A Comparative Study between Pitch Detection Techniques on Reverberant Speech Signals

Shaimaa E. A. Aziz Hassan
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
engshaimaaelsayed@gmail.com

Adel S. El-Fishawy
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
aelfishawy@hotmail.com

Fathi I. Abd El-Samie
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
fathi_sayed@yahoo.com

A.M. Sharshar
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
A_m_sharshar@yahoo.com

Gaber El-Sayed Mohamed El-Abyad
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
Gaber_elabyad@yahoo.com

Moawd I. Dessouky
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt,
dr_moawad@yahoo.com

El-Sayed M. El-Rabaie
Communications and Electronics Department
Faculty of Electronic Engineering,Manoufia University:
Menouf, Egypt
El-sayed_Rabaie @yahoo.com

***Abstract*—Reverberation is one of the effects that occur regularly in closed room due to multiple reflections. This paper investigates the result of reverberation on both male and female speech signals. This effect is reflected in pitch frequency of speech signals. This parameter is important as it is usually used for speaker identification. Hence, several methods for pitch frequency estimation are investigated and compared on clear and reverberant male and female speech signals to select the one that is not affected so much by the reverberation effect.**

Keywords—Reverberation,pitch frequency.

I. INTRODUCTION

Reverberation is the survival of the sound after removing it [1]. The sound in the closed room will be reflected from surfaces. These reflections mix with each other for creating the phenomenon of reverberation, these reverberations are reduced as sound hit absorbent surface. The reverberation impulse response can be fractionated to echoes and late reverberation, these echoes (early reflections) depends on enclosure's size and also position of source and reception, those arriving after the direct sound about 50 to 80 msec [2]. Reverberation has an effect on the perceptibility of sound in an enclosure, especially if the reverberation time is impressively long, where RT is the time the sound pressure spends to minimize its level by 60 db. This is because the sound heard by person is a mixture of direct sound and time-delayed reflections [3]. Too much reverberation has a negative effect on the intelligibility of speech as reverberation deviates harmonic structure in voiced signals so, this impact increase the greatness of hearing deterioration.

Pitch or fundamental frequency (F_0) is an important parameter that characterizes the speech signal. It transfers multiple characteristics of the information transmitted by speech signal [4], so the pitch is the auditory quality of sound; it is a perceived fundamental frequency of sound. Pitch can be objectively characterized as the rate of vibration of the acoustic folds, so the pitch strength of voiced signal parts is an indicator of the level of reverberation [5].

The systematization of this paper is as follows; Reverberation model is illustrated in section II. Pitch frequency estimation algorithms will be demonstrated in section III. Results will be produced in section VI. Finally, the conclusions will be introduced in Section V.

II. Reverberation Model as a comb filter

In closed room, sound is reflected of all of the walls. The reverberation phenomena can be reduplicated by bolstering back the output through a postpone component and added with the input.

A. Comb filter structure

Comb filter is the traditional reverberator that can be generated when a signal is time postponed and added back to itself. Some frequencies will be declined and others will be reinforced, which can radically change the sound tonal shade. Comb filter structure is shown in Fig.1, Fig. 2 shows the magnitude response of comb filter and Fig. 3 shows the comb filter response that provides exponentially decaying impulse sequence. This decay time ("reverberation time") is used to define the gain of the feedback loop; the reflection density depends on the length of the delay line [6]. This paper investigates an artificial reverberation based in MTF formula that can be created by convolving the input signal with specific impulse response of an acoustic space, this achieves an acceptable reverberation.

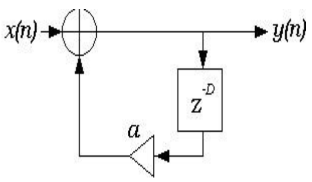

Fig.1: Comb filter structure.

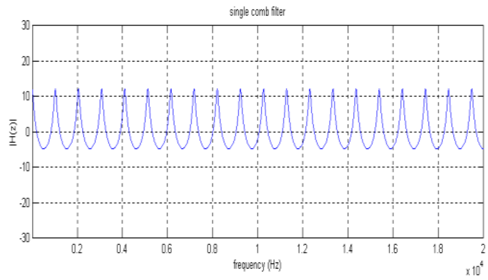

Fig. 2: Magnitude response of comb filter.

