Computing and Informatics, Vol. 36, 2017, 449–469, doi: 10.4149/cai_2017_2_449

THREE-PARAMETRIC CUBIC INTERPOLATION FOR ESTIMATING THE FUNDAMENTAL FREQUENCY OF THE SPEECH SIGNAL

Zoran Milivojević, Nataša Savić

College of Applied Technical Sciences
Aleksandra Medvedeva 20
18000, Niš, Serbia
e-mail: {zoran.milivojevic, natasa.savic}@vtsnis.edu.rs

Darko Brodić

University of Belgrade
Technical Faculty Bor
Vojske Jugoslavije 12
19210 Bor, Serbia
e-mail: dbrodic@tfbor.bg.ac.rs

Abstract. In this paper, we propose a three-parametric convolution kernel which is based on the one-parameter Keys kernel. The first part of the paper describes the structure of the three-parameter convolution kernel. Then, a certain analytical expression for finding the position of the maximum of the reconstructed function is given. The second part presents an algorithm for estimating the fundamental frequency of the speech signal processing in the frequency domain using Picking Picks methods and parametric cubic convolution. Furthermore, the results of experiments give the estimated fundamental frequency of speech and sinusoidal signals in order to select the optimal values of the parameters of the proposed convolution kernel. The results of the fundamental frequency estimation according to the mean square error are given by tables and graphics. Consequently, it is used as a basis for a comparative analysis. The analysis derived the optimal parameters of the kernel and the window function that generates the least MSE. Results showed a higher efficiency in comparison to two or three-parameter convolution kernel.

Keywords: Frequency estimation, interpolation, signal processing, speech analysis, speech processing

Mathematics Subject Classification 2010: 65D05

1 INTRODUCTION

The estimation of the speech fundamental frequency f_0 has received an immense interest from different speech research areas, such as speech segregation, speech synthesis, speech coding, speech and speaker recognition, and speech articulation training for the deaf [1–3]. In many multimedia applications, it is necessary to process audio records in order to improve the quality and intelligibility of speech. A typical example is the quality improvement of the speech signal by reducing dissonant frequencies [4–6]. Accordingly, it is possible to classify:

1. The emotional state of humans (sadness, anger, joy, ...) [7],
2. the evaluation of health status, and
3. the hypoxia (manifested as a decrease in the concentration of oxygen in the blood due to incidents during a flight, working in mines, tunnels, etc.) [8].

A number of algorithms for determining the fundamental frequency have been developed. Their processing is performed in the time-domain (TD) and frequency-domain (FD) methods [9–14]. In TD methods, one or more speech features (the fundamental harmonic, a quasi-periodic time structure, an alternation of high and low amplitudes, and points of discontinuities in the speech waveform) are identified first, and then pitch markers or epochs are obtained in a pitch synchronous manner. In FD methods, a short-time frame or block of speech samples is transformed into spectral or frequency-domain in order to enhance the periodicity information contained in the speech. These methods determine an average pitch from several contiguous periods in the analysis frame. The performance of TD methods compared to FD methods depends more on the shape of the time waveform of speech [15]. The autocorrelation function (ACF) [16] and the average magnitude difference function (AMDF) [17] have been commonly employed for pitch estimation. In [9], an estimator named YIN has been proposed, where a series of modifications (a difference function formulation, normalization and parabolic interpolation) has been introduced to decrease the error rates in pitch estimation from a clean speech [18].

The widespread method for determination of the fundamental frequency is based on Picking Peaks of the amplitude characteristic in a specific frequency range. This method is used for analyzing the signal values in the spectrum at frequencies on which Discrete Fourier Transform (DFT) is calculated. Usually, the real value of the fundamental frequency is not there on the frequencies where DFT is calculated.

In contrast, it lays between the two spectrum samples. That causes the frequency estimation error that lies in the interval $[-f_s/2N\,\mathrm{Hz}, f_s/2N\,\mathrm{Hz}]$, 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 the interval between 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, etc.).

The calculation of the interpolation function by using Parametric Cubic Convolution (PCC) was represented in [19, 20]. The special case of PCC interpolation applied in computer graphics was called Catmull-Rom interpolation [21, 22]. In [23] gives detailed analysis of the fundamental frequency estimation and shows the advantage of PCC interpolation. The application of PCC interpolation for determining the fundamental frequency in specific conditions is presented in [24]. The efficiency of the algorithm for evaluating the fundamental frequency is determined by the simulation. As a quality measure of the algorithm, the mean square error (MSE) has been used. The best results were shown by the algorithm with the implemented Blackman window. The analysis of the algorithm efficiency where the signal-to-noise relation (SNR) is changeable according to the presence of the important harmonics in the fundamental function is shown in [25]. It confirmed the efficiency of the algorithm with the Blackman window. In [26] an analysis of PCC interpolation algorithm efficiency is made for the case where Greville two-parametric cubic convolution kernel (G2P) was implemented. The window was determined and the kernel parameters (α, β) were calculated where the minimal MSE was generated (in relation to Caltmull-Rom kernel the error was smaller for 58.1 %). The results of the fundamental frequency estimation of the speech signal modeled by SYMPES method (Systematic Procedure to Model Speech Signals via Predefined Envelope and Signature Sequences) [27] are shown in [28]. Furthermore, the results of the fundamental frequency estimation for the speech signal coded by G.3.721 method are shown in [29]. The aforementioned papers showed that the accuracy of estimated fundamental frequency using PCC interpolation depends on: a) window functions and b) the interpolation kernel.

PCC interpolation is based on the parameter interpolation kernel. It is possible to minimize the estimation error of the fundamental frequency by the proper selection of the kernel parameter. The first kernel is proposed by Keys [19]. It is commonly referred as a Keys one-parametric (1P) kernel [20]. Detailed analysis of the Keys 1P kernel effect on the accuracy to the estimated fundamental frequency F_0 is given in [23]. It is shown that determination of the optimal parameter and minimization of the estimation error MSE is possible. To reduce the numerical complexity and increase the speed, the new formula for direct calculation of the position of the maximum in the spectrum of speech signal without the use of convolution is given [23]. In [21], the two-parameter (2P) convolution kernel called Greville 2P is proposed. It results in greater accuracy of the f_0 estimation compared to estimates with respect to Keys 1P kernel [31].

Perceiving the presented results of the study the following question is of the interest: *Is it possible to increase the accuracy of the estimates using the kernel with a larger number of parameters without increasing the numerical complexity compared to 1P and 2P kernel?* For this purpose, the authors constructed the 2P and 3P kernel based on Keys 1P kernel. Then, they conducted a comprehensive investigation in order to determine the optimum parameters and window functions.

