

## Research Letter

# Open Quotient Measurements Based on Multiscale Product of Speech Signal Wavelet Transform

Aïcha Bouzid<sup>1</sup> and Nouredine Ellouze<sup>2</sup>

<sup>1</sup> *Institut Supérieur d'Electronique et de Communication de Sfax, Route Menzel Chaker, km 0.5, BP 868, Sfax 3018, Tunisia*

<sup>2</sup> *Department of Electric Engineering, National School of Engineering of Tunis, BP 37, Le Belvédère, Tunis 1002, Tunisia*

Correspondence should be addressed to Aïcha Bouzid, bouzidacha@yahoo.fr

Received 20 July 2007; Accepted 17 December 2007

Recommended by Yannis Stylianou

This paper describes a multiscale product method (MPM) for open quotient measure in voiced speech. The method is based on determining the glottal closing and opening instants. The proposed approach consists of making the products of wavelet transform of speech signal at different scales in order to enhance the edge detection and parameter estimation. We show that the proposed method is effective and robust for detecting speech singularity. Accurate estimation of glottal closing instants (GCIs) and opening instants (GOIs) is important in a wide range of speech processing tasks. In this paper, accurate estimation of GCIs and GOIs is used to measure the local open quotient (Oq) which is the ratio of the open time by the pitch period. Multiscale product operates automatically on speech signal; the reference electroglottogram (EGG) signal is used for performance evaluation. The ratio of good GCI detection is 95.5% and that of GOI is 76%. The pitch period relative error is 2.6% and the open phase relative error is 5.6%. The relative error measured on open quotient reaches 3% for the whole Keele database.

Copyright © 2007 A. Bouzid and N. Ellouze. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Voiced speech is produced through the excitation of the vocal tract with quasiperiodic vibrations of the vocal folds at the glottis constituting the glottal volume velocity waveform or the glottal airflow. These vibrations produce the glottis opening and closing within each pitch period. The glottal closing is related to the maximum excitation energy, enabling the glottal pulse to exhibit a maximum magnitude, which after being distorted and delayed by the vocal tract can be visible in the speech waveform. The glottal closing instant (GCI) is a significant moment in a pitch period, and it is therefore treated as an event or epoch [1, 2]. This marks the beginning of the closed phase that can be essential in speaker verification, speech recognition, speech synthesis and modification, and low-bit-rate coder [3].

An important parameter of the glottal flow is the open quotient (Oq) defined as the ratio of the duration of the open phase to that of the pitch period. A change of the Oq substantially changes the spectrum of the excitation; it is also highly correlated to physiological constraints, as is the case in different phonation types. Accurate measurement of the

open quotient may require precise opening and closing instants. These instants can be located by filming the vocal fold vibration with a high-speed motion camera. Although this method was useful in quantifying the timing characteristics of vocal fold vibration, the process has proved uncomfortable, time-consuming, and expensive [4]. Inverse filtering of speech signal and electroglottography (EGG) also provide information about vocal fold motion avoiding the pitfalls of high-speed motion filming.

Over the last decades, attempts have been made to locate the GCI, but little work has been done in the detection of the GOI. These approaches can be classified as single or dual channel analysis. Previous methods had mostly focused on the dual channel approach. In those methods, the laryngograph is used to detect the closed phase by locating the GOI. In each period, the derivative of EGG signal (DEGG) shows two opposite peaks. The stronger peak corresponds to the GCI and the weaker one corresponds to the GOI [5]. However, the technique is not appropriate for speech analysis outside laboratory conditions. Single channel analysis uses only the speech signal to locate the GCI and GOI. Most of the single channel methods that rely on

using the speech signal alone have proved unreliable in locating the closed phase. Because of the difficulty in locating the GOI, many of these techniques merely rely on estimating the GCI and assuming that an ad hoc choice of post-GCI interval length will lie within the closed phase. These lengths are generally chosen to be either a fixed constant length (e.g., 2 milliseconds) or a percentage of the pitch period (e.g., 30%). The work presented in [6] outlines a method used for automatic closed phase location by excluding the intervals that are not included in the closed phase; the indicator used is the log determinant of the Kalman filter which estimates error covariance matrix. Recent work has evaluated the DYPSA algorithm for GCI detection using APLAWD and SAM databases [7]. In [8], open quotient is estimated by second-order linear prediction of the inverse filtered speech signal. The estimation is presented with rough precision.

