DOI 10.31891/2307-5732-2024-339-4-48
УДК 621.395.6

BRYSIN PETRO
National Aerospace University – KhAI
https://orcid.org/0000-0003-2598-3907
e-mail: p.v.brysin@khai.edu
LUKIN VOLODYMYR
National Aerospace University – KhAI
https://orcid.org/0000-0002-1443-9685
e-mail: v.lukin@khai.edu

DCT-BASED DENOISING OF SPEECH SIGNALS

This paper considers a traditional task of processing speech signals embedded in noise. A model of additive white Gaussian noise is employed as simple and reasonable case. Existing approaches to denoising applicable to speech signals are briefly reviewed. An opportunity to use denoising based on discrete cosine transform is investigated. Advantages of this approach consists in operation in blocks that allows to obtain filtered values with a reasonable delay with respect to input data. The method performance depends on several factors including block size and block overlapping. Block full overlapping is considered since it is the most efficient whilst fast enough. Block size is fixed and equal to 32 samples. Both conventional signal-to-noise ratio and PESQ metric that reflects peculiarities of speech perception and understanding by humans are used in analysis of filtering efficiency. This is done in order to describe the processing performance more adequately. It is also demonstrated that the denoising performance for the analyzed test signals also depends on, at least, three factors: input signal-to-noise ratio, type of threshold, and parameter β used in threshold setting. To consider practical aspects, wide limits of input signal-to-noise ratio from 0 dB to 40 dB are explored with the main emphasis on the range from 10 to 30 dB. An example showing input and output signals is presented. The main observations based on the obtained results are the following: 1) improvement of signal-noise ratio due to denoising is the largest for low input SNR and can reach 10 dB; 2) there are optimal values of β that have the tendency to increase if input SNR decreases; 3) the combined threshold that is the first time tested for speech signals performs better than the hard threshold. Optimal values of parameter β are recommended based on the analysis carried out for several test signals. The directions of further studies are discussed.

Keywords: additive noise, DCT-based filtering, performance analysis

БРИСІН ПЕТРО, ЛУКІН ВОЛОДИМИР
Національний аерокосмічний університет ім. М.Є. Жуковського "ХАІ"

ЗНЕШУМЛЕННЯ МОВНИХ СИГНАЛІВ З ВИКОРИСТАННЯМ ДКП-ФІЛЬТРАЦІЇ

У цій статті розглядається традиційна задача придушення адитивного білого гаусового шуму в мовних сигналах. Досліджено можливість використання шумозаглушення на основі дискретного косинусного перетворення. Для аналізу ефективності фільтрації використовуються як традиційне відношення сигнал/шум (ВСШ), так і метрика PESQ. Показано, що для аналізованих тестових сигналів ефективність залежить принаймні від трьох факторів: вхідного відношення сигналу до шуму, типу порогу та параметра β, який використовується для встановлення порогу. Основні спостереження такі: 1) покращення співвідношення сигнал/шум за рахунок придушення шумів є найбільшим для низького вхідного ВСШ; 2) існують оптимальні значення β, які мають тенденцію до зростання при зменшенні вхідного ВСШ; 3) комбінований поріг, який вперше тестується для мовних сигналів, працює краще, ніж жорсткий поріг. Обговорюються напрямки подальших досліджень.

Ключові слова: адитивний шум, ДКП-фільтрація, аналіз ефективності

Problem overview

A general problem of noise removal in speech and audio signals has been under interest for several decades [1]. A great number of approaches has been proposed and tested starting from linear finite impulse response and adaptive filters (see [1, 2] and references therein), continued by orthogonal transform based denoising (see [4, 5] and references therein), and completing by modern auto-encoders and neural networks [6]. There are many reasons for the interest to noise removal. First, noise is almost always present in environment where audio signals are recorded; moreover, noise can be quite intensive as this happens in hydroacoustics [7] or in crowded rooms [8]. Second, noise spectral characteristics and intensity can be unknown in advance and they can vary in time [8-10]. Third, it is often desired to carry out noise removal in real time, i.e. with minimal (appropriate) time delay with respect to the signal reception (registration) [11].

This means that a denoising method to be efficient has to be adaptive to signal and noise properties and, simultaneously, quite simple to be implementable. Moreover, it should be local in the sense that a mixture of signal and noise should be processed not after recording the entire signal but in scanning windows or blocks of a certain size [12] to produce an output (filtered data) with delay not larger than a certain threshold. In the case of speech denoising, aspects of its perception have to be taken into account since it is known that SNR is not strictly connected with processed speech perception and people often use criteria other than SNR (or its improvement) to characterize speech quality [12, 13].

Although most papers that concern speech denoising deal with wavelets (see [4, 5, 12] and references therein), orthogonal transform based denoising that exploits discrete cosine transform (DCT) [14] is worth considering as well. Note that signal and image processing based on DCT is widely used in denoising [15, 16] and lossy compression [17, 18]. The advantages of DCT are excellent energy compaction property and high computational efficiency. These positive properties alongside with operation in blocks of limited size [14] make the

DCT-based denoising a good candidate for noise removal in speech signals. Note that performance of the DCT-based filters applied to denoising of one-dimensional (1D) signals can be predicted [19].

Meanwhile, performance of the DCT-based filters depends on several factors including the signal properties, noise intensity, block size, threshold type and value, etc. Thus, **the paper goal** is to analyze the performance characteristics of the DCT-based denoising applied to speech signals corrupted by additive white Gaussian noise (AWGN). The main attention is paid to: 1) considering the filter performance in wide limits of variation of signal-to-noise ratio (SNR), 2) analyzing two types of thresholds with optimization of thresholds for them; 3) studying the filter performance with respect to speech perception criteria.

