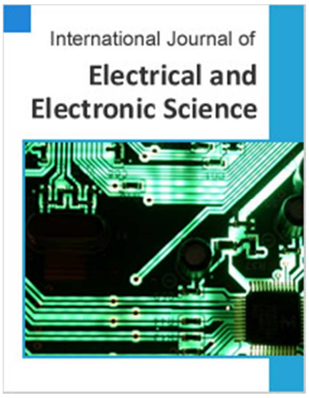
International Journal of Electrical and Electronic Science
2015; 2(2): 31-36
Published online August 30, 2015 (http://www.aascit.org/journal/ijees)
ISSN: 2375-2998

American Association for Science and Technology



Keywords
Segmental Signal-to-Noise Ratio,
Estimator Stability,
Estimation Error,
Time-Alignment Error,
Sampling Rate

Received: July 31, 2015
Revised: August 10, 2015
Accepted: August 11, 2015

Reducing Sensitivity of Segmental Signal-to-Noise Ratio Estimator to Time-Alignment Error

Arkadiy Prodeus

Acoustic and Electroacoustic Department, Faculty of Electronics, National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine

Email address
aprodeus@gmail.com

Citation
Arkadiy Prodeus. Reducing Sensitivity of Segmental Signal-To-Noise Ratio Estimator to Time-Alignment Error. *International Journal of Electrical and Electronic Science*. Vol. 2, No. 2, 2015, pp. 31-36.

Abstract
In this paper, quantitative assessment of influence of the time-alignment error on segmental signal-to-noise ratio (SSNR) estimation is made.It is shown that an effective way to reduce sensitivity of SSNR estimator to time-alignment error is to increase the sample rate of the compared signals in 2...4 times by means of their interpolation. It was founded also that when distorted signal is a result of adjusting FIR filtering, the filter order must be odd for minimizing the SSNR estimation error.

1. Introduction

Speech quality assessment is an urgent task in communication systems, in automatic speech recognition systems, hearing aids. A variety of objective (instrumental) speech quality measures is used because of their ability of minimizing time and cost of estimation procedure [1,2,3,4,5,6,7]. The number of objective measures is huge and reaches several thousand [2]. It is therefore understandable desire to choose a quality measure that is easy to calculate. On the other hand, there is a real danger that the selected quality measure does not provide a high-reliability evaluation.

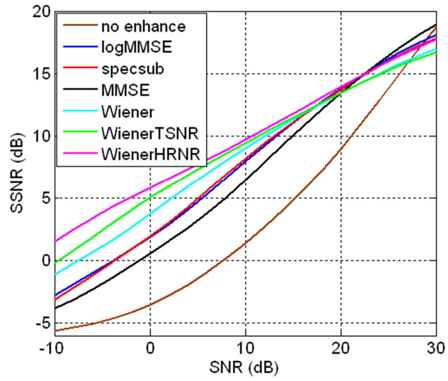
A widely used instrumental parameter of speech quality, easy to compute and well understood, is the segmental signal-to-noise ratio

$$SSNR = \frac{1}{L}\sum_{l=1}^{L} 10\lg\left[\frac{\sum_{n=Rl}^{Rl+N-1} x^2(l,n)}{\sum_{n=Rl}^{Rl+N-1} [x(l,n)-y(l,n)]^2}\right] \tag{1}$$

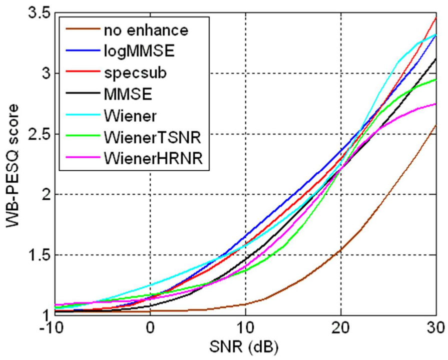
where $x(l,n)$ and $y(l,n)$ are n th sample of l th frame of reference and degraded signals $x(n)$ and $y(n)$, respectively, L is number of frames, N is frame length, R is frame number [1,5,7].

