AUDIO SIGNALS CLIPPING DETECTION USING KURTOSIS AND ITS TRANSFORMS

Article *in* International Journal of Computing · September 2020

DOI: 10.47839/ijc.19.3.1890

Arkadiy Prodeus
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Maryna Didkovska

computing@computingonline.net
www.computingonline.net

Print ISSN 1727-6209
On-line ISSN 2312-5381
International Journal of Computing

AUDIO SIGNALS CLIPPING DETECTION USING KURTOSIS AND ITS TRANSFORMS

Arkadiy Prodeus [1], Maryna Didkovska [2]

[1] Department of Acoustics and Acoustoelectronics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, aprodeus@gmail.com
[2] Department of Mathematical Methods of System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, maryna.didkovska@gmail.com

Paper history:
Received 10 February 2020
Received in revised form 11 May 2020
Accepted 08 July 2020
Available online 27 September 2020

Keywords:
speech clipping;
music clipping;
clipping detection;
clipping value measure;
kurtosis.

Abstract: This paper compares the results of subjective and objective assessments of the quality of speech and music signals distorted during clipping when large instantaneous signal values are replaced by a certain threshold constant or by values close to it. It was proposed in recent works to use kurtosis and some of its simple functional transforms such as reciprocal of kurtosis and square root of reciprocal of kurtosis as objective (instrumental) clipping value measures. This paper clarifies the results of a subjective assessment of the quality of speech and music signals distorted by clipping. A comparison of the obtained estimates allows one to conclude that the human auditory system is slightly more sensitive to the clipping of musical signals than to the clipping of speech signals, but this difference is small. Similarly, objective quality measures of clipped signals are almost equally sensitive to the clipping value of speech and music signals. An analysis of the variability of the kurtosis estimates, depending on the time of estimation, showed that the relative standard deviation of the kurtosis estimates is close to 10% for the analysis time interval of 1–40 s.

1. INTRODUCTION

Full use of the dynamic range when speech or music signals are transmitted or recorded is highly desirable, since it allows minimizing effects of background noise. However such mode involves risk of nonlinear signal distortion due to clipping, when large instantaneous signal values $x(n)$ are replaced by a certain threshold constant:

$$y(n) = \begin{cases} x(n), & |x| < A, \\ A \cdot \text{sign}[x(n)], & |x| \geq A, \end{cases} \qquad (1)$$

where n is signal sample number, A is the clipping threshold $(0 < A < C = max\,|\,x(n)|)$, sign$(\cdot)$ is sign function, and $|\cdot|$ is the modulus sign.

To minimize signal distortion caused by clipping, automatic gain control (AGC) systems are commonly built into the transmission and recording paths of audio signals. Clipping detection subsystems are important parts of such AGC systems [1].

A small clipping value is accompanied by quite small non-linear distortions of the signals that rarely cause a negative reaction from the audience. Therefore, it seems reasonable to construct a clipping detection algorithm such that the decision on presence or absence of clipping perceived by the listeners was preceded by an assessment of the clipping value.

A number of known methods for clipping detection is based on exactly this approach, and in most cases, it is proposed to use a degree of difference in the shape or parameters of the probability density function (PDF) between analyzed and undistorted signals as a measure of clipping value [1–7].

In particular, the US patent [1] discloses embodiments of clipping detection method based on analysis of the shape of preliminary PDF estimate for an analyzed signal.

On the contrary, the Russian patent [2] proposes to detect clipping using evaluated PDF parameters such as variance, mean square deviation, half-period average value, and average number of outliers. The most serious drawback of this method that prevents its mass implementation is the use of unnormalized parameters.

This drawback was eliminated when the signal-to-noise ratio as a measure of the clipping value was proposed to use [3]. In this publication, undistorted instantaneous values of audio signal are implied as the ‘signal’ while the audio signal values beyond the acceptable limits are implied as the ‘noise’. Since instantaneous values of such ‘noise’ are unknown, it is proposed to estimate its power by extrapolated PDF tails of the analyzed signal. However, this method has another obvious drawback, which is its enormous computational complexity.

