Estimation of the Fundamental Frequency of the Speech Signal Using PCC Interpolation With the BL Kernel

Zoran N. Milivojević, Zoran S. Veličković***

Keywords: *Fundamental frequency; Interpolation; Kernel.*

Abstract: This paper describes the construction of a parametric cubic interpolation kernel based on the blending method. The blending kernel is constructed by mixing two third-order interpolation kernels, the one-parameter Keys kernel (parameter α) and oMoms^3 kernel. The proportionality of the participation of the oMoms^3 kernel in blending kernel is represented by the blending factor (w). Created blending kernel has two parameters (α, w) which adjusting affects the accuracy of interpolation, that is, the reduction of interpolation error. The characteristics of the blending kernel are shown graphically in both, the time and spectral domains. After that, the algorithm for estimating the optimal parameters of the blending kernel is presented. The algorithm is described using a pseudo code. Subsequently, the precision of estimation of the fundamental frequency of the speech signal in the spectral domain, was tested experimentally, using an estimation algorithm. First, the speech test signal is processed in the time domain using some window (Hamming, Hann, Kaiser and Triangular). Subsequently, the speech test signal was transformed into the spectral domain using fast Fourier transformation (FFT). Then, using Peak-picking algorithm, a dominant spectral component (fundamental frequency) was determined. Most often the real fundamental frequency is between the two dominant spectral components. The real fundamental frequency is determined by applying a parametric convolution with a blending interpolation kernel. The estimation precision is represented by the mean square error (MSE) between the estimated and the real fundamental frequency. The optimal kernel parameters are determined by minimizing the mean square error, and the appropriate window is selected. The results are presented by tables and graphics.

* Zoran Milivojević is with the College of Applied Technical Sciences Niš, Serbia, (phone +381 63 860 9206, e-mail: zoran.milivojevic@vtsnis.edu.rs).

** Zoran Veličković is with the College of Applied Technical Sciences Niš, Serbia, (phone +381 69 468 7155, e-mail: zoran.velickovic@vtsnis.edu.rs).

1. INTRODUCTION

Interpolation is often used in the digital processing of multimedia signals (image, video, audio, speech,...) [1], [2]. The term *interpolation* derives from the Latin verb *interpolare.* In modern multimedia systems there is a need for: a) geometric transformations of the image, b) improving the quality and image reconstruction, c) compression, d) recognition of the object in the image and other [3], [4]. Digital processing of music signals enable: detecting instruments, recognition of chords and their transcription, recognition of tempo and rhythm, isolation and transcription of solo and bass lines, detection and evaluation of the quality of vibrates, inharmonicity of the instruments, etc.). In digital processing of speech signals there are algorithms for quality improved and comprehensibility of speech, verification of speakers, echo reduction, language recognition, understanding in semantic sense, emotional state of speakers recognition and others [5], [6].

In many of these algorithms, application of interpolation is required. Tasks by interpolation algorithm in modern multimedia systems are: a) large precision and b) execution speed. In polynomial interpolations, precision is increased by increasing the order of polynomial interpolation function, which leads to an increase in numerical complexity and execution time. Consequently, a convolution interpolation, which is based on the use of a low-order convolutional polynomial kernel ($n \leq 7$), is used extensively. The theoretically observed, ideal interpolation kernel is $sin(x)/x$, and is denoted by *sinc* symbol, where $-\infty < x < \infty$. Its amplitude characteristic is a rectangular, or box function [4]. Because of the endless limit, the *sinc* kernel is unrealizable, and it is necessary to limit the length of the kernel. The kernel length limitation leads to a significant deviation of the amplitude characteristics in relation to the box functions.

In the literature, a number of convolution polynomial kernels have been proposed, which reduce the problems of shortening the *sinc* kernels. Cubic convolution kernel represent a compromise between the precision of interpolation and numerical complexity, or the execution speed. The convolutional interpolation with parameter kernels third-order is marked with PCC (Parametric Cubic Convolution). A one-parameter (1P) interpolation convolution kernel is proposed in [7]. Later, this kernel is called Keys 1P kernel. Kernel optimization for the application of image interpolation is shown in [7]. It is determined that the optimal value of the kernel parameter is $\alpha = -0.5$. Greville 1P convolution interpolation kernel is described in [8]. The kernel parameter can be adapted to a particular problem in accordance with the defined criterion, most often by minimizing the interpolation error. In order to increase adaptability, kernels with more parameters are constructed (Keys 2P [9], Grevile 2P, Keys 3P [10]).

This paper describes the principle of creating a cubic convolution kernel, using the Blending (BL) method [11]. By using the BL method, interpolation kernel is formed from two or more interpolation kernels (parents kernels). An interpolation BL kernel was created from Keys 1P [7] and oMoms (optimal-Maximal-order-minimal-support) third-order ($oMoms^3$) [12], [13]. The new BL kernel contains two parameters: a) α from Keys kernel, and b) the blending factor *w*. An experiment was performed to determine the optimal parameters α_{opt} i w_{opt} for estimating the fundamental frequency of signal in frequency domain. In the experiment was used: a) sine test signal [14] and b) speech test signal [15]. The criterion for determining the optimal values of the kernel parameters is to minimize the mean square error (MSE). Testing was performed for some standard, time symmetric, window functions that modified the test signals in the time domain.