SSNR measure, among nine objective speech quality measures, had been used in [8] for comparing of six noise reduction algorithms. Comparison of SSNR and perceptual evaluation speech quality (PESQ) measure graphs had shown sufficiently good reliability of SSNR indicator (Fig. 1).

Analysis of features of some objective indicators of dereverberation algorithm quality had been performed in [9,10], and it was shown that SSNR indicator may be useful as speech quality measure. At the same time it was shown in [8,9,10] that SSNR measure is not as good as competitive measures.



a



b

Fig. 1. *SSNR (a) versus PESQ (b) [8].*

However, a lot of authors noticed the relatively low efficiency both traditional overall SNR and segmental SNR [1,2,3,4,5]. In particular, it was shown in [2] that correlation coefficient between overall SNR and Diagnostic Acceptability Measure (DAM) subjective measures about 0.24, and the same coefficient between SSNR and DAM is approximately 0.77. In [3], the SSNR measure was found to be a significantly better predictor of speech quality than the overall SNR. It was shown in [4,5] existence of a set of speech quality indicators which are much better than SSNR. It was noted in [1,6] that SSNR is highly dependent on the time-alignment and phase shift between the reference and degraded speech signals. Unfortunately, the quantitative assessment of this impact was not given in [1,6].

Four speech quality measures - SSNR, log-spectral distortion (LSD), bark-spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ) - were compared in [7] in accordance with scheme shown in Fig. 2. As a result, the dependences of these measures on the bandwidth Δf of the FIR low-pass filter had been received.

It was naturally suppose that speech quality increases monotonically when bandwidth Δf is increasing. It turned out that the functions $BSD(\Delta f)$ and $PESQ(\Delta f)$ behaviour is consistent with the requirement of monotony. The behaviour of the function $LSD(\Delta f)$ was also satisfactory, despite the small local disturbances of monotony.

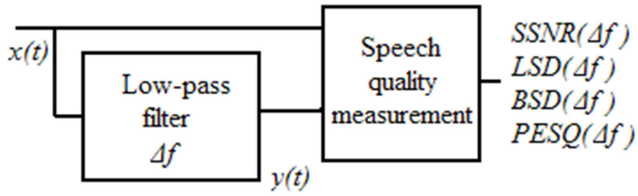


Fig. 2. *Research scheme in [7].*

However, although the disturbances of $SSNR(\Delta f)$ monotony were anomalously large (Fig. 3), the reasons of this appearance were not analysed in [7]. The object of this paper is the study of this phenomenon and developing recommendations on its suppression.

It is hoped that scientific and practical usefulness of this problem solution will be the better understanding of the limitations imposed on the use of SSNR measure in studying and developing the noise and reverberation suppression algorithms [8,9,10]. In addition, thanks to the simplicity of the studies scheme shown in Fig. 1, it is hoped that this scheme will be useful in the study of the reliability of other speech quality measures which are insufficiently studied, despite their popularity.

2. Analysis

2.1. Simple Analytical Model

The quantitative assessment of influence of the time-alignment error on SSNR estimation can be easily made analytically using sinusoidal signal. In this case, signals waveform is unvarying in the observation interval T, so the values of overall and segmental SNR are the same. Therefore, when setting

$$x(t) = \cos 2\pi f_0 t\,, \quad y(t) = \cos 2\pi f_0 (t - \tau) \tag{2}$$

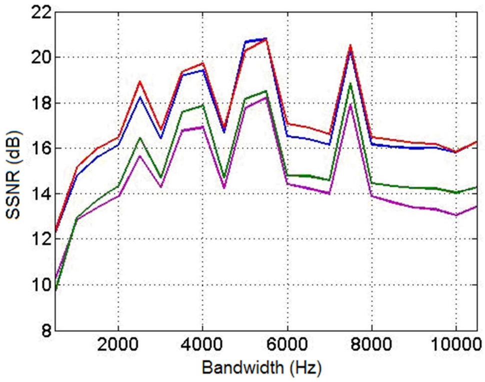
where f_0 is sinusoid frequency, τ is time-alignment error, we obtain

$$SSNR = SNR = 10\lg \frac{\int_0^T x^2(t)\,dt}{\int_0^T [x(t) - y(t)]^2\,dt} = 10\lg \frac{0,5}{1 - \cos 2\pi f_0 \tau}. \tag{3}$$

