WAVELET DECOMPOSITION OF VOICED SPEECH AND MATHEMATICAL MORPHOLOGY ANALYSIS FOR GLOTTAL CLOSURE INSTANTS DETECTION

Amel Ben Slimane Rahmouni[(1),(2)], *Aicha Bouzid*[(1),(3)], *Noureddine Ellouze*[(1)]
(1) ENIT (LSTS), (2) ESSTT, (3) ISET Sfax
BP. 37, Le Belvédère 1002 Tunis, Tunisie
Tel: (216) 71874700; fax: (216) 71872729
e-mail: Amel.Rahmouni@esstt.rnu.tn, Aicha.bouzid@enit.rnu.tn

ABSTRACT

This paper presents a robust algorithm for glottal closure instants (GCIs) detection of speech signals. The algorithm uses a multi-scale analysis based on a dyadic wavelet filterbank. Significant minima and maxima of the filtered signals are localized at each scale using adaptive mathematical morphology transformation of erosion. With reference to the GCIs detected from the laryngograph signal, a robust strategy for GCI localization was deduced. Each GCI is determined as the position of a minimum suitably chosen on one of the outputs of the different filters. This choice aims to ensure the best accuracy and reliability even for weak glottal effort.

1 INTRODUCTION

Pitch detection research shows a great interest in analyzing voiced speech period by period over an interval delimited by two successive instants of glottal closure. The glottal closure instants carry important information on the speech signal. Prosodic parameters like voicing degree and voicing frequency (F_0) can be derived from glottal closure instants. Efficient detection and estimation of pitch has many applications in audio signal processing. For example, pitch is very useful in speech processing applications as speech and language recognition, speaker identification and speech synthesis. So as determination of GCIs allows pitch synchronous processing of speech signals.

Glottal closure instants are often points of sharp variations or singularities in the speech signal. According to Mallat [1], the wavelet transform demonstrated excellent capabilities for detection of singularities in signals. Furthermore, in the last years, wavelet transforms have been intensively applied in different pitch detection algorithms [2] [3] [4]. Most of those algorithms are based on the dyadic wavelet transform, what it means a constant dilation factor equal to 2. Vu Ngoc [5] proposes speech representation in the time-scale domain by wavelet transform and a filterbank implementation. The main idea presented is that all dyadic scales are used for speech analysis. As a result, not only high frequency features are analyzed with accuracy but also smooth singularities in the signal can be detected. The present work, explores similar concept and proposes a robust strategy for glottal closure instants detection. The proposed strategy uses significant minima and maxima time localization of the filterbank outputs. A specific erosion mathematical morphology transformation is used for minima and maxima detection. The proposed algorithm takes decision from different scale minima giving the best estimation of the GCIs.

This paper is organized as follows. After an introduction, we describe in section 2 the dyadic wavelet transform used to analyze speech signals. Then, section 3 focuses on the peaks detection algorithm. In section 4, we present the strategy of GCI determination and some experimental results. Finally, concluding remarks are given in section 5.

2 DYADIC WAVELET TRANSFORM OF SPEECH SIGNAL

Wavelet transform is a powerful mathematical tool for hierarchical function decomposition . It allows a function to be described in terms of a coarse over all shape, plus details that range from broad to narrow. So it offers an elegant technique for representing different levels of details. Signal characteristics can be efficiently located in the space and frequency domains. Thus, unlike the Short Time Fourier Transform (STFT), wavelets are adequate for the study of non-stationary and unpredictable signals with both low frequency components and sharp transitions. The wavelet transform is a multiresolutional and multi-scale analysis which has been shown to be very well suited for speech processing.

We used a specific filterbank [5] in order to make a multi-scale analysis of the voiced speech signal. The mother wavelet g(t) used is given by equation 1 :

$$g(t) = -\cos(2\pi f_0 t).exp(-t^2/2\tau^2). \quad (1)$$

The transformed signal $y_i(t)$ at scale i of $x(t)$ is given by equation 2 :

$$y_i(t) = x(t) * g(t/s_i). \quad (2)$$

This is a convolution with a zero phase filter impulse response, even and non causal. The input signal and its response are in phase, then the phase of the input signal can be read in the phases of the output filters at each scale. The filterbank is implemented using scales $s_i = 2^i; \quad i = 0, 1, 2, 3, 4, 5, 6$. In the experiments the signals are sampled at $f_e = 20\ kHz$, $f_0 = fe/2$ and $\tau = 1/2f_0$. 7 band-pass filters centered at frequencies f_i : $10000, 5000, 2500, 1250, 625, 312, 156\ (Hz)$ are obtained. The frequency band of scale 5 and 6 filters covers the pitch range.

