International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Development of Android Application for speech Quality Measurement

Laishram Rahul[1], Vinoth Kumar[2]

Veltech University, Department of Information Technology, Chennai-600062, India
[1]*laishramrahulib@gmail.com*
[2]*vinothkumar.r@veltechuniv.edu.in*

Abstract: ***Speech is a vocalized form of human communication. Clarity of the voice is a major dependent factor for a good quality speech. It is necessary to have the means for measuring the clarity of the voice and train to improve it. In this paper, we describe the development of a system for measuring the speech quality through clarity of voice production. An Android system is built for the purpose where the user gives speech input and the resulted output is the measurement of the speech quality. The system also differentiates human speech to other sounds. If a non-speech sound is given as an input to the system, the user will be prompted to re-record again.***

Keywords: Epoch location, epoch strength, zero band filtering, voice quality, Android**.**

1. Introduction

Speech is the ability to express thoughts and feelings by articulated sounds. Normally it is generated with pulmonary pressure provided by the lungs that generates sound by phonation in the glottis inside the larynx that is then modified by the vocal tract. In this process the major source of the speech production is due to the vibration of vocal folds i.e. the opening and closing of glottis. Considering the time varying nature of the vocal tract characteristics and the voiced excitations, estimating the epoch location accurately is quite a challenging task. Various characteristics of speech can be measured through analysis of signal produced by glottal vibration [1]. This analysis is used to find the fundamental frequency (F0) of a speech signal [4]. The F0 parameter is used in applications such as prosody modification [5] and speech synthesis [6]. This study can also be used for segmenting foreground speech regions from rest of the background regions in noisy environments collected naturally [7].

This paper is an attempt to measure the quality of speech, based on the clarity of voice. This characteristic is one of the most important aspects of good quality speech. It is indeed very necessary that we could measure the quality of our speech and make an effort to improve it. We have chosen the Android platform for developing this system. Handheld devices are the most suitable platform for building such an application. Since, Android is the most widely used mobile platform and the number has been doubling every year. So, this platform is the right choice for initiating the development of such system. The user will provide speech input by speaking on the Android device and the speech will be analyzed for its quality and a result will be displayed.

2. Approach for Voice Quality Measurement

Speech is produce by the pulmonary pressure generated by the lungs that passes through various phases of the vocal tract. However, the primary mode of speech production is due to the impulse like excitation in each glottal cycle. Anchoring the speech analysis around the glottal closure instants yields significant benefits for speech analysis [1]. The instant of significant excitation of the vocal tract system, which resulted from the glottal vibration is referred to as epoch.

There are various techniques proposed for estimating the epoch location for a speech signal without using Electroglottography (EGG). These methods are as follows (i) HE-based (ii) GD-based (iii) DYPSA (iv) Zero-frequency-based. Among these available techniques, the Zero-frequency- based methods has the highest accuracy for estimating the epoch location[1,2,3]. Considering these facts, we used Zero-Band-Filtering (ZBF), a bounded input bounded output stable realization of Zero-Frequency-Filtering(ZFF). The advantages of using such a stable filter is that the filter output is bounded and has no precision related problem associated with the output for lengthy speech files, also, the method does not require to remove trend procedure that needs initial pitch estimation[2]. The procedure for extraction of ZBF signals are as follows:

1) The speech signal s[n] as shown in fig.1 is differenced to minimize any low frequency fluctuations present.

$$x[n] = s[n]-s[n-1] \quad (1)$$

2) The differenced speech signal (x[n])is than passed through zero band filter two times as show below.

$$y_1[n] = \sum_{k=0}^{\infty}(k+1)r^k\, x[n-k] \quad (2)$$

$$y_2[n] = \sum_{k=0}^{\infty}(k+1)r^k\, y_1[n-k] \quad (3)$$

Where, r is the pole location and in this paper we used r=0.99.

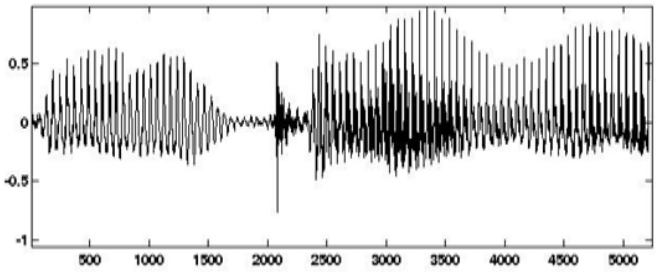
3) The zero band filtered output consist of low frequency fluctuations and to eliminate it, 4th order high-pass butterworth filter is used with cutoff frequency of 100Hz which produced signal $y_3[n]$ as shown in fig.2.

