Performance Evaluation of Novel AMDF-Based Pitch Detection Scheme

Sandeep Kumar

A novel average magnitude difference function (AMDF)-based pitch detection scheme (PDS) is proposed to achieve better performance in speech quality. A performance evaluation of the proposed PDS is carried out through both a simulation and a real-time implementation of a speech analysis-synthesis system. The parameters used to compare the performance of the proposed PDS with that of PDSs that are based on either a cepstrum, an autocorrelation function (ACF), an AMDF, or circular AMDF (CAMDF) methods are as follows: percentage gross pitch error (%GPE); a subjective listening test; an objective speech quality assessment; a speech intelligibility test; a synthesized speech waveform; computation time; and memory consumption. The proposed PDS results in lower %GPE and better synthesized speech quality and intelligibility for different speech signals as compared to the cepstrum-, ACF-, AMDF-, and CAMDF-based PDSs. The computational time of the proposed PDS is also less than that for the cepstrum-, ACF-, and CAMDF-based PDSs. Moreover, the total memory consumed by the proposed PDS is less than that for the ACF- and cepstrum-based PDSs.

Keywords: DSP, pitch detection, autocorrelation, average magnitude difference function.

I. Introduction

The estimation of pitch of a speech signal is a fascinating topic in the field of speech processing.

Accurately estimated pitch helps to improve the quality of a synthesized speech signal. Several pitch detection schemes (PDSs) based on either an autocorrelation function (ACF), an average magnitude difference function (AMDF) [1]–[3], a cepstrum [4], [5] or a wavelet [6], [7] have been developed over the past few decades.

Among the different methods for pitch detection of speech signals, the simplest is that of an AMDF-based method. Due to its simplicity, AMDF is used in real-time processing. However, the major drawback of this method is that the background noise of a speech signal has a negative effect on the minimum amplitude of a speech frame, which in turn negatively affects the accuracy of a pitch detector [8].

To improve the performance of AMDF, high resolution AMDF (HRAMDF) was proposed in [9]. However, this proposed PDS suffered from a phenomenon known as "double pitch error."

Further, a circular AMDF (CAMDF) was proposed in [10]. It was observed that the CAMDF-based PDS performed better than the HRAMDF-based PDS. However, the CAMDF-based PDS introduced a new type of "pitch error"; that is, octave error, which is caused by the presence of enhanced magnitudes at each pitch multiple.

PDSs based on a combination of AMDF and ACF have also been proposed [8], [11]. Although the performance of an ACF-AMDF-based PDS is better than that of either an AMDF- or ACF-based PDS [3], [8], [11], there is no significant difference in the computational complexity of such a scheme compared to an AMDF-based scheme.

In this paper, we propose a novel AMDF-based PDS to achieve better performance in speech quality. The novelty lies in the fact that the proposed AMDF-based PDS is based upon a modification of the AMDF occurring in [8].

Prior to being able to compute an AMDF, a speech signal must first be filtered by a low-pass elliptical filter and then by a numerical filter [9] so as to remove any high-frequency noise and formants. Thus, we choose to adopt the voiced/unvoiced decision scheme of [10] but modify it for use with the proposed novel AMDF-based PDS.

Previous efforts to compare the performances of PDSs have tended to focus solely on utilizing "accuracy of estimated pitch" as a performance measure. Furthermore, these previous efforts do not measure the performance of PDSs in real time. In fact, real-time performance comparisons of PDSs is little reported in the related literature. For a proper real-time performance comparison of PDSs, an analysis-synthesis scheme must include a "pitch detection" component block built with special-purpose processors (digital signal processors) — if it is to meet the "real-time" requirement (that is, there must not be any appreciable delay between the input and the output of the system).

Apart from a PDS's computation time, certain other issues, such as speech quality, also need to be investigated. Therefore, a speech analysis-synthesis system that makes use of the proposed novel AMDF-based PDS is first simulated and then implemented in real time using TMS320C6713 DSK within a MATLAB platform.

The performance of the proposed novel AMDF-based PDS is compared with that of cepstrum-, ACF- [6], [11], [12], AMDF- [1], and CAMDF-based [10] PDSs in terms of percentage gross pitch error (%GPE), a synthesized speech waveform, a mean opinion score (MOS) listening test, a Perceptual Evaluation of Speech Quality (PESQ) score, a diagnostic rhyme test (DRT), computation time, and memory consumption.

This paper is organized as follows. The proposed novel AMDF-based PDS is discussed in Section II. Results for the performance comparison of different PDSs is presented in Section III. Finally, Section IV provides some concluding remarks.

II. Proposed Novel AMDF-Based PDS

The AMDF of a speech signal is defined as follows [1]:

$$Y(k) = \frac{1}{N-k-1} \sum_{n=0}^{N-k-1} |x(n+k) - x(n)|, \qquad (1)$$

where $x(n)$ is a frame of the speech signal, N is the length of a frame of the speech signal, k is a lag number, and $1/(N-k-1)$

is a normalization factor. The range of k is $(0, N)$.

The pitch period of a speech signal, often denoted by T_P, can be determined by finding the position of the global minimum amplitude of the AMDF curve with respect to the origin. Equation (1) gives the locations of local minima amplitudes, of which the global minimum amplitude defines the pitch period of the speech signal. Since the global minimum amplitude of a signal is influenced by background noise [8], this alone (calculated from (1)) is insufficient for accurate pitch detection.

To improve the performance of a PDS, the AMDF of a speech signal is redefined by the researchers in [8] as follows:

$$Y(k) = \frac{1}{N-1} \sum_{n=2N-1}^{N} |x(n-k) - x(n)|, \qquad (2)$$

where two speech frames of $x(n)$ (current and previous frames) are instead used in the calculation of an AMDF value. In [8], it was observed that the local minima of the AMDF curve from (2) have a greater periodic nature than those calculated from (1).

