FACTA UNIVERSITATIS (NIŠ)
SER.: ELEC. ENERG. vol. 22, no. 3, December 2009, 391-404

Enhancement of the Perceptive Quality of the Noisy Speech Signal by Using of DFF-FBC Algorithm

Zoran N. Milivojević and Dragiša Z. Balanesković

Abstract: This paper presents an algorithm for enhancement of the noisy speech signal quality. This algorithm is based on the dissonant frequency filtering (DFF), F#, B and C# in relation to the frequency of the primary tone C (DFF-FBC algorithm). By means of the subjective Mean Opinion Score (MOS) test, the effect of the enhancement of the speech signal quality was analyzed. The analysis of the MOS test results, presented in the second part of this paper, points out to the enhancement of the noisy speech signal quality in the presence of superimposed noises. Especially good results have been found with Husky Voice signal.

Keywords: Fundamental frequency estimation, dissonant frequency, speech quality enhancement.

1 Introduction

SUPERIMPOSING of an acoustic background noise (Babble noise, Car noise, Factory noise, Computer fan noise, acoustic echo, ...) leads to deterioration of the speech signal quality which is, among the rest, manifesting as decreasing of intelligibility [1–3]. The speech signal with superimposed acoustic disturbancies is transmited by means of communication lines, so that on the receptive side degrading may appear so exsessively that the reproduced speech becomes unintelligible or rather unpleasant. In addition to the acoustic disturbancies, on the occasion of transferring of speech by communication lines, speech degradation is caused by the appearance of an echo. A number of algorithms have been developed based on compensation, i.e. echo decreasing [4–6].

Manuscript received on April 1, 2009

Z. Milivojević is with Technical College, Aleksandra Medvedeva 20, 18000 Niš, Serbia (e-mail: zoran.milivojevic@jotel.co.rs). D. Balanesković is with Faculty of Arts, University of Niš, 18000 Niš, Serbia (e-mail: balanesko@bankerinter.net).

In speech signal processing for the purpose of signal compressing, as well as in systems for speech identification, the superimposed acoustic noise considerably degrades performances of the processing algorithm. In order to improve performaces of the processing algorithm it is necessary to do preprocessing of the speech signal to decrease the noise components. Enhancement of the speech signal quality is a actual problem and a great number of algorithms have been developed. Algorithms can be classified into three groups:

a) speech enhancement algorithms based on the short-time spectral estimation such as the spectrum subtraction ([7, 8]) and Wiener ([9]) filtering techniques,
b) comb filtering and adaptive noise canceling techniques which exploit the quasi-periodic nature of the speech signal [10], and
c) algorithms that are based on the statistical model of the speech signal and use hidden Markov model (HMM) or expectation and maximization (EM) for speech enhancement [11, 12].

The completely new speech enhancement algorithm for filtering of dissonant frequencies is presented in [13]. This algorithm is based on speech signal processing in the spectral domain: a) determining of fundamental frequency and b) filtering of dissonant frequency in relation to the fundamental frequency in all octaves. Filtered are dissonant frequencies which in relation to the fundamental frequency stand as the tone F# in relation to the tone C. F# is known in music as the Devil's interval [14]. The results, based on the application of Mean Opinion Score (MOS) test, show that the proposed method provides a significant gain in audible improvement especially for speech contaminated by Gaussian noise and a Husky Voice.

In [15] the authors broaden the activity range of the algorithm described in [13] by increasing the number of the filtered dissonant frequencies (B and F# in relation to C). The subjective test results indicated that the proposed method delivered improvements in terms of both speech intelligibility and perceived quality when compared with the unprocessed case. Therefore when the filter is employed as a prefilter for speech enhancement, the output speech quality is more enhanced perceptually.

The efficiency of the algorithms described in [13, 15] depends of how precisely the speech signal fundamental frequency is estimated. A number of algorithms were developed for determination of the fundamental frequency where the analysis is performed in the time and frequency domain [16, 17]. 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 $[-F_S/(2N)$ Hz, $F_S/(2N)$ 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...). Calculation of the interpolation function by using of Parametric Cubic Convolution (PCC) was represented in [18]. The detailed analysis of the fundamental frequency estimation, as well as the advantage of the PCC interpolation, which can be seen in the speed of determining of the interpolation function parameters, is described in the paper [19]. 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 [20].

