UDC 004.934

N.V. Bogdanova, Ph.D., **A.M. Prodeus**, Dr.Sc.
National Technical University of Ukraine "Kyiv Polytechnic Institute",
off. 233, Politekhnichna Str., 16, Kyiv, 03056, Ukraine.

Objective quality evaluation of speech band-limited signals

Dependence of objective quality evaluation of speech band-limited signals is experimentally obtained. As part of this task, a comparison of the considered indicators of the speech quality had been made. It is shown that computationally simple indicators, such as segmental SNR (SSNR) and log-spectral distortion (LSD), may not adequately respond to changes in bandwidth. More complex computationally perceptual indicators, such as bark spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ), behave much more correct and, in the end, clarify the real needs of the human auditory system to speech perception.
Reference 14, figures 5.

Keywords: *signal bandwidth, voice quality, quality indicators.*

Introduction

It is supposed to use super-wideband (50 Hz - 14 kHz) signal at a sampling frequency of 48 kHz in the standard ITU-T Rec. P.863 (POLQA) [9, 11] for a modern commercial communication. Speech signal may be transformed for transmission in wide band (50 Hz – 7 kHz) and narrow band (300 Hz – 3,4 kHz) after proper band-pass filtering and sampling down to 16 or 8 kHz, accordingly. Obviously, the inclusion super-wideband (SWB) in the modern standards of commercial communications stems from a desire to improve the quality of communication. This is evidenced by the following circumstantial evidence presented in [9]: the maximum quality of the speech signal in a narrow band is estimated to be 4,5 points on MOS scale, and super-wideband maximum quality is 4,75 points. Unfortunately, it is difficult to find in literature information about dependence of estimates of real (i.e. no maximum) speech quality on the signal bandwidth [2, 3, 7, 13]. Meanwhile, the issue, in our opinion, is of undoubted theoretical and practical interest, as paired with the clarification of the real needs of the human auditory system.

The other side of the raised issue is the choice of the quality index of the speech signal. Subjective assessment methods are very resource intensive, so the attention of researchers is aimed at finding objective (instrumental) indicators of speech quality. Today, the best solution would be to use the standard ITU-T P.863 (POLQA), which most fully takes into account the effect of confounding factors and features of the human auditory system. However, the use of this standard for scientific purposes is practically impossible, since access to the source code of the corresponding software is closed. So you have to either use outdated index PESQ [1, 10, 14], or to look for alternative, more computationally simple indicators, allowing for the possibility of reduced effectiveness. Unfortunately, there is no clear evaluation of the potential of objective measures of speech quality in the solution of certain problems in the literature.

The object of the paper is filling, at least in part, the above-mentioned gaps.

1. Objective quality measures of speech signals

To get the dependence of speech quality estimates on the frequency band occupied by the signal, let us use a series of low-pass filters instead of the exact models of the band-pass filters used in narrowband (NB), wideband (WB) and SWB modes. Successively increasing the cut-off frequency of the filter, one would expect the growth of the quality of the filtered speech signal. Obviously, the used quality indicators must, as a minimum, adequately reflect this growth. Otherwise, the quality indicators should recognize ineffective.

Subjective methods for evaluating the speech quality, suggesting the participation in the experiments of several speakers and several auditors, have the undoubted advantage that real human auditory system is used in this estimation. Obvious drawback of subjective methods is their high requirement to resources.

Objective (instrumental) methods for speech quality estimation are largely free of these shortcomings. There are two approaches to estimation and, consequently, two kinds of speech quality indicators, when using objective methods [3]:

1) with use of a reference signal (intrusive indicators);

2) without the use of a reference signal (non-intrusive indicators).

Only intrusive indicators, providing the greatest proximity to the results of the subjective evaluation, had been used in this paper.

From the set of the currently known indicators of this kind [2, 3, 7, 8, 13], we consider four. They are segmental signal to noise ratio (SSNR), logarithmic spectral distortion (LSD), bark spectral distortion (BSD) and perceptual evaluation of speech quality (PESQ). In justifying this choice, we note that the first two indicators - SSNR and LSD - are very attractive due to ease of computation, while the other two indicators, referred to as "perceptual" - BSD and PESQ - have the advantage that they allow to take into account, with varying degrees of accuracy, features of the human auditory system.

