Instantaneous Harmonic Analysis: Techniques and Applications to Speech Signal Processing

Alexander Petrovsky and Elias Azarov

Belarusian State University of Informatics and Radioelectronics,
Department of Computer Engineering, Minsk, Belarus
{palex,azarov}@bsuir.by

Abstract. Parametric speech modeling is a key issue in various processing applications such as text to speech synthesis, voice morphing, voice conversion and other. Building an adequate parametric model is a complicated problem considering time-varying nature of speech. This paper gives an overview of tools for instantaneous harmonic analysis and shows how it can be applied to stationary, frequency-modulated and quasiperiodic signals in order to extract and manipulate instantaneous pitch, excitation and spectrum envelope.

Keywords: Speech processing, instantaneous frequency, harmonic model.

1 Introduction

There are many speech processing applications that require parametric representation of the signal. One of the most popular multipurpose approaches for flexible speech processing is hybrid stochastic/deterministic parameterization [1,2]. According to it the signal is decomposed into two parts of different nature: stochastic part (unvoiced speech) can be modeled as a random process with given power spectral density, while deterministic part (voiced speech) is a quasiperiodic signal that can be represented using harmonic modeling. The harmonic model assumes that the signal is a sum of sines with slowly varying parameters.

In this paper we briefly describe some methods for harmonic parameters estimation that can be applied for speech analysis. To be consistent with term ‘instantaneous’ it is assumed that the signal is a continuous function $s(t)$ that can be represented as a sum of P harmonic components with instantaneous amplitude $A_p(t)$, frequency $\omega_p(t)$ and phase $\varphi_p(t)$ [3]:

$$s(t) = \sum_{p=1}^{P} A_p(t) \cos\varphi_p(t),$$

where $\varphi_p(t) = \int_0^t \omega_p(t)dt + \varphi_p(0)$ and $\omega_p \in [0, \pi]$ (for discrete-time signals π corresponds to the Nyquist frequency). In speech processing it is assumed that frequency trajectories of separate harmonics are close to integer multiplies of pitch

A. Ronzhin et al. (Eds.): SPECOM 2014, LNAI 8773, pp. 24–33, 2014.

(or fundamental frequency), i.e. $\omega_p(t) \approx p\omega_1(t)$. Since parameters of the model vary slowly it is possible to assume that each component is narrow-band.

An alternative to instantaneous modeling is frame-based modeling, i.e. when harmonic parameters are assumed to be stationary over whole analysis frame. The simplest way to show the difference between these two approaches is to extend the modeling signal beyond analysis window as shown in figure 1. In frame-based modeling analysis frame is repeated while in the other case the signal is extended according to the parameters corresponding to a specified moment of time.

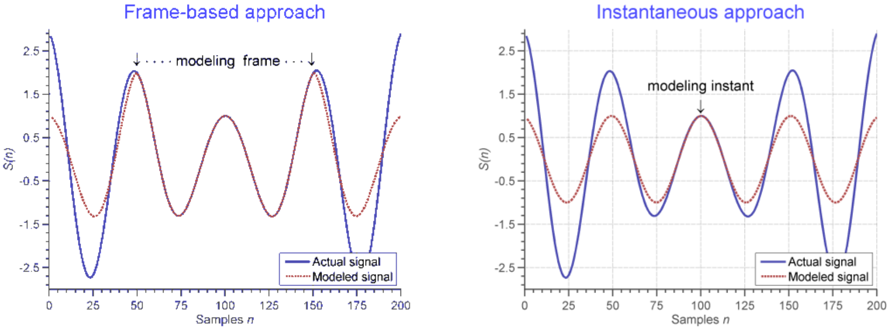

Fig. 1. Signal extension using frame-based and instantaneous modeling

An issue of frame-based approach is aliasing that emerges during synthesis stage. A classical overlap and add method applied in different speech processing systems [4,5] reduces amount of aliasing noise by using concatenation windows. However this effect is not avoided completely because the method cannot ensure that each harmonic is a narrow-band component. Instantaneous harmonic modeling allows filtering and manipulating of each harmonic and therefore can be theoretically more beneficial for voiced speech synthesis.

The present work gives a review of recent approaches to instantaneous harmonic analysis of voiced speech. Despite that we consider input signal as a continues-time function all the analysis techniques presented in the paper can be applied to discrete-time signals as well. We also present some approaches to pitch and spectral envelope extraction based on the harmonic model.

