ESTIMATION OF THE FUNDAMENTAL FREQUENCY OF THE SPEECH SIGNAL COMPRESSED BY G.723.1 ALGORITHM APPLYING PCC INTERPOLATION

Zoran N. MILIVOJEVIĆ [*] — Darko BRODIĆ [**]

In this paper the results of the estimation of the fundamental frequency of the speech signal modeled by the G.723.1 method are analyzed. The estimation of the fundamental frequency was performed by the Peaking-Peaks algorithm with the implemented Parametric Cubic Convolution (PCC) interpolation. The efficiency of PCC was tested for Keys, Greville and Greville two-parametric kernel. Depending on MSE a window that gives optimal results was chosen.

K e y w o r d s: speech processing, signal representation, speech compression, fundamental frequency, VoIP

1 INTRODUCTION

ITU (International Telecommunication Union) has defined the standard H.324 for specification of components, protocols and procedures which make multimedia communication services in networks with packets communication possible [1]. Besides, H.324 defines audio and video coders-decoders. G.723.1 is an audio codec intensively used in VoIP (Voice over IP) communications [2]. G.723.1 provides the speech compression to 5.3 kb/s and 6.3 kb/s. In the first case ACELP (Algebraic Code Excited Linear Prediction) algorithm and in the second MP-MLQ (Multipulse Maximum Likelihood Quantization) algorithm are applied. The frame size is 30 ms with 240 samples with sampling of 8 kHz. Each frame is being processed in the following 7.5 ms. The frame is divided into smaller frames whose duration is 7.5 ms. Delaying (6797 ms), delaying variations, packets losses and echo influence on the quality of speech transmitted by means of packet network. Evaluation of MOS (Mean Opinion Score) test is acceptable 3.4 (5.3 kbps) and good 3.8 (6.3 kbps) [3]. With G.723.1 coding the fundamental frequency estimation is firstly being done, and than the estimated fundamental frequency is used in the following steps of the speech signal coding. In the decoding process the speech signal is reconstructed which implies renewal of the fundamental frequency [5].

After the transfer of the speech signal by the VoIP service, on the reception side there is a need for the signal processing (speech and speaker recognition, echo canceling, improvement of the quality and the speech articulateness etc.). The characteristic example is the correction of the speech signal quality by reducing of dissonant frequencies [6, 7]. In the speech signal processing there is a need for determination of the fundamental frequency. A number of algorithms were developed for determination of the fundamental frequency where the analysis is performed in the time and frequency domain [8–11].

The frequently applied method for determination of the fundamental frequency is based on the peaking peaks of the amplitude characteristic in the specific frequency range. This method is used for analyzing of the signal values in the spectrum on frequencies on which the Discrete Fourier Transform (DFT) was calculated. Most often the real value of the fundamental frequency is not there on the frequencies where DFT is calculated, but lies between the two spectrum samples. That causes the frequency estimation error that lies in the interval $[-(F_S/(2N)\,\text{Hz}, (F_S/(2N)\,\text{Hz}]$, where F_S is the sampling frequency and N is the DFT window size. One way of reducing the error is determination of the interpolation function and estimation of the spectrum characteristics in an interval between the two samples. This procedure gives the reconstruction of the spectrum on the base of DFT. The spectrum parameters are then determined by analytic procedures (differentiation, integration, extreme values, ...).

Determination of spectral characteristic values between DFT samples is taken as the local convolution of the neighboring DFT samples and interpolation kernel (piece-wise polynomial interpolation) [12–14]. Two methods are frequently used for interpolation: a) the cubic B-spline interpolation [12] and b) the Parametric Cubic Convolution (PCC) interpolation. From the point of view of the fastness of the maximum position estimation the application of PCC interpolation kernel is more suitable, because it is possible to find the maximum position directly (by means of a formula) using the sampled data without convolution being applied. The detailed analysis of the fundamental frequency estimation by means of PCC interpolation is described in the paper [15]. In this paper there is a formula intended for analytical calculation of the fundamental frequency without convolution.

* Technical College, Aleksandra Medvedeva 20, Niš, Serbia, zoran.milivojevic@jotel.co.rs, ** University of Belgrade, Technical Faculty Bor, Vojske Jugoslavije 12, 19210 Bor, Serbia
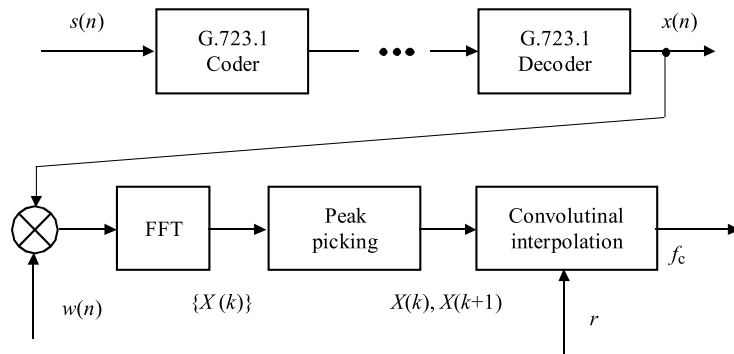
Fig. 1. Algorithm of estimate of fundamental frequency

The results of the application of PCC interpolation for determining of the fundamental frequency in the conditions of application of some window in the processing of the discrete speech signal are presented in [16]. Through some simulation procedures algorithm efficiency analyses have been done where, as a quality measure of an algorithm, the Mean Square Error (MSE) has been used. The best results were shown by the algorithm with the implemented Blackman window. The results of the fundamental frequency estimation of the speech signal modeled by SYMPES (Systematic Procedure to Model Speech Signals via Predefined "Envelope and Signature Sequences) method are shown in [17].

Further in this paper there will be presented some results of fundamental frequency estimation in audio and speech signals compressed by G.723.1 algorithm in order to determine the fundamental frequency using PCC interpolation kernels (Keys, Greville and two parametric Greville kernel) and some window functions (Hamming, Hanning, Blackman, Rectangular, Kaiser and Triangular window). The fundamental frequency estimation was performed on the base of the analytical expression from [15] for Keys kernel. For Greville and two-parametric Greville kernels we have determined analytical form for determination of the fundamental frequency position. As a measure of the quality of interpolation algorithm MSE (Mean Square Error) will be applied. On the base of minimum values of MSE optimum kernel parameters and the corresponding window function will be determined.