This paper is organized in the following way. Section II shows the BL kernel. Section III presents the algorithm for estimating optimal parameters of the BL kernel. Section IV describes an experiment that determines the optimal parameters of the BL kernel. Section V is a conclusion.

2. DESIGN OF BL KERNEL

In this section, we describe the BL interpolation kernel formed by blending the Keys 1P and oMoms[3] kernels [11]. The Keys 1P kernel is defined with [7]:

$$r_{Kejs\_1P} = \begin{cases} (\alpha+2)|x|^3 - (\alpha+3)|x|^2 + 1, & 0 \le x < 1 \\ \alpha|x|^3 - 5\alpha|x|^2 + 8\alpha|x| - 4\alpha, & 1 \le x < 2 \\ 0, & 2 \le x \end{cases}, \tag{1}$$

where α is the kernel parameter. oMoms[3] kernel is defined with [12]:

$$r_{oMom^3} = \begin{cases} \frac{1}{2}|x|^3 - |x|^2 + \frac{1}{14}|x| + 1, & 0 \le x < 1 \\ -\frac{1}{6}|x|^3 + |x|^2 - \frac{85}{42}|x| + \frac{29}{21}, & 1 \le x < 2 \\ 0, & 2 \le x \end{cases}. \tag{2}$$

By blending the Keys 1P and oMoms[3] kernels, the BL kernel formed:

$$r_{BL}(x) = (1-w)\, r_{Kejs\_1P} + w\, r_{oMom^3}(x) \tag{3}$$

where $w \in (0,1)$ is the blending factor. For example, in Fig. 1 are shown the time characteristics: a) ideal *sinc*, b) Keys 1P (α = -0.5), c) oMoms[3] and d) the newly formed BL kernel (3) (α = -0.3, w = 0.3). The spectral characteristics of these kernels are shown in Fig. 2. Fig. 2 shows the characteristic of the ideal *sinc* kernel in interval [-2, 2] ($sinc_{\mathrm{win}}$) . It is created by shortened of an infinite *sinc* kernel in time domain by used rectangular window. It is noted that in the passband and stopband the amplitude characteristic is wiggles while in the transient band amplitude characteristic is with the finite slope. The BL kernel depends on the kernel parameter α and the blending factor w (3). It is possible to select the values of the kernel parameters α and w so, that in the PCC interpolation, the smallest error of interpolation is obtained, according to a certain criterion. In this paper, the application of parametric cubic convolution interpolation in estimation of the fundamental frequency of the speech signal in the spectrum domain is analyzed.

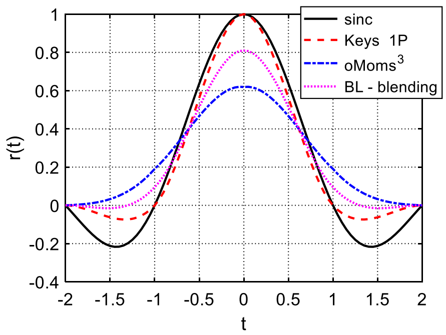

Fig. 1. Time characteristics: a) *sinc*, b) Keys 1P (α = -0.5), c) oMoms^3 and d) BL kernels (α = -0.3, w = 0.3).

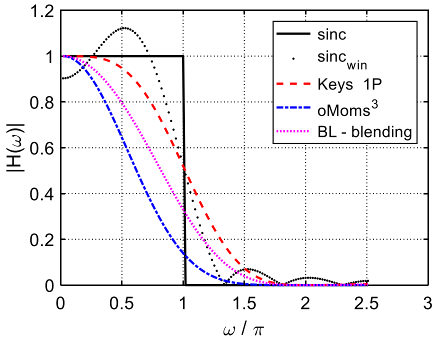

Fig. 2. Spectral characteristics: a) *sinc*, b) shortened $sinc_{\mathrm{win}}$, c) Keys 1P (α = -0.5), d) oMoms^3 and e) BL kernels (α = -0.3, w = 0.3).

3. ALGORITHM FOR ESTIMATES OF OPTIMAL PARAMETERS OF THE BL KERNELS

The optimal parameters of the BL interpolation kernel for estimation of the fundamental frequency of speech signal were determined experimentally. First, the speech Test signal is processed by a window in the time domain. After that, spectrum of the speech signal was calculated using Fast Fourier Transformation. Used Peak-picking algorithm the dominant component in the spectrum, witch representing the fundamental frequency, was determined [10], [14]. Then, in order to increase the precision of fundamental frequency estimation, a parametric cubic convolution with BL interpolation kernel was applied. Finally, by minimizing the fundamental frequency estimation error, the optimal parameters of the BL kernel are determined. The algorithm for estimation the optimal parameters of the BL kernel is realized in the following steps:

Input: r_1, - Keys 1P kernel,
r_2, - Moms^3 kernel,
K, - kernel length,
N, - test signal length,
$\mathbf{w}_P$, - window,
$NFFT$, - FFT length,
F_{0d}, F_{0g}, - fundamental frequency boundaries,
ΔF_0, - fundamental frequency iterative step,
α_d, α_g, - kernel parameters boundaries,
$\Delta\alpha$, - kernel parameters iterative step,
Δw, - blending factor iterative step.
Output: w_{opt}, α_{opt}.

