Estimation of instantaneous parameters of speech signals
with Teager-operator and Hilbert-Huang-transform

István Pintér

College of Kecskemét, GAMF Faculty, Department of Automation and Applied Informatics
pinter.istvan@gamf.kefo.hu

***Keywords:** Teager-operator, HHT, instantaneous amplitude and frequency, reconstruction from instantaneous parameters*

In order to analyze the fine structure of speech signals, methods for determining nonlinear and non-stationary characteristics of speech are necessary. In this paper the Teager-operator and the Hilbert-Huang Transform (HHT) are presented as speech processing methods suitable for the estimation of instantaneous amplitude and instantaneous frequency. **(In: 2006/8, pp.28–37.)**

1. Introduction

In digital speech processing the so-called quasi-stationary signal model is often used in solutions to a lot of practical problems. It means that it is possible to process the speech signal with the sequence of overlapped speech segments in order to complete the computations necessary for the algorithm in question. It is presumed that the parameters of the applied speech model do not vary during the segment under processing. According to the relevant literature, the suitable segment duration is 2...5 times longer than the fundamental period, and the sufficient overlap is 1...3 times of pitch period [1].

During the progress of digital speech processing a demand has arisen for analyzing methods suitable for investigating speech signal changes of lower duration than the pitch period itself. These types of speech signal changes constitute the fine structure of speech. The phenomenon of the small fluctuation of the fundamental period during nonlinear vocal chord vibration – among many others – is an example which calls for methods necessary to represent the fine structure of speech. These methods ought to derive physically meaningful parameters from few speech samples only. Consequently, for these purposes the speech processing methods based on the quasi-stationary assumption are not appropriate [2].

Summing up succinctly the essence of the problem, it could be said that it is not possible to increase the time-resolution of the analysis, while keeping the detailed frequency-domain representation because of the uncertainty-relation of Dennis Gabor. Nowadays, the wavelet-transform is widely used in applications requiring increased time-resolution. However, the time-resolution of the wavelet-method is also limited by the time-scale uncertainty relation, which replaces the time-frequency uncertainty mentioned above [2,3].

A possible method, suitable for analyzing the fine structure of speech, is the Teager-operator-based Energy Separation (ES) algorithm. The wavelet-based analysis and the Teager-operator have recently led to successful applications [4]. Another possibility in determining the instantaneous parameters is the application of the Hilbert-Huang-transform [5]. Because we have not found published results in the relevant literature available to us on the comparison of the Teager-operator-based methods and HHT-based ones, their comparison has been chosen as the subject of this article.

2. The Teager-operator and the ES-algorithm

2.1. The continuous-time Teager-operator and the estimation of instantaneous parameters

The definition of the Teager-operator has become possible after detailed investigation of the nonlinear physical phenomena of human speech production. In order to describe the fast changes in the speech-signal's energy during the fundamental period, it is useful to determine the overall energy of the system producing the speech. This overall energy can be estimated by applying a suitable operator to the speech signal – the operator is termed as Teager-operator [2]:

$$\Psi\{x(t)\} = \left(\frac{dx(t)}{dt}\right)^2 - x(t)\cdot\frac{d^2x(t)}{dt^2}, \quad (1)$$

where Ψ{.} denotes the Teager-operator. In the case of the signal x(t)=a·cos(ω·t+φ) we get:

$$\frac{dx(t)}{dt} = -a\cdot\omega\cdot\sin(\omega\cdot t+\varphi),$$
$$\frac{d^2x(t)}{dt^2} = -a\cdot\omega^2\cdot\cos(\omega\cdot t+\varphi), \quad (2)$$

which leads to

$$\Psi\{x(t)\} = a^2\cdot\omega^2 \quad (3)$$

It can be checked that the result will be the same when the operator is applied to the signal x(t)=a·sin(ω·t+φ), as it is expected. It is interesting to note that the next equation also holds:

$$\Psi\{a\cdot e^{j(\omega t+\varphi)}\} = 0\,. \quad (4)$$

A possible generalization of the signal x(t)=a·cos(ω·t+φ) is the case when both the amplitude and the phase are time-dependent, that is the form of the resulting AM-FM signal is the following:

$$x(t)=a(t)\cdot\cos(\varphi(t)). \qquad (5)$$

By direct algebraic manipulation it is easy to check that in the case of arbitrary amplitude- and phase-function the application of the operator in (1) results in a formula which is not easy to manipulate further. However, in the case of slowly varying amplitude- and phase-functions by using the approximations below:

$$\frac{da(t)}{dt}\approx 0, \quad \frac{d\varphi(t)}{dt}\approx \text{const}, \quad \frac{d^2\varphi(t)}{dt^2}\approx 0, \qquad (6)$$

and by applying the Teager-operator to the signal in (5) we get:

$$\frac{dx(t)}{dt}\approx -a\cdot\frac{d\varphi(t)}{dt}\cdot\sin\varphi(t),$$

$$\frac{d^2x(t)}{dt^2}\approx -a\cdot\left(\frac{d\varphi(t)}{dt}\right)^2\cdot\cos\varphi(t),$$

$$\Psi\{x(t)\}\approx a^2(t)\cdot\left(\frac{d\varphi(t)}{dt}\right)^2. \qquad (7)$$

The operator can also be applied to the derivative of the signal:

$$\Psi\left\{\frac{dx(t)}{dt}\right\}=\left(\frac{d^2x(t)}{dt^2}\right)^2-\frac{dx(t)}{dt}\cdot\frac{d^3x(t)}{dt^3}. \qquad (8)$$

By using the approximations in (6), after detailed calculations we finally get the relation below for the AM-FM signal in (5):

$$\Psi\left\{\frac{dx(t)}{dt}\right\}\approx a^2(t)\cdot\left(\frac{d\varphi(t)}{dt}\right)^4. \qquad (9)$$

Therefore, both the value of the magnitude and the magnitude of the derivative of the phase (which is by definition the absolute value of the angular frequency) can be estimated by the formulae below:

$$\frac{\Psi\{x(t)\}}{\sqrt{\Psi\left\{\frac{dx(t)}{dt}\right\}}}=|a(t)|, \qquad (10)$$

$$\sqrt{\frac{\Psi\left\{\frac{dx(t)}{dt}\right\}}{\Psi\{x(t)\}}}=\left|\frac{d\varphi(t)}{dt}\right|. \qquad (11)$$

So, based on equations (1), (10) and (11) the slowly time-varying envelope and the slowly time-varying instantaneous angular frequency can be estimated from the signal itself. It is easy to check that for the signal $x(t)=a\cdot\cos(\omega\cdot t+\varphi)$ these estimations give exactly the same values of the (constant) amplitude and (constant) angular frequency.

2.2. Discrete-time Teager-operator and the ES-algorithm

After proper sampling and suitably approximating the derivation with differences equations (1), (10) and (11) can be considered as the basis of the computations. According to our numerical experiments the five-point Savitzky-Golay smoothing derivative algorithm [6] gives acceptable results. This type of computation is further called as direct computation. The discrete-time version of the Teager-operator can also be derived from the continuous form in (1) by approximating the differentiation with the $d(n)=x(n)-x(n-1)$ difference. This leads to the next definition of the discrete-time Teager-operator:

$$\Psi_D\{x(n)\}=x^2(n)-x(n-1)\cdot x(n+1). \qquad (12)$$

After some algebraic manipulation it calls forth that by applying the Teager-operator to the signal $x(n)=a\cdot\cos(\omega\cdot n+\varphi)$ discrete-time sequence the result is below:

$$\Psi_D\{x(n)\}=a^2\cdot\sin^2\omega, \qquad (13)$$

where ω denotes digital angular frequency. In the case of the discrete-time Teager operator it can be shown that, starting from the $x(n)=a(n)\cdot\cos(\varphi(n))$ sequence the estimation formulae of the slowly varying instantaneous parameters are given below [2]:

$$a(n)\approx\frac{2\cdot\Psi_D\{x(n)\}}{\sqrt{\Psi_D\{x(n+1)-x(n-1)\}}}. \qquad (14)$$

$$\omega(n)\approx\arcsin\left(\sqrt{\frac{\Psi_D\{x(n+1)-x(n-1)\}}{4\cdot\Psi_D\{x(n)\}}}\right). \qquad (15)$$

