International Journal of Engineering Research in Electronic and Communication Engineering (IJERECE) Vol 3, Issue 6, June 2016

Noise Dereveberation in Speech Signal

[1] Neharaj K J [2] Pramila B
[1] Student [2] Lecturer
[1] Dept.of Electronics and Communication
East West Institute of Technology
Bengaluru, India

Abstract: **There are many artifact and different distortions present in the recording. The reflections of sound depend on the geometry of the room and it causes the smearing of the recording called asreverberation. The background noise depends on the unwanted audio source activities present in the evidential recording. The acoustic reverberation depends on the shape and the composition of the room, and itfor digital media to be considered as proof in a court its authenticitymust be verified. A technique proposed is based on spectral subtraction to estimate the amount of reverberation. Also nonlinear filtering based on particle filtering is used to estimate the background noise. Feature extraction is by using MFCC approach. The feature vector is the addition of features from acoustic reverberation and background noise. SVM classifier isused for classification of the environments. Acoustic environment identification(AEI), audio forensics, and ballistic settings. We describe a statistical technique based on spectral subtraction to estimate the amount of reverberation and nonlinear filtering based on particle filtering toestimate the background noise. The effectiveness of the proposed method is tested using a data set consisting of speech recordings of two human speakers (one male and one female) made in eight acoustic environments using four commercial grade microphonesis robust to MP3 compression attack.**

***Keywords*-MFCC Mel Frequency Cepstral Coefficient ,SVM-Support Vector Machine, ENF-Electric Network Frequency**,

I. INTRODUCTION

The ENF discontinuity method provides a visual aid to detect the phase changes in the original audio recording. The sudden phase changes in the waveform provide us the information about the Editing points where the original recording is tempered; however it fails to perform if the recording is done with high quality audio devices. In existing system **Acoustic Reverberation Estimation Method:** The Reverberation Time (RT) is calculated without the prior information of microphones used or the dimensions of the room. Here the tail of the reverberation waveform was modeled as an exponentially damped Gaussian white Noise process. The RT estimate was calculated by maximum likelihood estimation

II. PROPOSED SYSTEM

The proposed system can be divided into three subsystems:
1) The background noise estimation subsystem,

2) The blind dereverberation subsystem, and
3) The feature extraction, fusion, and classification subsystem

III. BLIND DE-REVERBERATION SUBSYSTEM

This subsystem estimates reverberation signal,

$hRIR(t)\ .hMic(t)\ .s(t)$

from the audio recording, $y(t)$ which embodiesacoustic environment characterization. It is therefore reasonableto focus on $hRIR(t)$. The proposed audio recordingmodel indicates that it is hard to estimate, *hRIR* directly fromaudio recording, as

$y(t)=s(t)+hRIR(t)*hMic(t).s(t)+\text{hMic}(t).\eta(t)+\eta\text{Mic}(t)+\eta\text{TC}(t).$

To get rid of microphone impulse response and microphone and transcending distortion, flat microphone impulse response and negligible distortions are assumed. This is a reasonable assumption, as have that identification performance is *m* independent of the microphone used. The $y(t)$ now can be expressed as

$y(t) = s(t)+r(t) = s(t)+hRIR(t)* s(t),$

where(t) represents direct sound component, also referred as dry signal and $r(t)$ represents the reverberant component. The reverberant signal $r(t)$ can expressed as

$r(t) = s(t) * hRIR(t).$

In proposed system the acoustic environment signature is related to acoustic reverberation and background noise. Reverberation is\ the extended effect of sound after it is generated from the source. The acoustic reverberation is estimated by BD algorithm. In BD original dry signal is separated from the reverberation signal. The signal y(t) is the addition of dry signal s(t) and reverberation signalr(t). The main aim of dereverberation is to extract r(t)from an enhanced recording

