Sensitivity of Pitch Frequency Estimation to Reverberation Effect

Badawi Neyazi Badawi Ghandour
Productivity and Vocational Training Department, Ministry of Industry, Egypt.

Adel Zaghloul Mahmoud
Electronic and Communication Department, Faculty of Engineering, Zagazig University, Egypt.

Hoda Soliman Seddeq
Acoustic Department, Housing and Building National Research Center, Egypt.

El-Sayed M.El-Rabie
Electronic and Communication Department, Faculty of Engineering, Menofia University, Egypt.

Moawad I. Dessouky
Electronic and Communication Department, Faculty of Engineering, Menofia University, Egypt.

Fathi E. Abd El-Samie
Electronic and Communication Department, Faculty of Engineering, Menofia University, Egypt.

Abstract

This paper investigates the effect of reverberation on pitch frequency estimation. The autocorrelation function (ACF) method is used for pitch frequency estimation with and without reverberation effect. This paper modeled the reverberation effect on speech signal using a comb filter. The estimation error percentage of a comb filter at the length mild 8, moderate 10, and severe 12 in these scenarios have been investigated. The accuracy of the pitch frequency estimation is evaluated for the different scenarios for further accurate speech processing and verification.

1. Introduction

Speech recording can be performed in open areas or closed areas. Open area recording does not suffer from reverberation effect so, features can be extracted directly from speech signals for further processing. On the other hand, closed room speech recording is subject to some sort of reverberation due to multiple reflections. The direct signal combined with the multiple reflections constitute, there for Called reverberant signal [1].

Speech recording is normally performed for further speech processing application so, knowing the effect of reverberation is very necessary for further accurate processing of speech signals. Reverberation effect is in fact some sort of multiple reflections with decaying energy. A very important parameter that characterized reverberation is the reverberation time. It is defined as the time taken by the signal to decay to 60 dB from its initial value at detection. The long reverberation time is an indication of the severity of the reverberation effect and the poor quality of the recorded speech signal So, This degree of severity, in taken, affects the further signal processing tasks applied to speech signal [2].

Figure (1, 2), gives an illustration of the reverberation effect and signal processing representation of this effect. To discriminate between echo and reverberation, the echo is a signal reflection of the original speech signal, while the reverberation accounts for multiple reflection, and in some cases reinforcement of the signal. This can be modeled as additional sound sources added to the system. The simple simulation of echo [3, 4] is given by

Output = Input + Delayed input × Gain

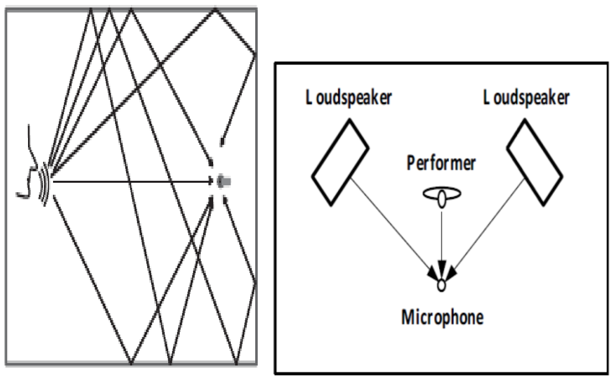


Fig. 1 reverberant recording environments the signals arriving at the microphone after one or more reflection.

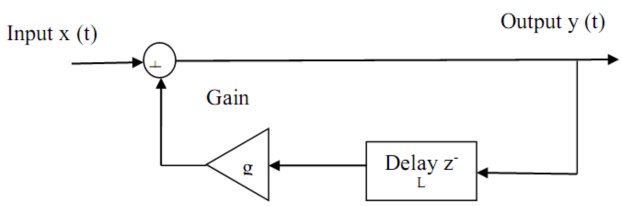
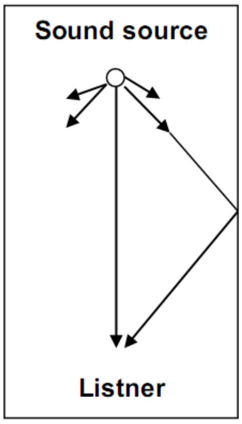


Fig.2 Simplified block diagram representing reverberation [5].

Figure (3), gives an illustration of the echo effect some phase changes occur in both echo and reverberation effects. These changes are audible to listeners [6].

Output = Input + Delayed Input × Gain



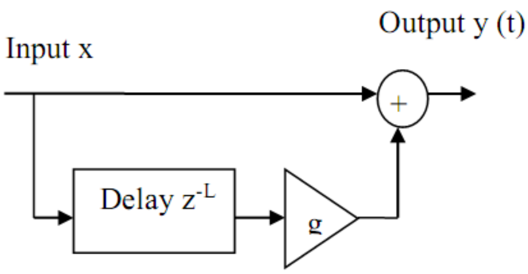


Fig. 3 represented echo feed forward signal processing.

