Computation of Immittance and Line Spectral Frequencies Based on Inter-frame Ordering Property

Vasyl Semenov
Delta SPE, Kiev, Ukraine
Email: semenov@deltacorp.net

***Abstract*— Line spectral frequencies (LSF) and immittance spectral frequencies (ISF) are widely used in modern speech codecs based on autoregressive model. This paper addresses LSF and ISF calculation problems. The investigation of LSF/ISF placement on adjacent quasi-stationary frames is performed. It is shown that in majority of cases the inter-frame ordering property takes place. On this basis a new approach to LSF/ISF calculation is proposed. It is shown that LSF/ISF localization task is mainly reduced to verification of inter-frame ordering property. The computational expenses are reduced in 3.4 times in comparison with widely used Kabal's method. Besides, the maximum number of operations is lower than the minimum expenses of accelerated Kabal's method. The method is implemented on fixed-point DSP and showed stable performance.**

***Index Terms*— speech coding, autoregressive model, line spectral frequencies, rootfinding procedures, inter-frame ordering, immittance spectral frequencies**

I. Introduction

The development of cellular, satellite and IP telephony has led to the necessity of efficient low-bit-rate speech coders. The main reasons for the compression of the speech signals are the limited bandwidth of communication channels and the necessity of information security provision. Along with narrow-band speech coders (with upper frequency below 3.4 kHz) [1], [2], wide-band speech coders (with upper frequency below 7 kHz) also become wide-spread [3], [4].

The development of speech compression methods requires for simple and effective parametric models of speech based on knowledge of speech production and perception mechanisms. A majority of modern speech processing methods are based on autoregressive (AR) model of speech [5]

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + w(n), \qquad (1)$$

where $s(n)$ is a speech signal; $w(n)$ is an excitation modeling air flow at the glottis; $a_k, k = 1, 2, \ldots, p$ are AR coefficients modeling the shape of vocal tract. Traditionally AR coefficients are calculated by linear prediction methods [5]. The order p is usually taken from 8 to 20. It is evident that wide-band speech coding requires higher AR model orders in comparison with narrow-band one.

It is known that AR parameters change relatively slowly and can be considered as constant on time intervals (frames) of approximately 20 ms. Thus, the task of speech encoding is separated on 2 subparts performed once per frame: coding of excitation and coding of AR parameters $a_k, k = 1, 2, \ldots, p$.

However, AR coefficients are not directly used in speech coding devices due to their high spectral sensitivity and wide dynamical range. That is why, AR parameters need to be transformed to some alternative set of parameters for which efficient scalar and vector quantization procedures can be applied. The examples of these parameters are reflection coefficients or log-area ratios [6].

However, the most efficient alternative set of AR parameters became line spectral frequencies introduced by Itakura [7]. LSF are closely connected with formant frequencies and can be coded taking into account specific features of human sound perception [8]. LSF are used in 3G systems [1], IP-telephony [2], vocoders [9]. For example, in 12200 bit/s mode of AMR-NB codec the frames of 20 ms are coded by 244 bits, from which 38 bits are used for the coding of 10 LSF and the rest are used for the excitation.

Immittance spectral frequencies (ISF) were introduced some later [10]. It was shown that they possess more efficient quantization properties compared to LSF [11]. This resulted in their implementation in wide-band AMR [3] and its modifications [12].

Computation of LSF and ISF is connected with rootfinding procedures which are undesired for a majority of computational devices (especially fixed-point DSPs) since it may cause unpredictable delays and errors accumulation. That is why the issues of LSF calculation attracted the attention of many researches [13]–[15]. In work [16] a new LSF computation methods based on developed algorithm for the solution of non-linear equations were introduced. In [17] a property of LSF inter-frame ordering property was introduced and used for the improvement of LSF calculation methods performance.