This paper is organized as follows: In Section 2 there is a description of the PCC algorithm. In Section 2.a there are definitions of interpolation kernels. In Section 2.b the algorithm for determination of optimal kernel parameters is presented. In Section 3 numerical MSE results in the estimation of fundamental frequency of the speech signal modeled by the G.723.1 method are presented. The comparative analysis of the results and the choice of the optimal kernel and window function are shown in Section 4. Section 5 represents the conclusion.

2 ALGORITHM OF FUNDAMENTAL FREQUENCY ESTIMATION

Algorithm for estimation of the fundamental frequency, based on the algorithm from [15], is presented in Fig. 1. This algorithm is realized in the following steps:

Step 1: Audio or speech signal s(n) is coded by G.723.1 coder.

Step 2: Coded signal is decoded by G.723.1 decoder and generises signal $x(n)$.

Step 3: Window $w(n)$ whose length is N applies to decoded signal $x(n)$.

Step 4: Spectrum $X(k)$ is calculated by using DFT:

$$X(k) = DFT\big(x(n)\big), \qquad (1)$$

The spectrum is calculated in discrete points $k = 0, \ldots, N-1$, where N is the length of DFT. The real spectrum of signals $x(n)$ is continuous, whereas DFT defines the values of the spectrum at some discrete points.

Step 5: By using peak picking algorithm, the position of the maximum of the real spectrum that is between kth and $(k+1)$th samples is determined, where the values $X(k)$ and $X(k+1)$ are the highest in the specified domain.

Step 6: The position of the maximum of the spectrum is calculated by PCC interpolation. The reconstructed function is

$$X_r(f) = \sum_{i=k-L/2+1}^{k+L/2} p_i r(f-i), \quad k \le f \le k+1, \quad (2)$$

where $p_i = X(i)$, $r(f)$ is the kernel of interpolation and L the number of samples that participate in interpolation.

Step 7: By differentiation $X_r(f)$ and zero adjustment the position of maximum is determined; it presents the estimated fundamental frequency f_e.

The quality of the algorithm for the estimate of fundamental frequency can be also expressed by MSE

$$MSE = \overline{(f - f_e)^2}, \qquad (3)$$

where f is true fundamental frequency and f_e is fundamental frequency estimate.

2.1 Interpolation Kernel

Next, we give definitions of the interpolation kernels which are tested in this paper

a) *Keys* interpolation kernel [12, 13]

$$r(f) = \begin{cases} (\alpha+2)|f|^3 - (\alpha+3)|f|^2 + 1, & |f| \le 1, \\ \alpha|f|^3 - 5\alpha|f|^2 + 8\alpha|f| - 4\alpha, & 1 \le |f| \le 2, \\ 0 & , otherwise. \end{cases} \quad (4)$$

For $L = 4$ from (2) position of maximum is determined

$$f_{\max} = \begin{cases} k - \frac{c}{2b}, & a = 0, \\ k + \frac{-b \pm \sqrt{b^2 - ac}}{a}, & a \ne 0, \end{cases} \quad (5)$$

where:

$$a = 2\big(\alpha p_{k-1} + (\alpha+2)p_k - (\alpha+2)p_{k+1} - \alpha p_{k+2}\big),$$
$$b = -2\alpha p_{k-1} - (\alpha+3)p_k + (2\alpha+3)p_{k+1} + \alpha p_{k+2},$$
$$c = -\alpha p_{k-1} - \alpha p_{k+1}. \quad (6)$$

b) *Greville* interpolation kernel [17]

$$r(f) = \begin{cases} \big(\alpha+\frac{3}{2}\big)|f|^3 - \big(\alpha+\frac{5}{2}\big)|f|^2 + 1, \\ \qquad\qquad\qquad 0 \le |f| \le 1, \\ \frac{1}{2}(\alpha-1)|f|^3 - \big(3\alpha-\frac{5}{2}\big)|f|^2 \\ + \big(\frac{11}{2}\alpha-4\big)|f| - (3\alpha-2), \\ \qquad\qquad\qquad 1 \le |f| \le 2, \\ -\frac{1}{2}\alpha|f|^3 + 4\alpha|f|^2 - \frac{11}{2}\alpha|f| + 9\alpha, \\ \qquad\qquad\qquad 2 \le |f| \le 3, \\ 0, \qquad\qquad\qquad 3 \le |f|. \end{cases} \quad (7)$$

For $L = 6$ from (2) and (7) position of maximum is determined according to (5), where

$$a = -\frac{3}{2}\alpha p_{k-2} + \frac{3}{2}(\alpha-1)p_{k-1} + 3\big(\alpha+\frac{3}{2}\big)p_k$$
$$\qquad - 3\big(\alpha+\frac{3}{2}\big)p_{k+1} - \frac{3}{2}(\alpha-1)p_{k+2} + \frac{3}{2}\alpha p_{k+3},$$
$$b = -2\alpha p_{k-2} + (-3\alpha+2)p_{k-1} - (2\alpha+5)p_k \qquad (8)$$
$$\qquad + 4(\alpha+1)p_{k+1} - p_{k+2} - \alpha p_{k+3},$$
$$c = -\frac{1}{2}\alpha pk - 2 + \big(\alpha-\frac{1}{2}\big)p_{k-1} - \big(\alpha-\frac{1}{2}\big)p_{k+1}$$
$$\qquad + \frac{1}{2}\alpha p_{k+2}.$$

c) *Greville* two-parametric cubic convolution kernel (G2P) [17]

$$r(f) = \begin{cases} \big(\alpha-\frac{5}{2}\beta+\frac{3}{2}\big)|f|^3 - \big(\alpha-\frac{5}{2}\beta+\frac{5}{2}\big)|f|^2 + 1, \\ \qquad\qquad\qquad 0 \le |f| \le 1, \\ \frac{1}{2}(\alpha-\beta-1)|f|^3 - \big(3\alpha-\frac{9}{2}\beta-\frac{5}{2}\big)|f|^2+ \\ \big(\frac{11}{2}\alpha-10\beta-4\big)|f| - (3\alpha-6\beta-2), \\ \qquad\qquad\qquad 1 \le |f| \le 2, \\ -\frac{1}{2}(\alpha-3\beta)|f|^3 + \big(4\alpha-\frac{25}{2}beta\big)|f|^2- \\ \big(\frac{21}{\alpha}-34\beta\big)|f| + (9\alpha-30\beta), \\ \qquad\qquad\qquad 2 \le |f| \le 3, \\ -\frac{1}{2}\beta|f|^3 + \frac{11}{2}|\beta||f|^2 - 20\beta|f| + 24\beta, \\ \qquad\qquad\qquad 4 \le |f|. \end{cases} \quad (9)$$

