A New Pitch Estimation Method Based on AMDF

Huan Zhao and Wenjie Gan
School of Information Science and Engineering, Hunan University, ChangSha, China
Email: hzhao@hnu.edu.cn, wjgan@163.com

Abstract—In this paper, a new modified average magnitude difference function (MAMDF) is proposed which is robust for noise-corrupt speech pitch estimation. The traditional technology in pitch estimation can easily give rise to the problem of detecting error pitch period. And their estimation performance behaves badly with the occurrence of background noise. In the process of calculation on speech samples, MAMDF presented in this paper has the property of strengthening the characteristic of pitch period and reducing the influence of background noise. And therefore, MAMDF can not only decrease the disadvantage brought by the decreasing trend of pitch period but also overcome the error caused by severe variation between neighboring samples. The experiment which is implemented in CSTR database shows that MAMDF is greatly superior to AMDF and CAMDF both in clean speech environment and noisy speech environment, representing prominent precision and robustness in pitch estimation.

Index Terms—Average Magnitude Difference Function (AMDF), Circular Average Magnitude Difference Function (CAMDF), Modified Average Magnitude Difference (MAMDF), Pitch Estimation

I. INTRODUCTION

Pitch period is an important parameter in speech coding, speech synthesis, speech segmentation, speech recognition, speaker recognition and other fields in speech signal process [1-9]. The precision and robustness in pitch estimation plays a great role in the performance in speech signal analysis and process. Pitch period reflects the periodic characteristic of vocal cords vibration when voiced speech is issued. However, the voiced speech is not absolutely periodic in the whole stage whose period varies along with the speech signal. Existing traditional classic pitch estimation methods have been developed such as average magnitude difference function AMDF [10], autocorrelation function ACF, cepstrum pitch estimation CEP [11] and other improved methods [12-26].

Due to low computation complexity and fine estimation performance, AMDF can be realized without much difficulty and it is often applied in real-time environment such as speech coder lpc10e [12] and so on. Because of nonstationary of speech signal, AMDF is generally implemented in short time process form. Then it can easily cause error pitch period detection for the sample number’s decrease with the offset of speech signal increasing which renders descent of peak value along with the function value in AMDF. In view of the situation, short AMDF is redefined and it adopts two frames of sample data to fix the number of difference value. In the meanwhile, it decreases the tendency of peak value in the function with the offset of speech signal increasing [12]. However, the redefined function is not able to reduce pitch estimation error obviously. LVAMDF is proposed to agrandize the correct rate of pitch estimation [13]. In LVAMDF, calculation range of sample amplitude difference is widened with the increase of speech signal offset. This kind of process can focus the length of samples taken part in computation to around one or two pitch periods. Also the effective length can change along with the variety of different people. In the meanwhile, normalization process by employing a denominator item is carried out to avoid the effect of window length. Though LVAMDF can improve the performance, it is subjected to the influence of speech frame heading and the case of error estimation in pitch period is serious. CAMDF is a pitch estimation function which processes the sample data in the form of cyclic shift to enhance the precision in pitch estimation [14]. Besides, it adopts fixed number of difference value in order to meliorate the situation of function peak value decrease with the reducing number of difference value. But CAMDF still takes on serious error estimation in pitch period when it confronts severe fluctuation in speech waveform and has a poor performance in noisy environment.

By analyzing the flaw of traditional classic algorithms in pitch estimation, a new approach based on AMDF which is named MAMDF is proposed to reduce and even avoid these defects to some extent. It adopts fixed range of calculation on sample data and has an advantage of highlighting the period characteristic in voiced speech signal. This advantage can not only drop the influence of severe fluctuation in speech waveform when doing pitch determination but also lower the error rate caused by noise mixed into original clean speech even in very low SNR environment, showing a satisfying estimation result. In our experiment, CSTR database is employed and pitch estimation is carried out both in clean speech and a variety of noisy environment with different SNR produced by adding white noise. Experiments are carried out with male speech and female speech separately and in each condition we will observe the performance of them.
clearly. Besides experimental results on the two kinds of speech, we conclude the total outcome on both female and male speech. A large number of experimental results demonstrate that MAMDF is greatly superior to AMDF and CAMDF, performing wonderful estimation precision and robustness.