An important parameter to characterize the reverberation effect is the critical distance. It is the distance at which the energy of the direct path is equal to the combined energy of the early and late reflection [7].

The effect of reverberation on speech signal parameters is the issue of this paper one of the main parameters of speech signals that characterize. Speech signal is the pitch frequency or the fundamental frequency. It is very important in applications such as speech coding and speaker identification. It is affect with the reverberation effect. This paper is mainly concerned with this effect to compensate for it in further speech processing tasks [8].The rest of the paper is organized as follows. Section 1 covers the reverberation effect. Section 2 summarizes some principles of room acoustics. Section 3 discusses the reverberation effect which is modeled by the speech through a comb filter. Section 4 discusses the methodologies for reverberation time estimation. Section 5 gives procedures for pitch frequency estimation that can be applied with and without reverberation. Section 6 discusses the sensitivity of estimated pitch frequencies to the reverberation effect. Section 7 gives some simulation results. Section 8 gives the concluding remarks.

2. Reverberation Effect

Reverberation effect on the human auditory system is tow ford coloration and echo. Coloration is defined by estimating the direct-to-reverberation ratio. Audible temporal smearing is induced due to early reverberation. The reverberation leads to sever effect on the performance of automatic speaker recognition system [9].

3. Room Acoustic and Reverberation Effect:

Normally, reverberant speech signals are recorded in closed rooms. These signals can be modeled as follows [9]:

$$y(n) = x(n) * h(n) \qquad (1)$$

where x (n) is the original speech signal and h (n) is the room impulse response. If the impulse response h (n) is long, it destroys the original speech signal characteristics [10].

4. Reverberation as a Comb Filter

The reverberation as a comb filter is in fact a multi band filter represented as [11, 12]:

$$H(z) = 1 - Z^{-L} = Z^{L}-1 / Z^{L} \qquad (2)$$

Discrete domain, it is represented as:

$$y(n) = x(n) - x(n - L) \qquad (3)$$

where L is filter length, which is proportional to the reverberation time. Both magnitude and frequency responses of the comb filter of order 8 are given in figure (4) [13, 14]:

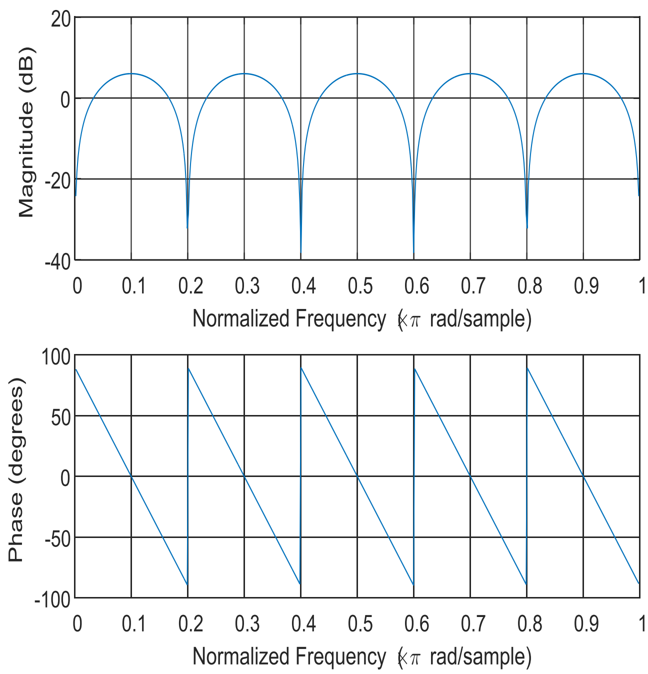


Figure (4) Magnitude and phase responses of a comb filter [15].

The comb filter is also represented by two sections as shown in figure (5).

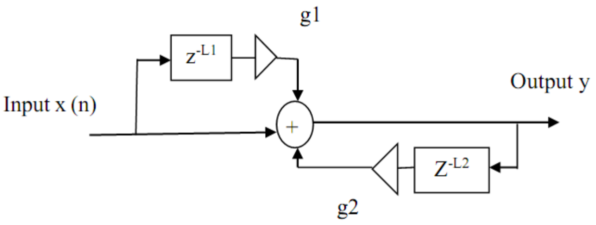


Fig.5 the general comb filters feedback and feed forward respectively [16, 17].

