Cogn Comput (2010) 2:151–159
DOI 10.1007/s12559-010-9048-1

Autocorrelation of the Speech Multi-Scale Product for Voicing Decision and Pitch Estimation

Mohamed Anouar Ben Messaoud · Aïcha Bouzid · Noureddine Ellouze

Received: 30 December 2009 / Accepted: 10 May 2010 / Published online: 26 May 2010

Abstract In this work, we present an algorithm for voiced/unvoiced decision and pitch estimation from speech signals. Our approach is based on classifying the peaks provided by the autocorrelation of the speech multi-scale product. The multi-scale product is based on making the product of the speech wavelet transform coefficients at three successive dyadic scales. The autocorrelation function of the multi-scale product is calculated over frames of a specific length. The experimental results show the robustness and the effectiveness of our approach. Besides, the proposed method outperforms some existing algorithms in a clean and noisy environment.

Keywords Wavelet transform · Multi-scale product · Autocorrelation · Voicing decision · Pitch

Introduction

The classification of the speech signal into voiced, unvoiced and silence provides a preliminary acoustic segmentation of the speech, which is important for speech analysis. The nature of the classification is to determine whether a speech signal is present and if so, whether the production of speech involves the vibration of the vocal folds. The vibration of vocal folds produces periodic or quasi periodic excitations to the vocal tract for voiced speech whereas pure transient and/or turbulent noises are aperiodic excitations to the vocal tract for unvoiced speech [1].

This type of classification [2] finds other applications mainly in fundamental frequency estimation, formant extraction or syllable marking and so on. In fact, pitch detectors for speech signal can only work correctly if the fundamental frequency estimation is linked with a reliable voiced–unvoiced decision.

Moreover, the fundamental frequency is an important parameter in the speech analysis and synthesis. It plays an eminent role in the speech production and perception. In application areas such as speech enhancement, analysis and prosody modelling, low-bit rate coding, and speaker recognition, reliable pitch estimation is required [3].

A wide variety of sophisticated voicing classification and pitch detection algorithms (PDAs) have been proposed in the speech processing literature [4–8, 13].

Most voicing decision algorithms exploit almost any elementary speech signal parameter that may be computed independently of the type of input signal: energy, amplitude, short-term autocorrelation coefficients, zero-crossings count, ratio of signal amplitudes in different sub-bands or after pre-processing as, linear prediction error, or the salience of a pitch estimate. Voicing decision algorithms can be grouped into three essential categories: (1) simple threshold analysis algorithms, which exploit only a few basic parameters; (2) more-complex algorithms based on pattern recognition methods; and (3) integrated algorithms for both voicing and pitch determination.

Besides, the pitch estimation from the speech signal only is basically done by relying on different types of speech transformation. This transformation can be operated following three domains:

M. A. Ben Messaoud (✉) · A. Bouzid · N. Ellouze
Department of Electrical Engineering, Tunis El Manar University, ENIT, BP. 37 Le Belvédère, 1002 Tunis, Tunisia
e-mail: anouar.benmessaoud@yahoo.fr

A. Bouzid
e-mail: bouzidacha@yahoo.fr

N. Ellouze
e-mail: n.ellouze@enit.rnu.tn

The first approach works in the time domain. The common transformation is the autocorrelation function (ACF) like the YIN algorithm, the Praat Software application [9–12]. The second approach works in the frequency domain. The frequently used transformation is the spectrum [13, 14]. The third approach combines both time and frequency domains, using the short time Fourier transform (STFT) and the wavelet transform (WT) [15].

Although many PDAs were proposed, there is still no reliable algorithm that can be used for various speech processing applications. The difficulty of accurate and robust pitch estimation of speech is due to several reasons as the fast variation of the instantaneous pitch and formants.

In this paper, we detail and evaluate our improved algorithm called Multi-Scale Product Autocorrelation for voicing decision and fundamental frequency estimation from both clean and noisy speech.

The proposed algorithm was originally inspired by our works reported in [16, 17] where we used the speech multi-scale product spectrum (SMP) for pitch estimation and voicing decision.

The paper is presented as follows. After the introduction, we present the multi-scale product (MP) method used in this work to provide the derived speech signal. Section "Autocorrelation of the Speech Multi-Scale Product" introduces the multi-scale product autocorrelation (MPA) approach for the voicing detection and fundamental frequency estimation. In section "Voicing Decision and Pitch Estimation", we evaluate our approach and compare it to other well-known algorithms. Evaluation results are also presented for speech corrupted by real noise at various SNR levels.

Multi-Scale Product

The WT is a multi-scale analysis which has been shown to be very well suited for speech processing as glottal closure instant (GCI) detection, pitch estimation, speech enhancement and recognition and so on. Moreover, a speech signal can be analysed at specific scales corresponding to the range of human speech [18–21].

One of the most important WT applications is the signal singularity detection. Continuous WT produces modulus maxima at signal singularities allowing their localisation. However, one-scale analysis is not accurate. So, decision algorithm using multiple scales is proposed by different works to circumvent this problem [22, 23].

