

Electroglottographic Measures Based on GCI and GOI Detection Using Multiscale Product

Aicha Bouzid, Nouredine Ellouze

Abstract: This paper deals with glottal parameter estimation such as local pitch and open quotient from electroglottographic signal (EGG). This estimation is based on glottal closing instants and glottal opening instants determined by a multi-scale product of this signal. Wavelet transform of EGG signal is made with a quadratic spline function. Wavelet coefficients calculated on different dyadic scales, show modulus maxima at localized discontinuities of EGG signal. The detected maxima and minima correspond to the glottal opening and closing instants called GOIs and GCIs. To improve the estimate precision, we operate the multi-scale product of wavelet transform coefficients of three successive dyadic scales. This processing enhances edge detection. A Multi-scale product is a nonlinear combination of successive scales; it reduces noise and spurious peaks. We apply cubic root amplitude on the product to improve the representation of weak amplitudes. The method has a good representation of GCI and a best detection of GOI. The method was tested on the Keele University database; it is effective and robust in multiple cases even for a typical signal showing undetermined GOIs and multiple peaks at GCIs. Finally precise measurement of these instants allows accurate estimation of prosodic parameters as local pitch and open quotient.

Keywords: wavelet transform, multi-scale product, electroglottographic signal, glottal closing instant, glottal opening instant

1 Introduction

Electroglottography is a non-invasive medical exploration technique of a glottal activity. The resulting signal called electroglottogram (EGG) is a common and efficient reference signal for pitch estimation. EGG signal can be used to determine glottal closing instant (GCI) and glottal opening instant (GOI). GCI is commonly used in speech processing like voiced/unvoiced classification, accurate source parameter estimation and robust instant detection; it is very useful in synchronous speech analysis and synthesis. GOI is useful for voice quality estimation and other voice and speaker characterization. In the present work, the estimated GCI and GOI are used to calculate the local fundamental frequency and the glottal open quotient. Referring to Childers [1], EGG signal presents an important amplitude variation at GCI, indicating a rapid behaviour change of the source (glottis). The derivative of the EGG signal, called DEGG shows strong peaks at closing instants and weak ones at opening instants. Referring to Mallat [2], important information lies in sharp transitions on the signal or its derivative. These singularities are detected by following the wavelet transform modulus maxima at fine scales. As we have shown in previous work [3], wavelet transform is efficient in most cases for detecting singularities in EGG signal at closing instants, but not so for glottal opening instant detection. In fact, large variations of amplitude on EGG signal can be observed at GCIs, however at GOIs the discontinuities are less obvious and can hardly be detected on the signal [4]. Referring to Sadler [5], [6], the multiscale product method (MPM) is the product of wavelet transform coefficients on different scales. The scale multiplication gives better results of discontinuity detection than any single scale, especially on the localization performance [7], [8], [9]. This method is used for edge detection in image processing. The aim of this paper is to present a new measurement method of glottal parameters from EGG signal, such as GOI, GCI, the signal fundamental frequency F_0 and the open quotient O_q . Different time and spectral based methods have been proposed for estimating the pitch period and the open quotient O_q . In this work, we propose to apply the

multiscale product method (MPM) on the EGG signal, in order to improve the accuracy of GOI and GCI measurement, so as to obtain better estimation of pitch and open quotient measures.

The present paper is organised as follows. After introduction, section 2 focuses the well-known methods applied on EGG signal for GCI and GOI detection and their related problems. The first method uses various thresholds (or crossing levels) of the EGG signal, the second method is based on the derivative of EGG signal. Besides, it is shown that in many cases, these methods present undetermined GOI and in fewer cases ambiguous GCI. Section 3 presents the multiscale product method and its application to EGG signal, in order to improve the detection of GOIs and GCIs. Section 4 presents a multiscale product enhancement for GCI and GOI detection based on the cubic root of the MP (CRMP) and its comparison to the MP and the DEGG methods. Section 5 deals with EGG parameter measurements given by the CRMP, the crossing level and the DEGG extrema methods. Finally, section 6 concludes this work.

2 GCI and GOI Detection Methods and Limitations

Glottal opening instants (GOI) and closing instants (GCI) estimation can be carried out on EGG signal, by different methods using time or frequency based methods. Here, we present two basic methods: the first method is based on EGG amplitude crossing level (with different levels), and the second method is based on maxima and minima detection on the DEGG signal.

EGG threshold method

Rothenberg et al. [10] use the 50% crossing level of the signal amplitude from a base line. This method gives a direct value of GOI and GCI from EGG signal in the case of modal and tense voice. 35% of the maximum amplitude of the EGG has been also used as a threshold; this ratio gives also direct values of the GCI and GOI. Howard [11] uses the 3/7 maximum EGG amplitude threshold for GOI detection. Illustrations of these methods are depicted in figure 1. The major drawback of such methods is the lack of accuracy for the GCI detection and a missing of GOI.