The aim of this work is to construct a computationally efficient LSF and ISF calculation methods which opti-

This paper is based on "A novel approach to calculation of line spectral frequencies based on inter-frame ordering property" by V. Semenov, which appeared in the Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toulouse, France, May 2006. © 2006 IEEE.

mally exploit the features of their time distribution. The second section reviews existing approaches to LSF/ISF computation. In the third section the investigation of mutual LSF/ISF location on adjacent frames is performed and the condition of inter-frame ordering is analyzed. The fourth section presents a new algorithm of LSF/ISF calculation based on inter-frame ordering property. In the experimental section the proposed methods are compared with method [16] and standard grid approaches.

II. Existing approaches to calculation of line and immittance spectrum pairs

Consider a transformation of initial AR coefficients to LSF or ISF [13]. Both in LSF and ISF calculation procedures initial polynom of AR coefficients

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (2)$$

is transformed to a pair of symmetric polynomials $G^{(m)}(z), m = 1, 2$:

$$G^{(m)}(z) = \sum_{k=0}^{n_g} g_k^{(m)} z^{-k}, m = 1, 2. \qquad (3)$$

The orders of these polynomials are equal to p except for the order of second ISF polynomial equal to $p - 2$. The coefficients of these polynomials are recursively calculated via the initial AR coefficients [13].

All roots of polynomials $G^{(1)}, G^{(2)}$ lie on the unit circle, i.e. have a form $z_k = e^{i\omega_k}$. The angles of roots of these polynomials are called the line spectral frequencies $\omega_1, \omega_2, \ldots, \omega_p$ (or, respectively, immittance spectrum pairs $\eta_1, \eta_2, \ldots, \eta_{p-1}$).

An important property of LSF is their intra-frame ordering

$$\omega_1 < \omega_2 < \ldots < \omega_{p-1} < \omega_p, \qquad (4)$$

where the odd frequencies correspond to $G^{(1)}$ and the even frequencies correspond to $G^{(2)}$. The same property possess also ISF [11].

The polynomials (3) can be written as ($M = n_g/2$):

$$G^{(m)}(z) = \sum_{k=0}^{n_g} g_k^{(m)} z^{-k} = e^{-iM\omega} \sum_{k=0}^{M} r_k^{(m)} \cos(k\omega), \qquad (5)$$

where

$$\begin{cases} r_0^{(m)} &= g_M^{(m)}, \\ r_k^{(m)} &= 2g_{M-k}^{(m)}, k = 1, \ldots, M. \end{cases} \qquad (6)$$

Thus, the rootfinding task for functions (5) is reduced to the calculation of roots of two equations

$$\sum_{k=0}^{M} r_k^{(m)} \cos(k\omega) = 0, m = 1, 2. \qquad (7)$$

Equation (7) can be reduced to a polynomial one with the help of Chebyshev polynomials by substitution $x = \cos(\omega)$ [13]

$$\sum_{k=0}^{M} r_k'^{(m)} x^{M-k} = 0, m = 1, 2. \qquad (8)$$

So, the calculation of LSF and ISF is reduced to the searching of roots of polynomials (8). Below the main approaches for the calculation of roots of these equations are discussed.

Grid methods. These methods structurally consist of two parts: localization of roots and their refinement. The typical representative of the first subgroup is the method of Kabal and Ramachandran [13]. The single-root intervals are determined by evaluation of functions (7) or (8) on a large predefined grid of points. The main disadvantage of this approach is that for every AR order the step of grid must be estimated on a large speech database. At that, if two roots occur close than grid step, this method becomes inapplicable.

Methods of degree reduction. Another subgroup of methods [14], [15], exploits methodology of consequent deflation of polynomials (8) by their maximum root. However, the use of deflation leads to worse accuracy of each subsequent root. This becomes an especially serious problem for high-order polynomials and may lead to instability of synthesis filters [15]. Besides, the division operation is usually undesirable for fixed-point DSPs.

A principally another approach was proposed in paper [16] where authors proposed a new method of LSF calculation based on the developed universal algorithm for the solution of non-linear equations. Although the proposed approach had several essential advantages over existing ones, it did not use physical LSF features which, obviously, must contain a good reserve for the improvement of the method's efficiency.