These two sections are feed forward and feedback sections, Simulating direct and reflected paths of speech signals.

5. Reverberation Time Estimation (T_R or T_{60})

The reverberation time in closed room can be estimated from the Sabine's formula [18, 19]:

$$R_T = (0.161\ V) / (A\ \acute{\alpha})\ s \tag{4}$$

$$(\acute{\alpha} = S^{-1} \Sigma\, \alpha_i\, S_i) \tag{5}$$

$$R_T(\text{sec}) = \frac{0.161\ V}{-A * \ln(1 - \alpha)} \tag{6}$$

where V is the volume of the hall in m^3, (A) is the surface area in m^2, and ($\acute{\alpha}$) is the average absorption coefficient, α_i is the absorption coefficient of area S_i .This formula is valid with $\acute{\alpha} \leq 0.15$. The following Eyring's formula (s) is used where $\alpha \geq 0.15$ and R_T becomes;

In the case of α equal one the value of R_T is zero, so it gives the physical meaning of the dead room. The reverberation time at the frequencies 500 Hz and 1000 Hz ($R_{T\ 500}$ and $R_{T\ 1000}$) give the average reverberation time, where [20, 21]:

$$R_T = 0.5\ (R_{T\ 500} + R_{T\ 1000}) \tag{7}$$

6. Proposed Method

The pitch is estimated with the autocorrelation function method for reverberant speech modeled with comb filter as shown in figure (6)

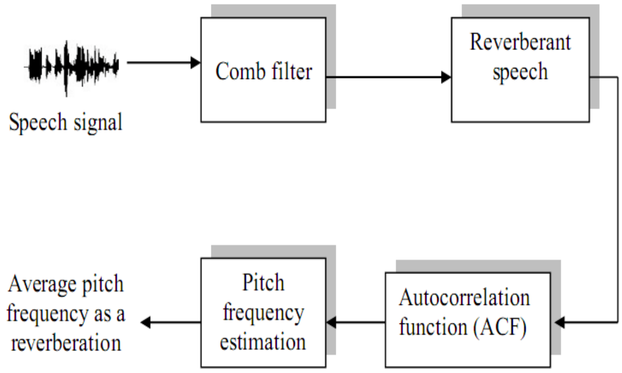


Figure (6) illustration the estimated average pitch frequency as a comb filter

6.1. Pitch Frequency Estimations

Pitch frequency is the fundamental frequency of the speech signal. Several techniques have been developed for this take. One of the most popular pitch frequency estimation methods is the AFC method represented with the following equation:

$$R\ (k) = \sum_{n=0}^{N-K-1} x(n)x(n+k) \tag{8}$$

The pitch is estimated by estimating the period between peaks in the auto correlation sequence.

6.2. Sensitivity of Pitch Frequency Estimation to Reverberation

It is expected that the multiple reflections in the reverberation effect will affect the fundamental frequency of the speech signal. This, in turn, will lead to some deviation the values of estimated pitch frequencies. The following section, will present some results regarding this issue.

7. Experimental Results

Several simulation experiments have been carried out to estimate the pitch frequencies from speech signals with and without reverberation. Moreover, the effects of a comb filter at the length mild 8, moderate 10, and severe 12 have also been considered.

Table (1), summarizes the average pitch frequencies estimated, and the average pitch frequency as a comb filter at the length mild 8, moderate 10, and severe 12 estimated at all scenarios from different speech signals.
The estimation errors for all scenarios are summarized in table (2):

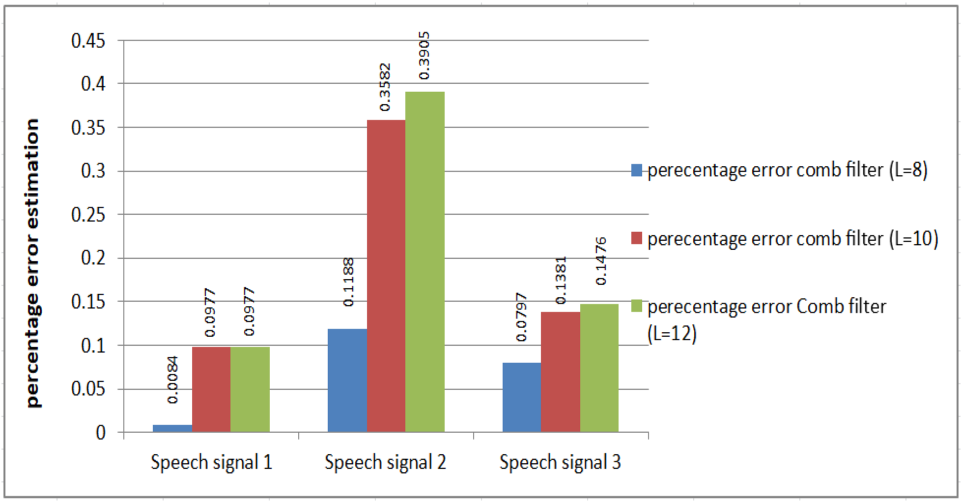
Table (2), Shows the estimated errors for all scenarios (i.e.., a comb filter at the length mild 8, moderate 10, and severe 12). The error percentage estimation can be compute it by the average pitch frequency estimated subtracted from the average pitch frequency as a comb filter for all scenarios and then divided on the average pitch frequency estimated.