This paper is organized as follows. A brief review of AMDF and CAMDF is described in section 2. In section 3 MAMDF in pitch estimation is proposed as a novel algorithm. Firstly the formula is presented and then a comparison between MAMDF and AMDF where CAMDF is illustrated both in clean and noisy environment. The experiment is implemented in section 4, where the process and results will be explained in detail. Finally, in section 5, the conclusion is given.

II. REVIEW OF AMDF AND CAMDF

A. AMDF

Average magnitude difference function which has minima at integer multiples of pitch period, has been commonly employed for pitch estimation and its original definition is described as:

$$D(t) = \frac{1}{L} \sum_{j=0}^{L-1} |S_j - S_{j-t}|, \text{ } t = 0, 1, \ldots, t_{\text{max}}.$$

(1)

where $S_j$ is the speech sample sequence and $S_{j-t}$ is the shifted sample which has the lag time $t$. $t_{\text{max}}$ is the longest shift time. The range of $t$ is between 0 and $t_{\text{max}}$. L is the length of samples calculated in this formula.

AMDF is generally implemented in short time process form because of the nonstationarity of speech signal and the commonly used formula is defined as:

$$D(k) = \frac{1}{n-k-1} \sum_{j=0}^{n-k-1} |x(j+k) - x(j)|.$$

(2)

where $x(j)$ is speech sample sequence multiplied by a rectangular window of length $n$. $k$ is the lag number whose value is between 0 and $n$. AMDF value is obtained by adding the signal difference of certain samples with certain delays.

As is seen in (2), sample number taken part in calculation is decreasing with the delay number increase so that (2) gives rise to a falling tendency with the increase of $k$. The true pitch period will not in the valley position of this function. Therefore, inaccurate pitch estimation will be obtained. To conquer the defect we define the format of AMDF in another form which has a better performance as follows in this paper:

$$D(k) = \frac{1}{n-2k} \sum_{j=0}^{n-2k} |x(j) - x(j-k)|.$$

(3)

where we use one more frame samples which avoid the number of difference value decreasing. And therefore it has a better performance than that in form described in (2).

In (3) varied from original AMDF, we can find that the number of difference value involved in computation is fixed. With the increase of delay number the value of this function will not take on a falling tendency brought by decreasing addition number. The pitch period will be in the valley position of this function. The difference of this form from original AMDF lies in the sample number adopted in calculation of adding corresponding difference value.

B. CAMDF

To avoid the falling tendency of AMDF and improve the performance of AMDF in pitch estimation, CAMDF as a different function was proposed and defined as:

$$D(k) = \frac{1}{n} \sum_{j=0}^{n-k} |x(\text{mod}(j+k,n)) - x(j)|.$$

(4)

where $x(j)$ is speech sample sequence and $\text{mod}(j+k, n)$ is a kind of modulo operation. One implicated property in (4) is that $D(k)$ is symmetric around $k=n/2$, satisfying the equation which is expressed by $D(k)=D(n-k)$. Therefore, to get enough lag number of $k$, we compute two frame samples when taking advantage of formula (4) in this paper.

The values of the valley points are in the position of pitch period and multiple pitch period. They have increasing tendency as the delay number increase in CAMDF which can demonstrate that the CAMDF can be used to estimate pitch period as the AMDF does. Different from the AMDF mentioned above, all speech samples in current frame are used and only used once in CAMDF to calculate the difference value for every lag. The total number of items summed together to obtain $D(k)$ for different $k$ is identical to each other. These different changes can cause the result that the function curve develops along a horizontal line. There is another characteristic that the peak values are almost kept in the same level.