Analysis of recent sources and fundamentals of DCT-based denoising

Starting from the pioneering papers by Donoho and Johnstone [20], wavelet-based denoising got prime attention. Quite many papers deal with wavelet-based denoising of speech signals. Ali et al [4] proposed double-density dual-tree discrete wavelet transform (DDDTWT) using a level dependent threshold algorithm. Output SNR improvement compared to earlier designed counterparts was demonstrated. Wishwakarma et al [5] showed good performance of Coiflet wavelets in audio denoising and dependence of output indicators, SNR and mean square error (MSE), on the threshold type and settings. The Haar wavelet based denoising was studied in [21] where the authors showed that performance depended on the threshold type. Hidayat et al [22] studied the influence of preliminary denoising on feature extraction for speech recognition. Denoising was shown useful where the Fejer-Korovkin 6 wavelet was the best denoising for input signals with SNR of 5-15 dB. Singh, Aggarwal et al [23] considered soft and hard thresholding and designed modified universal threshold using output SNR as one of performance criterion.

Thus, the following can be concluded. First, the wavelet type plays an important role in denoising efficiency. Second, the noise removal efficiency depends on the threshold type and value, which, in turn, is a function of noise standard deviation (SD) or input SNR (note that AWGN variance can be quite accurately controlled in a blind manner [24, 25]). Third, to be efficient for practice, denoising efficiency should be high for a wide range of input SNR, e.g., starting from 0-5 dB and ending with 35-40 dB (when noise can be hardly noticed and, thus, speech filtering becomes useless). Fourth, although standard criteria of signal denoising efficiency such as output MSE for a given AWGN variance or SNR improvement due to filtering are still widely used, other criteria that take into account peculiarities of speech and audio signal perception are employed as well.

The DCT-based denoising is performed in blocks where the block size is usually equal to 32, 64, 128 for the 1D case [14] and to 8×8 or 16×16 pixels for image processing [26]. The three main steps in DCT-based denoising are direct DCT, thresholding of the obtained AC coefficients, inverse DCT performed in each block. The blocks can be non-overlapping, partly overlapping, and fully overlapping where the latter variant is the most efficient in terms of SNR but the least efficient in terms of required computations. We will further consider the full-overlapping variant since anyway it performs quickly enough. For this variant, output values for a given data sample coming from all block positions that "cover" this sample are averaged.

In this paper, we do not consider the influence of the block size although it is known that a larger block size can be preferable for a larger intensity of the noise and, usually, there exists an optimal block size. Instead, we concentrate on the influence of the threshold type and value. Hard and soft thresholds have been studied in DCT denoising where hard was shown to perform better. Meanwhile, combined thresholding was proposed for image processing applications [27] and it has not been analyzed for 1D signal processing.

Presentation of the main material

To test the effectiveness of filtering, this study uses recordings of English speech in which a male voice utters the so-called Harvard phrases. The Harvard phrases are a standardised set of phrases widely used in testing the speech signal processing systems [28]. The recordings are taken from a set of speech signals created at McGill University, Montreal, Canada [29]. The duration of each signal is approximately two seconds, which at a sampling rate of 16 kHz provides enough information for analysis. This sampling rate is the standard for high quality audio recordings, providing good sound quality.

To conduct the study, a set of audio files was created, in each of which additive white Gaussian noise of different intensity was added to the speech signal. The noise level was chosen to produce a signal-to-noise mixture with signal-to-noise ratios of 0, 10, 20, 30, and 40 dB. Signal-to-noise ratio (hereinafter SNR) is a parameter defined as the ratio of signal power to noise power, which is measured in decibels (dB):

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{\mathrm{P}_{\mathrm{sig}}}{\mathrm{P}_{\mathrm{noise}}}\right). \tag{1}$$

The range of selected values allows to study in detail the effect of noise on speech perception and understanding, as well as to estimate the limiting conditions under which the signal component (useful signal) remains discernible to human hearing. Later, on the basis of measurements and listening, it is possible to select from all signals after the filter those processing variants for which noise reduction works most effectively.

Let us recall some basic principles of one-dimensional filtering based on DCT. Let $S(i)$, $i = 1,...,I$ – be the signal component to be estimated (i is the sample index, I denotes the total number of samples) from the observed realisation $S_n(i)=S(i) + n(i)$, $i=1,...,I$, where $n(i)$ is the noise in the i-th sample assumed to be additive, white and Gaussian with zero mean and a previously known or accurately estimated variance σ^2 . The estimation problem is solved by obtaining an estimate $S_f(i)$,$i=1,...,I$ at the filter output, which should be as close as possible to $S(i)$,$i=1,...,I$

according to the criterion used, which, for example, is often the mean square error (MSE), which for an effective filter should be substantially less than σ^2 .

We consider one-dimensional filtering based on DCT [33] as a noise reduction method for the following reasons:

1) This filter is very efficient and can be easily adapted to different types of noise [14].
2) It has a clear physical meaning and can be considered as an approximation of the local Wiener filter [30].
3) It can be efficiently implemented using block processing, which gives several advantages - processing the long term signal piece-wise, changing the filter parameters locally when needed and the filter results can be obtained before the full signal is recorded.