The MP is essentially introduced to improve signal edge detection. It is based on the multiplication of WTC at some scales. The non-linear combination of wavelet transform coefficients (WTC) attempts to enhance the peaks of the gradients caused by true edges, while suppressing the spurious peaks.

This method was first used in image processing. Xu et al. [24] rely on the variations in the WT decomposition level. They use multiplication of WT of the image at adjacent scales to distinguish important edges from noise. Sadler and Swami [25] have studied the MP method of a signal in presence of noise.

The choice of the mother wavelet is crucial to detect discontinuities. It depends essentially on the wavelet vanishing moment number and the wavelet support. The WT with n vanishing moments can be interpreted as a multi-scale differential operator of nth order of the smoothed signal. This provides a relationship between the differentiability of the signal and wavelet modulus maxima decay at fine scales.

It has been demonstrated that wavelet with n vanishing moments can be expressed as follows:

$$\Psi(t)=(-1)^n\frac{\mathrm{d}^n\theta(t)}{\mathrm{d}t^n} \tag{1}$$

where θ is a smoothing function. So, the WT of a function f can be written as:

$$Wf(u,s)=s^n\frac{\mathrm{d}^n}{\mathrm{d}u^n}(f^*\bar{\theta}_s)(u) \tag{2}$$

with

$$\bar{\theta}_s(t)=\frac{1}{\sqrt{s}}\theta\left(\frac{-t}{s}\right) \tag{3}$$

So if the wavelet is chosen to have one vanishing moment, modulus maxima appear at discontinuities of the signal and represent the maxima of the first derivative of the smoothed signal.

The MP [25] consists of making the product of WTC of the function $f(n)$ at some successive dyadic scales as follows:

$$p(n)=\prod_{j=j_0}^{j=j_L} w_{2^j}f(n) \tag{4}$$

where $w_{2^j}f(n)$is the WT of the function f at scale 2^j. The MP is taken at three levels to preserve the edge sign.

In this work, we are motivated by the MP use because it provides a derived speech signal which is simpler to be analysed. The Fig. 1 summarises the steps of the MP.

The voiced speech MP has a periodic structure with more reinforced singularities marked by extrema. It has a structure that reminds the derivative laryngograph signal. So, the autocorrelation function can be operated on the obtained signal.

Autocorrelation of the Speech Multi-Scale Product

We propose a new technique to determine voiced frames with an estimation of the fundamental frequency. The

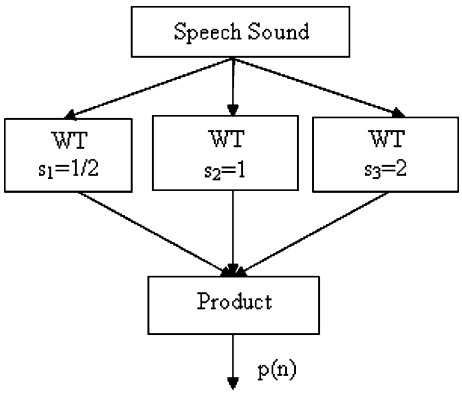

Fig. 1 Diagram of the speech multi-scale product

method is based on the autocorrelation analysis of the speech MP. It can be decomposed into three essential steps, as shown in Fig. 2. The first step consists of making the speech MP. Then, we decompose the obtained signal into overlapping frames. Each frame includes N samples and is weighted by a Hanning window $s_w(n)$, $n = 0, 1,\ldots, N-1$ ($N = 1{,}024$ samples with an overlapping of 512 points at a sampling frequency of 20 kHz). The wavelet used in this MP analysis is the quadratic spline function with a support of 0.8 ms at scales $s_1 = 2^{-1}$, $s_2 = 2^0$ and $s_3 = 2^1$. The second step consists of calculating the ACF of each frame extracted from the obtained signal. The third step consists of looking for the ACF maxima that are classified to make a voicing decision and then giving the fundamental frequency estimation for the voiced frames.

For the first step, the MP computing is detailed in the previous section. Then, the product $p[n]$ is divided into frames of N length by multiplication with a sliding analysis Hanning window $w[n]$:

$$P_{wi}[k] = p[k]w[k - iN/2] \tag{5}$$

where i is the window index, and $N/2$ is the overlap.

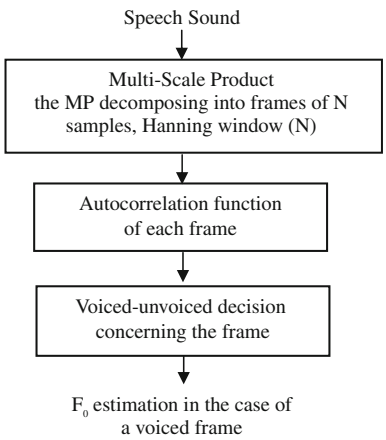

Fig. 2 Block diagram of the proposed approach for voiced/unvoiced decision and the fundamental frequency estimation

The weighting $w[n]$ is assumed to be non-zero in the interval $[0, N-1]$. The frame length N is chosen in such a way that, on the one hand, the parameters to be measured remain constant and, on the other hand, that there are enough samples of $p[n]$ within the frame to guarantee reliable frequency parameter determination.

The choice of the windowing function influences the values of the short-term parameters, the shorter the window, the greater is its influence [14].

