Speech Reverberation: A Review

GituGeevarghese
Dept. Electronics and Telecommunication
Fr.C.R.I.T, sector 9A, Vashi-400703
Navi Mumbai, India

Dr. Milnd S. Shah
Dept of Electronics and Telecommunication
Fr.C.R.I.T, sector 9A, Vashi-400703
Navi Mumbai, India

***Abstract*—The aim of thispaper is to study the degrading effects of reverberation on the speech signal. Reverberation is one of the most important feature existing in a closed room. By estimating the reverberation time, the speech enhancement technique aims toimprove the speech. Speech reverberation is the total decay in the actual signal which is in the form of reflections of the target signal which degrades the quality of the speech.The objective of this review paper is to give an overview of reverberation, its effects on speech signal and estimation of the same. Reverberation is typically represented by the parameter RT_{60} Section I gives the introduction. Section 2 gives a brief overview of reverberation and its types. Section 3 describes the various techniques to measure reverberation. Section 4 is the speech models used for the estimation of spectral parameters. Section 5 deals with the estimation of blind RT_{60} reverberation. Section 6 gives the summary and conclusion**

***IndexTerms*— Reverberation, Reverberation time and its estimation, speech system modeling, Blind RT_{60} estimation**

I. INTRODUCTION

We have all experienced difficulty in understanding someone speaking in a reverberant environment like gymnasium stairwells, squash courts etc. Speech signals which are being captured in an enclosed environment contain reverberation from the surrounding objects. This corrupts the original speech signal of its quality and intelligibility which can degrade the performance of many applications such as hands-free teleconferencing and automatic speech recognition. In many applications one important criteria to improve the system performance is by estimating the room acoustical parameters as explained in ref [3]. Among the various acoustical parameters reverberation time in particular should be estimated by using the most appropriate techniques mentioned in e.g., [16] and [17].

Reverberation according to psychoacoustics is the persistence of sound after a sound is being produced. This persistence is caused when the signal is reflected causing a large number of reflections to build up and then decay as the sound is absorbed by the surfaces of the objects in space such as furniture, people and air. These reflections continue to exist even after the source sound stops but the reflections continue until its amplitude decreases to zero amplitude [18]. Reverberation time is the most important parameter which defines the room impulse response (RIR) characteristics. The reverberation time, represented as RT_{60} is the time at which the remaining RIR power level falls by 60dB lower than the total RIR power for every doubling of distance from the lips. Reverberation time is a rough measure of the reverberant properties of the room.

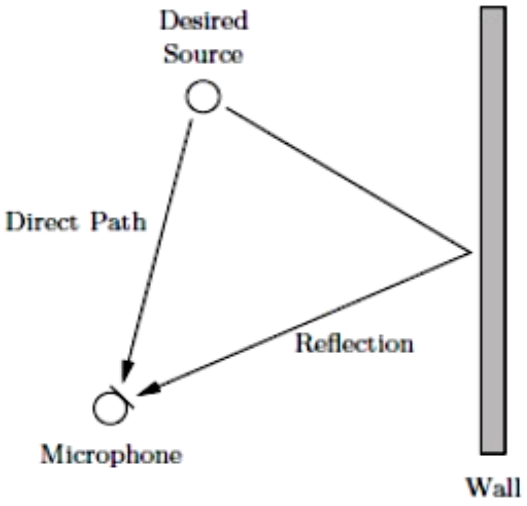

Figure1. Illustration of the direct path and a single reflection from the desired source to microphone. (source [13])

The fig 1 explains the way sound travels in a closed room. When the desired source and microphone are in the line-of-sight the sound travelled is called as the direct path.Ifthe microphone and the desired source are not in the line-of-sight then the reflection of the target sound is received at the microphone crating a reverberation sound after the initial sound.

II. SPEECH REVERBERATION

Reverberations are the reflections that arrive at less than 50ms. It is a process in which multipath propagation of the target signal exists from its source to microphone. The received sound generally consists of direct sound, reflections that arrive soon after the direct sound (early reverberation) and reflections that arrive shortly after the early reverberation (late reverberation). Direct sound is the sound which is received without reflection, if there is no line-of-sight between the

source and the observer there is no direct sound present. Early reverberationis the sound which arrives a little time latter due to the reflections off one or more surfaces. The reflected sounds produced are separated in both time and direction from the direct sound. Early reverberation will vary as the source and the microphone moves within the space, which is not supposed to be a separate sound to the direct sound as long as the delay of the reflected sound does not exceed a limit of approx 80-100 ms with respect to the arrival time of the direct sound. Early reverberation actually supposed to emphasize the direct sound and therefore it is considered useful in regard to speech intelligibility. Early reverberation also causes a spectral distortion called colouration. 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.

