Trend Analysis Based AMDF for Robust Pitch Detection of Speech Signals

Weihua Zhang[1], Yingying Lu[2], and Pingping Xu[3]

[1,2,3]School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing, China
E-mail: 150916907@qq.com, xhluyingying@163.com, xpp@seu.edu.cn

Received: August 15, 2021 / Revised: September 18, 2021 / Accepted: September 28, 2021

Abstract—**In this paper, we focus on improving the AMDF pitch detection algorithm (PDA) rather than designing a complete pitch detection system including many complex modification stages. As a hot classical PDA, generating half or multiple pitch errors is a usual defect of AMDF, especially in noisy conditions. Based on a deep analysis of many existing improvements of AMDF, we summarize two modified frameworks and classify the most outstanding improvements into them. Then we propose a novel and simple modified framework for AMDF to conquer the defect of AMDF. For our framework, we also present two kinds of falling trend extraction methods to obtain the proposed Trend Analysis based AMDF (TAAMDF). Finally, Experiments on the *Keele* database are conducted to evaluate our framework. Compared with some outstanding modified AMDFs and well-known ACF, modified AMDF based on our framework shows the best performance especially its robustness to different noises.**

Index Terms—**Pitch Detection Algorithm, AMDF, Falling Trend Analysis**

I. Introduction

Pitch (or fundamental frequency) plays an important role in many fields of speech signal processing such as speech coding, speech recognition, speech enhancement, etc. This fact has motivated researchers to think of how to detect the pitch from speech signals accurately and effectively. As is known to all, breakthroughs of PDAs emerged decades ago. Since then, there are many classical pitch detection algorithms (PDAs) [1]-[3] and their improvements have been proposed. In spite of this, developing accurate and reliable PDAs is still challenging. There are still many excellent works reported in recent years [4], [5], and one of the most important features we notice is that nowadays researchers devote themselves to designing a complete pitch detection system for high accuracy and noise robustness that adds many pre-processing and post-processing stages to enhance the key part of the system i.e., their PDAs. However, a lot of outstanding software for speech signal analysis still adopts classical PDAs mentioned above to design their pitch detection module. For example, the Autocorrelation Function (ACF) based pitch detection module is included in the Praat [6] software. YIN [7], an excellent pitch estimator, is also based on ACF with some additional modifications. This situation indicates that these classical PDAs are still valuable and powerful. Hence, we hold a viewpoint that it is still meaningful and worthwhile to make these classical PDAs more powerful as well as develop a complete pitch detection system.

In this paper, we pay attention to another classical PDA, namely Average Magnitude Difference Function (AMDF) [2]. AMDF is also one of the most widely used PDAs because of its low computation and high precision. However, Zhang et al. [8] pointed out that a falling trend presents as a global feature in AMDF such that some detection errors that the estimated pitch is half or multiple of the actual sometimes happened. This is not only due to a single cause but a combination of complex factors such as formant, noise, and framing setup of speech signals. Furthermore, noise is the most usual and unavoidable factor for PDA. Therefore, it is important to improve AMDF to eliminate the falling trend and enhance the robustness to noise. To this end, Zhang et al. [8] proposed Circular AMDF (CAMDF) by introducing modulo operation to redefine the calculation AMDF of speech frame at each lag. CAMDF prevents the falling trend and achieves excellent detection performance. Another state-of-the-art modification of AMDF which is worth mentioning is Extended AMDF (EAMDF) presented by Muhammud [9]. EAMDF extends the length of the speech frame to supply the loss of overlap with lag increasing. Thus, EAMDF can overcome the falling trend effectively and outperforms the classical AMDF.

Although these modified AMDFs achieve promising performance, they are not very satisfactory due to changes of either definition or speech frame and will also bring estimated errors sometimes. In this paper, we propose a novel modified framework for AMDF. Different from many existing improvements, this is a simple and distinctive framework that can overcome the shortcoming of AMDF more effectively and is considerably robust to different types of noise.

The rest of this paper is organized as follows: Section 2 reviews AMDF and its representative improvements. Section 3 describes the proposed framework for AMDF. After that, experiments on the *Keele* database are conducted for testing and verifying the proposed framework in Section 4. Finally, the paper is concluded in Section 5.

