UDC 004.934:621.391.8 **DOI:** 10.31673/2412-9070.2026.023601

P. BRYSIN, PhD Student;
ORCID: 0000-0003-2598-3907
V. LUKIN, DSc (Engin), Prof,
ORCID: 0000-0002-1443-9685
National Aerospace University "Kharkiv Aviation Institute", Kharkiv

NEURAL-BASED LOSSY COMPRESSION OF NOISY AUDIO SIGNALS AND THEIR DCT-BASED POST-FILTERING

A recently proposed method of lossy compression of one-dimensional audio signals is considered with application to musical and speech signals corrupted by additive white Gaussian noise. It is shown that excellent compression ratios exceeding 100 are obtained. Properties of decompressed signals and introduced distortions depend on input signal-to-noise ratio (SNR). For medium ratios, lossy compression is able to partly suppress noise, whilst, for low input signal-to-noise ratios, noise is left. Then, it can be suppressed after decompression by a filter based on discrete cosine transform able to provide improvement by up to 10 dB. Spectral analysis of distortions is carried out and it is demonstrated that the largest distortions are observed for low frequencies.

Keywords: audio signal, noise, lossy compression, DCT-based filtering, distortions.

Introduction

Many multimedia and telecommunication applications deal with compression, transferring, and processing of one-dimensional (1-D) signals such as music and speech [1], [2]. Although there exist several transform based methods for audio compression [3], [4], a new generation of audio compression methods based on trained neural networks is under interest nowadays [2], [5], [6] demonstrating impressive results.

In most considered cases, it is supposed that audio signals to be compressed have high quality. Meanwhile, signals subject to compression and further transferring can be imperfect, e.g., degraded by noise [7], [8]. Therefore, several questions arise: 1) what happens to noisy audio signals if they are compressed in a lossy manner by modern neural-based methods? 2) is it worth denoising such signals after transferring and decompression? 3) what is efficiency of such post-filtering?

Problem statement

In this paper, we focus on neural-based technique of lossy compression proposed and considered in [5]. The TSAC codec [9], built using this technology with an additional Transformer model, signifycantly outperforms many existing audio compression methods in terms of its parameters. Our first goal is to analyze compression ratio (CR) values for signals corrupted by noise of different level (for a wide set of input SNRs) and to study distortions introduced into compressed signals.

Another goal of this paper deals with noise suppression after decompression. As known, lossy compression might have specific noise filtering effect [10], [11]. Then, noise can be partly removed and, besides, its statistics can change. This can lead to problems in finding an appropriate filter for further noise removal and setting its parameters. Therefore, the second goal is to determine is it possible to get certain improvement of decompressed signal by post-filtering, when post-filtering use is expedient and what can be its efficiency.

Current approaches to audio compression and filtering

The orthogonal-based methods of audio data lossy compression employ transform coefficient quantization with quite many coefficients assigning zero values after quantization with further use of this property by exploiting run-length encoding or other lossless compression techniques [1], [12].

They are able produce CR considerably larger than lossless compression techniques by the expense of compressed signal quality where the general tendency is that a larger CR (a smaller bitrate) is associated with worse quality. Then, CR should be restricted to provide appropriate quality.

Recently, neural audio codecs have appeared as alternatives to conventional codecs [5], [6] and they are based on the creation of audio autoregressive generative models [13], [14]. In particular, in [5, 9], the authors combine recent advances in high-fidelity audio generation and vector quantization with improved adversarial and reconstruction losses. This allows obtaining CR about 100 and larger while keeping high quality of compressed audio of different types. An obvious advantage of this approach is that the authors offer aforementioned open-source code and already trained model weights that allows additional testing and studying the properties of this codec.

For audio contaminated by noise, its removal, if necessary, can be carried out before compression and after decompression. Note that numerous denoising techniques have been developed so far [15], [16]. There are methods based on Kalman filtering [7], wavelets [15], LMS adaptive filtering [16], discrete cosine transform (DCT) [17], and convolutional neural networks [18]. Below, we concentrate on considering DCT-based techniques keeping in mind their computational efficiency [19] and ability to adapt to noise characteristics assuming that they are either known in advance or accurately pre-estimated [20].