In this paper, a new approach for open quotient estimation is put forward. The method uses the detected GCI and GOI from the speech signal, obtained as the product of wavelet transform coefficients at three dyadic scales. The open quotient ratio is derived from the estimation of the pitch period and the corresponding opening time obtained from GCI and GOI.

The paper is organized as follows. In Section 2, we present the method of multiscale product of wavelet transforms. In Section 3, we evaluate the multiscale product method for GCI and GOI detection. In Section 4, we determine the pitch period from the detected GCIs. The open time interval, derived from successive GOI and GCI, is divided by the pitch period to calculate the open quotient. Finally, in Section 5 we give the conclusion.

## 2. GCI AND GOI DETECTION FROM SPEECH

According to Mallat [9], the wavelet transforms have shown excellent capacities for the detection of singularities in signals. At the glottal closing instant, the glottal pulse exhibits a peak which can be regarded as transient phenomenon carrying information about the vibration of the vocal folds. Motivated by the use of wavelet transform to detect and characterize signal singularities, various works have developed algorithms based on wavelet transform to locate GCI. However, one-scale analysis does not give a good precision for GCI detection. So, decision algorithm using multiple scales is proposed by different works to circumvent this problem [10–12]. These works confirm that wavelet transform modulus maxima coincide in most cases with GCI, which is given by the derivative minima of EGG signal. Besides, the methods developed in [10, 11] are not able to detect the GOI. In [12], we propose a new approach based on multiscale analysis for GOI detection; however, for a better localization, the used wavelet needs to be modified. In [13], we have shown that opening instant on EGG signal is less abrupt than the closing instant. Thus, the glottal flow presents approximately the same behavior in the neighborhood of these instants. These singularities are smoothed by the effect of the vocal tract causing the speech signal to present smoothed singularities at these instants. It is then more difficult to detect them especially at

GOI. Thus, the GOI localization is generally more challenging than that of GCI.

In this paper, we propose a novel approach in speech signal analysis. This method is used in image processing to improve edge detection. This approach is based on the multiplication of wavelet transform coefficients at some scales. It attempts to enhance the peaks of the gradients caused by true edges, while suppressing false peaks caused by noise. Xu et al. [14] rely on the variations in scale of the wavelet transform. They use multiplication of wavelet transform of the image at adjacent scales to distinguish important edges from noise. Sadler and Swami [15] have studied multiscale product method of signal in presence of noise. In wavelet domain, it is well known that edge structures are present at each subband, while noise decreases rapidly along the scales. It has been observed that multiplying the adjacent scales could sharpen edges while smoothing out noise [14, 15].

Consider a multiscale analysis of a function  $f(n)$  by generating the product of the wavelet transforms at some dyadic adjacent scales given by

$$p(n) = \prod_j W_{2^j} f(n). \quad (1)$$

Here,  $W_{2^j} f(n)$  represents the wavelet transform of the function  $f(n)$  at scale  $2^j$  given by the following equation:

$$W_{2^j} f(n) = f * \bar{\psi}_{2^j}(n), \quad (2)$$

with

$$\bar{\psi}_{2^j}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{-t}{2^j}\right). \quad (3)$$

The expression given by (1) is distinctly a nonlinear function. The product  $p(n)$  reveals peaks at signal edges and has relatively small values elsewhere. Although particular smoothing levels may not be optimal, the nonlinear combination tends to reinforce the peaks while suppressing spurious peaks.

The signal peaks will align across scales for the first few scales. This is because increasing the amount of smoothing will spread the response and cause singularities separated in time to interact. Thus, choosing very large scales will result in misaligned peaks in the product. An odd number of terms preserve the sign of the edge in the product. The efficiency of the multiscale product method (MPM) to improve the edge detection has encouraged us to apply it on speech signals to outperform the wavelet transform precision in weak singularity detection as GOI.