II. Review and Analysis of AMDF and Its Improvements AMDF

The conventional AMDF [2] is defined as follows:

$$D_{AMDF}(\tau) = \sum_{n=0}^{N-\tau-1} \left| x(n) - x(n+\tau) \right| \tag{1}$$

where $x(n)$ denotes a voiced speech frame multiplied by a rectangular window of length N and τ denotes the lag number. For a periodic or quasi-periodic signal with a period of T_p, its AMDF $D_{AMDF}(\tau)$ should exhibit valleys at lag nT_p, where n is an integer. Generally, we can estimate the raw pitch f_p from AMDF according to Equation (2).

$$f_p = fs / \arg\min_{\tau} (D_{AMDF}(\tau)) \tag{2}$$

where *fs* denotes sample frequency of speech signals. As less data is used to calculate AMDF at higher lags, a falling trend may present as a global feature in the AMDF curve sometimes. Thus, the valley with true pitch information may not be the lowest and the multiple pitch errors may be produced according to Equation (2). Fig. 1 (b) shows a typical example of the double pitch error of classical AMDF. In the figure, the corresponding speech signal is a clean voiced frame (see Fig. 1 (a)).

In order to overcome this shortcoming of AMDF, many outstanding modified AMDFs have been proposed. Among these modifications, CAMDF [8] and EAMDF [9] mentioned in the previous section are two representative ones. CAMDF was proposed to overcome the falling trend by a modulo operation and is defined as follows:

$$D_{CAMDF}(\tau) = \sum_{n=0}^{N-1} \left| x(\mathrm{mod}(n+\tau, N)) - x(n) \right| \tag{3}$$

where mod(n + τ, *N)* represents that n + τ modulo N. Muhammad [9] proposed EAMDF and define it as:

$$D_{EAMDF}(\tau) = \frac{1}{N - \tau} \sum_{n=-N/2}^{N+N/2-\tau} \left| x(n) - x(n+\tau) \right| \tag{4}$$

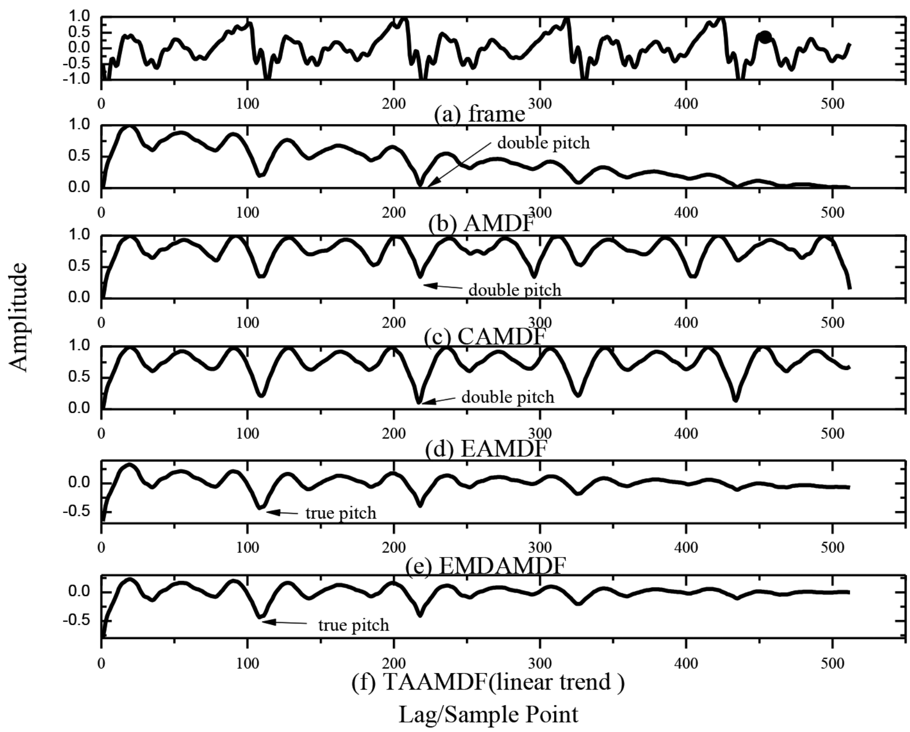

Fig. 1. Comparison between (b) AMDF, (c) CAMDF, (d) EAMDF, (e) EMDAMDF and (f) TAAMDF of (a) a typical speech frame. EMDAMDF and TAAMDF detects the true pitch, while all the other produce double pitch errors.