One-dimensional filtering based on DCT is performed as follows. The data are processed in blocks, where a block includes values $S_n^{bl}(l)=\{S_n(l + j - 1)\}$, j= 1,...,N , N is the block size, usually chosen equal to a power of two and l = 1,...,I-N+1(below we will consider a variant of the DCT filter with the so-called full overlap, which is the most effective in terms of noise suppression and also provides less noise reduction artifacts), l is the index of the leftmost (initial) sample included in the block. For each block, a direct DCT is performed and the result is the DCT coefficients D(k),k=1,...,N, where D(1) is related to the mean in the block and is not involved in further threshold processing. In this study, two types of threshold are used – a hard threshold and a combined threshold. Simple hard threshold processing is performed according to the algorithm:

$$\mathrm{D_{thr}}(\mathrm{k}) = \begin{cases} D(k), & if\,|\mathrm{D}(\mathrm{k})| > T \\ 0, & if\,|\mathrm{D}(\mathrm{k})| \le T \end{cases}, \ \mathrm{k} = 2,...\mathrm{N}, \tag{2}$$

Processing using the combined threshold [31] uses the following expression:

$$\mathrm{D_{thr}}(\mathrm{k}) = \begin{cases} D(k), & if\,|\mathrm{D}(\mathrm{k})| > T \\ \mathrm{D}^3(k)/\mathrm{T}^2, & if\,|\mathrm{D}(\mathrm{k})| \le T \end{cases}, \ \mathrm{k} = 2,...\mathrm{N}, \tag{3}$$

where T is the threshold value, which is generally set equal to βσ (β – is a selectable coefficient, which depends on the type of threshold and for a hard one is usually set to 2.7 by default). Processing using a combined threshold has not been previously applied for filtering of one-dimensional signals. For image processing, the recommended values of β were about 1.5 times higher than for the hard thresholding [32].

After such thresholding, the inverse DCT is applied to $D_{thr}(k)$,k=1,...,N and the filtered values for the given block $S_f^{bl}(l) = \{S_f(l + j - 1)$, j= 1,...,N are obtained. As can be seen, for each sample there can be from one (for the first and last blocks) to N filtered values belonging to overlapping blocks. There are different variants of their processing [33]. However, more complex variants do not provide a significant gain in efficiency, so let us focus on the simplest variant - averaging of the obtained estimates.

This paper considers the DCT filtering with a block size of N = 32 samples. Using small block sizes reduces overall system complexity and delay, which is critical for real-time applications such as video conferencing or voice control systems. Using small block sizes allows for more detailed signal processing, which is important for preserving high-frequency components and speech details; in addition, a small block size reduces the probability of artifacts appearing at block boundaries, especially when using full overlap filtering. Processing smaller blocks requires less memory to store temporary data, which is especially useful for resource-constrained devices such as mobile phones or embedded systems.

The efficiency of the DCT-based filter in terms of noise reduction obviously depends on many factors [14, 33]: properties of the signal component, filter parameters (N, β), signal-to-noise ratio at the input (see formula (1)). Filtering efficiency can be quantitatively characterized by either MSE/σ^2 ratio where:

$$\mathrm{MSE} = \sum_{i=1}^{I} \frac{\left[\mathrm{S}(\mathrm{i}) - \mathrm{S_f}(\mathrm{i})\right]^2}{\mathrm{I}-1}, \tag{4}$$

or the improvement in the signal-to-noise ratio (in dB):

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

In this study, to evaluate the effectiveness of filtering, we use two metrics, the speech quality metric PESQ [34] and ISNR. The Perceptual Evaluation of Speech Quality (PESQ) metric is recommended by the International Telecommunication Union (ITU-T) standard ITU-T P.862 for evaluating speech quality. PESQ evaluates speech quality from a human perceptual perspective, considering aspects such as clarity, crispness and naturalness. The metric is based on psychoacoustic models of human perception of sound and attempts to simulate the rating a person would give after listening. PESQ measures the quality of one-way voice transmission: a signal is fed into the system under test and the degraded output signal is compared by PESQ with the input (reference) signal (see the block-diagram in Fig. 1 taken from [34]).

Signal processing performed in the PESQ algorithm contains several stages - equalization of the levels of the reference and target signals, input filtering, time aligning and equalization, auditory transformation, and disturbance processing.

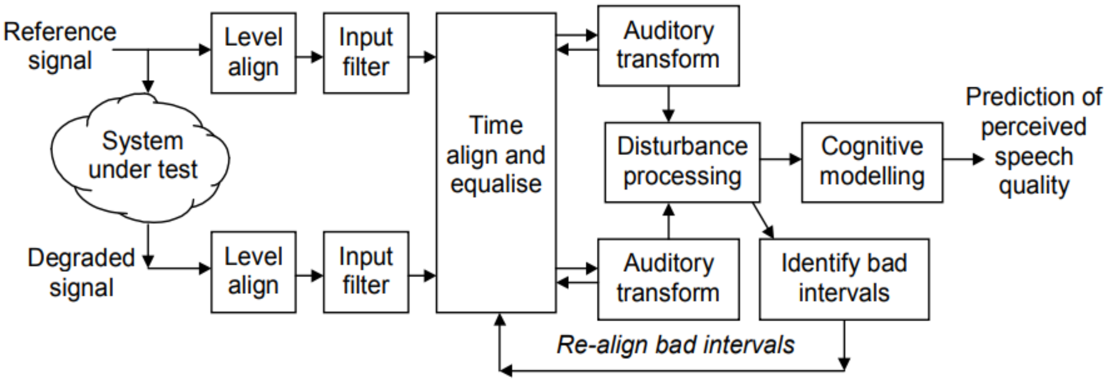

Fig. 1. PESQ algorithm

PESQ results are evaluated using the MOS-PESQ scale and are expressed in a numerical range from -0.5 to 4.5, where a higher value indicates better speech quality. In our study, we use the standard MOS-LQO (Mean Opinion Score - Listening Quality Objective) scale where the score ranges from 1 to 5. To convert MOS-PESQ values to MOS-LQO values, the additional recommendation ITU-T P.862.1[34] is used. Figure 2 shows the MOS-LQO scale and the correspondence between this scale and user satisfaction level [35].