The proposed multiscale analysis is operated with a quadratic spline wavelet whose spectrum  $\hat{\psi}(w)$  is given by

$$\hat{\psi}(w) = iw \left( \frac{\sin(w/4)}{w/4} \right)^4. \quad (4)$$

At the top of Figure 1, we show the cubic spline smoothing function followed by its quadratic spline wavelet derivative. Wavelet transforms are operated for the following dyadic scales 1/2, 1, and 2 corresponding to the respective  $j$  values  $-1, 0, \text{ and } 1$ .

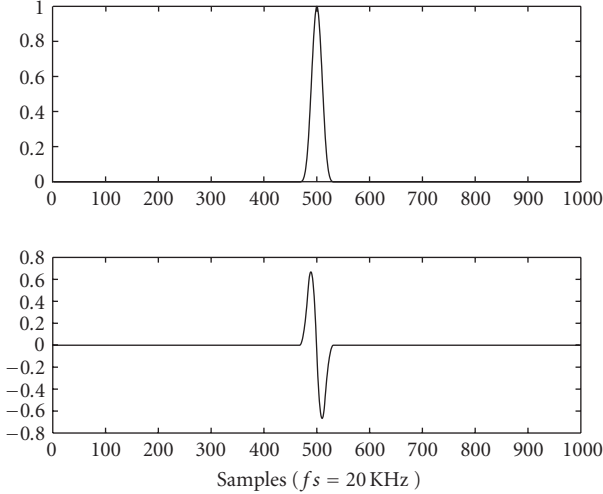


FIGURE 1: A cubic spline smoothing function (at the top) followed by its quadratic spline wavelet derivative.

Figure 2 depicts a vowel /o/ uttered by a female speaker followed by its wavelet transforms at three successive dyadic scales, the multiscale product and finally the DEGG. This figure depicts a simple case of a vowel without noise where GCI and GOI given by the MP match perfectly with GCI and GOI given by the DEGG. We note in the cross-scale product two types of peaks; minima corresponding to GCIs are the most distinguishable, and maxima related to GOIs are considerably weaker but discernible. The GCI is detected using the MATLAB codes of “findpeaks” file for minima detection available in [16]. The GOI is the maximum detected in the interval  $[t_1, t_2]$ , where  $t_1 = 0, 8 \text{ GCI}(k) + 0, 2 \text{ GCI}(k + 1)$  and  $t_2 = \text{GCI}(k + 1)$ .

Figure 3 depicts a complex case corresponding to a voiced fricative /z/ being naturally noisy. We can clearly see the effect of the product on suppressing the additional noise peaks, and consequently the best detection of GOI. We note the efficiency of the MPM to strengthen the GCI and its ability to detect GOIs that become clear. A block diagram describing the GCI and GOI detection is given by Figure 4.

### 3. EVALUATION RESULTS

To experiment with the proposed method, we use the Keele University database. Estimation of GCIs and GOIs is operated for voiced speech signal. Keele database includes both acoustic speech and laryngograph signals. Five adult female and five adult male speakers are recorded in low ambient noise using a sound proof room. Each utterance consists of the same phonetically balanced English text. In each case, the acoustic and laryngograph signals are time-synchronized, and they share the same sampling rate value of 20 KHz [17]. The EGG signal proves to be the best estimate of the vocal cord activity, because it is not modified by the vocal tract. The glottal closing presents a discontinuity in the kinetics of the vocal cord obtained by the DEGG that is used as a reference in all vocal experiments. Based on studies on pitch detection,