2 Estimation of Instantaneous Harmonic Parameters of Speech

2.1 The Fourier and Hilbert Transform

Most of the analysis techniques require separation of individual harmonics before extraction of instantaneous components. A combination of the Fourier and Hilbert transform can do both separation and extraction. Let us assume that harmonic components do not intersect in frequency domain and therefore can be separated by narrow-band filtering. A good practical approach is to use linear phase filters that can be implemented as a filter bank. Let ω_1 and ω_2 are normalized frequencies from range $[0,\pi]$ that specify bottom and top edges of a pass-band. Then continuous impulse response of the correspondent filter $h(t)$ can be derived as follows:

$$h_{\omega_1,\omega_2}(t) = \frac{1}{\pi}\int_0^{\omega_2} e^{-j\omega t}d\omega - \frac{1}{\pi}\int_0^{\omega_1} e^{-j\omega t}d\omega =$$

$$= \frac{e^{-j\omega t}}{-jt\pi}\bigg|_0^{\omega_2} - \frac{e^{-j\omega t}}{-jt\pi}\bigg|_0^{\omega_1} = \frac{e^{-j\omega_1 t} - e^{-j\omega_2 t}}{jt\pi}.$$

Substituting ω_1 and ω_2 with center frequency ω_c and wideness of the pass-band $2\omega_\Delta$ i.e. $\omega_1 = \omega_c - \omega_\Delta$ and $\omega_2 = \omega_c + \omega_\Delta$ the equation becomes:

$$h_{\omega_1,\omega_2}(t) = \frac{e^{-j\omega_c t}e^{j\omega_\Delta t} - e^{-j\omega_c t}e^{-j\omega_\Delta t}}{jt\pi} = \frac{e^{-j\omega_c t}(e^{j\omega_\Delta t} - e^{-j\omega_d t})}{jt\pi} =$$

$$2\frac{\sin(\omega_\Delta t)}{t\pi}e^{-j\omega_c t}.$$

If the output of the filter is a one periodic component with time-varying parameters then it can be written in the following way:

$$s_{\omega_1,\omega_2}(t) = s(t) * h_{\omega_1,\omega_2}(t) = A(t)e^{j\varphi(t)},$$

where $\mathrm{A}(t)$ is instantaneous amplitude, $\varphi(t)$ – instantaneous phase and $\omega(t)$ – instantaneous frequency. Considering that $s_{\omega_1,\omega_2}(t)$ is a complex analytical signal its parameters can be calculated directly using the following equations:

$$A(t) = \sqrt{R^2(t) + I^2(t)},$$

$$\varphi(t) = \arctan\left(\frac{-I(t)}{R(t)}\right),$$

$$\omega(t) = \varphi'(t),$$

where $R(t)$ and $I(t)$ are real and imaginary parts of $s_{\omega_1,\omega_2}(t)$ respectively.

To get an impulse response with finite length it is possible to use window tion $w(t)$:

$$h_{\omega_1,\omega_2}(t) = 2\frac{\sin(\omega_\Delta t)}{t\pi}w(t)e^{-j\omega_c t}.$$

The method that has been shortly described above applies Hilbert transform to each subband signal. If the filters are uniform (which is generally the case for quasi periodic signals) in real-life applications analysis routine can be implemented very efficiently using fast Fourier transform (FFT). That makes this approach very popular for speech processing applications [6,7].

The accuracy of harmonic separation significantly degrades in case of pitch modulations. The technique requires long analysis window that results in spectral smoothing when frequencies of harmonics change too fast. One of possible solutions to the problem is to use the filter with frequency-modulated impulse response:

$$h_{\omega_1,\omega_2}(t) = 2\frac{\sin(\omega_\Delta(t - t_0))}{(t - t_0)\pi} w(t - t_0)e^{-j\varphi_c(t,t_0)},$$

where $\varphi_c(t, t_0) = \int_{t_0}^{t} \omega_c(t)dt$ and t_0 – the instant of harmonic parameters extraction. In real-life applications required trajectory of center pass-band frequency $\omega_c(t)$ can be estimated from pitch contour. Direct recalculation of impulse response for each estimation instant and each subband is computationally inefficient. Another way to get a similar effect of improving frequency resolution for pitch-modulated signals is to use time-warping. A warping function is applied to the input signal:

$$s_{wrp}(t) = s\left(\varphi_c^{-1}(t, 0)\right),$$

which adaptively warps time axis of the signal and eliminates pitch modulations [8,9]. Since pitch becomes constant it is possible now to apply an efficient FFT-based analysis scheme that has been described above.