A ‘clipping coefficient’ was proposed in [4] as parameter for making a decision about clipping:

$$R_{cl} = 2 \cdot max(D_l, D_r)/D$$

where D_l and D_r are distances between left and right outermost outliers and central peak of PDF, D is difference between maximum and minimum undistorted signal values. However, it was subsequently noted that the clipping coefficient is insufficiently reliable when using for preliminary estimating the clipping value, although it is suitable for clipping detection [5].

Methods for detecting clipping proposed in [6, 7] consist in use of rough (20 bins) or detailed (6000 bins) histograms. The mutual disadvantage of these methods is the lack of normalization of the histogram, which makes it difficult to use the proposed methods when changing the signal parameters and the histogram constructing algorithm parameters.

None of above-mentioned publications has considered normalized fourth-order moment known as kurtosis [8]

$$\beta_4 = \frac{\mu_4}{(\mu_2)^2} \qquad (2)$$

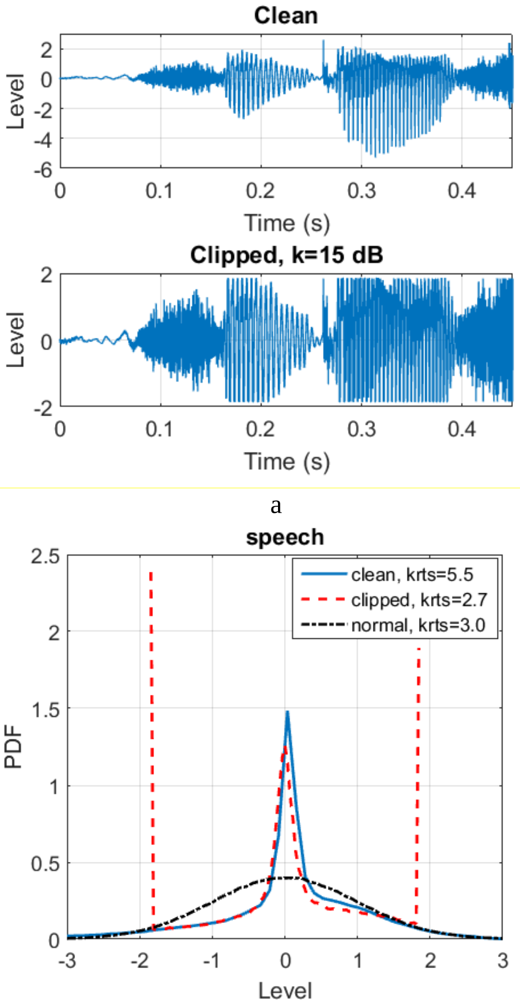
where μ_k is a central moment of the k-th order, or closely related coefficient of kurtosis $\varepsilon_4 = \beta_4 - 3$, as a possible clipping measure.

This gap was filled in [9] where usefulness of kurtosis and its transforms for speech signals clipping value assessment was shown. Similar conclusion for musical signals was made in [10]. Note that this utility does not consist in reducing the number of calculations (on the contrary, the amount of calculations grows by about half), but in obtaining a smooth and monotonous dependence of the objective quality measure on the sound signal clipping value, which allows one to more accurately assess the degree of degradation of the audio signal.

The present paper is aimed at comparing the quality estimates of clipped speech and music. Subjective estimates and objective ones based on kurtosis and its transforms are under consideration. The practical usefulness of such a comparison is the ability to adjust the transmission or recording channel to the type of signals that are more sensitive to non-linear distortion caused by clipping. Another object of the paper is to analyze the sensitivity of kurtosis and its transformations estimates to the estimation time interval and signal sample.

2. SOME FEATURES OF STUDIED PARAMETERS

Waveforms of clean and clipped speech signals are shown in Fig. 1a. As can be seen, clipping a signal leads to a significant change in its waveform.



a

b

Figure 1 – Clean and clipped speech signals: (a) waveform; (b) PDF estimates

The clipping value

$$k = 20\, lg(max\,|\,x(n)|/A) \qquad (3)$$

was taken equal to 15 dB in this case.