The following part of the paper deals with:

1. The construction of three-parameter convolution kernel which is based on the one-parameter Keys kernel [19]. In addition, the analytical expression is determined by estimating the position of the reconstructed function maximum.
2. The algorithm for estimating the fundamental frequency of speech signal based on the application of Picking Picks algorithm in the spectral domain and parametric cubic convolution.
3. The results of experiments conducted in order to choose the optimal parameters of the kernel and analyze the efficiency of the proposed three-parameter compared to one and two-Keys kernel. The analysis is performed for a standard window functions. In the experiment, the fundamental frequency of the following signals is determined: i) mathematically generated Sine signal [23], and ii) real Speech signals recorded in a real environment [29]. The results were compared to the results previously published in [30].
4. The results of the fundamental frequency for three-parameter convolution kernel in terms of superposed Additive White Gaussian Noise (AWGN) are presented.

The organization of paper is as follows. Section 2 shows the three-parameter PCC kernel. Section 3 denounces algorithm that estimates the fundamental frequency of the speech signal. Section 4 presents the experimental results. Section 5 presents a comparative analysis. Section 6 makes conclusions.

2 KEYS 3P-PCC KERNEL

The three-parameter cubic convolution (3P-PCC) Keys kernel $r(f)$ is defined as the composition of the third degree polynomials for the parts of the interval $[-4,-3)$, $[-3,-2)$, $[-2,-1)$, $[-1,-0]$, $[0,1]$, $(1,2]$, $(2,3]$, $(3,4]$:

$$r(f) = \begin{cases} a_3|f|^3 + a_2|f|^2 + a_1|f| + a_0, & 0 \le |f| \le 1, \\ b_3|f|^3 + b_2|f|^2 + b_1|f| + b_0, & 1 < |f| \le 2, \\ c_3|f|^3 + c_2|f|^2 + c_1|f| + c_0, & 2 < |f| \le 3, \\ d_3|f|^3 + d_2|f|^2 + d_1|f| + d_0, & 3 < |f| \le 4, \\ 0, & |f| > 4. \end{cases} \tag{1}$$

Let suppose that the interpolation kernel is continuous and differentiable. The kernel must satisfy the criteria $r(0) = 1$ and $r(k) = 0$ when k is any nonzero integer. Then:

$$\begin{aligned} r(0) = 1 \quad &\Rightarrow \quad a_0 = 1, \\ r(1) = 0 \quad &\Rightarrow \quad a_3 + a_2 + a_1 + a_0 = 0, \\ r(2) = 0 \quad &\Rightarrow \quad 8b_3 + 4b_2 + 2b_1 + b_0 = 0, \\ r(3) = 0 \quad &\Rightarrow \quad 27c_3 + 9c_2 + 3c_1 + c_0 = 0, \\ r(4) = 0 \quad &\Rightarrow \quad 64d_3 + 16d_2 + 4d_1 + d_0 = 0. \end{aligned} \tag{2}$$

From the condition of continuity and differentiability follows that:

$$\lim_{f\to 1^-} r(f) = \lim_{f\to 1^+} r(f) \Rightarrow a_3 + a_2 + a_1 + a_0 = b_3 + b_2 + b_1 + b_0, \tag{3}$$

$$\lim_{f\to 2^-} r(f) = \lim_{f\to 2^+} r(f) \Rightarrow 8b_3 + 4b_2 + 2b_1 + b_0 = 8c_3 + 4c_2 + 2c_1 + c_0, \tag{4}$$

$$\lim_{f\to 3^-} r(f) = \lim_{f\to 3^+} r(f) \Rightarrow 27c_3 + 9c_2 + 3c_1 + c_0 = 27d_3 + 9d_2 + 3d_1 + d_0, \tag{5}$$

$$r'(0) = 0 \Rightarrow a_1 = 0, \tag{6}$$

$$\lim_{f\to 1^-} r'(f) = \lim_{f\to 1^+} r'(f) \Rightarrow 3a_3 + 2a_2 + a_1 = 3b_3 + 2b_2 + b_1, \tag{7}$$

$$\lim_{f\to 2^-} r'(f) = \lim_{f\to 2^+} r'(f) \Rightarrow 12b_3 + 4b_2 + b_1 = 12c_3 + 4c_2 + c_1, \tag{8}$$

$$\lim_{f\to 3^-} r'(f) = \lim_{f\to 3^+} r'(f) \Rightarrow 27c_3 + 6c_2 + c_1 = 27d_3 + 6d_2 + d_1, \tag{9}$$

$$r'(4) = 0 \Rightarrow 48d_3 + 8d_2 + d_1 = 0. \tag{10}$$

The system of Equations (2)–(10) has 13 equations and 16 unknowns. Therefore, the three variables may be arbitrary chosen ($b_3 = \alpha, c_3 = \beta, d_3 = \gamma$). The solutions of this system of equations are:

$$a_1 = 0; \quad a_2 = -\alpha + \beta - \gamma - 3; \quad a_3 = \alpha - \beta + \gamma + 2, \tag{11}$$

$$b_0 = -4\alpha + 2\beta - 2\gamma; \quad b_1 = 8\alpha - 3\beta + 3\gamma; \quad b_2 = -5\alpha + \beta - \gamma; \quad b_3 = \alpha, \tag{12}$$

$$c_0 = -18\beta + 6\gamma; \quad c_1 = 21\beta - 5\gamma; \quad c_2 = -8\beta + \gamma; \quad c_3 = \beta, \tag{13}$$

$$d_0 = -48\gamma; \quad d_1 = 40\gamma; \quad d_2 = -11\gamma; \quad d_3 = \gamma, \tag{14}$$

where α, β and γ are arbitrary real numbers. Substituting the solution (11)–(14) into (1) gives the final form of the 3P-PCC Keys kernel:

$$r(f)=\begin{cases}(\alpha-\beta+\gamma+2)|f|^3+(-\alpha+\beta-\gamma-3)|f|^2+1, & 0\leq |f|\leq 1,\\ \alpha|f|^3+(-5\alpha+\beta-\gamma)|f|^2+(8\alpha-3\beta+3\gamma)|f|, & \\ +(-4\alpha+2\beta-2\gamma), & 1<|f|\leq 2,\\ \beta|f|^3+(-8\beta+\gamma)|f|^2+(21\beta-5\gamma)|f|, & \\ +(-18\beta+6\gamma), & 2<|f|\leq 3,\\ 0, & |f|>3.\end{cases} \tag{15}$$

3 ALGORITHM FOR THE ESTIMATION OF THE FUNDAMENTAL FREQUENCY

The algorithm for the estimation of the fundamental frequency (ALGORITHM 1) is based on the algorithm from [23]. This algorithm is realized as follows:

ALGORITHM 1:
Input: Speech signal $\mathbf{x}\in\mathbb{R}^N$ where N is a number of speech samples.
Output: The estimated fundamental frequency f_e.
Step 1: Window $\mathbf{w}\in\mathbb{R}^N$ has been applied to modifying signal $\mathbf{x}$. *Step 2*: Spectrum $\mathbf{X}$ is calculated by using DFT:

$$\mathbf{X}=DFT(\mathbf{x}). \tag{16}$$

The spectrum $\mathbf{X}$ is calculated in discrete points $k=0,\dots,N-1$, where N is the length of DFT.
Step 3: The maximum of the real spectrum that is between k^{th} and $(k+1)^{\text{th}}$ samples is determined using the picking peak algorithm. The values $X(k)$ and $X(k+1)$ are the highest in the specified domain.
Step 4: The maximum of the spectrum is calculated by PCC interpolation. The reconstructed function is:

$$X_r(f)=\sum_{i=k+1-\frac{L}{2}}^{k+\frac{L}{2}} p_i r(f-i), k\leq f\leq k+1, \tag{17}$$

where $p_i=X_r(i)$, $r(f)$ is the kernel of interpolation, and L is the number of samples that participate in the interpolation.
Step 5: By differentiation $X_r(f)$ and zero adjustment, the position of the maximum is determined. It presents the estimated fundamental frequency f_e.

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

$$MSE=\overline{(f_0-f_e)^2}, \tag{18}$$

where f_0 is true fundamental frequency and f_e is estimated fundamental frequency. It is possible to adjust the parameter for reductions estimation error by applying the parametric interpolation of the convolution kernels.

3.1 Interpolation Kernels

The definitions of the interpolation kernels, which are tested in this paper, are:

3.1.1 1P-PCC Keys Kernel

Keys interpolation kernel is given as [19, 20]:

$$r(f) = \begin{cases} (\alpha+2)|f|^3 - (-\alpha+3)|f|^2 + 1, & 0 \le |f| \le 1, \\ \alpha|f|^3 - 5\alpha|f|^2 + 8\alpha|f| - 4\alpha, & 1 \le |f| \le 2, \\ 0, & |f| > 2. \end{cases} \tag{19}$$

The maximum of the reconstructed function $X_r(f)$ (see Equation (17)) is found by differentiating in spectrum domain and equalizing the first derivative with zero. For $L = 4$ from (17) the position of the maximum is:

$$f_{max} = \begin{cases} k - \frac{c}{2b}, & \text{a=0}, \\ k + \frac{-b \pm \sqrt{b^2 - ac}}{a}, & a \neq 0, \end{cases} \tag{20}$$

where k is the position of the maximum component in the spectrum, whereas:

$$\begin{aligned} a &= 2(\alpha p_{k-1} + (\alpha+2)p_k - (\alpha+2)p_{k+1} - \alpha p_{k+2}), \\ b &= -2\alpha p_{k-1} - (\alpha+3)p_k - (2\alpha+3)p_{k+1} - \alpha p_{k+2}, \\ c &= -\alpha p_{k-1} - \alpha p_{k+1}. \end{aligned} \tag{21}$$

3.1.2 2P-PCC Keys Kernell

We designed the 2P-PCC Keys kernel as:

$$r(f) = \begin{cases} (\alpha-\beta+2)|f|^3 + (-\alpha+\beta-3)|f|^2 + 1, & 0 \le |f| \le 1, \\ \alpha|f|^3 + (-5\alpha+\beta)|f|^2 + (8\alpha-3\beta)|f| + (-4\alpha+2\beta), & 1 < |f| \le 2, \\ \beta|f|^3 - 8\beta|f|^2 + 21\beta|f| - 18\beta, & 2 < |f| \le 3, \\ 0, & |f| > 3. \end{cases} \tag{22}$$

The position of maximum is determined by Equation (17) and Equation (20) for $L = 6$, where the following is valid:

$$\begin{aligned}
a &= 3(\beta p_{k-2} + \alpha p_{k-1} + (\alpha - \beta - 2)p_k \\
&\quad -(\alpha - \beta - 2)p_{k+1} - \alpha p_{k+2} - \beta p_{k+3}), \\
b &= -4\beta p_{k-2} - (4\alpha - 2\beta)p_{k-1} - (2\alpha - 2\beta + 6)p_k \\
&\quad +(4\alpha - 4\beta + 6)p_{k+1} + (2\alpha + 2\beta)p_{k+2} + 2\beta p_{k+3}, \\
c &= \beta p_{k-2} + (\alpha - \beta)p_{k-1} - (\alpha - \beta)p_{k+1} - \beta p_{k+2}.
\end{aligned} \tag{23}$$

3.1.3 3P-PCC Keys Kernel

The 3P-PCC Keys kernel is given in Equation (15). The position of maximum is determined by Equation (17) and Equation (20) for $L = 8$, where the following is valid:

$$\begin{aligned}
a &= 3(\gamma p_{k-3} + \beta p_{k-2} + \alpha p_{k-1} + (\alpha - \beta + \gamma - 2)p_k \\
&\quad -(\alpha - \beta + \gamma - 2)p_{k+1} - \alpha p_{k+2} - \beta p_{k+3}) - \gamma p_{k+4}), \\
b &= -4\gamma p_{k-3} + (-4\beta + 2\gamma)p_{k-2} + (-4\alpha + 2\beta - 2\gamma)p_{k-1} \\
&\quad +(-2\alpha + 2\beta - 2\gamma - 6)p_k - (\alpha - \beta + \gamma - 2)p_{k+1} \\
&\quad +(2\alpha + 2\beta - 2\gamma)p_{k+2} + (2\beta + 2\gamma)p_{k+3} + 2\gamma p_{k+4}), \\
c &= \gamma p_{k-3} + (-\gamma + \beta)p_{k-2} + (\alpha - \beta + \gamma)p_{k-1} \\
&\quad -(\alpha - \beta + \gamma)p_{k+1} - (\beta - \gamma)p_{k+2} - \gamma p_{k+3}.
\end{aligned} \tag{24}$$

In the Equations (19)–(21), (23) and (24) there are α, β and γ parameters. The optimal values of these parameters will be determined by the minimal value of MSE, for 1P-PCC (Equation (19)), 2P-PCC (Equation (22)) and 3P-PCC (Equation (15)) Keys kernel.

The optimal parameters for 1P-PCC Keys kernel is given as:

$$\alpha_{opt} = \arg\min_{\alpha}(MSE), \tag{25}$$

for 2P-PCC Keys kernel is given as:

$$(\alpha_{opt}, \beta_{opt}) = \arg\min_{\alpha,\beta}(MSE), \tag{26}$$

and for 3P-PCC Keys kernel is given as:

$$(\alpha_{opt}, \beta_{opt}, \gamma_{opt}) = \arg\min_{\alpha,\beta,\gamma}(MSE). \tag{27}$$

The detailed analysis in [22-28] 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.