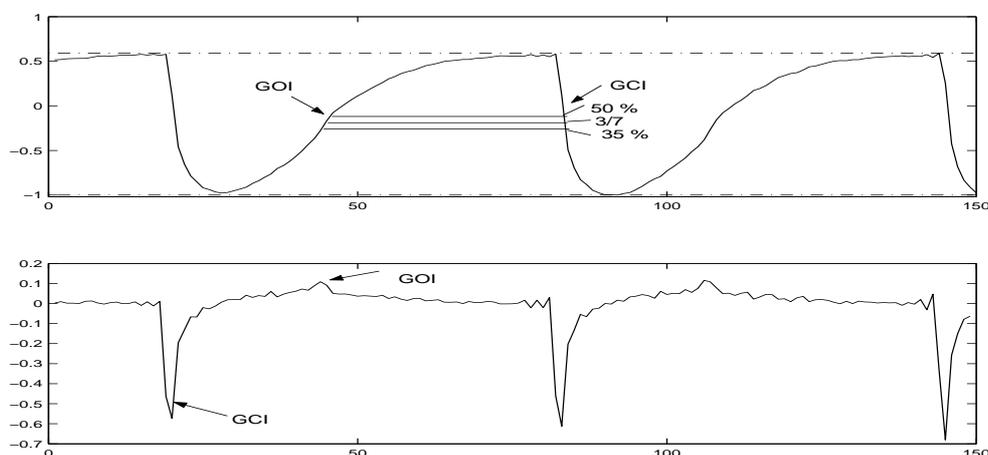


Figure 1: EGG signal (female voice, speaker f4, vowel [o]), 3 crossing levels, and the DEGG.

The top of figure 1 depicts the case of EGG signal of a female voice pronouncing the vowel [o], the crossing levels of 50%, 3/7 and 35% of the amplitude allowing the estimation of GOI and GCI.

DEGG maximum and minimum detection method

A more precise method for GCI detection uses the DEGG signal [12]. Childers characterises the EGG inflexion points by the DEGG [13], [14]. Experimental investigation shows that DEGG signal shows two opposite peaks in each period. The strong peak corresponds to GCI and the weak one corresponds to GOI [15], [16], [17]. An illustration of the derivative method is depicted at the bottom of figure 1.

GOI and GCI detection related problems

Methods based on EGG and DEGG, for GOI and GCI detection have different problems, such as missing events or duplication. Examples are taken from the Keele University database. It is a speech database containing acoustics and EGG signals simultaneously recorded in a soundproof room. It is made by five adult females and five adult males speakers. Each utterance consists of the same phonetically balanced English text. EGG and speech signals are given with the same sampling frequency of 20 KHz [18].

Many cases of missing peaks appear on DEGG at glottal closing instants [19]. Some GCIs and GOIs are indiscernible on the DEGG. This glottal behaviour is observed by Anastaplo and Karnell [20]. Pérez et al. underline the difficulty of GOI detection from the derivative of EGG signal [21].

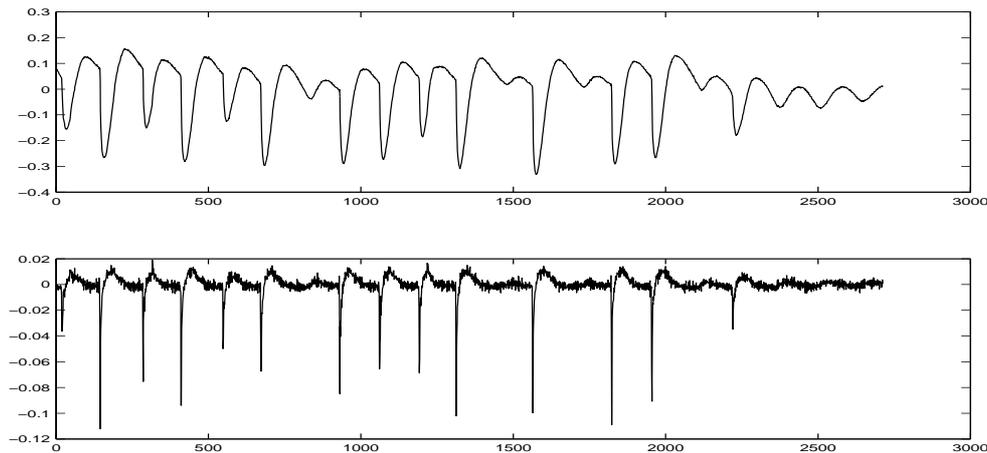


Figure 2: EGG signal of a male voice (speaker m2, voiced sound) and DEGG depicting the case of GCI and GOI missing.

GCI missing has a considerable effect on the time period measure. GOI missing influences the O_q measure.

Figure 2 illustrates problematical cases where threshold methods on EGG or DEGG fail to detect GCIs and GOIs. This example shows EGG signal of a voiced sound of the male speaker m2 and its derivative.

In many cases multiple peaks appear on DEGG at glottal closing instants and opening ones [19]. A typical example is presented in figure 3. In this case, the DEGG signal shows undetermined open instant and a double closing instant. EGG signal used for this example is a vowel [i] of the female speaker f2. We can distinguish that the EGG signal presents noise due to the subject movement during the recording.