For $L = 8$ from (2) and (9) position of maximum is determined according to (5), where

$$a = -\frac{3}{2}\beta p_{k-3} - \frac{3}{2}(\alpha-3\beta)p_{k-2} + \frac{3}{2}(\alpha-\beta-1)p_{k-1}$$
$$\quad + 3\big(\alpha-\frac{5}{2}\beta+\frac{3}{2}\big)p_k - 3\big(\alpha-\frac{5}{2}\beta-\frac{3}{2}\big)p_{k+1}$$
$$\quad - \frac{3}{2}(\alpha-\beta-1)p_{k+2} \pm \frac{3}{2}(\alpha-3\beta)p_{k+3} + \frac{3}{2}\beta p_{k+4},$$
$$b = -2\beta p_{k-3} - (2\alpha-7\beta)p_{k-2} + (-3\alpha+6\beta+2)p_{k-1}$$
$$\quad - \big(2\alpha-5\beta+\frac{5}{2}\big)p_k + (4\alpha-10\beta+1)p_{k+1} \qquad (10)$$
$$\quad + (3\beta-1)p_{k+2} + (-\alpha+2\beta)\alpha p_{k+3} - \beta p_{k+4},$$
$$c = -\frac{1}{2}\beta p_{k-3} + \big(-\frac{1}{2}\alpha+2\beta\big)p_{k-2} + \big(\alpha-\frac{5}{2}\beta-\frac{1}{2}\big)p_{k-1}$$
$$\quad - \big(\alpha+\frac{5}{2}\beta+\frac{1}{2}\big)p_{k+1} + \big(\frac{1}{2}\alpha-2\beta\big)p_{k+2} + \frac{1}{2}\beta p_{k+3}.$$

In (4)–(10) there are α and β parameters. The optimal values of these parameters will be determined by the minimal value of MSE, for Keys, Greville and G2P kernel. For the first two of them

$$\alpha_{\text{opt}} = \arg\min_{\alpha} MSE, \quad (11)$$

and for the G2P kernel:

$$(\alpha_{\text{opt}}, \beta_{\text{opt}}) = \arg\min_{\alpha,\beta} MSE, \quad (12)$$

The detailed analysis in [15–17] showed that the minimal value of MSE depends on the application of window by which signal processing $x(n)$ is carried out in time domain. MSE will be defined for: a) Hamming, b) Hanning, c) Blackman, d) Rectangular, e) Kaiser and f) Triangular window.

2.2 Interpolation Kernel Parameters

Algorithm for determination of interpolation kernel parameters α and β is realized in the following steps:

Step 1: signal $x(n)$, which was previously coded and decoded by G.723.1 algorithm, is modified by the window function $w(n)$ whose length is N,

*Step 2*: spectrum $X(k)$ is determined by the application of DFT,

Step 3: reconstruction of the continual function that represents spectrum $X(f)$ is performed by the application of PCC interpolation,

Step 4: MSE is calculated for various values of parameters α and β depending on the implemented window,

Step 5: α_{opt} and β_{opt} are determined for which minimal value of MSE is obtained.
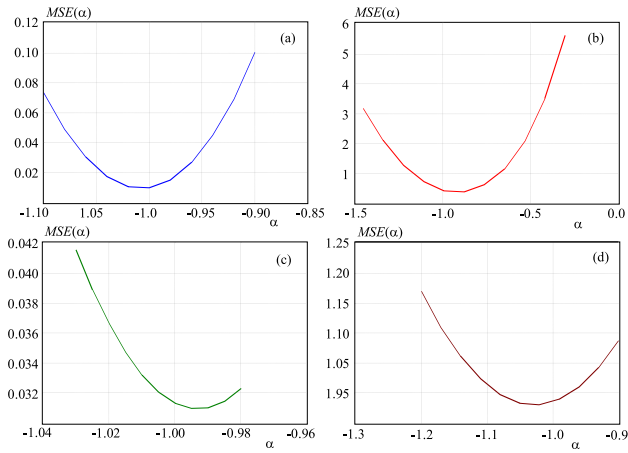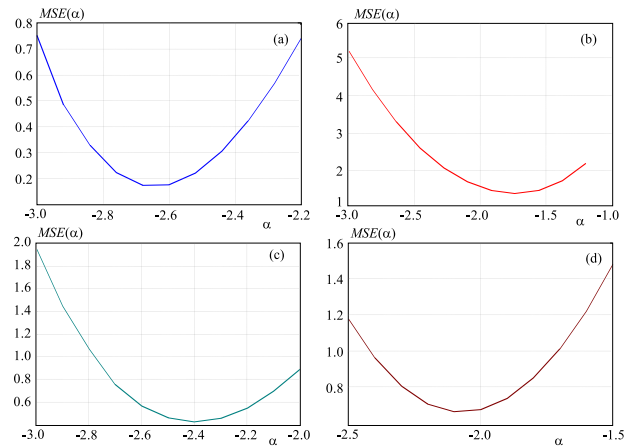
Fig. 2. $MSE(\alpha)$ for Keys kernel and Hamming window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal
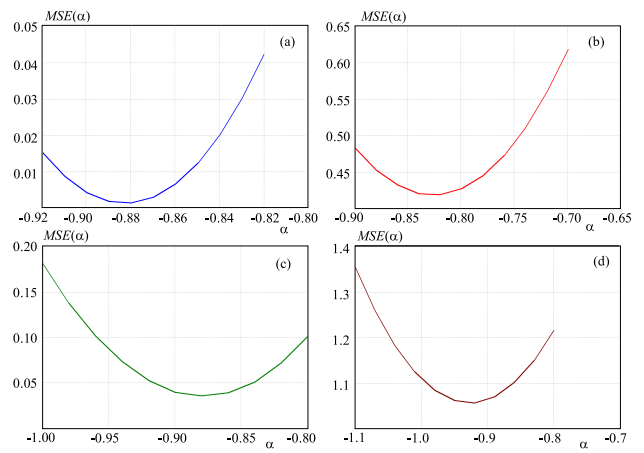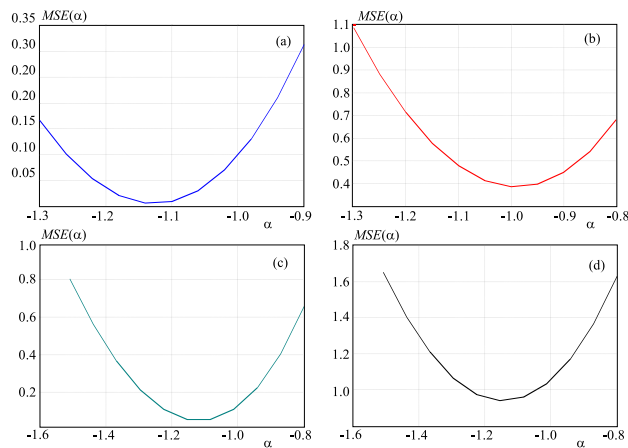
Fig. 3. $MSE(\alpha)$ for Keys kernel and Hann window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal
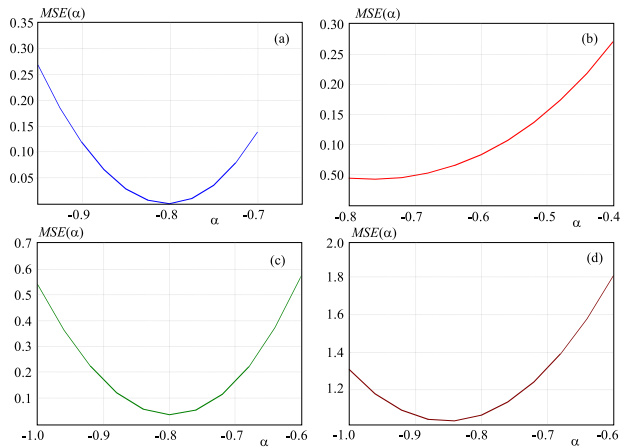
Fig. 4. $MSE(\alpha)$ for Keys kernel and Blackman window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, c) uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

