

Dereverberation of Speech using LPC based Approaches

K V S Manoj Kumar, A S N Murthy, Dr. D Elizabeth Rani

K V S Manoj Kumar, *Master of Technology, Digital Systems and Signal Processing, GITAM University, Visakhapatnam.*

A S N Murthy, *Sr. Assistant Professor, ECE Department, GITAM University, Visakhapatnam.*

Dr. D Elizabeth Rani, *Professor, EIE Department, GITAM University, Visakhapatnam.*

Abstract - In this paper, we proposed a method of speech processing degraded by reverberation. The processing method involves identifying and manipulating the linear prediction residual signal using Itakura Entropy Weighted Algorithm (IEWA). The Linear Prediction (LP) residual contains the original excitation impulses along with several other peaks due to reverberation. The weighted residual signal samples are used to excite a time-varying all-pole filter to obtain perceptually enhanced speech. The performance is evaluated through Signal to Reverberant component Ratio (SRR), speech waveforms, Spectrograms, and Inverse filter characteristics.

Key words—All-pole filter, Itakura Entropy Weighted Algorithm, Linear Prediction residual, Reverberation, Signal to Reverberant component Ratio (SRR).

I. INTRODUCTION

The quality of speech signal in enclosed spaces is degraded by additive noise and reverberation. In this paper we consider enhancement of speech under reverberant conditions. Reverberation occurs due to reflections of direct path of sound wave from surrounding walls and objects.

A reverberant sound is created in an enclosed space when a sound is produced causing a large number of echoes to build up and then slowly decay as the sound source stops but the reflections continue, decreasing in amplitude until they die. Normally, degraded speech is processed assuming that the degradation has long term stationary characteristics relative to speech.

The following Figure 1 shows the direct sound and its reflections. Reverberation is the collection of all those reflected sounds.

A reverberant signal is quite different from an echo signal. Echo is the case where the reflection of direct sound is heard or recorded after the sound of first syllable from direct path is heard or recorded, whereas in reverberation, it is heard before the completion of direct path sound.

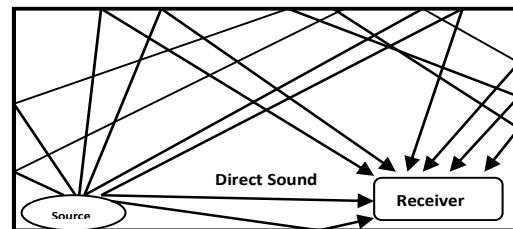


Figure 1: Direct and reflected speech signals reaching receiver

Reverberation Time RT_{60} is the time required for the reflections of a direct sound to decay 60dB. Basic factors that affect room's reverberation time include the size and shape of the enclosure as well as the materials used. Even people and their belongings also affect RT_{60} .

Reverberation is the process of multi-path propagation of an acoustic signal $s(n)$ from its source to one or more receivers. The observed signal at the receiver can be written as

$$x(n) = \sum_{m=0}^{\infty} h(m) s(n - m) \quad (1.1)$$

where $h(m)$ is the room impulse response.

In noise suppression and dereverberation, there is more emphasis on improving the overall SRR of the degraded speech. While attempting to reduce the degradation effects, the natural characteristics of the speech may change. In order to improve the overall SNR or SRR, it is necessary to reduce the noise in the low SNR regions.

Several microphone methods have been proposed for enhancement of speech degraded by room reverberation. The microphone array based methods enhance the signal in particular direction and suppress signals from other directions.

Methods focusing on characteristics of speech also have been proposed for enhancement of degraded speech. Two such algorithms used for enhancement are Linear Prediction Algorithm (LPA) and Itakura Entropy Weighted Algorithm (IEWA). These methods are mainly dependent on periodicity property.

LPA concentrates on calculating the LP coefficients and synthesizing the output frames from those coefficients via inverse filter designed depending on the characteristics of LP coefficients whereas the IEWA calculates the LP residual signal from the LP coefficients and modifies the LP residual, from

which the output dereverberated speech is synthesized.

The clean and reverberant speech signals are shown in Figures 2 and 3. The clean speech is, the utterance “One Two Three Four Five Six Seven Eight Nine Ten” by a male speaker.