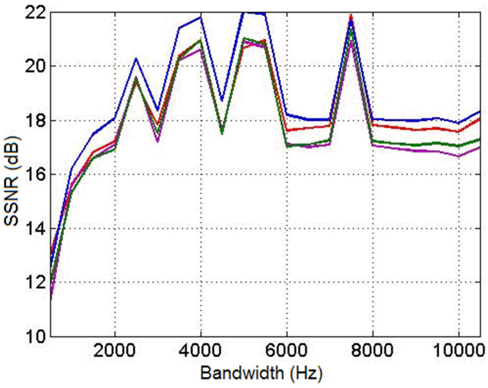
For $f_0 = 1$ kHz and sample rate $F_s = 22050$ Hz, we obtain from (3)

$$SSNR = \begin{cases} 10,93\,dB, & \tau = \Delta t, \\ 16,93\,dB, & \tau = 0,5\Delta t \end{cases} \tag{4}$$

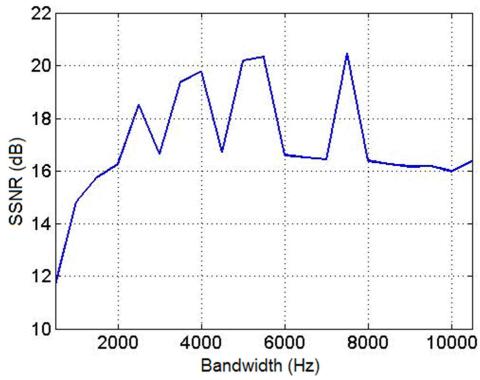
where $\Delta t = 1/F_s$ is sampling period.



a



b



c

Fig. 3. *SSNR estimation: male (a), female (b) and averaged (c) [7].*

It follows from (4) that even when approximately 22 samples are placed on the period of the harmonic signal, the alignment error, which is comparable with the sampling interval, leads to significant fluctuating of SSNR. It follows also from (4) that for $f_0 \approx F_s/2 = 11025$ Hz, when only 2 samples are placed on the harmonic signal period, SSNR value is minimum:

$$SSNR = \begin{cases} -6,02\ dB, & \tau = \Delta t, \\ -3,01\ dB, & \tau = 0,5\Delta t \end{cases} \tag{5}$$

Speech and music spectra are enclosed essentially in the frequency range from 80 Hz to 11 kHz. As it follows from the above calculations, it is possible reduce the effect of the alignment error at the SSNR calculation results by means of sampling rate increasing or, what is much cheaper, by means of signals' interpolation. Matlab function *resample* can be used for this purpose.

2.2. Computer Simulation

Estimation of SSNR had been made using (1) and scheme Fig. 2. Four signals were used for the computer simulation:

- sinusoid at 1 kHz;
- speech signal;
- music fragment (*Bolero* by M. Ravel);
- band-limited white noise.

The duration of signals and frames were 60 s and 32 ms, respectively (it was noted in [3] that SSNR isn't highly sensitive to the frame length: when frame length varied between 8 and 32 ms, the difference in the computed SSNR was found to be less than 1 dB). Three sample rate values had been used in the study when interpolating: 22050 Hz, 44100 Hz and 96000 Hz. Bandwidth Δf of low-pass Remez filter varied from 500 Hz to 10500 Hz with increment value 500 Hz. Transfer function parameters were as follows: the transition bandwidth $0.05\Delta f$, the passband ripple 1 dB, the attenuation in the stopband -80 dB. Results of SSNR calculations are shown in Fig. 4.

3. Proposed Solution

3.1. Interpolation of Compared Signals

As can be seen, all SSNR graphs include jumps in the same places. But the graphs for all the signals (except sinusoid) are becoming more smooth and monotonically increasing with the increase of sampling rate in 2…4.4 times in comparison with the initial value of $F_s = 22050$ Hz. Matlab function *resample* had been used for interpolation in the paper. It follows that upon SSNR calculation, sampling frequency should be increased, by interpolation, several times to obtain correct results.