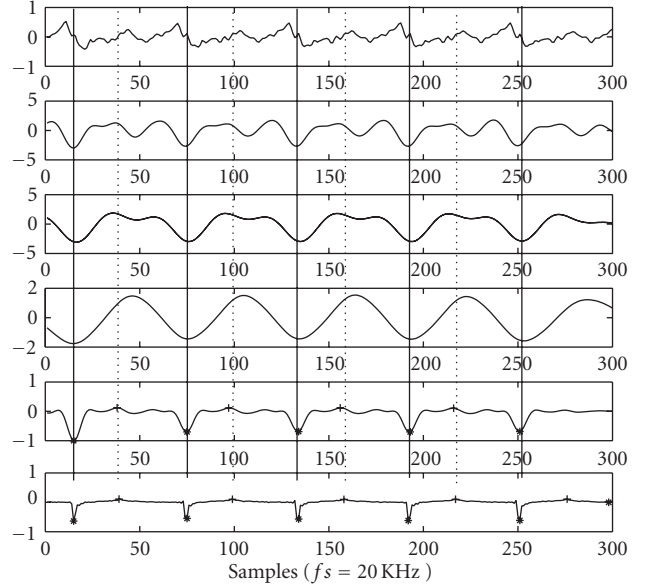


FIGURE 2: GOI and GCI detection for a vowel /o/ uttered by a female speaker. At the top, speech signal followed by its wavelet transforms at scales 1/2, 1, and 2, the multiscale product, and the DEGG signal. Solid lines indicate GCI and dotted ones indicate GOI.

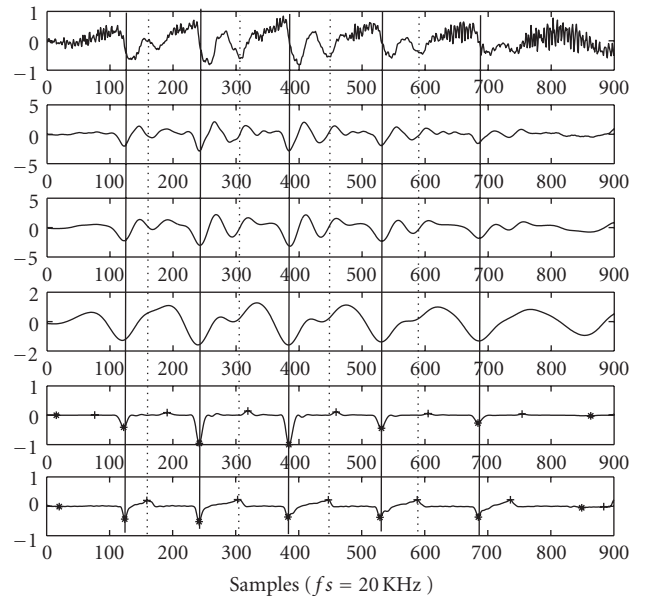


FIGURE 3: GOI and GCI detection for a voiced fricative /z/ uttered by a male speaker. At the top, speech signal followed by its wavelet transforms at scales 1/2, 1, and 2, the multiscale product, and the DEGG signal. Solid lines indicate GCI and dotted ones indicate GOI.

the evaluation uses four criteria: the number of missing measures, the number of gross errors, the number of fine errors, and the number of good detection processes. Referring to Krishnamurthy and Childers, the DEGG signal gives the glottal closing instant with an accuracy of two sample points at 10 KHz [18]. Tolerance of 5 samples is generally accepted at

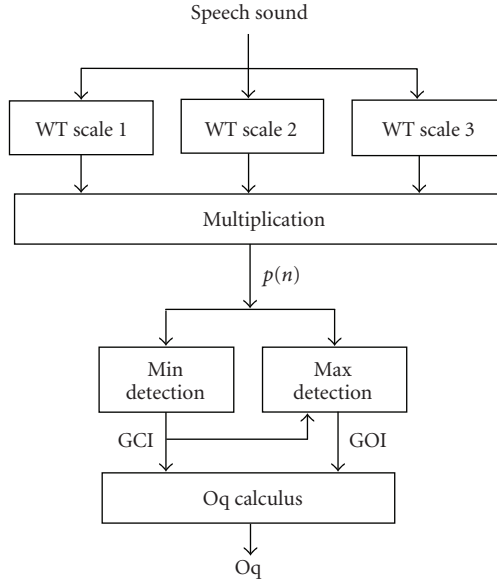


FIGURE 4: Block diagram describing the process to get the GCI, the GOI, and the open quotient from the multiscale product  $p(n)$ .

TABLE 1: GCI and GOI detection performances on Keele database.