In this paper, we modify (2) to obtain the following:

$$Y(k) = \frac{1}{k+1} \sum_{n=N}^{2N-1} |x(n-k) - x(n)|. \qquad (3)$$

We have changed the normalization factor, $1/(N-1)$, to $1/(k+1)$, where the range of k is $[1, N-1]$. We note that a further culprit of error in pitch detection is the falling trend of the AMDF peaks at higher lags.

Upon implementing the proposed novel AMDF-based PDS (whose steps are outlined below) in an analysis-synthesis system, we have observed that the performance of pitch detection is improved. The reason for this is that the proposed novel AMDF-based PDS uses (3) as opposed to (2).

The following are the steps of the proposed novel AMDF-based PDS:

1) First, a speech signal is filtered by a fifth-order low-pass elliptical filter with cut-off frequency of 800 Hz. This filter is used to eliminate high-frequency noise and formants; components pertaining to high frequencies do not contain significant information regarding pitch frequency and the "fundamental frequencies" region in the case of most men and women lies in the range 50 Hz to 500 Hz. The filter also preserves the first and second harmonics for a range of high pitch frequencies. The transfer function for the fifth-order low-pass elliptical filter is given by [9]

$$H(z) = \frac{0.008233 - 0.004879z^{-1} + 0.007632z^{-2}}{1 - 3.6868z^{-1} + 5.8926z^{-2} - 5.0085z^{-3} + 2.2518z^{-4} - 0.4271z^{-5}}$$
$$+ \frac{0.007632z^{-3} - 0.004879z^{-4} + 0.008233z^{-5}}{1 - 3.6868z^{-1} + 5.8926z^{-2} - 5.0085z^{-3} + 2.2518z^{-4} - 0.4271z^{-5}}. \qquad (4)$$

2) The accuracy of the PDS suffers if the first and second formants are present in a speech filtered by the fifth-order

low-pass elliptical filter. Therefore, an additional ninth-order numerical filter [9] is used to attenuate the first and second formants to improve the accuracy of the pitch detection method. The transfer function for this filter is as follows [9]:

$$H(z) = \frac{1}{9} \sum_{n=0}^{8} z^{-n}. \tag{5}$$

3) Now, calculate AMDF values using (3).
4) Find local minima for these AMDF values and set *Count* to the total number of local minima.
5) If *Count* $\leq$ 1, then set $T_p = 0$.
6) If *Count* > 1, then find the global minimum value, denote it by A, and denote its position by T, in the range between 16 to 160 samples with respect to the origin. The position T then signifies the pitch period (T_p) of the speech signal.
7) For the voiced/unvoiced decision scheme: Calculate the average (A_{vg}) of the local minima over the entire speech frame. For a certain frame, if $A > 0.70 A_{vg}$, then mark the frame as "unvoiced" and modify the pitch period T_p to be zero.
8) Modify the voicing decision for the current frame of the speech signal as follows:
   (a) If the preceding and succeeding frames are marked as "unvoiced," then mark the current frame as "unvoiced." Likewise, if the preceding and succeeding frames are marked as "voiced," then mark the current frame as "voiced."
   (b) If the preceding and succeeding frames have an approximately equal pitch period and the pitch for the current frame differs by more than 60%, then modify the pitch of the current frame to equal that of the average of the preceding and succeeding frames.

To remove a half/double pitch error in the proposed novel AMDF-based PDS, the pitch period of the current frame must be compared with that of the previous frame. If a half/double pitch error is found, then the pitch period of the current frame is modified by a factor of 2 or 0.5.

III. Results and Discussion on Performance Evaluation of PDSs

A model for a speech analysis-synthesis system using the proposed novel AMDF-based PDS was created using SIMULINK®. The digital signal processor (DSP) "TMS320C6713" was chosen to implement this model in real-time.

A description of the analysis and synthesis procedures used is as follows. At the analysis stage, a speech signal in ".wav" format with 8 kHz sampling frequency is divided into frames of length 20 ms. Then, filter parameters (such as voicing, gain,

filter coefficient, and pitch period) are extracted from the speech signal using a linear predictive analysis. Autocorrelation LPC is then used to represent the vocal tract parameter (in terms of reflection coefficient). For vocal tract extraction, the order of prediction is 15. At the synthesis stage, an impulse train is generated based on the estimated pitch period of the "voiced" frame. For the "unvoiced frame," random noise–like excitation is used. A "voiced"/"unvoiced" decision switch is used to select the proper excitation signal. Finally, the proper excitation signal, gain, and filter coefficients are used to reconstruct the speech signal.

The model was tested with different speech files selected from PTDB-TUG (clean speech database) [13], NOIZEUS (noisy speech database with various SNR levels) [14], and Keele (pitch reference database) [15]. These three speech databases consist of different speech files uttered by both male and female speakers. The experiment using different PDSs was performed in a quiet environment. The quality of the synthesized speech was listened to. The performance of the proposed novel AMDF-based PDS was compared with cepstrum-, ACF-, AMDF-, and CAMDF-based PDSs. The parameters used for performance comparison are as follows: %GPE; a subjective listening test (MOS); an objective speech quality assessment (PESQ); a speech intelligibility test (DRT); a synthesized speech waveform; computation time; and memory consumption. The computation time of the different PDSs is calculated using both a tool named Profiler (for simulation) and a method named Breakpoint (for real-time implementation). However, a CCS® DSP/BIOS configuration file is used for memory calculation.

1. Gross Pitch Error

The %GPE for the different PDSs is evaluated using a pitch reference database named Keele. If a measured pitch differs from the reference pitch by more than 1 ms, then it is termed a "gross pitch error." The reference pitch values are obtained from the original database. Table 1 shows the %GPE for the

Table 1. Results for %GPE.