Analytical description of the above-mentioned indicators is next:

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

$$LSD = \frac{2}{KL}\sum_{l}\sum_{k=0}^{\frac{K}{2}-1}\left|G\{X(l,k)\}-G\{Y(l,k)\}\right|, \quad (2)$$

$$G\{X(l,k)\} = \max\{20\lg(|X(l,k)|), \delta\},$$
$$\delta = \max_{l,k}\{20\lg(|X(l,k)|)\} - 50,$$

$$BSD = \frac{\sum_{l=1}^{L}\sum_{k=0}^{\frac{K}{2}-1}\left[B\{X(l,k)\}-B\{Y(l,k)\}\right]^2}{\sum_{l=1}^{L}\sum_{k=0}^{\frac{K}{2}-1}\left[B\{X(l,k)\}\right]^2}, \quad (3)$$

where $x(l,n)$ and $y(l,n)$ are n-th samples of l-th frame of input and output filter signals $x(n)$ and $y(n)$, respectively; $X(l,k)$ and $Y(l,k)$ are amplitude spectrums of l-th frame of signals $x(n)$ and $y(n)$, respectively; $B\{X(l,k)\}$ and $B\{Y(l,k)\}$ are bark-spectrums of l-th frame of signals $x(n)$ and $y(n)$, respectively.

Analytical description of a very cumbersome algorithm of a PESQ calculation, which is significantly improved, compared with BSD, to incorporate features of the human auditory system, is presented in [1].

2. Some features of BSD and PESQ calculation

BSD and PESQ are the most computationally complex indicators among objective indicators considered in this work. However, this computational complexity is compensated with high quality estimation: relatively high Pearson correlation coefficient was achieved between results of objective and subjective evaluation ($r = 0{,}85$-$0{,}95$) [2, 3, 7, 13]. In this regard, it is interesting how one can overcome the estimation difficulties of indexes BSD and PESQ in Matlab.

Comparing the definitions of "bark spectrum" and "PLP-spectrum" (perceptual linear predictive spectrum), given in [3, 5], it is easy to come to a conclusion about the identity of these concepts, since in both cases it is assumed that the following computational steps are made:

- calculation of signal power spectrum $P(\omega)$;
- transformation of frequency scale ω to bark scale Ω;
- formation of bark filter frequency response $\Psi(\Omega)$;
- convolving of power spectrum $P(\Omega)$ with frequency response $\Psi(\Omega)$;
- the obtained result is multiplied by the isophone $E(\omega)$;
- the loudness scale is corrected by means of cubic root calculation of the previous step result (phone is translated in sone).

This identity can be used to compute the bark spectrum by means of ready programs from the library rastamat [4]. They are rastaplp, powspec, audspec, fft2barkmx, hz2bark, bark2hz, postaud, spec2cep, lifter.

However, some correction of these programs was required before their using. Firstly, a modern function spectrogram need be used instead of the obsolete function specgram in program powspec. Secondary, frame length (32 ms was used in the paper) and frame shift (16 ms was used) must be specified in the program rastaplp when calling the program powspec.

Specifying the input data when the program rastaplp starts, we must reject the RASTA-spectrum calculation, and must specify the zero order model. Bark spectrum assessment is obtained as the result of the command executing:

```
[cepstra, spectra] = rastaplp(x, fs, 0,0)
```

The results of cepstrum calculation are discarded as unusable in the future.

Note that PESQ calculation can be realised in accordance with early algorithm version (standard ITU-T P.862), and the later version (standard ITU-T P.862.2) [10]. For brevity, we will call them PESQ and PESQ-2, respectively. In this paper both versions are used, allowing to compare the results of their operation.

Although the PESQ index is only designed for narrowband telephony, but the sampling frequency of the analyzed signals can be used either 8 or 16 kHz when the PESQ calculations are realised in the Matlab [6].

Indicator PESQ-2 is designed for both narrowband and wideband telephony. It can be calculated in Windows, using the console application pesq.exe, which is result of compiling the source code written in C and available in the public domain [14]. Another, more convenient way of calculating the PESQ-2 is to control the pesq.exe application from Matlab. To implement this method function pesq2_mtlb, presented in [12], was used.

3. Experimental results

When evaluating speech quality, there were recorded 1 minute length speech signals of each for 4 speakers female and 4 male speakers reading text on juridical topics. Signal recording had been made at the Department of Acoustic of National Technical University of Ukraine "Kyiv Polytechnic Institute", in anechoic room with a reverberation time of 0,15 s, with a sampling rate of 22050 Hz and a bit depth of 16 bits.