Fig. 5. $MSE(\alpha)$ for Keys kernel and Rectangular window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal
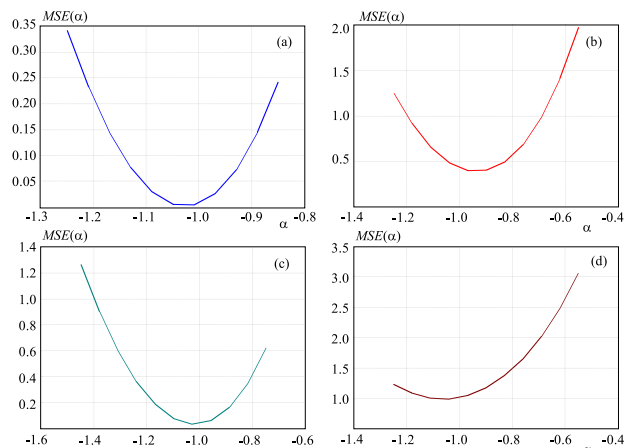
Fig. 6. $MSE(\alpha)$ for Keys kernel and Kaiser window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

Fig. 7. $MSE(\alpha)$ for Keys kernel and Triangular window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

2.3 Test Signals

PCC algorithm of the fundamental frequency estimation will be applied to:

a) simulation sine test signal, and

b) speech test signal.

Simulation sine signal for testing of PCC algorithm is defined in [15]

$$s(t) = \sum_{i=1}^{K} \sum_{g=0}^{M} a_i \sin\left(2\pi i \left(f_0 + g\frac{f_s}{NM}\right)t + \theta_i\right), \qquad (13)$$

where f_0 is fundamental frequency, θ_i and a_i are phase and amplitude of the i-th harmonic, respectively, K is the number of harmonics, M is the number of points between the two samples in spectrum where PCC interpolation is being made. The speech test signal is obtained by recording of a speaker in the real acoustic ambient. For further comparative analysis by interpolation (Matlab function **interp 1**) it is suitable for the fundamental frequency of the speech test signal and the one of the sine test signal to be equal.

PCC algorithm will be applied to:

a) uncoded simulation sine and speech test signals, and

b) by G.723.1 algorithm coded and decoded sine and speech test signals.

3 EXPERIMENTAL RESULTS AND COMPARISON

3.1 Testing Parameters

In the simulation process f_0 and θ_i are random variables with uniform distribution in the range $[G2(97.99\,\text{Hz}), G5(783.99\,\text{Hz})]$ and $[0, 2\pi]$. Signal frequency of sampling is $f_S = 8\,\text{kHz}$ and the length of window is $N = 256$, which assures the analysis of subsequences that last $32\,\text{ms}$. The results presented further in this paper relate to $f_0 = 125\text{–}140.625\,\text{Hz}$ (frequencies between the eighth and ninth DFT components). Number of frequencies in the specified range for which the estimation is done is $M = 100$. The sine test signal is with $K = 10$ harmonics. All further analyzes will relate to a) Hamming, b) Hamming, c) Blackman, d) Rectangular, e) Kaiser and f) Triangular window.

3.2 Experimental Results

3.2.1 Keys Kernel

Applying an algorithm for determination of parameters of Keys interpolation kernel some diagrams are drawn $MSE(\alpha)$, the minimum value MSE_{Kmin} determined, and on the base of it the optimum value of Keys kernel α_{opt} determined for: a) – Hamming (Fig. 2), b) – Hann (Fig. 3), c) – Blackman (Fig. 4), d) – Rectangular (Fig. 5),

e) – Kaiser (Fig. 6) and f) – Triangular (Fig. 7) window functions. Values MSE_{Kmin} and α_{opt} are presented in Tab. 1 (uncoded sine test signal MSE_{Kmin}, G.723.1 coded sine test signal MSE_{KGmin}) and Tab. 2 (speech test signal MSE_{KSPmin}, G.723.1 coded sine test signal MSE_{KGSPmin}).

On the base of the results presented in Tabs. 1 and 2 it is obvious that:

a) – at speech test signal the greatest precision of fundamental frequency estimation is when Blackman window ($MSE_{\text{Kmin}} = 0.000423$) is applied. At G.723.1 coded sine test signal the greatest precision of estimation is in Rectangular ($MSE_{\text{KGmin}} = 0.3855$) window. When G.723.1 coding is applied precision in fundamental frequency estimation is lowered for $MSE_{\text{KGmin}}/MSE_{\text{Kmin}} = 0.3855/0.00619 = 62.277$ times.

b) at speech test signal the greatest precision is in triangular window ($MSE_{\text{KSPmin}} = 0.0277$). At G.723.1 coded speech signal the greatest precision is in rectangular window ($MSE_{\text{KGSPmin}} = 0.6752$). When coding is applied precision in fundamental frequency estimation is lowered for $MSE_{\text{KGSPmin}}/MSE_{\text{KSPmin}} = 0.6752/0.0277 = 24.375$ times.