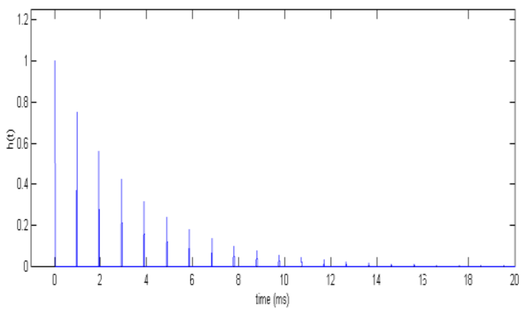

Fig. 3: Impulse response of comb filter.

Comb filter has a postpone D with a feedback of gain a, and as in Fig.3 the impulse response of comb filter has the same shape of the impulse response of a reverberation that can be partitioned into: the early reflexions and the late reverberation.

B. Comb Filter Equations

The output of the comb filter:

$$y[n] = x[n] + ay[n - D] \quad (1)$$

The transfer functions of comb filter:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - az^{-D}} \quad (2)$$

C. Reverberation Model Equation (MTF)

Convolve the input signal with impulse response of an acoustic space that represents how digital sample will behave in the acoustic space will create artificial reverberation.

The reverberant signal can be explained as shown in (3) and (4) represents the room impulse response of the closed room.

$$y(t) = x(t) * h(t) (3)$$

$$h(t) = q exp\left(-\frac{6.9t}{T_R}\right) n(t) (4)$$

where: $y(t)$ points to resulting reverberant signal, $x(t)$ represents the original signal, $n(t)$is white noise carrier, $h(t)$ is a stochastic (randomly) approximation of RIR, i.e., synthetic reverberant impulse response, T_R is the reverberation time, q is the amplitude of the speech signal [7].

III. PITCH FREQUENCY ESTIMATION

Pitch estimation points to evaluate the fundamental frequency $\boldsymbol{F_0}$for speech segments, Pitch is the quality that judgment sounds as "higher" and "lower" in the sense related with musical melodies. Many papers evaluate the pitch frequency of speech signal, using many algorithms but this paper will use Normalized Correlation Function (NCF), Cepstrum Pitch Determination (CEP), Summation of Residual Harmonics (SRH) andPitch Estimation Filter (PEF)

A. Normalized Correlation Function (NCF)

NFC is used to reveal the highest value of autocorrelation function of the specified speech signal [8]. NCF is estimated by:

$$R(m) = \frac{1}{N} \frac{\sum_{n=0}^{N-1-m} x(n).x(n+m)}{\sqrt{\sum_{n=0}^{N-1-m} x^2(n).\sum_{n=0}^{N-1-m} x^2(n+m)}}, 0 \leq m < M_0 \quad (5)$$

where N represents length of tested range and M_0points tothe evaluated autocorrelation points, m is the delay; pitch is the value of m which leads to the superior R (m).

B. Cepstrum Pitch Determination (CEP)

CEP is based on Cepstral signal analysis to find the pitch of signal. Cepstral coefficients are calculated as shown in(6)[9]:

$$c(\tau) = F^{-1}\{\log(|F\{x[n]\}|^2\} (6)$$

where F^{-1} denotes the inverse Fourier transform and x[n] is the discrete signal and $|F\{x[n]\}|^2$ is the power spectrum estimatedof the signal. Symbol $\tau$is the frequency and the magnitude of c (τ) is called the magnitude.A peak in the Cepstrum denotes that the signal is a linear combination of multiples of the pitch frequency. The pitch period can be calculated as the number of the coefficient where the peak occurs and the peak is located at the period of the fundamental frequency CEP block diagram is shown in fig.4.