Set of FIR low-pass filters was synthesized by Remez method by means of Matlab (fdatool). Filters features are:

- cutoff frequency varies from 0,5 kHz to 10,5, incrementing of 0,5 kHz;
- the size of the transition zone is 5% from bandwidth;
- ripple in the pass band is 1 dB;
- transfer coefficient in the stop band is minus 80 dB.

Speech signals quality calculation results at the filter output are shown in Fig. 1-4.

Note that SSNR index is clearly inefficient since its values are non-monotonic and fluctuate significantly when bandwidth increasing. This conclusion agrees with the findings of [3] about the unsuitability of SSNR index to assess the distortion caused by filtration.

LSD index is much better "on average", however, and its drawback is local violations of monotonic dependence on the frequency band. Therefore LSD should be also recognize as ineffective index.

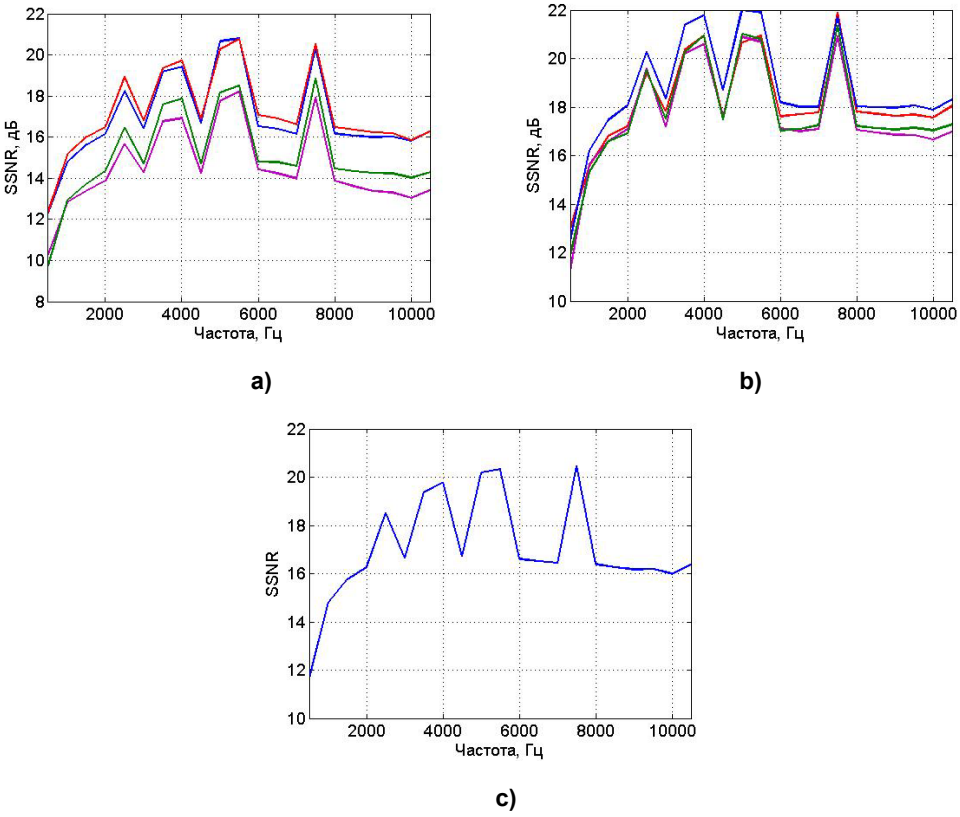


Fig. 1. SSNR index: female (a), male (b), averaged (c)