Another typical example is shown on figure 4. Here the DEGG signal shows glottal opening and closing instants with a poor precision. The EGG signal used in this example is a voiced fricative [z] of the a male speaker m5. We can distinguish that the EGG signal presents natural noise of the fricative vowel.

3 GOI and GCI Detection by Multiscale Products

It is commonly known that wavelet transform is an efficient tool for detecting and characterizing signal singularities [2]. Singularities of the signal are detected by finding abscissa, where the wavelet modulus maxima converge at fine scales. The singularity type of the signal is characterised by wavelet vanishing moments and the decay of maxima across scales. This is explained by the fact that wavelet transform with n vanishing moments can be interpreted as a differential operator of n th order of the signal, smoothed by the primitive function of the wavelet called smoothing function. So if the wavelet is chosen to have one vanishing moment, modulus maxima appear at discontinuities of the signal, and

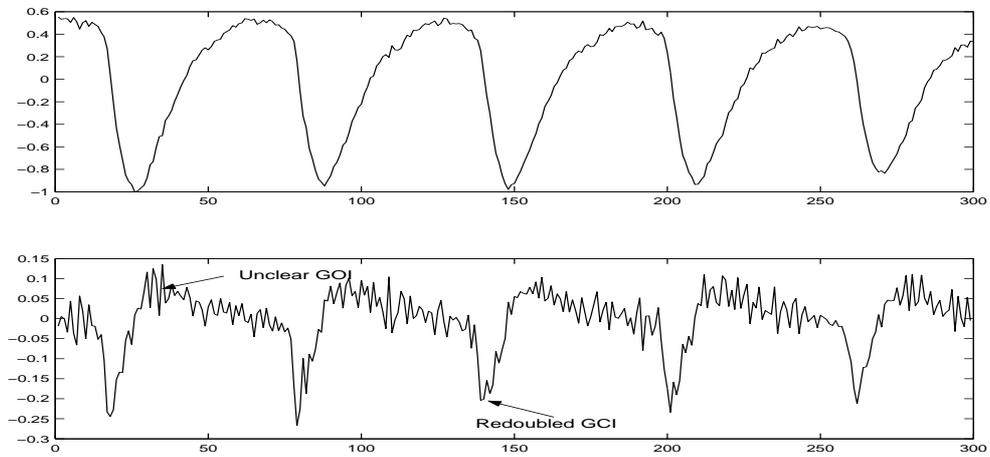


Figure 3: EGG signal (speaker f2, vowel [i]) and the DEGG depicting the case of double and imprecise peaks at GCI and GOI.

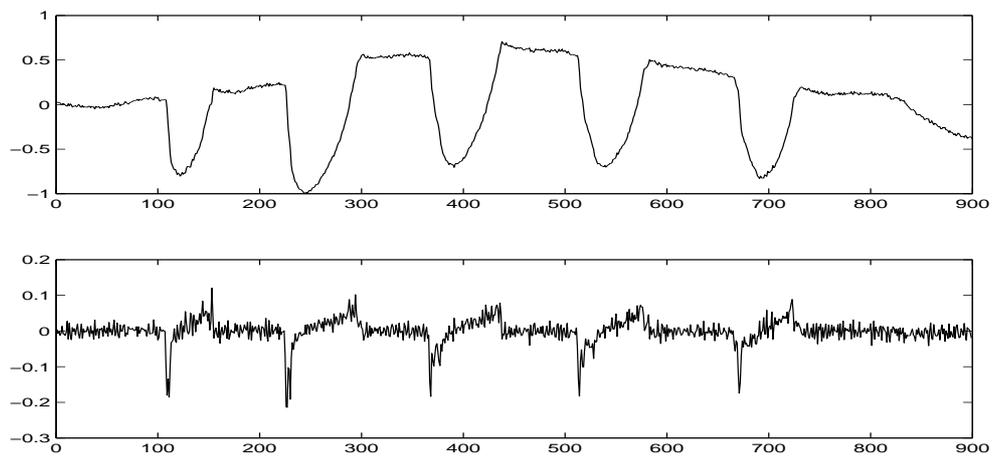


Figure 4: EGG signal (speaker m5, voiced fricative [z]) and the DEGG depicting the case of double and imprecise peaks at GCI and GOI.