In the second step, we compute the short-term autocorrelation function of each weighted block $p_{wi}[n]$ as follows:

$$\begin{aligned} R_i(k) &= \sum_{1=0}^{N-1} p_{wi}(1)p_{wi}(1+k) \\ \mathrm{ACF}_i(k) &= \frac{R_i(k)}{R_i(0)} \end{aligned} \tag{6}$$

The third step is detailed in the next section.

Voicing Decision and Pitch Estimation

After calculating the ACF of the speech MP in the ith frame, we store all the peak positions in the vector P_i corresponding to the frequencies. Peaks with very low value, below a fixed threshold T, are removed and T is fixed to $0.2 = \mathrm{Max(ACF)}/5$.

If there are no peaks, the frame is declared unvoiced, else we calculate the distance separating two successive peak positions $D_{ij} = P_{ij+1} - P_{ij}$ constituting the D_i vector elements. Where i is the frame index, j is the peak index ($j = 1, 2,\ldots, M$) and M is the peak number.

These elements are ranked in the growing order to compose the E_i vector. To make a voicing decision, we look for well-defined groups constituted from the E_{ij} set. The groups are sorted as follows:

If $E_{i1} - E_{i2} < S$, where S is the threshold chosen to be 12. So E_{i1} and E_{i2} belong to the same group G_{i1} and we calculate $E_{i1} - E_{i3}$, else, E_{i1} is in G_{i1} and E_{i2} is in G_{i2}. Then, we calculate $E_{i2} - E_{i3}$ and so on until reaching the last elements in the E_i vector. Once the groups are formed, we look for their number N_i. If $N_i = 1$, the ith frame is voiced, else, the frame is unvoiced.

Figure 3 shows a voiced speech signal followed by its MP. The MP has a periodic structure and reveals maxima corresponding to the glottal opening instant (GOI) and clear minima corresponding to the GCI. The Fig. 4 shows the autocorrelation function of the speech MP depicted in Fig. 3. The calculated function is obviously periodic and has the same period than the speech signal. Its first maximum at the non-zero index value corresponds to the pitch period.