III. Property of inter-frame ordering

Taking into account the connection of LSF with vocal tract resonances [8], one can assume that LSF on adjacent quasi-stationary frames must not differ too much. In work [16] it was shown that in majority of situations (from 88.9 to 96.8% for different orders of AR model) LSF inter-frame ordering property takes place

$$\omega_{i-1}^{(n-1)} < \omega_i^{(n)} < \omega_{i+1}^{(n-1)}, i = 2, \ldots, p-1, \qquad (9)$$

where the upper indices denote frame numbers.

To illustrate this property, consider a speech signal pronounced by a male speaker and digitized with sampling frequency of 8000 Hz. The spectrogram of this signal is represented at Figure 1. Consider three parts of this signal marked at Figure 1 as ”A”, ”B”, ”C”. Time plots of these segments and corresponding LSF plots are presented at figures 2, 3, and 4 respectively. The calculation of LSF was performed on 20 ms frames for the order of AR model $p = 10$. The LSF were normalized to the Nyquist frequency: $f_k = \omega_k f_s/(2\pi), k = 1, \ldots, p$.

Figure 2 shows the situation corresponding to the end of vowel ”e” and the beginning of consonant ”t”. It can

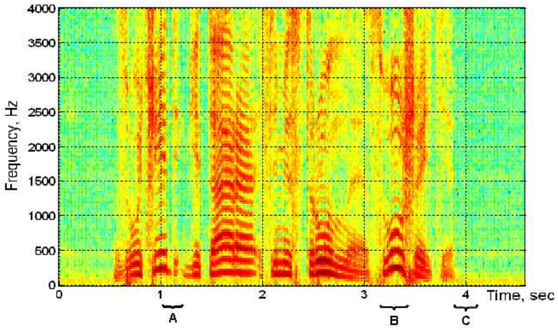

Figure 1. The spectrogram of test speech signal.

be seen that though each LSF vary in quite a wide range, almost in all situations LSF with number i lies between LSF of previous frame with numbers $i-1$ and $i+1$. The only exclusion is the third LSF of seventh frame which is slightly higher (on 4 Hz) than a forth LSF of a previous frame. As is seen from the time plot, this situation corresponds to the beginning of a new sound (consonant ”t”).

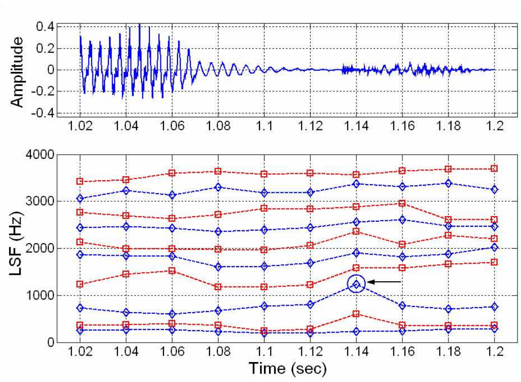

Figure 2. Example of LSF distribution on boundary of two sounds.

Figure 3 shows LSF distribution during a continuous sound ”l” with a very little difference between LSF in the pitch region. On all frames of this speech fragment LSF i lies between the previous frame LSF with numbers $i-1$ and $i+1$, i.e. the inequality (9) is satisfied.

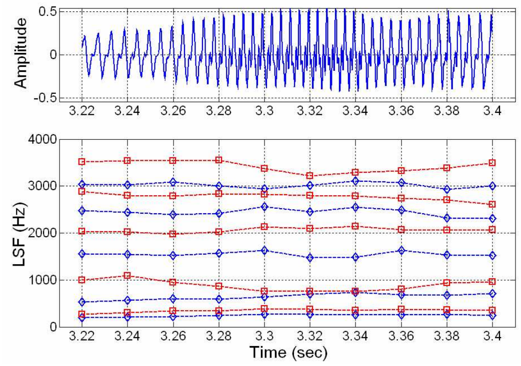

Figure 3. LSF distribution during a continuous sound ”l”.