FOR $\alpha = \alpha_d$ **TO** α_g **STEP** $\Delta\alpha$

FOR $F_0 = F_{0d}$ **TO** F_{0g} **STEP** ΔF_0

Step 1: test signal $\mathbf{x}_T$ is generated (11),

Step 2: test signal $\mathbf{x}_T$ was modified using the window $\mathbf{w}_P$:

$$\mathbf{x}_w = \mathbf{x}_T \cdot \mathbf{w}_P , \tag{4}$$

Step 3: The spectrum of $\mathbf{x}_w$ was calculated in NFFT points using Fast Fourier Transformation (FFT):

$$\mathbf{X} \xleftarrow{FFT} \mathbf{x}_w , \tag{5}$$

Step 4: The position of the dominant component X_{p_max} and its neighboring spectral components using the Peak-picking algorithm is calculated:

$$\mathbf{X}_p = \left\{ X_{p_{\max}-L}, \ldots, X_{p_{\max}}, \ldots, X_{p_{\max}+L} \right\},$$

where is $L = (K-1)/2$.

FOR $w = 0$ **TO** 1 **STEP** Δw

Step 5: The BL kernel r_{BL} was formed (3):

$$r_{BL} = (1-w)\cdot r_1 + w \cdot r_2 , \tag{6}$$

Step 6: Estimation of the fundamental frequency using parametric cubic interpolation (PCC):

$$X_{PCC}(f) = \sum_{i=p_{\max}-L}^{p_{\max}+L+1} X_P(i) \cdot r_{BL}(f-i), \qquad p_{\max} \le f < p_{\max}+1 , \tag{7}$$

where is $F_{0e} = arg\ \underset{f}{max}\left(X_{PCC}(f)\right)$.

Step 7: The fundamental frequency estimation error:

$$e\left(\left\lfloor \frac{w}{\Delta w} \right\rfloor + 1, \left\lfloor \frac{F_0 - F_{0d}}{\Delta F_0} \right\rfloor + 1\right) = F_0 - F_{0e} , \tag{8}$$

ENDFOR

ENDFOR

Step 8: Mean-square error estimation:

$$MSE(i,j) = \frac{1}{Q \cdot Y} \sum_{q=1:}^{Q} \sum_{y=1}^{Y} \left| e(q,y) \right|^2, \quad 1 \le q \le Q, \; 1 \le y \le Y \tag{9}$$

where Q and Y are the dimensions of the matrix **e**.

ENDFOR

Step 9: The optimal values of the BL kernel parameters:

$$(w_{opt}, \alpha_{opt}) = \arg\min_{w,\alpha}(MSE) \tag{10}$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment

An experiment, in which the fundamental frequencies F_0 of the audio signal **x** were estimated of: a) sine test signal and b) speech test signal was realized. The test signal is sampled with F_s. After that, test signal divided into frame lengths N by applied the window function w_p. Spectrum of each frame are calculated by using FFT length NFFT. In this way, the spectral components F_k are defined, where $0 \le k \le$ NFFT-1. In order to perform the test, F_0 was changed to the range $F_k \le F_0 \le F_{k+1}$. Because the real F_0 are different from the frequency at which FFT is calculated, the spectrum will have an leakage effect of the spectrum. In order to estimate the frequency position of the maximum of the spectrum and, therefore, the estimation of the fundamental frequency, PCC interpolation is applied. The F_0 estimate is realized by the algorithm described in Section III. By analyzing MSE (9), the optimal kernel parameters w_{opt}, α_{opt} (10) and the window in which the smallest MSE is generated, are determined.

The optimal parameters of BL kernel (3) are determined by used algorithm from Section 3. Time symmetric window: a) Hamming, b) Hann, c) Kaiser and d) Triangular are used. In Fig. 3.a are shown the time characteristics, and in Fig. 3.b are shown the spectral characteristics of used windows. Sampling frequency is F_s=8 kHz, T_s=0.125 ms, length of frame is N = 512, duration of frame is t_b = 32 ms, FFT length NFFT = 512, frequency resolution is ΔF = 15.625 Hz. The fundamental frequency of the test signal changed in the range F_0 = 125 - 140.625 Hz with step ΔF_0 = (140.625 - 125) / 100 = 0.15625 Hz (M=100), between spectral components k = 8 (f = 125Hz) i k = 9 (f = 140.625 Hz). The algorithm is performed over the Sinus test signal (11) with K = 10 harmonics, whose amplitudes a = {0.9800, 0.5541, 0.1120, 0.6982, 0.7511, 0.2498, 0.2491, 0.9642, 0.3478, 0.4739, 0.8610}, and phases θ = (0.2232, 0.7886, 1.7065, 1.4787, 1.5250, 0.4709, 1.5141, 0.4643, 0.9217, 0.0708)*π.