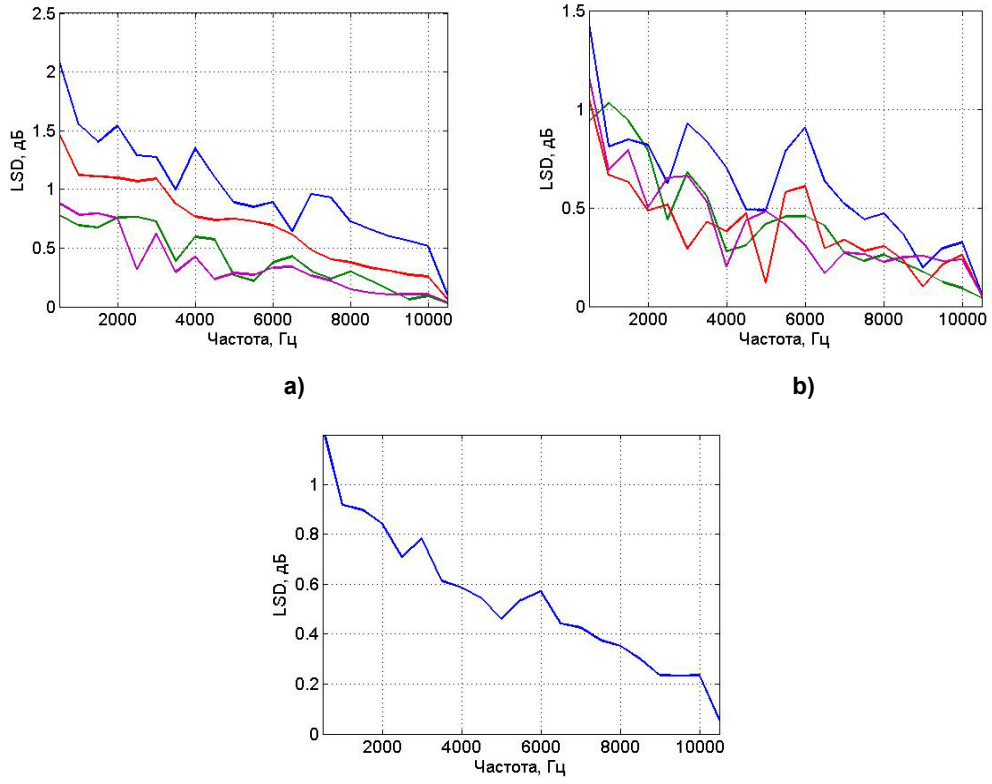


Fig.2. LSD index: female (a), male (b), averaged (c)

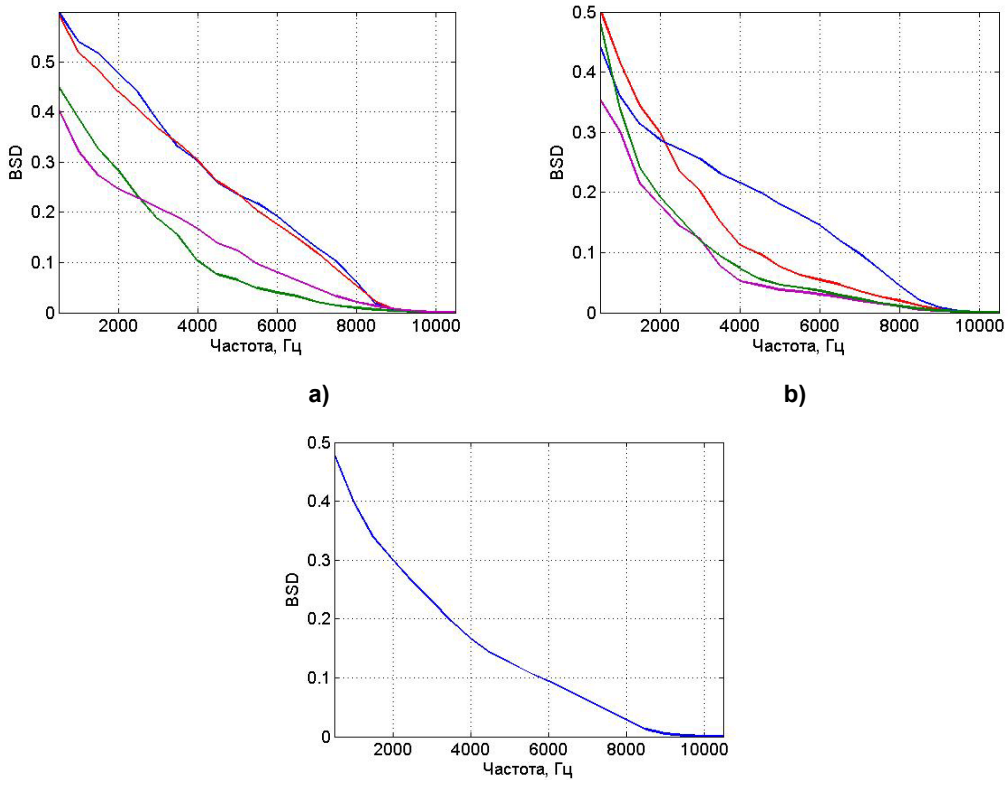


Fig. 3. BSD index: female (a), male (b), averaged (c)