General signal/noise model and preliminary analysis of compression characteristics

We consider conventional model of a signal contaminated by noise as:

$$S_n(i) = S(i) + n(i), \qquad i = 1, \ldots, I, \tag{1}$$

where {S(i), i=1,...,I} is the noise-free signal, i denotes the sample index, I is the total number of registered samples, {n(i), i=1,...,I} denotes the zero mean additive white Gaussian noise (AWGN) with variance σ^2.

If lossy compression is applied, one can be interested in two kinds of metrics for the case of noisy signal/image compression. A metric can be measured between original noisy and compressed data as well as between compressed and noise-free (true) data if one deals with simulations where noise is artificially added to noise-free signal and then lossy compression follows. Below, we concentrate on the latter type of metrics since it is desired to have a decompressed signal as close to the noise-free one as possible [11].

In our experiments, we used three musical fragments corresponding to classical (file W1), pop (file W2) and rock (file W3) music from a set of files for conducting public multiformat listening test [9]. The sampling frequency is equal to 44.1 kHz and the fragment duration is 5 s for all three fragments. Noise has been generated in such a manner that five fixed input SNR equal to 0, 5, 10, 15, and 20 dB have been provided. All compression and decoding experiments with the codec were performed using the Transform model (the -f parameter was disabled) and setting the -q parameter equal to 9 to obtain the maximum compression ratio (minimum bit rate). The obtained CR values defined as the ratio of the file sizes before and after compression are given in table 1.

Table 1

CR values for three test signals corrupted by AWGN of different intensity

Test musical fragment	Input SNR, dB					
	0	5	10	15	20	No noise
Classical	129.0	137.2	148.1	154.8	158.8	155.6
Pop	111.0	114.4	116.4	118.4	119.9	120.3
Rock	119.9	125.4	129.3	130.9	130.2	124.8

The following tendencies and observations can be mentioned. First, all CR values are larger than 100, i.e. they are very high. Second, CR for classical music is slightly larger than for pop and rock musical fragments. Third, CR values for clean signals are about the same as for input SNR equal to 20 dB where there is the tendency of CR increasing for input SNR growing from 0 to 20 dB.

Availability of {S(i), i=1,...,I}, {S_n(i), i=1,...,I}, and compressed {S_c(i), i=1,...,I} signals as well as an opportunity to determine Fourier spectra for realizations of {n(i), i=1,...,I} and {n_c(i)= S_c(i)-S(i), i=1,...,I} denoted as $F_n(f)$ and $F_c(f)$ (f is frequency where the maximal frequency equals to 22.05 kHz), respectively, allows analyzing noise suppression and introduced distortions (in aggregate) in spectral domain. Consider now the plots:

$$\Delta F(f) = 10log_{10}(|F_n(f)| - |F_c(f)|) \quad (2)$$

Such a plot for input SNR equal to 20 dB is shown in fig. 1. A small noise reduction takes place for frequencies from 5000 Hz to 21000 Hz. A larger reduction takes place for f>21 000 Hz. Meanwhile, due to introduced distortions, ΔF(f) are mainly smaller than zero for f<5000 Hz. If input SNR is smaller, the situation changes. An example of ΔF(f) for input SNR equal to 0 dB for the same musical fragment is given in fig. 2. Some noise suppression takes place for f from about 1000 Hz till 17000 Hz and, especially, for f>21 kHz. However, there are distortions for f<1 kHz.