III. PROPOSED APPROACH MAMDF

A. Introduction of MAMDF

As speech signal is produced by the human organs, the waveform in the form of samples will not be a certain function related to time parameter. It is of instability and uncertainty in the magnitude and frequency even when the signal is issued by the same person and the same word or sentence. Seen from a total voiced speech, it is not absolute periodic in its shape because of the nature of speech signal. In the rough appearance of speech signal, approximate contour of pitch period is still visualized.

The methods of AMDF and CAMDF are based on this kind of thought: in periodic signal the samples of original shape and these samples of delayed signal have the same value in corresponding relative sequences. With different delay, we can get different function value produced by the addition of difference value. The function value will be zero in periodic position and multiple pitch period. In the periodic position of voiced speech which is not absolutely periodic, the function will has a valley value which has the minimum during certain pitch range.

By analyzing the algorithms of AMDF and its improvement we find that their magnitude difference
function is carried out directly between certain samples. As the speech signal is not absolute periodic during frames of signal sample and fluctuation usually occurs in waveform, the determination of pitch period will be not agree with true pitch. In allusion to these features mentioned above, we propose a thought that carry out calculation of difference value between corresponding samples indirectly instead of directly. The thought can avoid the disadvantage brought by the nature of speech signal. To ameliorate the situation produced by AMDF and CAMDF we propose a new method modified average magnitude difference function abbreviated as MAMDF which is able to improve the precision and robustness in pitch estimation.

MAMDF is defined as:

\[
D(k) = \frac{1}{2a+1} \sum_{j=0}^{n-1} \sum_{i=j}^{j+q} |x(i) - \sum_{i=j-a}^{j+a} x(i)|
\]

\[
x1 = \text{mod}(j + k, n) - a \quad \text{(5)}
\]

\[
x2 = \text{mod}(j + k, n) + a
\]

where \(x(i)\) denotes speech sample sequence, \(n\) is the length of rectangle window and \(\text{mod}(j + k, n)\) represents the modulo operation. \(k\) stands for the lag number of speech signal while \(\alpha\) which is set to be 5 in this paper is relevant to sample frequency. This function implicates a property that it is symmetric around \(k=n/2\). Therefore, to estimate the pitch period within enough samples, two frame samples are needed. From the detailed computation on speech signal samples we can find a fact that the addition on difference value is carried out indirectly between certain samples. This process can strengthen periodic characteristic of voiced speech.

The number of samples to be computed in this method is fixed which help avoiding falling tendency of valley value. This characteristic of MAMDF is convent for pitch estimation so that we can make the decision on pitch period for one time instead of complex logic. MAMDF has the feature of owning distinct valley value in the position of pitch period and multiple pitch period. These valley values have an increasing tendency.

The pitch period is usually determined by:

\[
TP = \arg \text{MIN}(D(k)). \quad (6)
\]

The final value obtained from (6) is the value of \(k\) who makes the minimum \(D(k)\). \(TP_{\text{min}}\) and \(TP_{\text{max}}\) correspond to possible minimum and maximum pitch periods whose value depends on sampling frequency and possible pitch frequency. In this paper pitch frequency considered is between 60Hz to 400Hz with sampling frequency 20KHz. Therefore, \(TP_{\text{min}}\) we set to be 50 while \(TP_{\text{max}}\), to be 350. \(TP\) indicates the estimated pitch periods.

From the analysis above in the form of MAMDF, the signal difference value of MAMDF is usually minimum at delay of pitch period and exhibits deep nulls at delays corresponding to the pitch period of voiced speech. Both female voice and male voice have a range of pitch period which can help determining the estimation range of values obtained from MAMDF. It is convenient for us to reduce the calculation in judgment on pitch estimation and improve the precision of pitch estimation in contrast to that without using possible maximum and minimum pitch period.

**B. Observation and Analysis on MAMDF**

Error rate of pitch estimation using AMDF will increase seriously when the amplitude or frequency changes rapidly in speech signal while MAMDF can help avoiding this phenomenon.