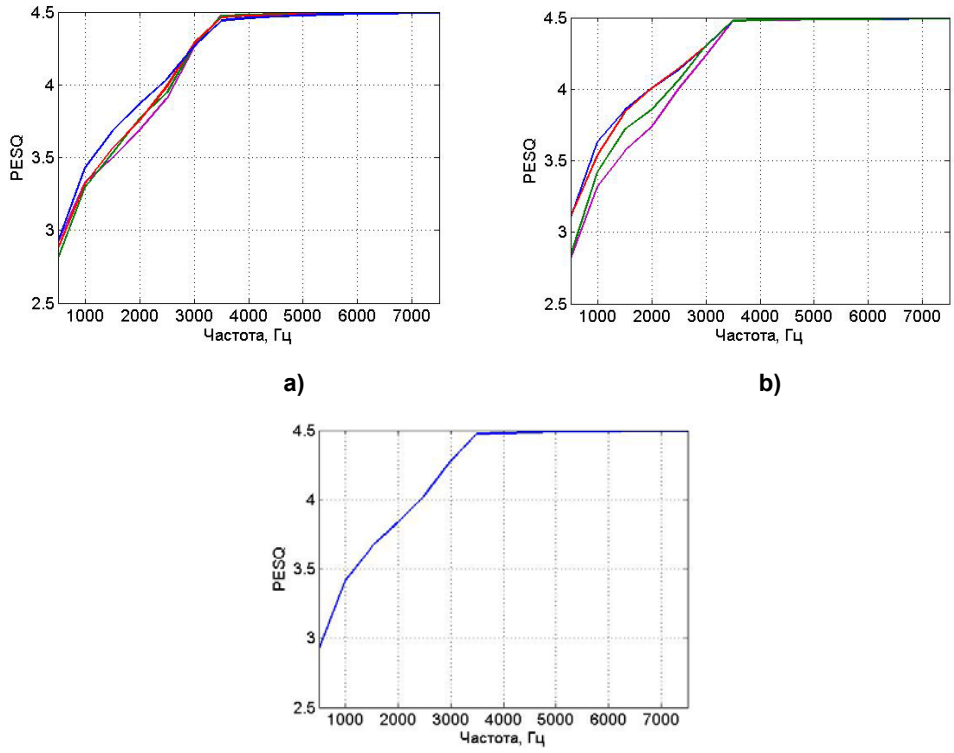
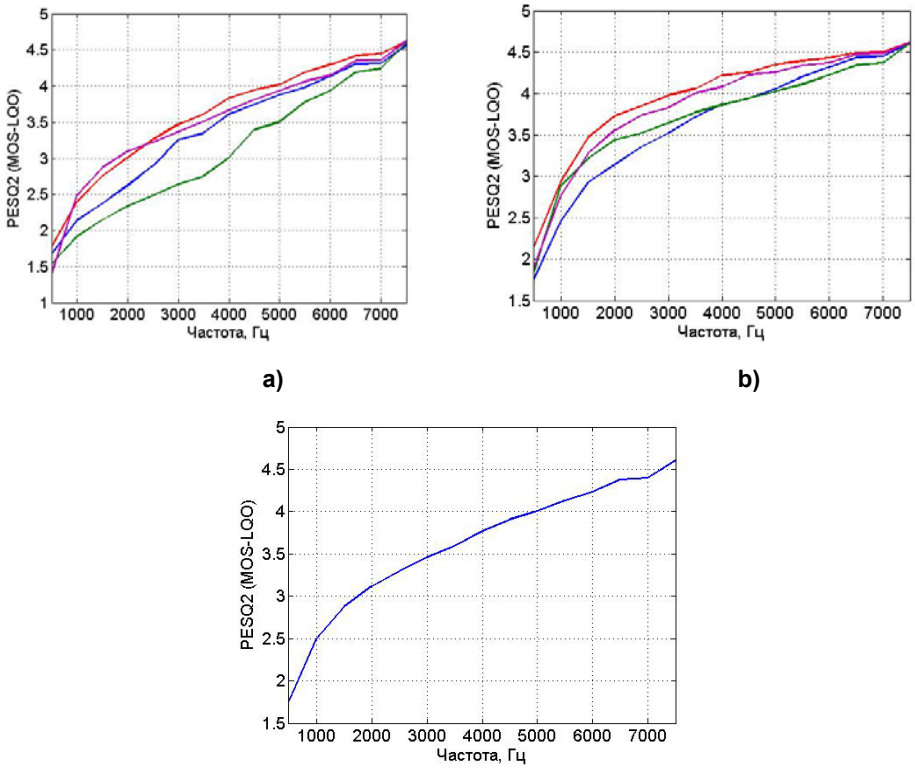