c) at coded speech signal in relation to coded sine signal non precision in fundamental frequency estimation is greater for $MSE_{\text{KGSPmin}}/MSE_{\text{KGmin}} = 0.6752/0.3855 = 1.75$ times.
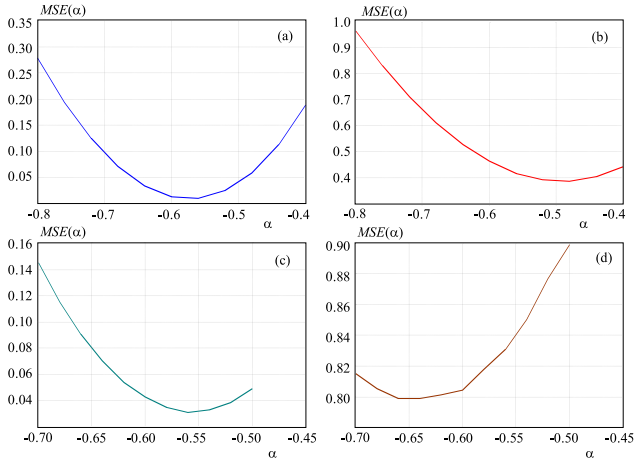
3.2.2 Greville Kernel

Applying algorithm for determination of Greville interpolation kernel some diagrams $MSE(\alpha)$ are drawn, minimum value MSE_{Gmin} determined, and on the base of it optimum value of Greville kernel parameters α_{opt} determined for: a) – Hamming (Fig. 8), b) – Hann (Fig. 9), c) – Blackman (Fig. 10), d) – Rectangular (Fig. 11), e) – Kaiser (Fig. 12) and f) – Triangular window. Values MSE_{min} and α_{opt} are presented in Tab. 3 (uncoded sine test signal MSE_{Gmin}, coded sine test signal MSE_{GGmin}) and Tab. 4 (speech test signal MSE_{GSPmin}, coded speech test signal MSE_{GGSPmin}).

On the base of the results presented in Tabs 3 and 4 it is obvious that:

a) – at sine test signal the greatest precision of fundamental frequency estimation is when Blackman ($MSE_{\text{Gmin}} = 0.0002441$) window is applied. At G.723.1 coded sine test signal the greatest precision of estimation is in rectangular window ($MSE_{\text{GGmin}} = 0.3274$). When coding is applied, precision of fundamental frequency estimation is lowered for $MSE_{\text{GGmin}}/MSE_{\text{Gmin}} = 0.3274/0.0002441 = 1341.25$ times.

b) at speech test signal the greatest precision is in triangular window ($MSE_{\text{GSPmin}} = 0.0255$). At coded speech signal the greatest precision is in rectangular window ($MSE_{\text{GGSPmin}} = 0.5178$). When coding is applied, precision of the fundamental frequency estimation is lowered for $MSE_{\text{GGSPmin}}/MSE_{\text{GSPmin}} = 0.5178/0.0255 = 20.3$ times.

Fig. 8. $MSE(\alpha)$ for Greville kernel and Hamming window: a) – uncompressed sine test signal, b) – G.723.1 compressed sine test signal, c) – uncompressed speech test signal, d) – G.723.1 compressed speech test signal
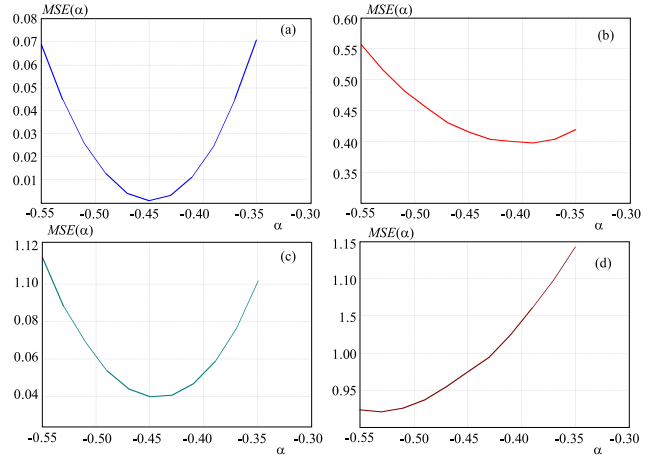
Fig. 9. $MSE(\alpha)$ for Greville kernel and Han window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal
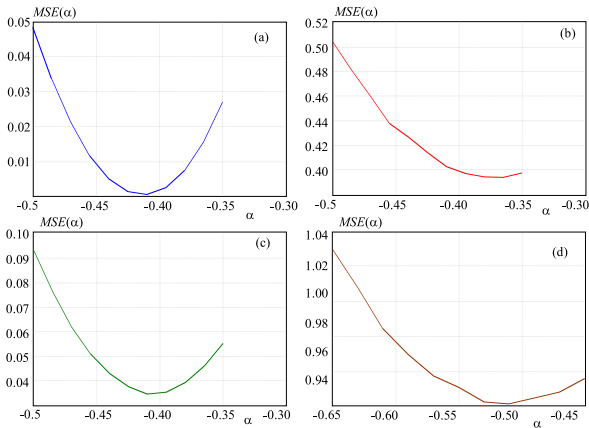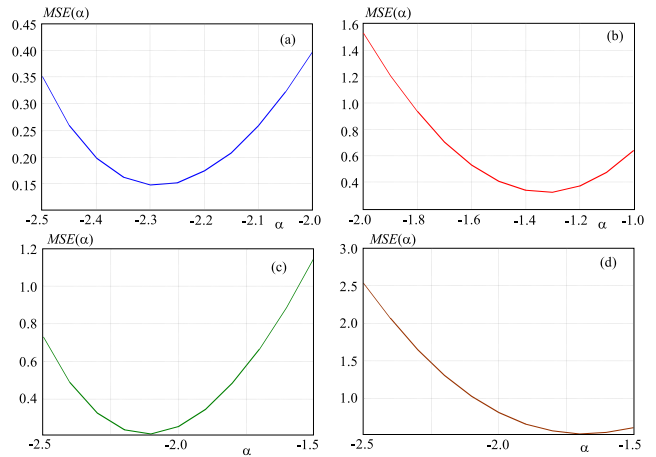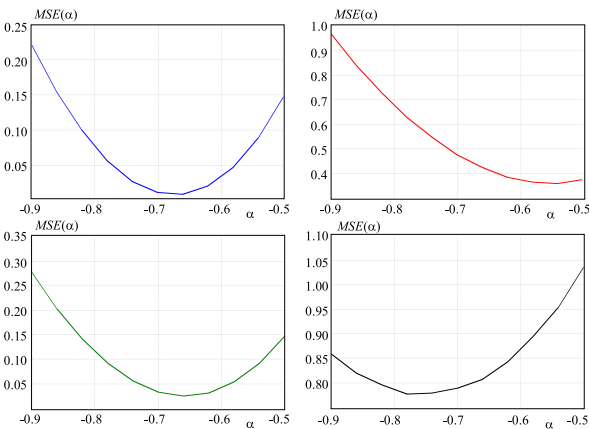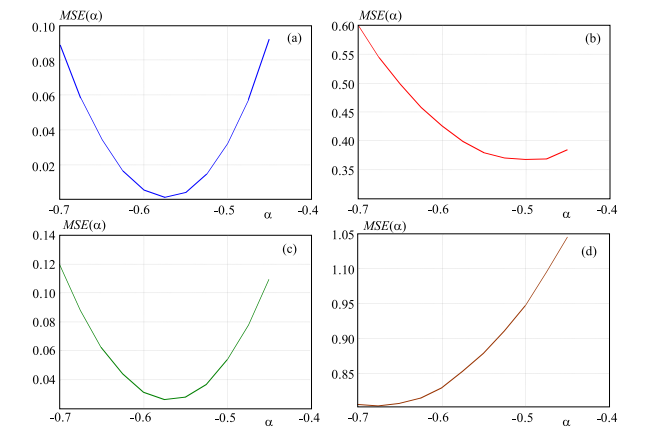