At Figure 4 one can see the end of the vowel and a following pause filled by background noise. During the decaying of the vowel, the LSF span a relatively wide range. Nevertheless, LSF with number $i-1$ constantly lies between LSF of previous frame with numbers $i-1$ and $i+1$. On the time fragment containing background noise LSF have a uniform almost time-invariant distribution and, obviously, satisfy to inequality (9). The fact that the pauses usually take not less than 40-50 % of speech duration, additionally enforces the assumption that condition (1) is met in a majority of practical situations.

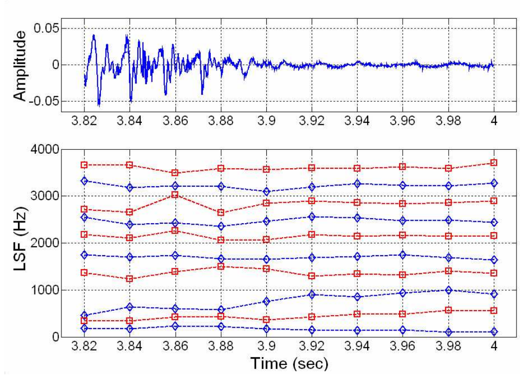

Figure 4. Example of LSF distribution on boundary of vowel and pause.

The inter-frame ordering property was also verified for the immittance spectrum pairs of wideband speech signals (with sampling rate of $f_s = 16000$ Hz). This property was verified on the database of nine speakers with general duration of approximately 16 minutes. AR model orders from 14 to 20 were considered. Table I shows the percentage of cases when property

$$\eta_{i-1}^{(n-1)} < \eta_i^{(n)} < \eta_{i+1}^{(n-1)}, i = 2, \dots, p-2. \qquad (10)$$

is true for all LSF on a current frame (n_0) or it failed for one, two or three LSF (n_1, n_2 and n_3 respectively).

As follows from Table I, inequality (10) is true for a majority of cases. Corresponding percentage lies in the range from 93.2 % at $p = 20$ to 95.2% at $p = 14$. It can therefore be stated that LSF and ISF localization problems are primarily reduced to verification of their inter-frame ordering property.

IV. Resulting algorithm of LSF/ISF computation

According to the above mentioned investigation of LSF and ISF mutual placing, the following algorithm for their computation is proposed. Since there is no principal difference between algorithms for LSF and ISF, the algorithm is formulated for LSF only.

Initial data. AR coefficients of current frame $a_k, k = 1, 2, \dots, p$; values of LSF cosines $x_1^{(n-1)}, x_2^{(n-1)}, \dots, x_{p-1}^{(n-1)}, x_p^{(n-1)}$ calculated at previous frame.

Step 1. Calculation of coefficients of equations (8). Initial AR coefficients $a_k, k = 1, 2, \dots, p$ are transformed to coefficients $r_k'^{(m)}$ of equations (8).

TABLE I.
PERCENTAGE OF CASES WHEN CONDITION (10) IS MET (n_0) AND WHEN IT IS NOT MET (n_1, n_2, n_3)

	$p=14$	$p=16$	$p=18$	$p=20$
n_0	95.21	94.54	93.68	93.17
n_1	3.37	3.66	4.21	4.32
n_2	0.96	1.10	1.20	1.42
n_3	0.32	0.47	0.54	0.60

Step 2. Localization of LSF cosines. The condition (9) is verified by evaluation of signs of first function (8) in points $\{-1, x_2^{(n-1)}, x_4^{(n-1)}, \ldots, x_{p-2}^{(n-1)}, x_p^{(n-1)}, 1\}$ at the calculation of odd roots or by evaluation of signs of second function (8) on grid $\{-1, x_1^{(n-1)}, x_3^{(n-1)}, \ldots, x_{p-3}^{(n-1)}, x_{p-1}^{(n-1)}, 1\}$ at the calculation of even roots.

If for some LSF with number i condition (9) is not met, the presence of root at interval $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$ is verified by the universal method for the solution of non-linear equations [16].

Step 3. Refinement of roots. After the localization of LSF cosines, their exact values are determined by Newton's method.