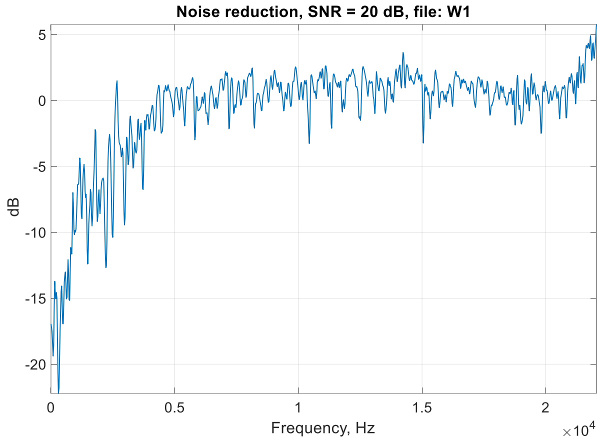

Fig. 1. ΔF(f) for input SNR equal to 20 dB for the classical music fragment

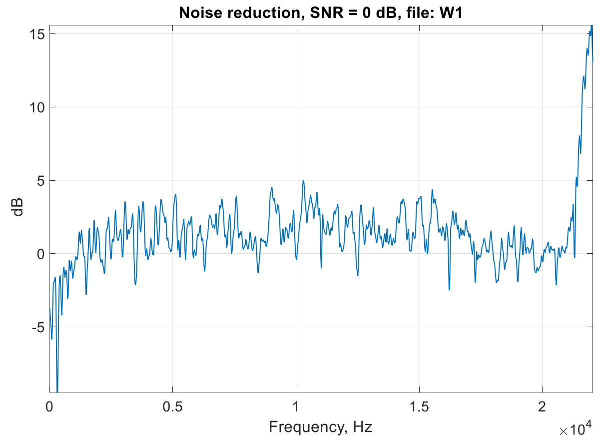

Fig. 2. ΔF(f) for input SNR equal to 0 dB for the classical music fragment

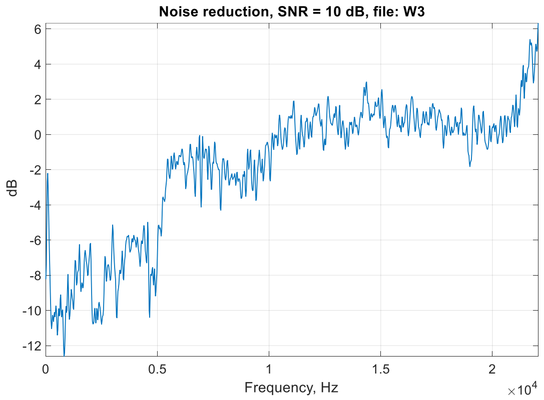

Fig. 3. ΔF(f) for input SNR equal to 10 dB for the rock music fragment

The general tendency that ΔF(f) increases if f grows take place for other test signals (see data in fig. 3). However, the distortions in low frequencies are larger than for classical music. Thus, it is possible to state that distortions introduced by the codec [9] act in a specific way for noisy signals leading to certain noise suppression for higher frequencies and deterioration of signal content for low frequencies.

It is also possible to control powers of the noise {n(i), i=1,...,I} and {n_c(i), i=1,...,I} denoted as P_n and P_d, respectively. Let us calculate $\Delta P=10\log_{10}(P_n - P_d)$, dB. The obtained data are given in table 2.

Table 2

ΔP values for three test signals corrupted by AWGN of different intensity

Test musical fragment	Input SNR, dB				
	0	5	10	15	20
Classical	0.79	0.47	-0.39	-2.03	-4.97
Pop	0.63	-0.64	-3.03	-6.45	-10.1
Rock	0.31	-1.14	-3.97	-7.01	-12.7

Analysis shows that even some improvement is possible for low input SNR (the positive effect of denoising is larger than the negative effect of introduced distortions) but for larger input SNR the introduced distortions prevail, especially for rock musical fragment.

Efficiency analysis for post-filtering

Let us briefly recall the main principles of DCT-based denoising. It is carried out using fixed size blocks where, to provide high computational efficiency, the block size is usually set equal to power of two. This can be 16, 32, or 64 allowing to benefit from algorithms of fast DCT. Denoising, for each l-th block, consists of the following three stages: 1) direct DCT is carried out for block with obtaining DCT coefficients $\{D_l(k), k=1,...,N\}$ where N here is the block size; 2) thresholding is then applied resulting in obtaining the modified coefficients $\{D_{l\,thr}(k), k=1,...,N\}$ (see details below); 3) inverse DCT is performed for $\{D_{l\,thr}(k), k=1,...,N\}$ producing signal values $S_{f\,l}(m=l,\dots,l+N-1)$ for all samples belonging to the considered block. The main idea of this scheme is that the DCT coefficients having large absolute values correspond to signal components, whilst other DCT coefficients relate to noise. The thresholding operation removes or reduces the non-informative spectral components.