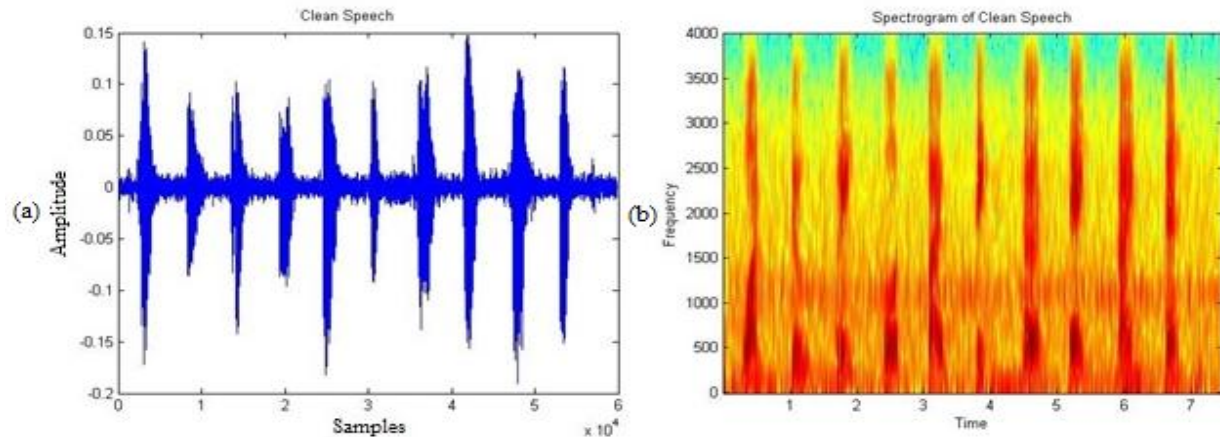


Figure 2: (a) Clean speech signal (b) It's Spectrogram

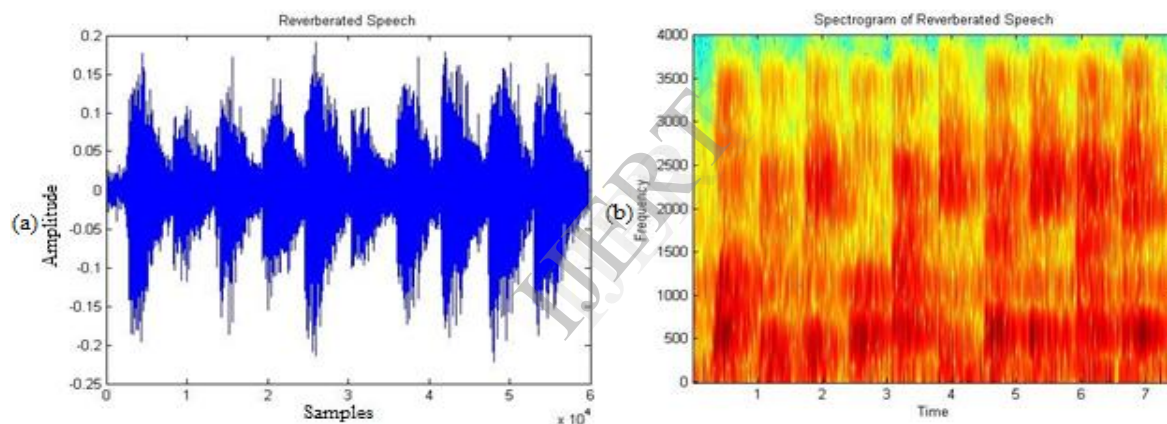


Figure 3: (a) Reverberant speech signal (b) It's Spectrogram

II. LINEAR PREDICTION AND LP RESIDUAL PROCESSING ALGORITHMS

The residual signal following LP analysis has been observed to contain the effects of reverberation, comprising peaks corresponding to excitation events in voiced speech together with additional peaks due to the reverberant channel. Several LP residual processing techniques have been developed using established models of speech production. These aim to suppress the effects of reverberation without degrading the original characteristics of the residual such that dereverberated speech can be synthesized using the processed residual and the all-pole filter resulting from LP analysis on the reverberant speech. The redundancy in the speech signal is exploited in the Linear Predictive (LP) analysis. The prediction of current sample as a linear combination of past ‘p’ samples form the basis of linear prediction analysis