Step 4. Transformation to ω-domain. LSF are finally calculated by transformation $\omega = \arccos(x)$.

Note 1. The localization step is illustrated by Figure 5, where the plot of the first polynomial (8) is depicted and the cosines of previous frame LSFs are shown. As can be seen, considered function changes sign in points $\{-1, x_2^{(n-1)}, x_4^{(n-1)}, \ldots, x_{p-2}^{(n-1)}, x_p^{(n-1)}, 1\}$ and, thus, necessary number of polynomial function calls is just $(p+2) = 12$ (which is a strong contrast to 111 calls for the Kabal-Ramachandran's method [13]).

Note 2. A "derivative-free version" of algorithm can be obtained if method [16] on the localization stage is replaced by verification of sign changes on a intersection of fixed grid with interval $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$. At the same time, secant or bisection method must be applied on the root refinement stage.

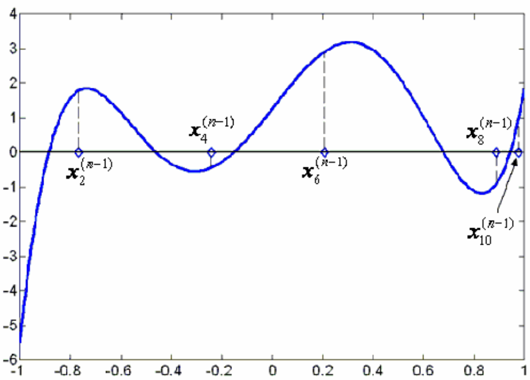

Figure 5. Example of LSF cosines' localization.

V. EXPERIMENTAL RESULTS

The effectiveness of the proposed LSF computation method was verified for different AR model orders p: $8, 10, \ldots, 20$. For this purpose we used a base of five-minute records of four male and two female speakers digitized with $f_s = 8000$ Hz. At first, AR coefficients were computed by autocorrelation method [5] and then LSF were computed for every 20 ms of these signals. Table I shows average computational expenses (in Mflops), corresponding to method described in [16] (in brackets) and to the method proposed in this paper. For the objectivity of comparison initialization of the Newton's method was taken the same as in [16]. The condition $|f(x)| < 10^{-6}$ was taken as a rule for algorithm stop.

From Table II it can be seen that proposed algorithm provides a reduction of computational expenses from 1.7 times (for $p = 8$) to 2.4 times (for $p = 10$). This is explained by more perfect LSF localization procedure proposed in this paper.

Since the test speech signals were characterized by a variety of speakers and were almost free of pauses, provided computational characteristics of proposed method can be considered as corresponding upper bounds. During a real work of speech coding devices expenses must be lower.

Now let's make a comparison of proposed algorithm with Kabal-Ramachandran's method [13], since it is most widely used in speech processing applications. Since Kabal-Ramachandran's method exploits bisection method of root refinement, we also considered its "accelerated" version corresponding to root refinement by Newton's method. Table III shows average, minimal and maximum numbers of operations per frame for the computation of ten LSF by Kabal-Ramachandran's method, its accelerated version, method [16] and the method proposed in this paper. The convergence criterion based on the uncertainty of root position was used: $|x_k - x_{k-1}| < \epsilon$, where x_k, x_{k-1} are the approximate root values obtained at successive iterations.

From Table III follows that proposed algorithm provides reduction of computational expenses in comparison with Kabal-Ramachandran's method in 3.42 times. The gain over accelerated Kabal-Ramachandran's algorithm is equal to 2.38 times. One of the main characteristics of method [16] was that its peak computational expenses were lower than (time-invariant) expenses of Kabal-Ramachandran's method. However, the method proposed in this paper has a stronger property: its maximum number of operations (1428) is lower than that for Kabal-

TABLE II.
COMPUTATIONAL EXPENSES (MFLOPS) OF METHOD [16] (IN BRACKETS) AND PROPOSED METHOD

	$p = 8$	$p = 10$	$p = 12$	$p = 14$	$p = 16$	$p = 18$	$p = 20$
Method [16]	0.046	0.075	0.124	0.172	0.236	0.314	0.381
Proposed method	0.029	0.044	0.061	0.082	0.104	0.129	0.158