Actually, it should be noted that compared with the AMDF (Equation (2)), we can clearly find that CAMDF (Equation (3)) is improved by modifying the definition of AMDF whereas EAMDF (Equation (4)) promotes by means of adjusting the length of the speech frame which is used to calculate AMDF. As Fig. 1 (c) and Fig. 1 (d) depict, CAMDF and EAMDF can all achieve eliminating the falling trend. However, they still produce unexpected double pitch errors because of the adverse impact brought by the changes of speech frame or definition. We think that they represent two typical modified frameworks for improving AMDF, namely modifying the definition of AMDF and adjusting the length of the speech frame. Furthermore, our observation is that the vast majority of existing modified AMDFs can all be included in these two frameworks. For example, LVAMDF [10] adjusts the length of the speech frame and HRAMDF [11] both adjusts the frame and redefines AMDF by adding a normalized term.

III. A NOVEL FRAMEWORK FOR IMPROVING AMDF

Ideally, we want to find a framework for AMDF that eliminates the falling trend effectively and produces no estimated errors because of adjustment. In our previous work [12], we employ Empirical Mode Decomposition (EMD) [13] to address the problem and improve AMDF. More specifically, let $D_{AMDF}(\tau)$ be AMDF of a voiced speech frame. $D_{AMDF}(\tau)$ can be decomposed into a series of Intrinsic Mode Functions (IMFs) $c_i(\tau)$ and a residue $r_N(\tau)$. Based on the principle of EMD, AMDF can be reconstructed by all IMFs and the residue, which can be expressed as:

$$D_{AMDF}(\tau)=\sum_{i=1}^{N}c_i(\tau)+r_N(\tau) \tag{5}$$

where N is the number of IMFs. We find that the residue $r_N(\tau)$ represents the trend of AMDF data points namely the falling trend which is mentioned in many literatures. Therefore, we spontaneously consider to reconstruct AMDF abandoning the residue, and obtain our EMD-based AMDF (EMDAMDF) [12]:

$$D_{EMDAMDF}(\tau)=\sum_{i=1}^{N}c_i(\tau) \tag{6}$$

It is worth noting that EMDAMDF can also be written as:

$$D_{EMDAMDF}(\tau)=D_{AMDF}(\tau)-r_N(\tau) \tag{7}$$

For the speech frame in Figure 1(a), EMDAMDF can both eliminate the falling trend of AMDF and estimate the true pitch effectively as shown in Fig. 1 (e). It must be emphasized that although the two formulas above turn out the same results, their thoughts are distinct that the former is a reconstruction method and the latter is a removing one.

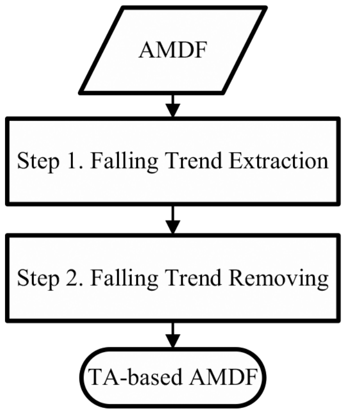

Fig. 2. Trend Analysis-based framework for improving AMDF.

Due to the fine performance of EMDAMDF and inspired by Equation (7) to calculate EMDAMDF, we propose a novel modified framework for AMDF as shown in Fig. 2. This is a simple framework and differs from the two ones mentioned in Section 2. As the figure describes, it consists of two steps, namely, Step 1. falling trend extraction, and Step 2. falling trend removal. More specifically, given an AMDF of speech frame denoted by $D_{AMDF}(\tau)$ we first use some methods such as EMD mentioned above to analyze its mathematical form of the falling trend $r_{trend}(\tau)$ and then remove the falling trend from AMDF to obtain the modified AMDF that we call Trend Analysis based AMDF (TA-based AMDF or TAAMDF) as following:

$$D_{TAAMDF}(\tau)=D_{AMDF}(\tau)-r_{trend}(\tau) \tag{8}$$

Compared with the other two frameworks mentioned before, our framework aims to analyze the falling trend based on AMDF and then remove it from AMDF instead of modifying the definition of AMDF and changing the length of the speech frame.