The computation procedure defined by equations (12), (14) and (15) is called ES-algorithm in the literature. The benefit of the ES-algorithm is that it needs only three samples for the estimation, while the direct method above needs five samples, however, in the latest case the evaluation of the arcsin(.) function is not needed for determination of instantaneous digital angular frequency.

3. The Hilbert-Huang-transform and the computation of the instantaneous parameters

In the previous part it has been shown when several well-defined conditions are fulfilled, the computation of the instantaneous parameters is possible. These conditions can be guaranteed e.g. with a suitable band-pass filtering before estimation.

A natural question could arise: is there a much more general method for estimating the physically meaningful instantaneous parameters? The question was answered positively in 1988 in a paper by Norden E. Huang et al. [5]. The authors elaborated a signal decomposition algorithm which results in signal components having positive instantaneous frequencies, so the instantaneous parameters can be estimated using these components.

They proposed the so-called EMD-algorithm (Empirical Mode Decomposition), they termed the component signals as IMFs (Intrinsic Mode Function), and the instantaneous parameters of IMFs can be estimated by using the so-called canonical representation of analytic signals.

3.1. The empirical mode decomposition algorithm and the intrinsic mode functions

The intrinsic mode functions ought to have two basic properties [5]:

– The number of extrema and the number of the zero-crossings are the same or their difference equals 1,

– The mean value between the envelopes of local maxima and local minima is zero.

Details of the algorithm for determination of the intrinsic mode functions can be found in [5]. When determining an IMF only local speech sample values are used, that is the IMFs are computed with a locally adaptive manner. Moreover, the original signal can be reconstructed by summing up the IMFs, that is:

$$x(n)=r(n)+\sum_{k=0}^{K-1}m_k(n), \tag{16}$$

where *r(n)* denotes the residual signal, *$m_k(n)$* denotes the k-th IMF, and K denotes the number of IMFs. There is no estimation for the number of IMFs in [5], so it has to be determined experimentally.

3.2. The canonical representation of the signal and the instantaneous parameters

As it is well known from the work of Dennis Gabor [7], the x(t)=a(t)·cos(φ(t)) signal model is not unambiguous, but the so-called canonical representation – which can be derived from the complex analytic signal – is unambiguous. This latter signal is composed from the signal itself and also from its Hilbert-transform as given below:

$$\hat{x}(t)=H\{x(t)\}=\frac{1}{\pi}\cdot P\int_{-\infty}^{+\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{17}$$

$$Z(t)=x(t)+j\cdot\hat{x}(t)=A(t)\cdot e^{j\Phi(t)} \tag{18}$$

and the canonical representation is defined as:

$$x(t)=A(t)\cdot\cos(\Phi(t)) \tag{19}$$

The instantaneous parameters in (19) are defined as:

$$A(t)=\sqrt{x^2(t)+\hat{x}^2(t)} \tag{20}$$

$$\omega(t)=\frac{d\Phi(t)}{dt}=\frac{d}{dt}\left(\arctan\left(\frac{\hat{x}(t)}{x(t)}\right)\right). \tag{21}$$