where ‘p’ is the order of prediction. The predicted sample $\hat{s}(n)$ can be represented as

$$\hat{s}(n) = - \sum_{k=1}^p a_k s(n-k) \quad (2.1)$$

where a_k , $k=1,2,\dots,p$ are the prediction coefficients and $s(n)$ is the windowed speech obtained by multiplying short time speech frame with a hamming or similar type of window which is given by

$$s(n) = x(n) \cdot w(n) \quad (2.2)$$

where $w(n)$ is the windowing sequence.

The prediction error $e(n)$ can be computed by the difference between actual sample $s(n)$ and the predicted sample $\hat{s}(n)$ which is given by

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^p a_k s(n-k) \quad (2.3)$$

The primary objective of LP analysis is to compute the LP co-efficients which minimizes the prediction error $e(n)$.

(A) LP Residual

LP residual is the prediction error $e(n)$ obtained as the difference between the predicted samples $\hat{s}(n)$ and the current sample $s(n)$ which is as given in equation 2.3.

In the frequency domain, the above equation 2.3 can be represented as

$$E(z) = S(z) + \sum_{k=1}^p a_k S(z)z^{-k} \quad (2.4)$$

The transfer function of the LP error filter can be obtained as

$$A(z) = \frac{E(z)}{S(z)} = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.5)$$

(B) Linear Prediction Algorithm (LPA)

1. Input: Reverberant speech signal acquired through one distant microphone.
2. Divide the input speech into short 30ms frames and perform windowing using Hamming window.
3. Perform the LPC analysis on each windowed frame and calculate the LP Co-efficients.
4. Compute the Linear Prediction (LP) residual from the calculated LP co-efficients of each separate frame.
5. Synthesize the speech signal frames using the inverse filter designed using LP residual along with LP co-efficients.
6. Combine the synthesized frames to form the entire output speech.
7. Output: The dereverberated speech signal.

This algorithm uses the LP co-efficients and LP residual, calculated for each frame and synthesize the output dereverberated speech using the inverse filter designed using those LP parameters. As the LPC Analysis method predicts the current sample from the previous samples, the effect of reverberation can be estimated from the LP co-efficients and LP residual and is reduced using the inverse filter designed according to the data from LP parameters and filtering each speech frame.

The schematic diagram of the above explained algorithm is shown in the below Figure 4.

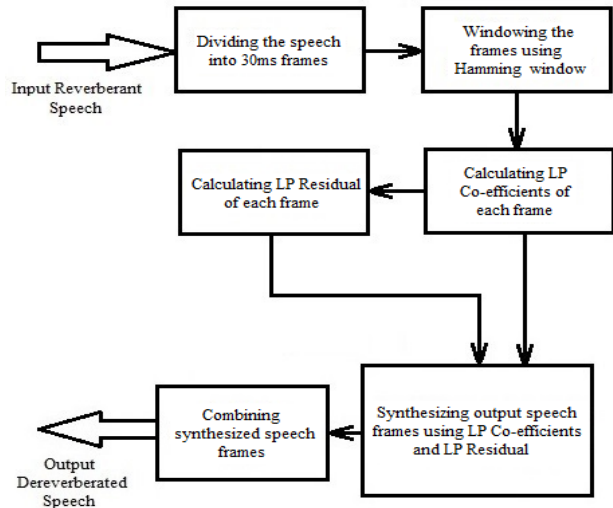


Figure 4: Schematic Diagram of the proposed algorithm with LPC Co-efficients and LP Residual

(C) Itakura Entropy Weighted Algorithm (IEWA)