Table1

Speech waves	Average Pitch Frequency estimation	Average Pitch Frequency Of comb filter with mild reverb L=8	Average Pitch Frequency Of comb filter with moderate reverb L=10	Average Pitch Frequency Of comb filter with severe reverb L=12
Speech signal 1	983.6066	1.0084e+03	983.6066	975.6098
Speech signal 2	1.1215e+03	860.2151	853.3344	860.2151
Speech signal 3	819.1126	926.6409	1.0667e+03	∞
Speech signal 4	960	1.0435e+03	956.1753	960
Speech signal 5	1.1163e+03	1.0526e+03	948.6166	952.3810
Speech signal 6	836.2369	898.8764	833.3333	839.1608
Speech signal 7	952.3810	909.0909	971.6599	1.0213e+03
Speech signal 8	851.0638	895.5224	851.0638	898.8764
Speech signal 9	857.1429	1.0526e+03	1.0526e+03	1.0526e+03
Speech signal 10	979.5918	991.7355	975.6098	975.6098
Speech signal 11	902.2556	1.0390e+03	1.1215e+03	1000
Speech signal 12	1.0860e+03	1.0480e+03	805.3691	805.3691
Speech signal 13	1.0345e+03	1.0909e+03	1.0300e+03	1.0169e+03
Speech signal 14	892.1933	916.0305	1000	995.8506
Speech signal 15	1.0526e+03	952.3810	991.7355	987.6543
Speech signal 16	1.0256e+03	1.0256e+03	971.6599	1.1594e+03
Speech signal 17	912.5475	923.0769	1.1268e+03	912.5475
Speech signal 18	1.0390e+03	875.9124	875.9124	869.5652
Speech signal 19	845.0704	879.1209	885.6089	909.0909
Speech signal 20	944.8819	860.2151	987.6543	991.7355
Speech signal 21	863.3094	1.0345e+03	1.0390e+03	1.0169e+03
Speech signal 22	1.0909e+03	1.1429e+03	1.0573e+03	1.0435e+03
Speech signal 23	948.6166	948.6166	1.0390e+03	963.8554
Speech signal 24	1.0169e+03	1.0084e+03	1.1163e+03	1.1163e+03
Speech signal 25	821.9178	919.5402	1.1163e+03	1.1429e+03
Speech signal 26	1.0390e+03	956.1753	895.5224	885.6089
Speech signal 27	963.8554	987.6543	987.6543	987.6543

Table 2

speech reverberation test wave	Error percentage estimation comb filter (L=8)	Error percentage estimation comb filter)L=10(	Error percentage estimation comb filter)L=12(
Speech signal 1	0.0252	0	0.0081
Speech signal 2	0.2330	0.2391	0.2330
Speech signal 3	0.1313	0.3023	∞
Speech signal 4	0.0870	0.0040	0
Speech signal 5	0.0571	0.1502	0.1468
Speech signal 6	0.0749	0.0035	0.0035
Speech signal 7	0.0455	0.0202	0.0724
Speech signal 8	0.0522	0	0.0562
Speech signal 9	0.2280	0.2280	0.2280
Speech signal 10	0.0124	0.0041	0.0041
Speech signal 11	0.1516	0.2430	0.1083
Speech signal 12	0.0350	0.2584	0.2584
Speech signal 13	0.0545	0.0043	0.0170
Speech signal 14	0.0267	0.1208	0.1162
Speech signal 15	0.0952	0.0578	0.0617
Speech signal 16	0	0.0526	0.1305
Speech signal 17	0.0115	0.2348	0
Speech signal 18	0.1570	0.1570	0.1631
Speech signal 19	0.0403	0.0480	0.0758
Speech signal 20	0.0896	0.0453	0.0496
Speech signal 21	0.1983	0.2035	0.1779
Speech signal 22	0.0477	0.0308	0.0454
Speech signal 23	0	0.0953	0.0161
Speech signal 24	0.0084	0.0977	0.0977
Speech signal 25	0.1188	0.3582	0.3905
Speech signal 26	0.0797	0.1381	0.1476
Speech signal 27	0.0247	0.0247	0.0247



FIR filters

Figure (7) comparison between the estimation errors for three different speech signals

The data provided consists of a training set, a development test set, and a (final) evaluation test set. The evaluation test set will be made available on Nov 5, 2013. Before distribution of the evaluation test set, the participants can develop their systems based on the training and development test sets. The development test set and the final evaluation test set each consist of different parts, namely simulated data, real recordings (Real Data) [22].