This paper presents an algorithm for noisy speech signal quality enhancement based on the filtering of three dissonant frequencies and their harmonics in seven octaves of audio range. The proposed algorithm is based on the algorithm described in [13] and [15] and represents broadening of the activity range. In contrast to the algorithm from [13], where one dissonant frequency which in relation to the fundamental frequency F_0 stands as the tone F# in relation to C was filtered and the algorithm from [15], where two dissonant frequencies which in relation to the fundamental frequency F_0 stand as tones F# and B in relation to C were filtered, we have specified one more dissonant frequency which in relation to the fundamental frequency F_0 stands as the tone C# in relation to C. The filtering algorithm proposed in this paper includes dissonant frequencies which in relation to the fundamental frequency stand as the tones F#, B and C# in relation to C. The efficiency of the proposed algorithm was tested by processing of speech signals which are superimposed by: a) White Gaussian Noise (WGN), b) Computer Fan Noise, c) Babble Noise, d) Car Noise, and e) Husky Voice (clean speech for Male and Female).

The paper is organized as follows: Section 2 presents a musicological definition of dissonant frequencies. Section 3 presents algorithms for dissonant frequency filtering. The MOS test results performed on the filtered speech signal with superimposed noises, are presented in Section 4. The analysis of the MOS test results is performed in Section 5. Concluding remarks are given in Section 6.

2 Musicologic Definition of Dissonant Frequencies

The theory of music defines the fundamental features of the sound: a) duration, b) intensity and c) *color*. The expression *color* applies to the sound in a metaphorical way, which points out to the complexity of this feature of the sound. The source of a sound generates a sound with the fundamental frequency (the primary tone) as well as the overtones (aliquoties in relation to the primary tone). Different number of the present aliquoties (**lat**. *aliquoties* - several times) and their various relative intensity within the total sounding, determine the color of a sound.

The frequency of the musically defined tones in relation to the primary tone in an interval of one octave is determined by:

$$F_k = F_0 \cdot 2^{\frac{k}{12}}, \quad k = 0, 1, ..., 12, \tag{1}$$

where F_0 is the frequency of the primary tone and F_k the frequency of the k-th halftone. In relation to the primary tone, the halftones form intervals. An interval is defined by the relation of the frequency of a halftone and the frequency of the primary tone. Fractions F_k/F_0, for k=0,1,...,12, which present individual intervals (1/1, 135/128, 9/8, 6/5, 5/4, 4/3, 45/32, 3/2, 8/5, 27/16, 9/5, 15/8, 2/1) present approximation of the real value (Eq.1). Interval classification according to their sounding is realized on the base of the fraction it describes it. If the fraction is simpler, the interval, as an ackord of tones, more stable, i.e. more consonant. If the fraction is more complex, stability of the interval is smaller, so that dissonance is greater.

Consonace and dissonance are not sharply delimited but make together one differentiated scale, from total stability on one end of the scale to total instability on the other end. Within the scale we distinguish: a) perfect (complete) consonaces (prima (1/1), octave (2/1), quinta (3/2) and quarta (4/3)), b) unperfect (incomplete) consonaces (big tierce (5/4), big sixth (5/3), small tierce (6/5) and small sixth (8/5)), c) unperfect (incomplete) disonaces (small seventh (9/5) and big second (9/8) and d) perfect (complete) disonances (small second (135/138), threetones or excessive quarta (45/32) and big seventh (15/8). From the view-point of experience, i.e. perception of the sound, the musical interval is defined as being consonant if the sound is pelasant or restful. The musical interval is dissonant to a great extent if the sound is unpleasant or rough.

In relation to the primary tone, half tones frequencies, which together with the primary tome make consonances in all octaves within the audible range, are defined:

$$F_d = F_0 \cdot 2^{n+\frac{k}{12}}, \quad n = 0, 1, ..., 7; \quad k = \{1, 6, 11\}, \tag{2}$$

where F_0 is the frequency of the primary tone, n is the number of the octave and

k is the number of half tones in individual octaves. Considering the tone C as the referent one, i.e as the primary tone, then its dissonant half tones are B, F# and C# as well as their harmonics in all the octaves.

3 Filtering of Dissonant Frequencies

The speech is created by excitation of the vocal tract of a man [21]. According to the analogy of the speech signal with the musicological definition of the sound, there can be established the correspondence of the primary tone and its appropriate half tones an aliqotes, with the fundamental frequency F_0 and accompanying frequencies of the speech signal. Hereby it is possible to define dissonant frequencies in relation to F_0.