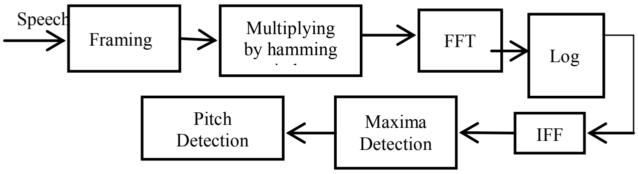

Fig. 4: CEP Block Diagram.

C. *Summation of Residual Harmonics (SRH)*

SRH is based on the harmonicsin the residual signal spectrum by using harmonicssummation to detect the fundamental frequency.

- Firstly estimation of the spectral envelope from S(t).
- Secondly the residual signal e(t) is obtained using inverse filtering.

As Hamming window frame covering several cycles of the resulting residual signal e (t) the amplitude spectrum $E(f)$ has flat envelope and presents peaks of fundamental frequency at harmonics. For each frequency of range $|F_{0,min}, F_{0,max}|$, the Summation of Residual Harmonics (SRH) is computed as following [10]:

$$SRH(f) = E(f) + \sum_{k=2}^{N_{harm}} \left[E(k.f) - E\left(\left(k - {}^{1}\!/_{2}\right).f\right)\right] (7)$$

This expression gets maximum at $f = F_0$ this is true for harmonics present in range$|F_{0,min}, F_{0,max}|$.

For this reason subtraction by $E\left(\left(k - {}^{1}\!/_{2}\right).f\right)$ allows reducing the importance of SRH at even harmonics. The estimated pitch value F_0^* is thus the SRH (F) at that time.

D. *Pitch estimation filter (PEF)*

PEF based on using comb-filter to calculate a weighted sum of harmonic amplitude in linear frequency domain, comb filter essential frequency initially unknown must match the pitch. The essential frequency of each frame is evaluated by convolving frame's power spectral density in the log frequency domain with the filter which sums the energy of the pitch harmonics while neglecting additive noise.

In the power spectral density domain,the signal at time t is [11]:

$$Y_t(f) = \sum_{k=1}^{k} a_{k,t}\delta(f - kf_0) + N_t(f) (8)$$

where $N_t(f)$ is the unwanted noise power spectral density, $a_{k,t}$ represents power of k^{th} harmonics. In log frequency domain the signal model is:

$$Y_t(q) = \sum_{k=1}^{k} a_{k,t}\delta(q - logk - logf_0) + N_t(q); q = logf (9)$$

Impulse response of the filter is $h(q) = \sum_{k=1}^{k} \delta(q - logk)$ and this is convolved with $Y_t(q)$, this convolution includes a peak at $q_0 = logF_0$.

IV. RESULTS AND DISCUSSION

In this section the reverberation effect is studied for two speech signals, one for female speech signal and the other for male speech signal at a reverberation time of .5 sec ,The speech signals is sampled at frequency 10KHZ , time is taken for (0:5) sec, amplitude is 5.The Matlab R2018a program is used for simulation. The first investigation concerns in effect of reverberation on the male and female speech, and then simulation of pitch frequency with different algorithms.

A. *Reverberation effect*

Fig. 5 and 6 shows the effect of reverberation on female speech signal and male speech signal respectively.

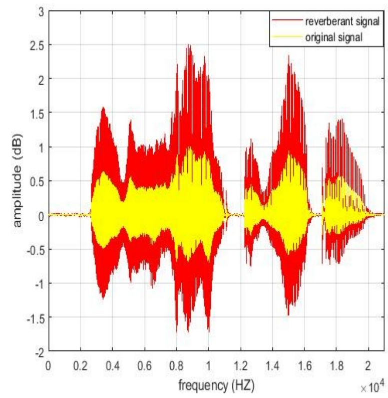

Fig. 5: Reverberation effect on female speech.

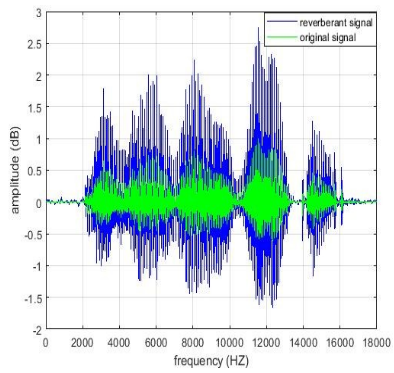

Fig. 6: Reverberation effect on male speech.