2.2 Energy Separation Algorithm and Prony's Method

Energy Separation Algorithm

The Hilbert transform that has been used in the previous subsection for harmonic parameters extraction is not the only one possible option. Another popular approach is the energy separation algorithm (ESA) [10] which is based on the nonlinear differential Teager-Kaiser Energy Operator (TEO) [11]:

$$\Psi[s(t)] \triangleq \dot{s}^2(t) - s(t)\ddot{s}(t),$$

where $\dot{s}(t) = ds(t)/dt$.

According to ESA two TEO's outputs are separated into amplitude modulation and frequency modulation components. As shown in [12] the third-order energy operator

$$\Upsilon_3[s(t)] \triangleq s(t)s^{(3)}(t) - \dot{s}(t)\ddot{s}(t),$$

where $s^{(3)}(t) = d^3s(t)/dt^3$, can be used for estimating damping factor.

Considering that for a periodical signal with constant amplitude and frequency $s(t) = A\cos(\omega t + \theta)$ the following equations are true:

$$\Psi[s(t)] = A^2\omega^2,$$

$$\Psi[\dot{s}(t)] = A^2\omega^4,$$

instantaneous frequency and absolute value of amplitude can be obtained as follows:

$$\omega(t) = \sqrt{\frac{\Psi[\dot{s}(t)]}{\Psi[s(t)]}},$$

$$|A(t)| = \frac{\Psi[s(t)]}{\sqrt{\Psi[\dot{s}(t)]}}.$$

These equations constitute energy separation algorithm for continuous signals.

Prony's Method for Continuous-Time Signals

Despite the fact that Prony's method is originally intended for discrete-time data it is possible to apply it to continuous-time signals as well. Let us consider a continuous signal $s(t)$ which can be represented as a sum of damped complex exponents:

$$s(t) = \sum_{k=1}^{p} h_k\, z_k^t,$$

where p is the number of exponents, $h_k = A_k e^{j\theta_k}$ is an initial complex amplitude and $z_k = e^{\alpha_k + j\omega_k}$ is a time-dependent damped complex exponent with damping factor α_k and normalized angular frequency ω_k. Then let us introduce a time shift t_0 and obtain n-th order derivatives of $s(t)$ [13]:

$$s^{(n)}(t) = \left(\sum_{k=1}^{p} h_k\, z_k^{t-t_0}\right)^{(n)} = \sum_{k=1}^{p} h_k\,(\alpha_k + j\omega_k)^n z_k^{t-t_0} = \sum_{k=1}^{p} l_k(t)\, y_k^n,$$

where (n) denotes order of derivative, $l_k(t) = h_k z_k^{t-t_0} = A_k e^{\alpha_k(t-t_0)+j(\theta_k+\omega_k(t-t_0))}$, $y_k = (\alpha_k + j\omega_k) = e^{(\ln|y_k| + j\arg(y_k))}$.

According to the equation for any fixed moment of time $t = t_0$ series of derivatives $s, \dot{s}, \ddot{s}, \dots, s^{(n)}$ can be represented as a sum of damped complex exponents with initial complex amplitudes $l_k(t_0) = h_k$, damping factors $\ln|y_k|$ and normalized angular frequencies $\arg(y_k)$. The required parameters of the model h_k and y_k can be found using original Prony's method as it is briefly summarized below.

In order to estimate exact model parameters $2p$ complex samples of the sequence are required. The solution is obtained using the following system of equations:

$$\begin{pmatrix} y_1^0 & y_2^0 & \cdots & y_p^0 \\ y_1^1 & y_2^1 & \cdots & y_p^1 \\ \vdots & \vdots & & \vdots \\ y_1^{p-1} & y_2^{p-1} & \dots & y_p^{p-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{pmatrix} = \begin{pmatrix} s \\ \dot{s} \\ \vdots \\ s^{(p-1)} \end{pmatrix}.$$