PDF estimates of the clean (solid line) and clipped (dashed line) speech signals, and Gaussian white noise (dash-dotted line), are shown in Fig. 1b. Here it can be seen that clipping the signal leads to the appearance of specific tails in the PDF plot.

Comparison between β_4 estimates in Fig. 1b indicates that clipping leads to decrease in β_4 values.

Measure β_4 values are theoretically unlimited from above and cannot be less than +1. β_4 values can reach 50 for real unclipped music signals and 12 for real unclipped speech signals, and parameter β_4 values close to 1 corresponds to heavily clipped signals [10].

Since the "fuzziness" of upper bound of measure β_4 is inconvenient in engineering applications, it was proposed in [10] to substitute β_4 with the quantities:

$$\gamma_4 = 1/\beta_4, \tag{4}$$

$$\eta_4 = 1/\sqrt{\beta_4}, \tag{5}$$

with possible values lying within the interval [0; 1] and values close to zero corresponding to unclipped signal. More detailed information on features of parameter (2) can be found in [11] and some known speech distributions have been tested as hypotheses for different genres of music in [12].

As can be seen, $\eta_4 = 1/\sqrt{\beta_4} = \mu_2/\sqrt{\mu_4}$ is signal variance normalized by the square root of the fourth-order central moment. Though the idea of using signal variance to detect clipping was initially proposed in [1], unfortunately, this idea was not developed up to a level sufficient for technical implementation, since nothing was said about the need to normalize the variance of the analyzed signal. Thus, measure (4) and the related measure (5) are devoid of this drawback.

Dependencies of parameters (4) and (5) on the clipping value (3) can be obtained for real speech and music signals, as well as the obtained dependences can be compared with the results of subjective quality assessment of clipped signals. Unfortunately, a comparison of the quality ratings of clipped speech and music has not been made until recently, so this drawback is eliminated in this paper.

3. EXPERIMENTAL SETUP

Speech signals were recorded in an anechoic room with reverberation time of 0.15 s at the signal-to-noise ratio of 38 dB. The same legal text was read by 8 speakers (4 men and 4 women) at a normal reading pace. All speech signals were digitalized at the sampling frequency of 22050 Hz and the bit depth of 16 bits.

Musical signals included fragments of 8 musical compositions with one half belonging to genre of popular music, and the other half belonging to genre of classical music. All musical signals were digitalized at the sampling frequency of 44100 Hz and the bit depth of 16 bits.

Duration of studied signal record fragments was from 15 to 20 seconds, which is sufficient for subjective and objective assessment of clipping value.

In order to simulate heavily clipped signals (1), the clipping value was varied using a non-negative parameter (3) which value $k = 0$ corresponds to the unclipped signal.

Subjective assessment of clipped signal quality was carried out by comparing of aural perception of distorted and clean signals and rating them using a 5-point Degradation Mean Opinion Score (DMOS) scale [13]. Percipients, aged 19 to 35, having no hearing impairments, scored 5 points if they did not perceive any distortion or 1 point if they perceived a heavily distorted and very annoying signal. The quality of speech signals was evaluated by 32 percipients, whereas quality of musical signals was evaluated by 36 percipients.

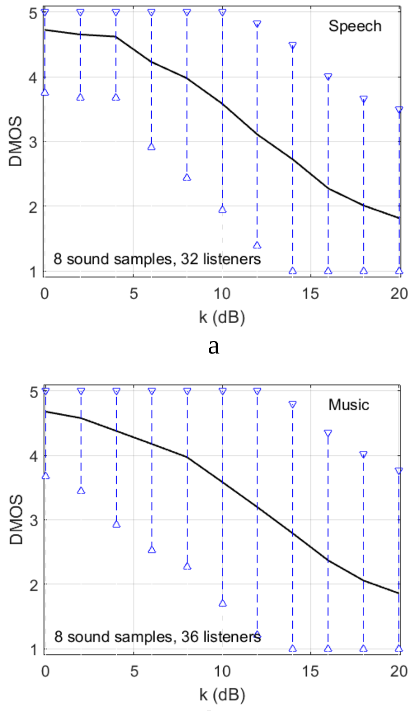
An unbiased estimate was used to calculate the β_4 value [9]. Estimates of parameters $\gamma_4 = 1/\beta_4$ and $\eta_4 = 1/\sqrt{\beta_4}$ were calculated taking into account (4) and (5).