MOS	Quality	User satisfaction
5	Excellent	4,3 - 5,0 Very satisfied
4	Good	4,0 - 4,3 Satisfied
3	Fair	3,6 - 4,0 Some users dissatisfied
		3,1 - 3,6 Many users dissatisfied
2	Poor	2,6 - 3,1 Nearly all users dissatisfied
1	Bad	1,0 - 2,6 Not recommended

Fig. 2. Relation between MOS and user satisfaction

Figure 3 shows the dependence of PESQ on the parameter β for the second speech signal (file F2) for different SNR at the filter input. For each SNR, a set of three graphs is shown - the solid horizontal line shows the measured PESQ value at the filter input, the dashed line shows the PESQ graph after filtering using the hard threshold, the dotted line shows the PESQ graph after filtering using the combined threshold. The graphs are shown in red for a signal with SNR = 0 dB, green for 10 dB, blue for 20 dB, black for 30 dB, and cyan for 40 dB.

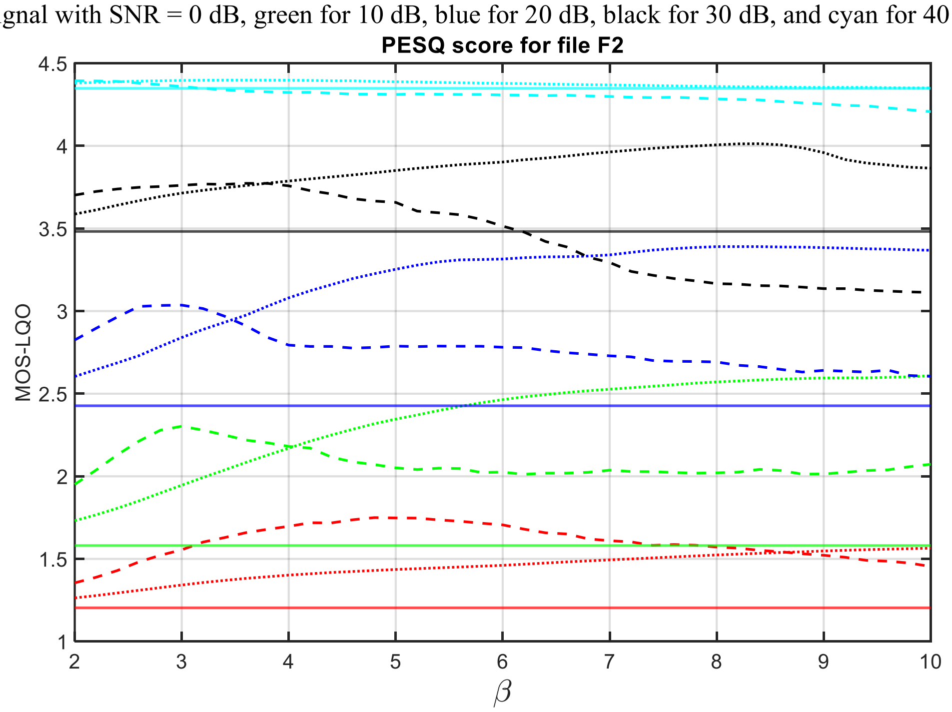

Fig. 3. PESQ metric for the file F2

Figure 4 shows the dependence of SNR improvement at the filter output on the parameter β for the same speech signal with different SNR at the filter input. For each SNR a set of two graphs is shown - the dashed line shows the graph of SNR improvement after filtering using a hard threshold, the dotted line shows the graph of SNR improvement after filtering using a combined threshold. Red colour shows the graphs for a signal with SNR = 0 dB, green - 10 dB, blue - 20 dB, black - 30 dB, cyan - 40 dB.

From the presented graphs, we can see that both metrics show a positive effect of filtering for all input SNRs. However, for some values of signal input SNR, the use of filtering does not make sense. Consider these cases - for SNR = 40 dB the improvement in the metrics is very small, the PESQ score of the input signal has a high value and the noise in the input signal is almost indistinguishable by ear. For SNR = 30 dB, the improvement of the metrics is slightly greater, but also by ear the input signal with this SNR is initially of a rather high quality and the noise level at this SNR allows perfect understanding of what is being said. For SNR = 0 dB, the improvement in metrics is the most significant of the whole set of signals at the filter input, but initially the speech signal with this SNR has a very low PESQ score and although this score increases after filtering, but still in terms of speech intelligibility remains poor. Therefore, it is further proposed to investigate the filter using signals with SNR equal to 10 and 20 dB, where noise filtering works quite effectively and its application is appropriate.