3.2. Filter Order Correction

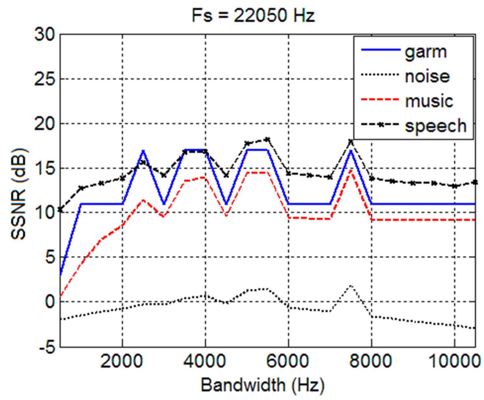
Despite the achievement, in general, positive results upon signals interpolation, it remains unexplained abrupt behaviour ("jumps") of $SSNR(\Delta f)$ curves. As can be seen in case of sinusoid, function $SSNR(\Delta f)$ in Fig.4a takes two possible values of approximately 11 dB and 17 dB (except for case Δf = 500 Hz when the sinusoid is suppressed by filter). Good agreement of these values with ones of (4) makes it possible to suggest that used software contains features that lead to time alignment errors, which are equal to $\tau = \Delta t$ and $\tau = 0.5\Delta t$.

Checking the validity of this assumption has shown that such features are present. When aligning the signals, it has been previously necessary to calculate half duration of the filter impulse response. To do this, it was used the following

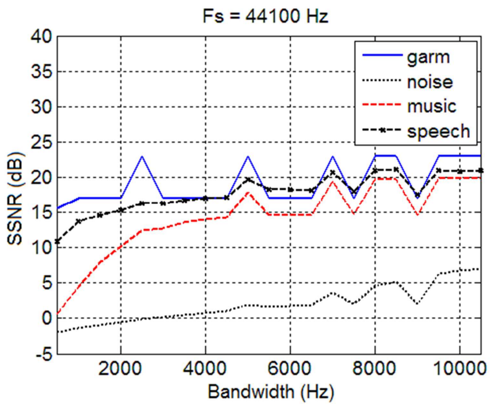
Matlab command:

$$\text{len\_a05} = \text{round}(0.5*\text{len\_a}) \quad (6)$$

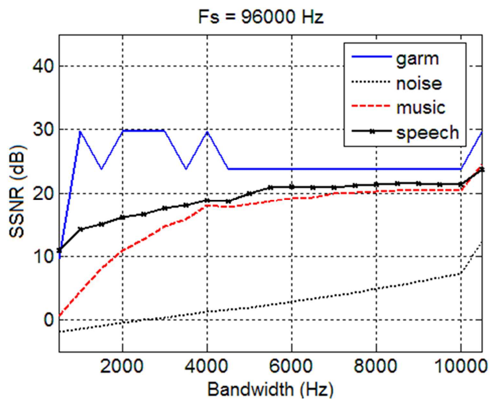
where *len_a* is a filter order.



a



b



c

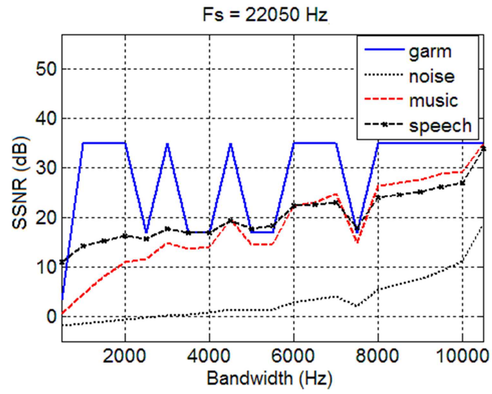
Fig. 4. *SSNR(Δf) for alignment error (0.5…1) Δt : F_S =22050 Hz (a), F_S =44100 Hz (b), F_S =96000 Hz (c).*

When calculating the coefficients of the lowpass Remez filter (Parks-McClellan algorithm), filter order can be either even or odd, and property "even-odd order" is practically uncontrollable.