4. EXPERIMENTAL RESULTS

4.1 SUBJECTIVE ASSESSMENT

Results of subjective assessment of clipped speech and music quality are shown in Fig. 2.

Averaged DMOS estimates both over listeners and speech (music) samples of signals are represented by solid lines, and 95% confidence intervals are indicated by segments of vertical dashed lines. It can be seen that quality of clipped signals remains subjectively high ($DMOS \geq 4.5$) at $k \leq 5$ dB for speech and at $k \leq 3.5$ dB for music. At $5 < k \leq 8$ dB for speech and at $3.5 < k \leq 8$ dB for music, quality of clipped signals may be considered subjectively good ($4 \leq DMOS < 4.5$). In the range 8 <k <20 dB, the $DMOS(k)$ dependences practically coincide. Summarizing the results presented above, we can conclude that the human auditory system is slightly more sensitive to the clipping of musical signals than to the clipping of speech signals, but this difference is small. In the future, it will be useful to compare these results with ones of [14, 19-21] and other papers in order to find out how common the identified phenomenon is.



a

b

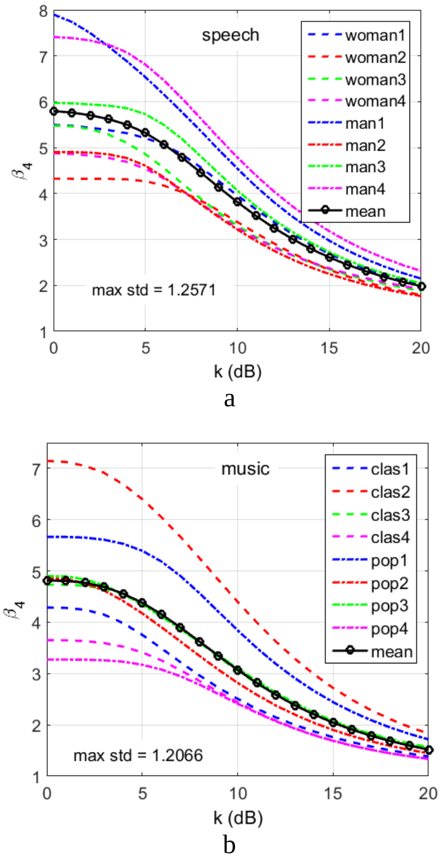
Figure 2 – $\boldsymbol{DMOS}$ versus $\boldsymbol{k}$: (a) speech; (b) music

4.2 OBJECTIVE ASSESSMENT

Estimates of β_4, $\gamma_4 = 1/\beta_4$, and $\eta_4 = 1/\sqrt{\beta_4}$ in the form of dependences $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$, and $\overline{\eta_4}(k)$ averaged over listeners are presented in Figures 3, 4, and 5, respectively. The result of additional averaging over the signal samples is shown in these figures by a bold line with circles.

As can be seen, the dependences $\overline{\beta_4}(k)$ and $\overline{\gamma_4}(k)$ only slightly vary in interval $0 \le k \le 5$dB, that is, at low clipping values, where quality of speech and music stays subjectively high. Meanwhile, in the most interesting for practical use interval $5 < k \le 15$ dB, where speech quality subjectively drops from 4.5 points to 2 points in the DMOS scale, dependences $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$, and $\overline{\eta_4}(k)$ vary with a quite considerable and almost constant rate. This means that parameters β_4, $\gamma_4 = 1/\beta_4$, and $\eta_4 = 1/\sqrt{\beta_4}$ are good as clipping value measures.