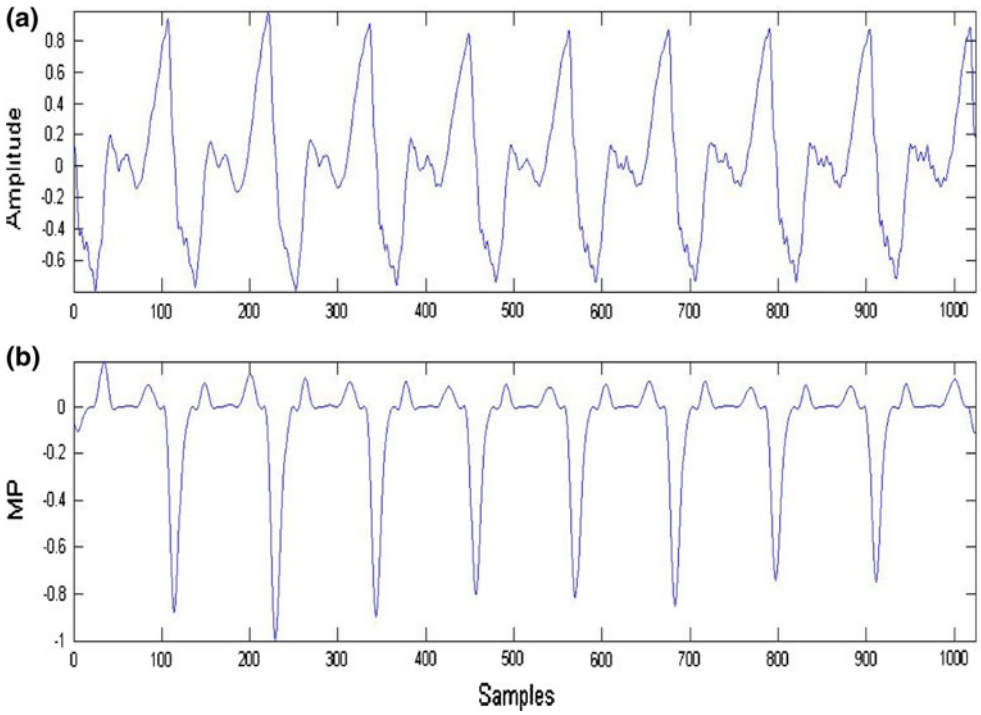

Fig. 3 **a** Voiced speech signal. **b** Its multi-scale product

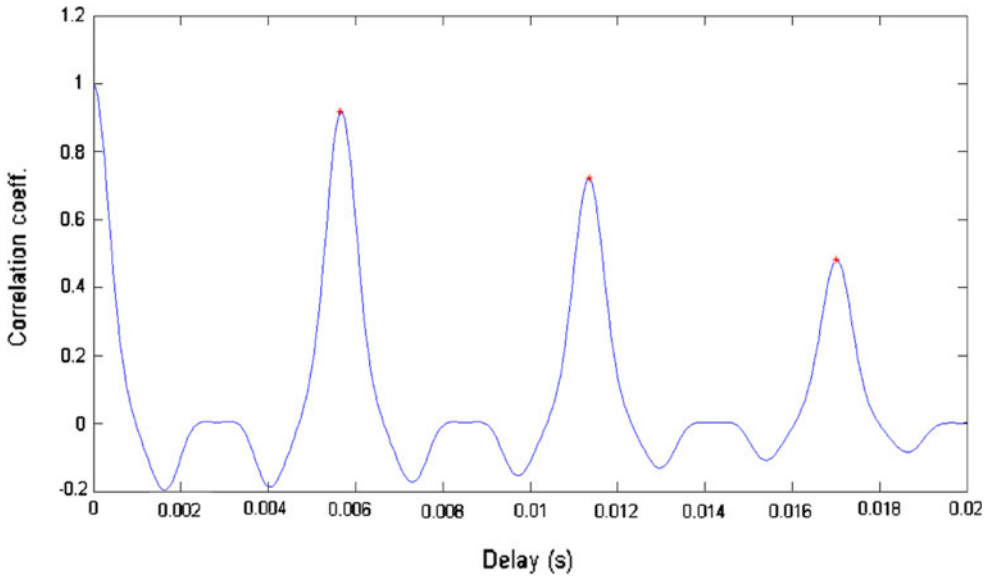

Fig. 4 Autocorrelation of the voiced speech multi-scale product

On the other hand, the Fig. 5 illustrates the MP of the unvoiced speech signal. The MP shows maxima and minima randomly separated.

Figure 6 illustrates the autocorrelation function of the unvoiced speech signal MP. This function shows extrema that are also randomly separated with weak amplitude. These two different behaviours (voiced and unvoiced cases) allow us to operate a voicing decision.

Now we try to underline the effect of the MP to reduce noise when added to a speech signal.

Figure 7 depicts a noisy voiced speech signal with an SNR of -5 dB followed by its MP. The MP lessens the noise effects leading to an autocorrelation function with clear maxima comparing to the one calculated directly on the noisy speech signal as shown in Fig. 8.

Evaluation

To evaluate the performance of our algorithm, we use the Keele pitch reference database [26, 27]. This database consists of speech signals of five male and five female English speakers each reading the same phonetically balanced text with varying duration between about 30 and