It is seen that reverberation spread speech energy and so reduce the information of the speech signal that reaches human ears.

B. *Pitch estimation*

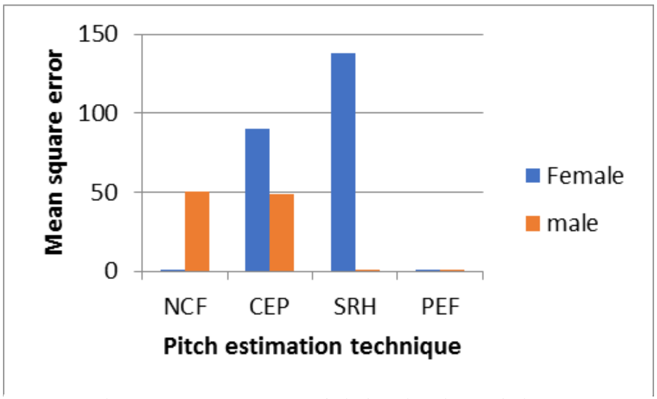

Fig.7: Mean square error of pitch estimation technique.

Table 1: Mean square error of pitch estimation technique.

Algorithm	MSE	
	Female	Male
NCF	0.0397	50.5542
CEP	89.8790	48.7143
SRH	138.2219	1.4477
PEF	0.0136	0.0197

From Fig.7 and Table 1 that represents Mean Square Error of pitch estimation techniques, it is seen that PEF algorithm is the best for pitch frequency detection for male and female speech signals in reverberation.

V. CONCLUSION

Reverberation affects the structure of audio signal; hence affects its pitch frequency. With reverberation effect all pitch frequency estimation algorithms have been investigated. From theobtained results, it is clear that sensitivity of the PEF method is the least with reverberation.In addition, the error in pitch frequency estimation for females is larger than that with males. So, some sort of compensation for the pitch frequency error is needed for applications such as speaker identification.

References

[1] Arkadiy Prodeus, "Late Reverberation Reduction and Blind Reverberation Time Measurement for Automatic Speech Recognition", IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 634 – 639, 2017.

[2] Fons Adriaensen, "Design of a Convolution Engine Opti- Missed for Reverb", In 4th International Linux Audio Conference, Karlsruhe, Germany, pp. 49-53, 27-30 April 2006.

[3] Vikramjit Mitra ,Andreas Tsiartas andElizabeth Shriberg , "Noise and Reverberation Effects on Depression Detection from Speech ", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5795 – 5799, 2016.

[4] Mohammed Kamal Khwaja, Sunil Sivadas and P. Arulmozhivarman1. "Pitch Tracking in Reverberant Environments", IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 192 – 196, 2015.

[5] Kenichiro Miwa and Masashi Unoki, "Study on Method for Estimating F0 of Steady Complex Tone in Noisy Reverberant Environments", Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 456 – 459, 2013.

[6] Oldrich Ondrácek and Miroslav Hagara, "Reverberator's Late Rreflections Parameters Calculation", International Conference Radioelektronika (RADIOELEKTRONIKA), pp. 1 – 4, 2018.

[7] 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, pp. 1 - 5 August 25-29, 2008.

[8] K Pratibha andH M Chandrashekar, " Estimation and Tracking of Pitch for Noisy Speech Signals using EMD Based Autocorrelation Function Algorithm", 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), pp. 2071 – 2075, 2017.

[9] Karishma Kolhatkar, Mahesh Kolte and Jyoti Lele, "Implementation of Pitch Detection Algorithms for Pathological Voices", International Conference on Inventive Computation Technologies (ICICT), pp. 1 - 5 2016.

[10] Thomas Drugman and Abeer Alwan "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics" Interspeech, pp. 1973-1976, 2011.

[11] Sira Gonzalez and Mike Brookes "A PITCH ESTIMATION FILTER ROBUST TO HIGH LEVELS OF NOISE (PEFAC) "19th European Signal Processing Conference, pp. 451 – 455, 2011.