	GD %	FE %	GE %	MR %
GCI	95.5	2.1	0.9	1.5
GOI	76	9.5	4.5	10

20 KHz for good detection (GD). As for the limit between gross error (GE) and fine error (FE), we take 1 millisecond according to Rabiner et al. [19]. The speech and EGG signals are time-aligned to compensate for the observed larynx-to-microphone time delay.

Results reported in Table 1 show that the multiscale product method presents rates of good detection reaching 95.5% for GCIs and 76% for GOIs, with an accuracy of 0.25 millisecond, determined on the Keele database. The auditory model and inverse filtering have been also applied on Keele database; they give less GCI detection performance than MPM.

#### 4. OPEN QUOTIENT MEASURES

The open quotient (Oq) is the ratio of the duration of the open glottal phase to the duration of the pitch period. Pitch period is the time separating two successive GCIs. Hence, the  $k$ th pitch period of the speech signals is given by the following expression:  $T(k) = \text{GCI}(k+1) - \text{GCI}(k)$ . The measures of pitch period  $T(k)$  are obtained with a relative error of 2.6% for the entire Keele database, corresponding to an RMS of 0.21 millisecond. The duration of the open phase is given by the following expression:  $TO(k) = \text{GCI}(k+1) - \text{GOI}(k)$ . The open phase is obtained with a relative error of 5.6% corresponding to an RMS value of 0.27 mil-

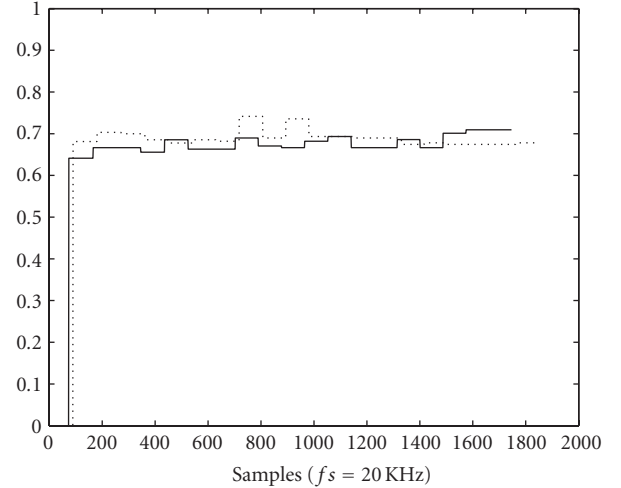


FIGURE 5: Oq estimation from speech signal in dotted line and EGG signal in solid line of vowel /o/ of speaker f1.

lisecond. Given the pitch period  $T(k)$  and the open phase  $TO(k)$ , the open quotient is given by the following equation:  $Oq = TO(k)/T(k)$ . The open quotient is obtained with a relative error of 3% in the case of no GCI or missing GOI.

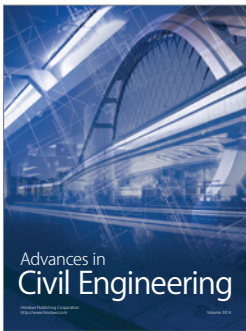
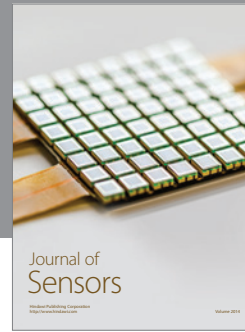
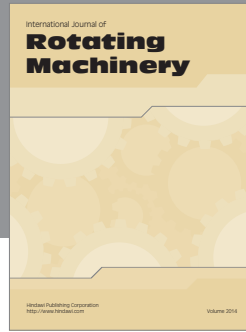
Figure 5 gives an example (utterance of vowel /o/ of speaker f1) of comparative measures of Oq parameter determined by the multiscale product of speech signal and the DEGG signal.