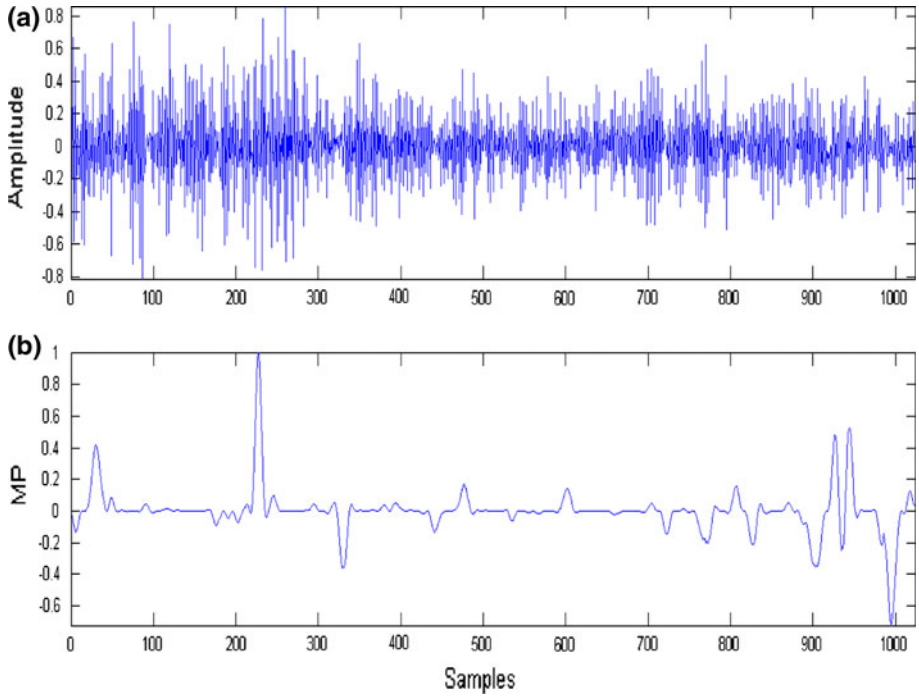

Fig. 5 **a** Unvoiced speech signal. **b** Its multi-scale product

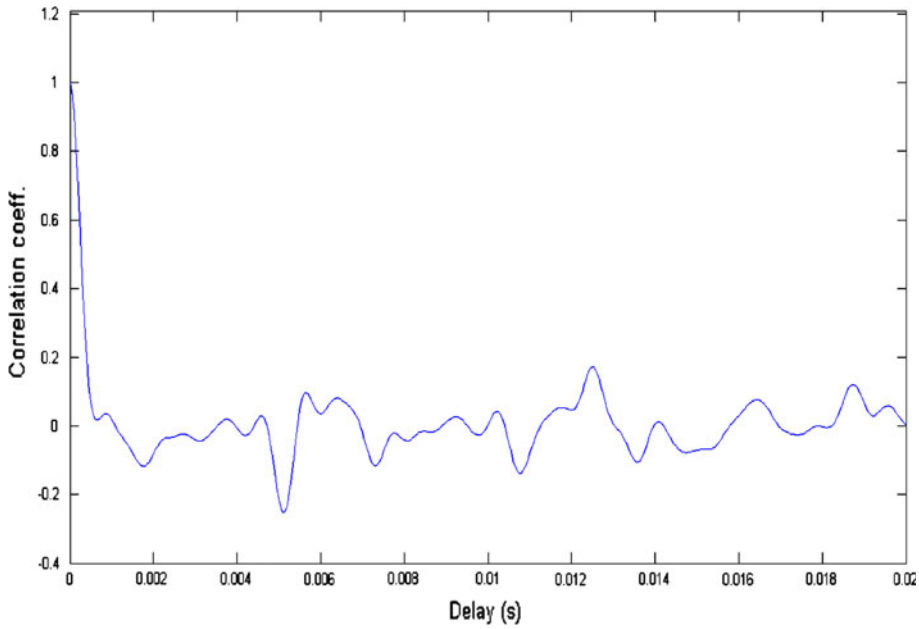

Fig. 6 Autocorrelation of the unvoiced speech multi-scale product

40 s. All the speech signals were sampled at a rate of 20 kHz. The Keele database includes reference files containing a voiced–unvoiced segmentation and a pitch estimation of 25.6 ms segments with 10 ms overlapping. The reference files also mark uncertain pitch and voicing decisions. The reference pitch estimation is based on a simultaneously recorded signal of a laryngograph. Unvoiced frames are indicated with zero pitch values, and negative values are used for uncertain frames.

The commonly used criteria for evaluating pitch estimation performance are the gross pitch error (GPE) and the root mean square error (RMS). A GPE is identified when the estimated fundamental frequency F_0 value is 20% higher or lower than the reference one. The RMS is computed as the root mean square difference in Hertz between the reference F_0 and the estimation for all frames having no GPE.

To evaluate a voicing decision algorithm, we calculate the V-UV error corresponding to the percentage of voiced

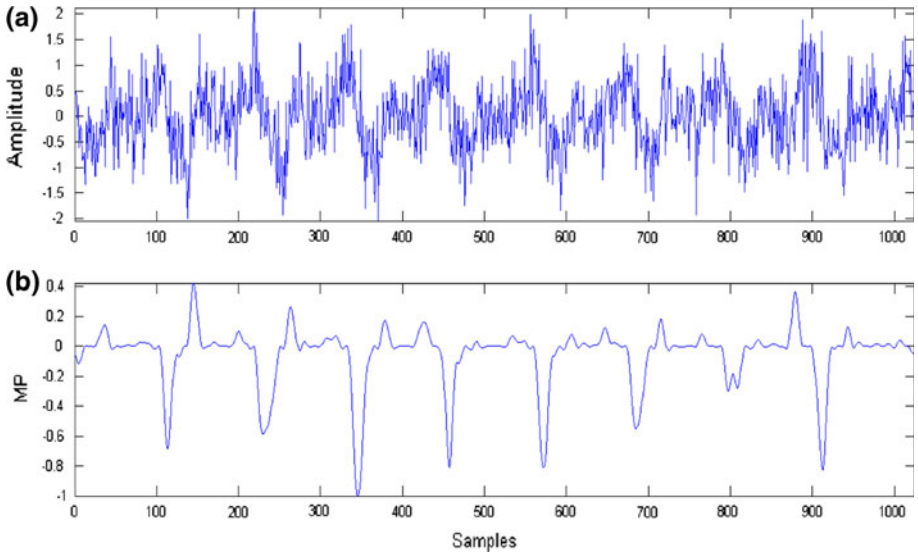

Fig. 7 **a** Voiced and noisy speech signal (SNR = −5 dB). **b** Its multi-scale product