B. Test signals

In experiments were used Test signals:

a) sine test signal definition in [14]:

$$s(t) = \sum_{i=1}^{K}\sum_{g=0}^{M} a_i \sin\left(2\pi i\left(f_o + g\frac{f_s}{NM}\right)t + \theta_i\right), \tag{11}$$

where f_0 is fundamental frequency, a_i and θ_i are amplitude and phase 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 (Fig. 4).

b) speech test signal (vowel 'a' in Serbian) defined in [15] (Fig. 5).

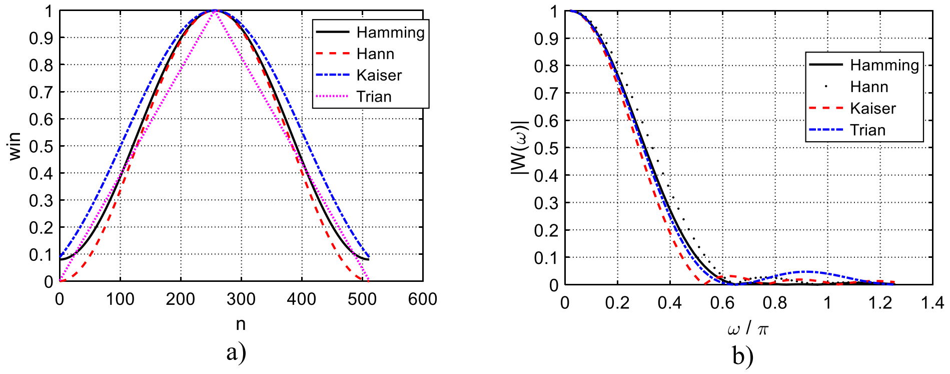

Fig. 3. Window functions: a) time domain and b) spectrum.

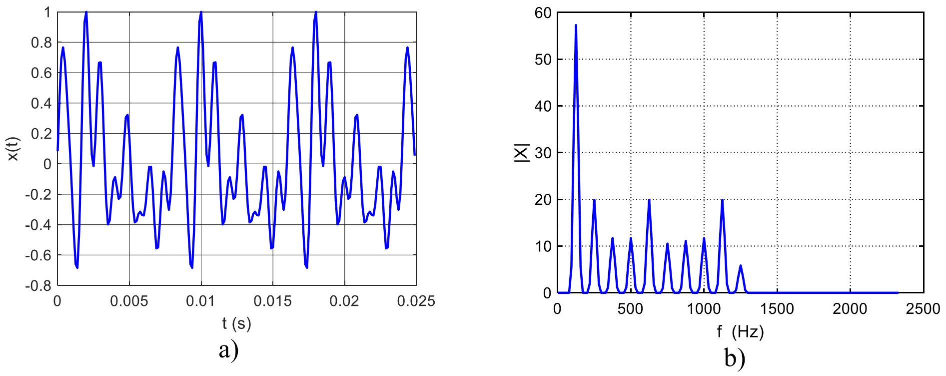

Fig. 4. Sine test signal: a) time domain and b) spectrum.

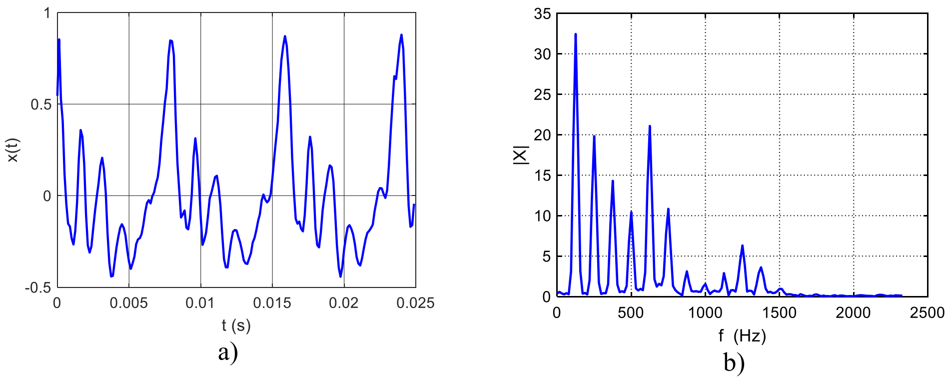

Fig. 5. Speech test signal (vowel 'a' in Serbian): a) time domain and b) spectrum.

C. Experimental results

Sine test signal: The MSE results for processing obtained using the Keys 1P kernel for the tested windows were obtained from (9) and (10) for w = 0, are shown in Fig. 6. Minimum MSE_{min} and corresponding optimal values of α_{opt} are shown in Table I. The results for oMoms^3 kernel (w = 1) are shown in Table I. The results for the application of the BL kernel (3) are shown in Fig. 7 (dependence of MSE on blending factor w).

