct sound. b) Early reflections: The sounds which have undergone reflection to one or more surfaces such as walls, floors, furniture are received after a short time delay. These sounds are called as the reflected sounds and all these reflected sounds combine to form a sound component called as early reverberation. This type of reverberation provides the details about the size and position of the source in space as it varies when the source or microphone moves in the space. As long as delay of reflections do not exceed the limit 50-60 milliseconds approximately. Early reflections provide information such as size of the room and position of the speaker in the room.

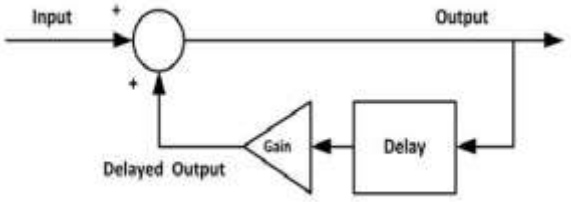
c) Late reverberation: Late reverberation results from reflections which arrive with larger delays after the arrival of the direct sound. They are perceived either as separate echoes, or as reverberation, and impair speech intelligibility.

1.1 Model of Reverberation



Reverberation in an enclosed space.

Due to the difference in length of the desired source to the microphone and in the amount of sound energy absorbed by the walls, each wave front arrives at the microphone with different amplitude and phase. The term reverberation entitle the presence of delayed and attenuated copies of the source signal.



BLOCK DIAGRAM OF REVERBERATION

$$x(n) = s(n) + N\sum k=1\ bks(n- nk)$$

The above equation is the mathematical model for the reverberation. This equation shows that in this model, the first term on the right hand side is the signal component and the second term is the component due to degradation. In the case of reverberation, the degrading component is dependent on previous speech data, whereas in the case of noisy speech the degrading component is independent of speech. That is, in the reverberation the degrading component is speech-like. The relative strength of the reverberant component over the direct component depends on the energy of the speech signal in a short segment around the current instant. This strength can be called signal-to-reverberant component ratio (SRR) at that instant. Likewise, the ratio of the signal energy to the noise energy in a short segment around the current instant is called SNR at that instant. To study the characteristics of SRR (Signal to Reverberant Ratio) and SNR(Signal to Noise Ratio) as a function of time, these ratios are computed for short (2 ms) segments of degraded speech. Due to non-stationary nature of speech, the signal energy varies with time.

1.2 Contributions

In this work various speech dereverberation methods that uses removal of reverberation is surveyed. A better approach with promising level of efficiency is considered and the problems associated with it is studied. And a new system for speech dereveberation is proposed with efficiency and accuracy.

2. Related works

In the existing method, it address the problem of separating the signals of an unknown number of speakers from multi-channel recordings in a reverberant room. Iterative hard thresholding (IHT) is an effective approach to estimate the sparse vectors. Since the computation of the exact bounds is combinatorially hard, the assumptions made for constant step size selection strategies are unverifiable even for moderate-sized random matrices. To improve stability, an adaptive scheme is mandatory. The constant step size using gradient update steps or pseudo inversion optimization techniques provides signal reconstruction efficiency, but more computational power is needed per iteration. In estimation of sparse vector model constant step size selection scheme is accompanied with strong constant conditions but empirical evidence reveal signal reconstruction vulnerabilities even for small deviations from the initial problem assumptions. While convergence derivations of adaptive schemes are characterized by weaker bounds, the performance gained by this choice, both in terms of convergence rate and data recovery, is quite significant. Memory-based methods lead to convergence speed with no extra cost on the complexity of hard thresholding.

Recovery of speech signals from an acoustic clutter of unknown competing sound sources plays a key role in many applications involving distant-speech recognition, scene analysis, video-conferencing, hearing aids, surveillance, sound field equalization and sound reproduction. Here it exploits structured sparsity models to perform speech recovery and room acoustic model ling from recordings of concurrent unknown sources. The speakers are assumed to lie on a two dimensional plane and the multipath channel is characterized using the image model.

