SPEECH PITCH DETECTION

Chon Lam LIO, Runzhu CHEN
University of Rochester

ABSTRACT – Pitch, as well as the fundamental frequency, has been proved as a critical parameter of all areas of speech research. People make use of the pitch to do a lot of synthesis work, such as automatic music transcription, tone changing and speech recognition. Therefore, pitch detection is important in the audio signal processing. Because of the importance of pitch, many algorithms have been proposed to extract the accurate pitch from a period of speech; especially from the noisy ones. One of a traditional method of pitch detection is AMDF[1], however, no matter the traditional AMDF or some other AMDF-based algorithms basically have two errors, which are the "falling tendency" and the "double/half pitch". In this paper, we propose a new algorithm to eliminate these two errors, and it gives a nice required calculation time as well, which allows us to implement it in the real-time analysis.

KEY WORDS
Pitch Detection, Average Magnitude Difference Function (AMDF), High Resolution AMDF (HRAMDF), Circular AMDF (CAMDF), Additive White Gaussian Noise (AWGN), Autocorrelation

1. Introduction

Pitch detection has been proved as a crucial task in speech research. Pitch, defined as a fundamental frequency in a periodic signal, is caused by the vibrato of human vocal cords. However, human voiced is not exactly periodic which is more like a quasi-periodic may introduce some problems in the process of pitch detection.

A plenty of algorithms has been developed to extract the pitch. Basically, there are three domains that allow us to do the pitch detection, which are the time domain, frequency domain and the cepstrum domain. They all try to find the important parameter, the fundamental frequency for pitch detection. In this project, we will mainly focus on the time domain since it can be easily applied in the real-time applications. There are some basic pitch detection algorithms (PDA) like autocorrelation function (ACF)[2], average magnitude difference function (AMDF) and some AMDF-based variations like circular-AMDF (CAMDF)[3], high resolution-AMDF(HRAMDF)[4] which can be used in the real-time applications for their low computation complexity. We implement the original AMDF, its improving version HRAMDF and CAMDF to detect the pitch of a series of vowel sounds aim to compare their pitch detection accuracy, time efficiency and the noisy immunity. The advantage of AMDF is the low computation complexity since it turns the multiplications in ACF to subtractions, which makes AMDF more

time efficiency for the implementation on real-time analysis. However, AMDF contains two errors, the "falling tendency" and the "double/half pitch". In order to eliminate these two errors, HRAMDF and CAMDF were proposed. HRAMDF successfully eliminate the "falling tendency", however it increases the "double pitch" error. CAMDF can also eliminate the "falling tendency" completely, but it also cannot completely eliminate the "double pitch" error, although it has better performance than HRAMDF, it decreases the magnitude difference between the real pitch and the double pitch. In order to truly avoid these two errors, we proposed a new algorithm, which is based on the CAMDF. Furthermore, we test the noise immunity for each algorithm by adding different level of additive white Gaussian noise (AWGN) on the input signals. Finally, we compare these four proposed algorithms to observe the results.

1.1 Falling Tendency

Falling tendency is the phenomenon where the self-correlation magnitude decrease as the delay increases within each frame. The falling tendency aggravates the double pitch effect (introduced in 1.2), which largely reduces pitch detection accuracy as well as robustness w.r.t. noise. Falling tendency is mainly caused by uneven numbers of samples adopted in self-correlation computations in AMDF. The problem is restrained by AMDF modifications like HRAMDF or CAMDF. Our proposed algorithm also takes falling tendency into account and supresses the falling tendency into an acceptable amount.

1.2 Double Pitch

Double pitch effect is the local minimum of self-correlation appears at integer multiples of the true pitch. Double pitch may be detected as true pitch if simple “minimum finding“ pitch detection algorithm is used which may significantly lower pitch detection accuracy. Double pitch effect is sensitive to noise. As noise level increases, double pitch effect appears more frequently. The hinge of solving double pitch problem is making it more robust to noise. Our proposed algorithm adopts a mean filter to each three successive frames to conquer double pitch problem under high level of AWGN. The improvements are quite out-standing as demonstrated in chapter 3.