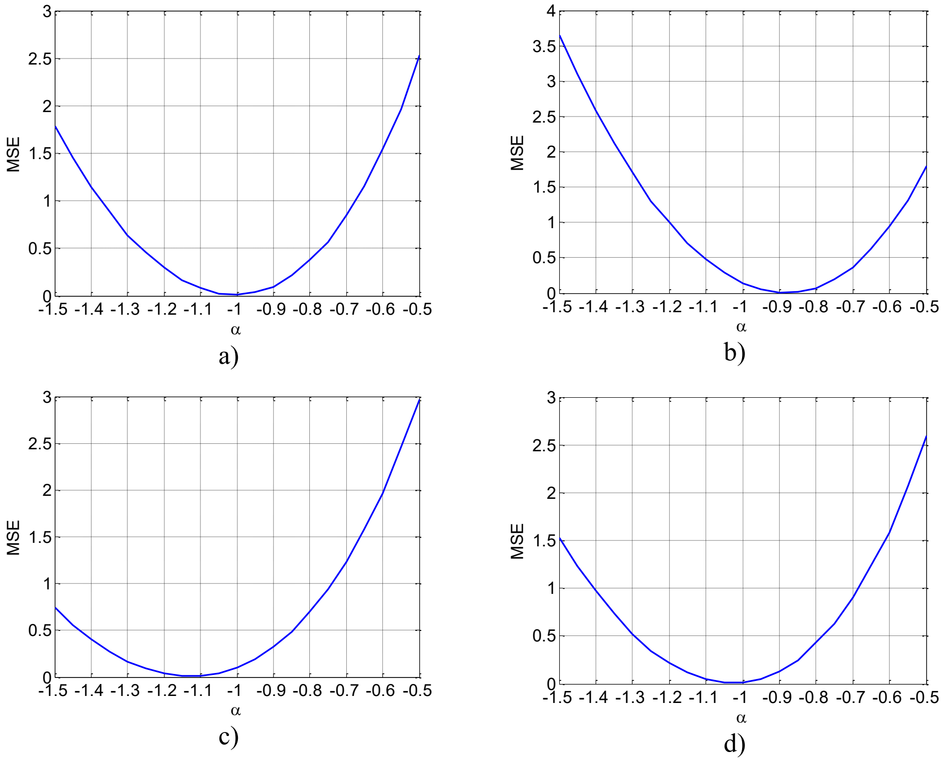

Fig. 6. MSE results for interpolation with Keys 1P kernel (w = 0) in the case of application: a) Hamming, b) Hann, c) Kaiser and d) Triangular windows for the sine test signal.

Table I
MSE results of the application of Keys 1P, oMoms^3 and BL kernels for sine test signal.

Window	**Kejs 1P**		**oMoms^3**	**BL**		
	α_{opt}	MSE_{min}	MSE_{min}	α_{opt}	w_{opt}	MSE_{min}
Hamming	-1.000	0.0094	0.1073	-1.000	0.1400	0.0092
Han	-0.900	0.0073	0.0520	-0.9000	0.1800	0.0039
Kaiser	-1.100	0.0126	0.1315	-1.1500	0.1600	0.0114
Triangular	-1.050	0.0094	0.1114	-1.0500	0.3200	0.0056