Two approaches to thresholding are considered by us. The hard thresholding is defined as:

$$D_{l\,thr}(k) = \begin{cases} D_l(k), & \text{if } |D_l(k)| > T \\ 0, & \text{if } |D_l(k)| \le T \end{cases}, k = 2, \dots N \tag{3}$$

and the combined thresholding is performed as:

$$D_{l\,thr}(k) = \begin{cases} D_l(k), & \text{if } |D_l(k)| > T \\ \dfrac{D_l^{\,3}(k)}{T^2}, & \text{if } |D_l(k)| \le T \end{cases}, k = 2, \dots N, \tag{4}$$

where T is the threshold. It is usually proportional to σ - T = βσ, where β is the proportionality factor set by a user or optimized in some way, for example about 2.6 for hard thresholding and around 4.5 for the combined thresholding.

Noise suppression efficiency depends on several factors, namely, the block size N, signal complexity, input SNR, used threshold, block overlapping, and correctness of assumption concerning noise properties [11], [17]. Below we consider the case of fully overlapping blocks. This means that, for neighboring blocks, l values differ by 1 and N-1 samples are the same. This mode of DCT based filtering provides the best efficiency of noise suppression but requires more computations. However, it is anyway fast enough. It is worth recalling here that each signal sample has filtered values coming from different positions (N positions for most samples) of blocks that contain this sample. Then, the filtered values are averaged.

Other known properties of the DCT-based filtering are the following. First, the denoising is more efficient for smaller input SNR where denoising efficiency can be characterized in different ways. In particular, it is possible to use improvement of SNR defined as:

$$\text{ISNR} = 10log_{10}\left(\frac{\sigma^2}{\text{MSE}}\right) = SNR_{\text{out}} - SNR_{\text{inp}} \tag{5}$$

Here, MSE denotes the mean square error at ilter output, whilst SNR_{out} and SNR_{inp} are output and input SNRs, respectively. Second, denoising using hard and combined thresholding produce approximately the same efficiency under condition that the corresponding optimal β values are applied. A larger β leads to better noise suppression by the expense of worse preservation of signal details. Just this fact explains the existence of optimal β. Third, in previous experiments [17], N=64 has resulted in slightly better outcomes than the use of N=32 or N=16. This is the reason why we applied N=64 in our experiments in this paper. Fourth, if noise is not AWGN, the assumption that it is AWGN leads to less efficient filtering compared to the case of AWGN for the same variance. It is possible to adapt thresholding to the noise spectrum and make the thresholds frequency dependent but, for this purpose, one has to know the noise spectrum a priori or to estimate it from signal+noise mixture at hand. This complicates processing. Because of this, in our experiments here, we have applied the DCT-based filters with thresholds determined by formulas (3) or (4).

Let us present some results obtained for post-filtering applied after decompression. fig. 4 shows dependences of ISNR on β for DCT-based filtering with hard and combined thresholding to the test classical music fragment after compression and filtering for five values of input SNR. As seen, ISNR can be positive for input SNR equal to 0, 5, and 10 dB, but it is negative for input SNR equal to 15 and 20 dB. Similar dependences are presented in fig. 5 for the pop music fragment. In this case, improvement is observed only for input SNRs equal to 0 and 5 dB. This is not surprising since ΔP values in this case are smaller than for classical music fragment (see data in table 2).

The results for hard and combined thresholding are approximately the same for given signal fragment and input SNR and under condition of optimal β setting. Concerning optimal β, it has the tendency to decrease if input SNR becomes smaller.