Fig. 10. $MSE(\alpha)$ for Greville kernel and Blackman window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

Fig. 11. $MSE(\alpha)$ for Greville kernel and Rectangular window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

Fig. 12. $MSE(\alpha)$ for Greville kernel and Kaiser window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal

Fig. 13. $MSE(\alpha)$ for Greville kernel and Triangular window: (a) – uncompressed sine test signal, (b) – G.723.1 compressed sine test signal, (c) – uncompressed speech test signal, (d) – G.723.1 compressed speech test signal
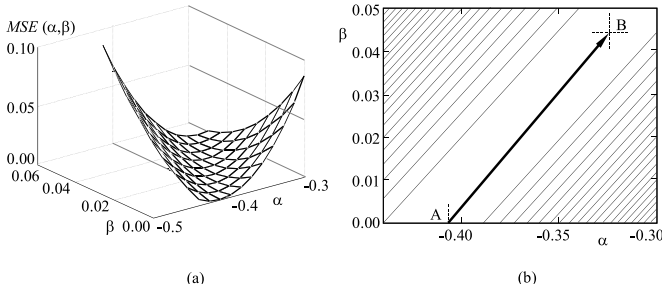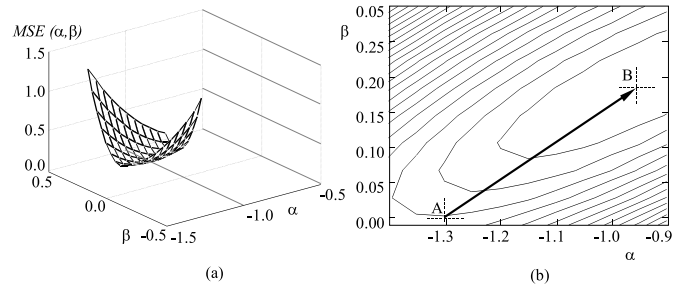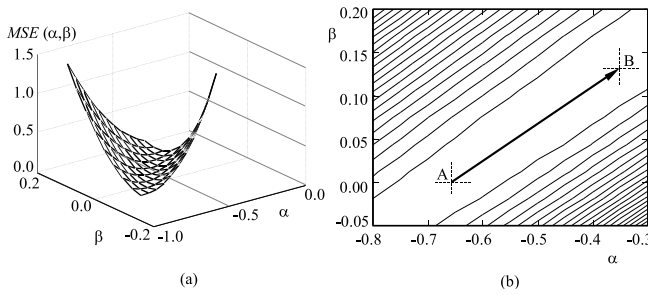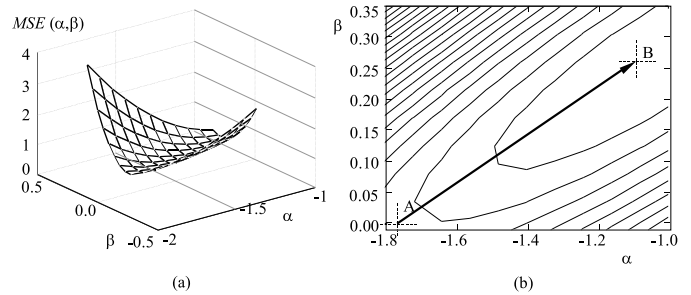
Fig. 14. Sine test signal without compression with the application of Blackman window: (a) – $MSE(\alpha, \beta)$ for G2P PCC interpolation, (b) – positions of $\min(MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}}))$ in plane $(\alpha\beta)$ for Greville (point A) and G2P PCC (point B) interpolation

Fig. 15. Sine test signal with G.723.1 compression with the application of Rectangular window: (a) – $MSE(\alpha, \beta)$ for the application of G2P PCC interpolation, (b) – Positions of $\min(MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}}))$ in plane $(\alpha\beta)$ for Greville (point A) and G2P PCC (point B) interpolation

Fig. 16. Speech test signal without compression with the application of Kaiser window: (a) – $MSE(\alpha, \beta)$ for G2P PCC interpolation, (b) – positions of $\min(MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}}))$ in plane $(\alpha\beta)$ for Greville (point A) and G2P PCC (point B) interpolation

Fig. 17. Speech test signal with G.723.1 compression with the application of Rectangular window: (a) – $MSE(\alpha, \beta)$ for the application of G2P PCC interpolation, (b) – Positions of $\min(MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}}))$ in plane $(\alpha\beta)$ for Greville (point A) and G2P PCC (point B) interpolation

c) at coded speech signal in relation to coded sine signal non-precision of the fundamental frequency is greater for $MSE_{\mathrm{GGSPmin}}/MSE_{\mathrm{GGmin}} = 0.5178/0.3274 = 1.58$ times.

3.2.3 G2P Kernel

For window (Blackman, Kaiser, Rectangular) which showed the best results with Greville kernel, an analysis was performed by means of G2P kernel. Three-dimensional $MSE(\alpha, \beta)$ graphics are drawn (Figs. 14.a, 15.a, 16.a, and 17.a), the shift of minimum MSE_{min} in (α, β) level determined, and α_{opt} and β_{opt} values determined and presented in Tab. 5.