TABLE III.
COMPARISON OF LSF COMPUTATION METHODS

	Average number of operations	Minimum number of operations	Maximum number of operations
K.-R. method	2150	2150	2150
Accelerated K.-R. method	1498	1491	1557
Method [16]	1309	1066	1936
Proposed method	629	428	1428

Ramachandran's method (2150), but is also lower than the minimum number of operations for the accelerated Kabal-Ramachandran's method. This fact additionally suggests the advantage of application of proposed method in real-time systems. The method was implemented into a fixed-point ADSP2191 narrow-band vocoder (2.4 kbps). During an hour of continuous testing with different speakers the algorithm failed only in 2 of 144000 cases due to insufficient accuracy of polynomial function evaluation in fixed-point 16-bit arithmetic.

Figure 6 shows the numbers of operations of different methods for the calculation of LSF of test speech signal, the spectrogram of which was shown at Figure 1. This clearly shows the advantage of the proposed method in the context of the number of operations. The minimum computational expenses take place in pauses filled by stationary background noise, while the peaks in the distribution of operations occur at abrupt transitions from one sound to another.

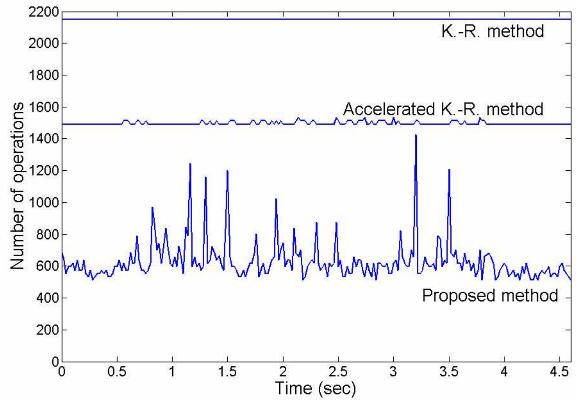

Figure 6. Example of computational expenses distribution for different LSF calculation methods.

Figure 7 demonstrates the application of proposed ISF calculation method ("derivative-free" version) for the test signal digitized with $f_s = 16000$ Hz. It shows the number of Chebyshev function evaluations required for proposed method and grid method used in [3]. It can be seen that proposed method reduces the average number of polynomial functions evaluations in 3.4 times in comparison with grid approach.

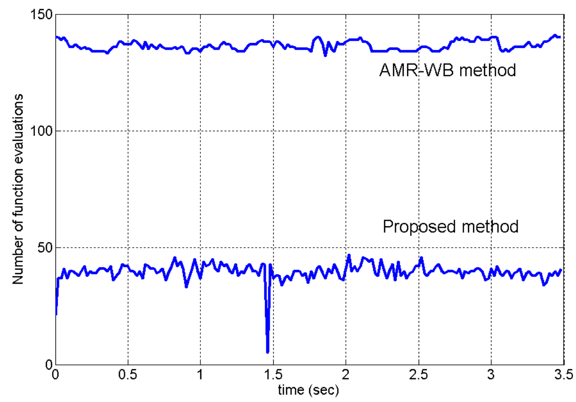

Figure 7. Distribution of polynomial function evaluations for ISF calculation methods.

VI. CONCLUSION

In this paper a simple and effective method of LSF/ISF calculation was proposed. It was shown that for different orders of AR model a property of "inter-frame ordering" takes place in 88.9% to 96.8 % of cases for LSF of narrow-band speech and in 93.2% to 95.2% of cases for ISF of wideband speech. This means that LSF/ISF localization task can be mainly reduced to the verification of inter-frame ordering property.

During experimental verification of the proposed method for different speakers and AR model orders the following results were obtained.