![Figure 1](image_url)

(a) a frame of clean voiced speech, (b) AMDF value, (c) MAMDF value.

In clean speech signal, the speech samples in a certain range are not completely equal to these samples generated by delay of pitch period in original speech signal, but the waveform represent a periodic shape fluctuated in an acceptable region. However the variation and uncertainty of speech pitch period samples in each given region can easily give rise to the problem that the
value of AMDF will accumulate error added by these difference value of corresponding samples in time alignment. On the other hand, MAMDF will reduce the accumulated error brought by inequality of corresponding samples when the delay is pitch period. The figures below can describe the difference of AMDF and MAMDF in clean speech signal environment. And they are of representative characteristic to explain the advantage of the method MAMDF which has a better performance because of its inner nature.

As is shown in Fig. 1(a), the true pitch period is 116 samples when sampling frequency is 20KHz. Marked by little circle in Fig. 1(b), minimum peak value between possible maximum pitch period and minimum one is 231 samples, locating wrong pitch period. In contrast, determined pitch period by MAMDF is in accordance with authentic one.

As is depicted in the figures above, we can find the fact that the frame of voiced speech has a periodic characteristic when the corresponding sample are rough measured in a relative range and the amplitude of samples whose interval is a pitch period can are extended in a certain range. But when these samples are measured in a high precision and time alignment is implemented strictly without even a time error of one sample, the periodic characteristic of voiced speech is so vague that usually error pitch estimation is obtained. The function value of AMDF has valley values when they are in the position of pitch period and multiple pitch period. But these valley values have a falling tendency as the delay grows, rending that the estimated pitch is double the true pitch period. In contrast to AMDF, the function value of MAMDF takes on valley values in the position of pitch period and its multiple pitch periods. The obvious improvement lies in that the value of MAMDF has its valley values in a rising tendency which has the minimum in the true pitch period in the given range between possible maximum pitch period and minimum pitch period samples.

In the condition of voiced speech with obvious pitch characteristic the function value of AMDF is able to detect the pitch period in a correct way with relative right result. At the same time, CAMDF can make up the shortage brought by AMDF when double pitch estimation occurs. But when speech signal samples fluctuate rapidly during neighboring position both the methods AMDF and CAMDF are invalid in pitch estimation correctly. That is because in this condition the samples which have a time interval of one period are not equal to each other and have a great difference value. When adding numbers of items of such sample difference value to generate the function value in traditional classic pitch estimation methods, error judgment will appear.

Speech signal is not as smooth as a certain continuous function and its waveform is crucial to the application of pitch estimation algorithms. Especially as neighboring samples fluctuate rapidly, error estimation of pitch period will be brought about in CAMDF. In Fig. 2(a) a frame of voiced clean speech undulate severely and the roughly contour of waveform shape is periodic. It renders that estimated pitch period is mistaken encircled by CAMDF in Fig. 2(b). Furthermore, value of CAMDF is not so smooth in its shape and the valley value is also obscure. In comparison to CAMDF, in Fig. 2(c) where MAMDF is put into use represents that the function value is very smooth and its valley value is quite distinct. Result obtained by MAMDF is proper.

![Figure 2](image-url)

From the figures presented above we can observe the typical example of a frame of speech signal which causes error when CAMDF is applied. The shape of function value of CAMDF is not as smooth as the corresponding samples in different pitch period have larger difference values than that in smooth speech signal. In comparison with function values and shape using CAMDF, the function value of MAMDF appears a more legible waveform in its shape and the valley value is so distinct that it reveals the inner periodic characteristic of voiced speech signal. One of the virtues of MAMDF is to decrease the influence of waveform appearance which causes error pitch estimation in other methods.
The difference between CAMDF and MAMDF lies in the thought of adding difference values in different forms of samples. In the addition of CAMDF, it carries out modulo operation on delayed speech signal samples directly and this kind of calculation is also influenced greatly by the feature of vibration during neighboring samples. While in the process of MAMDF, the addition is implemented on the difference values of samples in modulo operation indirectly and this process can eliminate the influence of unsmooth original speech signal samples.