represent the maxima of the first derivative of the smoothed signal. In previous work [3], we showed that the local regularity of the EGG signal can be characterised by the wavelet transform modulus maxima. Modulus maxima of EGG signal present two types of located singularities indicating glottal closing and glottal opening instants. The greater peak correspond to GCI and the weak one to the GOI. Modulus maxima give an estimation of the events with a better precision for small scales. But no scale can give an accurate value of GCI and GOI, so as singularities are too smoothed for large scales, and too weak for small scales, to be well located. Figure 5 shows the EGG signal of a voiced of speaker m2 followed by its wavelet transforms at the following scales 1/2, 1 and 2. Here wavelet transform (as well as DEGG), can't detect some singularities of the signal. Missed events still exist. That's why multiscale analysis seems to be necessary to improve the EGG edge localisation. The products of coefficients across scales are frequently used for image analysis. Witkin [20] provided the foundation for scale space theory by generalizing Rosenfeld's work [22], in which smoothing filters at dyadic scales are used. Based essentially on forming multiscale products of smoothed gradient estimates, this approach attempts to enhance the peaks of the gradients caused by true edges, while suppressing false peaks due to noise. The wavelet transform acts as an edge detector, and the detail coefficients should be equivalent to the estimated gradients. To distinguish edge maxima from noise and inappropriate maxima, Mallat and Zhong [23] analyze the singularity properties of wavelet transform domain maxima across various scales. The first derivative of a Gaussian and the quadratic spline are used to play this role. Xu et al. rely on the variations in scale of the wavelet transform. Direct multiplication of wavelet transform data at adjacent scales are used to distinguish important edges from noise [9]. Sadler and Swami [6] studied multiscale product method of a 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 diluting noise [6], [9]. The expression of multiscale product is given by:

$$p(n) = \prod_{j=1}^3 \omega_{s_j}(f(n)) \quad (1)$$

Where $\omega_{s_j}(f(n))$ is the wavelet transform of the function for the dyadic scale s_j . The product $p(n)$ has the property to reveal peaks at signal edges, and has relatively small values elsewhere. Thus, singularities produce peaks along scale in wavelet transform. These peaks are reinforced by the product $p(n)$. The signal peaks will just align across the first few scales, not all of them because increasing the amount of smoothing will spread the response and cause singularities separated in time to interact. Thus, choosing too large scales will result in misaligned peaks in $p(n)$. An odd number of terms in the product preserves the sign of the edge. Choosing three dyadic successive scales is an optimal solution in multiscale product for detecting small peaks.

As multiscale product improves the edge detection, we apply MPM to EGG signal to outperform the detection of GCI and GOI and to improve the measure precision particularly for weak singularities at GOI.

At the bottom of figure 5, we find the MP of the EGG signal of a voiced sound of speaker m2. We can clearly see the effect of the product on cancelling the additional noise peaks present at its derivative depicted in figure 2 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 enough. However, some missing events still exist. That's why multiscale product needs to be enhanced for an accurate GCI and GOI localisation.

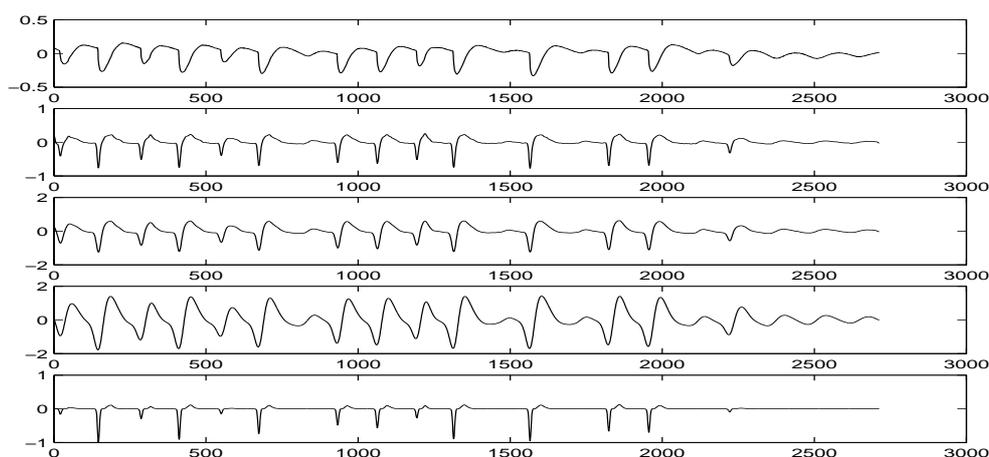


Figure 5: EGG signal of a voiced sound of speaker m2, the 3 wavelet transforms and their product.

4 Multiscale Product Enhancement for GOI and GCI detection

We have seen that multiscale product gives better results than wavelet coefficients. We use the cubic root of multiscale product to improve results for weak peaks. The cubic root of multiscale product has a zooming effect which enhances peak representation. This section deals with this basic enhancement and the comparative result detection of GCI and GOI detection by DEGG, MP and cubic root of MP. Specific examples from the Keele University database are presented below. Figure 6 shows an example of EGG signal and its derivative, which has precise peaks at GOI and GCI. The example is a frame of vowel [o] uttered by the speaker f4. The figure respectively shows the EGG signal, its derivative, the multiscale product of the following scales 1/2, 1 and 2 and the cubic root of the products. The multiscale product depicts the resulting cross-scale product $p(n)$ for three scales and shows clean peaks aligned with the DEGG signal edges. First, we note two types of peaks in the cross-scale product; those corresponding to GCI are more distinguishable than those related to GOI. The modulus cubic root shows maxima at GOI and GCI, the weak peaks at GOIs given by the modulus cubic root of the product are better represented and effectively reinforced than those obtained by the multiscale product. Figure 7 illustrates the example of a female utterance of vowel [i] where the GOIs are undetermined and the peaks of GCI present irregular structures. We note double peaks of glottal closing that bring about inaccurate measurements. Besides, we can clearly see the effect of the product in cancelling the additional peaks, which, then, gives a better detection of GOI. The modulus cubic root is used to reinforce the small peaks corresponding to GOI. This figure underlines the importance of the modulus cubic root of the multiscale product.