3.2 Interpolation Kernels Parameters

The algorithm for determination of interpolation kernel parameters α, β and γ is realized as follows:

ALGORITHM 2:
Input: Speech signal $\mathbf{x} \in \mathbb{R}^N$ where N is a number of speech samples. Fundamental frequency f_0.
Output: Kernel parameters α, β and γ.
Step 1: Window $\mathbf{w} \in \mathbb{R}^N$ has been applied to modify signal $\mathbf{x}$.
Step 2: Spectrum $\mathbf{X}$ is determined by the application of DFT.
Step 3: Reconstruction of the continual function X_r that represents spectrum $\mathbf{X}$ is performed by the application of PCC interpolation.
Step 4: $MSE(f_0, f_e)$ is calculated for various values of parameters α_{opt}, β_{opt} and γ_{opt} depending on the implemented window.
Step 5: α_{opt}, β_{opt} and γ_{opt} are determined for which the minimal value of MSE is obtained.

4 EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experiment

An experiment was carried out in order to determine the optimal value of parameters 1P-PCC, 2P-PCC, and 3P-PCC Keys kernels for a) Hamming, b) Hanning, c) Blackman, d) rectangular, e) Kaiser, and f) triangular window. For this purpose we use the algorithm described in Section 3. In the second part of the experiment, the effectiveness of the fundamental frequency estimation is analyzed, when the AVGM is superposed. For various values of SNR, the MSE is determined in accordance to the fundamental frequency estimation. All aforementioned determines the performance of the algorithm.

4.1.1 Test Signals

In the experiment, the test signals are given as

1. simulation Sine test signal, and
2. real Speech test signal.

Sine test signal is defined in [22]:

$$s(t) = \sum_{i=1}^{K} \sum_{g=0}^{M} a_i \sin\left(2\pi i \left(f_0 + g\frac{f_0}{KM}\right) t + \theta_i\right), \tag{28}$$

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, and M is the number of points between the two samples in spectrum where PCC interpolation is being made. The real Speech test signal is obtained by recording of a speaker in the real acoustic ambient [30].

4.1.2 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 $[0, 2\pi]$ with Sine and real Speech test signals. Signal frequency of sampling is $f_s = 8\,\text{kHz}$, and the length of window is $N = 256$, which assures the analysis of subsequences with the length of 32 ms. Furthermore, the results will relate to $f_0 = (125\text{–}140.625)\,\text{Hz}$ (frequencies between the 8^{th} and 9^{th} 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 analyses will relate to: a) Hamming, b) Hanning, c) Blackman, d) rectangular, e) Kaiser, and f) triangular window.

4.2 Experimental Results

4.2.1 1P-PCC Keys Kernel

Applying the algorithm for the parameters determination of Keys 1P-PCC interpolation kernel some diagrams MSE(α) are drawn (Figure 1 Sine test signal and Figure 2 Speeach test signal), the minimum value MSE is determined ($\text{MSE}_{\text{K\_1P\_SIN\_min}}$ Sine test signal, $\text{MSE}_{\text{K\_1P\_SP\_min}}$ Speech test signal), and on the base of it, the optimum value of Keys 1P-PCC kernel α_{opt} is determined for: a) Hamming, b) Hann, c) Blackman, d) Kaiser, and e) triangular window functions. Values of the MSE and α_{opt} are presented in Table 1 ($\text{MSE}_{\text{K\_1P\_SIN\_min}}$ Sine test signal) and Table 2 ($\text{MSE}_{\text{K\_1P\_SP\_min}}$ Speech test signal).

According to the results presented in Table 1 and Table 2, it is obvious that:

1. In Sine test signal, the greatest precision is given by the Blackman window ($\text{MSE}_{\text{K\_1P\_SIN\_min}} = 4.3616 \cdot 10^{-4}$). The minimum precision is obtained by the rectangular window in ($\text{MSE}_{\text{K\_1P\_SIN\_min}} = 0.1805$).
2. In Speech test signal, the greatest precision is given by the triangular window ($\text{MSE}_{\text{K\_1P\_SP\_min}} = 0.0271$). The lowest accuracy is obtained by rectangular windows ($\text{MSE}_{\text{K\_1P\_SP\_min}} = 0.6087$).

3. The estimated accuracy of the Sine (Blackman window) in relation to Speech (triangular window) is larger $\mathrm{MSE_{K\_1P\_SP\_min}}/\mathrm{MSE_{K\_1P\_SIN\_min}} = 0.0271/4.3616 \cdot 10^{-4} = 62.3331$ times.

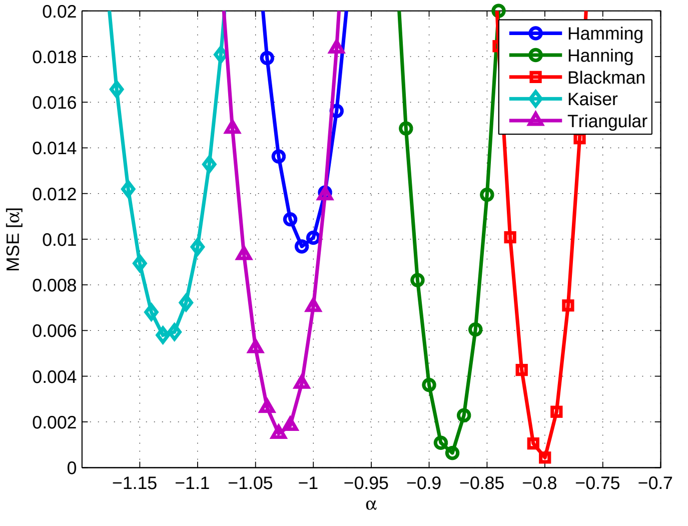


Figure 1. $\mathrm{MSE_{K\_1P\_SIN\_min}}(\alpha)$ for Keys kernel for Sine test signal

Window	α_{opt}	$\mathrm{MSE_{K\_1P\_SIN\_min}}$
Hamming	-1.0100	0.0097
Hann	-0.8800	$6.3836 \cdot 10^{-4}$
Blackman	-0.8000	$4.3616 \cdot 10^{-4}$
rectangular	-2.6400	0.1805
Kaiser	-1.1300	0.0058
triangular	-1.0300	0.0015

Table 1. Minimum MSE and α_{opt} for Sine test signal

4.2.2 2P-PCC Keys Kernel

By applying the algorithm to the 2P PCC Keys interpolation kernel, the optimal values α_{opt} and β_{opt} are determined for: a) Hamming, b) Hann, c) Blackman, d) Kaiser, and e) triangular window functions. The optimal values of the parameters and the minimum MSE values are shown in Table 3 ($\mathrm{MSE_{K\_2P\_SIN\_min}}$ Sine test signal) and Table 4 ($\mathrm{MSE_{K\_2P\_SP\_min}}$ Speech test signal). In both cases, the minimum value