4) The high pass filtered output ($y_3[n]$) is sinusoidal in nature and its positive zero crossings gives epoch location.

5) To find out the strength at epoch locations, slope around the [9] epoch locations is taken which gives approximate strength of excitation (SoE) around the epoch location and it is given by

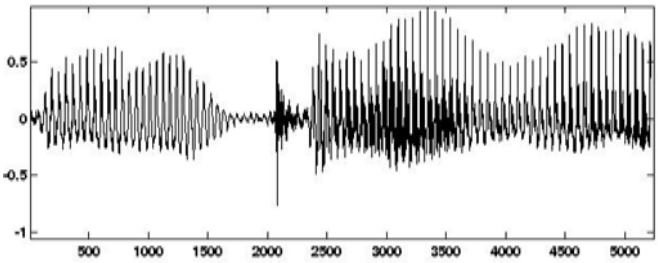
$$SoE[k]= y_3[k+1]-y_3[k-1] \quad (4)$$

6) The resulted epoch strength (fig.3) is used to measure the quality of the speech.



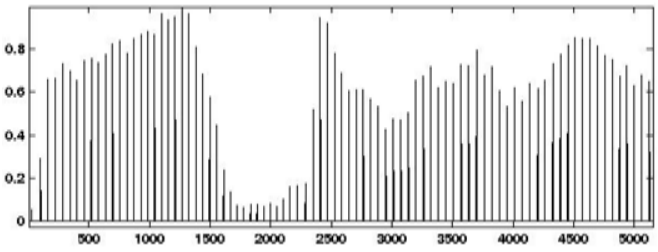
Time(10^{-4})

Figure 1: A segment of peech signal (s[n])



Time(10^{-4})

Figure 2: Zero band filtered signal ($[y_3[n]$) of s[n]



Time(10^{-4})

Figure 3: Strength of epoch around the detected epoch location

In order to validate that the user input is a voiced speech, the zero band filtered signal resulted in the step.3 is passed through the autocorrelation function to identify the maximum peak around the pitch period. The peak value is higher near the voiced speech and for rest of the sounds the peak value is low.

3. Development of Android Platform

In this system we emanate a client server model. On the client side a user interface in Android is designed where the user can input speech signal and perform speech quality measurement analysis. The speech is recorded using Adaptive Multi-Rate audio codec which is an audio compression format optimized for speech coding. AMR speech codec consists of a multi-rate narrow speech codec that encodes narrowband (200–3400 Hz) signals at variable bit rates ranging from 4.75 to 12.2 kbit/s with toll quality speech starting at 7.4 kbit/s. The file format for the recorded file is in 3GPP. The recorded audio file is sent to the server using HTTP protocol. The file handling on the server side is carried out using php script. Once the audio file successfully received on the server, the php script calls the external script for speech quality measurement which is written in C language for faster execution. If the user gives a non-speech input such as a silence or a noise, the user will be asked to re-record. When a speech input is given, the calculated result for voice quality measurement is sent back to the Android device as JSON object which is further processed to display in a graphical format. The results are displayed as very poor, poor, acceptable, good or excellent in a graphical format based on the voice quality.

4. Results and Analysis

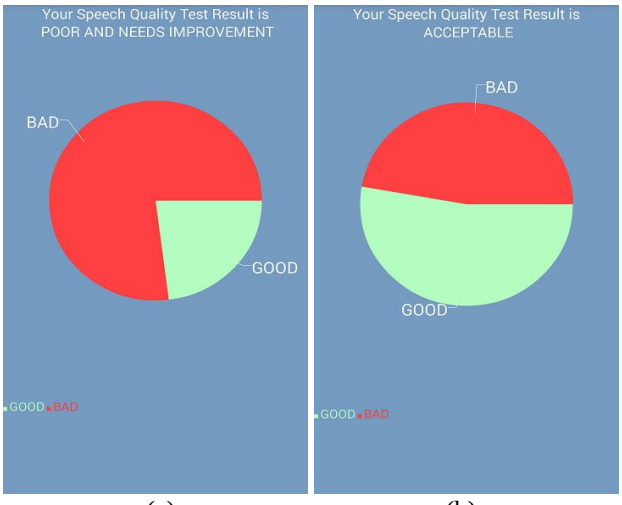
Based on the value of the mean epoch strength of a speech signal, various categorization of voice quality measurement has been made as shown in the table.1 below.

Table 1: Categorizing different speech quality based on mean epoch strength.