Method used	%GPE					
	Clean	10 dB	5 dB	0 dB	–5 dB	–10 dB
Cepstrum	12.36	15.52	19.38	23.68	38.45	74.12
AMDF	10.55	14.48	18.25	21.42	36.82	72.14
CAMDF	8.23	11.38	16.68	18.79	31.98	67.68
ACF	8.15	11.27	16.14	18.24	30.26	66.13
Proposed novel AMDF	7.86	10.86	15.68	16.96	29.35	65.42

Table 2. Results for PESQ scores.

Method used	Average PESQ score		
	Male speakers	Female speakers	Overall (male & female)
Original unprocessed speech	4.50	4.50	4.50
Cepstrum	2.05	1.88	1.97
AMDF	2.30	2.03	2.17
CAMDF	2.44	2.18	2.31
ACF	2.43	2.28	2.36
Proposed novel AMDF	2.61	2.30	2.45

different PDSs using clean and noisy speech with different SNR levels (10 dB, 5 dB, 0 dB, −5 dB, −10 dB). From the results, it can be seen that the %GPE obtained for the proposed novel AMDF-based PDS is less than the cepstrum-, AMDF-, CAMDF-, and ACF-based PDSs.

2. Objective Test Results

An objective assessment of the quality of the synthesized speech signals through ITU-T P.862 PESQ scores has been made [16]. PESQ compares a processed speech signal against the original speech signal, and the resulting PESQ score is mapped to a MOS-like scale with a range between −0.5 and 4.5. Twenty-five different speech sounds uttered by both males and females (randomly selected from the PTDB-TUG database) were used in the test. The PESQ scores for the different PDSs are shown in Table 2. The PESQ scores obtained for the original speech material are also provided as a benchmark score for the speech processed by the cepstrum-, AMDF-, CAMDF-, ACF- and proposed novel AMDF-based PDSs.

From the PESQ scores presented in Table 1, it is clear that the proposed novel AMDF-based PDS performs better than the cepstrum-, AMDF-, CAMDF-, and ACF-based PDSs.

3. Subjective Test Results

The subjective quality of the synthesized speech signals for different PDSs was evaluated by an MOS listening test [17]. In this test, 20 normal-hearing listeners were chosen and trained (before the test, a training on the listening test was given to every listener so that they can be familiar with the test) to rate their subjective impression on a five-point scale for different speech signals. An MOS rating of "5" and of "1" indicates "excellent speech quality" and "worst speech quality," respectively. We have used the same material for the MOS test as was used for the PESQ assessment. The results of the MOS

Table 3. Results for MOS test scores.

Method used	Average MOS score		
	Male speakers	Female speakers	Overall (male & female)
Original unprocessed speech	4.25	4.05	4.15
Cepstrum	2.75	2.58	2.67
AMDF	3.02	2.80	2.91
CAMDF	3.08	2.92	3.00
ACF	3.11	2.96	3.04
Proposed novel AMDF	3.24	3.00	3.12

Table 4. MOS test results for noisy speech (SNR = 0 dB).

Method used	Average MOS score		
	Male speakers	Female speakers	Overall (male & female)
Original unprocessed speech	2.60	2.10	2.35
Cepstrum	1.75	1.40	1.58
AMDF	1.85	1.50	1.68
CAMDF	1.95	1.65	1.80
ACF	2.00	1.65	1.83
Proposed novel AMDF	2.10	1.75	1.93

Table 5. MOS test results for noisy speech (SNR = 5 dB).

Method used	Average MOS score		
	Male speakers	Female speakers	Overall (male & female)
Original unprocessed speech	2.85	2.45	2.65
Cepstrum	1.85	1.55	1.70
AMDF	2.00	1.75	1.88
CAMDF	2.10	2.00	2.05
ACF	2.10	2.00	2.05
Proposed novel AMDF	2.25	2.05	2.15

test are presented in Table 3. The MOS scores obtained for the original speech material are also provided as a benchmark score for the speech processed by the different PDSs and presented with the MOS test results.

From the MOS scores presented in Table 3, it is clear that the proposed novel AMDF-based PDS performs better than the cepstrum-, AMDF-, CAMDF-, and ACF-based PDSs, which is an agreement with the result (PESQ scores) presented in Table 1. We also created a subjective listening test using noisy

speech database (NOIZEUS) to test the reliability and performance of the proposed novel AMDF-based PDS. The MOS test results for noisy speech with SNR 0 dB and 5 dB are presented in Tables 4 and 5, respectively.

From the MOS scores presented in Tables 4 and 5, it is clear that the performance of the proposed novel AMDF-based PDS is better than the cepstrum-, AMDF-, CAMDF-, and ACF-based PDSs for noisy speech, which is similar to the results presented in Table 1 in the case of clean speech.

4. Synthesized Speech Waveform

The SIMULINK model was simulated with different clean and noisy speech files having a sampling frequency of 8 kHz. The results for the case of synthesized speech have been noted and presented in two speech files — one noisy speech file (F13.wav) with SNR 0 dB and corrupted by noise from a train

station, and one clean speech file (s16.wav) that contains the sentences "Smoke poured out of every crack.," spoken by an English-speaking female, and "One validate acts of school districts.," spoken by an English-speaking male. The length of the speech files were 2.54 s and 6.87 s, respectively. The original and synthesized speech waveforms using different methods for pitch detection are shown in Figs. 1, 2, 3, 4, 5, and 6.

From Figs. 2, 3, 4, 5, and 6, it is observed that the synthesized speech waveforms (for both clean and noisy speech signals) obtained with the proposed novel AMDF-based and ACF-based PDSs are remarkably better than those obtained with the cepstrum-, CAMDF-, and AMDF-based PDSs.