Although equation (21) defines the instantaneous angular frequency as a derivative of the phase of the analytic signal it can also be computed as the following partial derivative:

$$\omega(t)=\mathrm{Im}\left\{\frac{\partial}{\partial t}\ln(Z(t))\right\} \tag{22}$$

which leads to the next relation:

$$\omega(t)=\frac{x(t)\cdot\frac{d\hat{x}(t)}{dt}-\frac{dx(t)}{dt}\cdot\hat{x}(t)}{x^2(t)+\hat{x}^2(t)}, \tag{23}$$

which can also be reached after the completion of the derivation in (21). Both (21) and (23) can be used for deriving an algorithm for the estimation of instantaneous angular frequency. From the point of view of realization we have to note there is an important relation between the Hilbert-transform and Fourier-transform of the signal, which is the following:

$$\hat{X}(j\omega)=-j\cdot\mathrm{sgn}(\omega)\cdot X(j\omega), \tag{24}$$

moreover, the relation below also fulfils:

$$F\{x(t)+j\cdot\hat{x}(t)\}=\begin{cases}2\cdot X(j\omega) & \omega>0\\ 0 & 0\leq\omega\end{cases}, \tag{25}$$

where F{.} denotes the Fourier-transformation.

3.3. The computation of the discrete-time Hilbert-transform and the estimation of the instantaneous parameters

The discrete-time Hilbert-transform of a signal can be computed either starting from (24) and by applying a suitable digital filtering operation [8] or by using an FFT-based algorithm (25). After determining the Hilbert-transform of the speech, the estimation of the instantaneous amplitude can be given as follows:

$$A(n)=\sqrt{x^2(n)+\hat{x}^2(n)} \tag{26}$$

For the computation of the instantaneous frequency two algorithms can be derived depending on the use of (21) or on the use of (23). By using (21) the phase-sequence can be given with the equation below:

$$\Phi(n)=\mathrm{arctg}\left(\frac{\hat{x}(n)}{x(n)}\right), \tag{27}$$

During the evolution of the signal the phase-change can be given as:

$$\Phi_u(n)=\Phi(n)+r(n)\cdot 2\pi, \qquad \Phi(n)\in[-\pi;\pi], \tag{28}$$

where *r(n)* is a positive integer. After computing the instantaneous phase by appling a suitable phase-unwrapping algorithm, the instantaneous digital angular frequency can be approximated by the difference below:

$$\omega(n)=\Phi_u(n)-\Phi_u(n-1). \tag{29}$$

Another procedure can be derived using equation (23) and by approximating the derivation with a suitable manner. As in the previous part, the five-point Savitzky-Golay smoothing derivative algorithm can also be applied in this case.

4. Comparison of the instantaneous parameters computed with the Teager-operator and the HHT

4.1. The reconstruction of the signal from its instantaneous parameters

As it was presented in the second part of this paper, the absolute value of the amplitude and the frequency of the slowly varying signal can be estimated by using two algorithm-pairs. In the third part the basis of the estimation of the instantaneous amplitude was the analytic signal computed from the IMFs and the estimation of the instantaneous frequency was determined either directly or from the instantaneous phase sequence. For these estimations two algorithm-pairs have also been given. Because these algorithms have been derived using very different signal models, it is necessary to compare the similarity or dissimilarity of the estimations of the instantaneous parameters. There are four corresponding algorithm-pairs to be compared. In order to compare these algorithm-pairs, it is necessary to estimate the instantaneous parameters and from these to re-estimate the original speech signal by using the same reconstruction algorithm. In the case of the original signal *x(n)* and the estimated signal $\tilde{x}(n)$, the performance of the algorithm-pairs can be characterized by the noise-to-signal ratio below:

$$\mathrm{NSR} = 10 \cdot \log\left(\frac{\sum_{n=4}^{N-5} \left[x(n) - \tilde{x}(n)\right]^2}{\sum_{n=4}^{N-5} x^2(n)} \right) \qquad (30)$$

Because only one algorithm estimates the phase directly and all the others estimate the instantaneous frequency, the basis sequence in reconstruction was in all cases the estimated instantaneous frequency and the determination of the phase sequence was the following:

$$\tilde{\Phi}(k) = \tilde{\Phi}(-1) + \sum_{n=0}^{k} \tilde{\omega}(n) \qquad k = 0,1,\ldots,N-1 \qquad (31)$$

According to our numerical experiments there is a phase-jitter between the original and the reconstructed signal, so the best initial phase value Φ(-1) has been determined by searching the best NSR using π/180 (1°) phase-step.