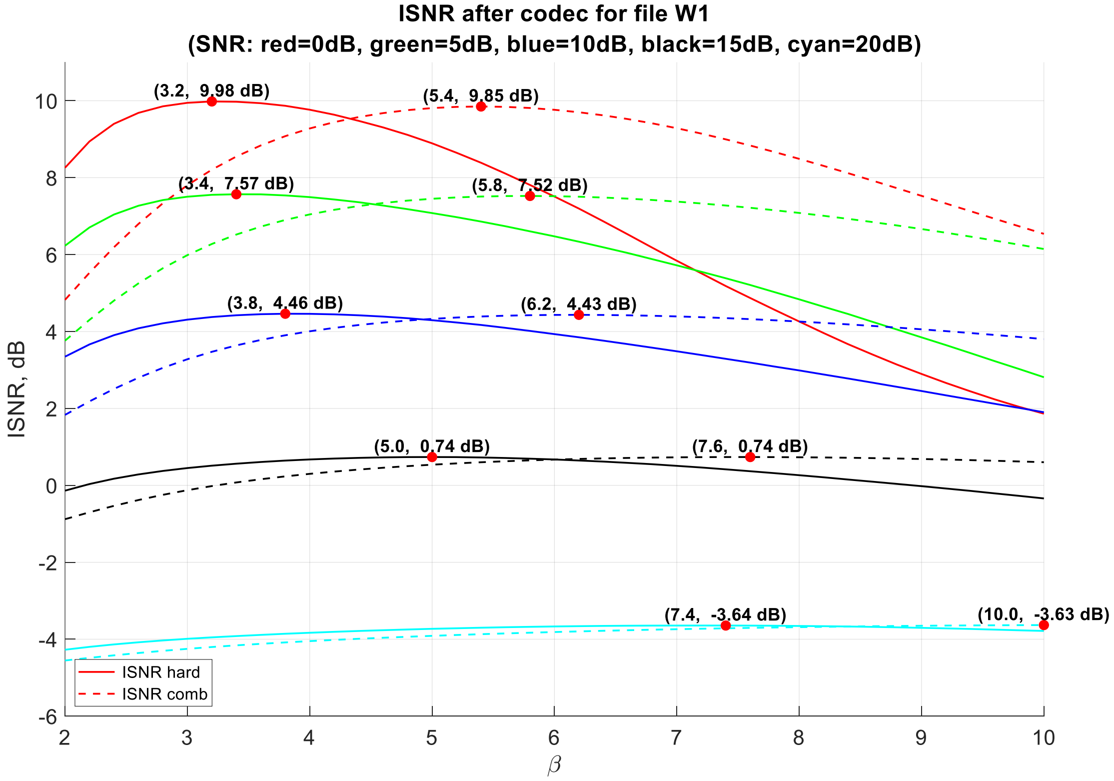

Fig. 4. Dependence of ISNR on β for processing decompressed classical fragment

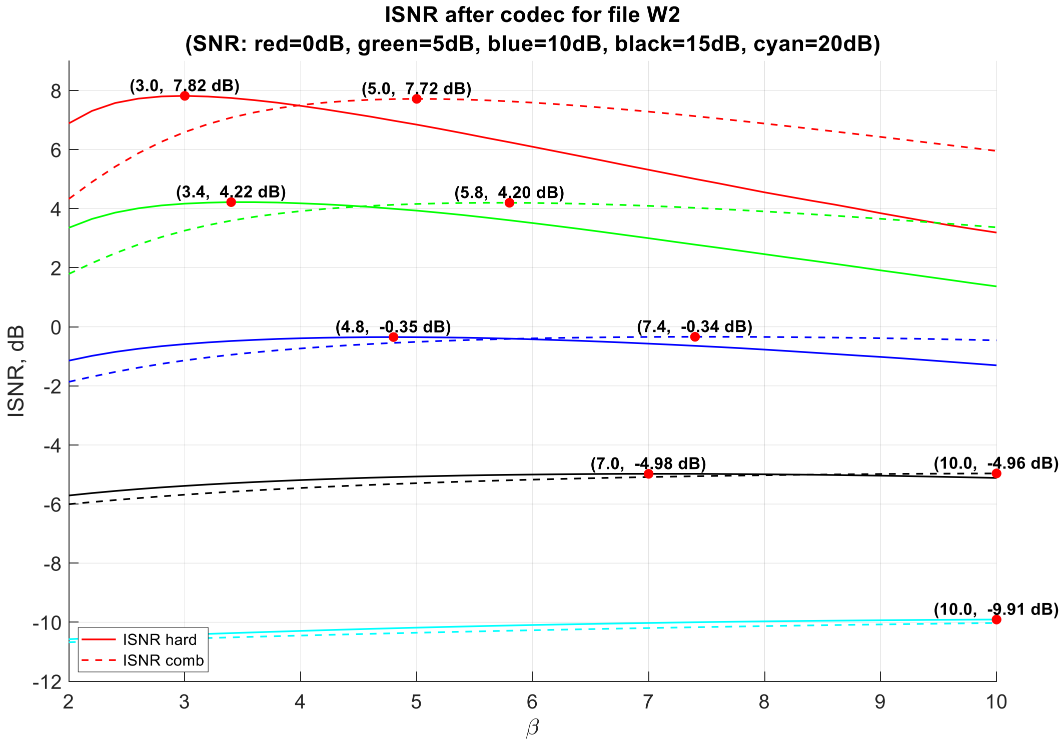

Fig. 5. Dependence of ISNR on β for processing decompressed pop fragment

Conclusions

This paper studies the performance characteristics of the modern neural-based lossy audio compression method applied to signals corrupted by AWGN. The research is focused on analyzing the compression ratio, the nature of introduced distortions, and the efficiency of DCT-based post-filtering for residual noise suppression.

The key findings are as follows:

The neural-based codec demonstrates exceptional robustness, achieving very high compression ratios (CR > 100) across all tested input SNRs. While the presence of intense noise slightly degrades the CR, the compression remains highly effective, confirming the method's suitability for imperfect signal conditions.

The codec exhibits a frequency-dependent effect on noise. For signals with medium to high SNRs (e.g., 20 dB), it provides partial noise suppression, primarily in the high-frequency range (above 5 kHz). However, for signals with low SNRs (e.g., 0 dB), the codec introduces significant distortions in the low-frequency range, which can deteriorate the perceived quality, while still providing some noise suppression at higher frequencies.

The application of DCT-based post-filtering after decompression proved to be a highly effective strategy for low-SNR scenarios. For signals with input SNRs between 0 and 10 dB, post-filtering can significantly improve the output signal quality, yielding a noise reduction of up to 8-10 dB. This confirms that a hybrid approach, combining the high compression efficiency of neural codecs with the targeted denoising capabilities of classical transform-based methods, is a viable and powerful solution for processing noisy audio signals.