In [13] we find a description of an algorithm for speech signal enhancement by filtering of dissonant frequencies. This algorithm consists of the following steps: a) division of the speech signal into sequences whose length is N and calculation of FFT for every sequence, b) determining of the fundamental frequency F_0, c) determining of the dissonant frequencyF_d in relation to the fundamental frequency F_0 (according to the relation C to F#), d filtering of the dissonant frequencies and e) generating of the speech signal sequence by using of IFFT. In [15] an algorithm for the speech signal enhancement is described which was obtained by broadening of the activity range of the algorithm from [13]. Broadening of the algorithm refers to eliminating od dissonant tones F# and B in relation to C.

Further in this paper it is presented an algorithm (Figure 1) for dissonant frequency filtering (DFF), F#, B and C# in relation to the frequency of the primary tone C. DFF-FBC algorithm is based on the algorithms from [13, 15]. DFF-FBC algorithm consists of the following steps:

Step 1: Speech signal division into sequences of length N and determining of FFT for each sequence,

Step 2: Estimation of the fundamental frequency F_0 by using of PCC,

Step 3: Determining of the dissonant frequency F_{d1}, F_{d2} and F_{d3} in relation to the fundamental frequency F_0 (according to relation of the tone C to F# and B to C#) as:

$$F_{d1} = F_0 \cdot 2^{n+\frac{6}{12}}, \quad n = 0, 1, ..., 7, \tag{3}$$

$$F_{d2} = F_0 \cdot 2^{n+\frac{11}{12}}, \quad n = 0, 1, ..., 7, \tag{4}$$

$$F_{d2} = F_0 \cdot 2^{n+\frac{1}{12}}, \quad n = 0, 1, ..., 7, \tag{5}$$

Step 4: Filtering of the dissonant frequencies from the range (which corresponds to the half tone):

$$F_0 \cdot 2^{n+\frac{11}{24}} < F_{d1} < F_0 \cdot 2^{n+\frac{13}{24}}, \quad n = 1, 2, \ldots, 7, \tag{6}$$

$$F_0 \cdot 2^{n+\frac{21}{24}} < F_{d2} < F_0 \cdot 2^{n+\frac{23}{24}}, \quad n = 2, 3, \ldots, 7, \tag{7}$$

$$F_0 \cdot 2^{n+\frac{1}{24}} < F_{d3} < F_0 \cdot 2^{n+\frac{3}{24}}, \quad n = 2, 3, \ldots, 7, \tag{8}$$

Step 5: Generating of the time sequence of the speech signal by using of IFFT.

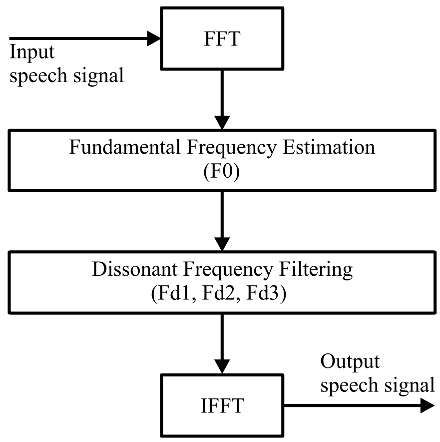

Fig. 1. The algorithm for speech signal processing by filtering of the dissonant frequencies.

4 Experimental Results

For the testing of the effect of DFF-FBC algorithm to the perceptive characteristics of the speech signal and for the comparative analysis with the results of speech signal processing by algorithms from [13, 15] the file bases of speech signals were formed according to the bases from [15] and the subjective MOS test was performed.