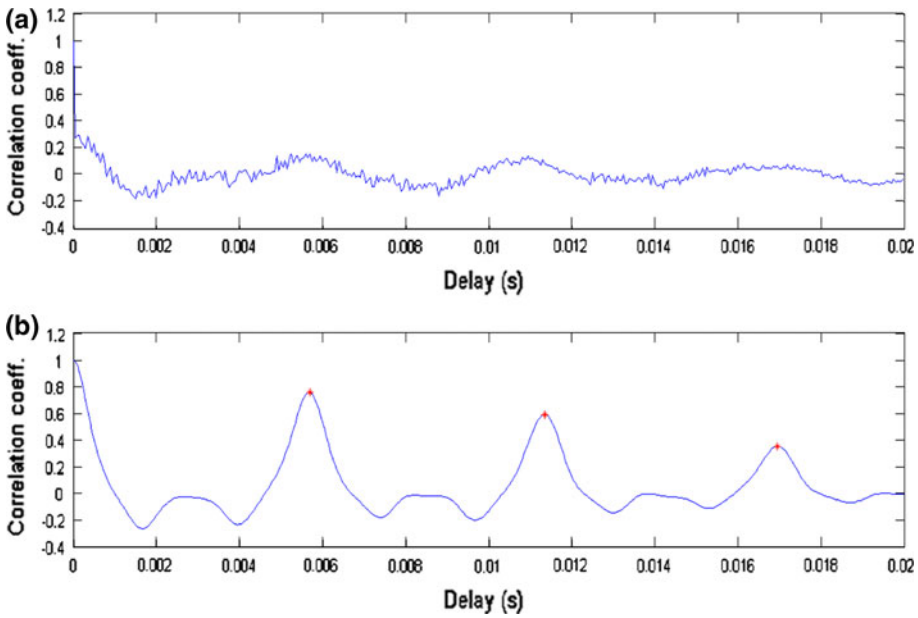

Fig. 8 **a** Autocorrelation of the noisy voiced speech. **b** Autocorrelation of the noisy voiced speech multi-scale product

frames misclassified as unvoiced, and the UV-V error defined as the unvoiced frames considered as voiced, it is about the rate of false alarms.

Evaluation in a Clean Environment

Table 1 reports evaluation results for voicing classification of the proposed method in a clean environment. We compare our method to other state-of-the-art algorithms [8, 17, 28–30] that are based on the same reference database.

As can be seen, our method yields interesting performance in comparison with well-known approaches with the lowest V-UV and UV-V rates of 1.8 and 2.7%, respectively. Moreover, the autocorrelation of the speech MP outperforms our previous proposed approach SMP.

Table 1 Voicing decision performances in a clean environment

Method	V-UV (%)	UV-V (%)
MPA	*1.8*	*2.7*
SMP [17]	2.3	3.8
RAPT [8]	3.2	6.8
NMF [29]	7.7	4.6
MLS [30]	7.0	7.9
NMF-HMM-PI [28]	8.4	8.8

Table 2 presents the evaluation results of the proposed algorithm (MPA) for pitch estimation in a clean environment and compared with the other state-of-the-art algorithms [8, 11, 12, 17, 28–31].

Table 2 Evaluation results of the MPA algorithm and others for pitch estimation in a clean environment

Method	GPE (%)	RMS (Hz)
MPA	*0.61*	*1.72*
SMP [17]	0.75	2.41
NMF-HMM-PI [28]	1.06	3.7
NMF [29]	0.9	4.3
MLS [30]	1.5	4.68
Yin [11]	2.35	3.62
RAPT [8]	2.2	4.4
PRAAT [12]	3.1	4.5
RCEPS [31]	3.66	3.12

The MPA shows a reduced GPE rate of 0.61% and an interesting RMS of 1.72 Hz. It is obviously more accurate than the SMP that has 0.75% of GPE rate and a RMS of 2.41 Hz.

Evaluation in a Noisy Environment

To test the robustness of our algorithm, we add various background noises (white, babble and vehicle) at four SNR levels to the Keele database speech signals. The noise is taken from the noisex-92 database [32].

Table 3 presents evaluation results for voicing decision of the proposed method in a noisy environment.

As reported in Table 3, when decreasing the SNR level, the performances of the proposed approach decrease but remain robust and more performing than the SMP and NMF-HMM-PI methods.

Table 4 illustrates the GPE of the proposed approach, the SMP [17], the PRAAT [12], the YIN [11], the RCEPS [31] and the NMF-HMM-PI [28] in a noisy environment. As depicted in Table 4, when the SNR level decreases, the MPA algorithm remains robust even at −5 dB and appears

Table 3 Performance comparison of the MPA algorithm and others for voicing decision in a noisy environment

Type of noise	SNR level	V-UV (%)			UV-V (%)		
		MPA	SMP [17]	NMF-HMM-PI [28]	*MPA*	SMP [17]	NMF-HMM-PI [28]
White	10 dB	*5.32*	5.37	11.6	*2.67*	3.43	2.97
	5 dB	*5.4*	5.8	15.4	*3.1*	4.1	3.8
	0 dB	*7.38*	7.73	25.2	*5.2*	5.9	3.2
	−5 dB	*9.72*	10.2	38.45	*6.9*	7.6	5.43
Babble	10 dB	*5.18*	5.89	10.43	*3.02*	4.52	3.56
	5 dB	*5.79*	6.78	12.82	*4.33*	5.23	4.02
	0 dB	*6.76*	7.33	19.53	*4.89*	5.98	5.28
	−5 dB	*8.44*	9.84	21.37	*5.23*	7.52	5.75
Vehicle	10 dB	*6.33*	8.89	9.65	*4.49*	6.25	3.98
	5 dB	*9.44*	10.55	18.26	*5.55*	8.21	4.23
	0 dB	*11.13*	12.43	24.78	*6.71*	10.44	5.28
	−5 dB	*13.26*	15.98	32.55	*7.33*	11.18	9.18