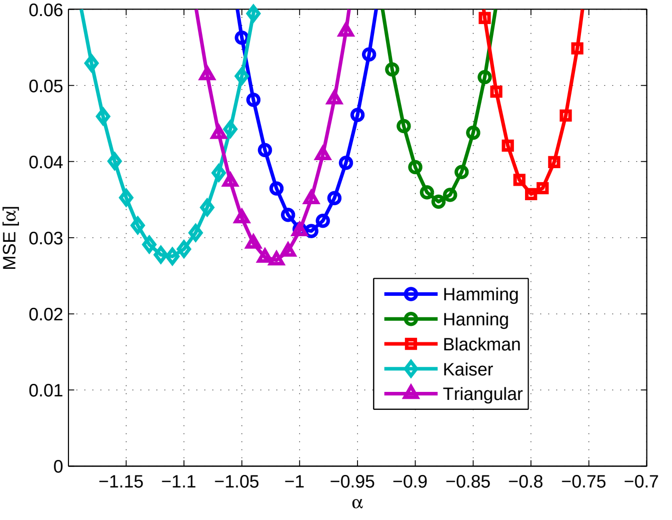


Figure 2. $\mathrm{MSE}_{\mathrm{K\_1P\_SP\_min}}(\alpha)$ for Keys 1P-PCC kernel for Speech test signal

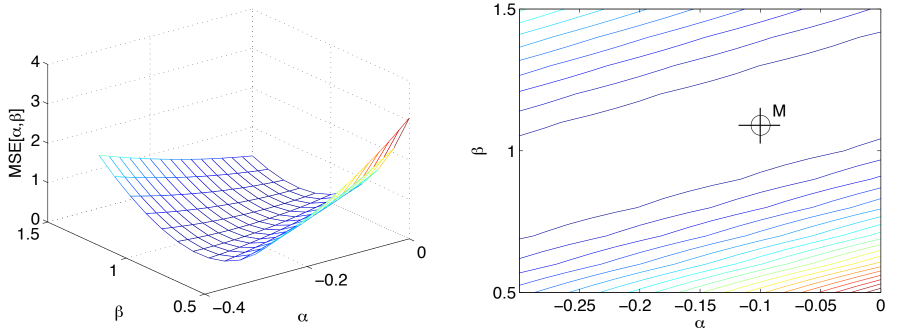
Window	α_{opt}	$\mathrm{MSE}_{\mathrm{K\_1P\_SP\_min}}$
Hamming	−0.9900	0.0309
Hann	−0.8800	0.0347
Blackman	−0.8000	0.0357
rectangular	−2.2700	0.6087
Kaiser	−1.1100	0.0275
triangular	−1.0200	0.0271

Table 2. Minimum MSE and α_{opt} for Speech test signal

of MSE is obtained for the cases of triangular windows. Charts for $\mathrm{MSE}(\alpha, \beta)$ are shown in Figure 3 a) (Sine test signal) and Figure 4 a) (Speech test signal). Minimum positions of the $\mathrm{MSE}(\alpha_{opt}, \beta_{opt})$ in (α, β) plane (point $\mathbf{M}$) are shown in Figure 3 b) (Sine test signal) and Figure 4 b) (Speech test signal).

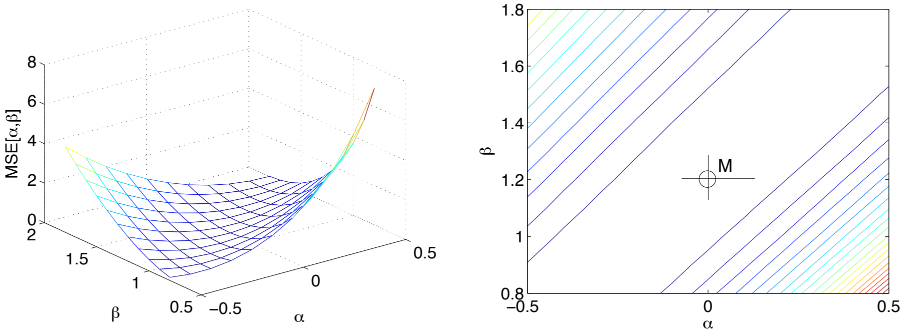
Window	α_{opt}	β_{opt}	$\mathrm{MSE}_{\mathrm{K\_2P\_SIN\_min}}$
Hamming	2.5500	4.6000	0.0013
Hann	−1.4500	−0.8000	$3.0273 \cdot 10^{-4}$
Blackman	−0.7200	0.1800	$1.8042 \cdot 10^{-4}$
rectangular	−1.8000	0.9600	0.1514
Kaiser	−1.0200	0.1200	0.0053
triangular	−0.1000	1.1000	$7.8770 \cdot 10^{-5}$

Table 3. Minimum MSE, α_{opt} and β_{opt} for Sine test signal



a) b)

Figure 3. Sine Test Signal with the application of triangular window; a) MSE(α, β) for the application of 2P-PCC Keys interpolation, b) positions of min (MSE$(\alpha_{opt}, \beta_{opt})$) in plane (α, β) for 2P-PCC Keys interpolation (point **M**)



a) b)

Figure 4. Speech Test Signal with the application of triangular window; a) MSE(α, β) for the application of 2P-PCC Keys interpolation, b) positions of min (MSE$(\alpha_{opt}, \beta_{opt})$) in plane (α, β) for 2P-PCC Keys interpolation (point **M**)

Window	α_{opt}	β_{opt}	$\mathrm{MSE}_{\mathrm{K\_2P\_SP\_min}}$
Hamming	−2	−1.3000	0.0284
Hann	−1.3000	−0.6000	0.0302
Blackman	−0.7000	0.2000	0.0357
rectangular	0.1000	2.9000	0.2884
Kaiser	−0.5500	0.6500	0.0257
triangular	0	1.2000	0.0255

Table 4. Minimum MSE, α_{opt} and β_{opt} for Speech test signal

1. In the Sine test signal, the greatest precision are in the triangular window ($MSE_{K\_2P\_SIN\_min} = 7.8770 \cdot 10^{-5}$). The minimum precision is given by the rectangular window ($MSE_{K\_2P\_SIN\_min} = 0.1514$).
2. In Speech test signal, the greatest precision is given by the triangular window ($MSE_{K\_2P\_SP\_min} = 0.0255$). The lowest accuracy is in the rectangular windows ($MSE_{K\_2P\_SP\_min} = 0.2884$).
3. The estimation accuracy of the Sine (triangular window) relative to Speech (triangular window) is larger $MSE_{K\_2P\_SP\_min}/MSE_{K\_2P\_SIN\_min} = 0.0255/7.8770 \cdot 10^{-5} = 323.727$ times.