4.1 Speech database

The database for performance evaluation consists of 14 speech files collected from 7 speakers (5 males and 2 fe-males), each one delivering 2 Serbian sentences. Also we obtained 6 Husky Voice files from 3 speakers (2 males and 1 female) who have worse quality than normal speakers. All utterances were sampled at 8 kHz with 8 bit resolution. Speech input is windowed by 256-point Hanning window and padded with 256 point zeroes. Hanning window is halfoverlapped 256 point window corresponds to 32 ms in which speech is considered to be stationary. The used types of noises are adequate to noises from [15] (White Gaussian noise, Babble noise, Car noise, Computer Fan noise). In our experiments we used a wide range SNR=0, 5, 10, 15, 20 dB. Special attention in this experiment was paid to babble noise (i.e. the summed waveform of several simultaneous talkers). The speech signal for communication is often formed in an environment where a number of speakers are active. Babble is often used as a masker in studies of everyday speech perception in noise. However, the masking effect of Babble is heavily dependent on the number (N) of simultaneous talkers in the mixture [22]. Single-talker maskers (N=1) and speech-shaped noise ($N = \infty$) are the extremes of the babble continuum. In [1] they suggest that babble is a more effective masker than speech-shaped noise. In this paper an analysis was performed for at SNR=0,5,10,15,20 dB. Car noise recorded inside a car moving approximately at a speed of 90 km/h. Because of an unequal energy distribution in the babble noise spectrum, car noise and computer noise are classified into a group of colored noise [3].

4.2 MOS test

The quality of the reproduced speech was tested by the subjective Mean Opinion Score (MOS) test. Twenty listeners participated in this test. The listeners were sitting comfortably in a quiet room and listening to sentences played randomly. Then the listeners on the base of their individual experience of the quality made an evaluation in range from 1 (very bad) to 5 (excellent). The results of MOS test are presented in Table 1 (White Gaussian noise), Table 2 (Computer Fan noise), Table 3 (Car noise), Table 4 (Babble noise) i Table 5 (Husky Voice). In the tables the results of MOS tests are presented for: a) the unprocessed speech signal ($\mathrm{MOS_{UPSS}}$), b) the speech signal with eliminated F# intervals ($\mathrm{MOS_{F\#}}$) (the algorithm from [13]), c) the speech signal with eliminated F# and B intervals ($\mathrm{MOS_{F\#B}}$) (the algorithm from [15]) and d) the speech signal with eliminated F#, B and C# intervals ($\mathrm{MOS_{F\#BC\#}}$) (DFF-FBC algorithm proposed in this paper).

Table 1. MOS test results for speech signal with superimposed White Gaussian noise.

SNR [dB]	$\mathrm{MOS_{UP}}$	$\mathrm{MOS_{F\#}}$	$\mathrm{MOS_{F\#B}}$	$\mathrm{MOS_{F\#BC\#}}$
0	1.85	1.98	2.02	2.03
5	2.64	2.81	2.83	2.84
10	2.88	3.06	3.11	3.13
15	3.05	3.24	3.31	3.33
20	3.89	4.03	4.07	4.08
Clean Speech	4.8	4.85	4.89	4.9

Table 2. MOS test results for speech signal with superimposed Computer Fan Noise.

SNR [dB]	Unprocessed	F#	F#,B	F#,B,C#
0	1.93	2.04	2.06	2.09
5	2.34	2.47	2.51	2.53
10	2.78	2.94	2.91	3.01
15	2.96	3.12	3.16	3.19
20	3.16	3.33	3.37	3.41

Table 3. MOS test results for speech signal with superimposed Car Noise.

SNR [dB]	Unprocessed	F#	F#,B	F#,B,C#
0	1.87	2.05	2.09	2.10
5	2.21	2.41	2.44	2.48
10	2.89	3.13	3.17	3.21
15	2.93	3.10	3.14	3.17
20	3.05	3.20	3.22	3.25

Table 4. MOS test results for speech signal with superimposed Babble Noise.

SNR [dB]	N	Unprocessed	F#	F#,B	F#,B,C#
0	1	2.67	2.84	2.86	2.87
	4	2.65	2.82	2.84	2.85
	8	2.64	2.80	2.82	2.84
5	1	2.74	2.87	2.89	2.90
	4	2.71	2.88	2.90	2.91
	8	2.69	2.86	2.88	2.89
10	1	2.83	3.00	3.03	3.04
	4	2.82	2.99	3.02	3.03
	8	2.79	2.96	2.98	2.99
15	1	2.86	3.03	3.06	3.07
	4	2.82	2.99	3.01	3.03
	8	2.81	3.98	3.00	3.01
20	1	2.98	3.15	3.16	3.18
	4	2.96	3.13	3.14	3.16
	8	2.95	3.11	3.13	3.15

Table 5. MOS test results for Husky Noise.