2.1 Reverberant Speech Recordings

The problem of separating the signals of an unknown number of speakers from multichannel recordings in a rev erberant room. Consider an approximate model of the acoustic observation as a linear convolutive mixing process, stated concisely as

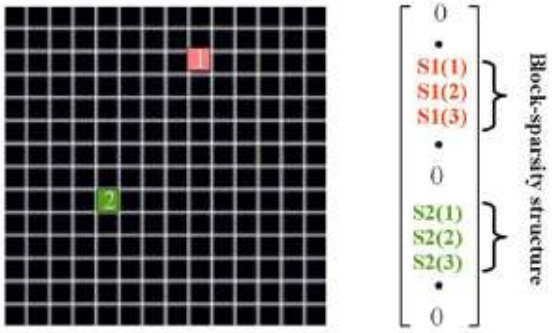
$$Xm=N\sum n=1\ Hmn\ Sn;\ m=1\ldots.M$$

- Xm and Sn denote the time domain signal of mth microphone and nth source
- Hmn denotes the acoustic channel between signal and microphone
- M and N :- Total number of microphones and sources

To represent it in a sparse domain, apply the discrete Short-Time Fourier Transform (STFT) on speech signals.

2.2 Spatio-Spectral Sparse Representation

To obtain the sparse representation of multiparty speech sources, consider a scenario in which N speakers are distributed in a planar area (at the same height in three dimensional space) spatially discretized into a grid of G cells. It assume to have a sufficiently dense grid so that each speaker is located at one of the cells thus N << G. The spatial spectrum of the sources is defined as a vector with a sparse support indicating the components of the signal corresponding to each cell of the grid is given in figure.



3. PROPOSED WORK

The Iterative Hard Thresholding algorithm (IHT) is a powerful and versatile algorithm for sparse inverse problems. The Iterative Hard Thresholding algorithms is a simple yet powerful tool to reconstruct sparse signals. Not only does it

give near optimal recovery guarantees under the RIP, it is also very versatile and can be easily adapted to a range of constraint sets as well as to non-linear measurement systems. The standard IHT implementation faces two challenges when applied to practical problems. The step size parameter has to be chosen appropriately and, as IHT is based on a gradient descend strategy, convergence is only linear. The choice of the step size can be done adaptively and as a result the use of acceleration methods to improve convergence speed. Based on recent applications on IHT it shows that a host of acceleration methods are also applicable to IHT. Importantly, it shows that

These modifications not only significantly increase the observed speed of the method, but also satisfy the same strong performance guarantees enjoyed by the original IHT method.

IHT is a very simple algorithm and under certain conditions, IHT can recover sparse and approximately sparse vectors with near optimal accuracy .

There are two issues with this simple scheme

• The step size has to be chosen appropriately to avoid instability of the method.

• IHT has only a linear rate of convergence.

In normalized IHT algorithm, the chooses adaptively in each iteration. This shows to guarantees the stability of normalized IHT.

3.1 Iterative Hard Thresholding

Accelerated IHT has split into two categories:

- method that only update the non-zero elements
- method that are allowed to update all elements

The second thresholding step in above method guarantees the new estimation of the k-sparse.

4. Results and Analysis

. This chapter deals with details regarding the implementation of both the MCA based detection system and its modified version. Later the performance of both detection systems are evaluated and discussed with the help of corresponding graphs.

4.1 Implementation Details

The existing and proposed systems are implemented using MATrix LABoratory and different tool boxes in MATHLAB. Data set (challenges in reverberation) and green function values are obtained, by using this data set the room impulse response and estimation of room geometry are formulated. The result are measured from the standard composite quality measurement. The selected dataset contains the values of room dimensions relative distance between source (speaker) and microphones. The overall implementation process is detailed as follows. First the structured sparsity model based constant step size IHT is implemented. Second the modified version is implemented and the performance, quality aspects of both the systems are compared. Graphs are generated based on the results.