1. Input: Reverberant speech signal acquired through one distant microphone.
2. Divide the input speech into short 30ms frames and perform windowing using Hamming window.
3. Perform the LPC analysis on each windowed frame and calculate the LP Co-efficients.
4. Compute the Linear Prediction (LP) residual from the calculated LP co-efficients of each separate frame.
5. Synthesize the speech signal frames using the inverse filter designed using LP residual along with LP co-efficients.
6. Compute the M -bin histogram of the samples in each frame of the LP residual signal.
7. Compute the entropy $H = -\sum_{i=1}^M p_i \log_2(p_i)$ and smooth entropy for each frame, where p_i is the estimated probability in the i^{th} bin of the histogram.
8. Compute the gross and fine weight functions by mapping the smoothed entropy to weight values using the functions

$$w_{gross} = \left(\frac{1-b}{2}\right) \tanh\left(-a_g * 3.14 * (H_s - a)\right) + \left(\frac{1+b}{2}\right) \quad (2.6)$$

$$w_{fine} = \left(\frac{1-b}{2}\right) \tanh(-a_g * 3.14 * H_s) + \left(\frac{1+b}{2}\right) \quad (2.7)$$

Where $a = 1.55$, $b = 0.05$, $a_g = 1.5$ are fixed parameters in Itakura weights calculation and H_s is the smoothened entropy.

9. Compute the overall weight function by multiplying the gross and fine weight functions.
10. The LP residual is modified by multiplying the overall weight function with the LP residual calculated earlier from LP co-efficients.
11. Synthesize the speech frames using the inverse filter designed using modified LP residual, using Itakura Entropy weighted algorithm, along with LP co-efficients.
12. Combine the synthesized frames to form the entire output speech.
13. Output: The dereverberated speech signal.

In this algorithm the LP residual signal calculated from LP co-efficients are modified using the Itakura Entropy Weighted algorithm. Here, the entropy in each frame is calculated from which the gross and fine weight functions are computed which further gives the overall weight function to be multiplied with the LP residual signal to form the modified LP residual signal. This modified LP residual signal along with the LP co-efficients is used to synthesize the dereverberated speech frames, by inverse filtering, which are to be combined to form the final dereverberated speech.

This algorithm gives better SRR compared to the former as the LP residual is modified according to the speech characteristics instead of blind inversion. The schematic diagram of the above explained IEWA is shown in the below Figure 5.

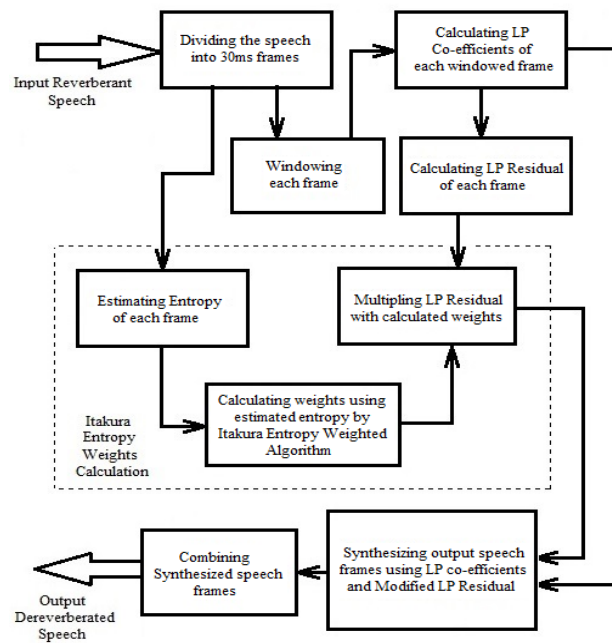


Figure 5: Schematic Diagram of the Itakura Entropy Weighted algorithm

III. SIMULATION RESULTS

In this section the performance of the proposed method is examined for processing speech data degraded by reverberation. The performance of the method is evaluated through subjective and objective analysis. For this purpose the reverberant speeches with different reverberation times were considered. As a part of the objective analysis, the Signal to Reverberant component Ratio of input reverberant speech and output dereverberated speech were evaluated and recorded. The corresponding results were shown in the Table 1.