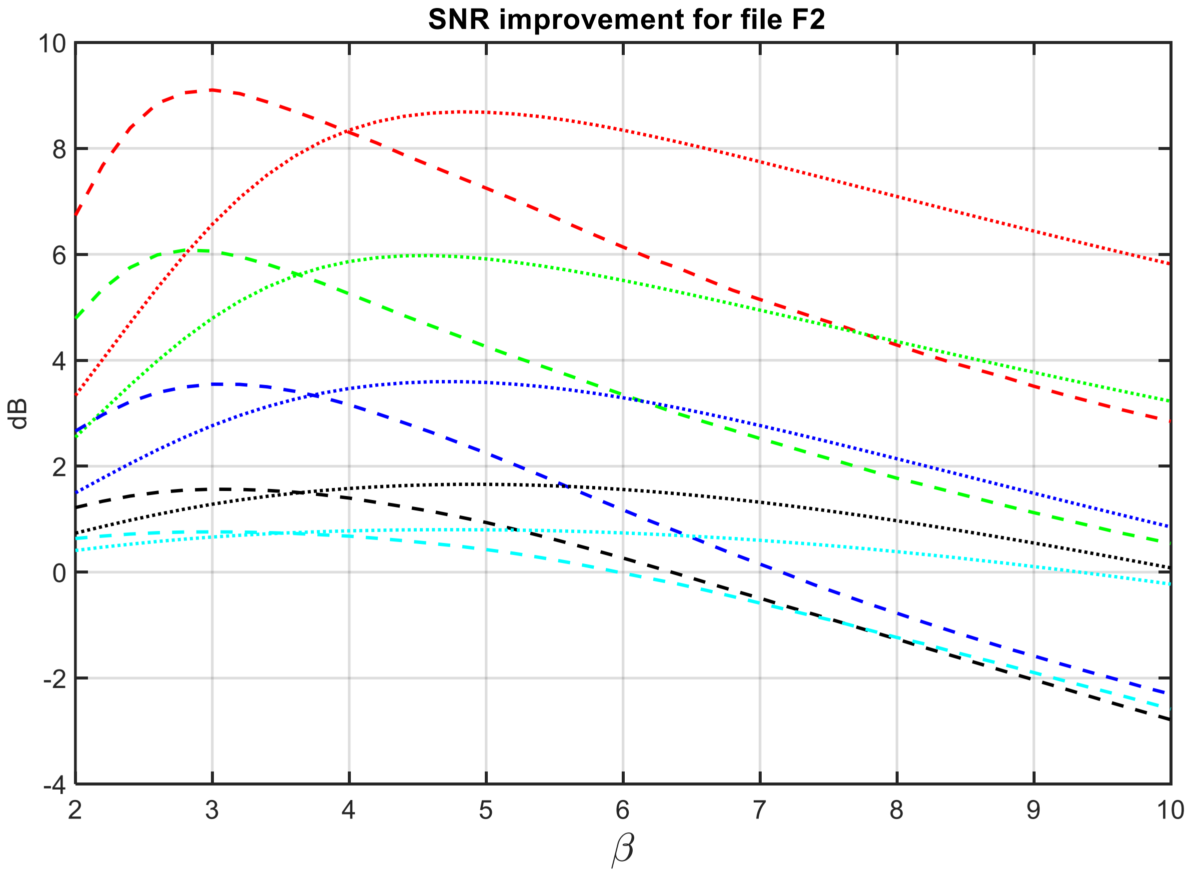

Fig. 4. ISNR metric for file F2

For these input SNRs, filtering using the combined threshold gives (we compare for optimal values of β) a noticeable gain according to the PESQ metric. By ear, these differences are also noticeable. According to ISNR, the performance of the two analyzed filtering variants is approximately the same. Figures 5a - 5d show the time plots of the half-second speech signal at all processing stages. Figure 5a shows the signal without noise, Figure 5b presents the signal at the filter input with SNR = 10 dB. Figure 5c shows the filtered signal using hard thresholding with parameter β = 2.8. Figure 5d represents the filtered signal using a combined threshold with parameter β = 4.6. The values of β are chosen based on the maximum ISNR value. Noise reduction is considerable.

The study was also conducted for several speech signals and in all cases the quality metrics showed an improvement in the signal characteristics after filtering. Figures 6 and 7 show plots of the metrics at the filter output for another than F2 speech signal. It can be noted the similarity in improvement of the PESQ and ISNR after filtering.

It is clear from these graphs that there is a certain value of β, at which the quality metrics at the filter output are maximised. At the same time, for hard and combined thresholds the value of metrics is different - the optimal value of β for combined threshold is significantly higher than for hard threshold.