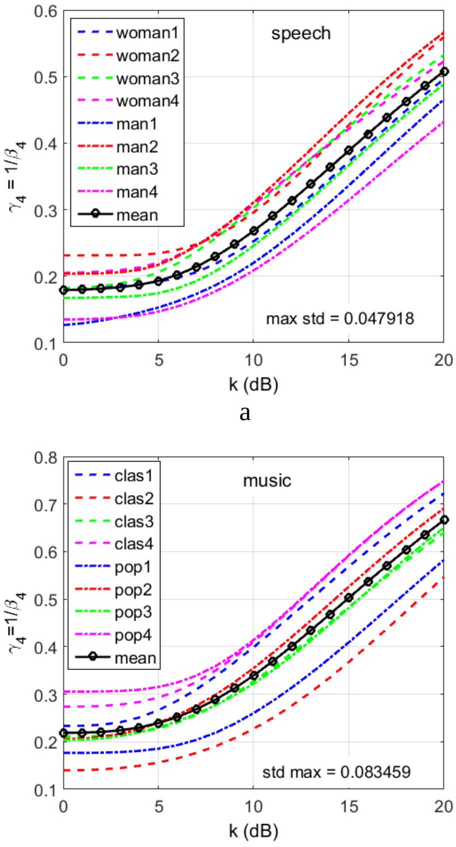
The next interesting question is: how sensitive are the objective measures mentioned above to the difference between speech and music? This question is more difficult to answer, since, as can be seen, the average values of these parameters are different for undistorted speech and undistorted music. To solve this problem for the β_4 parameter, one can calculate the ratio of $\beta_4(k)/\beta_4(0)$ for specific value of k parameter. For example, we have almost equal values for speech and music, near 0.63-0.66, for $k = 10$ dB. Similarly, we have values 1.5-1.53 for the parameter γ_4 and 1.23-1-25 for the parameter η_4.



a

b

Figure 3 – $\overline{\boldsymbol{\beta_4}}$ versus $\boldsymbol{k}$: (a) speech [11]; (b) music [10]



a

b

Figure 4 – $\overline{\boldsymbol{\gamma_4}}$ versus $\boldsymbol{k}$: (a) speech [11]; (b) music [10]

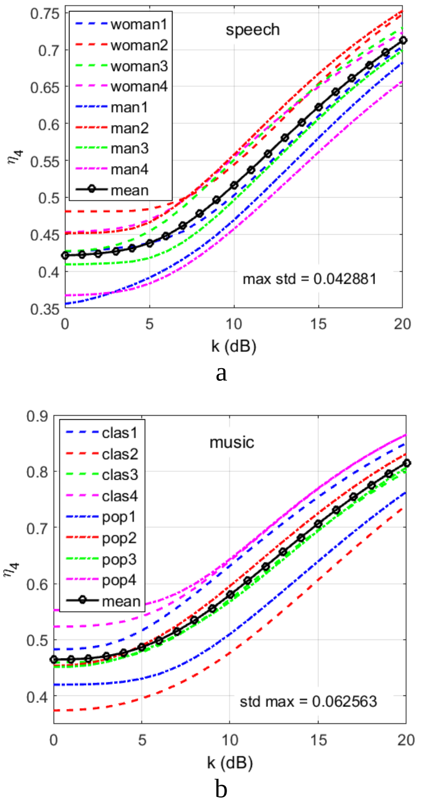


Figure 5 – $\overline{\eta_4}$ versus k: (a) speech [11]; (b) music

Thus, we can conclude that studied objective measures β_4, $\gamma_4 = 1/\beta_4$, and $\eta_4 = 1/\sqrt{\beta_4}$ are practically insensitive to a kind of acoustic signal. Note that a similar situation was previously discovered in studies of phase distortion of speech and music signals [14].

5. ESTIMATES VARIABILITY

As was noticed in section 3, the length of the analyzed segments of acoustic signals was 15–20 s. In this case, at least two questions inevitably arise. Firstly, how correct are these actions, given that speech and musical signals are not stationary random processes. Secondly, the problem of the statistical stability of the kurtosis estimate (and related parameters) to changing the length of the analyzed signal segment is of undoubted interest.