3 PEAKS DETECTION ALGORITHM

Mathematical morphology transformation of erosion is used, in order to detect significant peaks that could delimit periods of a pseudo-periodic signal [6]. The structuring element of the erosion, which is an important parameter, is a segment defined by its origin and length. The size of the erosion is the length of that segment. Let $E_{L,M}$ be the erosion of size $P = L + M$, the structuring element has its origin at L samples from the first extremity of the segment. When applied to a sampled signal x_k, this transformation gives the eroded signal y_k so that :

$$y_k = min[x_{k-L}, \cdots, x_{k+M-1}]. \tag{3}$$

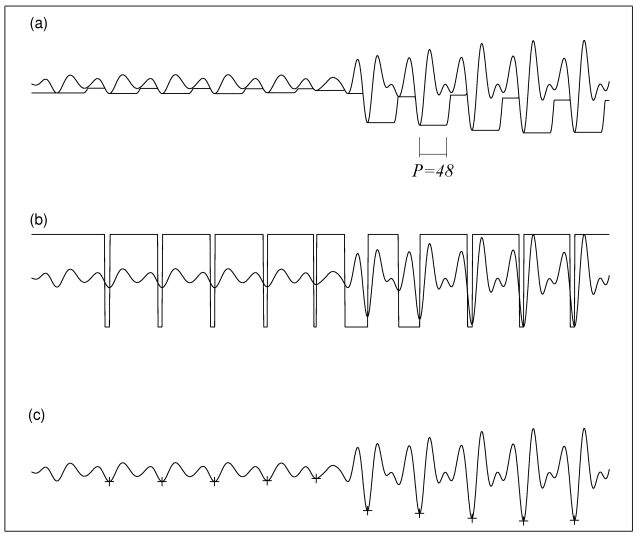

Figure 1: Minima detection on a quasi-periodic signal
(a) erosion $E_{0,48}$
(b) square transformation
(c) minima detected (+)

The structuring element used has its origin at its first extremity [6]. The eroded signal of a quasi-periodic signal is shown in figure 1-a. The erosion applied reduces the maxima and creates bearings next to the minima. In order to detect those minima a transformation called square transformation [6] is applied to the eroded signal. As shown in figure 1-b the square signal has two states corresponding to the increasing and the decreasing states of the eroded signal. Note that bearings which are between strictly decreasing and strictly increasing states of the eroded signal are counted with the increasing states. As it appears in figure 1-b ascending fronts of the square signal correspond to the minima positions of the original signal. Thus the minima are detected as shown in figure 1-c. The signal used in the previous illustration is a quasi-periodic signal. It is possible to detect the local minima that delimit a period, because the size of the erosion is adequately chosen in the inferior order of each period. Choice of the erosion size is done as in [7]. Henceforth the local minima detection system used needs a rough estimation of the period over a stationary interval. As the erosion size has to be chosen over a finite length segment, the minima detection algorithm is applied separately on such segments. Overlapping of the segments refer to the last but one location of the minima detected. The mean period value serves as a reference to guide the choice of the erosion size. In fact this size will be tuned until the right detection is achieved. The initial value of the erosion size is chosen equal to half of the mean period value in samples. Then it is incremented by one sample until the mean period value, calculated from the minima detected, correspond to the reference value. This extrema detection method is easy to implement and shows robustness to large amplitude and period variations of the signal.

4 GCI DETECTION ALGORITHM

Sharp glottal fold closure happens in case of important vocal effort, this results in peaks in the glottal flow signal derivative and in the speech wave [8]. The speech signal contains high frequency components in this case and GCI is well localized in time. Glottal folds closure can also be soft for low vocal effort, in this case the speech signal has low amplitudes, contains mainly low frequency components and the GCI is not well localized in time. The strategy proposed is GCI determination from both high frequency and low frequency behavior. For this purpose the outputs of the filterbank are investigated and the mathematical morphology peak detector is used.

In experiments, voiced speech segments are extracted from the Keele University database [9]. This database provides speech signal simultaneously registered with corresponding laryngogram signal (EGG signal : Electroglottogram signal). The EGG as well as speech signal are sampled at 20 kHz with 16 bit resolution. Band pass filtered EGG signal ($3-400\ Hz$) is derived (DEGG signal) and GCI reference values are located at the local minima of the DEGG signal.