Future work will focus on optimizing the post-filtering parameters based on the specific distortion characteristics introduced by the neural codec and exploring adaptive filtering techniques that take into account noise properties.

Внесок авторів

Петро БРИСІН - аналіз джерел та огляд літератури; розробка та реалізація програмного забезпечення для проведення експериментів; проведення експериментів та збір, обробка та візуалізація отриманих результатів; Володимир ЛУКІН – концептуалізація дослідження та постановка задачі; наукове керівництво та загальна методологія дослідження; критичне редагування та наукова редакція рукопису.

Декларація про штучний інтелект

Автор не використовував штучний інтелект при створенні матеріалів статті.

Конфлікт інтересів

Автор заявляє про відсутність конфлікту інтересів та підтверджує, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

References

1. Audio Signal Processing / V. J. Mathews, J.-H. Lee, eds. — Basel, Switzerland: MDPI, 2018. — 188 p.

2. Muller T. Speech quality evaluation of neural audio codecs / T. Muller, S. Ragot, L. Gros, P. Philippe, P. Scalart // INTERSPEECH: conf. proc., Kos Island, Greece, 1-5 Sept. 2024. — 2024. — P. 1–5. — DOI: 10.21437/Interspeech.2024-1072

3. Zhurakovskyi B. Comparative Analysis of Modern Formats of Lossy Audio Data Compression / B. Zhurakovskyi, O. Tsopa // Computing, Telecommunications and Control. — 2019. — Vol. 12, No. 2. — P. 39–51. — URL: https://api.semanticscholar.org/CorpusID:221462262 (date of access: 09.12.2025).