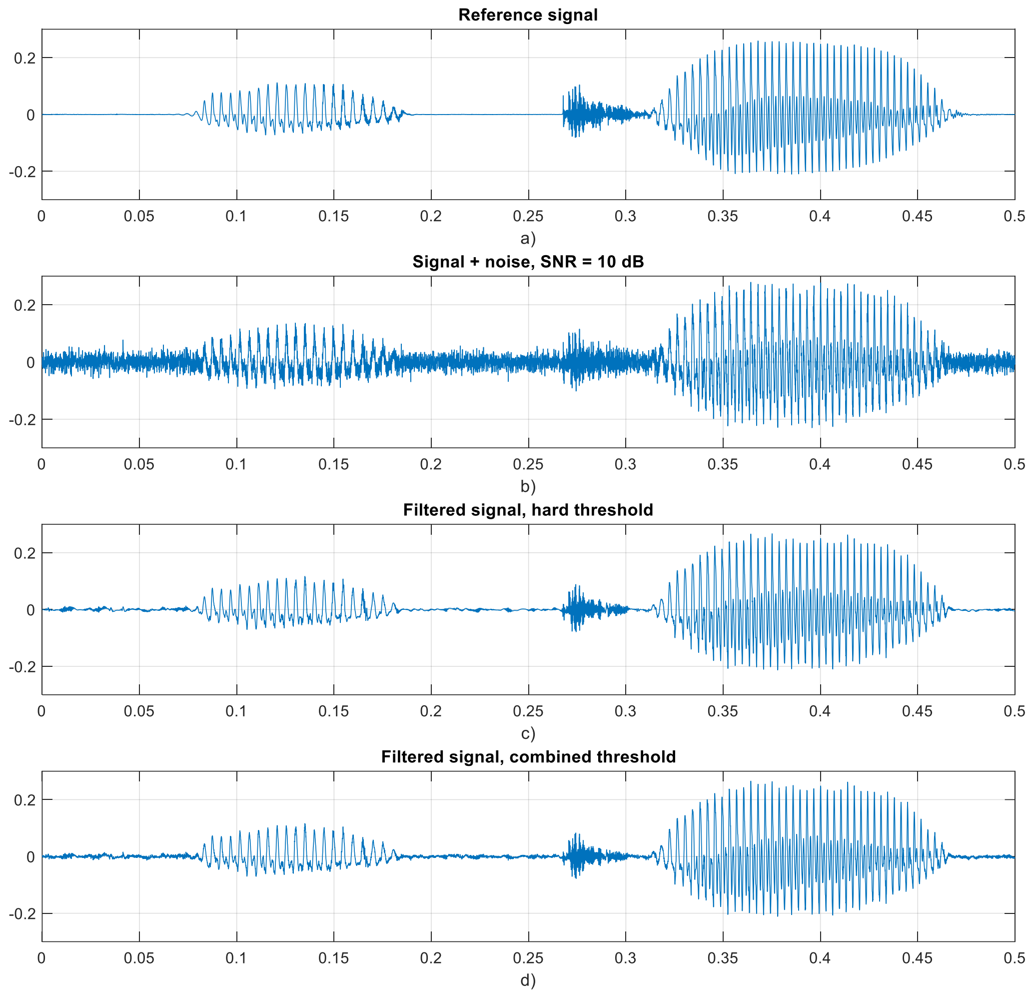

Fig. 5. Time diagrams of the F2 signal processing. a) - reference, b) - signal + noise (SNR =10dB), c) - filtered signal (hard threshold, β = 2.8), c) - filtered signal (combined threshold, β = 4.6)

