Raising of the Computational Efficiency in Glottal Closure Instant Determination through Wavelet Transform*

Zdravko Nikolov, Atanas Gotchev

Institute of Information Technologies, 1113 Sofia

I. Introduction

The determination of Glottal Closure Instant (GCI) is an important task in voice signal analysis. At GCI the vocal folds close the glottis for the airstream coming from the lungs. By opening and closing the airstream passage the vocal folds generate quasi periodic impulse source for the vocal tract system. GCI determine the instantaneous voice signal pitches. They are used for speaker identification, speech recognition, voice pathology investigations, etc. GCI is of considerable importance for modelling the vocal tract too. The latter can be assumed a resonant system in free oscillating mode during closed glottis interval (CGI). In this case the vocal tract can be modeled applying linear prediction methods [1].

Some of the new methods for GCI determination use wavelet transform (WT) apparatus. It is well-known that WT are very appropriate for signal abrupt changes detection [2, 3]. Several types of wavelets have been applied for GCI determination, real [4, 5] and complex valued too [6]. The latter determine the signal transients through transform modulus maxima and equiphase lines. The most commonly used wavelet in this case is the Morlet wavelet [3, 5, 7]. Unfortunately, its application is complicated because of large computational costs. Several algorithms for improving the computational efficiency have been proposed. The most popular among them is known as the algorithm "a trous" [8]. Another approach uses an approximation of a wavelet with close to Morlet wavelet features [9].

Novel results for GCI determination with exponentially modulated (EM) wavelet are reported in the present work. The wavelet has a close to Morlet wavelet form and allows considerable computational costs reduction without loss of GCI determination accuracy.

II. Wavelet Transform

1. Definitions

WT is an inner product of the analyzed signal with a basis function which is scaled and translated to obtain the time-scale (time-frequency) signal representation. Let $s(t)$ be the analyzed signal and $\varphi(t)$ – the analyzing wavelet. Then WT can be defined as:

$$(1) \qquad W(\tau, a) = \frac{1}{\sqrt{a}} \int s(t) \varphi * \left(\frac{t - \tau}{a} \right) dt,$$

where $*$ is the complex conjugate operator, a – scale factor; τ – time shift. The equation can be presented in the frequency domain as follows:

$$(2) \qquad W(\tau, a) = \sqrt{a} \int S(w) \psi(aw) e^{jw} dw$$

Here $S(w)$ and $\Psi(w)$ are Fourier transforms on $s(t)$ and $\varphi(t)$, respectively.

Equations (1) and (2) can be interpreted as time-frequency analysis of $s(t)$ with filters with impulse response $\varphi(t)$ and constant relative frequency resolution $\Delta w/w$.

For discrete signals WT assumes the form:

$$(3) \qquad S(kT_s, a) = \frac{1}{\sqrt{a}} \sum_{i=k-n/2}^{k+n/2} s(i) g * \left(\frac{i-k}{a} \right),$$

where T_s is the sampling period, n – the wavelet width in samples at $a=1$.

The analyzing wavelet must satisfy the following conditions [7]:

    a) absolute quadratic integrability (function with finite energy);

    b) one-side spectrum $\Psi(w)=0$ for $w \leq 0$.

2. Morlet wavelet

The Morlet wavelet is a modulated Gaussian added some correction terms for zero mean value:

$$(4) \qquad \varphi(t) = (e^{ict} - e^{-c^2/2}) e^{-t^2/2}.$$

Its frequency characteristics is shifted Gaussian:

$$(5) \qquad \psi(t) = \sqrt{2\pi} [e^{(w-c)^2/2} - e^{w^2/2} - e^{-c^2/2}].$$

The parameter c is usually chosen by the condition requiring the two highest maxima ratio to be equal to zero. The choice of c close to 5 makes the second term negligible and it can be practically omitted.

Judging from the wavelet equations and its frequency characteristics one can see that it is well localized both in time and frequency domains. Its symmetric form assures linear phase characteristics. If the second term in (4) is omitted the wavelet loses its zero mean value and does not satisfy the admissibily condition b) [7]. For $c = 5$, $\psi(0) = \sqrt{2\pi} e^{-5^2/2} \approx 9 \times 10^{-6}$ which assures a good approximation. The wavelets for 4 successive dyadic scales are presented in Fig. 1 together with the corresponding spectra.

The Morlet wavelet is complex valuated and implemented to a real valuated signal leads to decomposition with complex coefficients. Thus a two time-frequency representations (in modulus and phase) can be performed [10].