Table 4 GPE rate for some pitch estimation algorithms in a noisy environment

Type of noise	SNR level	GPE (%)					
		MPA	SMP [17]	PRAAT [12]	YIN [11]	RCEPS [31]	NMF-HMM-PI [28]
White	10 dB	*0.78*	0.92	3.65	3.15	4.04	1
	5 dB	*0.87*	1	4.6	3.9	5.52	1.2
	0 dB	*1.1*	1.2	6.1	5.1	8.43	1.28
	−5 dB	*1.3*	1.4	6.2	5.9	11.43	1.43
Babble	10 dB	*0.92*	1.53	9.24	7.35	10.03	1.36
	5 dB	*1.23*	2.61	16.15	15.74	18.65	1.97
	0 dB	*2.86*	4.56	29.08	31.85	32.64	3.21
	−5 dB	*3.45*	7.62	45.11	48.71	42.56	5.68
Vehicle	10 dB	*3.28*	4.79	6.50	4.45	8.43	3.55
	5 dB	*4.26*	6.41	9.88	7.89	11.45	4.16
	0 dB	*4.75*	7.04	17.68	14.24	12.28	5.3
	−5 dB	*5.25*	8.98	32.56	27.95	19.46	9.58

Table 5 RMS (in Hz) for different pitch estimation algorithms in a noisy environment

Type of noise	SNR level	RMS (Hz)					
		MPA	SMP [17]	PRAAT [12]	YIN [11]	RCEPS [31]	NMF-HMM-PI [28]
White	10 dB	*2.14*	3.06	4.81	3.71	3.75	3.9
	5 dB	*2.6*	3.23	5.13	4.13	3.98	4.5
	0 dB	*3.1*	3.73	5.72	5.75	4.23	4.9
	−5 dB	*4.2*	4.67	6.44	7.82	4.77	5.04
Babble	10 dB	*2.88*	3.57	5.33	3.96	3.78	3.71
	5 dB	*3.42*	4.28	5.84	4.78	4.42	4.34
	0 dB	*4.21*	4.93	6.78	5.45	5.58	5.26
	−5 dB	*4.87*	6.38	9.42	7.84	6.15	6.78
Vehicle	10 dB	*2.56*	4.56	5.87	3.67	4.78	3.52
	5 dB	*3.42*	5.67	8.12	4.56	5.89	4.43
	0 dB	*3.56*	7.89	12.33	5.67	6.31	4.87
	−5 dB	*5.93*	11.57	14.67	7.81	8.57	6.76

as the most efficient approach for pitch estimation. The SMP method has a greater GPE than the NMF-HMM-PI in the case of babble and vehicle noises.

Besides, the MPA method presents the lowest RMS values showing its convenience for pitch estimation in hard situations.

As depicted in Table 5, when the SNR level decreases, the MPA algorithm remains reliable even at −5 dB and appears as the most accurate approach for pitch estimation.

Moreover, the voiced/unvoiced decision and the pitch estimation accuracy are closely related to the threshold T. We have studied the GPE rate versus the T value and we find as depicted in Fig. 9 that for the used $T = 0.2 = \text{Max(ACF)}/5$, the GPE rate is the lowest.

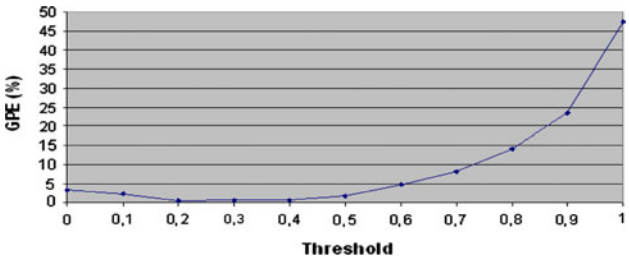

Fig. 9 GPE rate variation versus the threshold T

Conclusion

In this paper, we present a voicing classification and pitch estimation method that relies on the autocorrelation analysis of the speech multi-scale product. The proposed approach can be summarised in three essential steps. First, we compute the product of the speech WTC at three successive dyadic scales. The obtained signal is divided into frames of 1,024 samples, and each frame is weighted by a Hanning window having the same length. Second, we calculate the autocorrelation function of each weighted frame. Thirdly, we detect the peaks given by this function. A peak classification is operated respecting some defined rules to be able to make a voiced/unvoiced decision concerning the frame. For voiced frame, the pitch period is the index non-zero corresponding to the first maximum. The fundamental frequency can be estimated as the inverse of the pitch period.

The experimental results show the efficiency of our proposed approach for voicing detection and pitch estimation from clean speech, and its robustness in the noisy environment compared with the state-of-the-art algorithms. The MPA approach outperforms the cited algorithms in this work not only for voiced/unvoiced decision but also for pitch estimation.

Future work concerns the extension of the proposed approach to the multi-pitch estimation.

Acknowledgments The authors are very grateful to Alain de Cheveigné for providing the fundamental frequency estimation software (Yin algorithm).

References

1. Qi Y, Hunt BR. Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier. IEEE Trans Speech Audio Process. 1993;1(2):250–6.
2. Martin A, Charlet D, Mauuary L. Robust speech/non-speech detection using LDA applied to MFCC. IEEE Int Conf Acoust Speech Signal Process. 2001;1:237–40.
3. Shaughnessy DO. Speech communications: human and machine. 2nd ed. Piscataway, NJ: IEEE Press; 1999.
4. Childers DG, Hahn M, Larar JN. Silent and voiced/unvoiced/mixed excitation classification of speech. IEEE Trans Acoust Speech Signal Process. 1989;37(11):1771–4.
5. Liao L, Gregory M. Algorithms for speech classification. IEEE Int Conf Signal Process Appl. 1999;2:623–7.

