https://doi.org/10.31891/2307-5732-2025-355-61
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

EFFICIENCY ANALYSIS FOR DCT-BASED DENOISING OF SPEECH SIGNALS

This paper investigates the problem of noise suppression in speech signals, where interference is modeled using the Additive White Gaussian Noise (AWGN) model. A key challenge for this case is to effectively reduce noise without introducing audible artifacts that degrade the perceptual quality of the speech. We employ a denoising method based on the Discrete Cosine Transform (DCT), which is applied to fully overlapping signal blocks of fixed sizes (16, 32, and 64 samples). The effectiveness of the proposed approach is comprehensively evaluated using both objective and perceptual criteria. The improvement in the output Signal-to-Noise Ratio (SNR) compared to the input (ISNR) serves as the objective measure. The perceptual quality of the processed speech is assessed using the standard Perceptual Evaluation of Speech Quality (PESQ) metric. We investigate the dependence on the input Signal-to-Noise Ratio (SNR), the processing block size, the type of threshold applied (hard and combined), and the parameter β employed in threshold calculation. The analysis, conducted on a set of standard Harvard sentence test signals, yielded highly consistent results and revealed the following tendencies: 1) A block size of N=64 consistently provides the best denoising efficiency according to both metrics compared to sizes N=16 and N=32; 2) The greatest gain in the ISNR metric is observed at low input SNR values, which is particularly important for highly noisy signals; 3) The optimal value of the parameter β depends strongly on both the input SNR0 (generally decreasing as SNR increases for ISNR optimization) and the chosen evaluation metric; 4) The combined threshold demonstrates an advantage over the hard threshold according to the perceptual PESQ metric, provided β is selected appropriately, whereas their ISNR performance characteristics are approximately the same for the respective optimal β values. Ultimately, this study underscores the necessity of an adaptive approach to parameter selection, tailored to both the specific noise conditions and the primary application's performance metric, whether objective or perceptual. Moreover, the computational complexity of the DCT-based method remains manageable, making it suitable for real-time applications. Examples of signal processing are presented and discussed.

Keywords: additive white Gaussian noise, DCT-based denoising, efficiency analysis, block size

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

АНАЛІЗ ЕФЕКТИВНОСТІ ЗНЕШУМЛЕННЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ ДКП

Наша стаття присвячена традиційній задачі видалення шумів у мовних сигналах. Перешкоди моделюються як адитивний білий гаусів шум. Його придушення базується на дискретному косинусному перетворенні (ДКП), застосованому для блоків фіксованого розміру, що повністю перекриваються. Як кількісні критерії ефективності придушення шуму використовуються покращення вихідного відношення сигнал/шум (ВСШ) порівняно з вхідним, а також метрика PESQ. Проаналізовано кілька тестових мовних сигналів, і результати, що отримані для них, дуже схожі. Ця ефективність залежить від кількох факторів, у тому числі відношення вхідного сигналу до шуму, розміру блоку, типу використовуваного порогу (жорсткий і/або комбінований) і параметра β, який використовується при обчисленні порога. Спостережувані тенденції такі: 1) найбільше покращення ВСШ за рахунок фільтрації відбувається при малих вхідних ВСШ; 2) оптимальні значення β зазвичай зменшуються, якщо вхідне ВСШ збільшується; 3) є випадки, коли комбінований поріг перевершує підхід до обробки мовних сигналів, який спирається на жорсткий поріг; 4) використання розміру блоку, що дорівнює 64, призводить до кращої ефективності видалення шуму порівняно з випадками розміру блоку, що дорівнює 32 і 16. Наведено та обговорено приклади обробки сигналів.

Ключові слова: адитивний білий гаусів шум, ДКП-фільтрація, аналіз ефективності, розмір блока

Стаття надійшла до редакції / Received 11.05.2025
Прийнята до друку / Accepted 26.06.2025

Problem overview

Speech and other types of audio signals are often noisy due to several reasons [1]. Because of this, great attention has been paid to filtering (denoising) of such signals over the past 40 years. Initially, linear finite impulse response (FIR) and various adaptive filters [1-3] were developed. Later and in parallel, orthogonal transform techniques based on wavelets and DCT were studied [4, 5]. The use of auto-encoders and convolutional neural networks has become popular recently [6].

For different applications, there are different reasons for the presence of noise in registered audio and speech signals. This noise might be quite intensive – this takes place in crowded rooms and places [7] as well as in hydroacoustics [8]. In addition, the noise intensity and spectral characteristics might vary in time in quite wide limits [7, 9]. Recall that there are several applications where noise removal should be carried out in real time [10]. Summarizing the aforesaid, a speech signal denoising method should be able to adapt to the properties of the registered

signal and noise. At the same time, such a method and the corresponding algorithm have to be simple and fast enough to be a good candidate for practical application. This imposes restrictions on the size of scanning windows or blocks used for producing the output signal with a suitable delay with respect to the input [11]. One should also keep in mind that, in speech denoising, the perception of input and/or output signals is important where the SNR and its improvement are not strictly connected with speech perception; due to this, special metrics of speech quality are often employed [11, 12].