4. Kim B. Lossy audio compression identification / B. Kim, Z. Rafii // 26th European Signal Processing Conference (EUSIPCO): conf. proc., Rome, Italy, 3-7 Sept. 2018. — 2018. — P. 2459–2463. — DOI: 10.23919/EUSIPCO.2018.8553611

5. Kumar R. High-Fidelity Audio Compression with Improved RVQGAN. / R. Kumar, P. Seetharaman, A. Luebs, I. Kumar // ArXiv preprint arXiv:2306.06546. — 2023. — DOI: 10.48550/arXiv.2306.06546

6. Lahrichi Z. QINCODEC: Neural Audio Compression with Implicit Neural Codebooks. / Z. Lahrichi, G. Hadjeres, G. Richard, G. Peeters. // QINCODEC: Neural Audio Compression with Implicit Neural Codebooks. — 2025. — DOI: 10.48550/arXiv.2503.19597

7. Muthu R. Denoising of Speech Signal using Empirical Mode Decomposition and Kalman Filter / R. Muthu, P. Bharath // International Journal of Innovative Technology and Exploring Engineering. — 2020. — Vol. 9, Iss. 8. — P. 83–87. — DOI: 10.35940/ijitee.H6313.069820

8. Kowalski P. Review and comparison of smoothing algorithms for one-dimensional data noise reduction / P. Kowalski, R. Smyk // International Interdisciplinary PhD Workshop (IIPhDW): conf. proc., Świnoujście, Poland, 9-12 May 2018. — 2018. — P. 277–281. — DOI: 10.1109/IIPHDW.2018.8388373

9. URL: https://bellard.org/tsac/ (date of access: 11.12.2025).

10. Lukin V. Preliminary Processing and Lossy Compression of Multichannel Information Data / V. Lukin, M. Zriakhov, A. Popov, O. Pogrebnyak // Industrial Informatics, Research in Computing Science. — Mexico, 2007. — Vol. 31. — P. 105–114.

11. Kryvenko S. Post-Filtering of Noisy Images Compressed by HEIF / S. Kryvenko, V. Rebrov, V. Lukin, V. Golovko, A. Sachenko, A. Shelestov, B. Vozel // Applied Sciences. — 2025. — Vol. 15, Iss. 6. — P. 2939. — DOI: 10.3390/app15062939

12. Valin J.-M. High-Quality, Low-Delay Music Coding in the Opus Codec / J.-M. Valin, G. Maxwell, T. B. Terriberry, K. Vos // 135th AES Convention: conf. proc., New York, USA, 17-20 Oct. 2013. — 2013. — DOI:10.48550/arXiv.1602.04845

13. Borsos Z. AudioLM: A Language Modeling Approach to Audio Generation / Z. Borsos, R. Marinier, D. Vincent, E. Kharitonov // ArXiv preprint arXiv:2209.03143. — 2023. — DOI: 10.48550/arXiv.2209.03143

14. Défossez A. Moshi: a speech-text foundation model for real-time dialogue / A. Défossez, L. Mazaré, M. Orsini, A. Royer // ArXiv preprint arXiv:2410.00037. — 2024. — DOI: 10.48550/arXiv.2410.00037

15. Upadhay P. Denoising 1D signal using wavelets / P. Upadhay, K. Shukla, S. K. Upadhyay // International Journal of Intelligent Systems Technologies and Applications. — 2020. — Vol. 19, Iss. 5. — P. 517–530. — DOI: 10.1504/IJISTA.2020.10034664

16. Xie X. H. An Improved LMS Adaptive Filtering Speech Enhancement Algorithm / X. H. Xie, W. C. Wang // 5th International Conference on Natural Language Processing (ICNLP): conf. proc., Guangzhou, China, 24-26 Mar. 2023. — 2023. — P. 146–150. — DOI: 10.1109/ICNLP58431.2023.00033