6. Hess W. Pitch determination of speech signals: algorithms and devices. New York: Springer; 1983.
7. Bagshaw PC, Hiller SM, Jack MA. Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching. In: The 3rd European conference on speech communication and technology; 1993.
8. Talkin D. A robust algorithm for pitch tracking. In: Kleijn WB, Paliwal KK, editors. Speech coding and synthesis. Amsterdam: Elsevier; 1995. p. 495–518.
9. Rabiner L. On the use of autocorrelation analysis for pitch detection. IEEE Trans Acoust Speech Signal Process. 1977;25(1): 24–33.
10. Krubsack DA, Niederjohn RJ. An autocorrelation pitch detector and voicing decision with confidence measures developed for noise-corrupted speech. IEEE Trans Signal Process. 1991; 39(2):319–29.
11. De Cheveigné A, Kawahara H. YIN, a fundamental frequency estimator for speech and music. J Acoust Soc Amer. 2002;111(4): 1917–30.
12. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc Inst Phon Sci. 1993;17:97–110.
13. Noll AM. Cepstrum pitch determination. J Acoust Soc Amer. 1967;41(2):293–309.
14. Shimamura T, Takagi H. Noise-robust fundamental frequency extraction method based on exponentiated band-limited amplitude spectrum. IEEE Int Conf Midwest Symposium on Circuits and Systems. 2004;47(2):141–4.
15. Shahnaz C, Zhu WP, Ahmad MO. A spectro-temporal algorithm for pitch frequency estimation from noisy observations. In: IEEE international symposium on circuits and systems. Seattle, WA; 2008. p. 1704–7.
16. Ben Messaoud MA, Bouzid A, Ellouze N. Spectral multi-scale product analysis for pitch estimation from noisy speech signal. In: Solé-Casals J, Zaiats V, editors. Advances on non-linear speech processing, International conference on non-linear speech processing, NOLISP'09, LNAI, vol. 5933. Berlin: Springer; 2010. p. 95–102.
17. Ben Messaoud MA, Bouzid A, Ellouze N. A new method for pitch tracking and voicing decision based on spectral multi-scale analysis. Signal Process: An Int J. 2009;3(5):144–9.
18. Burrus CS, Gopinath RA, Guo H. Introduction to wavelets and wavelet transforms: a primer. Englewood Cliffs: Prentice Hall; 1998.
19. Mallat S. A wavelet tour of signal processing: the sparse way. 3rd ed. Burlington, VT: Academic Press; 2008.
20. Berman Z, Baras JS. Properties of the multiscale maxima and zero-crossings representations. IEEE Trans Signal Process. 1993;41(12):3216–31.
21. Kadambe S, Boudreaux-Bartels GF. Application of the wavelet transform for pitch detection of speech signals. IEEE Trans Inf Theory. 1992;38(2):917–8.
22. Bouzid A, Ellouze N. Voice source parameter measurement based on multi-scale analysis of electroglottographic signal. Speech Commun. 2009;51(9):782–92.
23. Bouzid A, Ellouze N. Open quotient measurements based on multiscale product of speech signal wavelet transform. New York: Hindawi Publishing Corp, Res Lett Signal Process; 2007. p. 1–6.
24. Xu Y, Weaver JB, Healy DM, Lu J. Wavelet transform domain filters: a spatially selective noise filtration technique. IEEE Trans Image Process. 1994;3(6):747–58.
25. Sadler BM, Swami A. Analysis of multi-scale products for step detection and estimation. IEEE Trans Inf Theory. 1999;45(3): 1043–9.
26. Meyer G, Plante F, Ainsworth WA. A pitch extraction reference database. The 4th European conference on speech communication and technology, EUROSPEECH. Madrid, Spain; 1995. p. 837–40.
27. Keele Pitch Database. In: Psychology Home page-human machine perception. University of Liverpool. 1995. http://www.liv.ac.uk/Psychology/hmp/projects/pitch/speech/keele_pitch_database.html. Accessed 24 April 2010.
28. Joho D, Bennewitz M, Behnke S. Pitch estimation using models of voiced speech on three levels. IEEE Int Conf Acoust Speech Signal Process. 2007;4:1077–80.
29. Sha F, Saul LK. Real time pitch determination of one or more voices by nonnegative matrix factorization. In: Saul LK, Weiss Y, Bottou L, editors. Advances in neural information processing systems. Cambridge: MIT Press; 2005. p. 1233–40.
30. Sha F, Burgoyne JA, Saul LK. Multiband statistical learning for f0 estimation in speech. IEEE Int Conf Acoust Speech Signal process. 2004;5:661–4.
31. Nakatani T, Irino T. Robust and accurate fundamental frequency estimation based on dominant harmonic components. J Acoust Soc Amer. 2004;116(6):3690–700.
32. Noisex92. In: Signal Processing Information Base (SPIB). The Signal Processing Society and the National Science Foundation. 2007. http://spib.rice.edu/spib/select_noise.html. Accessed 24 April 2010.