The comparison and analysis mentioned above is in clean speech signal environment. It has a conclusion that the method of MAMDF has a superior performance conquering some defect which is brought about by the inner nature of speech than the methods of AMDF and CAMDF. In reality application, pitch estimation is often carried out in noisy environment where the original speech signal is corrupted by background noise. In noisy speech environment with very low SNR, the performance of AMDF and CAMDF is very poor while MAMDF can have a satisfying estimation result. That is because the performance of methods AMDF and CAMDF is depended on the original shape of speech signal. These figures below are a comparison of performance in pitch estimation by the three methods in noisy speech signal.

These comparisons mentioned in previous section reflect superior performance of MAMDF in clean speech environment. Many applications in reality are based on the background with noise where traditional pitch estimation algorithms behave badly. Nevertheless, MAMDF can acquire relatively accurate outcome in spite of noisy speech with very low SNR, representing marvelous robustness. Fig. 3(a) is a frame of voiced clean speech where we can perceive the pitch period is 140 samples and Fig. 3(b) is a frame of noisy speech which is added white noise into original clean speech. And we can get the SNR to be -5dB where pitch period in the noisy speech is made obscure. Both the shape of AMDF and CAMDF depicted in Fig. 3(c) and Fig. 3(d) is so poor in smoothness and their valley values are in the points of double the pitch period instead of in the position of pitch period. The results of pitch estimation on the noisy speech signal using methods AMDF and CAMDF are 280 samples. Superior to traditional classic algorithms, MAMDF has a virtue of being little pregnable of background noise which can be proved by truly estimated pitch period encircled in Fig. 3(e). Pitch period detected by MAMDF is 140 samples which is in accordance with correct one shown in clean speech signal environment. So one merit of MAMDF is that it can reduce the influence of background noise.

This example is representative to illustrate the features of AMDF, CAMDF and MAMDF in noisy speech signal environment. Though background noise is added to clean speech signal, one of the property of MAMDF is that it can drop the influence of extra amplitude and energy from background noise. While methods of AMDF and CAMDF are subjected to this kind of interference to change the original waveform. In the meanwhile,
MAMDF can also strengthen the periodic characteristic of speech signal. Observing from the figures in Fig. 3 we can find that the original clean speech signal has a strong periodic characteristic and the contour is so distinct. After adding white noise to the original clean speech we obtain the noisy speech signal whose SNR is -5dB. From the rough shape we still can perceive the approximate pitch period when the measurement is not implemented in high precision. But in detail the periodic feature has been destroyed as the amplitude of the original speech signal has been mixed with the amplitude of the noise. And therefore, the plot represented by the function value of AMDF and CAMDF is not smooth and influenced greatly by the noisy speech. The valley value of their function value is also too vague in pitch period and multiple pitch periods. A falling tendency of valley value is taken on which results in error pitch estimation. Calculated by MAMDF, the function value plot is smoother and the valley value in corresponding points is just in right pitch period. The tendency of minimum peaks is ascending with the increase of number in x-coordinate which can give an obvious and right pitch period value in MAMDF.

IV. EXPERIMENT AND RESULTS

A. CSTR Database and Experimental Detail

In our simulation experiment, we utilize the CSTR database of real speech signals which is publicly available provided by the Center for Speech Technology Research at University of Edinburgh, Scotland, UK. There are totally 100 long utterances consist of 50 female voice and 50 male voice in English respectively. The database is biased towards utterances which contain voiced fricatives, nasals, liquids and glides, because PDAs usually find these difficult to analyze. Database is of studio quality which is sampled at 20kHz with 16-bit resolution. CSTR provides a reference pitch and it is obtained from a simultaneously recorded laryngograph trace.

In order to implement our experiment in a noisy environment, we add white noise with different signal-to-noise ratios (SNR) to the original clean speech signal. The white noise sequence is available in the Noisex92 database and it is used to imitate white noise-corrupted speech. Altogether we acquire 5 kind of SNR as 10dB, 5dB, 0dB, -5dB and -10dB.