4.2.3 3P-PCC Keys Kernel

By applying the algorithm in order to determine the parameters of 3P-PCC interpolation Keys kernel, the optimal values $\alpha_{opt}, \beta_{opt}$ and γ_{opt} are obtained for a) Hamming, b) Hann, c) Blackman, d) rectangular, e) Kaiser, and f) triangular window functions. The optimal values of the parameters and the minimum MSE values are shown in Table 5 ($MSE_{K\_3P\_SIN\_min}$, Sine test signal) and Table 6 ($MSE_{K\_3P\_SP\_min}$, Speech test signal).

Window	α_{opt}	β_{opt}	γ_{opt}	$MSE_{K\_3P\_SIN\_min}$
Hamming	2.5800	4.7400	0.1000	0.0013
Hann	−1.9500	−1.6000	−0.0900	$9.1211 \cdot 10^{-5}$
Blackman	−0.6200	0.2800	−0.1000	$8.2038 \cdot 10^{-5}$
rectangular	−1.4500	1	−0.3400	0.1485
Kaiser	−0.7000	0.1000	−0.3900	0.0039
triangular	−0.0800	1.4200	0.2900	$3.0849 \cdot 10^{-5}$

Table 5. Minimum MSE, α_{opt}, β_{opt} and γ_{opt} for Sine test signal

Window	α_{opt}	β_{opt}	γ_{opt}	$MSE_{K\_3P\_SP\_min}$
Hamming	−1.7000	−4.7000	−3.8000	0.0119
Hann	−2.3000	−2.5000	−0.5000	0.0202
Blackman	−2.3000	0.2000	3.2000	0.0182
rectangular	−0.2000	4.6000	2	0.1329
Kaiser	0.9000	−0.8000	−3.1000	0.0035
triangular	1.6000	1.7000	−1.4000	0.0212

Table 6. Minimum MSE, α_{opt}, β_{opt} and γ_{opt} for Speech test signal

1. In the Sine test signal, the greatest precision in given by the triangular window ($MSE_{K\_3P\_SIN\_min} = 3.0849 \cdot 10^{-5}$). The minimum precision is obtained by the rectangular window ($MSE_{K\_3P\_SIN\_min} = 0.1485$).

2. In Speech test signal, the greatest precision is given by the triangular window ($MSE_{K\_3P\_SP\_min} = 0.0035$). The lowest accuracy is obtained the rectangular windows ($MSE_{K\_3P\_SP\_min} = 0.1329$).

3. The estimated accuracy of the Sine (triangular window) relative to Speech (triangular window) is larger $MSE_{K\_3P\_SP\_min}/MSE_{K\_3P\_SIN\_min} = 0.0035/3.0849 \cdot 10^{-5} = 113.455$ times.

4.2.4 3P-PCC Keys Kernel for AWGN Speech

To the Sine and Speech test signal the Additive white Gaussian noise AWGN for SNR = 0-50 dB is superimposed. In order to estimate the fundamental frequency, 3P-PCC Keys kernel with optimal parameters $\alpha_{opt}, \beta_{opt}$ and γ_{opt} is used (Table 5 and Table 6). MSE values are shown for Sine test signal (Table 7, Figure 5) and Speech test signal (Table 8, Figure 6).

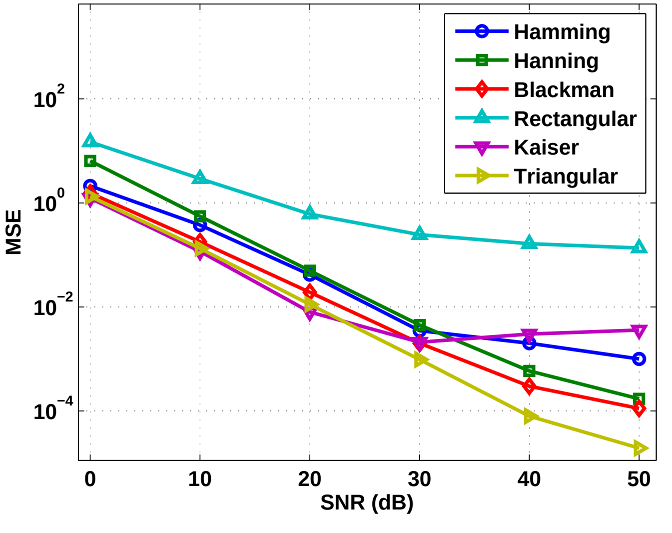


Figure 5. MSE according to the SNR of a window function for a test Sine signal

The analysis of the MSE shows that decreasing SNR leads to decreasing MSE. From the theoretical point of view, it is the expected result. In extremely adverse conditions (SNR = 0 dB), the minimum MSE is obtained by Kaiser ($\alpha_{opt} = -0.7, \beta_{opt} = 0.1$ and $\gamma_{opt} = 0.39$, Sine test signal) and Hamming window ($\alpha_{opt} = -1.7, \beta_{opt} = -4.7$ and $\gamma_{opt} = -3.8$, Speech test signal).

Window	SNR [dB]					
	0	10	20	30	40	50
Hamming	2.1195	0.3763	0.0416	0.0035	0.002	0.001
Hann	6.4151	0.5508	0.0499	0.0045	5.8770e−004	1.7148e−004
Blackman	1.5472	0.1790	0.0191	0.0020	3.0010e−004	1.1152e−004
rectangular	14.8808	2.9171	0.6077	0.2444	0.1641	0.1358
Kaiser	1.2260	0.1169	0.0080	0.0021	0.0030	0.0036
triangular	1.3318	0.1334	0.0112	9.7357e−004	7.9743e−005	1.9322e−005

Table 7. MSE according to of SNR for Sine test signal for $\alpha_{opt}, \beta_{opt}$ and β_{opt}

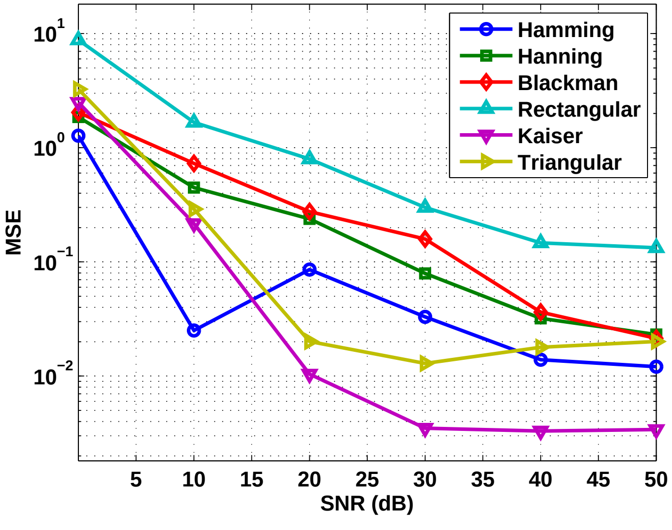


Figure 6. MSE according to the SNR for Speech test signal

Window	SNR [dB]					
	0	10	20	30	40	50
Hamming	1.2770	0.0251	0.0858	0.0330	0.0139	0.0121
Hann	1.8615	0.4479	0.2384	0.0793	0.0320	0.0231
Blackman	2.0423	0.7303	0.2757	0.1587	0.0363	0.0214
rectangular	8.8040	1.6701	0.7968	0.3002	0.1466	0.1332
Kaiser	2.4989	0.2169	0.0104	0.0035	0.0033	0.0034
triangular	3.2579	0.2907	0.0201	0.0129	0.0179	0.0201