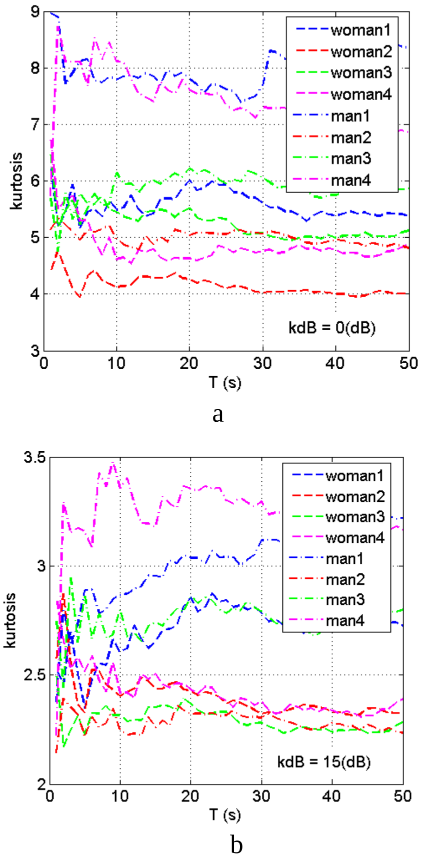
Some answers to these questions can be found in [15-18]. The first attempts to estimate kurtosis for the processes at the outputs of a set of narrow-band filters are described in [15, 16]. Arctic under-ice ambient noise was analyzed in the papers and the frequency dependence of kurtosis coefficients was called "frequency domain kurtosis" (FDK). In [17], for such a set of kurtosis coefficients, the other term "spectral kurtosis" (SK) was used and examples of the analysis of artificial (additive mixture of stationary Gaussian noise and several harmonic signals with constant and variable parameters) and real (noise of a rotating mechanism) signals are given. The examples demonstrate the usefulness of SK to identify both non-Gaussianity and non-stationarity of the analyzed processes.

An obvious drawback of the aforementioned papers is the lack of substantiation of the correctness of kurtosis measurements in the case of non-stationary random signals. This gap was filled in [18], where paradigm of conditionally non-stationary (CNS) processes was proposed. It was shown that for CNS processes, which, in particular, include speech and music signals, estimating kurtosis as a measure of non-Gaussianity generated by non-stationarity is quite correct.

Another important issue is the choice of the time interval at which the statistical stability of kurtosis estimates is ensured. In the experiments described in [15], the SK and other parameters were measured with 1, 0.5, 0.17, and 0.1 s segments duration at the output of a short-time Fourier transform (STFT) with processing times from 2 to 14 minutes. High importance of the segments duration choice in STFT-based SK estimation was pointed out in [18]. If segments duration is too small, it causes excessive bias of kurtosis estimate. On the other hand, if segments length value is too large, the SK tends to Gaussian process values in accordance with the central limit theorem.

In our studies, SK is not evaluated, but a "classical" kurtosis in the time domain is estimated, since clipping the signal leads to distortion of almost all frequency components of the signal. Thus, there is no need for segmentation of the analyzed process. Nevertheless, the question remains how sensitive the kurtosis estimates are to the choice of the length of the segment of the analyzed signal.