5. Diagnostic Rhyme Test

A speech intelligibility test (a subjective measurement technique to evaluate the speech quality) of a synthesized
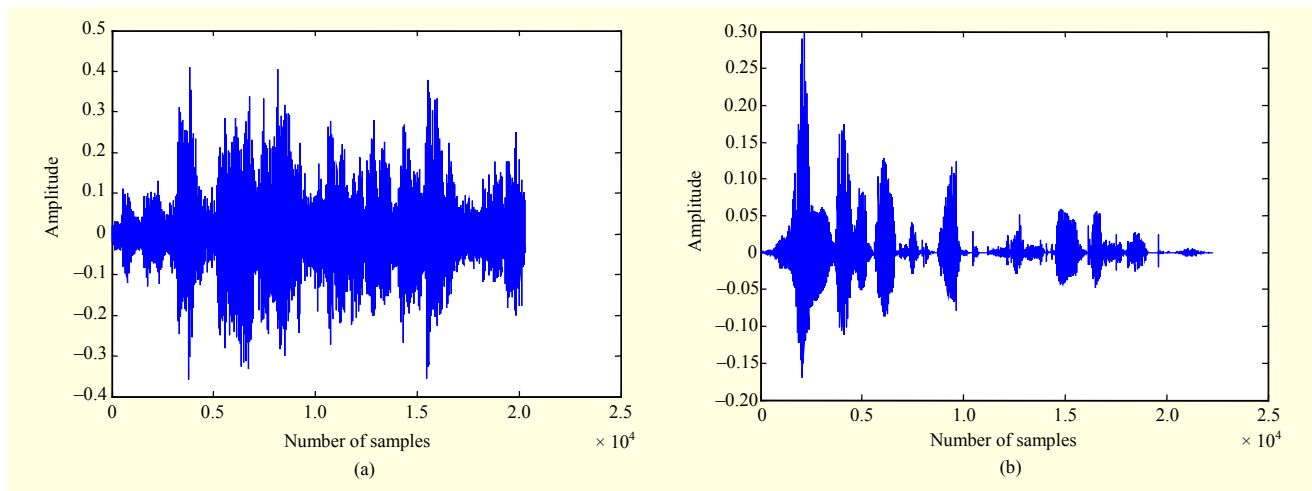

Fig. 1. Original speech signals: (a) noisy speech F13.wav and (b) clean speech s16.wav.
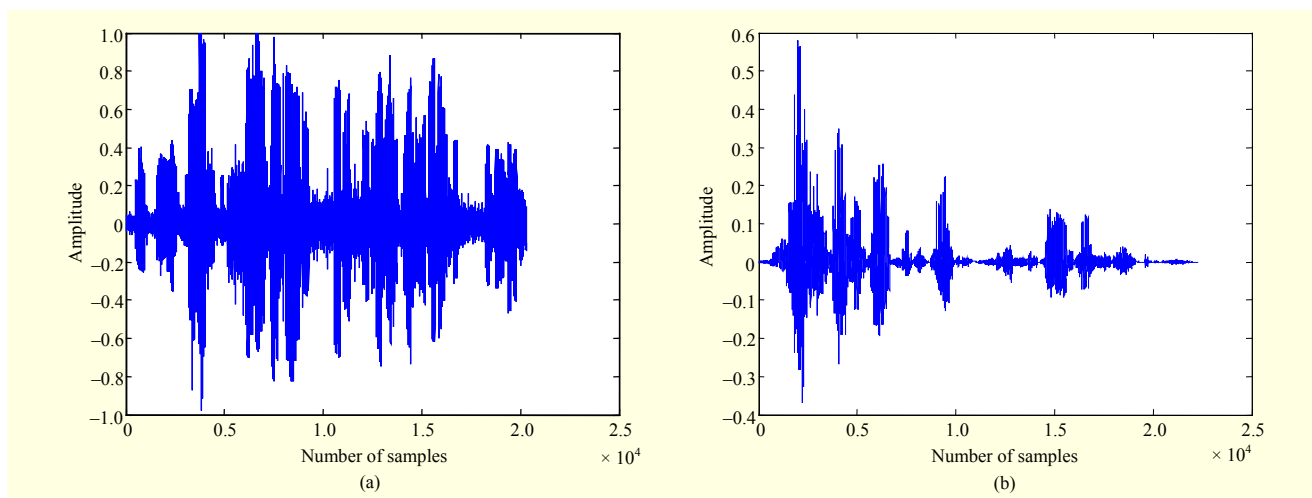

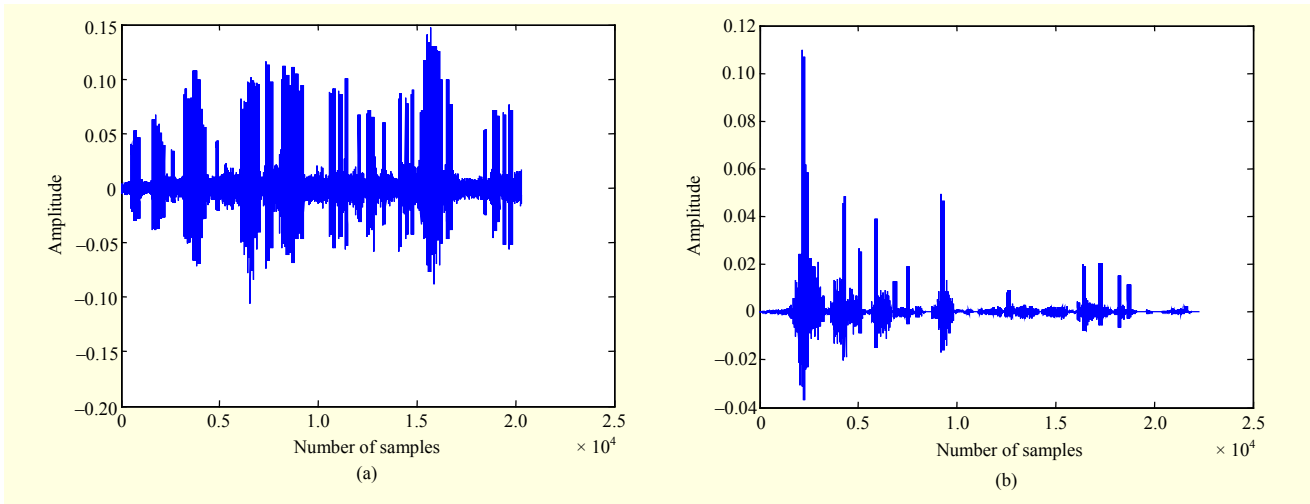Fig. 2. Synthesized speech signals for ACF-based pitch detection: (a) F13.wav and (b) s16.wav.