We fix the analysis window length to be 26.5ms at 20kHz sampling frequency. The considered pitch range both for male and female speakers is between 60-400Hz and there is no pre-processing stage to filter the speech signal.

In experiments to get readable file we first transfer the original voice files to certain format according to certain resolution of samples and sampling frequency. Then after reading voice files we calculate the function value to estimate pitch period separately. From each file we can get a part of experimental results. When all files have been processed, total data will be integrated together to obtain the performance of each method.

When calculating AMDF, we make use of two frames of speech to avoid falling tendency of $D(k)$ resulted by decreasing number of difference values. According to analysis mentioned in previous section, to get enough delay number in $k$ we also adopt two frames of signal samples in CAMDF and MAMDF and the parameter $a$ in MAMDF is set to be 5.

B. Performance Evaluation Criteria

What is used as the reference is the $f_0$ value from the laryngeal frequency contour. There is a time label responding to every $f_0$ value which can align the estimated pitch value $P_{est}$ with the reference pitch $P_{ref}$.

Criteria considered in our simulation for the performance evaluation of the proposed method is gross pitch error (GPE). An determined pitch $P_{est}$ is classified as incorrect when it falls outside 20% or -20% of the true pitch value $P_{ref}$ and it is judged to cause GPE. The percentage of GPE is used as a standard, which is calculated from the ratio of the number of frames $N_{GPE}$ yielding GPE to the total number of voiced frames $N$, namely as follows:

$$GPE(\%) = \frac{N_{GPE}}{N} \times 100.$$  

In the database CSTR, pitch period is provided in the form of frequency while we get the function value in the form of samples in our experiments implemented by AMDF, CAMDF and MAMDF. That is because we carry out the estimation in the form of waveform and judge the pitch period in the position which has minimum valley value in the given range converted by maximum frequency and minimum frequency to corresponding samples. Then we should have a comparison with the referenced pitch period when we want to judge whether a frame of speech signal cause GPE frequency from estimated pitch period. And to process the comparison we need convert our pitch period in the form of samples obtained through the function of pitch estimation methods to the form of frequency. In each piece of speech signal we can receive a number of frames which cause GPE. After all speech signal is taken part in the calculation we can get total number of $N_{GPE}$ and total number of frames in all speech signal involved in this calculation. At last, the value of performance evaluation criteria can be computed in the form of GPE (%).

C. Experimental Results and Comparison

The pitch estimation performance of the AMDF, CAMDF and the proposed MAMDF methods is investigated in the speech signal of the CSTR databases. The speech signal is consist of clean speech and noisy speech added by white noise. Results are depicted in Table I through Table III. Pitch estimation of male speech and female speech is carried out respectively in Table I and Table II. From the results we can find that the other two methods produce less accurate results. The proposed method successfully estimates the pitch with higher accuracy no matter in clean speech or noisy speech environment within even low SNR.
From the results represented in the pitch estimation on the male speech, we can see that in clean speech signal environment the method of AMDF can perform a relative fine efficacy. But in the lower SNR environment by adding white noise to original clean speech signal the value of GPE (%) is increasing rapidly. The error rate is up to 42.82 when the SNR is equivalent to -10dB, so the method of AMDF is nearly of no use. The background noise influences the performance of AMDF so greatly. Meanwhile the performance of CAMDF is almost equivalent to that in AMDF no matter in clean speech signal environment or in noisy environment. We can find that on male speech the method of CAMDF has no improvement in the performance of pitch estimation. The estimation result using MAMDF reduces the value of GPE (%) remarkably and has lower GPE value. The error rate reduced by MAMDF has been about 10 percent from 42 percent to 31 percent in noisy environment of -10dB.