2. Reviews of Algorithms

2.1 ACF

ACF is one of the most basic pitch detection methods that have been prevalently used in a wide range of fields. This method aims to detect the pitch by find the value when the largest autocorrelation value occurs in a certain range. Since that the autocorrelation of a periodic signal is also a periodic signal with the period,

the period of the autocorrelation function reflects the period of input signal. In this way, we can use ACF to detect the pitch of the signal. The autocorrelation function is defined as:

$$R(\tau) = \frac{1}{N}\sum_{n=0}^{N-1-\tau} x(n)x(n+\tau)$$

$$0 \le \tau \le N-1$$

where $x(n)$ is the signal, N is the length of $x(n)$, N is the total number of points involved in the calculation.

2.2AMDF

The AMDF is actually another kind of the autocorrelation function. In AMDF, it takes the absolute value of the difference between the original signal and the delay signal instead of the product of them to decrease the computation complexity which make AMDF more suitable for the real-time applications.

The difference function of Average Magnitude Difference Function (AMDF) is defined as:

$$D(\tau) = \frac{1}{N-1-\tau}\sum_{n=0}^{N-1-\tau} |x(n) - x(n+\tau)|$$

$$0 \le \tau \le N-1$$

where *x(n)* is the processed speech signal with the length N , and τ is the lag number, which ranges from *0* to *N-1*. $\frac{1}{N-1-\tau}$ is the weighting efficient used to normalize the function.

The pitch in this function is defined as:

$$T_P = \arg\underset{\tau}{MIN}\big(D(\tau)\big)$$

where τ ranges from τ_{min} to τ_{max}.

The pitch T_P equals to value of τ which make $D(\tau)$ minimum.

From the formula, we can see that as the time lag τ is getting larger, less data is used to calculate *D*, since the speech signal outside the window is *0*. Thus the "falling tendency" error occurs at the later part of the signal, in other words, at higher lags. Moreover, the AMDF algorithm is sensitive to noise and intensity. Therefore, the "double/half pitch" error occurs in noisy condition.

2.2HRAMDF

HRAMDF was proposed in[ ] which eliminates the "falling tendency" successfully. The HRAMDF is defined as:

$$D(\tau) = \sum_{n=(N/2-\tau)/2+1}^{(N/2-\tau)/2+N/2} |x(n) - x(n+\tau)|$$

The HRAMDF involves up to three frames to calculate *D*. The changes of summation range make the time lags better averaged by which the “falling tendency” is eliminated. However, due

to the repeated additions on the same period of signal, this algorithm emphasizes the pitch multiples, which leads to the increased "double pitch" error.

2.3CAMDF

Another algorithm called CAMDF was proposed to prevent the “falling tendency” and “double pitch” error. CAMDF is defined as:

$$D(\tau) = \sum_{n=0}^{N-1} |x(mod(n+\tau, N)) - x(n)|$$

From the formula, we can see that this algorithm also averages all the time lags using modulo operation, so that it can eliminate the "falling tendency" problem. Moreover, this function is symmetric around □ = $N/2$. It has better performance than HRAMDF, although the "double pitch" error still exists occasionally, the magnitude difference between the real pitch and the double pitch is reduced.

2.4GCAMDF

Our proposed algorithm is created based on CAMDF. Since the "double pitch" error occurs in noisy condition, therefore, we add the AWGN to the input signal to create this noisy condition. Basically, the idea of the algorithm is that we apply a mean filter to the CAMDF. This will basically eliminate the "double pitch" error. Furthermore, the proposed algorithm uses three samples, which are the current sample, the previous sample and the next sample to calculate D. The result is that we can also successfully eliminate the "falling tendency". The proposed method could be formulated as equation (6)

$$D(\tau) = \frac{\sum_{n=0}^{N-1} |x_K(mod(n+\tau,N)) - x(n)|}{mean\left(\sum_{k=K-2}^{k=K} G(k) * x_k\right)} \quad (6).$$

Where $G(k)$ represents 1-D Gaussian filter function. The mean filter is applied on three successive frames starting from the current frame K to $K-2$ frame. This innovative approach could largely offsets the negative influence introduced by AWGN, hence improving the pitch detection robustness under high level of noise.

3. Experimental Results and Comparisons

3.1Assumption

The algorithms are based on the Additive White Gaussian Noise model where input signal is contaminated by various levels of AWGN.