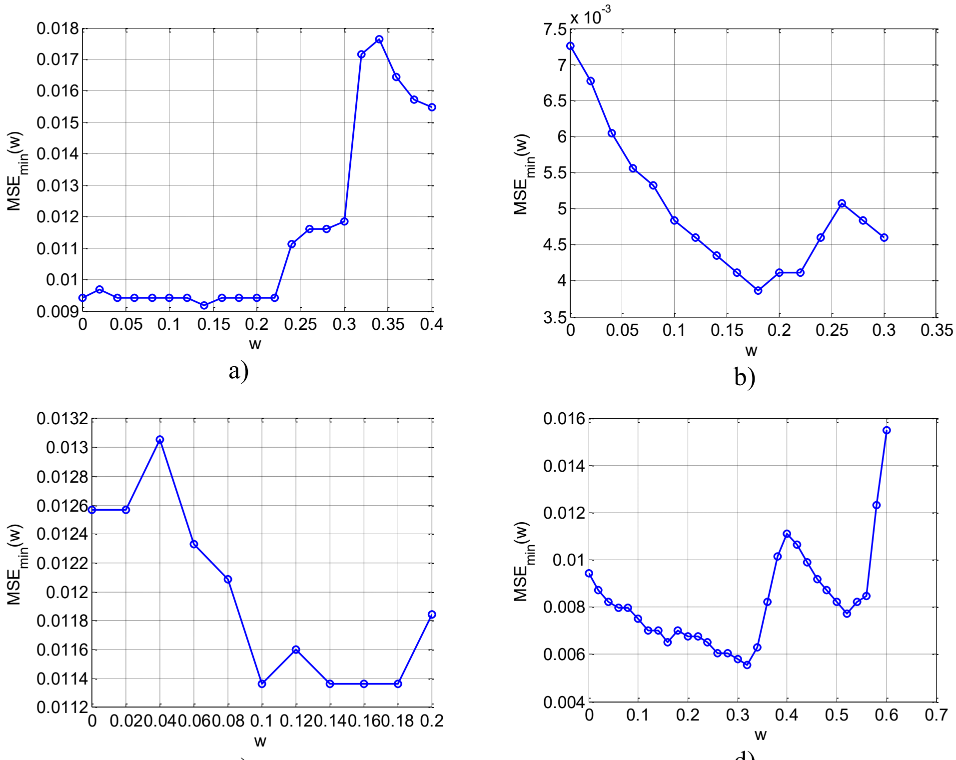

Fig. 7. Interpolation with the BL kernel: the dependence of the MSE on the blending factor *w* in the application: a) Hamming, b) Hann, c) Kaiser and d) Triangular windows for a sine test signal.

Speech test signal: The MSE results for processing obtained using the Keys 1P kernel for the tested window were obtained from (9) and (10) for w = 0, are shown in Fig. 8. Minimum MSE_{min} and corresponding optimal values of α_{opt} are shown in Table II. The results for oMoms^3 kernel (w = 1) are shown in Table II. The results for the application of the BL kernel (3) are shown in Fig. 9 (dependence of MSE on blending factor w).

Table II
MSE results of the application of Keys 1P, oMoms^3 and BL kernels for speech test signal.

Window	**Keys 1P**		**oMoms^3**	**BL**		
	α_{opt}	MSE_{min}	MSE_{min}	α_{opt}	w_{opt}	MSE_{min}
Hamming	-1.000	0.0375	0.1257	-1.000	0.1200	0.0338
Han	-0.900	0.0423	0.0834	-0.9000	0.3800	0.0331
Kaiser	-1.100	0.0324	0.1431	-1.1000	0.0200	0.0321
Triangular	-1.000	0.0331	0.1339	-1.0500	0.3600	0.0300

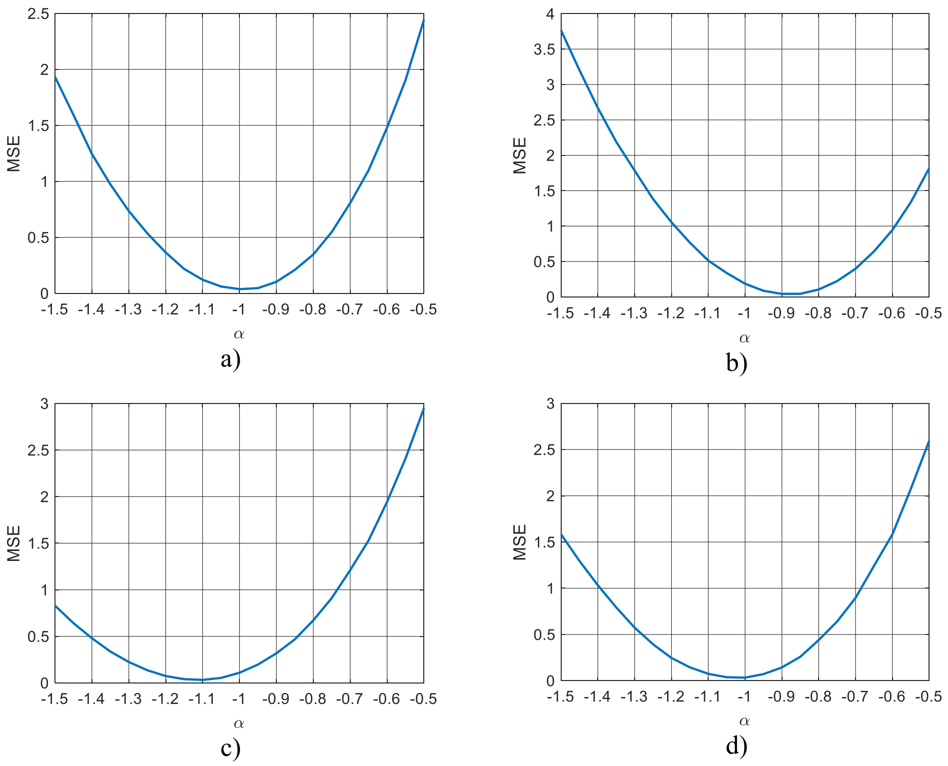

Fig. 8. MSE results for interpolation with Keys 1P kernel ($w = 0$) in the case of application: a) Hamming, b) Hann, c) Kaiser and d) Triangular windows for the speech test signal.

D. Comparative analysis

According to the results presented in Tbl. I and Tbl. II, and Fig. 6-9, it is obvious that, with the sine test signal, using: a) Keys 1P kernel is the smallest error with Hann window (α_{opt} = -0.9, MSE_{min} = 0.0073), b) oMoms^3 kernel is the smallest error with Hann window (MSE_{min} = 0.052), c) Keys 1P kernel relative to the oMoms^3 kernel precision of estimation is increased by 0.052 / 0.0073 = 7.123 times, and d) BL kernel the smallest error estimation is with Hann window (α_{opt} = -0.9, w_{opt} = 0.18, MSE_{min} = 0.0039). Compared to the Keys 1P kernel, the precision is increased 0.0073 / 0.0039 = 1.87 times. Compared to the oMoms^3 kernel, the precision is increased 0.052 / 0.0039 = 13.33 times.