	Unprocessed	F#	F#,B	F#,B,C#
Male	2.82	3.18	3.32	3.34
Female	2.86	3.16	3.25	3.26

For the purpose of a comparative analysis of the algorithms' effect on the speech signal quality, calculations were made for the percentage increments of MOS test values processed ($\Delta MOS_{\text{F\#}}$, $\Delta MOS_{\text{F\#B}}$, $\Delta MOS_{\text{F\#BC\#}}$) in relation to unprocessed (MOS_{UPSS}) speech signals according to:

$$\Delta MOS_{\mathrm{F\#}} = \frac{100 \cdot MOS_{\mathrm{F\#}}}{MOS_{\mathrm{UPSS}}} - 100, \tag{9}$$

$$\Delta MOS_{\mathrm{F\#B}} = \frac{100 \cdot MOS_{\mathrm{F\#B}}}{MOS_{\mathrm{UPSS}}} - 100, \tag{10}$$

$$\Delta MOS_{\mathrm{F\#BC\#}} = \frac{100 \cdot MOS_{\mathrm{F\#BC\#}}}{MOS_{\mathrm{UPSS}}} - 100, \tag{11}$$

where $\Delta\mathrm{MOS}_{\mathrm{F\#}}$, $\Delta\mathrm{MOS}_{\mathrm{F\#B}}$, and $\Delta\mathrm{MOS}_{\mathrm{F\#BC\#}}$ are percentage increments of MOS values for algorithms from [13, 15] and DFF-FBC algorithm, respectively.

Percentage increments of MOS values are graphically presented in Figure 2: (a) White Gaussian Noise, (b) Computer Fan Noise, (c) Car Noise, (d) Babble Noise *N*=1, (e) Babble Noise *N*=4 i (f) Babble Noise *N*=8. Percentage increments of MOS values in processing of Husky Voice are graphically presented in Figure 3.

Spectrogrames of the speech signal (the sentence '*High Technical School*' pronounced in Serbian) presented in Figure 4: a) Clean speech signal, b) Clean speech signal with superimposed WGN SNR=5 dB, c) filtered F#, d) F#B and e) F#BC# dissonances; and Figure 5.: a) with 5 dB Babble Noise *N*=8, b) filtered F#, c) F#B d) and F#BC# dissonances.

5 Analysis of the Results

To analyze the effects of DFF-FBC algorithms on the subjective quality of the speech signal, on the base of the data (percentage increments of MOS test results in relation to the result of an unprocessed signal (Eq. 9-11)) graphically presented in Figures 2 and 3, the mean values of the percentage increment of MOS test values are determined for all kinds of noises:

$$\overline{\Delta\mathrm{MOS}_d} = \frac{\sum_{SNR} \Delta\mathrm{MOS}(\mathrm{SNR})}{len(\mathrm{SNR})}, \tag{12}$$

where *d*=F#; F#B, F#BC#, SNR=0,5,10,15,20 dB and *len*(SNR) presents the number of elements in the sequence SNR, and presented in Table 6.

On the base of MOS test results graphically presented in Figures 2 and 3 and in Tables 1-6 it may be concluded that:

a) an algorithm for eliminating of dissonance F# in the speech signal with superimposed noise generates a speech signal whose MOS test result is 5.63-7.72% higher than that of an unprocessed signal. This results is in accordance with the result of the algorithm from [13];