III. MEASUREMENT OF REVERBERATION TIME

Reverberation time was traditionally determined analytically using the time domain RIR or the room geometry and wall absorption properties. The absorption property of a room is measured with the help of an absorption coefficient which is a number between 0 and 1, indicating the proportion of sound which is being absorbed by the surface compared to the proportion which is reflected back into the room. A large, fully open window would have an absorption coefficient of 1 which would offer no reflection as any sound reaching it would pass straight out and no sound would be reflected. On the contrary a thick, smooth painted concrete ceiling would be the acoustic equivalent of a mirror, and would have an absorption coefficient very close to 0. Most methods for the measurement of reverberation time try to measure the sound decay after switching off the excitation source. This requires a test signal by using an impulse measured by making a sufficiently large noise for which the decay is tracked after the sound dies. This method however is not possible in real-time signal processing. Semi-blind methods and statistical theory model have been developed which uses neural network methods to estimate the room characteristics. Various other methods are available which scans the speech signal to detect gaps allowing the smooth curve to be tracked. In most of the recent researches a robust and reliable method to blindly estimate the reverberation time from passively received microphone signals is suitable for improving the audio processing instruments.
However nowadays blind methods have been developed to perform RT_{60} estimation from the received speech signal.The work which is described in [2] uses blind reverberation suppression algorithm and its ideal counterpart to suppress the reverberation effect and to obtain robust speech.

IV. REVERBERATION MODELING IN SPEECH PROCESSING

The impulse response between two points in a room is normally divided into two parts, one constituting pf the direct sound and early reflections and the other of the late reverberation. In room acoustics modeling, these two parts are often implemented in different ways. This is done to make the simulation computationally efficient in order to render real-time processing of sound.

A. *Signal model:*

Lollmann and Vary [6] in their study on early and late reverberation created a speech enhancement model in which the distorted signal *x(k)* is given by the superposition of the reverberant signal *z(k)* and additive noise *v(k)*, where *k* denotes the discrete time index. The received signal *x(k)* and the original speech signal *s(k)* is given by:

$$x(k) = z(k) + v(k) = \sum_{n=0}^{L_R} s(k-n) h_R(n,k) = v(k) \quad (1)$$

Where $h_R(n,k)$ denotes the time varying RIR of length LR between source and receiver. The reverberant signal can be written as

$$z(k) = \sum_{n=o}^{L_e-1} s(k-n) h_R(n,k) + \sum_{n=Le}^{L_R-1} s(k-n) h_R(n,k) \quad (2)$$

The signal $z_e(k)$ is termed as early speech component and constitutes the source signal of the speech enhancement algorithm and $z_i(k)$ and additive noise *v(k)* denotes suppression of late reverberant speech by modeling both of them as uncorrelated random noise process so that the spectral enhancement techniques can be applied.

B. *Time-domain model:*

In the paper by E. S. Jan and H. Richard[7] they have assumed that the observed noisy and reverberant speech signal *x* to be the sum of a source signal *s* convolved with a RIR *h* and additive noise *d*, independent of *s*:

$$x(n) = \sum_{l=0}^{\infty} h(l) s(n-l) + d(n) = y(n) + d(n) \quad (3)$$

Where *n* is a discrete time sample index and *y* is the noise-free reverberant signal. RIR usually consists of a number of impulses for the early reflections and an exponentially decaying tail with a noise like appearance giving rise to the late reverberation.

C. *REMOS Modelling*:

REMOS(Reverberation Modeling for Speech Recognition) is basically a structure created for the robustness of distance-talking automatic speech recognition(ASR). The main idea for implementation of REMOS is to consider a reverberant utterance in the time frequency domain as the convolution of the utterance and the time-frequency representation of the room impulse response as shown in fig (2).

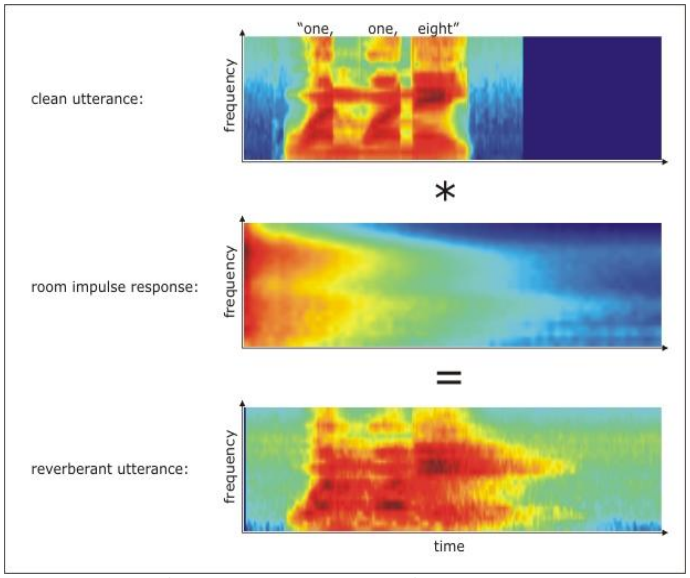

Fig.2: Illustration of the convolution in the time-frequency domain. Source [19]