Fig. 3. Synthesized speech signals for cepstral analysis: (a) F13.wav and (b) s16.wav.
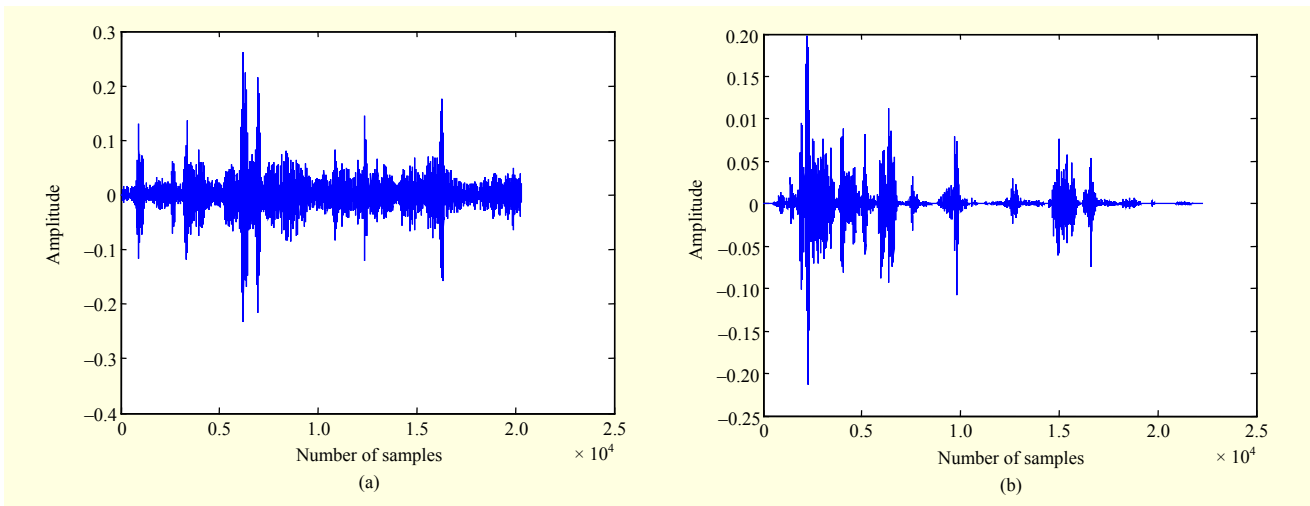

Fig. 4. Synthesized speech signals for AMDF-based pitch detection: (a) F13.wav and (b) s16.wav.
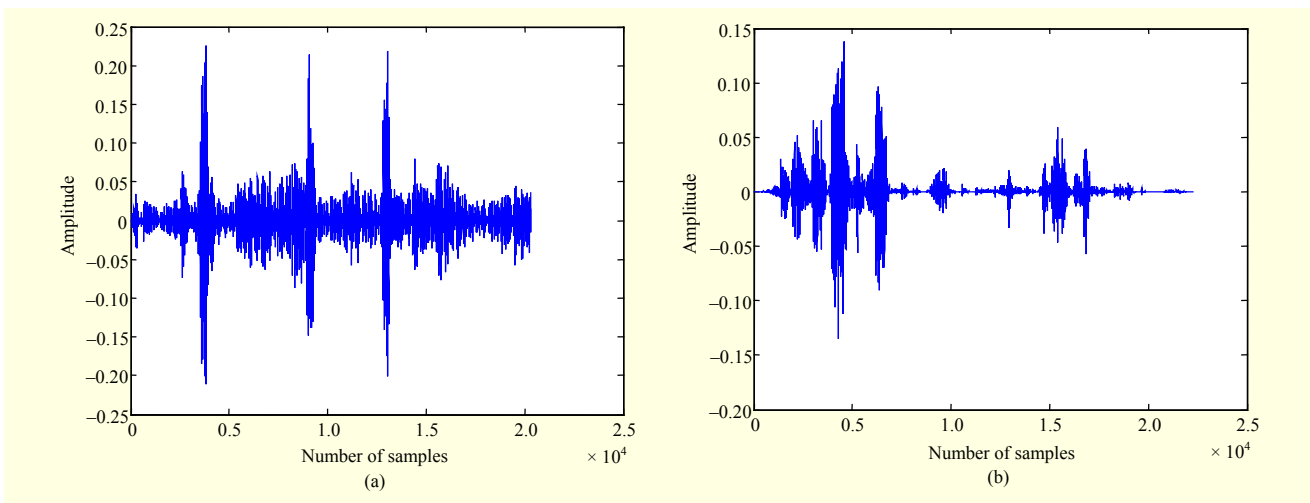