WT computation using Morlet wavelet can be performed replacing formula (3) by the discrete convolution of the discrete signal and sampled wavelet for each scale a. Scale growth causes wavelet dilation which increases the multiplications number for each scale. For dyadic scale change ($a=a_0, 2a_0, 4a_0, \ldots$) the multiplications increase with powers

of two. For example, the eight scale corresponds to 256 times the multiplications of the starting scale. So, the need for more effective algorithms is obvious. One of the most popular and commonly used algorithms is the algorithm "a trous". It reduces the multipications in formula (3) retaining just the even wavelet samples in each scale transition [8]. For compensation of the information lost due to omitting of every odd sample a signal low-pass filtering with second order Lagrange interpolator is performed [8]. A disadvantage of the "a trous" algorithm is that for a desired scale decomposition one must calculates low-pass filtering for all previous dyadic scales.

3. Exponentially modulated wavelet

The exponentially modulated (EM) wavelet has been proposed in [9] as a result from the search for a form close to Morlet wavelet.

It is an even function:

(6) $$\psi(t) = (1 + \sigma|t|)e^{-\sigma|t|}e^{ict}.$$

Its frequency characteristics is the following:

(7) $$\psi(w) = \frac{4\sigma^3}{[\,\sigma^2 + (w-c)^2\,]^2}.$$

The admissibility condition verification shows that this wavelet has higher zero frequency value than the Morlet wavelet. This might cause problems in signal restoration from its wavelet decomposition. For instance, if $\sigma=1$ and $c=8$, $\Psi(0)=9\times10^{-4}$. However, for analysis purposes it gives acceptable results due to its good time and frequency localization.

For easier implementation EM wavelet can be presented as a superposition of two semi-wavelets: causal and non-causal, respectively, in view of current scale

(8a) $$\psi(t,a) = \begin{cases} (1+\sigma t)\,e^{-\sigma t/a}e^{ict/a} \text{ for } t\geq 0, \\ 0 \text{ for } t < 0, \end{cases}$$

(8b) $$\psi_+(t,a) = \begin{cases} (1-\sigma t)\,e^{-\sigma t/a}e^{ict/a} \text{ for } t < 0, \\ 0 \text{ for } t \geq 0. \end{cases}$$

The two semi-wavelet transforms in z-domain are as follow:

(9a) $$\psi_+(z) = \frac{1 + a_1 z^{-1}}{1 + b_1 z^{-1} + a_2 z^{-2}},$$

(9b) $$\psi_-(z) = \psi^*_+(z^{-1}) - \psi(0) = \frac{a'_1 z + a'_2 z^2}{1 + b'_1 z + b'_2 z^2},$$

where:

$$a_1 = (\sigma T_S/a - 1)\,e^{-(\sigma-ic)T_S/a},$$

$$b_1 = -2\,e^{-(\sigma-ic)T_s/a},$$

$$b_2 = e^{-2(\sigma-ic)T_s/a},$$

$$a'_1 = a^*_1 - b^*_1,$$

$$a^*_2 = -b^*_2,$$

$$b'_1 = b^*_1,$$

$$b'_2 = b^*_2,$$

If we represent equation (3) in z-domain replacing convolution with multiplication and the wavelet with its z-images, we should obtained a recurrent formula for WT calculation:

$$S_+(k,a) = s(k) + a_1 s(k-1) - b_1 S_+(k-1) - b_2 S_+(k-2),$$

(10) $$S_-(k,a) = a'_1 s(k+1) + a'_2 s(k+2) - b'_2 S_-(k+1) - b'_2 S_-(k+2),$$

$$S(k,a) = \sqrt{1/a} \, [S_+(k,a) + S_-(k,a)].$$

The complexity analysis leads to the following conclusions:

1. Multiplications number does not depend on the scale factor a.

2. To obtain a wavelet coefficient we need 11 multiplications, so the complexity is relative to signal length N.

3. The algorithm does not require a wavelet cutting and allows arbitrary choice of desired scale without calculation of higher scales.

III. Experimental results

Synthesized and natural signals corresponding to voiced sounds for applicability investigation of EM wavelet to GCI determination have been used.

For signal synthesis the Fujisaki-Ljngqvist (FL) model for differentiated glottal waveform (DGW) generation has been used [11]. FL model gives the opportunity to change the pitch periods of generated DGW as well as the close-to-open phase ratios. DGW have been used with pitch periods from 71 to 50 samples (from 162 Hz to 230 Hz at sampling frequency 11500 Hz). They cover the middle range of male and female pitches.

The natural signals have been recorded through direct analog-to-digital conversion at 11500 Hz sampling frequency and 16 bit resolution using sound card Z1 of Antex Electronics. A microphone AKG D330BT with high-frequency filtering at 50 Hz has been used. In this case the low-pass noise due to breathing has been excluded.

Four natural vowels /a/, /e/, /i/, /u/ for different pitches have been examined.

Apart from wavelet method a Frobenius norm approach has been applied. It is known as the most accurate method for GCI determination and has been used for making comparisons of GCI determination accuracy via WT for natural signals.

The time-frequency plane tilings of WT phase characteristics for six octaves and three "voices" per octave have been obtained. They have been compared to the corresponding tilings got using the Morlet wavelet [6]. The comparison has shown that there is any noticeable difference between the tilings and in both cases equiphase lines lead to GCI.