In Figures 14.b, 15.b, 16.b, and 17.b the positions of $MSE_{\mathrm{min}} = MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}})$ minimum in (α, β) plane for Greville (point A) and G2P (point B) interpolation kernel, are shown. Vector **AB** shows the position change of minimum ($MSE(\alpha_{\mathrm{opt}}, \beta_{\mathrm{opt}})$).

On the base of the results presented in Tab. 5 it is obvious that:

a) – at sine test signal the greatest precision of the fundamental frequency estimation is when Blackman ($MSE_{\mathrm{G2Pmin}} = 0.00014$) window is applied. At G.723.1 coded sine test signal the greatest precision of estimation

is in rectangular window ($MSE_{\mathrm{G2PGmin}} = 0.175$). When coding is applied, precision of fundamental frequency estimation is lowered for $MSE_{\mathrm{G2PGmin}}/MSE_{\mathrm{G2Pmin}} = 0.175/0.00014 = 12500$ times.

b) at speech test signal the greatest precision is in Kaiser window ($MSE_{\mathrm{G2PSPmin}} = 0.0112$). At G.723.1 coded speech signal the greatest precision is in rectangular window ($MSE_{\mathrm{G2PGSPmin}} = 0.2898$). When G.723.1 coding is applied, precision of fundamental frequency estimation is lowered for $MSE_{\mathrm{G2PGSPmin}}/MSE_{\mathrm{G2PSPmin}} = 0.2898/0.0112 = 25.875$ times.

c) at G.723.1 coded speech signal in relation to G.723.1 coded sine signal, non-precision of fundamental frequency estimation is greater for $MSE_{\mathrm{G2PGSPmin}}/MSE_{\mathrm{G2PGmin}} = 0.2898/0.175 = 1.656$ times.

4 COMPARATIVE ANALYSIS

Comparative precision analysis of the estimated fundamental frequency of the sine test signal and the speech test signal, without and with G.723.1 compression will be performed on the base of the minimal values of MSE. The minimal value of MSE is determined on the base of a diagram in the Figs. 2–7 (Keys), Figs. 8–13 (Greville) and Figs. 14–17 (G2P) and presented in the Table 1 ($MSE_{\mathrm{Kmin}}, MSE_{\mathrm{KGmin}}$), Table 2 ($MSE_{\mathrm{KSPmin}},$

Table 1. Minimum MSE and α_{opt} for sine test signal (Keys kernel)

	Uncoded signal		Signal coded by G.723.1 algorithm	
	α_{opt}	MSE_{Kmin}	α_{opt}	MSE_{PKGmin}
Ham	-1	0.0098	-0.875	0.4203
Hann	-0.880	0.000619	-0.820	0.4170
Black	-0.800	0.000423	-0.760	0.4137
Rect.	-2.680	0.1767	-1.740	0.3855
Kais.	-1.140	0.0066	-1	0.3887
Trian.	-1.010	0.0036	-0.970	0.3982

Table 2. Minimum MSE and α_{opt} for speech test signal (Keys kernel)

	Uncoded signal		Signal coded by G.723.1 algorithm	
	α_{opt}	MSE_{KSPmin}	α_{opt}	MSE_{KGSPmin}
Ham	-0.995	0.0310	-1.020	0.9351
Hann	-0.880	0.0349	-0.920	1.0523
Black.	-0.800	0.0358	-0.840	1.0213
Rect.	-2.400	0.4323	-2.100	0.6752
Kais.	-1.080	0.0339	-1.150	0.9390
Trian.	-1.030	0.0277	-1.040	0.9865

Table 3. Minimum MSE and α_{opt} for speech test signal (Keys kernel)

	Uncoded signal		Signal coded by G.723.1 algorithm	
	α_{opt}	MSE_{Gmin}	α_{opt}	MSE_{GGmin}
Ham.	-0.560	0.0089	-0.4800	0.3841
Hann	-0.450	0.0006573	-0.3900	0.3969
Black.	-0.410	0.0002441	-0.3650	0.3938
Rect.	-2.300	0.1472	-1.3000	0.3274
Kais.	-0.660	0.0059	-0.5400	0.3598
Trian.	-0.575	0.0007793	-0.5000	0.3659

Table 4. Minimum MSE and α_{opt} for speech test signal (Greville kernel)

	Uncoded signal		Signal coded by G.723.1 algorithm	
	α_{opt}	MSE_{GSPmin}	α_{opt}	MSE_{GGSPmin}
Hamm	-0.560	0.0310	-0.6400	0.7983
Hann	-0.450	0.0363	-0.5300	0.9201
Black.	-0.410	0.0344	-0.5100	0.9209
Rect.	-2.100	0.2016	-1.7000	0.5178
Kais.	-0.660	0.0255	-0.7800	0.7787
Trian.	-0.575	0.0256	-0.6750	0.8011

MSE_{KGSPmin}), Table 3 (MSE_{Gmin}, MSE_{GGmin}), Table 4 (MSE_{GSPmin}, MSE_{GGSPmin}) and Table 5 (MSE_{G2Pmin}, MSE_{G2PGmin}, MSE_{G2PSP}, $MSE_{\mathrm{G2PGSPmin}}$) respectively.

Comparing the values MSE_{min} from Tabs. 1–5 it can be concluded that:

(a) – the optimum choice for sine test signal is Blackman window for all interpolation kernels. G2P interpolation kernel, which generates 66.91 % less than Keys and 42.65 % less than Greville kernel, showed the best results.

(b) – the optimum choice for speech test signal is G2P kernel with Kaiser window, which generates 59.57 % less than Keys kernel (Triangular window) and 56.08

c) the optimum choice for sine test signal coded by G.723.1 algorithm is rectangular window for all interpolation kernels. G2P interpolation kernel, which generates 54.61 % less than Keys and 46.55 % less than Greville kernel, showed the best results.

d) the optimum choice for speech test signal coded by G.723.1 algorithm is rectangular window for all interpolation kernels. G2P interpolation kernel, which generates 57.21 % less than Keys and 44.04 % less than Greville kernel, showed the best results.

e) comparing MSE for G2P kernel for uncoded speech test signal (Kaiser window, $MSE_{\mathrm{G2PSPmin}} = 0.0112$) and G.723.1 coded speech signal (Rectangular window, $MSE_{\mathrm{G2PGSPmin}} = 0.2898$) in relation to $MSE_{\mathrm{G2PGSPmin}}/MSE_{\mathrm{G2PSPmin}} = 0.2898/0.0112 = 25.875$ has been obtained.

In accordance to the derived conclusion, the application of the algorithm for further processing of G.723.1 speech signal with algorithms based on the estimated fundamental frequency (automatic verification of a speaker, recognition of the speech, *etc*) would not bring satisfactory results.