It is clear that the key part of our framework is Step 1, i.e., how to extract the falling trend. Therefore, what we focus on is converted to a trend analysis problem in time series analysis. In time series analysis, trend analysis is not an easy question. For many complex uncertain trends, it is difficult to estimate their concrete form. Based on the further analysis of EMDAMDF, we can convince that the falling trend of AMDF is nearly a linear trend and our framework need not pursue preciseness of trend analysis. As shown in Fig. 1 (f) TAAMDF adopting linear trend and least square can eliminate the falling trend and obtain the accurate pitch as well as EMDAMDF. Therefore, we believe that although we do not know the mathematical form of the falling trend of AMDF, many existing conventional trend analysis methods are available for our framework.

As the trend analysis methods for time series analysis have no strict classification, we summarize two types of falling trend analysis methods for the proposed framework. Generally, one can be called decomposition method such as EMD and the other can be called estimation method such as least square. Fig. 3 is an incomplete list of falling trend extraction methods we summarized. As shown in decomposition methods, we are also able to employ other signal analysis methods such as wavelet analysis to extract the falling trend instead of EMD. In estimation methods, least-square is a representative effective method to estimate the falling trend. Usually, the falling trend is assumed as a specified form such as Linear, Polynomial, Gaussian, etc. Then least square is used to estimate the parameters of the falling trend based on all data points of AMDF such that we can obtain the concrete form of the falling trend of AMDF.

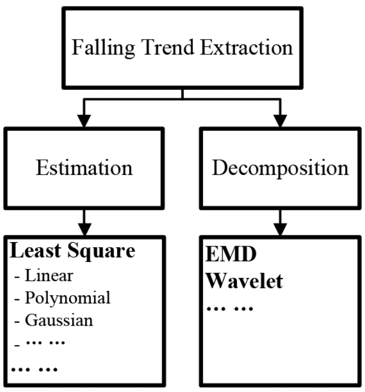

Fig. 3. Two types of falling trend extraction methods for the proposed framework.

For completely understanding our framework, now we describe how to calculate TAAMDF using the estimation method with least square plus polynomial, for example. Suppose we have an AMDF of a voiced speech frame $D(\tau)$ and its lag $\tau = 1, 2, \cdots, n$. Accordingly, all the data points of AMDF are $(1, D(1))$, $(2, D(2))$, $\cdots$, $(n, D(n)$. We use polynomial with the degree m to estimate the falling trend of AMDF and denote it as:

$$r_{trend}(\tau) = a_0 + a_1\tau + a_2\tau^2 + \cdots + a_m\tau^m = \sum_{j=0}^{m} a_j\tau^m \tag{9}$$

Substituting all the data points into (9), we obtain that

$$\begin{cases} D(1) = a_0 \cdot 1^0 + a_1 \cdot 1^1 + a_2 \cdot 1^2 + \cdots + a_m \cdot 1^m \\ D(2) = a_0 \cdot 2^0 + a_1 \cdot 2^1 + a_2 \cdot 2^2 + \cdots + a_m \cdot 2^m \\ \quad \cdots\cdots \\ D(n) = a_0 \cdot n^0 + a_1 \cdot n^1 + a_2 \cdot n^2 + \cdots + a_m \cdot n^m \end{cases} \tag{10}$$

Let $A = \begin{bmatrix} 1 & 1 & 1^2 & \cdots & 1^m \\ 1 & 2 & 2^2 & \cdots & 2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & n & n^2 & \cdots & n^m \end{bmatrix}$, $x = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $b = \begin{bmatrix} D(1) \\ D(2) \\ \vdots \\ D(n) \end{bmatrix}$ As is known that $rank(A) = m+1<n$ the least square solution of the parameter vector $\hat{x}$ can be calculated by

$$\hat{x} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_n \end{bmatrix} = (A^T A)^{-1} A^T b \tag{11}$$

Thus, we obtain the falling trend:

$$\hat{r}_{trend}(\tau) = \hat{a}_0 + \hat{a}_1\tau + \hat{a}_2\tau^2 + \cdots + \hat{a}_m\tau^m = \sum_{j=0}^{m} \hat{a}_j\tau^m \tag{12}$$