Figure 8 shows the EGG signal of fricative [z] of speaker m5, the DEGG signal, the multiscale product and its modulus cubic root. Figure 9 shows the same representation for a voiced sound of speaker m2. These examples illustrate the modulus cubic root applied on the multiscale product of EGG signal. In figure 9, continuous lines indicate GCI and dotted lines indicate GOI for missing events. In fact the cubic root takes out peaks at GOI and GCI that don't exist either in DEGG signal or in the product. It's a complex case. Thus, we can see that modulus cubic root of multiscale product can give a better detection of GCI and GOI from the EGG signal than the threshold methods and the DEGG approach. The MP cubic root reinforces the GCI minima indiscernible not only at DEGG but also at MP. Consequently the MPM associated with the cubic root gives the most efficient reference measures of GOI and GCI.

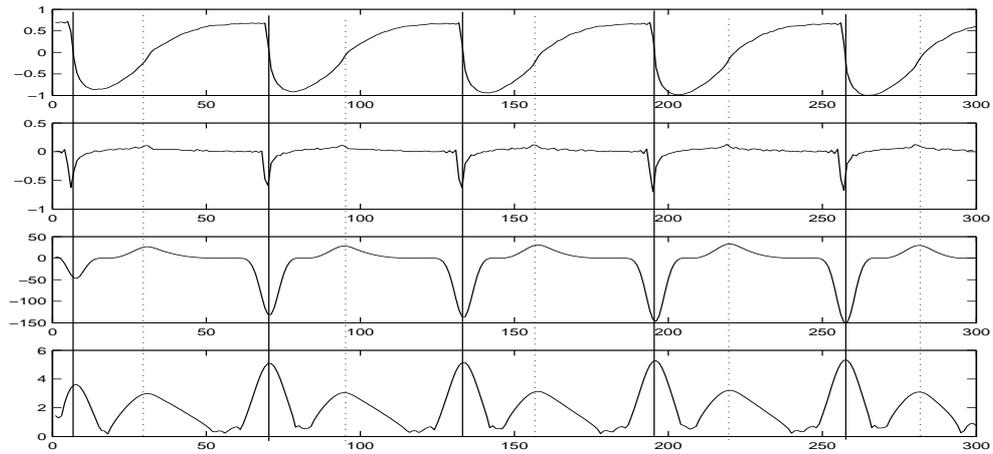


Figure 6: EGG signal of vowel [o] (speaker f4), DEGG, MP and its modulus cubic root.

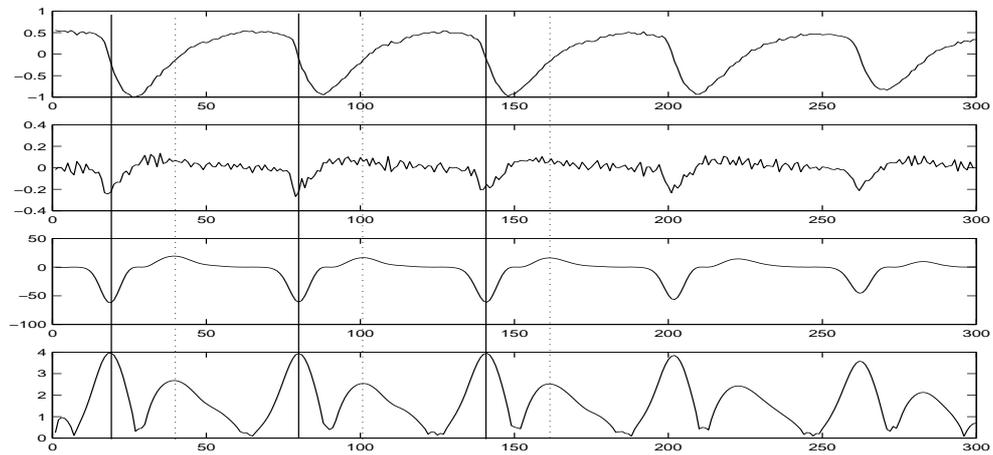


Figure 7: EGG signal of vowel [i] (speaker f2), DEGG, MP and its modulus cubic root.