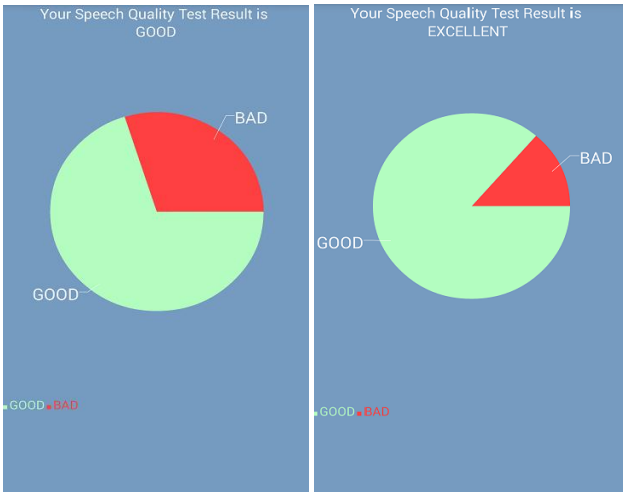
Mean Epoch Strength	*Voice Quality*
Mean epoch strength <=0.1	Very poor
0.1<Mean epoch strength <= 0.2	Poor and need improvement
0.2<Mean epoch strength <=0.3	Acceptable
0.3<Mean epoch strength <=0.4	Good
Mean epoch strength > 0.4	Excellent

In fig.4 (a), (b), (c) and (d) results for poor, acceptable, good and excellent voice quality are shown respectively.

In case of a non-speech input the user will be prompted to re-record again. The differentiation of voiced speech to other sound is determined by setting a threshold value of the normalized peak strength of the ZBF signal. If the normalized peak strength is less than 0.6, it is categorized as a non-speech input else the process continues for voice quality measurement.



(a) (b)



(c) (d)

Figure 4: Results of speech quality on Android Interface (a) Poor speech quality (b) Acceptable speech quality (c) Good speech quality (d) Excellent speech quality.

Various testing results for validation of speech to non-speech are shown in table 2. In the table, type signifies that the given input is a speech or a non-speech, the value is the normalized peak strength of the ZBF signal, the result shows if the input sound is correctly classified.

5. Conclusion and Future Works

An Android system for speech quality measurement is built successfully and tested in various environment. In the future measurement of speech quality for different modes of speech such as conversation or lecture like public speaking can be added. A better method for setting the threshold for identifying speech and non-speech and speech quality has to be implemented. The classification of speech quality in only good and bad regions need improvement.

Table 2: Validation testing for speech with the rest of sounds

Type	*Value*	*Result*
Non-speech input	0.4746	True
Non-speech input	0.5616	True
Non-speech input	0.5161	True
Non-speech input	0.4636	True
Non-speech input	0.4498	True
Non-speech input	0.5028	True
Non-speech input	0.4970	True
Non-speech input	0.4724	True
Non-speech input	0.4915	True
Non-speech input	0.5324	True
Speech input	0.6435	True
Speech input	0.6427	True
Speech input	0.6240	True
Speech input	0.7360	True
Speech input	0.5971	False
Speech input	0.6195	True
Speech input	0.7379	True
Speech input	0.6385	True
Speech input	0.6374	True
Speech input	0.6452	True

References

[1] B Yegnanarayanaand SuryakanthV Gangashetty,"Epoch-based analysis of speech signals",Sadhana, Vol. 36, Part 5, October 2011.
[2] K. T. Deepak and S. R. M. Prasanna, "Epoch Extraction Using Zero Band Filtering from Speech Signal",Springer Science, December 2014.
[3] K. Sri Rama Murty and B. Yegnanarayana, "Epoch Extraction From Speech Signals", IEEE transaction on audio,speech and language processing, Vol. 16, No. 8, November 2008.
[4] B. Yegnanarayana, S. R. M. Prasanna, and S. Guruprasad, "Study of robustness of zero frequency resonator method for extraction of fundamental frequency", ICASSP, 2011.
[5] D. Govind, S. R. M. Prasanna, and B. Yegnanarayana,"Neutral to target emotion conversion using source and suprasegmental information", Interspeech, 2011.
[6] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "Simultaneous modelling of spectrum, pitch and duration in HMM-based speech synthesis", Eurospeech, 1999.
[7] K. T. Deepak, B. D. Sarma, and S. R. M. Prasanna, "Foreground speech segmentation using zero frequency filtered signal", Interspeech 2012.
[8] http://developer.android.com/index.html
[9] K. Sri Rama Murty, B. Yegnanarayana and M. Anand Joseph "Characterization of Glottal Activity From Speech Signals",IEEE Signal processing letters, Vol. 16, No. 6, June 2009.
[10] D. Govind, S. R. Mahadeva Prasanna and Ramesh K,"Improved Method for Epoch Extraction in High Pass Filtered Speech",Annual IEEE India Conference,2013.