4.2. The comparison of the method using a test signal

For the test signal the following AM-FM signal, which can be found in the relevant literature, has been used [2]:

$$s(n) = (0{,}998)^n \cdot \left[1 + 0{,}8 \cdot \cos(2 \cdot \pi \cdot f_3 \cdot n)\right] \cdot \cos\left[2 \cdot \pi \cdot \left(f_1 \cdot n + \frac{1}{2 \cdot \pi} \cdot \sin(2 \cdot \pi \cdot f_2 \cdot n)\right)\right] \qquad (32)$$

$$f_s = 10000\mathrm{Hz} \quad f_1 = \frac{1000\mathrm{Hz}}{f_s} \quad f_2 = \frac{100\mathrm{Hz}}{f_s} \quad f_3 = \frac{50\mathrm{Hz}}{f_s}$$

By examining the time-domain figure of this signal, it is clearly an IMF, so it is expected from the EMD-algorithm to give back only one IMF. This is also the case, as it can be seen in *Figure 1*. The difference between the reconstructed and the original signal can be characterized numerically, as it can be found in *Table 1*.

4.3. The comparison of the methods in the case of band-limited speech

In the case of the speech signal it is necessary to ensure the slow changes of the parameters to be estimated, which can be solved by suitable band-pass filtering. Although – to our best knowledge – there is no accepted method for designing the suitable filter, it can be deduced from the relevant literature that a member of some 1 CB filter bank is suitable for the application of the Teager-operator [4] which is the basis for the above mentioned two methods. In this part the estimation of the instantaneous parameters of band-limited speech signal is illustrated.

The speech samples stem from the utterance of the Hungarian word '*igen*', uttered by a native male speaker, using 8 kHz sampling frequency and 16 bit linear quantization. The original utterance was band-limited to 300 Hz...3400 Hz by using a linear-phase FIR filter. After examining the spectrogram, because of the presence of a strong formant near 500 Hz, a member of a perceptual wavelet filter-bank has been used for further FIR-filtering [9]. By examining the band-limited signal, it can also be clearly seen that it is an IMF too, so it is expected from the EMD-algorithm to give back only one IMF. This is also the case, as it can be seen in *Figure 2*. The difference between the reconstructed and the original signal can be characterized numerically, as it can be found in *Table 2*. The best result has been given by the HHT (phase-difference) method.

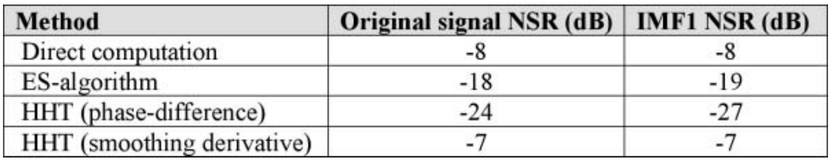
In Figure 2 and in the case of the Teager-operator-based methods the deep valley at 177 ms is caused by the realized program, because it gives back 0 value for the instantaneous frequency in order to avoid the square root from the negative value (see equations (10),(11), (14) and (15)).

4.4. The comparison of the methods in the case of speech signals

The results presented in the previous part show that the Teager-operator-based estimations are very similar to those computed by the HHT-based methods. In the following, our results in analyzing one uttered word are presented. However, there was no band-pass filtering in the investigations below. The EMD process itself serves as an adaptive band-pass filter-bank. The adaptive nature of the algorithm stems from the iterative computing of the upper and lower envelopes. That is, in the case of the first IMF these envelopes are fitted to the

Table 1.