At the speech test signal, using: a) Keys 1P kernel is the smallest error with Triangular window (α_{opt} = -1.00, MSE_{min} = 0.0331), b) oMoms^3 kernel is the smallest error with Hann window (MSE_{min} = 0.0834), c) Keys 1P kernel relative to the oMoms^3 kernel precision of estimation is increased by 0.0834 / 0.0331 = 2.52 times, and d) BL kernel the smallest error estimation is with Triangular window (α_{opt} = -1.05, w_{opt} = 0.36, MSE_{min} = 0.03). Compared to the Keys 1P kernel (Triangular window), the precision was increased 0.0331 / 0.03 =

1.103 times. Compared to the oMoms^3 kernel, the precision is increased 0.0834 / 0.03 = 2.78 times.

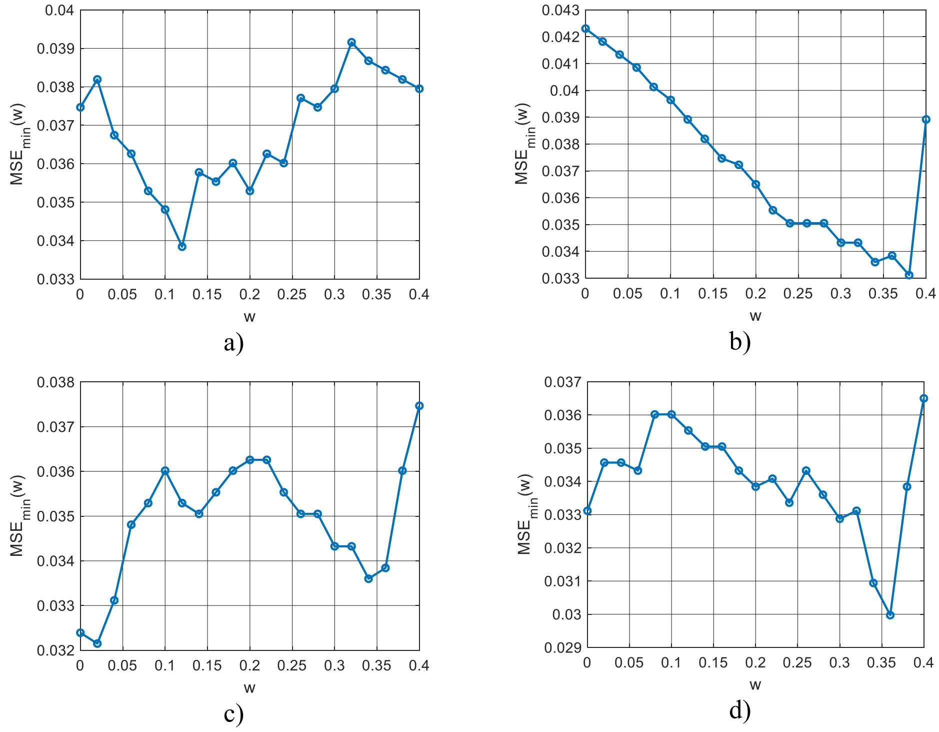

Fig. 9. Interpolation with the BL kernel: the dependence of the MSE on the blending factor *w* in the application: a) Hamming, b) Hann, c) Kaiser and d) Triangular windows for a speech test signal.

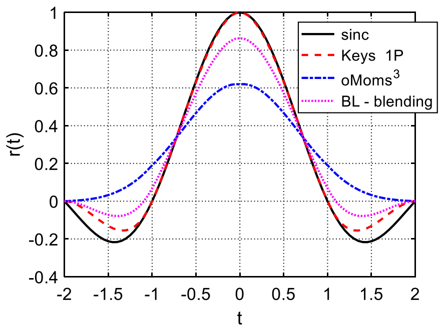

Fig. 10. Time characteristics: a) sinc, b) Keys 1P (α_{opt} = -1.05), c) oMoms^3 and d) BL kernels (α_{opt} = -1.05, w_{opt}=0.36).

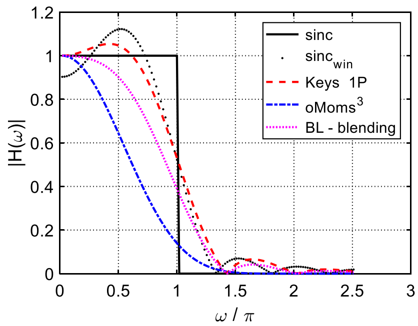

Fig. 11. Spectral characteristics: a) sinc, b) shortened $sinc_{win}$, c) Keys 1P (α_{opt} = -1.05), d) oMoms^3 and e) BL kernels (α_{opt} = -1.05, w_{opt}=0.36).