The results on female speech depict a fact that the method of AMDF in voiced speech with high frequency has a bad performance because of its inner nature. The performance of AMDF is worse when it is used on female speech than that on male speech. So we can conclude that AMDF is not suitable in this kind of speech environment which has a high frequency. With the lower of SNR the performance of AMDF is so poor that the pitch estimation usually is wrong and the method is no more effective. In the worst condition, error rate gets to almost 62 percent when SNR is -10dB. Though the method of CAMDF has a better performance than that in AMDF, the results is not satisfying as its value of GPE(%) is also very high. Error rate increases rapidly by using CAMDF with the lower SNR. Here MAMDF still performs very well and it reduces the value of GPE(%) greatly in comparison to AMDF and CAMDF. Even in very low SNR environment it is still valid in pitch estimation.

Though the existing improved method can overcome some disadvantage of AMDF and has a better performance to reduce the difference of true pitch and estimated pitch, pitch error estimation is not eliminated. Besides, the existing approaches mentioned above have a disadvantage that in noisy conditions it can’t output minimum at the pitch period position which brings about error pitch estimation.

On the whole, when the results are summarized both on male speech signal and on female speech signal, the performance of MAMDF is very prominent. As we can see that in -10dB environment, the value of GPE (%) is 52.4 using AMDF and 45.6 by CAMDF while it is only 34.47 through our proposed method MAMDF. The value of GPE (%) has been greatly decreased by the application of MAMDF in many signal environments.

<table>
<thead>
<tr>
<th>Method</th>
<th>Clean</th>
<th>10dB</th>
<th>5dB</th>
<th>0dB</th>
<th>-5dB</th>
<th>-10dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMDF</td>
<td>6.16</td>
<td>7.53</td>
<td>9.36</td>
<td>14.41</td>
<td>25.13</td>
<td>42.82</td>
</tr>
<tr>
<td>CAMDF</td>
<td>7.81</td>
<td>8.97</td>
<td>10.98</td>
<td>15.53</td>
<td>25.73</td>
<td>41.54</td>
</tr>
<tr>
<td>MAMDF</td>
<td>5.94</td>
<td>7.46</td>
<td>8.83</td>
<td>11.71</td>
<td>18.69</td>
<td>31.12</td>
</tr>
</tbody>
</table>

### V. Conclusions

A new robust method has been presented to deal with the problem in pitch estimation from the noise-corrupted speech signals. Firstly we analyze the form and shortcoming of the original AMDF and its improvement CAMDF. Then we describe MAMDF in detail to analyze its virtue. It can overcome the error estimated pitch brought by vibration during neighboring samples and signal feature of not being periodic absolutely along with the speech signal waveform. During the process of pitch estimation, MAMDF has the advantage of strengthening the periodic characteristic of voiced speech and being less influenced by background noise. These characteristic can improve the phenomenon of error pitch estimation in traditional classic methods. By carrying out our experiment in CSTR database, it turns out that MAMDF can obviously decrease error rate of pitch estimation and has a superior performance in comparison to AMDF and CAMDF no matter in female and male speech signal. Even in very low SNR environment, marvelous precision and robustness is to be obtained as compared to other methods in the literature.

### Acknowledgment

This work was supported by National Science Foundation of China (Grant No. 61173106), the Key Program of Hunan Provincial Natural Science Foundation of China (Grant No.10JJ2046).

### References


Huan Zhao was born in Changsha, China, in 1967. She obtained her B.Sc. degree, M.S. degree and Ph.D. in Computer Science and Technology in Hunan University in 1989, 2004 and 2010, respectively. She served as visiting scholar at the University of California, San Diego (UCSD), USA during the period of March 2008 to September 2008. The visiting scholarship was appointed and sponsored by the China Scholarship Council (CSC).

She is a professor at the School of Information Science and Engineering, Hunan University. Her current research interests include speech information process, embedded system design and embedded speech recognition. She has published more than 40 papers and 9 books.

Prof. Zhao is a Senior Member of China Computer Federation, Governing of Hunan Computer Society, China and China Education Ministry Steering Committee Member of Computer Education on Arts.