Fig. 5. Synthesized speech signals for CAMDF-based pitch detection: (a) F13.wav and (b) s16.wav.
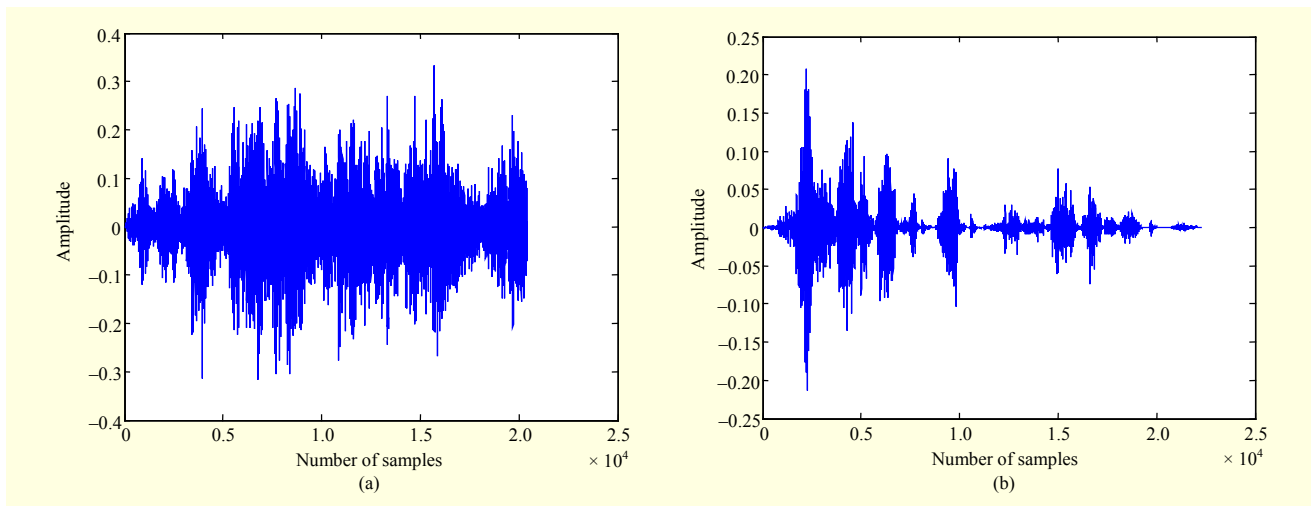
Fig. 6. Synthesized speech signals for proposed novel AMDF-based pitch detection: (a) F13.wav and (b) s16.wav.

speech signal has been performed. The DRT is a popular, widely used test for intelligibility of speech [17]. In this test, six phonetic attributes (voicing, nasality, sustention, sibilation, graveness, and compactness) are tested using a corpus of 192 words in 96 rhyming pairs (a word pair differs by only one phoneme, usually a consonant). Ninety-six of the 192 words (only one word from each word pair) were read out and recorded. This process has been carried out for three male and three female speakers.

Six listeners were chosen for this test — they did not know which word was from which word pair. After listening to the synthesized speech signal using the different PDSs, the listener is asked to identify and mark the correct word from a word pair. Ten speech files (two different synthesized speech files for each PDS) processed by the system were given to each listener. Each speech file contained 96 recorded words for the various speakers. The overall DRT score was calculated from the correct/incorrect response marked by the listener, as follows:

$$DRT\ score\ =\ \frac{100\ (correct\ response - incorrect\ response)}{total\ number\ of\ tests}.$$

(6)

For a given speech signal, a DRT score above 85 would indicate that it is of good quality [17]. The results of the DRT test for the different PDSs are presented in Table 6. The DRT scores obtained for the original unprocessed speech material are also provided as a benchmark score for the speech processed by the different PDSs and presented with the DRT test results in Table 6.

From the results of the DRT test, it is clear that the synthesized speech using the proposed novel AMDF-based PDS is more intelligible than that from the cepstrum-, AMDF-,

Table 6. DRT scores for different PDSs.

Method used	DRT score		
	Male speakers	Female speakers	Overall (male & female)
Original unprocessed speech	97.89	96.84	97.36
Cepstrum	75.87	73.96	74.91
AMDF	87.50	81.94	84.72
CAMDF	88.19	82.98	85.59
ACF	89.06	83.85	86.46
Proposed novel AMDF	91.49	86.97	89.23

ACF- and CAMDF-based PDSs.

6. Computation Time

A. Time Taken During Simulation

The average simulation time for the analysis-synthesis system with different PDSs was calculated with the help of a tool named Profiler [18]. The results obtained from the different PDSs are marked as ten different speech files (s6.wav, s7.wav, s8.wav, s9.wav, s10.wav, s16.wav, s17.wav, s18.wav, s19.wav, and s20.wav) and are presented in Table 7. The first five speech files (s6.wav, s7.wav, s8.wav, s9.wav, s10.wav), corresponding to five English-speaking females, contained the sentences "She had your dark suit in greasy wash water all year.," "Jane may earn more money by working hard.," "Don't ask me to carry an oily rag like that.," "At twilight on the twelfth day will have Chablis." and "Cut a small corner off each edge.," respectively. The length of the speech files were

Table 7. Average simulation time for speech analysis-synthesis system with different PDSs.