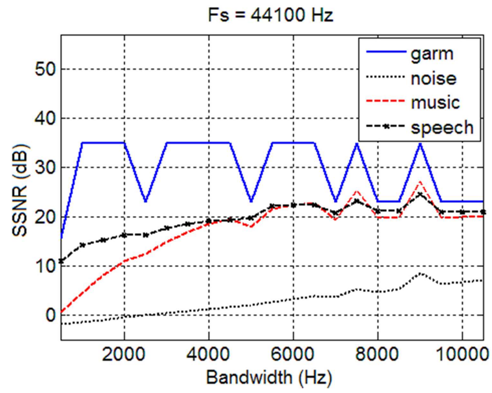
The alignment error $\tau = \Delta t$ is caused by the fact that, when filter order is odd, calculation in accordance with (6) leads to parameter *len_a* value which is per unit larger than true value. To correct this error, it is necessary use the *floor* function instead of the *round* function (Fig. 5).

As can be seen from Fig. 5, now the SSNR value for sinusoid is 35 dB in some cases. It means that alignment error is zero in these cases (35 dB threshold value was set in computer routine in order to avoid exponent overflow).

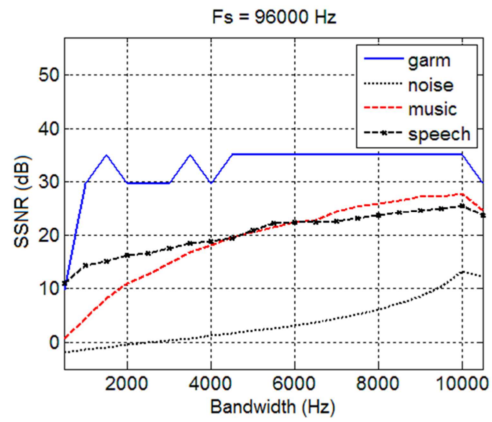
However, as we see, it remained uncompensated alignment error $\tau = 0.5\Delta t$ due to fundamental inability to accurately align the signals in cases of the even-order filters.