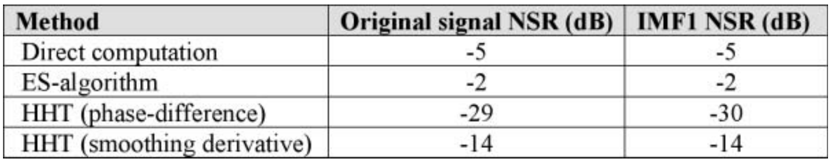
Characterization of the algorithm-pairs in the case of the test signal



Method	Original signal NSR (dB)	IMF1 NSR (dB)
Direct computation	-8	-8
ES-algorithm	-18	-19
HHT (phase-difference)	-24	-27
HHT (smoothing derivative)	-7	-7

Table 2.

Characterization of the algorithm-pairs in the case of the band-limited speech signal



Method	Original signal NSR (dB)	IMF1 NSR (dB)
Direct computation	-5	-5
ES-algorithm	-2	-2
HHT (phase-difference)	-29	-30
HHT (smoothing derivative)	-14	-14

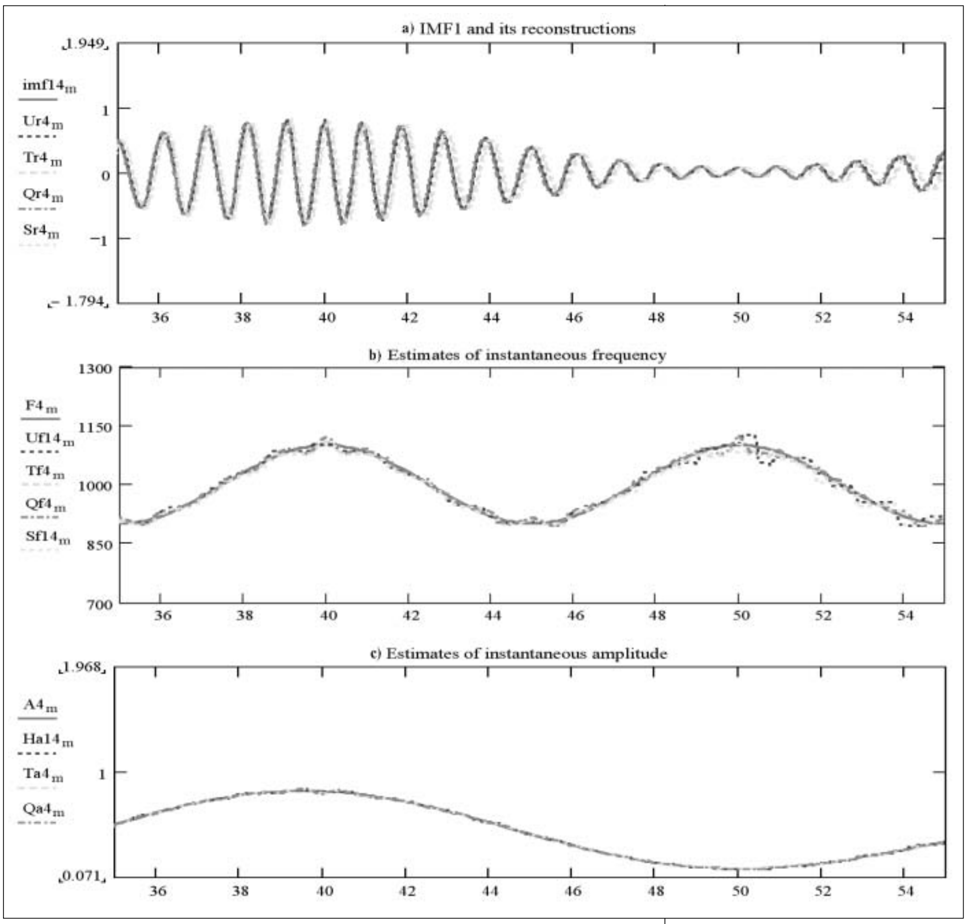


Figure 1.