Wav file used	Length of speech (s)	Total number of frame processed	Time taken with autocorrelation based PDS (ms)		Time taken with cepstral-based PDS (ms)		Time taken with AMDF based PDS (ms)		Time taken with CAMDF based PDS (ms)		Time taken with proposed novel AMDF based PDS (ms)	
			Total time	Time/frame	Total time	Time/frame	Total time	Time/frame	Total time	Time/frame	Total time	Time/frame
S6	7.99	400	8,602	21.51	9,210	23.03	8,298	20.74	8,330	20.83	8,314	20.79
S7	8.06	403	8,806	21.85	9,406	23.33	8,250	20.47	8,414	20.88	8,392	20.82
S8	6.88	344	7,210	20.96	8,314	24.16	7,010	20.37	7,089	20.61	7,078	20.57
S9	7.41	371	7,651	20.62	8,595	23.16	7,414	19.98	7,511	20.24	7,500	20.21
S10	6.75	338	7,226	21.37	7,909	23.39	6,902	20.42	7,001	20.71	6,938	20.52
S16	6.87	344	7,198	20.92	7,929	23.04	6,889	20.02	6,997	20.34	6,921	20.11
S17	7.14	357	7,498	21.00	8,502	23.81	7,240	20.28	7,297	20.43	7,288	20.41
S18	6.87	344	7,231	21.02	8,002	23.26	7,022	20.41	7,088	20.60	7,082	20.58
S19	9.43	472	9,798	20.75	10,990	23.28	9,598	20.33	9,631	20.41	9,612	20.36
S20	9.16	458	9,628	21.02	10,920	23.84	9,502	20.74	9,563	20.88	9,562	20.87

7.99 s, 8.06 s, 6.88 s, 7.41 s, and 6.75 s, respectively. The next five speech files (s16.wav, s17.wav, s18.wav, s19.wav and s20.wav), corresponding to five English-speaking males, contained the sentences "One validate acts of school districts.," "Two other cases also were under advisement.," "Their props were two stepladders, a chair and a palm fan.," "Selecting bunks by economic comparison is usually an individual problem.," and "Bright sunshine shimmers on the ocean.," respectively. The length of these speech files were 6.87 s, 7.14 s, 6.87 s, 9.43 s, and 9.16 s, respectively.

From the results presented in Table 7, it is observed that the average simulation time per frame for the different speech files is least in the case of the AMDF-based PDS, and the performance of the proposed novel AMDF-based PDS is close to this PDS. Moreover, the performance of the proposed novel AMDF-based PDS, in terms of simulation time, is better than that for the ACF-, CAMDF-, and cepstrum-based PDSs. This can be justified by the fact that the AMDF-based PDSs involve only addition and modulus operations; hence, they are computationally simpler in comparison to the other PDSs. On the other hand, the ACF-based PDS involves the summation of products; hence, it is computationally more complex than the AMDF-based PDSs [7], [11]. The cepstrum-based PDS is computationally more complex compared to the other PDSs, because it involves computation of a Fourier transform, the logarithm of the power spectrum, and the inverse Fourier transform. However, due to the involvement of a number of pre-processing and post-processing steps, the computational complexity of the proposed novel AMDF-based PDS is slightly higher than that of the standard AMDF-based PDS.

From Table 7, it can also be seen that for all PDSs, the computation time per frame is greater than the actual frame duration; that is, 20 ms. However for the real-time application, the computation time per frame should be less than the actual frame duration. Therefore, implementation of a real-time model of a speech analysis-synthesis system that makes use of various PDSs with a fast DSP is required to fulfill the real-time need.

B. Execution Time (for Real-Time Implementation)

The execution time for a real-time model such as that mentioned in the previous subsection has been calculated by inserting breakpoints into the generated C code of the analysis-synthesis system [19]. The average number of cycles (N) for each frame (frame size is equal to 20 ms) was calculated between the breakpoints. The execution time (ET) was calculated as follows:

$$ET = T_{pc} \cdot N, \qquad (7)$$

where T_{pc} is the execution time per cycle. Here, T_{pc} is 4.44 ns, since the clock frequency of the TMS320C6713 processor is 225 MHz. The execution times for the different PDSs are presented in Table 8. From the results, it can be observed that the execution time for the AMDF-based PDS is least as compared to the other PDSs. However, the execution time for the proposed novel AMDF-based PDS is also less than that of the ACF-, cepstrum-, and CAMDF-based PDSs. Moreover, the execution time per frame for all PDSs is less than the frame duration (20 ms), which meets the real-time requirement.

7. Memory Consumed

The memory consumptions of the different PDSs were

Table 8. Execution times for different PDSs.

PDS used	Avg no. of cycles (N)	Execution time (ms)
ACF	4,431,888	19.678
Cepstrum	4,497,168	19.967
CAMDF	4,217,861	18.727
AMDF	4,096,922	18.190
Proposed novel AMDF	4,209,709	18.690

Table 9. Memory consumption for implementation using different PDSs.

PDS used	Program memory (bytes)	Data memory (bytes)	Stack memory (bytes)	Total memory (bytes)
ACF	250,608	14,805	640	266,053
Cepstrum	359,344	19,057	640	379,041
AMDF	69,876	23,309	640	93,825
CAMDF	67,844	23,309	640	91,793
Proposed novel AMDF	74,124	27,561	640	102,325

calculated using *CCS® DSP/BIOS configuration file* [12] and are presented in Table 9. The total memory consists of three memory spaces — program memory, data memory, and stack memory.

From the results presented in Table 9, it can be seen that the stack memory is the same for all PDSs. The data memory for the ACF-based PDS is the least of all the PDSs. However, the program memory for the proposed novel AMDF-based, CAMDF-based, and AMDF-based PDSs is less compared to that for the ACF-based and cepstrum-based PDSs. Moreover, the total memory required for the proposed novel AMDF-based PDS is less than that for the ACF-based and cepstrum-based PDSs.

In summary, the performance comparison results of the different PDSs shows that the proposed novel AMDF-based PDS is better than the AMDF-, CAMDF-, ACF-, and cepstrum-based PDSs in terms of %GPE, synthesized speech quality, and intelligibility. In addition, the computational complexity of the proposed novel AMDF-based PDS is less than that of the ACF-, CAMDF-, and cepstrum-based PDSs. Moreover, the memory consumed by the proposed novel AMDF-based PDS is less than that for the ACF- and cepstrum-based PDSs.