a

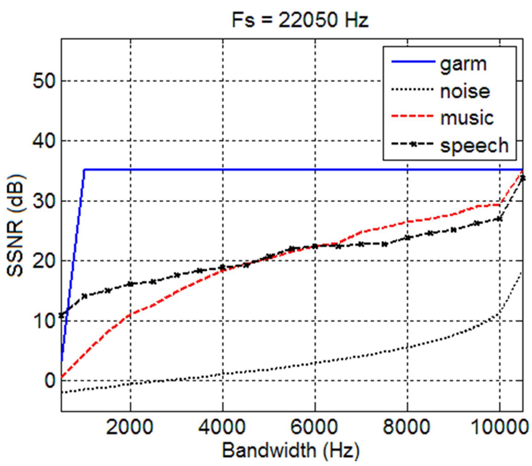


b

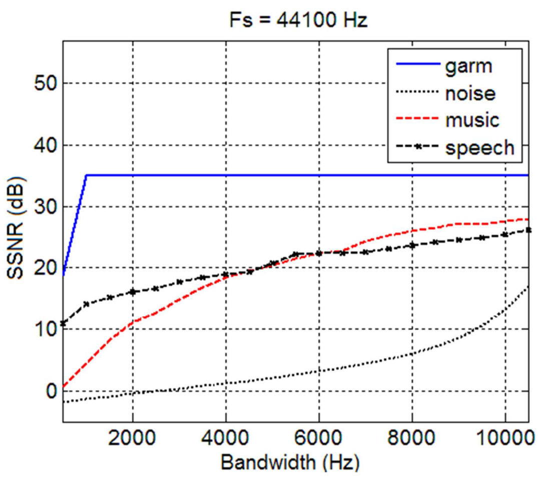


c

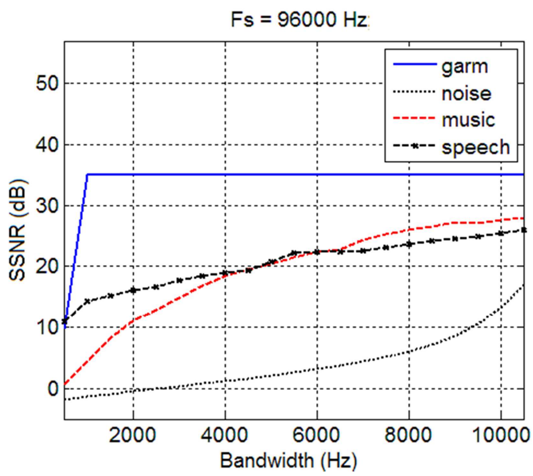
Fig. 5. *SSNR(Δf) for alignment error (0…0.5) Δt : F_S =22050 Hz (a), F_S =44100 Hz (b), F_S =96000 Hz (c).*



a



b



c

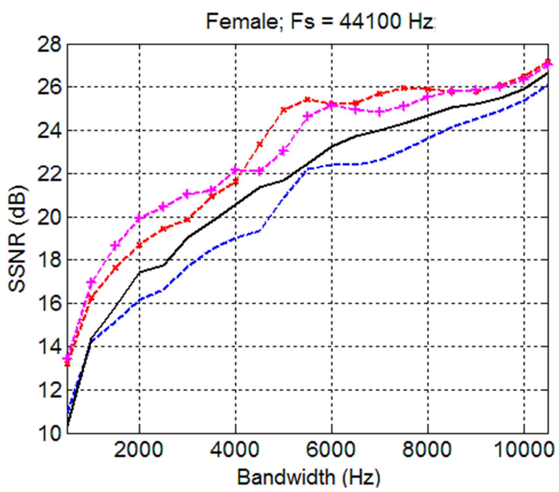
Fig. 6. *SSNR*(Δf) *for zero alignment error:* F_S *=22050 Hz (a),* F_S *=44100 Hz (b),* F_S *=96000 Hz (c).*

To eliminate this shortcoming, we propose in this paper to extrapolate 2n filter coefficients to the 2n+1 filter coefficients. For example, it may be used next commands in computer program:

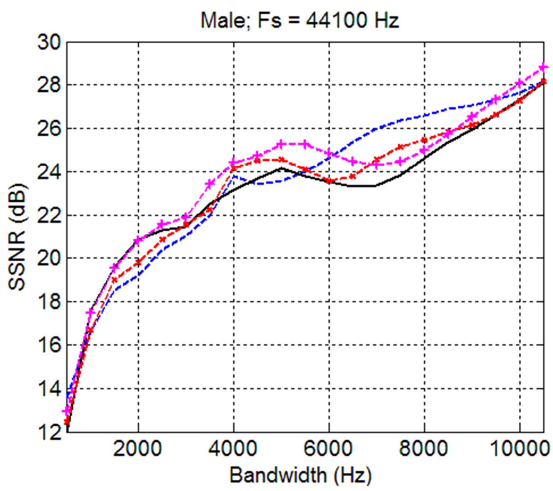
```
if mod(len_a,2)==0
len_a1 = len_a + 1;
  arg_a = 1:len_a;
arg_a1 = 0.5:len_a+0.5;
  a1 = interp1(arg_a,a,arg_a1,'spline');
  a = a1;
len_a = length(a);
end
```

Corrected results are shown in Fig. 6. As follows from Fig. 6, the measures taken have led to a positive result. For a sinusoid, which has the simplest waveform, function $SSNR(\Delta f)$ is "the upper limit" of $SSNR(\Delta f)$ graphs for signals with more complex waveform. In contrast, for white noise with the most complex waveform, function $SSNR(\Delta f)$ represents the "lower limit". Situations of speech and music occupy an intermediate position between harmonic signal and white noise, so their respective $SSNR(\Delta f)$ curves are concluded between the borderline cases.