The required exponents $y_1, y_2, \dots, y_p$ are estimated as roots of the polynomial

$$\psi(z) = \sum_{m=0}^{p} a_m z^{p-m}$$

with complex coefficients a_m which are the solution of the system

$$\begin{pmatrix} s^{(p-1)} & s^{(p-2)} & \dots & s \\ s^{(p)} & s^{(p-1)} & \cdots & \dot{s} \\ \vdots & \vdots & & \vdots \\ s^{(2p-2)} & s^{(2p-3)} & \dots & s^{(p-1)} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = -\begin{pmatrix} s^{(p)} \\ s^{(p+1)} \\ \vdots \\ s^{(2p-1)} \end{pmatrix}$$

and $a_0 = 1$. Each damping factor α_k and frequency ω_k are calculated using the following equations:

$$\alpha_k = \mathrm{Re}(y_k), \omega_k = \mathrm{Im}(y_k).$$

Using the extracted values of $y_1, y_2, \dots, y_p$ the initial system is solved with respect to $h_1, h_2, \dots, h_p$. From each of these parameters initial amplitude A_k and phase θ_k are calculated as:

$$A_k = |h_k|, \theta_k = \arctan\left[\frac{\mathrm{Im}(h_k)}{\mathrm{Re}(h_k)}\right].$$

For real-valued signals the solution gives pairs of complex conjugate exponents. In order to identify parameters of b real-valued sinusoids we should calculate $4b-1$ derivatives.

Considering $s(t)$ as a single real-valued damped sinusoid it is possible to identify its parameters using its actual value and three derivatives. Using the equations that have been given above we can formulate the following estimation algorithm.

1) Calculate three derivatives of the signal: $\dot{s}, \ddot{s}, s^{(3)}$;

2) Calculate coefficients of the polynomial:

$$a_1 = \frac{s s^{(3)} - \dot{s}\ddot{s}}{\dot{s}^2 - s\ddot{s}} = \frac{\Upsilon_3[s]}{\Psi[s]},$$

$$a_2 = \frac{\ddot{s}^2 - \dot{s}s^{(3)}}{\dot{s}^2 - s\ddot{s}} = \frac{\Psi[\dot{s}]}{\Psi[s]};$$

3) Calculate roots of the polynomial:

$$y_{1,2} = \frac{1}{2}\left(-a_1 \pm \sqrt{a_1^2 - 4a_2}\right) = -\frac{\Upsilon_3[s]}{2\Psi[s]} \pm \sqrt{\frac{\Upsilon_3^2[s]}{4\Psi^2[s]} - \frac{\Psi[\dot{s}]}{\Psi[s]}};$$

4) Calculate initial complex amplitude:

$$h = \frac{s y_2 - \dot{s}}{y_2 - y_1} = \frac{1}{2}\left(s + \frac{\frac{\Upsilon_3^2[s]}{2\Psi[s]} s + \dot{s}}{\sqrt{\frac{\Upsilon_3^2[s]}{4\Psi^2[s]} - \frac{\Psi[\dot{s}]}{\Psi[s]}}}\right);$$

5) Calculate required parameters of the sinusoid:

$$\alpha = \mathrm{Re}(y_1) = -\frac{\Upsilon_3[s]}{2\Psi[s]},$$

$$\omega = \mathrm{Im}(y_1) = \sqrt{\frac{\Psi[\dot{s}]}{\Psi[s]} - \frac{\Upsilon_3^2[s]}{4\Psi^2[s]}},$$

$$A = 2|h|, \qquad \theta = \arctan\left[\frac{\mathrm{Im}(h)}{\mathrm{Re}(h)}\right].$$

Note that the resulting equation for damping factor is exactly the same as given in [12] and the equation for frequency can be derived from the case of cosine with exponential amplitude discussed in [10]. The equations show how ESA and Prony's method are connected in the case of one real-valued sinusoid.

3 Estimation of Pitch and Spectral Envelope from Instantaneous Harmonic Parameters

In this section we show how high-level speech characteristics such as pitch and spectral envelope can be estimated from instantaneous harmonic parameters.

3.1 Instantaneous Pitch Estimation

The most popular approach for period candidate generating is autocorrelation-based functions such as normalized cross-correlation function (NCCF). Let $s(m)$ be a discrete-time speech signal, z – step size in samples and n – window size. The NCCF $\phi(x,k)$ of K samples length at lag k and analysis frame x is defined as [14]:

$$\phi(x,k) = \frac{\sum_{i=m}^{m+n-1} s(i)s(i+k)}{\sqrt{e_m e_{m+k}}}, k = 0, K-1;\ m = xz;\ x = 0, M-1,$$

where $e_i = \sum_{l=i}^{i+n-1} s^2(l)$. Instantaneous parameters of harmonic model give a spectral representation of the current instant $s(t)$ that can be utilized in order to estimate momentary autocorrelation function $R_{inst}(t,\Delta t)$. Using the Wiener-Khintchine theorem:

$$R_{inst}(t,\Delta t) = \frac{1}{2}\sum_{p=1}^{P} A_p^2\,(t)\cos\big(\omega_p(t)\Delta t\big).$$