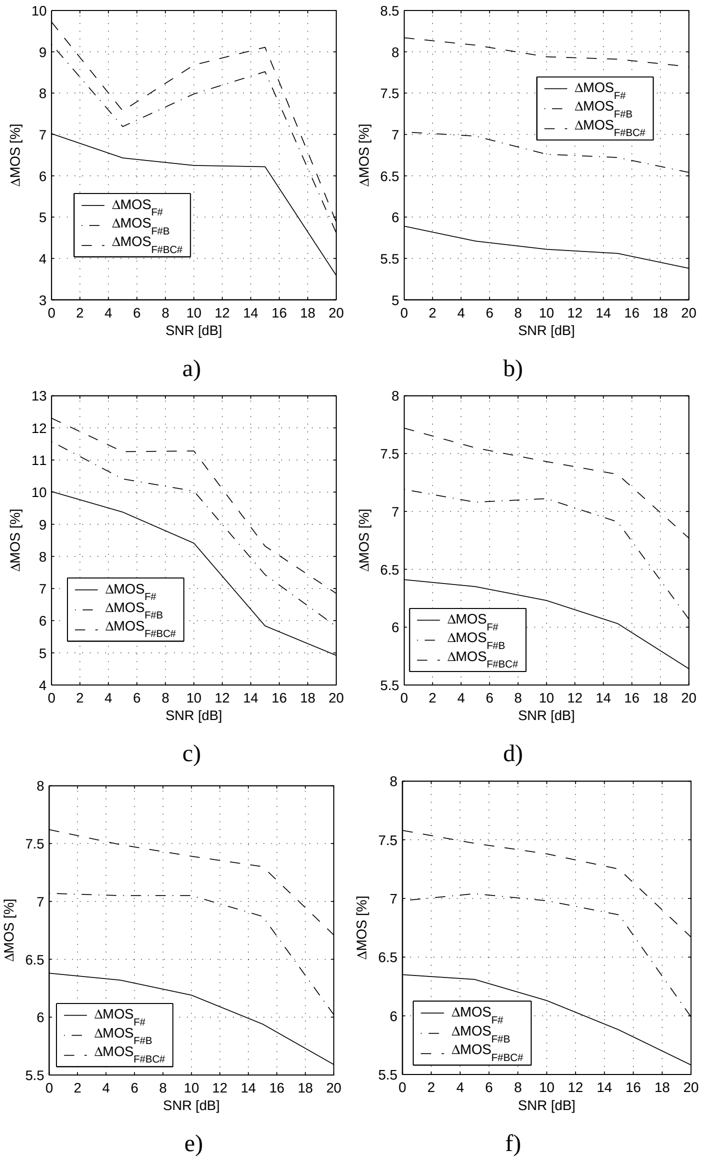

Fig. 2. Percentage increments of MOS values for: (a) White Gaussian Noise, (b) Computer Fan Noise, (c) Car Noise, (d) Babble Noise N=1, (e) Babble Noise N=4 i (f) Babble Noise N=8.

b) an algorithm for eliminating of dissonance F# and B in the speech signal with superimposed noise generates the speech signal whose MOS test result is 6.81-9.05% higher than that of an unprocessed signal. This result is in accordance with the result of the algorithm from [15];

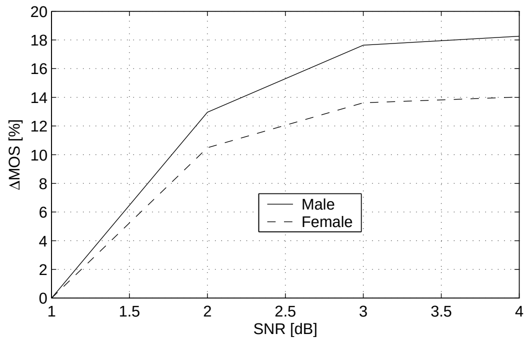

Fig. 3. Percentage increments of MOS values for Husky Voice.

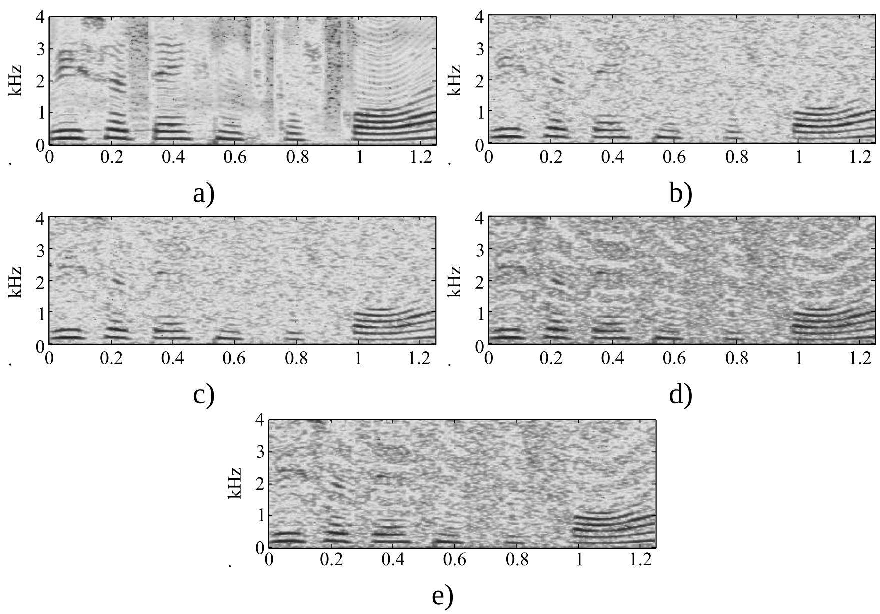

Fig. 4. Spectrogram of the speech signal: a) Clean speech signal, b) Clean speech signal with superimposed WGN SNR=5 dB, c) filtered F#, d) F#B and e) F#BC# dissonances.