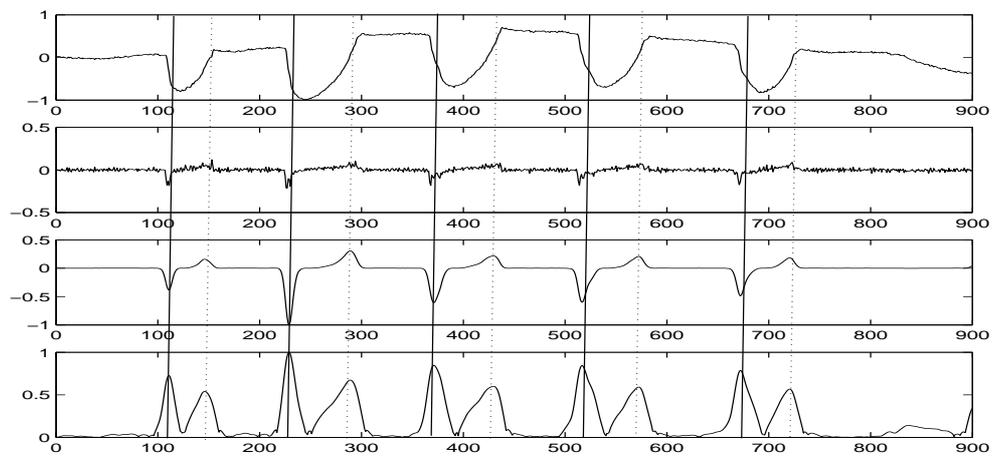


Figure 8: EGG signal of voiced fricative [z] (speaker m5), DEGG, MP and its modulus cubic root.

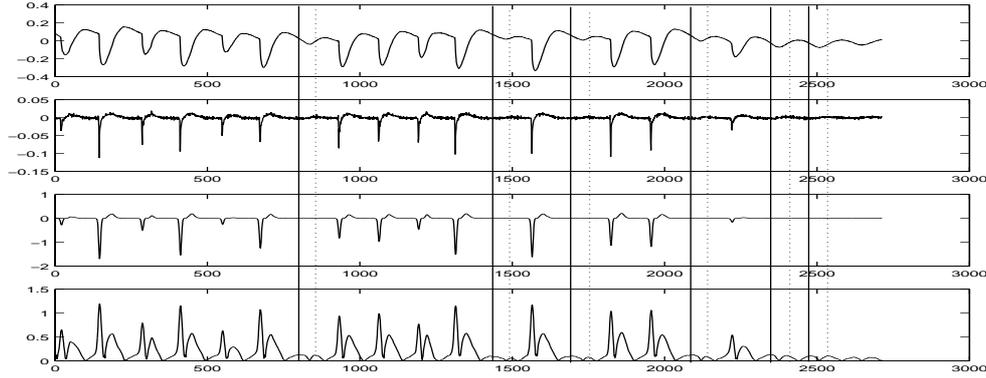


Figure 9: EGG signal of a voiced sound (speaker m2), DEGG, MP, and its modulus cubic root.

5 EGG Parameter Measurements

This section deals with the local pitch period and the local open quotient measurement. These can be calculated by using the $GOI(k)$ and $GCI(k)$. The glottal closing instant represent the beginning of the pitch period. The glottal opening instant $GOI(k)$ corresponds to the beginning of the open phase.

Measurement of F_0 and Oq

Local pitch period is given by the following formula

$$T_0(k) = GCI(k+1) - GCI(k) \quad (2)$$

The local fundamental frequency $F_0(k)$ is given by

$$F_0(k) = \frac{1}{T_0(k)} \quad (3)$$

The open quotient is defined as the ratio between the duration of the glottis open phase and the fundamental period. Open quotient is given by the following formula

$$Oq(k) = \frac{GCI(k+1) - GOI(k)}{T_0(k)} \quad (4)$$

In section 5, it is shown that the cubic root of MP outperforms the other methods in typical cases presented in this work. Moreover, this performance can be confirmed by evaluating prosodic parameter measures using the proposed method. Figure 10 depicts the local fundamental frequency F_0 of a voiced sound of speaker m2 corresponding to a case where the DEGG method gives imprecise GOI and the threshold method fails in determining some ambiguous GCI and GOI . Figure 11 depicts the local open quotient for the same utterance determined by the same methods. The threshold methods mentioned above fail in detecting some GCI s and GOI s leading to aberrant and missing measures of the fundamental frequency and the open quotient. Besides, imprecise detection of GOI from DEGG signal leads to aberrant measures of Oq as shown in periods 12 and 14 of figure 11.

6 Conclusion

Parameter characterisation of EGG signal by glottal closing instant GCI and glottal opening instant GOI detection from this signal is carried out by a new method called multiscale product MPM. The proposed method consists of computing the product of the wavelet transform of EGG signal at three successive dyadic scales. The wavelet used is the quadratic spline. This wavelet has one vanishing moment