The precision of the fundamental frequency estimation in the sine test signal relative to the speech test signal when applied: a) Keys 1P kernel (Hann window, Triangular window) is greater than 0.0331 / 0.0073 = 4.563 times, b) The oMoms^3 kernel (Hann window, Hann window) is greater than 0.0834 / 0.052 = 1.6 times, and c) BL kernel (Hann window, Triangular window) is greater than 0.03 / 0.0039 = 7.69 times.

From the conducted analysis, it is concluded that the optimal choice for the BL interpolation kernel is Triangular window. Optimal BL parameters are α_{opt} = -1.05 and w_{opt} = 0.36. In Fig. 10 shows the time characteristics of the *sinc* kernel and BL kernels with optimal parameter. In Fig. 11 shows the spectral characteristics of the BL kernel with optimal parameters. It can be seen that the spectral characteristic of the BL kernel is with less wiggles compared to Keys 1P with the optimal parameter (α_{opt} = -1.05), as well as with a smaller deviation from the ideal box characteristic compared to oMom^3 kernel.

As a global conclusion, the efficiency of estimating the fundamental frequency of the speech signal, used Triangular window and parametric cubic interpolation with implemented the BL interpolation kernel are indicated.

5. CONCLUSION

The paper presents the BL interpolation kernel witch is created by blending Keys 1P and oMoms^3 kernels. The efficiency of the BL kernel in estimation the fundamental frequency of the audio signala (sine and speech test signals) was tested experimentally. Testing was performed in the case application the parametric cubic interpolation. The test signal is modified in a time domain with Hamming, Hann, Kaiser and Triangular windows. A detailed analysis of the experimental results showed that the optimal choice is the Triangular window. In this case, the optimal parameters of the BL kernel are α_{opt} = -1.05 and w_{opt} = 0.36. BL kernel with optimal parameters generated less MSE compared to Keys 1P 1.103 times and compared to oMoms^3 2.78 times. In this way, it has been proven that the BL kernel generates a less error in relation to the kernels from which it was created by blending method. Therefore, it is recommended that the BL kernel will be implemented in the real-time systems.

REFERENCES

[1] P. Thvenaz, et al., *Image Interpolation and Resampling.* Handbook of Medical Imaging. Orlando, FL: Academic, 2000.

[2] D. F.Watson, *Contouring: A Guide to the Analysis and Display of Spatial Data*". New York: Pergamon, 1992.

[3] M. Unser, A. Aldroubi, and M. Eden, "Enlargement or reduction of digital images with minimum loss of information, " *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 247–258, 1995.

[4] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing, " *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, Nov. 1999.

[5] B. Joen, S. Kang, S.J. Baek, and K.M. Sung, "Filtering of a Dissonant Frequency Based on Improved Fundamental Frequency Estimation for Speech Enhancement," *IEICE Trans. Fundamentals,* vol. E86-A, no. 8, pp. 2063-2064, August 2003.

[6] Z. Milivojević, M. Milivojević, D. Brodić, and D. Milivojević, "The Acute Hypoxia Indication by the Dissonant Intervals of the Speech Signals," *International Journal of the Physical Sciences*, vol. 7, no. 3, pp. 394-399, January, 2012.
[7] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, no. 6, pp. 1153–1160, Dec. 1981.
[8] E. Meijering, and M. Unser, "A Note on Cubic Convolution Interpolation", *IEEE Transactions on Image Processing,*" vol. 12, no. 4, pp. 447-479, April 2003.
[9] R. Hanssen, and R. Bamler, "Evaluation of Interpolation Kernels for SAR Interferometry", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no.1, pp. 318-321, Jan. 1999.
[10] Z. Milivojević, N. Savić, and D. Brodić, "Three-Parametric Cubic Interpolation For Estimating The Fundamental Frequency Of The Speech Signal," *Computing and Informatics*, vol. 36, 2017, pp. 1001–1020, V 2017-May-22.
[11] L. Liang, "Image Interpolation by Blending Kernels, " *IEEE Signal Processing Letters*, vol. 15, , pp. 805-808, 2008.
[12] P. Thévenaz, T. Blu, and M. Unser, "Interpolation Revisited," *IEEE Transactions On Medical Imaging*", vol. 19, no. 7, pp. 739-758, July 2000.
[13] T. Blu, P. Thévenaz, and M. Unser, "MOMS: Maximal-Order Interpolation of Minimal Support," *IEEE Transactions On Image Processing*, vol. 10, no. 7, pp. 1069-1080, July 2001.
[14] H. Pang, S. Baek, and K. Sung, "Improved Fundamental Frequency Estimation Using Parametric Cubic Convolution," *IEICE Trans. Fund.*, E83-A, 12, pp. 2747-2750, 2000.
[15] Z. Milivojević, D. Brodić, "Estimation of the Fundamental Frequency of the Speech Signal Compressed by MP3 Algorithm, " *ARCHIVES OF ACOUSTICS*, Vol. 38, No. 3, pp. 363–373 2013.