3.2Database

The experiments are implemented based on the database *VChart* provided by *"Vowel Chart with Sound Files"*[5], which contains 29 sound files for different vowels. Speech data in this database has the sampling frequency of 16kHz and 32-bit resolution. Through applying the four

pitch detection algorithms on the 29 sound files, we pick one that can significantly emphasize our concerned problems

3.3Comparisons

Figures 1 to 4 show the AMDF, HRAMDF, CAMDF and our proposed algorithm GCAMDF of the noisy input signal. From Figure 1, we can see that the "falling tendency" occurs at the later part of the signal, which is what we expected. Also, the "double pitch" error occurs since the input signal is noisy. The AMDF algorithm is sensitive to noise, so that the algorithm cannot correctly detect the true pitch. Figure 2 shows the HRAMDF algorithm for the same input signal. The "falling tendency" is eliminated completely. However, the "double pitch" error still exists due to the noisy environment. Figure 3 shows the CAMDF algorithm for the same input signal. We can see that this function is symmetric as what we expected. The "falling tendency" is completely eliminated as well. From the figure, we can see that it is still a little bit confusing where the true pitch is. Although it has a better performance than HRAMDF, the "double pitch" error still occurs occasionally. Figure 4 shows our proposed algorithm of the same input signal. We can see that GCAMDF successfully avoid the "falling tendency" and the "double pitch" errors pretty well. We can easily find where the true pitch is by looking at the sample where the minimum value takes place. Moreover, the result also means that the noise immunity of the GCAMDF is the best among these algorithms. Here we also compare the noise immunity of each algorithm, which is shown in Figure 5.