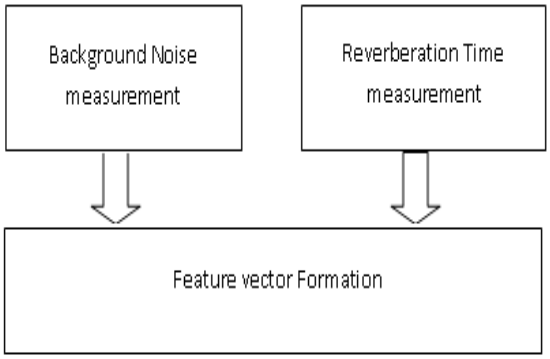


Fig4.1: General framework

V. ALGORITHM

- Calculation of reverberant signal.
- Spectral estimation is done by segmentation followed by temporal smoothing and conversion into frequency domain by MFCC and LMSC.
- Particle Filter initialization
- Particle evolution
- The prediction model is updated
- The noise is calculated in the form of samples and weights

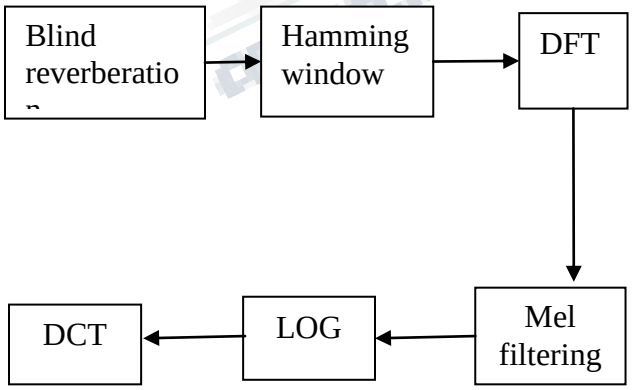


Fig5.1: Block diagram

Windowing the data makes sure that the ends match up while keeping everything reasonably smooth; this greatly reduces the sort of "spectral leakage" described in the previous paragraph. The FT of a finite length segment of sinusoid convolves the Fourier transform of the window against the sinusoid's frequency peak, since a property of the FFT is that vector multiplication in one domain is convolution in the other.

The FT of a rectangular window (which is what any unmodified finite length of samples in an FFT implies) is the messy looking Sinc function which splatters any signal that is not exactly periodic in the window over the entire frequency spectrum.

The FT of a Hamming shaped window concentrates this "splatter" much nearer to the frequency peak after the convolution, resulting in a fatter but smoother frequency peak, but a lot less splatter across frequencies far from the frequency peak.

This results in not only a cleaner looking spectrum, but also less interference from far away frequencies on any signal of interest. This interpretation (as opposed to the "infinitely repeating" interpretation) makes it more clear why differently shaped windows than Hamming may give you better results with even less "leakage". In particular, a Hamming window will reduce the size of the first Sinc side lobe of "leakage" right next to the frequency peak in exchange for actually more "leakage" (or convolution splatter) far from the frequency of interest. The idea of autocorrelation is to provide a measure of similarity between a signal and itself at a given lag.