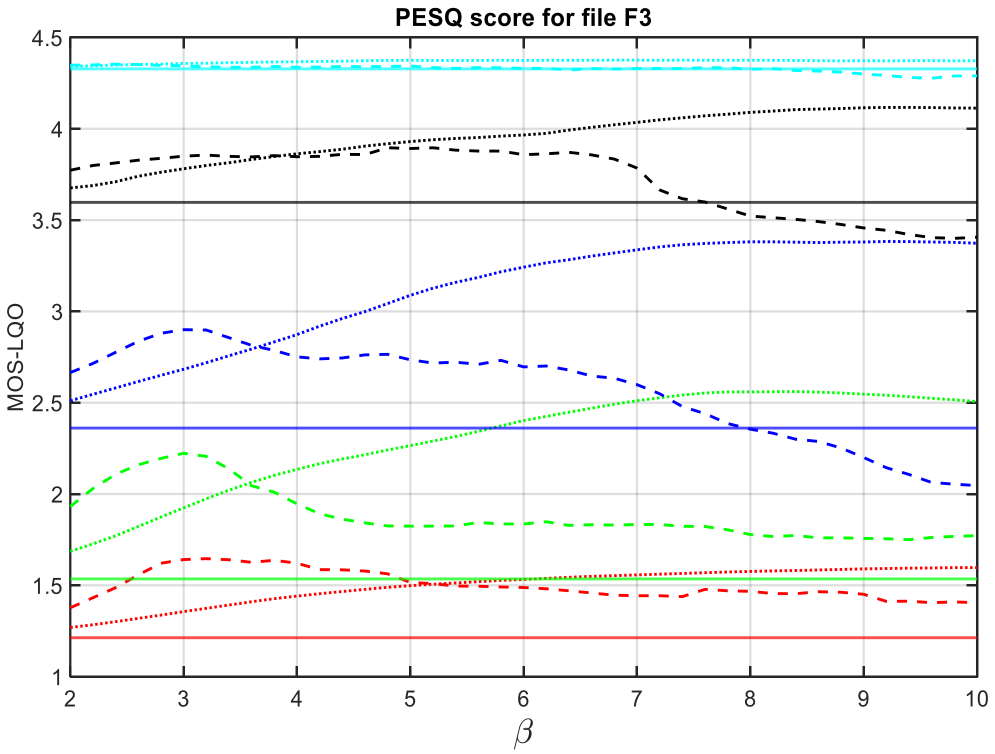

Fig. 6. PESQ metric for file F3

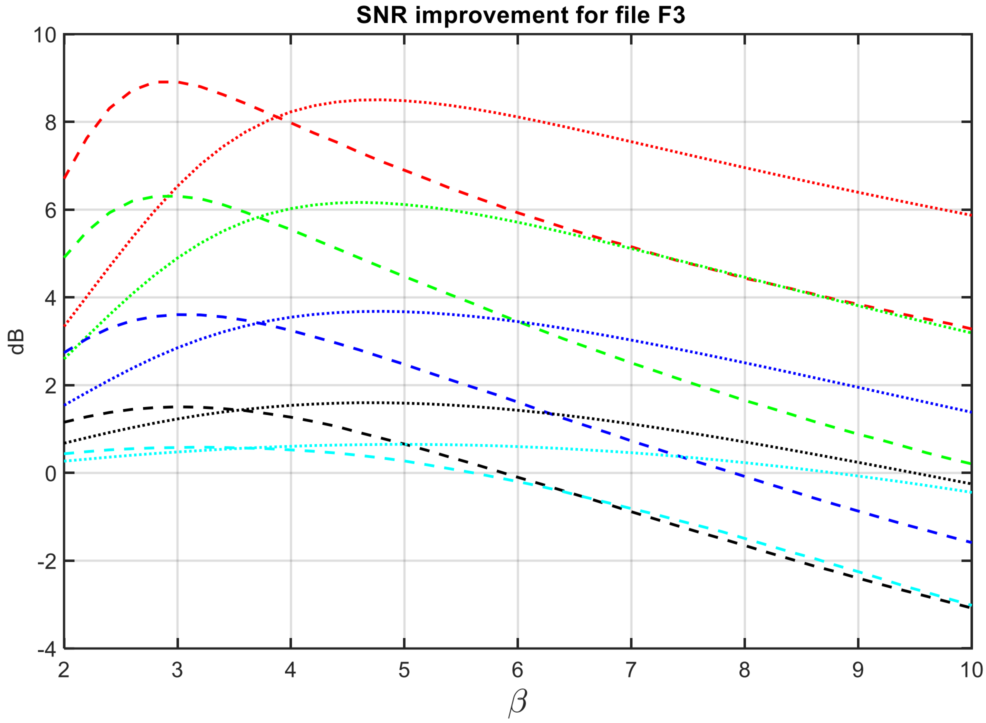

Fig. 7. ISNR metric for file F3

For signals with SNR equal to 10 and 20 dB, the PESQ value is higher in the case of using the combined threshold for filtering. We also note that the optimal β shifts towards higher values when the input SNR decreases and the optimal β is approximately the same when the SNR is the same for different test signals. Thus, it can be recommended to use the combined threshold. Knowing the optimal value of β it is possible to adapt the filtering to the known SNR at the filter input. In general, β≈8 seems to be a reasonable (quasi-optimal) solution for the filtering method with the combined threshold.

Conclusions

The problem of filtering human speech audio signals distorted by AWGN using the one-dimensional DCT filter with full overlap of blocks of size N=32 is considered. The PESQ speech quality metric was used to evaluate the performance. The efficiency of noise filtering using hard and combined types of threshold and parameter β lying in the range of 2 - 10 has been analysed. It is shown that in some cases the use of filtering does not make much sense, but in other cases a positive effect is observed. It is found that the greatest efficiency of filtering in accordance with the PESQ metric is achieved when using a combined threshold with β≈8.

References

1. Y. Hu, P. C. Loizou, Evaluation of Objective Quality Measures for Speech Enhancement, IEEE Transaction on Audio, Speech and Language Processing, Vol. 16, No. 1, 2008, pp. 229-238. doi:10.1109/TASL.2007.911054.
2. R. Muthu, P. Bharath, Denoising of Speech Signal using Empirical Mode Decomposition and Kalman Filter, International Journal of Innovative Technology and Exploring Engineering. 2020 9. 10.35940/ijitee.H6313.069820.
3. X. H. Xie, W. C. Wang, An Improved LMS Adaptive Filtering Speech Enhancement Algorithm, 2023 5th International Conference on Natural Language Processing (ICNLP), Guangzhou, China, 2023, pp. 146-150, doi: 10.1109/ICNLP58431.2023.00033.
4. M. A. Ali, P. M. Shemi, An improved method of audio denoising based on wavelet transform, 2015 International Conference on Power, Instrumentation, Control and Computing (PICC), Thrissur, India, 2015, pp. 1-6, doi: 10.1109/PICC.2015.7455802.
5. D. K. Vishwakarma, R. Kapoor, A. Dhiman, A. Goyal, D. Jamil, De-noising of Audio Signal using Heavy Tailed Distribution and comparison of wavelets and thresholding techniques, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2015, pp. 755-760.
6. M. Dogra, S. Borwankar, J. Domala, (2021), Noise Removal from Audio Using CNN and Denoiser, in: A. Biswas, E. Wennekes, TP. Hong, A. Wieczorkowska, Advances in Speech and Music Technology. Advances in Intelligent Systems and Computing, vol 1320. Springer, Singapore.
7. J. Vergoz, Y. Cansi, Y. Cano, et al, (2021), Analysis of Hydroacoustic Signals Associated to the Loss of the Argentinian, ARA San Juan Submarine. Pure Appl. Geophys. 178, pp. 2527–2556.

8. YH. Wu, E. Stangl, O. Chipara, SS. Hasan, A. Welhaven, J. Oleson, Characteristics of Real-World Signal to Noise Ratios and Speech Listening Situations of Older Adults With Mild to Moderate Hearing Loss, Ear Hear. 2018 Mar/Apr;39(2):pp. 293-304. doi: 10.1097/AUD.0000000000000486. PMID: 29466265; PMCID: PMC5824438.

9. R. Ondusko et al., Blind Determination of the Signal to Noise Ratio of Speech Signals Based on Estimation Combination of Multiple Features, APCCAS 2006 - 2006 IEEE Asia Pacific Conference on Circuits and Systems, Singapore, 2006, pp. 1895-1898, doi: 10.1109/APCCAS.2006.342229.

10. R. Ondusko, M. Marbach, R. Ramachandran, L. Head, (2017), Blind Signal-to-Noise Ratio Estimation of Speech Based on Vector Quantizer Classifiers and Decision Level Fusion, Journal of Signal Processing Systems. 89. 10.1007/s11265-016-1200-z.

11. D. Alexandre, S.Gabriel, A. Yossi, Real Time Speech Enhancement in the Waveform Domain, 2020. Interspeech 2020 Paper.