4.2 Quality Evaluation Metric

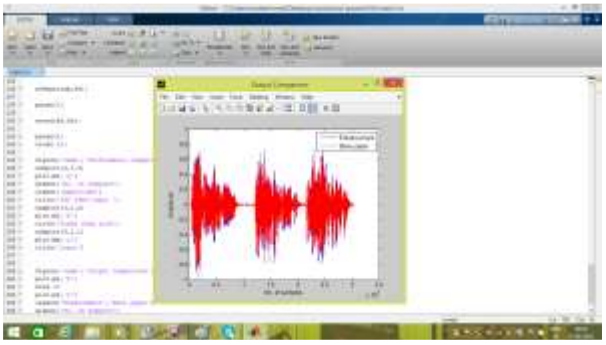
The project is implemented in Matlab. In speech communication, noises from surrounding environments affect the communication quality with various aspects. There are two aspects to speech quality; the perceived overall speech quality, and the speech intelligibility. In speech signal processing the quality of speech enhancement algorithms is validated using the subjective and objective distortion measures.

4.3 Subjective Quality Evaluations

Subjective quality measures are based on the opinion of a listener or a group of listeners. It provides the most accurate assessment of performance since the degree of perceptual quality and intelligibility is determined by the human auditory system. Subjective quality evaluations are done by a group of listeners. They are also called as test subjects. They can be classified into expert subjects (trained) and naive subjects (untrained) according to their knowledge about the selection of the processing conditions under test. The expert subjects will have knowledge about the speech characteristics and quality measurement. In speech quality, a specific unit, called Mean Opinion Score (MOS), is employed to define the resulting quality scores. It corresponds to an average of the individual scores for each processing condition. After listening to the processed speech signal listeners have to rate that particular speech signal.

The proposed adaptive iterative hard thresholding method are compared with the existing constant IHT for reverberant speech separation. The performance are measured by objective, subjective and composite quality measurements. The other results are taken from the waveforms fig 6.1 and 6.2.

In the above waveforms, the amplitude and number of samples are taken in x and y axis. The dark constant line in waveforms indicates high reverberation. That portion has more reverberant component. By iteratively minimizing the repeated values of the signal, can attain the output which having less reverberant content. So from the above figure when using the adaptive step size in IHT, the reverberant component from the signal is finely removed and thick dark lines also removed. Speech quality



can be measured only by hearing the audio signals rather than measuring graphical parameters as it is an audio signal. Therefore considering a survey method is the most suitable. The subjective measurement are evaluated from a group of 40 members. All the 40 members of the group belong to varying perspectives like age, qualification, sex etc... The reverberated input speech and both the output speech signals was subjected before the group for individual evaluation. All the members are utter illiterate about the reverberation input and output speech and all. Hence they were asked to rate the output speeches using MOS scale. The derived ratings of the individuals were random. The conclusions obtained are as follows:

- Some of individuals felt the output was perfect to hear.
- Some had an opinion to make the output much more perfect

As per the evaluation of subjective measurement MOS scale for Constant based IHT(CIHT), the obtained average result is rated 3 that means somewhat natural, some what degradation and for proposed Adaptive based IHT(AIHT) the average is rated as 4that means finely natural and little degradation.

The subjective measures are very costly and time consuming. It requires a set of discriminating listeners. It also needs complex setups like reference system, sending and receiving booths etc. Objective measures are based on mathematical measures and evaluate the quality using the original (clean speech signal) and processed speech signal (enhanced speech signal). It can be evaluated automatically from the speech signal. The four objective measures which are used for the evaluation are Segmental SNR (SNRseg), Weighted Slope Spectral distance (WSS), Perceptual Evaluation of Speech Quality (PESQ) and Log Likelihood Ratio (LLR). The most accurate method for evaluating speech quality is through subjective listening tests. Although subjective evaluation have many disadvantages, another promising approach to estimating the subjective quality is the use of composite objective measures, which is the result of the evaluation of the relationship between subjective analysis and the single objective measures. The reason behind the use of the composite measures is that different objective measures capture different characteristics of the distorted or enhanced signal, and therefore combining them in a linear or non-linear fashion can potentially yield a significant gain in correlations (i.e., correlation with subjective measures such as Mean Opinion Scores (MOS)). These values are obtained by linearly combining the existing objective measures by the following relations:

1. Csig = 3.093 1.029LLR + 0.603PESQ - 0.009WSS
2. Cbak = 1.634 + 0.478PESQ - 0.007WSS + 0.063segSNR
3. Covl = 1.594 + 0.805PESQ - 0.512LLR - 0.007WSS

where LLR, PESQ, WSS and segSNR represents the log likelihood ratio, perceptual evaluation of speech quality, weighted slope spectral distance and segmental SNR, respectively. The coefficients of the linear equations are determined by computing the correlation coefficient between the subjective quality measure (MOS) and the objective qualitymeasure. The above equations are taken from the composite quality measures matlab file.

4. Conclusion

The experimental results conclude that the proposed structured sparsity model based on adaptive IHT method produced better results when compared to the existing constant scheme IHT. In the estimation of sparse vector model constant step size selection scheme is accompanied with strong constant conditions but emprical evidence reveal signal reconstruction vulnerabilities even for small deviations from the initial problem assumptions. While convergence derivations of adaptive schemes are characterized by weaker bounds, the performance gained by this choice, both in terms of convergence rate and data recovery is quite significant. Furthermore, combining these acceleration methods with NIHT (Normal IHT) significant increased the algorithms convergence speed, making the accelerated IHT algorithm. Importantly, the accelerated NIHT method is extremely simple to implement and does not require the computation, storage and repeated use of matrix inverses. The proposed accelerated IHT has some following issues

- It have to update all the elements every time as the part of iteration.
- It should be time consuming.
- It became complex.

References

[1] Patrick A. Naylor , Nikolay D. Gaubitch, ”Speech Dereverberation”, IEEE Communications Tutorials,vol.15,No.1,First Quarter, 2011

[2] Itai Peer, Boaz Rafaely , Yaniv Zigel, ”Room Acoustics Parameters Affecting Speaker Recognition Degradation Under Reverberation”, Acoustics Study,2007

[3] B. Yegnanarayana , P. Satyanarayana Murthy, ”Enhancement of Reverberant Speech Using LP Residual Signal ”, IEEE Trans. on Speech and Audio Processing,2010

[4] Bradford W Gillespie , Henrique S. Malvar, ”Speech Dereverberation via Maximum-Kurtosis Subband Adaptive Filtering”, in proceedings of IEEE Global Telecommunications Conference(GLOBECOM),2012

[5]Marc Delcroix , Takafumi Hikichi , ” Blind dereverberation algorithm for speech signals based on multi-channel linear prediction” , (Acoustics Study), May,2011

[6] Raja Rajeswari, Kusma Kumari, "Noisy Reverberation Suppression Using AdaBoost Based EMD in UnderWater Scenario ", International Journal of Oceanography, 2014

[7] Mingyang Wu, D eLiang, "Two Stage Algorithm For One Microphone Reverberant Speech Signal", IEEE Transactions on Audio Speech and Language Processing, 2006

[8] P. Krishnamoorthy, S. R. Mahadeva Prasanna, "Temporal and Spectral Processing Methods for Processing of Degraded Speech: A Review", IETE JournalTechnical Review,2013

[9] Afsaneh Asaei, Mohammad Golbabaee, Herv Bourlard and Volkan Cevher,"Structured Sparsity Models for Reverberant Speech Separation", IEEE Transactionson Audio,Speech and Language Processing,2014

[10] Anastasios Kyrillidis and Volkan Cevher "Recipes on Hard Thresholding Methods", IEEE Transactions on Audio,Speech and Language Processing,2014