The developed algorithm allows arbitrary choice of the desired scale to obtain a WT modulus decomposition. According to the relation between scale and frequency (the reciprocal value of the scale is equal to the frequency) we can interpret the results from the physiological voice production point of view. It is well-known that the first formant resonance appears first after the glottis closure [1]. This fact enables the investigation of signals just for scales corresponding to the first formant and pitch frequency.

The first formant and pitch octave signals for synthesized signal analysis are presented in Fig. 1. The excitation source is presented too. The first formant signal has maxima corresponding to GCI. The pitch signal maxima are shifted towards the closed glottis intervals. The results for natural signals are similar. Fig. 2 represents such a WT together with the Frobenius measure curve.

EM wavelet decomposition reacts adequately for non-stationary signals too. A scale close to the first formant must be used. Such a case is presented in Fig. 3. Because of the pitch changes the scale close to the pitch does not detect GCI well enough. Like other wavelets [4] EM wavelet decomposition is not influenced by noises in the analyzed signal.

IV. Conclusions

EM wavelet is close in form to the Morlet wavelet. For this reason one can proceed as in case of Morlet wavelet decomposition obtaining complex (by modulus and phase) time-frequency tiling for successive octaves and "voices" per octave. The computational costs will be increased in this case but less than in the case with the Morlet wavelet.

The calculation algorithm is attractive because it allows using arbitrary scale. This is very useful in cases when the first formant is known or can be determined easy. Its value can be used for scale setting which leads to accurate results in GCI determination. It is not suitable to use scale close to the pitch frequency because such scale is sensitive to signal nonstationarities.

References

1. М а р к е л, Дж., А. Г р е й. Линейное предсказание речи. М., Связь, 1980.
2. M a l l a t, S., W. L. H w a n g. Singularity detection and processing with wavelets. – IEEE Trans. Inform. Theory, **38**, 1992, 617–643.
3. K r o n l a n d - M a r t i n e t, J. M o r l e t, A. G r o s s m a n. Analysis of sound patterns through wavelet transform. – J. Patern Recognition and Artifical Intell., **1**, 1987, 273–302.
4. Г о ч е в, А. Определяне интервалите на затворен глотис чрез вълнова трансформация. – Техн. мисъл, 1995, 28–42.
5. C a d a m b e, S., G. B o d r e a u x - B a r t e l s. A comparison of a wavelet functions for pitch detection of speech signals. – In: Proc. ICASSP, 1991, 449–452.
6. N i k o l o v Z., A. G o t c h e v. Detection of glottal closure instant via phase characteristics of wavelet transform. – In: Proc. Int. AMSE Conf. "Sys'94", 1994, Lyon, France, 133–144.
7. G r o s s m a n n, A., J. M o r l e t. Decomposition of Hardi functions into square integrable wavelets of constant shape. – SIAM J. of Math. Analysis, 1984, **15**, 725–736.
8. H o l s h n e i d e r, M., R. K r o n l a n d - M a r t i n e t, J. M o r l e t, P. T c h a m i t c h i a n. A real time algorithm for signal analysis with the help of the wavelet transform. – In: Wavelets, time-frequency methods and phase space. Springer-Verlag, 1989, N.Y., 286–297.
9. B a r r a t, M., O. L e p e t i t. Calcul rapide de la transformée en ondelettes. – Traitement du signal, 1991, **8**, 43–49.
10. G r o s s m a n n, A., R. K r o n l a n d - M a r t i n e t, J. M o r l e t. Reading and understanding continuous wavelet transforms. – In: Wavelets, Time-Frequency Methods and Phase Space. Proc. Int. conf.,1987, Marseille, France, 2–20.
11. F u j i s a k i, H., M. L j u n g q v i s t. Proposal and evaluation of models for the glottal source waveform. – In: Proc. ICASSP, 1986, 1605–1608.
12. M a, C., Y. K a m p, L. F. W i l l e m s. A frobenius norm approach to glottal closure detection from the speech signal. – In: IEEE Trans. Speech and Audio Proc., 1994, **2**, 258–265.

Повышение вычислительной еффективности при определении момента закрытого глотиса при помощи волновой транформации

Здравко Николов, Атанас Гочев

Институт информационных технологий, 1113 София

(Р е з ю м е)

Работа обсуждает актуальную проблемму исследования звуковых сигналов при помощи трансформации в разных пространстьвах. Выбрана волновая транс-формация, а основная цель работы повысить вычислительную еффективность, применяя комплексные волны и получая декомпозиции в модуле и фазе во времево-масштабной плоскости. Предложена волна, сходная известной волне Морлета, которая описана в z-области. При помощи предложенного алгоритма декомпозиции исследованы синтезированные и естественные сигналы. Экспериментальные результаты показывают точность метода при нестационарных сигналах и при наличии шума.