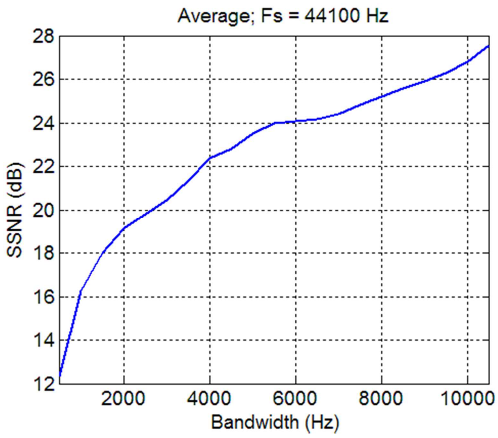
With regard to the accuracy of calculations, it was sufficient when doubling the Nyquist frequency.



a



b



c

Fig. 7. *SSNR(Δf) for doubled F_S (F_S =44100 Hz): male (a), female (b) and averaged (c).*

4. Corrected Graphs for Speech Signals

SSNR(Δ*f*) graphs adjusted in accordance with above recommendations are shown in Fig. 7. As can be seen, averaged over all speakers dependence *SSNR*(Δ*f*) (Fig. 7c) is monotonous and differs significantly from the Fig. 3c graph. However, the graphs for individual speakers contain small local defects of monotony (Fig. 7a, b). Natural to assume that these disturbances are caused exclusively by the properties of the SSNR measure, as well as by the features of the analyzed speech signals. It is advisable to check the validity of these assumptions in the future.

5. Conclusion

Indicator SSNR can be used to assess the quality of the filtered signals, but it should take into account its high sensitivity to the time alignment error. An effective way to reduce this sensitivity is to increase the sample rate by interpolation of the compared signals in 2...4 times. Further, the order of the filter must be odd for precise time-alignment of signals.

If the current time-alignment of signals is impossible due to asymmetric impulse response of the filter, this may require a more significant, than 2...4 times, upsampling. In this case, the specific value of the final sampling rate should be determined experimentally, by gradually increasing it and stopping when the SSNR value is stabilized.

It is hoped that simple studying scheme used in the paper will be useful in the examination of the reliability of speech quality measures.

References

[1] J. Benesty, M. Sondhi, Y. Huang (Ed.), Springer Handbook of Speech Processing.BerlinHeidelberg: Springer-Verlag, 2008.

[2] S. Quackenbush, T. Barnwell, and M. Clements, Objective Measures of Speech Quality, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[3] P. Mermelstein,"Evaluation of segmental SNR measure as an indicator of the quality of ADPCM coded speech," J. Acoust. Soc. Am., Vol. 66(6), pp.1664–1667, 1979.

[4] J.H. Hansen and B.L. Pellom, "An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms," In Proc. 5th Int. Conf. on Spoken Language Processing (ICSLP), Vol. 7, pp. 2819–2822, AUS–Sydney, 1998.

[5] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," IEEE Transactions on Speech and Audio Processing, 16(1), pp. 229-238, 2008.

[6] N. Cote, Integral and diagnostic intrusive prediction of speech, Springer-Verlag Berlin Heidelberg, 2011.

[7] N. Bogdanova, A. Prodeus. Objective quality evaluation of speech band-limited signals. Electronics and Communications, Vol.19, No.6(83), pp.58-65, 2014. [Online]. Available: http://aprodeus.narod.ru/About/Articles/elc.kpi.ua_images_pdf_Arhiv_1_6-83-.pdf

[8] A. Prodeus, "Assessment of speech enhancing algorithms quality," Proc. of V International Scientific Conference "Signal and non-Gaussian noise processing", Cherkasy State Technological University, May 20–22, 2015, Cherkasy, Ukraine (in Ukrainian).

[9] P. Naylor, N. Gaubitch, Speech Dereverberation, Springer, 2010.

[10] A. Prodeus, "Parameter Optimization of the Single Channel Late Reverberation Suppression Technique," Proc. of IEEE 35th International Scientific Conference on Electronics and Nanotechnology (ELNANO-2015), pp.269-274, 21-24 April 2015, Kyiv, Ukraine.