Results of the application of each algorithm-pair on the test signal.
a) the IMF1 and its estimations
b) the theoretical instantaneous frequency and its estimations
c) theoretical instantaneous amplitude and its estimations

rapid changes in the signal structure, which means the extraction of the higher frequency signal component. After subtracting the first IMF from the original signal, the procedure above is repeated in the lower frequency parts of the signal several times. It is not obvious however, whether this type of filtering is enough for the application of the Teager-operator or not.

This question has also been examined with the utterance of the Hungarian word analyzed in the previous part. It has been mentioned in the third part that there is no basis for the number of the IMFs. However, according to our numerical experiments, by using the first three IMFs the original speech can be reconstructed with NSR of -22 dB, so the instantaneous parameters have been estimated in the case of the first three IMF using the four methods mentioned in the previous part of the paper. The reconstruction itself has been accomplished for these three IMFs and the reconstructed speech has been computed by summation of the reconstructed IMFs.

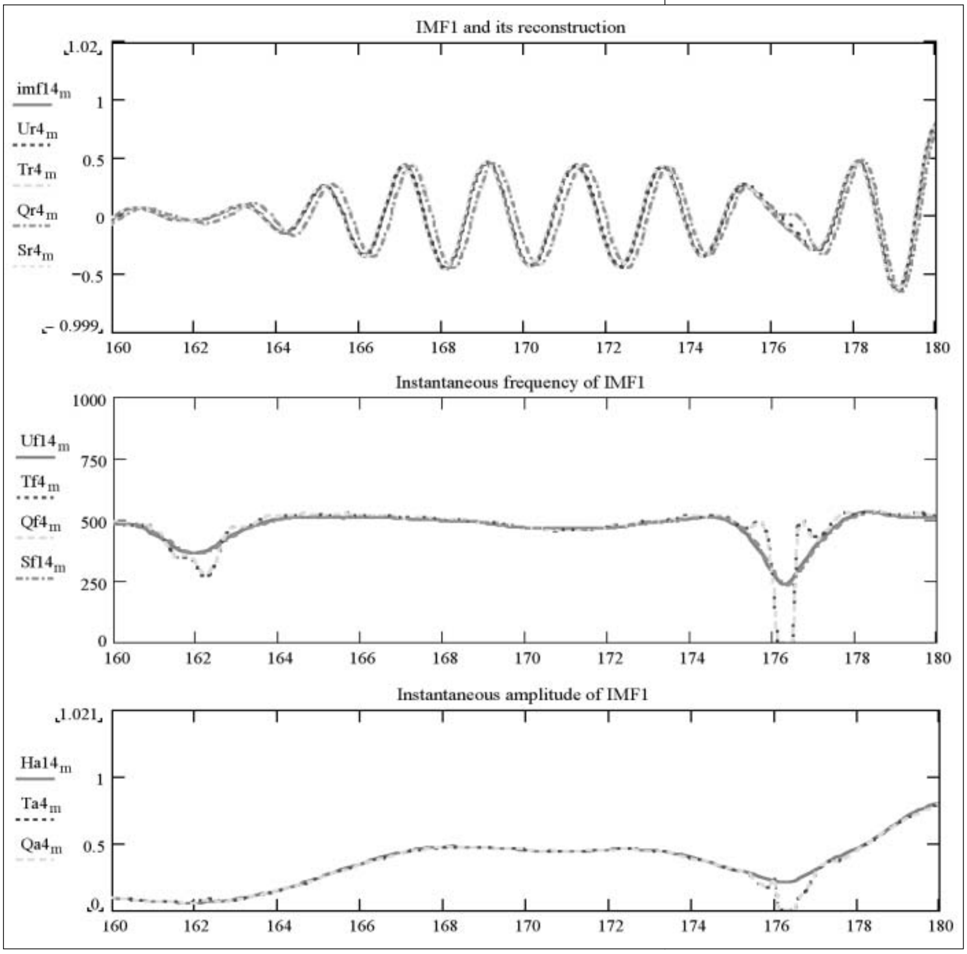


Figure 2.