Table 8. MSE according to of SNR for Speech test signal for $\alpha_{opt}, \beta_{opt}$ and β_{opt}

5 COMPARATIVE ANALYSIS

The comparative analysis of the estimated fundamental frequency for the Sine test signal and the real Speech test signal, will be performed on the base of MSE minimal values. The minimal value of MSE is determined on the base of the diagram in Figures 1–2 (1P-PCC) and Figures 3–4 (2P-PCC). It is presented in Table 1 ($MSE_{K\_1P\_SIN\_min}$), Table 2 ($MSE_{K\_1P\_SP\_min}$), Table 3 ($MSE_{K\_2P\_SIN\_min}$), Table 4 ($MSE_{K\_2P\_SP\_min}$), Table 5 ($MSE_{K\_3P\_SIN\_min}$) and Table 6 ($MSE_{K\_3P\_SP\_min}$). MSE according to of SNR are presented in the Table 7 ($MSE_{K\_3P\_SIN\_min}$) and Table 8 ($MSE_{K\_3P\_SP\_min}$).

Comparing the values MSE_{min} from Tables 1-6, it can be concluded that:

1. The optimal choice for Sine test signal is 3P-PCC and triangular window. Compared to the 1P-PCC (Blackman) and 2P-PCC (triangular), 3P-PCC (triangular) generates 92.928 % and 60.83 % lower MSE value, respectively.
2. The optimal choice for Speech test signal is 3P-PCC and Kaiser window. Compared to the 1P-PCC (triangular) and 2P-PCC (Kaiser), 3P-PCC (Kaiser) generates 87.084 % and 86.381 % smaller MSE value, respectively.
3. The accuracy of the assessment of speech signals (3P-PCC, Kaiser) in relation to the Sine signals (3P-PCC, triangular) is $MSE_{K\_3P\_SP\_min}/MSE_{K\_3P\_SIN\_min} = 0.0035/3.0849 \cdot 10^{-5} = 113.45$ times smaller.
4. The reduction of the SNR MSE is increased. In Sine test signals for SNR = 0 dB (Kaiser window) the accuracy is lower than the SNR = 50 (triangular window) $1.2260/3.0849 \cdot 10^{-5} = 39.8 \cdot 10^3$ times. In Speech test signal for SNR = 0 dB (Hamming window) the accuracy is lower than the SNR = 50 (Kaiser window) $1.277/0.0034 = 375.58$ times.
5. In respect to the Greville 2P kernel (Blackman window, $\alpha_{opt} = -0.42, \beta_{opt} = 0.002$, $MSE_{min} = 0.000377$) Keys 3P-PCC has a $37.7 \cdot 10^{-5}/3.0849 \cdot 10^{-5} = 12.22$ times greater precision [30].

Accordingly, the obtained results recommend the use of PCC algorithm with 3P-PCC kernel with the Kaiser windows in preprocessing stage of the speech signal. Hence, it is recommended for further processing by algorithms that require a precise determination of the fundamental frequency (automatic verification of a speaker, recognition of the speech, etc.).

6 CONCLUSION

This paper describes the design of the three-parameter (3P) convolution kernel. The kernel is specifically designed for the estimation of the speech signal fundamental frequency by interpolation. The experiment was carried out with the aim to determine the optimal convolution kernel parameters in order to minimize the estimation error. It is implemented on the example of some classical window function. Detailed analysis shows the superiority of the 3P-PCC kernel. The optimal choice for real speech signal is 3P PCC Keys kernel with the Kaiser window function. By comparing with 1P-PCC kernel with triangular window, MSE is lower by 87.084 %. Furthermore, MSE is lower 86.381 % by comparing to the 2P-PCC with a Kaiser windowed. Direct comparison of MSE for Sine test signal and Speech test signal, the precision of the fundamental frequency estimation is 113.45 times higher in the Sine test signals. The aforementioned results show the superiority of 3P-PCC kernel compared to 1P-PCC and 2P-PCC kernel.

REFERENCES

[1] Griffin, D.—Lim, J.: Multiband Excitation Vocoder. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 36, 1988, No. 8, pp. 1223–1235, doi: 10.1109/29.1651.

[2] Atal, B.: Automatic Speaker Recognition Based on Pitch Contours. Journal of the Acoustical Society of America, Vol. 52, 1972, No. 6, pp. 1887–1697.

[3] Kawahara, H.—Masuda-Katsuse, I. —Cheveigné, A.: Restructuring Speech Representations Using Pitch-Adaptive Time-Frequency Smoothing and an Instantaneous-Frequency-Based f0 Extraction: Possible Role of a Repetitive Structure in Sounds. Speech Communication, Vol. 27, 1999, No. 3-4, pp. 187–207, doi: 10.1016/s0167-6393(98)00085-5.

[4] Joen, B.—Kang, S.—Baek, S.—Sung, K.: Filtering of a Dissonant Frequency Based on Improved Fundamental Frequency Estimation for Speech Enhancement. IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E86-A, 2003, No. 8, pp. 2063–2064.

[5] Kang, S.: Dissonant Frequency Filtering Technique for Improving Perceptual Quality of Noisy Speech and Husky Voice. Signal Processing, Vol. 84, 2004, No. 8, pp. 431–433, doi: 10.1016/j.sigpro.2003.10.023.

[6] Kang, S.—Kim, Y.: A Dissonant Frequency Filtering for Enhanced Clarity of Husky Voice Signals. Springer, Berlin, Lecture Notes in Computer Science, Vol. 4188, 2006, pp. 517–522, doi: 10.1007/11846406_65.

[7] Ayadi, M.—Kamel, M.—Karray, F.: Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases. Pattern Recognition, Vol. 44, 2011, No. 8, pp. 572–587.

[8] Milivojević, Z.—Milivojević, M.—Brodić, D.: The Effects of the Acute Hypoxia to the Fundamental Frequency of the Speech Signal. Advances in Electrical and Computer Engineering AECE, Vol. 12, 2012, No. 2, pp. 57–60, doi: 10.4316/aece.2012.02010.

[9] Cheveigné, A.—Kawahara, C.: YIN, A Fundamental Frequency Estimator for Speech and Music. Journal of the Acoustical Society of America, Vol. 111, 2002, No. 4, pp. 1917–1930.

[10] Sekhar, S.—Sreenivas, T.: Effect of Interpolation on PWVD Computation and Instantaneous Frequency Estimation. IEEE Signal Processing, Vol. 84, 2004, pp. 107–116, doi: 10.1016/j.sigpro.2003.07.015.