Fig. 4. PESQ index: female (a), male (b), averaged (c)



Fif. 5. PESQ-2 index: female (a), male (b), averaged (c)

As it can be seen, monotonic behaviour of PESQ and BSD indicators say in their favour. However, these graphs indicate that PESQ is designed for narrowband telephony (although the calculations used PESQ signals sampled at 16 kHz). PESQ-2 is free of this drawback and allows analyzing the quality of speech signals transmitted in a narrow and in a wide band (see Fig. 5).

However, as follows from Fig. 5, PESQ-2 abilities are not sufficient for a final verdict as to the potential ability of objective indicators to assess the speech band limited signal quality, and to assess the real needs of the human auditory system to speech perception. It is necessary to use indicator POLQA for this purpose. But the measurements of POLQA, unfortunately, are only feasible on a commercial basis today.

Conclusions

When evaluating the speech band-limited signal quality, BSD and PESQ are the most informative indicators among examined ones in this paper. It should be noted that estimation results depend strongly on the choice of the estimation algorithm version when using the PESQ index.

Analysis of the BSD dependence on the speech signal bandwidth showed that increasing in the quality of the speech signal stops when band width reaching of 9-10 kHz. It is advisable to check the validity of this result with the POLQA usage in the future.

References

1. *Beerends J., Wijngaarden S., Buuren R.* “Extension of ITU-T Recommendation P.862 PESQ towards Measuring Speech Intelligibility with Vocoders. New Directions for Improving Audio Effectiveness”. Meeting Proceedings RTO-MP-HFM-123, Paper 10, P.10-1–10-6. Neuilly-sur-Seine, France: RTO. [Online]. Available: http://www.rto.nato.int/abstracts.aps
2. *Blauert J., ed.* (2005), “Communication acoustics”. Springer-Verlag Berlin Heidelberg, P. 385 p.
3. *Cote N.* (2011), “Integral and diagnostic intrusive prediction of speech”. Springer-Verlag Berlin Heidelberg, P. 267.
4. *Ellis D.* PLP and RASTA in Matlab. [Online]. Available: http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/
5. *Hermansky H.* (1990), “Perceptual Linear Prediction (PLP) analysis of speech”. J. Acoust. Soc. America. Vol. 87. Pp. 1738-1753.
6. *Loizou P.* “Matlab Software. PESQ and other objective measures for evaluating quality of speech”. [Online]. Available: http://ecs.utdallas.edu/loizou/speech/software.htm
7. *Moller S.* (2005), “Quality of Telephone-Based Spoken Dialogue Systems”. Springer Science + Business Media, Inc., P. 490 p.
8. *Naylor P., Gaubitch N.* (2010), “Speech Dereverberation”. Springer, P. 399.
9. Next-Generation (3G/4G) Voice Quality Testing with POLQA®. White Paper. Rohde & Schwarz, 2012. P. 22.
10. Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 October 2005.
11. Perceptual Objective Listening Quality Assessment (POLQA) ITU-T Recommendations P.863. January 2011.
12. *Prodeus A.* (2014), “PESQ Matlab Driver. MathWorks”. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/47333-pesq-matlab-driver
13. *Raake A.* (2006), “Speech Quality of VoIP. Assessment and Prediction”. John Wiley, P. 338.
14. Recommendation P.862. Amendment 2 (11/05), 2011. [Online]. Available: http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en

Поступила в редакцию 20 августа 2014 г.

УДК 004.934

Н.В. Богданова, канд.техн.наук, **А.М. Продеус**, д.-р.техн.наук
Національний технічний університет України «Київський політехнічний інститут»,
вул. Політехнічна 16, 03056, Київ, Україна.