The fig (2) shows the convolution process of the time-frequency room impulse response and the clean utterance and the resulting reverberant utterance.During the recognition process, REMOS inverts the convolution of the utterance andtime-frequency representation by separating each reverberant utterance by its room impulse response (RIR) and the clean part. To describe the acoustical parameters of the target room, one has to estimate the statistical reverberation model prior to the recognition process.

D. Room Impulse Response Model(RIR)

As mentioned in the paper [8] by Polack [9], he developed a time-domain model in which a room impulse response(RIR) is described as one realization of a non-stationary stochastic process. A simplified form of this can be expressed as:

$$h(t)=\begin{cases} b(t)e^{-\alpha t} & t\geq 0 \\ 0 & t<0 \end{cases} \tag{4}$$

Where *b(t)* is a white zero Gaussian stationary noise and α is linked to the reverberation time*Tr*through

$$\alpha \triangleq \frac{3In(10)}{T_r} \tag{5}$$

The energy envelope of the RIR can be expressed as:

$$E_h\{h^2(t)\}=\sigma^2 e^{-2\alpha t} \tag{6}$$

Where σ^2 denotes variance of *b(t)* and $E_h\{.\}$ denotes ensemble average over *h* i.e. over different stochastic process [10]. Different realizations of the stochastic process are obtained by either changing the position of the receiver or the source position keeping either of the two in a fixed position. The same stochastic process can be observed irrespective of the position, provided that the time origin be defined with reference to the signal emitted by the source and not with respect to the arrival time of the direct sound at the receiver. This means that we can assume ergodicity and evaluate the ensemble average according to [20]by spatial averaging.The RIR can be split into two main components as early echoes or early reverberant and later echoes or late reverberant as, $h_d(t)$ and $h_r(t)$ so that

$$h(t)=\begin{cases} 0 & t<0 \\ h_d(t) & 0\leq t\leq T \\ h_r(t) & t\geq T \end{cases} \tag{7}$$

The value *T* is chosen such that $h_d(t)$ consists of the direct signal and a few early echoes and *hr(t)* consists of all the later echoes, i.e. late reverberation. *T* usually ranges from 40-80 ms.

V. METHODS FOR BLIND RT_{60} ESTIMATION

A. Spectral Decay Distribution(SDD)

The method proposed by Wen et al.[12] is based on the assumption of a statistical model for the ASR and spectral decay distributions of the observed speech. By applying a least square linear fit to the log-energy envelope in each frequency band in the Discrete Fourier Transform (DFT) domain, frequency dependent decay rates can be estimated. To predict the reverberation time, the negative-side variance of the distribution of the decay rates is demonstrated to correlate with the room decay rate. This approach needs training so as to map these values from RT_{60} to negative-side variance. In the training period, the 2nd-order polynomial mapping function is calculated during reverberant speech signal with known RT_{60}. The 2nd-order polynomial mapping function is being calculates from a known RT_{60} of the reverberant speech.

B. Modulation Energy Ratio(MER)

Falk and Chan[13] proposed a non-intrusive quality measure for the dereverberated speech which is based on the speech-o-reverberation modulation energy ratio(SRMR). This method considers 23 acoustic frequency bands obtained from one gamma tone filter bank calculating the energy in eight modulation frequency bands varying logarithmically between 4 to 128 Hz. SRMR is the ratio of average energy in the low modulation frequencies to the high modulation frequencies. This leads to the observation that for low modulation frequency (4-18 Hz) energy is relatively insensitive to reverberation while for higher modulation frequencies (29-128 Hz), the energy is increasing almost linearly with RT_{60}.

Maximum Likelihood (ML)

Themethod which is proposed by L"ollmann et al. [14] uses a statistical model of the sound decay reverberant speech, similar to the reverberation model specified in method 1. This then is used to estimate RT_{60} by developing a maximum likelihood criterion. The speech samples are down-sampled before estimation in order to improve the computational efficiency. There is a pre-selection approach to detect plausible decays before these are used in ML estimation procedure. In order to increase the robustness, the estimated RT_{60} for each frame is used in a histogram and smoothing procedure. Unlike the previous two methods this method does not require training.