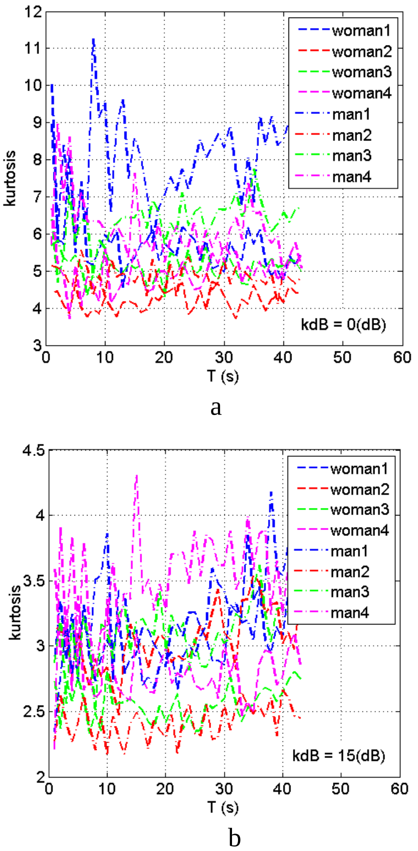
To study this problem, two experiments were performed using 8 records of speech signals lasting 55-60 s, mentioned in section 3. In both experiments, the duration T of the analyzed signal segments varied from 1 s to 50 s with a step of 1 s. In the first experiment, the results of which are presented in Fig. 6, all segments are started at time t=0, i.e., the signals in these segments were statistically dependent. In the second experiment (Fig. 7), the beginning of each subsequent segment coincided with the end of the previous segment. As a result, the signals in different segments were statistically independent of each other. As can be seen in Fig. 6, studying of statistical dependent segments is useful because makes more evident the strong influence of speaker's voice peculiar properties on the kurtosis value. At the same time, the graphs in Fig. 7 show that the relative standard deviation of kurtosis estimates varies little over the analysis time interval of 1–40 s and amounts to about 10%.



a

b

Figure 6 – Kurtosis versus T for dependent signals: (a) $k = 0$ dB; (b) $k = 15$ dB



a

b

Figure 7 – Kurtosis versus T for independent signals: (a) $k = 0$ dB; (b) $k = 15$ dB

It was recommended in [5] to measure the clipping coefficient on signal segments with a length of about 0.5-1 s. We can assume that the behavior of the graphs in Fig. 7 is in good agreement with this proposition, since a further increase in the analysis duration does not lead to a noticeable increase in the accuracy of kurtosis estimation. In the future, it will be useful to compare these results with the ones of [19-21] and other papers.

6. CONCLUSION

Subjective assessment of the quality of clipped speech and music signals showed that the human auditory system is slightly more sensitive to the clipping of musical signals than to the clipping of speech signals, but this difference is small. Similarly, considered in this paper objective measures of clipping value are almost equally sensitive to distortions of both speech and musical signals.

When implementing objective measures in real clipping detection algorithms, the length of the analyzed signal segment can be chosen close to 1 s, since the relative standard deviation of the kurtosis estimates is about 10% and changes little with increasing analysis time interval.

7. REFERENCES

[1] T. Otani, M. Tanaka, Y. Ota, S. Ito, *Clipping detection device and method*, Patent US 8,392,199 B2, Int. Cl. G10 19/00, 2013.

[2] G.R. Avanesyan, *Method and device for estimating and indicating distortions of output signal of audio frequency amplifiers (overload indication)*, Patent RU 2274868 C2, Int. Cl. G01R 23/20, G01R 19/165, 2006.

[3] X. Liu, J. Jia, L. Cai, “SNR estimation for clipped audio based on amplitude distribution,” *Proc. of the IEEE 9th Int. Conf. on Natural Computation (ICNC)*, Shenyang, China, 23-25 July, 2013, pp. 1434-1438.

[4] S.V. Aleinik, Yu.N. Matveev, and A.N. Rayev, “Evaluation method of speech signal clipping level,” *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, no. 3 (79), pp. 79–83, 2012.

[5] S.V. Aleinik, Yu.N. Matveev, and A.V. Sholokhov, “Detection of clipped fragments in acoustic signals,” *International Journal of Computer and Information Engineering*, vol. 8, no. 2, pp. 286-292, 2014.

[6] F. Bie, D. Wang, J. Wang, T.F. Zheng “Detection and reconstruction of clipped speech for speaker recognition,” *Speech Communication*, vol. 72, pp. 218-231, September 2015.

[7] C. Laguna, A. Lerch, "An efficient algorithm for clipping detection and declipping audio," *Proc. of the AES 141st Convention*, September 29-October 2, Los Angeles, USA, 2016, 10 p.