[11] Hussain, Z.—Boashash, B.: Adaptive Instantaneous Frequency Estimation of Multicomponent Signals Using Quadratic Time-Frequency Distributions. IEEE Transaction on Signal Processing, Vol. 50, 2002, No. 8, pp. 1866–1876, doi: 10.1109/tsp.2002.800406.

[12] Kacha, F.—Benmahammed, G.: Time-Frequency Analysis and Instantaneous Frequency Estimation Using Two-Sided Linear Prediction. IEEE Signal Processing, Vol. 85, 2005, pp. 491–503, doi: 10.1016/j.sigpro.2004.10.015.

[13] Veprek, P.—Scordilis, M.: Analysis, Enhancement and Evaluation of Five Pitch Determination Techniques. Speech Communication, Vol. 37, 2002, No. 3-4, pp. 2491–270, doi: 10.1016/s0167-6393(01)00017-6.

[14] Klapuri, A.: Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 11, 2003, No. 6, pp. 804–816, doi: 10.1109/tsa.2003.815516.

[15] Resch, B.—Nilsson, M.—Ekman, A.—Kleijn, W.: Estimation of the Instantaneous Pitch of Speech. IEICE Transactions on Audio, Speech, and Language Processing, Vol. 15, 2007, No. 3, pp. 813–822.

[16] Rabiner, L.: On the Use of Autocorrelation Analysis for Pitch Detection. IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 25, 1977, No. 1, pp. 24–33, 10.1109/tassp.1977.1162905.

[17] Ross, M.—Schafer, H.—Cohen, A.—Freudberg, R.—Manley, H.: Average Magnitude Difference Function Pitch Extractor. IEEE Transactions on Acustic, Speech, and Signal Processing, Vol. 22, 1974, No. 5, pp. 353–362.

[18] Shahnaz, C.—Zhu, W.—Ahmad, M.: Pitch Estimation Based on a Harmonic Sinusoidal Autocorrelation Model and a Time-Domain Matching Scheme. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, 2012, No. 1, pp. 310–323, doi: 10.1109/tasl.2011.2161579.

[19] Keys, R.: Cubic Convolution Interpolation for Digital Image Processing. IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. 29, 1981, No. 6, pp. 1153–1160.

[20] Park, S. K.—Schowengerdt, R. A.: Image Reconstruction by Parametric Cubic Convolution. Computing, Vision, Graphics, and Image Processing, Vol. 23, 1983, No. 1, pp. 258–272.

[21] Meijering, E.—Unser, M.: A Note on Cubic Convolution Interpolation. IEEE Transaction on Image Processing, Vol. 12, 2003, No. 4, pp. 447–479, doi: 10.1109/tip.2003.811493.

[22] Bozek, P.: Registration of Holographic Images Based on the Integral Transformation. Computing and Informatics, Vol. 31, 2012, No. 6, pp. 1369–1383.

[23] Pang, H.—Beak, S.: Improved Fundamental Frequency Estimation Using Parametric Cubic Convolution. IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, Vol. E83-A, 2000, No. 12, pp. 2747–2750.

[24] Milivojević, Z.—Mirković, M.—Rajković, P.: Estimating of the Fundamental Frequency by the Using of the Parametric Cubic Convolution Interpolation. Proceedings of International Scientific Conference (BG '04), Gabrovo, 2004, pp. 138–141.

[25] Mirković, M.—Milivojević, Z.—Rajković, P.: Performances of the System with the Implemented PCC Algorithm for the Fundamental Frequency Estimation, Proceedings of XII Telecommunications Forum TELFOR (SRB '04), Beograd, 2004.

[26] Milivojević, Z.—Mirković, M.—Milivojević, S.: An Estimate of Fundamental Frequency Using PCC Interpolation – Comparative Analysis. Information Technology and Control, Vol. 35, 2006, No. 2, pp. 131–136.

[27] Yarman, B.—Guz, U.—Gurkan, H.: On the Comparative Results of SYMPES: A New Method of Speech Modeling. International Journal of Electronics and Communications (AEU), Vol. 60, 2006, No. 2, pp. 421–427, doi: 10.1016/j.aeue.2005.08.003.

[28] Milivojević, Z.—Mirković, D.: Estimation of the Fundamental Frequency of the Speech Signal Modeled by the SYMPES Method. International Journal of Electronics and Communications (AEU), Vol. 63, 2009, pp. 200–208, doi: 10.1016/j.aeue.2007.12.006.

[29] Milivojević, Z.—Brodić, D.: Estimation of the Fundamental Frequency of the Speech Signal Compressed by G.723.1 Algorithm Applying PCC Interpolation. Journal of Electrical Engineering, Vol. 62, 2011, No. 4, pp. 181–189.

[30] Milivojević, Z.—Mirković, M.—Milivojević, S.: Fundamental Frequency Estimation of The Speech Signal Commpressed by MP3 Algorithm Using PCC Interpolation. Advances in Electrical and Computer Engineering AECE, Vol. 10, 2010, No. 1, pp. 18–22.

[31] Milivojević, Z.—Brodić, D.: Estimation of the Fundamental Frequency of the Real Speech Signal Compressed by MP3 AlgorithmSpeech Signal Compressed by G.723.1 Algorithm Applying PCC Interpolation. Archives of Acoustics, Vol. 38, 2013, No. 3, pp. 363–373.

Zoran Milivojević received his B.E., M.E., and Ph.D. degrees in electrical engineering from the Faculty of Electronic Engineering, University of Niš, in 1984, 1994 and 2002, respectively. He is Professor at the Faculty of Information Technology Alpha University of Belgrade and at the College of Applied Technical Sciences in Niš, Serbia. His primary research interests are digital signal processing: algorithms and applications. He is the author and coauthor of over 280 journal and conference papers (30 in journals indexed by Thomson SCI/SCIE JCR).

Nataša Savić received her B.Sc. degree in 1993 and M.A.Sc. degree in 2012 from the Faculty of Sciences and Mathematics, University of Niš. She is currently working as Professor at the College of Applied Technical Sciences in Niš, Serbia. Her research interests are numerical analysis, generating function, orthogonal polynomials, interpolation and approximation theory.

Darko Brodić received his B.E.E. and M.E.E. degrees from the Faculty of Electrical Engineering, University of Sarajevo in 1987 and 1990, respectively, and his Ph.D. from the Faculty of Electrical Engineering, University of Banja Luka in 2011. Currently, he is Assistant Professor at the University of Belgrade, Technical Faculty in Bor. His research interests include image processing, pattern recognition, artificial intelligence, algorithms. He is the author and coauthor of over 120 journal papers and conference papers (40 in journals indexed by Thomson SCI/SCIE JCR), and of few books and a textbook.