- The resulting computational savings in comparison with method [16], which did not use features of LSF distribution, are from 1.8 to 2.5 times (at different AR model orders).
- For AR model order the proposed approach reduces computational expenses in comparison with most widely used Kabal-Ramachandran's method in 3.4 times. Also there is a 2.4 times gain over the accelerated combination of Kabal-Ramachandran's method with Newton's method.

- It was found that the maximum number of operations of the proposed method is lower not only than the time-invariant expenses of Kabal-Ramachandran's method, but is also lower than the minimum number of operations of the accelerated combination of Kabal-Ramachandran's method and Newton's method. These facts show the advantage of application of proposed method in real-time systems. The method was realized on fixed-point 16-bit DSP and showed stable work.
- The application of proposed method to the computation of ISF have shown its advantage over the grid approach used in AMR-WB codec [3].

REFERENCES

[1] "3GPP TS 26.071. Adaptive multi-rate (AMR) speech codec," Dec. 2004.

[2] "ITU-T G.729. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction," Mar. 1996.

[3] "3GPP TS 26.171. Adaptive multi-rate-wideband (AMR-WB) speech codec," Dec. 2004.

[4] S. Ragot, B. Kovesi, D. Virette, R. Trilling, and D. Massaloux, "8-32 kbit/s scalable wideband speech and audio coding candidate for ITU-T G729EV standardization," in *Proceedings of ICASSP, Vol.1*, May 2006, pp. 1–4.

[5] L. R. Rabiner and R. Schafer, *Digital processing of speech signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[6] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proceedings of IEEE*, vol. 73, pp. 1551–1588, 1985.

[7] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal of Acoustic Society of America*, vol. 57, suppl.1, p. S.35, 1975.

[8] K. Paliwal and B. S. Atal, "Efficient vector quantization of lpc parameters at 24 bits/frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3–14, 1993.

[9] A. V. McCree and T. P. Barnwell, "A mixed excitation lpc vocoder model for low bit rate speech coding," *IEEE Trans. Speech Audio Proces.*, vol. 4, pp. 242–249, 1995.

[10] Y. Bistritz, H. Lev-Ari, and T. Kailath, "Immittance-domain levinson algorithms," *IEEE transactions on information theory*, vol. 35, pp. 675–682, 1989.

[11] Y. Bistritz and S. Peller, "Immittance spectral pairs (ISP) for speech encoding," in *Proceedings of ICASSP, Vol.2*, May 1993, pp. 9–12.

[12] H. Dong, A. Gersho, J. D. Gibson, and V. Cuperman, "A multiple description speech coder based on AMR-WB for mobile ad hoc networks," in *Proceedings of ICASSP, Vol.1*, May 2004, pp. 277–280.

[13] P. Kabal and R. P. Ramachandran, "The computation of line spectral frequencies using chebyshev polynomials," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, pp. 1419–1426, 1986.

[14] C.-H. Wu and J.-H. Chen, "A novel two-level method for the computation of lsp frequencies using a decimation-in-degree algorithm," *IEEE Transactions on Speech and Audio Processing*, vol. 5, pp. 106–115, 1997.

[15] J. Rothweiler, "A rootfinding algorithm on line spectral frequencies," in *Proceedings of ICASSP*, May 1999, pp. 661–664.

[16] V. Semenov, A. Kalyuzhny, and A. Kovtonyuk, "Efficient calculation of line spectral frequencies based on new method for solution of transcendental equations," in *Proceedings of ICASSP, Vol.2*, Apr. 2003, pp. 457–460.

[17] V. Semenov, "A novel approach to calculation of line spectral frequencies based on inter-frame ordering property," in *Proceedings of ICASSP, Vol.3*, May 2006, pp. 1072–1075.

V. Semenov received his PhD degree in acoustics from the Institute of Hydromechanics of NASU in 2004, his specialist and bachelor degrees in system analysis and control from the National Technical University of Ukraine "Kiev Polytechnic Institute" in 2000 and 1998 respectively.

He is currently a leading staff scientist at Scientific-Production Enterprise "DELTA", Kiev, Ukraine. His current research interests include speech coding, signal processing and wireless communications.