Table 6. Percentage increment of MOS test results.

SNR	$\overline{\Delta MOS}_{\mathrm{F\#}}$ [%]	$\overline{\Delta MOS}_{\mathrm{F\#}}$ [%]	$\overline{\Delta MOS}_{\mathrm{F\#}}$ [%]
WGN	5.91	7.49	7.99
CFN	5.63	6.81	7.98
Car Noise	7.72	9.05	10.01
Babble Noise	6.09	6.82	7.31
Husky Noise	11.72	15.62	16.14

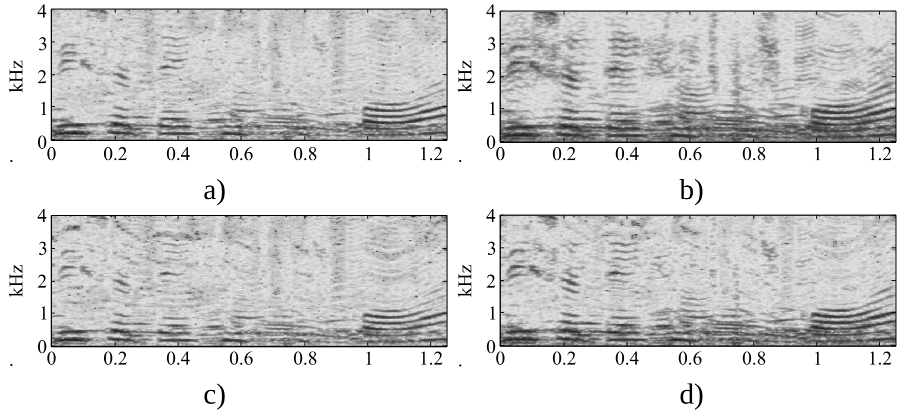

Fig. 5. Spectrogram of the speech signal: a) with 5 dB babble noise N=8, b) filtered F#, c) F#B d) F#BC# dissonances.

c) an algorithm for eliminating of dissonance F#, B and C# (DFF-FBC algorithm proposed in this paper) in the speech signal with superimposed noise generates the speech signal whose MOS test result is 7.31-10.01% bigger than that of an unprocessed signal;

d) the greatest effect DFF algorithms achieve with Husky Voice (Table 6), as follows: 11.72% (elimination of F#, the result from [13] is 11.955%), 15.625% (elimination of F# and B, the result from [15] is 15.277%) and 16.14% (DFF-FBC algorithm), respectively.

The special quality of DFF-FBC algorithm is the enhancement of the subjective quality effect of the sound with decreasing of SNR. This tendency is marked in all superimposed noises tested in this paper.

6 Conclusion

This paper presents DFF-FBC algorithm for eliminating of dissonant ranges in the speech signal spectrum. This algorithm is formed on the base of algorithms described in [13] and [15] by broadening of the activity range. The algorithm is based on the finding of the fundamental speech signal frequency and determining of frequency ranges where there are dissonant tones and their harmonics in all the octaves of the audio range. Dissonances which form the scale of perfect (complete) dissonances were analysed. The fundamental frequency F_0 is in relation to the perfect dissonances as the tone C is in relation to the tones F#, B and C#.

The effect of this algorithm on the speech signal was analysed on the base of the subjective MOS test results. Analyses were made on the speech signals to whom the White Gaussian Noise, Computer fan noise and Car noise were superimposed on the Husky Voice. MOS test results showed the subjective enhancement of the speech in range of 5.63-10.01%. The effect of the subjective quality enhancement increases if SNR decreases, which represents the special quality of this algorithm. The best marks DFF-FBC got for Husky Voice processing is 16.14%, which is better in comparison to the algorithm from [15]. The results presented in this paper spaek in favour of implementing of DFF-FBC algorithm for the speech signal preprocessing in algorithms for compressing, recognizing (identification) of speech etc.