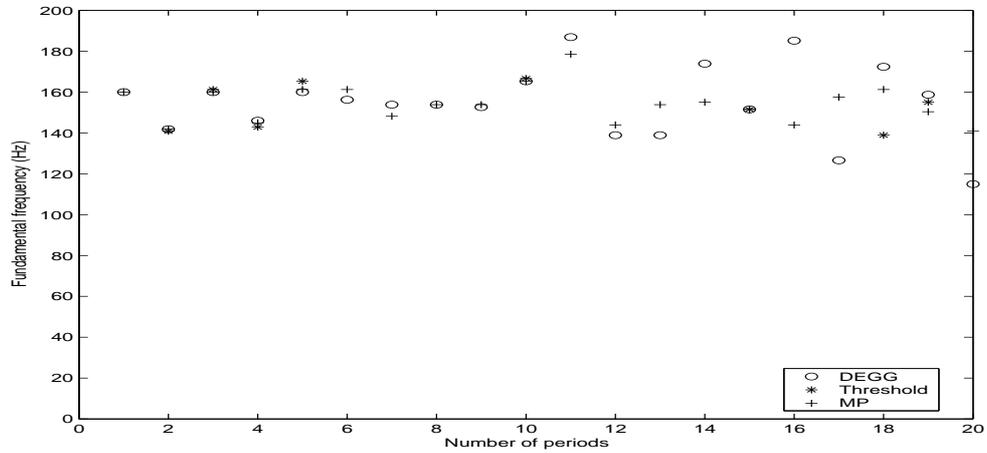


Figure 10: Fundamental frequency F_0 given by DEGG (o), 3/7 threshold (*) and MP (+) methods for a voiced sound of speaker m2.

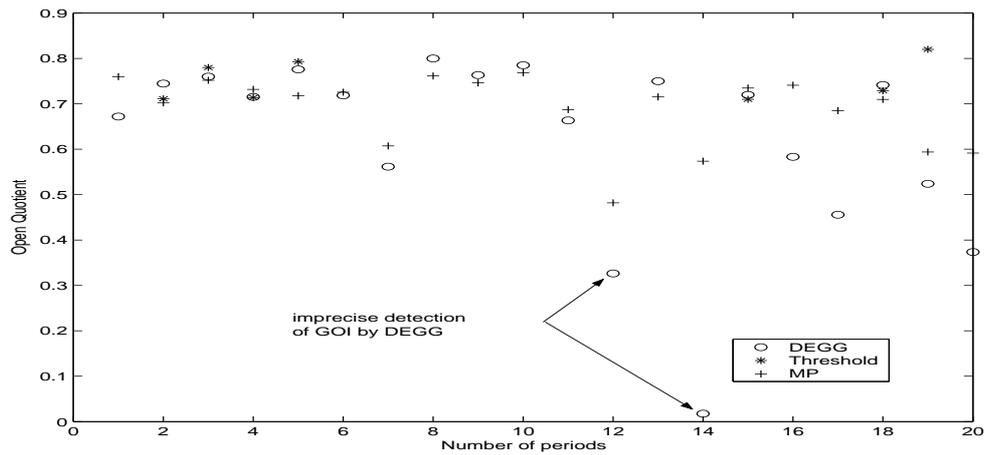


Figure 11: Open quotient O_q given by DEGG (o), 3/7 threshold (*) and MP (+) methods for a voiced sound of speaker m2.

and the calculated coefficients show modulus maxima at discontinuities of the signal. Wavelet transform is calculated at the following scales 1/2, 1 and 2. Then the cubic root amplitude of the product is calculated to enhance the resulting signal maxima. This method gives better GCI and GOI localisation than classical methods obtained by the crossing level or by the signal derivative, especially in typical cases of multiple peaks and undetermined GCI and GOI on DEGG signal. The non-linear products reinforce the cross-scale peaks produced at GCI and especially at GOI, and reduce spurious noisy peaks. Efficiency of the proposed method is proved. Comparative results are given with the threshold and derivative methods for glottal closing instants, glottal opening instants, and glottal parameter measurements like pitch frequency and open quotient. Locating GCI and GOI efficiently by using the cubic root of MP allows us to constitute a robust pitch and open quotient reference.

Bibliography

- [1] D. G. Childers, A. M. Smith and G. P. Moore, Relationships Between Electroglottograph, Speech, and Vocal Cord Contact, *Folia Phoniatr.*, Vol. 36, pp. 105-118, 1984.
- [2] S. Mallat, *A Wavelet Tour of Signal Processing*, Second Edition, Academic Press, San Diego 1999.
- [3] A. Bouzid, and N. Ellouze, Local Regularity Analysis at Glottal Opening and Closing Instants in Electroglottogram Signal Using Wavelet Transform Modulus Maxima, in Proc. Eurospeech 2003, Geneve, pp. 2837-2840, 2003.
- [4] N. Henrich, C. d'Alessandro, M. Castellongo, On the Use of the Derivative of Electroglottographic Signals for Characterization of Non-Pathological Phonation, *Journal of Acoustical Society of America*, Vol. 115, pp. 1321-1332, 2004.
- [5] B. M. Sadler, T. Pham, and L. C. Sadler, Optimal and Wavelet Based Shock Wave Detection and Estimation, *Journal of Acoustical Society of America*, Vol. 104, no. 2, pp. 955-963, 1998.
- [6] B. M. Sadler, and A. Swami, Analysis of Multiscale Products for Step Detection and Estimation, *IEEE Trans. Inform. Theory*, Vol. 45, no. 3, pp. 1043-1051, 1999.
- [7] L. Zhang, and P. Bao, Edge Detection by Scale Multiplication in Wavelet Domain, *Pattern Recognition Letters*, Vol. 23, no. 14, pp. 1771-1784, 2002.
- [8] P. Bao, L. Zhang, and X. Wu, Canny Edge Detection Enhancement by Scale Multiplication, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, no. 9, pp. 1485-1490, 2005.
- [9] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique, *IEEE Trans. Image Processing*, Vol. 3, no. 6, pp. 747-758, 1994.
- [10] M. Rothenberg, and J. J. Mahshie, Monitoring Vocal Fold Abduction through Vocal Fold Contact Area, *Journal of Speech and Hearing Research*, Vol. 31, pp. 338-351, 1988.
- [11] D. M. Howard, Variation of Electrolaryngographically Derived Closed Quotient for Trained and Untrained Adult Female Singers, *Journal of Voice*, Vol. 9, no. 2, pp. 1212-1223, 1995.
- [12] D. M. Howard, G. A. Lindsey, and B. Allen, Toward the Quantification of Vocal Efficiency, *Journal of Voice*, Vol. 4, no. 3, pp. 205-212, 1990.
- [13] D. G. Childers, and A. K. Krishnamurthy, A Critical Review of Electroglottography, *CRC Critical Reviews in Biomedical Engineering*, Vol. 12, pp. 131-161, 1985.