Об'єктивне оцінювання якості мовленнєвих сигналів, обмежених смугою частот

Експериментально отримані залежності об'єктивних оцінок якості мовленнєвого сигналу від смуги частот, що займає сигнал. У рамках даної задач виконано співставлення розглянутих показників якості мовленнєвого сигналу. Показано, що прості в обчислювальному відношенні по-

казники у вигляді сегментного відношення сигнал-шум (SSNR) і логарифмічно-спектральних спотворень (LSD) можуть неадекватно реагувати на зміну смуги частот. Значно коректніше поводяться більше складні в обчислювальному плані перцептуальні показники, такі як барк-спектральні спотворення (BSD) й перцептуальна оцінка якості мовлення (PESQ), що дозволяє, в остаточному підсумку, уточнити реальні потреби слухової системи людини до сприйняття мовлення. Бібл.14, рис. 5.

Ключові слова: *смуга частот сигналу, якість мовленнєвого сигналу, показники якості.*

.

УДК 004.934

Н.В. Богданова, канд.техн.наук, **А.Н. Продеус**, д.-р.техн.наук
Национальный технический университет Украины «Киевский политехнический институт»,
ул. Политехническая 16, 03056, Киев, Украина.

Объективное оценивание качества речевых сигналов, ограниченных по полосе частот

Экспериментально получены зависимости объективных оценок качества речевого сигнала от полосы частот, занимаемой сигналом. В рамках данной задачи произведено сопоставление рассмотренных показателей качества речевого сигнала. Показано, что простые в вычислительном отношении показатели в виде сегментного отношения сигнал-шум (SSNR) и логарифмически-спектральных искажений (LSD) могут неадекватно реагировать на изменение полосы частот. Значительно корректнее ведут себя более сложные в вычислительном плане перцептуальные показатели, такие как барк-спектральные искажения (BSD) и перцептуальная оценка качества речи (PESQ), что позволяет, в конечном счете, уточнить реальные требования слуховой системы человека к восприятию речи. Библ. 14 , рис. 5.

Ключевые слова: *полоса частот сигнала, качество речевого сигнала, показатели качества.*

Список использованных источников

1. *Beerends J., Wijngaarden S., Buuren R.* Extension of ITU-T Recommendation P.862 PESQ towards Measuring Speech Intelligibility with Vocoders. New Directions for Improving Audio Effectiveness // Meeting Proceedings RTO-MP-HFM-123, Paper 10, P.10-1–10-6. Neuilly-sur-Seine, France: RTO. [Online]. Available: http://www.rto.nato.int/abstracts.aps
2. *Blauert J., ed.* Communication acoustics. – Springer-Verlag Berlin Heidelberg, 2005. – 385 p.
3. *Cote N.* Integral and diagnostic intrusive prediction of speech - Springer-Verlag Berlin Heidelberg, 2011. – 267 p.
4. *Ellis D.* PLP and RASTA in Matlab // [Online]. Available: http://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/
5. *Hermansky H.* Perceptual Linear Prediction (PLP) analysis of speech // J. Acoust. Soc. America. – 1990. – Vol. 87. – P. 1738-1753.
6. *Loizou P.* Matlab Software. PESQ and other objective measures for evaluating quality of speech // [Online]. Available: http://ecs.utdallas.edu/loizou/speech/software.htm
7. *Moller S.* Quality of Telephone-Based Spoken Dialogue Systems – Springer Science + Business Media, Inc., 2005. – 490 p.
8. *Naylor P.,* Gaubitch N. Speech Dereverberation. – Springer, 2010. – 399 p.
9. Next-Generation (3G/4G) Voice Quality Testing with POLQA®. White Paper. – Rohde & Schwarz, 2012. – 22 p.
10. Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 – October 2005.
11. Perceptual Objective Listening Quality Assessment (POLQA) ITU-T Recommendations P.863 – January 2011.
12. *Prodeus A.* PESQ Matlab Driver // MathWorks, 2014. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/47333-pesq-matlab-driver
13. *Raake A.* Speech Quality of VoIP. Assessment and Prediction. – John Wiley, 2006. - 338 p.
14. Recommendation P.862. Amendment 2 (11/05), 2011. [Online]. Available: http://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en