$R_{inst}(t,\Delta t)$ corresponds to the autocorrelation function calculated on infinite window of periodic signal generated with specified harmonic parameters. As far as analysis window is infinite there is no difference between autocorrelation and cross-correlation functions. Considering this fact it is possible to propose the instantaneous version of NCCF $\phi_{inst}(t,\Delta t)$ in the following form:

$$\phi_{inst}(t,\Delta t)=\frac{\sum_{p=1}^{P}A_p^2\,(t)\cos(\omega_p(t)\Delta t)}{\sum_{p=1}^{P}A_p^2\,(t)}.$$

Unlike original time-domain NCCF lag Δt does not need to be an integer, valid values can be produced for any desired frequency. Function $\phi_{inst}(t,\Delta t)$ is immune to any rapid frequency modulations in the neighborhood of t provided that estimated instantaneous harmonic parameters are accurate enough. This period candidate generating function has been used in instantaneous pitch estimator [15], based on the harmonic model.

3.2 Estimation of Instantaneous Spectral Envelope

Let us use conventional linear-prediction (LP) technique for spectral envelope estimation of continuous-time signal $s(t)$. We assume that harmonic model of the signal is specified by the correspondent set of time-varying parameters. LP model approximates given signal sample $s(n)$ as a linear combination of the p past samples that leads to the following equality:

$$s(n)=\sum_{i=1}^{p}a_i\,s(n-i)+Gu(n),$$

where a_1, a_2, …, a_p are prediction coefficients, $u(n)$ is a normalized excitation and G is the gain of the excitation [16]. The prediction error $e(n)$ is defined as the difference between the source and predicted samples:

$$e(n)=s(n)-\tilde{s}(n)=s(n)-\sum_{k=1}^{p}a_k s(n-k).$$

The basic problem of LP is to find the set of predictor coefficients that minimize the mean-square prediction error. Let us consider a harmonic signal with constant amplitudes and constant frequencies of components. The relative residual energy can be evaluated as the following sum:

$$E_a^2=\sum_{k=1}^{K}A_k(n)^2\left(\left[1-\sum_{i=1}^{p}a_i\cos(\omega_k(n)i)\right]^2+\left[\sum_{i=1}^{p}a_i\sin(\omega_k(n)i)\right]^2\right).$$

In order to minimize E_a^2 it is possible to use the basic minimization approach by finding partial derivatives with respect to variables a_i and then solving the system of linear equations. Eventually the following system can be derived:

$$\sum_{i=1}^{p}a_i q(|i-j|)=-q(j),$$

where $j=1,2,\dots,p$ and $q(l)=\sum_{k=1}^{K}A_k(n)\cos(f_k(n)l)$, $(l\geq 0)$.

It is known that LP spectral representation tends to model individual harmonic components instead of the spectral envelope when the order of prediction becomes high. Using derived transformation system it is possible to represent exactly the specified envelope as a high-order filter by using amplitude and frequency vectors of infinite dimension.

The spectral envelope can be considered as a continuous function of frequency $A(\omega)$, specified on the interval $[0,\pi]$. Then the matrix elements $q(l)$ can be derived as the following integral:

$$q(l) = \int_{0}^{\pi} A(\omega)\cos(\omega l)d\omega.$$

If $A(\omega)$ contain discontinues in points $\omega_d = (\omega_1, \omega_2, \dots, \omega_I)$, then the equation can be expressed as:

$$q(l) = \sum_{i=1}^{I+1} \int_{\bar{\omega}_{d,i}}^{\bar{\omega}_{d,i+1}} A(\omega)\cos(\omega l)d\omega,$$

where $\bar{\omega}_d = (0, \omega_1, \omega_2, \dots, \omega_I, \pi)$.