Sl. No.	Input Speech	Input Speech SRR (dB)	Output Speech SRR (dB) LPA	Output Speech SRR (dB) IEWA
1	Malevoice	25.1281	39.5385	47.1161
2	Arena600	21.2372	22.7777	26.3143
3	Arena800	21.0868	22.3018	25.7898
4	Arena1300	18.3616	26.3200	27.3037
5	Arena2000	20.7818	23.2586	27.3982
6	Arena5000	24.1773	26.4172	29.8265

Table 1: Comparison between Input and Output Speech Signals for different inputs

From the analysis of the results, it is evident that the IEWA provides high SRR compared to LPA. Considering the malevoice referring to the above table, the waveforms and spectrograms of the input reverberant speech and the output dereverberated

speech from two algorithms are shown in the Figures 6 and 7 respectively. From the Figures it can be observed that LPA lacks sharpness and precision in the dereverberated speech whereas the IEWA provides dereverberated speech with sharp details and precision. From the subjective analysis it is confirmed that the quality and intelligence of IEWA is high. This is because LPA just uses the enhancement provided from LP analysis whereas the latter uses LP residual processing besides LP analysis.

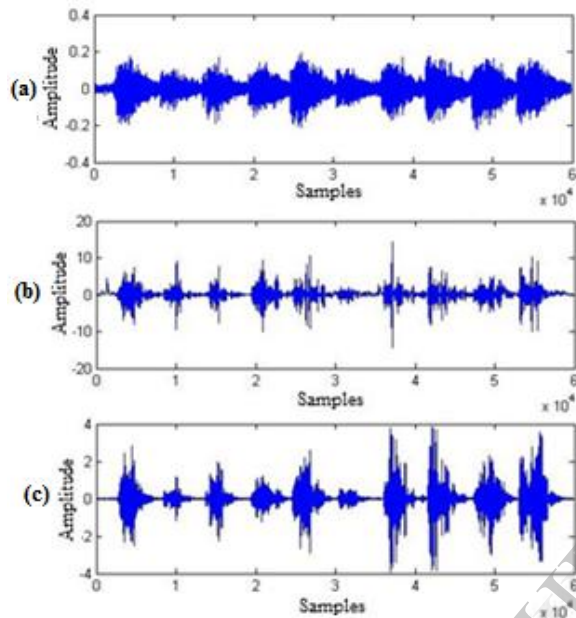


Figure 6: Wave forms:
(a) Input Reverberant speech (b) Output speech using LPA (c) Output speech using IEWA

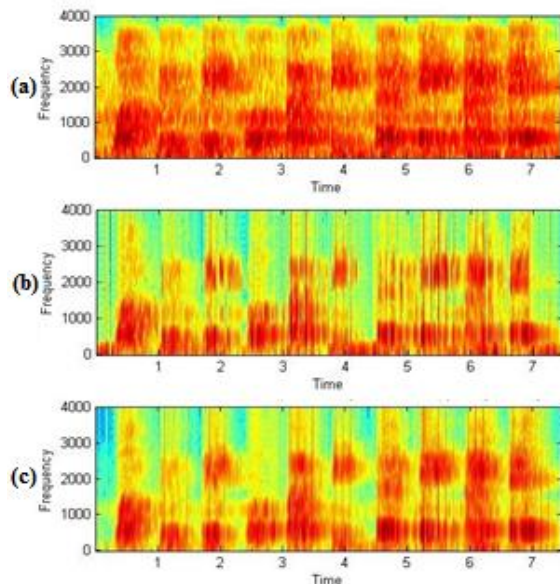


Figure 7: Spectrograms:
(a) Input Reverberant speech (b) Output speech using LPA (c) Output speech using IEWA

By subjective analysis, the performance of the two algorithms was verified for different reverberant speeches and is found and confirmed that there is increase in quality and intelligence of the processed reverberant speech. Both the algorithms improve the quality but the performance of IEWA is high compared to LPA which is evident from both subjective and objective analysis.

IV. CONCLUSION

In this paper, a new approach for processing reverberant speech via Itakura Entropy Weights is proposed. Experimental results show that the proposed method can be applied in speech recognition applications in which the speech signal is contaminated by reverberation. The processing was done by weighting the LP residual signal and the weight function was derived using the characteristics of the reverberant speech. The resulting signal shows reduction in the perceived reverberation without significantly affecting the quality. By adjusting the parameters used for obtaining the weight function, the comfort level in the processed signal can be traded with the distortion caused by the manipulation. Thus processing the LP residual signal provides an alternative approach for enhancing reverberant speech. In further work, we intend to evaluate the performance in real room responses considering an automatic speech segmentation strategy.