17. Brysin P. V. DCT-based denoising of speech signals / P. V. Brysin, V. V. Lukin // Herald of Khmelnytskyi National University. Technical sciences. — 2024. — № 4 (339). — P. 301–309. —DOI: 10.31891/2307-5732-2024-339-4-48

18. Dogra M. Noise Removal from Audio Using CNN and Denoiser / M. Dogra, S. Borwankar, J. Domala // Advances in Speech and Music Technology. Advances in Intelligent Systems and Computing. — Singapore: Springer, 2021. — Vol. 1320. — P. 135–144. —DOI: 10.1007/978-981-33-6881-14

19. Polyakova M. The Design of Fast Type-V Discrete Cosine Transform Algorithms for Short-Length Input Sequences / M. Polyakova, A. Witenberg, A. Cariow // Electronics. — 2024. — Vol. 13, Iss. 21. — P. 4165. — DOI: 10.3390/electronics13214165

20. Ondusko R. Blind Signal-to-Noise Ratio Estimation of Speech Based on Vector Quantizer Classifiers and Decision Level Fusion / R. Ondusko, M. Marbach, R. Ramachandran, L. Head // Journal of Signal Processing Systems. — 2017. — Vol. 89. — P. 415–425. — DOI: 10.1007/s11265-016-1200-z

П. В. Брисін, В. В. Лукін

НЕЙРОМЕРЕЖЕВЕ СТИСНЕННЯ З ВТРАТАМИ ЗАШУМЛЕНИХ АУДІОСИГНАЛІВ ТА ЇХ ПОСТФІЛЬТРАЦІЯ НА ОСНОВІ ДКП

Розглянуто нещодавно запропонований нейромережевий кодек TSAC для стиснення аудіосигналів із втратами, який застосовувався до музичних та мовних сигналів, спотворених адитивним білим гаусовим шумом (АБГШ). Основними цілями були: 1) Аналіз залежності коефіцієнта стиснення (КС) від вхідного відношення сигнал/шум (ВСШ); 2) Вивчення природи спотворень, що вносяться кодеком; 3) Оцінка доцільності та ефективності застосування пост-фільтрації на основі дискретного косинусного перетворення (ДКП) для придушення залишкового шуму в декодованих сигналах. Експериментально встановлено, що кодек має високу робастність: КС перевищує 100 для всіх рівнів вхідного шуму. При цьому спостерігається тенденція до збільшення КС із зростанням ВСШ від 0 до 20 дБ. Аналіз показав, що кодек має частотно-залежний вплив на шум: для сигналів з високим ВСШ (15-20 дБ) спостерігається часткове придушення шуму у високочастотній області (вище 5 кГц), але при цьому вносяться спотворення на низьких частотах. Для сигналів з низьким ВСШ (0 дБ) спотворення на низьких частотах стають більш вираженими, проте ефект шумозаглушення на високих частотах зберігається. Застосування постфільтрації на основі ДКП після декомпресії виявилося високоефективною стратегією саме для сильно зашумлених сигналів. Для вхідних ВСШ в діапазоні 0-10 дБ вдалося досягти значного поліпшення якості (позитивні значення метрики поліпшення ВСШ), з максимальним виграшем до 8-10 дБ при оптимальному виборі порогового коефіцієнта β. У той же час, для сигналів з високим ВСШ пост-фільтрація недоцільна, оскільки спотворення, що вносяться самим фільтром, переважають ефект шумозаглушення. Проведене дослідження підтверджує, що гібридний підхід, що поєднує високу ефективність стиснення сучасних нейромережевих кодеків з класичними методами шумозаглушення, є потужним і перспективним рішенням для обробки зашумлених аудіосигналів. Розглянутий кодек здатний ефективно стискати навіть сильно спотворені дані, а подальша фільтрація на основі ДКП дозволяє значно поліпшити якість відновленого сигналу в умовах низького відношення сигнал/шум.

Ключові слова: аудіосигнал, шум, стиснення із втратами, фільтрація на основі ДКП, спотворення.

Надійшла до редакції: 03.03.2026
Прийнята до друку: 21.04.2026
Опубліковано: 27.04.2026