Continuous spectral envelope can be estimated from amplitude and frequency vectors using linear interpolation. Single segments of the envelope $f_i \le \omega \le f_{i+1}$, $1 \le i \le K-1$ are described by linear equations of the form $A(\omega) = b_i\omega + c_i$. Parameters b_i and c_i are estimated from adjacent values of frequency and amplitudes. Finally elements of the required system can be derived in the following way:

$$q(l) = \sum_{i=1}^{K-1} D(l,i),$$

where $D(l,i) = \begin{cases} \frac{b}{l^2}[\cos(f_{i+1}l) + f_{i+1}l\sin(f_{i+1}l)] + \frac{c}{l}\sin(f_{i+1}l) - & \\ -\frac{b}{l^2}[\cos(f_il) + f_il\sin(f_il)] - \frac{c}{l}\sin(f_il) & l \neq 0 \\ \frac{1}{2}bf_{i+1}^2 + cf_{i+1} - \frac{1}{2}bf_i^2 - cf_i & l = 0. \end{cases}$

The presented technique is compared to original LP in [17] where was shown that it provides much more accurate envelope estimation compared to conventional time-domain method such as autocorrelation and covariance.

4 Conclusions

A short review of techniques for instantaneous harmonic analysis has been given in the paper. The techniques can be applied to voiced speech in order to extract time-varying parameters of each harmonic. The extracted parameters can be used for instantaneous pitch and envelope estimation.

References

1. Laroche, J., Stylianou, Y., Moulines, E.: HNS: Speech modification based on a harmonic+noise model. In: Acoustic, Speech, and Signal Processing: Proceedings of IEEE International Conference ICASSP 1993, Minneapolis, USA, pp. 550–553 (April 1993)
2. Levine, S., Smith, J.: A sines+transients+noise audio representation for data compression and time/pitch scale modifications. In: Proceedings of 105th AES Convention on Signal Processing, San Francisco, USA, p. 21 (1998) (preprint no 4781)
3. McAulay, R.J., Quatieri, T.F.: Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans. on Acoust., Speech and Signal Processing ASSP-34, 744–754 (1986)
4. Moulines, E., Charpentier, F.: Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication 9(5-6), 453–467 (1990)
5. Kawahara, H., Takahashi, T., Morise, M., Banno, H.: Development of exploratory research tools based on TANDEM-STRAIGHT. In: Proc. of the APSIPA, Japan Sapporo (2009)
6. Flanagan, J.L., Golden, R.M.: Phase vocoder. Bell System Technical Journal 45, 1493–1509 (1966)
7. Abe, T., Honda, M.: Sinusoidal model based on instantaneous frequency attractors. IEEE Trans. on Audio, Speech, and Language Processing 14(4), 1292–1300 (2006)
8. Nilsson, M., Resch, B.: Kim Moo-Young, Kleijn, W.B.: A canonical representation of speech. In: Proc. of the IEEE ICASSP 2007, Honolulu, USA, pp. 849–852 (2007)
9. Azarov, E., Vashkevich, M., Petrovsky, A.: GUSLAR: A framework for automated singing voice correction. In: Proc. of the IEEE ICASSP 2014, Florence, Italy, pp. 7969–7973 (2014)
10. Maragos, P., Kaiser, J.F., Quatieri, T.F.: Energy separation in signal modulations with application to speech analysis. IEEE Trans. Signal Processing 41, 3024–3051 (1993)
11. Kaiser, J.F.: On a simple algorithm to calculate the 'energy' of a signal. In: Proc. of the IEEE ICASSP 1990, Albuquerque, NM, pp. 381–384 (1990)
12. Maragos, P., Potamianos, A., Santhanam, B.: Instantaneous energy operators: applications to speech processing and communications. In: Proc. of the IEEE Workshop on Nonlinear Signal and Image Proc., Thessaloniki, Greece (1995)
13. Azarov, E., Vashkevich, M., Petrovsky, A.: Instantaneous harmonic representation of speech using multicomponent sinusoidal excitation. In: Proc. of the Interspeech 2013, Lyon, France, pp. 1697–1701 (2013)
14. Talkin, D.: A Robust Algorithm for Pitch Tracking (RAPT). In: Kleijn, W.B., Paliwal, K.K. (eds.) Speech Coding & Synthesis. Elsevier (1995) ISBN 0444821694
15. Azarov, E., Vashkevich, M., Petrovsky, A.: Instantaneous pitch estimation based on RAPT framework. In: Proc. of the EUSIPCO, Bucharest, Romania (2012)
16. Rabiner, L., Juang, B.H.: Fundamentals of speech recognition. Prentice Hall, New Jersey (1993)
17. Azarov, E., Petrovsky, A.: Linear prediction of deterministic components in hybrid signal representation. In: Proc. of the IEEE International Symposium on Circuits and Systems(ISCAS), Paris (2010)