Peak detector is applied to the outputs of the filterbank and to the opposite of the same outputs in order to detect respectively minima and maxima positions. Note that the erosion size required by the peak detector is

given by cepstral pitch period estimation on speech signal. This insures that the peaks detected should be at least distanced with the mean pitch period value. Cepstrum algorithm over approximately 52 ms segment of speech is used to determine the mean pitch period.

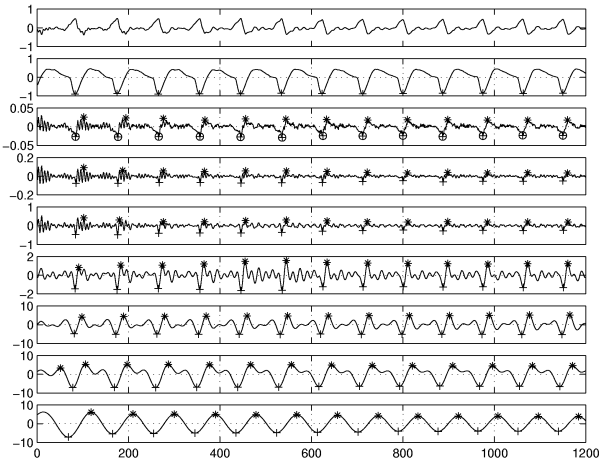

Figure 2: GCI detection for a voiced speech signal (female voice).
From top to bottom : speech signal - DEGG signal - filtered signals from scale 0 to scale 6.
symbols : + minimum, * maximum, o GCI

Figure 2 shows minima and maxima detected on the 7 outputs of the filterbank. Note that comparison between the peaks positions and the GCIs references conclude that the GCIs are located inside the meantime defined by minima and maxima of the scale 6 filter. As the scale decreases, the transitions between local minima and maxima become narrower and the accuracy on the GCIs localization increases. Besides, the minima and maxima positions converge to the reference GCIs for the channels where these extrema are detected and satisfy the condition of being included in the alternation minimum and maximum at the output of the pitch range filter. Thus the GCI is estimated as the position of the minimum given by the lowest scale which satisfies inclusion condition. In the worst case the mid instant of the minimum and maximum interval of the pitch range filter is chosen.

Figure 2 illustrates best conditions where abrupt glottal closure shows GCIs detected on the minima of the output of the scale 0 filter. As it can be seen in figure 3 (zoom on figure 2), a minimum and maximum alternation for one scale is included in the minimum and maximum alternation of the next one from small scales to greater ones. In this case, the GCIs are estimated from scale zero with the highest accuracy. Scale zero is the lowest scale satisfying the inclusion condition. Figure 4 illustrates an example where GCI detection fails at scale 0, but gives the best estimation at scale 1. Figure 5 gives an example of a voiced speech signal extinction

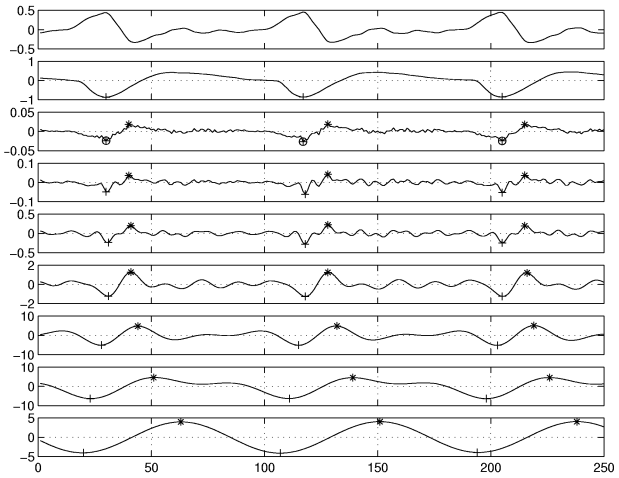

Figure 3: Zoom on figure 2 for samples 771 : 1020

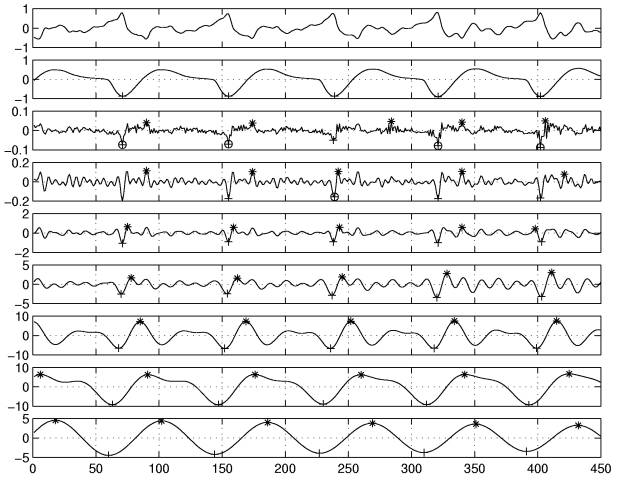

Figure 4: GCI detection for voiced speech signal (female voice)
GCI is detected from scales 0 and 1.
(same order of signals and symbols as in figure 2)