12. L. W. P. Biscainho, F. P. Freelanci, P. A. A. Esquef, P. S. R. Diniz, Wavelet shrinkage denoising applied to real audio signals under perceptual evaluation, 2000 10th European Signal Processing Conference, Tampere, Finland, 2000, pp. 1-4.

13. J. Beerends, A. Rix, M. Hollier, (2002), Perceptual evaluation of speech quality (PESQ) - The new ITU standard for end-to-end speech quality assessment - Part II - Psychoacoustic model, Journal of the Audio Engineering Society. Audio Engineering Society. 50.

14. V.V. Lukin, D.V. Fevralev, S.K. Abramov, S. Peltonen, J. Astola, Adaptive DCT-based 1-D filtering of poisson and mixed poisson and impulsive noise, CDROM Proceedings of LNLA, Switzerland, 2008. 8 p.

15. O. Rubel, V.Lukin, Improved prediction of DCT-based filter performance using regression analysis, Information and Telecommunication Sciences, 2014, Volume 5, Number 1, Kiev, Ukraine, pp. 30-41.

16. H. Hong, M. He, K. Wang, L. Wu. (2022), An Image Denoising Method for Real Scene Based on Pixel-Level Noise Estimation, 306-311. 10.1145/3569966.3570058.

17. F. Li, S. Krivenko, V. Lukin, An Approach to Better Portable Graphics (BPG) Compression with Providing a Desired Quality, 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2020, pp. 13-17, doi: 10.1109/ATIT50783.2020.9349289.

18. L. George, (2014), Audio Compression Based on Discrete Cosine Transform, Run Length and High Order Shift Encoding, International Journal of Engineering and Innovative Technology (IJEIT). 4. pp. 45-51.

19. S. Abramov, V. Abramova, V. Lukin, K. Egiazarian, Prediction of Signal Denoising Efficiency for DCT-Based Filter, Telecommunications and Radio Engineering, Vol. 78, No 13, 2019, pp. 1129-1142.

20. D.L. Donoho, (1995), Denoising by Soft-Thresholding. IEEE Transactions on Information Theory, 41, pp. 613-627.

21. J. S.Jakati, (2020), Efficient Speech De-noising Algorithm using Multi-level Discrete Wavelet Transform and Thresholding. International Journal of Emerging Trends in Engineering Research. 8. 2472-2480. 10.30534/ijeter/2020/43862020.

22. R. Hidayat, A. Bejo, S. Sumaryono, A. Winursito, Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System, 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Bali, Indonesia, 2018, pp. 280-284, doi: 10.1109/ICITEED.2018.8534807.

23. R. Aggarwal, J.Singh, ,V. Gupta, S. Rathore, M. Tiwari, A. Khare, (2011), Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold. International Journal of Computer Applications. 20. pp. 14-19. 10.5120/2431-3269.

24. A. Kharkov, V. Oliinyk, V. V. Lukin, S. S. Krivenko, Blind estimation of noise variance for 1D signal denoising, Telecommunications and Radio Engineering, Vol. 79, No 7, 2020, pp. 567-581.

25. D. Makovoz, (2006), Noise Variance Estimation In Signal Processing, pp. 364 - 369. 10.1109/ISSPIT.2006.270827.

26. O. Pogrebnyak, V. Lukin, Wiener DCT Based Image Filtering, Journal of Electronic Imaging. – 2012. – No 4. – 14 p.

27. D.V. Fevralev, S. S. Krivenko, V. V. Lukin, R. Marques, F. Medeiros, Combining level Sets and Orthogonal Transform for Despeckling SAR Images, Aerospace engineering and technology. – 2013. – No 2/99. – pp. 103-112.

28. IEEE Subcommittee on Subjective Measurements, "IEEE Recommended Practice for Speech Quality Measurements", IEEE Trans. Audio and Electroacoustics, vol. AU-17, no. 3, pp.225–246, Sept. 1969 (IEEE Standards Publication No. 297-1969).

29. TSP speech database. https://www.mmsp.ece.mcgill.ca/Documents/Data/TSP-Speech-Database/TSP-Speech-Database.pdf

30. A. Gotchev, N. Nikolaev, K. Egiazarian, (2001), Improving the transform domain ECG denoising performance by applying interbeat and intra-beat decorrelating transforms, Proc. of ISCAS 2001, Sydney, NSW, 2, pp. 17-20, doi: 10.1109/ISCAS.2001.920995.

31. V. Lukin, (2010), Speeding up DCT-Based Filtering of Images. Proc. of TCSET'2010, February pp. 23-27, 2010, Lviv-Slavske, Ukraine.

32. V.Lukin, O.Rubel, R.Kozhemiakin, S.Abramov, A.Shelestov, M.Lavreniuk, M.Meretsky, B.Vozel, K.Chehdi, (2018), Despeckling of Multitemporal Sentinel SAR Images and Its Impact on Agricultural Area Classification. Recent Advances and Applications in Remote Sensing. Edited by Ming-Chih Hung and Yi-Hwa Wu, 2017, Intech, DOI: 10.5772/intechopen.72577.

33. R.Oktem, L.Yarovslavsky, K.Egiazarian, Signal and image denoising in transform domain and wavelet shrinkage: A comparative study. In Proceedings of 9th European Signal Processing Conference, 1998, pp. 2269-2272.

34. "P.862 : Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs". https://www.itu.int/rec/T-REC-P.862

35. ITU-T G.107, “The E-model, a computational model for use in transmission planning”, https://www.itu.int/rec/T-REC-G.107