Further scientific researches will be directed toward determining of the effect of speech signal quality enhancement by eliminating of dissonant intervals which belong to the group of imperfect (incomplete) dissonances.

Acknowledgments

The authors of this paper thank to prof. Nevenka Balanesković, prof. Goran T. Djordjević and prof. Milica Krstić for the assistance in its realization.

References

[1] J. Festen and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," *J. Acoust. Soc. Am.*, vol. 88, pp. 1725–1736, 1990.

[2] G. Schmidt and T. Haulick, "Signal processing for in-car communication systems," *Signal Process.*, vol. 86, pp. 1307–1326, 2006.

[3] R. Udrea, N. Vizireanu, and S. H. S. Ciochina, "Nonlinear spectral subtraction method for colored noise reduction using multi-band bark scale," *Signal Process.*, vol. 88, pp. 1299–1033, 2008.

[4] T. Waterschoot, G. Rombouts, P. Verhoeve, and M. Moonen, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process*, vol. 55, pp. 846–858, 2007.

[5] Z. Milivojević and M. Mirković, "Performances of the near-and coherent speech detector," *Nonlinear Phenomena in Complex Systems*, vol. 9, pp. 370–379, 2006.

[6] G. Rombouts, A. Spriet, and M. Moonen, "Generalized sidelobe canceller based combined acoustic feedback- and noise cancellation," *Signal Process.*, vol. 88, pp. 571–581, 2008.

[7] B. Moor, "The singular value decomposition and long and short spaces of noisy matrices," *IEEE Trans. Signal Process*, vol. 41, pp. 2826–2838, 1993.

[8] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise," *IEEE Signal Process. Lett.*, vol. 10, pp. 104–106, 2003.

[9] L. Arslan, “Modified wiener filtering,” *Signal Process.*, vol. 86, pp. 267–272, 2006.

[10] J. Lim, A. Oppenheim, and L. Braida, “Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 26, pp. 354–358, Apr. 1991.

[11] P. Lanchantin, J. Lapuyade, and W. Pieczynski, “Unsupervised segmentation of triplet markov chains hidden with long-memory noise,” *Signal Process.*, vol. 88, pp. 1134–1151, 2008.

[12] Y. Ephraim, “Statistical-model-based speech enhancement systems,” *Proc. IEEE*, vol. 80, pp. 1526–1555, Oct. 1992.

[13] B. Joen, S. Kang, S. Baek, and K. Subg, “Filtering of a dissonant frequency based on improved fundamental frequency estimation for speech enhancment,” *IEICE Trans. Fundamentals*, vol. E86-A, pp. 2063–2064, Aug. 2003.

[14] J. Tennenbaum, “The foundations of scientic musical tuningt,” *FIDELIO Magazine*, vol. 1, 1991.

[15] S. Kang, “Dissonant frequency filtering technique for improving perceptual quality of noisy speech and husky voice,” *Signal Process*, vol. 84, pp. 431–433, 2004.

[16] A. Kacha, F. Grenez, and K. Benmahammed, “Time frequency analysis and instantaneous frequency estimation using two-sided linear prediction,” *IEEE Signal Process*, vol. 85, pp. 491–503, 2005.

[17] L. Qiu, H. Yang, and S. Koh, “Fundamental frequency determination on instantaneous frequency estimation,” *IEEE Signal Process*, vol. 44, pp. 233–241, 1995.

[18] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust., Speech and Signal Process*, vol. 29, pp. 1153–1160, 1981.

[19] H. Pang, S. Baek, and K. Sung, “Improved fundamental frequency estimation using parametric cubic convolution,” *IEICE Trans. Fundamentals*, vol. E83-A, pp. 2747–2750, 2000.

[20] Z. Milivojević and M. Mirković, “Estimation of the fundamental frequency of the speech signal modeled by the sympes method,” *AEUE - International Journal of Electronics and Communications*, vol. 63, pp. 200–208, 2009.

[21] P. Mokhtari, H. Takemoto, and T. Kitamura, “Single-matrix formulation of a time domain acoustic model of the vocal tract with side branches,” *Speech Commun*, vol. 50, pp. 179–190, 2008.

[22] S. Simpson and M. Cooke, “Consonant identification in n-talker babble is nonmonotonic function of n,” *J. Acoust. Soc. Am*, vol. 118, pp. 2775–2778, Nov. 2005.