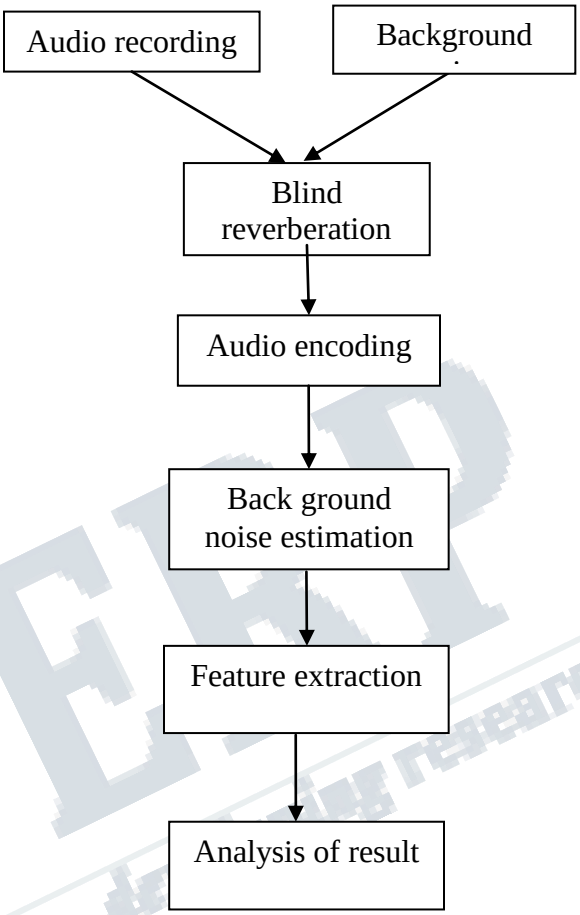
There are several ways to approach it, but for the purposes of pitch/tempo detection, you can think of it as a search procedure. In other words, you step through the signal sample-by-sample and perform a correlation between your reference window and the lagged window. The correlation at "lag 0" will be the global maximum because you're comparing the reference to a verbatim copy of itself. As you step forward, the correlation will necessarily decrease, but in the case of a periodic signal, at some point it will begin to increase again, then reach a local maximum. The distance between "lag 0" and that first peak gives you an estimate of your pitch/tempo. Twelfth-order autocorrelation coefficients are found, and then the reflection coefficients are calculated from the autocorrelation coefficients using the Levinson-Durbin algorithm. The original speech signal is passed through an analysis filter, which is an all-zero filter with

coefficients as the reflection coefficients obtained above. The output of the filter is the residual signal.
Spectral subtraction is a method for restoration of the power spectrum or the magnitude spectrum of a signal observed in additive noise, through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. The noise spectrum is usually estimated, and updated, from the periods when the signal is absent and only the noise is present.

The assumption is that the noise is a stationary or a slowly varying process, and that the noise spectrum does not change significantly in between the update periods. For restoration of time-domain signals, an estimate of the instantaneous magnitude spectrum is combined with the phase of the noisy signal, and then transformed via an inverse discrete Fourier transform to the time domain.
Feature extraction is a process that extracts data from the voice signal that is unique for each speaker. Mel Frequency Cepstral Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech. [6], [7], [8]

These extracted features are Vector quantized using Vector Quantization algorithm. Vector Quantization (VQ) is used for feature extraction in both the training and testing phases. It is an extremely efficient representation of spectral information in the speech signal by mapping the vectors from large vector space to a finite number of regions in the space called clusters. [6], [8] After feature extraction, feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features with the database using the DISTMIN algorithm

VI. FLOW DIAGRAM



Reverberation Time (T60) is an important measure of the acoustic properties of a room. It can provide information about the acoustic environment, the intelligibility, and quality of speech recorded in the room, and helps improve the performance of speech processing algorithms with reverberant speech. Where the acoustic impulse response of the room is not available, the T60 must be estimated no intrusively from reverberant speech. State-of-the-art non-intrusive T60 estimators have been shown to be strongly biased in the presence of noise. We describe a novel T60 estimation algorithm based on spectral decay distributions that provides robustness to additive noise for a range of realistic noise types for signal-to-noise ratios in the range 0 to 35dB and T60s between 200 and 950ms. The proposed method also has much reduced computational cost.
A speech signal, x(n) produced at a given position in a room will follow multiple paths to any observation point comprising the direct Pathas well as rejections from walls and other surface sin the room. The reverberant signal, y(n), captured by a microphone in the room is characterized by the Acoustic Impulse Response (AIR), h(n), of the acoustic channel between the source and microphone, such that

$y(n) = x(n) * h(n) + v(n)$….. (1)