- [14] D. G. Childers, D. M. Hooks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, Electroglottography and Vocal Fold Physiology, *Journal of Speech Hearing Research*, Vol. 33, pp. 245-254, 1990.
- [15] D. G. Childers, and J. N. Lara, Electroglottography for Laryngeal Function Assessment and Speech Analysis, *IEEE Trans. on Biomedical Engineering BME*, Vol. 31, No. 12, pp. 807-817, 1985.
- [16] S. Anastaplo, and M. P. Karnell, Synchronized Videoscopic and Electroglottographic Examination of Glottal Opening, *Journal of Acoustical Society of America*, Vol. 83, no. 5, pp. 1883-1890, 1988.
- [17] M. H. Hess, and M. Ludwigs, Strobophotoglottographic Transillumination as a Method for the Analysis of Vocal Fold Vibration Patterns, *Journal of Voice*, Vol. 14, no. 2, pp. 255-271, 2000.
- [18] F. Plante, G. F. Meyer, and W. A. Ainsworth, A Pitch Extraction Reference Database, in . Eurospeech 1995, pp. 837-840, 1995.
- [19] A. Bouzid, N. Ellouze, *Contribution à la Détection des Instants d'Ouverture et de Fermeture de la Glotte sur les Signaux de Parole Voisé par Transformée en Ondelettes*, Thèse de doctorat, ENIT, Juillet 2004.
- [20] A. Witkin, Scale-Space Filtering, Proc. Int. Joint Conf. Artif. Intell., pp. 1019-1021, 1983.
- [21] J. Pérez, and A. Bonafonte, Automatic Voice-Source Parametrization of Natural Speech, in Proc. ICSLP 2005, Lisboa, Portugal, 2005.
- [22] A. Rosenfeld, A Non Linear Edge Detection, *Proc. IEEE*, Vol. 58, pp. 814-816, 1970.
- [23] S. Mallat, and S. Zhong, Characterization of Signals from Multiscale Edges, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, no. 7, pp. 710-732, 1992.

Aicha Bouzid
ENIT
Signal Processing Lab
ENIT B. P. 37, le Belvédère 1002, Tunis, Tunisia
E-mail: aicha.bouzid@enit.rnu.tn, bouzidacha@yahoo.fr

Noureddine Ellouze
ENIT
Signal Processing Lab
ENIT B. P. 37, le Belvédère 1002, Tunis, Tunisia
E-mail: N.Ellouze@enit.rnu.tn
Received: September 1, 2007



Dr. Aicha Bouzid was born in 29 April, 1975. She has Diploma in Electrical Engineering at Ecole Nationale d'Ingénieurs de Tunis (ENIT) in 1998, Master degree on automatic and signal processing in 2000 and Ph. D. Thesis at signal processing laboratory (LSTS), in July 2004. She has joined the Institut Supérieur des Etudes Technologiques de Sfax as associate professor in 1999. Currently she is working as assistant professor in the department of electronic at the Institut Supérieur d'Electronique et de Communication de Sfax. She has more than 30 publications in international journals and conferences. Her research areas of interest are signal processing, speech processing and applied mathematics.



Pr. Nouredine Ellouze was born in 19 December, 1945. He received a Ph.D. degree in 1977 at INP (Toulouse- France), and Electronic Engineering Diploma from ENSEEIHT in 1968 University P. Sabatier. in 1978. Pr. Ellouze joined the Electrical Engineering Department at ENIT (Tunisia). In 1990, he became Professor in signal processing, digital signal processing and stochastic process. He was the head of the Electrical Department from 1978 to 1983 and General Manager and President of IR-SIT from 1987-1994. He is now Director of Research Laboratory LSTS at ENIT, and is in charge of ATS Master Degree at ENIT. Pr. Ellouze has directed multiple Masters and Thesis and published more than 300 scientific papers in journals and proceedings, in the domain of signal processing, speech processing, biomedical applications, pattern recognition, and man machine communication.