Based on the proposed framework in Fig. 2, the TAAMDF can be calculated as:

$$D_{TAAMDF}(\tau) = D(\tau) - \hat{r}_{trend}(\tau) \tag{13}$$

IV. Experiments and Analysis

We test our trend analysis-based framework for AMDF using the *Keele* pitch extraction reference database [14]. The *Keele* database consists of 5 mature females and 5 mature male speakers. Each speaker read a phonetically balanced text. The speech signals are sampled at 20 kHz with 16-bits resolution. The Database provides reference pitch values at 100 Hz frame rate with 26.5 ms rectangular window. Some frames with uncertain reference pitch recorded as '-1' are totally cut down. The whole samples of the database are all employed here.

We choose EMD from decomposition methods and polynomial from estimation methods to extract the falling trend to obtain TAAMDFs based on our framework denoted by TAAMDF (EMD) (*i.e.*, EMDAMDF in [12]) and TAAMDF (Poly) respectively. Note that according to large numbers of experiments and analyses we obtain a reliable formula to determine the degree of the polynomials m for TAAMDF, i.e., $m = \text{int}\,(L_{frame} \cdot f_{raw} / fs)$ where int is integer operation, L_{frame} is the frame length, fs is the sampling frequency and f_{raw} is "raw pitch". So-called "raw pitch" here actually refers to the empirical pitch ranges of females and males. Usually, we consider that in our framework it is feasible to set f_{raw} as 100 Hz for male and 200 Hz for female. Therefore, we can set the degree of polynomial of TAAMDF be 3 for male speech and 5 for female speech with regard to the *Keele* database. We compare the TAAMDF with CAMDF and EAMDF which are two state-of-the-art improvements of AMDF belonging to the other two frameworks discussed in the previous section. For showing the excellent performance of our framework, ACF, an outstanding classical PDA used by lots of speech analysis software, is included in the experiments as well. The experimental results are reported in terms of percentage GPE denoted as % GPE which is short for gross pitch error and defined by Rabiner et al. [15]. The definition of GPE is that the detected pitch period for a frame defers 1ms from the reference value. It should be noted that for fair comparison in all the experiments, both pre-processing and post-processing methods for error prevention and noise robustness such as band-pass filtering, half-wave rectification, center clipping and pitch smooth, are not employed. We want to show the most original performance of all PDAs.

TABLE I
PERFORMANCE OF DIFFERENT ALGORITHMS FOR CLEAN SPEECH OF THE WHOLE KEELE DATABASE

Method	***Keele***
AMDF	17.75
CAMDF	12.30
EAMDF	7.36
ACF	11.07
TAAMDF (EMD)	8.89
TAAMDF (Poly)	**7.13**

Table I gives % GPE of the whole *Keele* database detected by ACF, AMDF, CAMDF, EAMDF, TAAMDF (EMD), and TAAMDF (Poly). It can be observed that CAMDF, EAMDF, and two TAAMDFs all improve AMDF significantly. Although AMDF is not as good as ACF, its modifications can all turn the situation around (CAMDF is an exception but approximate). Besides, we notice that the proposed TAAMDF (Poly) has the least % GPE (7.13%) and owns the overall superiority among all PDAs while EAMDF performs excellently (7.36%) as well. TAAMDF (EMD) also achieves a fewer % GPE (8.89%) than CAMDF and ACF. Finally, from the experimental results, we notice that least square plus polynomial outperforms EMD for the proposed framework.

In order to further evaluate the performance especially the robustness of all PDAs, our experiments are also conducted by adding White, Babble, and Machinegun noise to the database at different Signal-to-Noise Ratio (SNR) set at 10, 5, 0, -5, and -10 dB respectively. Note that SNR is defined as, where S is the average power of the speech signal and N is the average power of the added noise. In our experiments, Babble and Machinegun noise are taken from NOISEX-92 database [16] and White noise is generated by awgn.m in MATLAB. Here we only choose the TAAMDF (poly) as the representative of our framework because of its performance shown in the former experiment. The experimental results are reported in Table II, Table III, and Table IV, respectively.