Results of the application of each algorithm-pair on the band-limited speech signal.
a) the IMF1 and its estimations
b) the estimations of the instantaneous frequency
c) the estimations of the instantaneous amplitude

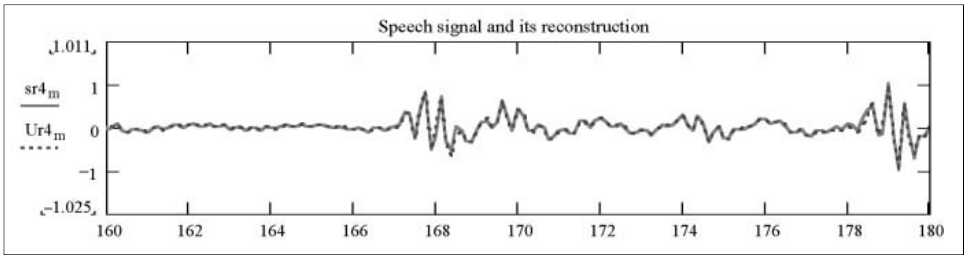


Figure 3.
Reconstruction of the speech signal from the first three reconstructed IMFs.

Method	Original signal NSR (dB)	IMF1 NSR (dB)	IMF2 NSR (dB)	IMF3 NSR (dB)
HHT (phase-difference)	-12	-10	-19	-24

Table 3.
Data in the case of best reconstruction

For the ease of survey *Figure 3* presents the parts of the speech computed in the best case, and in the *Table 3* the numerical values can be seen.

5. Conclusions

In this paper four methods have been proposed for the estimation of the instantaneous amplitude and instantaneous frequency of the speech signal. Two of these methods are based on the Teager-operator, and the others are based on the HHT. The methods have been illustrated with figures computed using a test signal and a speech signal, moreover, a reconstruction method has also been proposed in order to re-compute the speech from the instantaneous parameters. The reconstruction method was also the basis for the comparison of the methods mentioned above.

Our most important result is, that the Teager-operator based methods and the HHT-based methods give similar estimates for the instantaneous parameters of the IMFs of speech. It is planned to continue the work to discover the application areas of the algorithms presented in this paper.

Acknowledgement

The author would like to thank for their kind help and encouragement in his speech processing algorithm development work to Géza Gordos, Géza Németh and Péter Tatai (BME VIK TMIT).

References

[1] Gordos G., Takács Gy.:
Digitális beszédfeldolgozás (in Hungarian).
Műszaki Könyvkiadó, 1983.

[2] Quatieri, T.F.:
Discrete-time Speech Signal Processing:
Principles and Practice.
Prentice-Hall, 2001.

[3] Abbate, A., DeCusatis, M.C., Das, K.P.:
Wavelets and Subbands:
Fundamentals and Applications.
Birkhäuser, 2002.

[4] Chen, S-H., Wang, J-F.:
Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator.
Journal of VLSI Signal Processing 36, pp.125–139., Kluwer Academic Publishers, 2004.

[5] Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N-C., Tung, C.C., Liu, H.H.:
The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. A (1998) 454, pp.903–995.

[6] Valkó P. Vajda S.:
Műszaki-tudományos feladatok megoldása személyi számítógéppel (in Hungarian).
Műszaki Könyvkiadó, 1987.

[7] Gábor, D.:
Theory of communication.
Journal Inst. Electr. Eng. Vol. 93., pp.429–457., 1946.

[8] Simonyi E.:
Digitális szűrők – a digitális jelfeldolgozás alapjai.
Műszaki Könyvkiadó, 1984. (in Hungarian)

[9] Pintér, I.:
Perceptual wavelet-representation of speech signals and its application to speech enhancement.
Computer, Speech and Language, Vol. 10. No.1, pp.1–22., Academic Press, 1996.