#### 5. CONCLUSIONS

In this paper, we propose a method for estimating glottal closing and glottal opening instants from speech signal. This method is based on the detection of minima and maxima given by the product of wavelet transform coefficients of different scales. These minima and maxima correspond, respectively, to GCIs and GOIs. MP is obtained by the product of three successive scale wavelet coefficients using the quadratic spline wavelet. The obtained performance is up to 95.5% for GCI detection with an accuracy of 0.25 millisecond, determined on the Keele database. However, 76% of the GOIs are identified with the same accuracy. The described method for GCI detection outperforms other methods in the literature such as inverse filtering and auditory model. So, significant enhancement in performance is given by the MPM for identifying GCIs and GOIs in voiced speech. Furthermore, it offers a good measure of the pitch period, obtained as the distance between two successive GCIs with a relative error of 2.6%. It also offers a measure of the open phase, defined as the distance between the GOI and the next GCI with a relative error of 5.6%. Finally, the open quotient which is the ratio of the open phase to the pitch period is obtained with a good relative error of 3% determined on the whole Keele database.

## REFERENCES

- [1] I. Gavati, M. Zirra, and V. Enescu, "Pitch detection of speech by dyadic wavelet transform," in *Proceedings of the International Conference on Signal Processing Applications and Technology (ICSPAT '96)*, pp. 1630–1634, Boston, Mass, USA, October 1996.
- [2] Y. M. Cheng and D. O'Shaughnessy, "Automatic and reliable estimation of glottal closure instant and period," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1805–1815, 1989.
- [3] M. D. Plumpe, T. F. Quatieri, and D. A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 5, pp. 569–586, 1999.
- [4] C. M. Sapienza, E. T. Stathopoulos, and C. Dromey, "Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: effects of measurement criteria and sound pressure level," *Journal of Voice*, vol. 12, no. 1, pp. 31–43, 1998.
- [5] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo, "On the use of the derivative of electroglottographic signals for characterization of non pathological phonation," *Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, 2004.
- [6] J. G. McKenna, "Automatic glottal closed-phase location and analysis by Kalman filtering," in *Proceedings of the 4th ISCA Tutorial and Research Workshop on Speech Synthesis*, pp. 91–96, Pittlochrie, Scotland, August-September 2001.
- [7] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 34–43, 2007.
- [8] N. Henrich, B. Doval, and C. d'Alessandro, "Glottal open quotient estimation using linear prediction," in *Proceedings of the Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, September 1999.
- [9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 2nd edition, 1999.
- [10] T. Vu Ngoc and C. d'Alessandro, "Robust glottal closing detection using wavelet transform," in *Proceedings of the 6th European Conference on Speech Communication and Technology (EUROSPEECH '99)*, pp. 805–808, Budapest, Hungary, September 1999.
- [11] L. Janer, J. J. Bonet, and E. Lleida-Solano, "Pitch detection and voiced/unvoiced decision algorithm based on wavelet transforms," in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, vol. 2, pp. 1209–1212, Philadelphia, Pa, USA, October 1996.
- [12] A. Bouzid and N. Ellouze, "Glottal opening instant detection from speech signal," in *Proceedings of the 12th European Signal Processing Conference (EUSIPCO '04)*, Vienna, Austria, September 2004.
- [13] A. Bouzid and N. Ellouze, "Local regularity analysis at glottal opening and closing instants in electroglottogram signal using wavelet transform modulus maxima," in *Proceedings of the 8th European Conference on Speech Communication and Technology (EUROSPEECH '03)*, pp. 2837–2840, Geneva, Switzerland, September 2003.
- [14] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 747–758, 1994.
- [15] B. M. Sadler and A. Swami, "Analysis of multiscale products for step detection and estimation," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 1043–1051, 1999.
- [16] M. Brookes, "A speech processing toolbox for MATLAB," 2007, <http://www.ee.imperial.ac.uk/hp/staff/dmb/voicebox/voicebox.html>.
- [17] F. Plante, G. F. Meyer, and W. A. Ainsworth, "A pitch extraction reference database," in *Proceedings of the 4th European Conference on Speech Communication and Technology (EUROSPEECH '95)*, pp. 837–840, Madrid, Spain, September 1995.
- [18] A. Krishnamurthy and D. Childers, "Two-channel speech analysis," *IEEE Transactions on Signal Processing*, vol. 34, no. 4, pp. 730–743, 1986.
- [19] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 24, no. 5, pp. 399–418, 1976.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