However simulations based on the direct identification of the equalizer filter(s) seem more effective.Blindereverberation techniques based on kurtosis maximization have been shown to be practically usable for the reduction of the early reverberant part in a single channel dereverberation algorithm. As a general observation, it is difficult to draw conclusions on the comparative performanceof different dereverberation algorithms, since several dereverberation metrics are used.

VI. CONCLUSION

This is a review paper on the effects of reverberation on speech intelligibility. Traditional Reverberation time estimation methods were being described among which semi blind reverberation and blind reverberation estimation were discussed, among which blind reverberation techniques have been developed in the recent times. The different speech models were studied which is used for the basic construction of a linear system to estimate the reverberation time. The methodology involved in the estimation of reverberation time RT_{60} using the first two methods, the third method has less computational complexities. blind reverberation estimation is more appropriate for estimating the reverberation time.

REFERENCES

[1] S. Chukiet, "Effects Of Classroom Reverberation And Listeners Location To Speech Intelligibility" in *ECTI-CON* ,Phetchaburi , 2012, pp.1- 4.

[2] M. Roland, A. P. H. Emanuel, S. Armin and K. Walter,"OnThe Application Of Reverberation Suppression To Robust Speech Recognition," in *Proc. ICASSP*, Kyoto, 2012, pp. 297 - 300.

[3] K. Abbas, M. Saeed, B. Mehrzad, T. G. Aaron and E. Morteza, " Speech –Model Based Accurate Blind Reverberaion Time Estimation Using An LPCFilter,"IEEE *Trans. Acoust., Speech, Signal Processing*, vol. 20, no. 6, pp. 1884-1893, 2012.

[4] H. Y. Luo and P. N. Denbigh," a speech separation system that is robust to reverberation," in *Proc. Symposium on Speech, Image processing and Neural Networks*, (Hong Kong), 1994, pp. 339-342.

[5] B. F. Steven, F. Ercolino and P. Tracy, "Improving Synthetic Speech Quality Using Binaural Reverberation," in *Proc. ICASSP*, 1976, pp 705-708.

[6] L. W. Heinrich and V. Peter, "A Blind Speech Enhancement Algorithm For The Suppression Of Late Reverberation And Noise," in *Proc. ICASSP*, 2009, pp 3989 - 3992.

[7] E. S. Jan and H. Richard, "Single-microphone Late-reverberation Suppression In Noisy Speech By
Exploiting Long-term Correlation In The DFT Domain," in *Proc. ICASSP*, 2009, pp 3997 - 4000.

[8] E. A. P. Habets, "Multi-channel speech dereverberation based on a statistical model of late reverberation," in *Proc. ICASSP*, 2005, pp 173-1.

[9] J.D. Polack, *La transmission de l'energiesonoredans les salles*, These de doctoratd'etat, Universite du Maine, La mans, 1988.

[10] J. B. Allen,D. A. Brekley, and J. Blauret, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *journal of the acoustical society of america, vol. 62, no. 4*, pp.912-915, 1977.

[11]H. W. L¨ollmann and P. Vary, "Estimation of the Reverberation Time in Noisy Environments," *in Proc. of Intl. Workshop onAcoustic Echo and Noise Control (IWAENC)*, Seattle(Washington), USA, Sept. 2008.

[12] J. Y. C. Wen, E. A. P. Habets, and P. A. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. IEEE Intl. Conf. on Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, USA, Apr. 2008.

[13] P. Jaiswal and Dr. M. Kar, "A review of different approaches applied for the estimation of reverberation time",in*ProcIJSETR*, 2013 vol. 2, Issue. 8, pp. 1612-1615.

[14] H. W. Lollmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel-Aviv, Israel, Aug. 2010.

[15] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O'Brien,Jr., C. R. Lansing, and A. S. Feng, "Blind estimation of reverberation time," *J. Acoust. Soc. Am., vol. 114, no. 5*, pp. 2877–2892, Nov. 2003.

[16] L. Couvreur and C. Couvreur, "Blind model selection for automatic speech recognition in reverberant environments," J. VLSI Signal Process., vol. 36, no.2/3, pp.189-203, Feb 2004.

[17]J. Gammal and R. Gorbran, "Combating reverberation in speaker verification," in Proc. IEEE Conf. Instrum. Meas. Technol., May 2005, pp. 687-690.

[18] V. R. R. Datla, "Implementation and evaluation of spectral subtraction (SS) with minimum statistics and wiener beamformer combination," M.S. thesis, BTH, Sweden.

[19]Prof. Dr. I. W. Kellermann and R. Maas. *Reverberation modeling in speech recognition*[online].Availablehttp://www.lms.lnt.de/en/research/activity/audio/signal/nachhall.php.

[20] K. Lebart and J. M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *ActaAcoustica*, vol. 87, pp. 359-366, 2001.