IV. Conclusion

A novel AMDF-based pitch detection scheme (PDS) has been proposed. Using this PDS, a speech analysis-synthesis system has been simulated and also implemented in real time. The performance of the system has been tested with different speech files (for both clean and noisy speech signals) for both simulation and real-time implementation. From the results of the performance evaluation, it was found that the %GPE for the proposed novel AMDF-based PDS is less than that for the AMDF-, CAMDF-, ACF-, and cepstrum-based PDSs. The quality and intelligibility of synthesized speech for the proposed novel AMDF-based PDS is better than that for the AMDF-, CAMDF-, ACF-, and cepstrum-based PDSs. Moreover, the proposed novel AMDF-based PDS involves less computation time as compared to the ACF-, CAMDF-, and cepstrum-based PDSs. In addition, the proposed novel AMDF-based PDS required less memory as compared to the ACF- and cepstrum-based PDSs.

For future work, the evaluation of performance of the proposed novel AMDF-based PDS with different databases may be carried out.

References

[1] X.D. Mei, J. Pan, and S.-H. Sun, "Efficient Algorithm for Speech Pitch Estimation," *Proc. Int. Symp. Intell. Multimedia, Video Speech Process.*, Hong Kong, China, May 2–4, 2001, pp. 421–424.

[2] M.J. Ross et al., "Average Magnitude Difference Function Pitch Extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 22, no. 5, Oct. 1974, pp. 353–362.

[3] S. Kumar, S.K. Singh, and S. Bhattacharya, "Performance Evaluation of a ACF-AMDF Based Pitch Detection Scheme in Real Time," *Int. J. Speech Technol.*, vol. 18, no. 4, Dec. 2015, pp. 521–527.

[4] F. Wang and P. Yip, "Cepstrum Analysis Using Discrete Trigonometric Transforms," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 39, no. 2, Feb. 1991, pp. 538–541.

[5] H. Huang and J. Pan, "Speech Pitch Determination Based on Huang-Hilbert Transform," *Signal Process.*, vol. 86, no. 4, Apr. 2006, pp. 792–803.

[6] S. Kumar et al., "Performance Evaluation of a Wavelet-Based Pitch Detection Scheme," *Int. J. Speech Technol.*, vol. 16, no. 4, Dec. 2013, pp. 431–417.

[7] S. Kadambe and G.F. Boudreaux-Bartels, "Application of the Wavelet Transform for Pitch Detection of Speech Signals," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, Mar. 1992, pp. 917–924.

[8] L. Hui, B.-Q. Dai, and L. Wei, "A Pitch Detection Algorithm Based on AMDF and ACF," *IEEE Int. Conf. Acoust., Speech Signal Process.*, Toulouse, France, May 14–19, 2006, pp. 377–380.

[9] R. Cai, S. Shi, and Y. Zhu, "A Modified Pitch Detection Method

Based on Wavelet Transform," *Int. Conf. Multimedia Inf. Technol.*, Kaifeng, China, Apr. 2010, pp. 246–249.

[10] W. Zhang, G. Xu, and Y. Wang, "Pitch Estimation Based on Circular AMDF," *IEEE Int. Conf. Acoust., Speech Signal Process.*, Orlando, FL, USA, May13–17, 2002, pp. I.341– I.344.

[11] S. Kumar, S. Bhattacharya, and P. Patel, "A New Pitch Detection Scheme Based on ACF and AMDF," *Int. Conf. Adv. Commun. Contr. Comput. Technol.*, Ramanathapuram, India, May 18–10, 2014, pp. 1235–1240.

[12] S. Bhattacharya, S.K. Singh, and T. Abhinav, "Performance Evaluation of LPC and Cepstral Speech Coder in Simulation and in Real Time," *Int. Conf. Recent Adv. Inf. Technol.*, Dhanbad, India, Mar. 15–17, 2012, pp. 826–831.

[13] G. Pirker et al., "A Pitch Tracking Corpus with Evaluation on Multi-pitch Tracking Scenario," *Interspeech*, Florence, Italy, Aug. 27–31, 2011, pp. 1509–1512.

[14] Y. Hu and P. Loizou, "Subjective Evaluation and Comparison of Speech Enhancement Algorithms," *Speech Commun.*, July 2007, vol. 49, pp. 588–601.

[15] F. Plante, G.F. Meyer, and W.A. Ainsworth, "A Pitch Extraction Reference Database," *European Conf. Speech Commun. Technol.*, Madrid, Spain, Sept. 18–21, 1995, pp. 837–840.

[16] ITU-T P.862, *Perceptual Evaluation of Speech Quality (PESQ)*, June 2004.

[17] J.R. Deller, J.H.L. Hansen, and J.G. Proakis, "*Discrete-Time Processing of Speech Signal*," Piscataway, NJ, USA: John Wiley & Sons, 2000, pp. 570–579.

[18] *MATLAB Online Help*, Accessed Aug. 12, 2012. http://www. mathworks.in/help/toolbox/simulink/ug/f0–7640.html

[19] *Breakpoint Help*, Accessed July 10, 2012. http://processors. wiki.ti.com/index.php/Breakpoint

Sandeep Kumar received his B.Tech degree in electronics & instrumentation engineering from the Institute of Engineering & Technology, MJP Rohilkhand University, Bareilly, India, in 2006 and his M.Tech and PhD degrees in electronics & communication engineering from the Indian School of Mines, Dhanbad, India, in 2008 and 2015, respectively. From 2009 to 2013 and from 2013 to 2014, he worked as an assistant professor with the Department of Electronics & Communication Engineering, S. R. Group of Institution, Jhansi, India, and SRMS College of Engineering & Technology, Bareilly, India, respectively. Since 2015, he has been with the Department of Electronics & Telecommunication Engineering, Rungta College of Engineering & Technology, Bhilai, India, where he is now an associate professor. His research interests include digital signal processing and its application in speech, audio, and image processing.