REFERENCES

- [1] Patrick A. Naylor and Nikolay D. Gaubitch, "Speech Dereverberation", *Springer, London, 2010*.
- [2] J. Benesty, S. Makino, J. Chen, "Speech enhancement signals and communication technology," *Springer, New York, 2007*.
- [3] Jacob Benesty, M. Mohan Sondhi, and Yiteng Huang, "Springer Handbook of Speech Processing", *Springer, Berlin, 2008*.
- [4] A.U. Suryavamsi, "Blind deconvolution and adaptive algorithms for de-reverberation," *Blekinge Tekniska Högskola, 2012*.
- [5] Kinoshita, Delcroix, Nakatani and Miyoshi, "Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction", *IEEE Transactions on Audio, Speech, And Language Processing*, Vol. 17, No. 4, pp. 1-12, May 2012.
- [6] Thierry Dutoit, Ferran Marques, "Applied Signal Processing, A MATLAB Based Proof of Concept", *Springer Science+Business Media, 2009*.
- [7] B. Yegnanarayana and P. Satyanarayana Murthy, "Enhancement of Reverberant Speech using LPC Residual Signal", in *IEEE Transactions On Speech And Audio*

Processing, Vol. 8, No. 3, May 2000.

- [8] Rajan S. Rashobh, Andy W. H. Khong and Patrick A. Naylor, "Adaptive Blind System Identification for Speech Dereverberation Using Apriori Estimates", in *IEEE Transactions On Speech And Audio Processing, Vol. 1, No. 10, May 2010.*
- [9] V. A. Zverev, "Blind Dereverberation of a Speech Signal", in *Akusticheski Journal, Vol. 54, No.2, July 2008.*
- [10] Marco Jeub, Magnus Schäfer and Thomas Esch, "Model- Based Dereverberation Prerving Binaural Cues", in *IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 7, September 2010.*
- [11] Bradford W. Gillespie, Henrique S. Malvar and Dinei A. F. Florencio, "Speech Dereverberation via Maximum-Kurtosis Sub-band Adaptive Filtering", in *IEEE Transactions On Speech And Audio Processing, Vol. 4, No. 10, May 2001.*
- [12] Harish Padaki, Karan Nathwani and Rajesh M Hegde, "Single Channel Speech Dereverberation Using the LP Residual Cepstrum", in *IEEE Transactions on Speech and Audio Processing, Vol. 8, No. 13, 2013.*
- [13] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Multi-step linear prediction based speech dereverberation in noisy reverberant environment", in *Proc. Interspeech, 2007*, pp. 854–857.
- [14] B. Yegnanarayana, C. Avendano, H. Hemansky and P. Satyanarayana Murthy, "Processing Linear prediction Residual for speech enhancement," in *Proc. EUROSPEECH'97, Patras, Greece, Sept. 1997*
- [15] K. Kinoshita, T. Nakatani, and M. Miyoshi, "Spectral subtraction steered by multi-step linear prediction for single channel speech dereverberation", in *IEEE Int. Conf. Acoust., Speech, Signal Processing. (ICASSP), 2006*, vol. 1, pp. 817–820.
- [16] T. F. Quatieri, *Discrete-Time Speech Processing: Principles and Practice. Upper Saddle River, NJ: Prentice-Hall, 1997.*



Signal Processing, Network Theory.

K V S Manoj Kumar is presently pursuing M.Tech in the specialization of Digital Systems & Signal Processing. He received his B.Tech degree from JNTU Anantapur. His areas of interest are



26 years of teaching experience in India and abroad. His areas of interest are Digital Signal Processing, and Speech Signal Processing.

A S N Murthy is presently working as Sr. Asst. Professor in the dept. of ECE, GITAM University. He submitted his PhD in Speech Signal Processing. He received his ME and BE degrees from Andhra University. He has



from Bharathiar University and BE from Madhurai Kamaraj University. Her areas of interest are Signal Processing, Communication Systems and Image Processing.

Dr. D Elizabeth Rani is presently working as HOD for EIE Dept. in GITAM University. She received PhD from Andhra University in Radar Signal Processing. She has 26 years of teaching and 15 years of research experience. She received her ME