According to the results, something obvious can be observed. Firstly, different from the outcome for clean speech shown in Table I, ACF has remarkable progress in that it exceeds CAMDF and EAMDF and has a fewer % GPE than them. Then we can also find that although EAMDF has a very similar performance to TAAMDF for clean speech, its robustness to noise is so bad in noisy conditions. As is shown in Table II to Table IV, EAMDF always has the nearly most % GPE among these PDAs except AMDF in three noisy conditions. Finally, it is obvious that TAAMDF outperforms the other PDAs for every noise at any SNR. Besides, according to the definition of ACF, we know that ACF has the robustness to white noise (complete derivation can be seen in [17]). That is why for white noise ACF has a significant advantage over CAMDF and EAMDF. However, TAAMDF still has less % GPE than ACF in noisy conditions, especially for white noise. Therefore, we think that our framework is an efficient and reasonable way to improve AMDF. We also think that TAAMDF based on polynomial and least square within our framework is a more effective PDA than other modified AMDF and ACF.

TABLE II
PERFORMANCE OF DIFFERENT ALGORITHMS FOR THE WHOLE KEELE DATABASE POLLUTED BY WHITE NNOISE AT DIFFERENT SNR

Method	**10 dB**	**5 dB**	**0 dB**	**-5 dB**	**-10 dB**
AMDF	30.14	45.29	63.94	82.23	93.06
CAMDF	10.09	13.80	21.18	33.25	51.74
EAMDF	10.32	17.76	30.81	49.83	69.83
ACF	8.83	10.43	14.47	22.28	37.87
TAAMDF	**6.35**	**7.42**	**9.77**	**15.03**	**24.60**

TABLE III
PERFORMANCE OF DIFFERENT ALGORITHMS FOR THE WHOLE KEELE DATABASE POLLUTED BY BABBLE NOISE AT DIFFERENT SNR

Method	**10 dB**	**5 dB**	**0 dB**	**-5 dB**	**-10 dB**
AMDF	41.20	54.86	70.41	63.71	88.33
CAMDF	20.38	28.44	42.14	54.36	69.75
EAMDF	18.46	30.89	47.19	81.73	75.18
ACF	16.74	24.15	37.48	57.68	68.01
TAAMDF	**12.68**	**19.00**	**30.86**	**45.92**	**59.42**

TABLE IV
PERFORMANCE OF DIFFERENT ALGORITHMS FOR THE WHOLE KEELE DATABASE POLLUTED BY MACHINEGUN NOISE AT DIFFERENT SNR.

Method	**10 dB**	**5 dB**	**0 dB**	**-5 dB**	**-10 dB**
AMDF	30.81	36.37	43.28	50.77	58.64
CAMDF	19.28	23.86	29.71	36.13	42.97
EAMDF	15.94	22.36	29.67	37.74	45.50
ACF	17.15	21.86	28.12	34.99	42.18
TAAMDF	**11.36**	**14.89**	**19.83**	**26.11**	**32.83**

V. Conclusions

In this paper, we address the problem that multiple pitch errors sometimes appear in classical AMDF for pitch detection. We begin with a systematical review and analysis of its existing state-of-art modifications and sum up their improved ways as two kinds of frameworks. Then we propose a novel modified framework and two types of efficient falling trend extraction methods for the framework. Finally, experiments on the *Keele* database are conducted to test and validate the rationality and effectiveness of our modified framework. We can claim that the trend analysis based AMDF which chooses effective falling trend method owns the best performance especially for noisy speech and outperforms obviously modified AMDFs based on the other two frameworks summarized before and ACF which is an outstanding and well-known classical PDA.

Acknowledgement

This work was supported by the Teaching Reform Project of Nanjing Tech University Pujiang Institute under Grants njpj2019-1-03 and njpj2019-1-05.