corresponding to vocal effort decreasing. This behavior is accompanied by a rapid change of instantaneous pitch period. The algorithm shows a good robustness in such interesting case. Detection process gives alternatively best estimation from three different scales.

5 CONCLUSION

This paper presents a multi-resolution analysis method for detecting glottal closure instants of voiced speech signals. It was shown that when all scales of wavelet decomposition are used the analysis, which requires an exhaustive search, is more efficient. The instants of minima and maxima of filter outputs are computed thanks to mathematical morphology. The algorithm of GCIs localization chooses minima detected at the smallest scales satisfying the inclusion condition. The multi-resolution

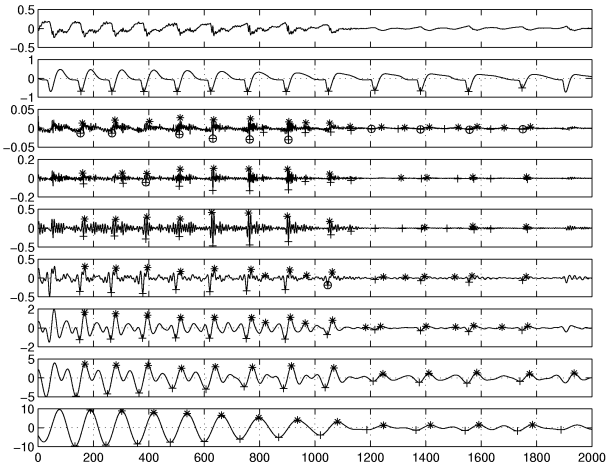

Figure 5: GCI detection for voiced speech signal (end of voicing, male voice).
(same order of signals and symbols as in figure 2)

analysis provides a hierarchical decomposition of the speech signal. The algorithm has shown a good robustness as the GCIs are in any case given by one scale filter. As the scale increases, the accuracy of GCIs estimation decreases and the reliability increases. Further work should be done on performance statistics of the proposed method.

References

[1] S. Mallat, Wen Liang Hwang, *Singularity detection and processing with wavelets*, IEEE trans. on IT, vol.38. No.2. pp.617-643, 1992.

[2] L. Janer, J.J Bonet, E.L Lleida-Solano, *Pitch Detection and voiced/Unvoiced Decision Algorithm based on Wavelet transforms*, Proc. ICSLP'96, Philadelphia, PA, USA, October 1996.

[3] S. Kadambe, G.F. Boudreaux-Bartels *A Comparison of wavelet functions for pitch detection of speech signals*, Proc. IEEE ICASSP'91, pp. 449-452, 1991.

[4] S. Kadambe, G.F. Boudreaux-Bartels *Application of the wavelet transform for pitch detection of speech signals*, IEEE trans. on IT, vol.38. No.2, pp. 917-924, 1992.

[5] T. Vu Ngoc, C. d'Alessandro, *Robust glottal closure detection using the wavelet transform*, Proc. Eurospeech'99, pp. 2805-2808, Budapest, september 1999.

[6] A. B.Slimane Rahmouni, B. Zouabi, N. Ellouze, *Traitement morphologique du signal de parole*, Les Annales Maghrbines de l'Ingnieur, Vol. 12. N Hors Srie. Tome.I. pp.383-387, Novembre 1998.

[7] A. B.Slimane Rahmouni, N. Ellouze, *Pitch tracking based on adaptive mathematical morphology transformation and cepstral estimation*, Proc. Journes Scientifiques Franco-Tunisiennes JSFT'2000, Monastir, Tunisie, 25-26 Octobre 2000.

[8] W.J. Hess, *Pich Determination of Speech Signals*, Springer-Verlag, 1983.

[9] F. Plante, G.F. Meyer, W.A. Ainsworth, *A pitch extraction reference datababase*, Proc. Eurospeech'95, volume 1, pages 837-840, 1995.