where v(n) is additive noise at the microphone. The AIR is a function of the room geometry, the receptivity of the walls and other surfaces, the location of the microphone, and the distance from the microphone to the source. Reverberation Time (T60) is defined as the time taken for a sound todecayby60dBafterthesourcehasabruptlyceased.Itcanprovide important information about the acoustic environment, the intelligibility and quality of speech recorded in the room, and can be used toimprovetheperformanceofspeechprocessingalgorithmswithr everberantspeechsuchasspeechrecognition [1]and de-reverberation [2,3,4]. T60 can be characterized by the Sabineor Syringe quations [5, 7], and in contrast to the AIR, T60 measured in the diffuse sound field is independent of the source to microphone configuration. Standardized methods exist for estimating T60 from a measured AIR [8] such as [9]. In many practical situations, the AIR is not available and so T60 must be estimated non-intrusively from reverberant speech. Existing algorithms for estimation of T60 nonintrusive include [4]. However, all of these methods except have been shown to give strongly biased estimates of T60 in the presence of high levels of additive noise , whilst was shown to be robust to noise but has a large variance with respect to different speakers. The key contribution soft hi spaperare: toproposea T60 estimation method employing Spectral Decay Distributions (SDD) which is substantially robust to additive noise thus avoiding problems of bias in existing estimators; to show how the cost of computing the SDD can be reduced.

VI. SUMMARY

In ENF method FFT is used to estimate the periodicity of a small time frame. In ENF discontinuity method the discontinuities in the phase waveform after DFT analysis provide the editing points in the form of insertion and deletions in the original recording. If the threshold is low the large number of samples is classified by guessing only and on the other hand if the threshold is high the amount of signal in the FFT results increases and the amount of noise decreases. This degrades the performance of the system in terms of accuracy. To overcome these difficulties, acquisition device identification method (ADIM) is used. In ADIM, idea is to find out the devices used for recording of the evidence. RSFs can identify acquisition devices in a better way than MFCC approach. This helps to increase the accuracy of RSF method. RSFs provide best performance in accuracy if given to SVM classifier. It also provides good accuracy using SRC classifier. This RT estimation method was used to extend the use of decay curve to scenarios where there are no input signals present to conduct a reverberation experiment, However it requires the high computational cost for implementation because of the iterative solution of MLE equation and suitable for passive sounds only. In AADI method recording environment identification accuracy is modeled as a function of number of iterations and microphone type. For increasing number of iteration the identification of locations is less provided that the number of actual locations is kept constant. In acoustic environment traces method the blind estimation of Acoustic Environment Identification uses five environments need to increase further. The full blind setting provides successful identification of the environments for original recordings.

VII CONCLUSION

Partial results of proposed algorithm are given below; still need to apply the algorithm for various databases to find optimal parameter setting to generalize it.
Acoustic reverberations and background noise are used to characterize the acoustic environment. Background noise is modeled using a dynamical system and estimated using particle filtering. The proposed system is strong against MP3 compression attacks. The audio recordings are taken from a database so they are not real time. It is essential to develop an algorithm for real time recording analysis.

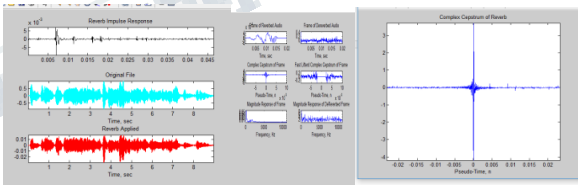


Fig 7.1 Simulated Output Results

REFERENCES

[1] m. delcroix et al.: dereverberation of speech signals based on linear prediction

[2]International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, No.5 (2013)

[3]Advanced Digital Signal Processing and Noise Reduction, Second Edition. Saeed V. Vaseghi

[4]ieee/acmtransactionsonaudio,speech,andlanguageprocessi ng,vol.24,no.2,february2016

[5] Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013

[6] D. Bees,M. Blostein, and P. Kabal, "Reverberant speech enhancement using cepstral processing," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., Apr. 1991, pp. 977–980.

[7] M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, "A method based on the MTF concept for dereverberating the power envelope from the reverberant signal," in Proc IEEE Int. Conf. Acoust., Speech, Signal Process., Apr. 2003, pp. 840–843.

[8] T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity based blind dereverberation for single-channel speech signals," IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 1, pp. 80–95, Jan. 2007.

[9] M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am., vol. 37, pp. 409–412, 1965.

[10] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien, Jr. C.R.Lansing,andA.S.Feng, "Blindestimationofreverberationtime," J.Acoust.Soc.Am.,vol.114,no.5,pp.2877–2892, Nov. 2003.