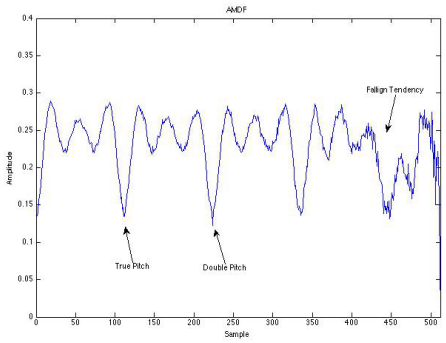

Figure 1 AMDF function of the signal

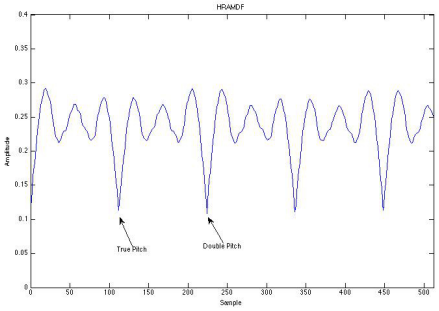

Figure 2 HRAMDF function of the signal

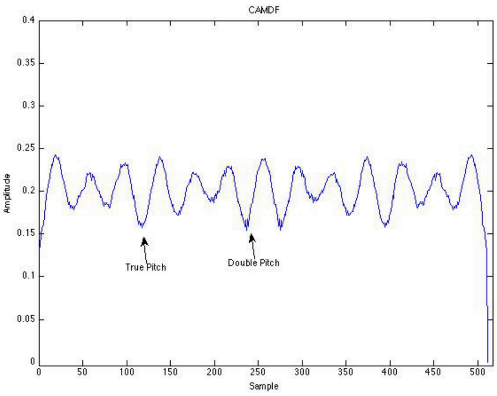

Figure 3 CAMDF function of the signal

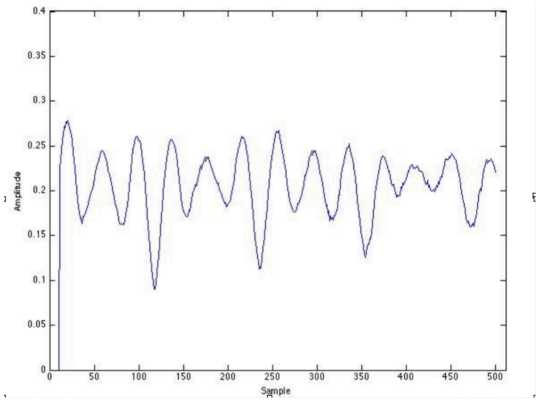

Figure 4 GCAMDF function of the signal

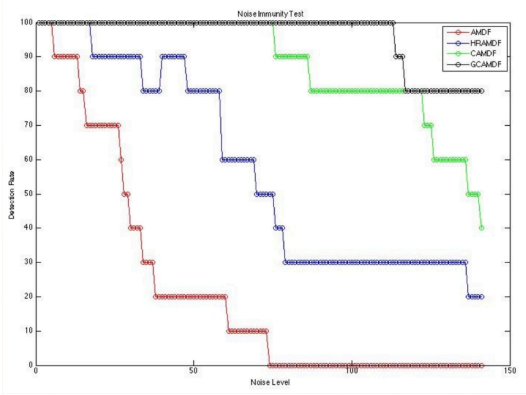

Figure 5 The noise immunity for all algorithms

The red line represents the AMDF algorithm, blue as HRAMDF, green CAMDF and black GCAMDF. For the same noise we added, we can see from the figure that our proposed algorithm has the best noise immunity. Basically, the order from worst to best is AMDF, HRAMDF, CAMDF, GCAMDF. Finally, if we want to implement the algorithm to do the real-time processing, the running time is also an important parameter that we need to consider. Here we also compare the running time of each algorithm, which is shown in Table 1.

Algorithm	Running Time (s)
AMDF	0.0029
HRAMDF	0.1624
CAMDF	0.0178
GCAMDF	0.0358

Table 1 The comparison of the running time of the algorithms

From Table 1, we can notice that the running time of the AMDF is the lowest one. It is reasonable because the AMDF algorithm is the easiest one. Notice that the running time of our proposed algorithm is about 36ms, comparing to the one of the CAMDF, which is about 18ms, GCAMDF needs twice the running time of the CAMDF to do the analysis. Although it needs this required time, it is also acceptable since 36ms is not so terrible if we implement it in the real-time processing, at least the running time of the HRAMDF is about 160ms. Anyway, there is always a trade-off between accuracy and time consumption. We think that it is worthy to get a more accurate result. It is better than getting a wrong result all the time.

4. Conclusion

In this paper, we firstly review the existing AMDF-based algorithms and get the problems that we need to improve, which are the "falling tendency" and the "double pitch" errors. Our proposed algorithm, GCAMDF successfully eliminates these two errors and gives us a pretty good running time. We also test the noise

immunity for each algorithm, using the same noise, which is the Additive White Gaussian Noise, throughout the experiment. From the noise immunity result, it turns out that our GCAMDF has the best noise immunity, so that it can eliminate the "double pitch" error, which is introduced by the noisy environment. Table 2 summarizes the problems that we concern.

	Falling Tendency	**Double Pitch Error**	**Time Efficiency**
AMDF	YES	YES	HIGH
HRAMDF	NO	YES	LOW
CAMDF	NO	YES	MEDIUM
GCAMDF	NO	NO	MEDIUM

Table 2 The summary of the problems for each algorithm

The above table shows the advantages and the disadvantages of each algorithm. We can see that GCAMDF does best among these algorithms. In general, our proposed algorithm successfully avoids the two existing problems and can be implemented in the real-time analysis, which is more practical. In the future, we want to increase the time efficiency of the GCAMDF so that it can use less time for the calculation.

References

[1]AMDF Ross M. J. et al, "Average Magnitude Difference Function Pitch Extractor", IEEE Trans. Speech and Audio Proc., vol. 22, pp 353-362, 1974.

[2]ACF Mei, X. D., Pan, J. and Sun, S-H., "Efficient algorithms for speech pitch estimation," Intelligent Multimedia, Video and Speech Processing, 2001, pp. 421 -424, 2001.

[3]CAMDF ZhangW.etal,"PitchestimationbasedoncircularAMDF", Proc. ofIEEE ICASSP'02, pp I 341-344,2002

[4] HRAMDF Gu L. and Liu R., “The government standard linear predictive coding algorithm”, Speech Technology, pp 40-49,1982.

[5] database“Vowel Chart with Sound Files database”, http://www.linguistics.ucla.edu/people/hayes/103/Charts/VChart/