Figure (7), gives a comparison for the estimation error on three different speech signals. It is clear that the error is large in the presence of reverberation. This error can be compensated prior to any further signal processing.

From all obtained results, it is clear that the error in estimated pitch frequency depends on the degree of reverberation severity, there for it is necessary to compensate for the effect of reverberation prior to any further processing of speech signals.

8. Conclusion

This paper has investigated a very important issue in speech processing including the effect of reverberation on speech signal characteristics where the pitch frequency has been taken as a case study. This study proved that the reverberation has a large effect on the pitch frequency estimated. There error in pitch frequency estimation is increased with the increase of the degree of reverberation. By taking this effect into account, it is possible to compensate for it in any further speech processing applications.

References

[1] Fundamental of marine acoustics edited by Jerald W. Caruthers.

[2] Kuttruff, Heinrich, 2007,(Acoustics An Introduction), S. HirzelVerlag.

[3] J. Eaton, N. D. Gaubitch, and P. A. Naylor, "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost," Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.(ICASSP), pp. 161–165, 2013.

[4] B. Cauchi, H. Javed, T. Gerkmann, S. Doclo, S. Goetze, and P. Naylor, "Perceptual and instrumental evaluation of the perceived level of reverberation," Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP), 2016.

[5] S.worral "Echo and Reverberation", based on an experiment devolved by Texas instrument Inc., 2007 and modified 2011.

[6] A.Álvarez, V. Nieto, P. Gómez, and R. Martínez, "Speech enhancement based on linear prediction error signals and spectral subtraction,"in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2003, vol. 1, pp.123–126.

[7] Techniques FOR Noise Robustness IN Automatic Speech Recognition, first published 2013 © 2013 John Wiley & Sons, Ltd.

[8] Patrick A. Naylor · Nikolay D. Gaubitch Editors Speech De-reverberation

[9] *Masashi Unoki and Sota Hiramatsu "*MTF-based method of blind estimation of reverberation time in room acoustics" 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008.

[10] Thomas D. Rossing Neville H. Fletcher, "Principles of Vibration and Sound", Second Edition, New York : Springer, - Springer Nature Book Archives Millennium (2000-2004).

[11]Schroeder, M. R. (1962). "Natural sounding artificial reverberation", Journal of the Audio Engineering Society 10(3): 219-223.

[13] Smith, J.O, "Physical Audio Signal Processing", Nature Book, on web site, http://ccrma.stanford.edu/~jos/pasp/, online book, 2010 edition, accessed 26/5/2014.

[14] tomohiro nakatani, takuya yoshioka, keisuke kinoshita, masato miyoshi, biing-hwang juang, "real-time speech enhancement in noisy reverberant multi-talker environments based on a location-independent room acoustics model", IEEE international conference on acoustics, 2009.

[15] JON DATTORRO, "Effect Design -Part 1: Reverberator and other Filters," J.Audio Eng. Soc. 45(19):660-684, (1997).

[16] A. Krueger and R. Haeb-Umbach, "A model-based approach to joint compensation of noise and reverberation for speech recognition," in Robust Speech Recognition of Uncertain or Missing *Data*, Eds. Berlin, Heidelberg: Springer, 2011.

[17] Mark Kahrs, karelheinz Brandenburg,"Applications of Digital Signal Processing to Audio and Acoustics", pp. 85–131. Kluwer, New York, USA (1998).

[18] Kiran P. Kamble, Manik K. Chavan**, "**Endorsement to Audio Recorded in Different Acoustic Environment with Feature as Reverberation Time with Blind Reverberation Time Estimation Method", IEEE International Conference on Acoustics, Dec 20-22, 2017

[19] Felicia Lim, Mark R. P. Thomas and Ivan J.Tashev,"blur kernel estimation approach to blind reverberation time estimation", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[20] H. Kuttruff, "*Room Acoustics"*, Nature Book, London: Taylor & Francis Group, 2009.

[21] J.Eargle, "*Handbook of Recording Engineering"*, *4th Ed,* Nature Book, New York: Springer, 2005

[22] https://reverb2014.dereverberation.com/data.html