[8] M. Kendall, A. Stuart, *The Advanced Theory of Statistics: Distribution theory*, London, Wiley, 1977, 700 p.

[9] A. Prodeus, I. Kotvytskyi, A. Ditiashov, "Assessment of clipped speech quality," *Electronics and Control Systems*, no. 4(58), pp. 11-18, 2018.

[10] A. Prodeus, I. Kotvytskyi, A. Grebin, "Using kurtosis for objective assessment of the musical signals clipping degree," *Proc. of the 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)*, October 2019, Kyiv, Ukraine, pp. 655-659.

[11] J. Moors, "The Meaning of Kurtosis: Darlington Reexamined," *The American Statistician*, 40:4, pp. 283-284, 1986.

[12] V. Arora and R. Kumar, "Probability distribution estimation of music signals in time and frequency," *Proc. of the IEEE 19th Int. Conf. on Digital Signal Processing (DSP-2014)*, August 2014, Hong Kong, pp. 409-414.

[13] N. Cote, *Integral and diagnostic intrusive prediction of speech*, Springer-Verlag: Berlin-Heidelberg, 2011, 250 p.

[14] A. Poorjamm J. Jensen, M. Little, M. Christensen, "Dominant Distortion Classification for Pre-Processing of Vowels in Remote Biomedical Voice Analysis," *INTERSPEECH 2017*, August 20-24, Stockholm, Sweden, 2017.

[15] R. Dwyer, "FRAM II single channel ambient noise statistics," *Proc. of the 101st Meet. Acoustical Soc. of America*, May 19, 1981, Published in *NUSC Tech. Doc. 6588*, 25 November 1981.

[16] R. Dwyer, "Use of the kurtosis statistic in the frequency domain as an aid in detecting random signals," *IEEE Journal of Oceanic Engineering*, vol. OE-9, no. 2, pp. 85-92, April 1984.

[17] V. Vrabie, P. Granjon, and C. Servière, "Spectral kurtosis: from definition to application," *Proc. of the 6th IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing*, Grado, Italy, 2003, pp. 1-5.

[18] J. Antoni, "The spectral kurtosis: a useful tool for characterizing non-stationary signals," *Mech. Syst. Signal Pr.*, vol. 20, pp. 282–307, 2006.

[19] S. Naida, "Acoustic theory problems of speech production in the light of the discovery of the formula for the middle ear norm parameter," *Proc. of IEEE 35th Int. Sc. Conf. on Electronics and Nanotechnology (ELNANO)*, 21-24 April, Kyiv, Ukraine, 2015, pp. 347-350.

[20] S. Naida, N. Naida, V. Didkovskyi, O. Pavlenko, "Spectral Analysis of Sounds by Acoustic Hearing Analyzer," *Proceedings of IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO)*, April 16-18, Kyiv, Ukraine, 2019, pp. 421-424.

[21] S. Lunova, O. Pedchenko, I. Rudenko, "Speech spectrum of the Ukrainian language," *Proc. of the IEEE 39th Int. Conf. on Electronics and Nanotechnology (ELNANO)*, April 16-18, Kyiv, Ukraine, 2019, pp. 444-448.

Arkadiy Prodeus, *received the BSc, MSc, PhD, and DSc degrees in electrical engineering from the Kyiv Polytechnic Institute (KPI), Kyiv, Ukraine, in 1970, 1972, 1982, and 2012, respectively. Now he is professor at the Acoustics and Acoustoelectronics Department, NTUU "Igor Sikorsky KPI". His current interests include digital signal processing, modeling and simulation, pattern recognition, image processing.*

Maryna Didkovska, *received the BSc, MSc, and PhD degrees in electrical engineering from the Kyiv Polytechnic Institute (KPI), Kyiv, Ukraine, in 1970, 1972, and 1982, respectively. Now she is senior lecturer at the Department of Mathematical Methods of System Analysis, NTUU "Igor Sikorsky KPI". Her current interests include software reliability, project management, artificial intelligence, digital signal processing.*