References

[1] D. A. Krubsack and R. J. Niederjohn, "An Autocerrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise Corrupted Speech," *IEEE*, vol. 39, no. 2, pp. 319-329, Feb. 1991.
[2] M. Ross, H. Shaffer, and R. Freudberg et al., "Average Magnitude Difference Function Pitch Extractor," *IEEE*, vol. 22, no. 5, pp. 353-362, Oct. 1974.
[3] S. Ahmadi and A. S. Spanias, "Cepsrum-Based Pitch Detection Using a New Statistical V/UV Classification Algorithm," *IEEE*, vol. 7, no. 3, pp. 333-338, May. 1999.
[4] F. Kurth, A. Cornaggia-Urrigshardt, and S. Urrigshardt, "Robust F0 estimation in Noisy Speech Signals Using Shift Autocorrelation," in *ICASSP 2014-39th IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014, pp. 1468-1472.
[5] P. Pelle and C. Estienne, "A Robust Pitch Detector Based on Time Envelope and Individual Harmonic Information Using Phase Locked Loops and Consensual Decisions," in *ICASSP 2014-39th IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, 2014, pp. 1483-1487.
[6] P. Boersma, "Accurate Short-Term Analysis of The Fundamental Frequency and The Harmonics-To-Noise Ratio 0f a Sampled Sound," in *Proc. IFA International Conference*, 1993, pp. 97-110.
[7] A. De Cheveigné and H. Kawahara, "YIN, a Fundamental Frequency Estimator for Speech and Music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917-1930, Apr. 2002.
[8] W. Zhang, G. Xu, and Y. Wang, "Pitch Estimation Based on Circular AMDF," in *ICASSP 2002*, Orlando, Florida, USA, 2002, pp. I-341-I-344.
[9] G. Muhammad, "Noise Robust Pitch Detection Based on Extended AMDF," in *ISSPIT 2008-IEEE International Symposium on Signal Processing and Information Technology*, Sarajevo, Bosnia and Herzegovina, 2008, pp. 133-138.
[10] L. Gu and R. Liu, "High Performance Mandarin Pitch Estimation," *Acta Electronica Sinica*, vol. 27, no. 1, pp. 8-11, Jan. 1999.
[11] T. E. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10," *Speech Technology*, vol. 1, no. 2, pp. 40-49, Feb. 1982.
[12] Y. Zong, Y. Zeng, and M. Li et al., "Pitch Detection Using EMD-Based AMDF," in *Proc. ICICIP 2013*, Beijing, China, 2013, pp. 594-597.
[13] N. E. Huang, Z. Shen, S. R. Long et al., "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis," in *Proc. the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 1998, pp. 903-995.
[14] G. Meyer, F. Plante, and W. Ainsworth, "A Pitch Extraction Reference Database," in *Proc. European Conference on Speech Communication and Technology*, *EUROSPEECH 1995*, Madrid, Spain, 1995, pp. 827-840.
[15] L. R. Rabiner, M. J. Cheng, and C. A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 399-417, Oct. 1976.
[16] A. Varga and H. J. M. Steeneken, "Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems," *Speech Communication*, vol. 12, no. 3, pp. 247-251. Jul. 1993.
[17] T. Shimamura and H. Kobayashi, "Weighted Autocorrelation for Pitch Extraction of Noisy Speech," *IEEE*, vol. 9, no. 7, pp. 727-730, Oct. 2001.

Weihua Zhang received a B.S. degree in Electronic Information Engineering from Southeast University, Nanjing, China, and an M.S. degree in Software Engineering from Nanjing University, Nanjing, China. He is currently a Lecturer and the Vice Dean with the School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing, China. His research interests include signal processing, artificial intelligence, and affective computing.

Yingying Lu received a B.A. degree in Japanese Language and Literature from Nanjing Tech University, Nanjing, China in 2012, and an M.A. degree in Teacher Education from Nanjing Normal University, Nanjing, China in 2016, respectively. From 2016 to 2020, she was a Lecturer with the Department of Humanities, Nanjing Normal University Zhongbei College, Zhenjiang, China. Currently, she has been working as a Lecturer with the School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing, China. Her research interests include teacher education, psychological health education, and affective computing.

Pingping Xu received the B.S., M.S., and Ph.D. degrees all from Southeast University, Nanjing, China. She is currently a Professor with the National Laboratory of Mobile Communications, School of Information Science and Engineering, Southeast University and the Dean of the School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing, China. She was working as a guest researcher with the Department of Information and Communication Engineering, Aichi Institute of Technology, Japan in 1997, and with the Mobile Communication Lab, Osaka University, Japan in 2000, respectively. Dr. Xu has published more than 50 scientific papers, 4 books, and 20 patents, and obtained 4 science prizes. Her interests include signal processing, artificial intelligence, and affective computing.