5 CONCLUSION

This paper presented the analysis of the fundamental frequency estimation results of the speech signal compressed by G.723.1 algorithm which has been intensively used in VoIP services. The estimation of the fundamental frequency has been made by Peaking Peaks algorithm with implemented PCC interpolation. Experiments have been performed with Keys, Greville and Greville two-parametric G2P kernels. In order to minimize MSE some windows have been implemented. The detailed analysis has shown that the optimal choice is Greville two-parametric kernel and the Rectangular window implemented in PCC algorithm. In relation to Keys and Greville kernels, Greville two-parametric kernel generates 54.61 % and 46.55 % less MSE, respectively. Comparing the obtained results to the results of the estimation of the fundamental frequency in the speech signal that is not modeled by G.723.1 method, a relation of minimal MSEs 25.875 has been obtained. This result leads to the conclusion that, although by the subjective tests (ACR-MOS tests) extremely good results have been obtained (3.4–3.8), in further processing of the signal based on the estimated fundamental frequency (automatic verification of the speaker, recognition of the speech, *etc*), the obtained results would not be satisfactory in regard to the reliability and precision.

Table 5. Minimum MSE, α_{opt} and β_{opt} (G2P kernel)

		α_{opt}	β_{opt}	MSE_{G2Pmin}
Sin test signal	Uncoded signal	−0.328	0.0438	MSE_{G2Pmin}
	Blackman win.	−0.328	0.0438	0.00014
	G.723.1 coded	−0.95	0.1838	MSE_{G2PGmin}
	Rectang. win.	−0.95	0.1838	0.1750
Speech test signal	Uncoded signal	−0.35	0.1375	MSE_{G2PSPmin}
	Kaiser win.	−0.35	0.1375	0.0112
	G.723.1 coded	−1.08	0.2600	$MSE_{\mathrm{G2PGSPmin}}$
	Rectang. win.	−1.08	0.2600	0.2898

Acknowledgement

References

[1] ITU-T Rec. G.723.1: Dual-rate Speech Coder For Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s, 1996.

[2] KARAPANTAZIS, S.—PAVLIDOU, F.: VoIP: A Comprehensive Survey on a Promising Technology, Computer Networks **53** (2009), 2050-2090.

[3] REGUERA, V.—PALIZA, F.—GODOY, W.—FERNANDEZ, E.: On the Impact of Active Queue Management on VoIP Quality of Service, Computer Communications **31** (2008), 73-87.

[4] KAOSAR, G.—SHELTAMI, T.: Voice Transmission over ad hoc Network Adapting Optimum Approaches to Maximize the Performance, Computer Communications **32** (2009), 634-639.

[5] ZOLGHADR, A.—GOODARZI, E.—MOOSAVINEZHAD, M.: Real-Time Implementation of G.723.1 Speech Coder, Iranian Journal of Information Science and Technology **2** No. 1 (2004), 3–12.

[6] JOEN, B.—KANG, S.—BAEK, S.—SUNG, K.: Filtering of a Dissonant Frequency Based on Improved Fundamental Frequency Estimation for Speech Enhancement, IEICE Trans. Fundamentals **E86-A** No. 8 (Aug 2003), 2063–2064.

[7] MILIVOJEVIC, Z.—BALANESKOVIC, D.: Enhancement of the Perceptive Quality of the Noisy Speech Signal by Using of DFF-FBC Algorithm, Facta Universitatis, Ser.: Elec. Energ. **22** No. 3 (Dec 2009), 379=-392.

[8] QIU, L.—YANG, H.—KOH, S.: Fundamental Frequency Determination on Instantaneous Frequency Estimation, IEEE Signal Process. **44** (1995), 233–241.

[9] CHEVEIGNE, A.—KAWAHARA, H.: YIN, a Fundamental Frequency Estimator for Speech and Music, J. Acoust. Soc. Am. **111** No. 4 (April 2002), 1917–1930.

[10] MURAKAMI, T.—ISHIDA. Y.: Fundamental Frequency Estimation of Speech Signal Using MUSIC Algorithm, Acoust. Sci. & Tech. **22** (2001), 293–297.

[11] KACHAA,—GRENEZ, F.—BENMAHAMMED, K.: TimeFrequency Analysis and Instantaneous Frequency Estimation Using Two-Sided Linear Prediction, IEEE Signal Processing **85** (2005), 491–503.

[12] KEYS, R.: Cubic Convolution Interpolation for Digital Image Processing, IEEE Trans. Acoust., Speech & Signal Process **29** (1981), 1153–1160.

[13] PARK, K.—SCHOWENGERDT, R.: Image Reconstruction by Parametric Cubic Convolution, Computer Vision, Graphics & Image Processing **23** (1983), 258–272.

[14] REICHENBACH, S.: Two-Dimensional Cubic Convolution, IEEE Transactions on Image Processing **12** No. 8 (Aug 2003), 857–865.

[15] PANG, H.—BAEK, S.—SUNG, K.: Improved Fundamental Frequency Estimation Using Parametric Cubic Convolution, IEICE Trans. Fundamentals **E83-A** (2000), 2747–2750.

[16] MILIVOJEVIC, Z.—MIRKOVIC, D.—MILIVOJEVIC, S.: An Estimate of Fundamental Frequency Using PCC Interpolation - Comparative Analysis, Information Technology and Control **35** (2006), 131–136.

[17] MILIVOJEVIC, Z.—MIRKOVIC, M.: Estimation of the Fundamental Frequency of the Speech Signal Modeled by the SYMPES Method, Int. J. Electron. Commun. (AEÜ) **63** (2009), 200-208.

Zoran N. Milivojević (Prof, PhD) was born in Svrljig, Serbia, in 1959. He received the BE, ME, and PhD degrees in electrical engineering from the Faculty of Electronic Engineering, University of Nis, Nis, in 1984, 1994 and 2002, respectively. He is a Professor at Technical Faculty Bor University of Belgrade and Technical College in Niš. His primary research interests are digital signal processing and speech processing: algorithms and applications. He is the author and coauthor of over 120 journal and conference papers.

Darko Brodić (Ing, MaSc) was born in Konjic, BiH, Yugoslavia, in 1963. He received his BASc., and MASc from Faculty of Electrical Engineering, University of Sarajevo, in 1987 and 1990, respectively. Currently, he is with University of Belgrade, Technical Faculty Bor. He is working toward PhD thesis. His research interests include document image processing. He is the author and coauthor of over 40 journal and conference papers as well as books.