As mentioned previously, there are quite many papers on speech denoising that consider the use of wavelets [4, 5, 11]. Meanwhile, other orthogonal transforms such as DCT [13-15] can be a useful tool as well. Having excellent energy compaction properties, DCT is widely exploited in signal/image filtering [14, 16, 17] and lossy compression [18, 19]. In addition, DCT is characterized by high computational efficiency and existence of fast algorithms [15]. Besides, DCT-based filtering is carried in blocks of a rather small size that allows obtaining the filter output with a quite small delay with respect to input data. Moreover, the DCT-based filter performance can be predicted in advance [20]. Taken together, these positive features allow expecting DCT-based filtering to perform well for noise suppression in speech and audio signals. In fact, this has already been shown in our paper [14]. Meanwhile, it has been demonstrated that the DCT-based filter performance depends on several factors, namely input SNR, threshold type and its value. The study has been performed for the fixed block size of 32 samples and it is not clear is this block size the best choice and are the tendencies discovered in [14] valid for other possible block sizes.

Therefore, **the goal of this paper** is to study the performance of the DCT-based filter for speech signals contaminated by AWGN for other block sizes. The main considered aspects are the following: 1) we analyze the filter characteristics over a wide range of input SNR values; 2) we study three sizes of blocks (16, 32 and 64) and compare the results; 3) we consider hard and combined thresholds with optimization of the thresholds for both cases; 4) we apply the speech perception criterion PESQ and analyze the filter performance according to it alongside with conventional SNR.

Analysis of recent sources on transform-based denoising

A starting point in orthogonal transform-based denoising was, probably, the paper [21] where wavelets were used. A common assumption behind such denoising is that the main part of information is contained in a limited number of large amplitude spectral coefficients whilst the noise is spread between all components and, thus, small amplitude spectral components, most probably, relate to the noise and can be neglected. Later it was shown that denoising efficiency depends on several factors including a used wavelet type, threshold type and its setting, etc. This has led to the widespread application of wavelet-based denoising in different signal/image processing areas including speech filtering.

Here are some examples. The authors of [5] designed double-density dual-tree discrete wavelet transform (DDDTWT) and applied a level dependent thresholding algorithm. They demonstrated improvement of output SNR in comparison with several earlier proposed analogs. Wishwakarma et al [4] studied the Coiflet wavelets for audio signal denoising. They showed that output SNR or, equivalently, mean square error (MSE) strongly depended on the threshold type and settings and could be significantly improved due to filtering. One more wavelet type, the Haar one, was considered by the authors of [22] where it was demonstrated an essential dependence of performance on the threshold type. Aggarwal et al [23] studied and compared soft and hard thresholding. Moreover, they proposed modified universal thresholding where output SNR was used as the main performance criterion. The low input SNR was considered in [24] where the authors analyzed the preliminary filtering impact on speech feature extraction. Two important conclusions were drawn. First, preliminary filtering was shown useful. Second, the best results were provided by the Fejer-Korovkin 6 wavelet based denoising.

The analysis carried out above shows the following. First, the denoising efficiency significantly depends on the wavelet type. Second, the threshold type and value are two more factors determining the filter performance. Note here that the threshold is usually set proportional to the noise standard deviation (STD) where the AWGN STD can be quite accurately estimated automatically [25, 26]). Third, the researchers mainly consider input SNR in the limits from 0 to 30 dB since, for the input SNR about 35-40 dB, the noise becomes hardly noticed and, therefore, it becomes useless to carry out noise removal. Fourth, in parallel to using conventional criteria of filtering efficiency (output MSE or SNR improvement due to denoising), the criteria characterizing signal perception are widely employed.

All these conclusions and observations were taken by us into account in [14] dealing with one-dimensional (1-D) DCT-based denoising in blocks of size 16. Note here that the block size equal to the powers of two such as 8, 16, 32, 64, and 128 for the 1D case [13] and 8×8 or 16×16 pixels in image denoising [27] is a common practice if one wants to provide high computational efficiency due to exploiting fast DCT algorithms. Recall here that DCT for each block is used twice: one first applies direct DCT, then thresholding of the determined AC DCT coefficients is performed, and, finally, inverse DCT is carried out. Then, time expenses considerably depend on the time spent on DCT since other operations are fast enough. Really, the thresholding is very simple (see the details in the next section) and data aggregation for fully overlapping blocks is based on averaging. Similarly to [14], we further consider the DCT-based denoising with fully overlapping blocks since this variant provides the highest efficiency in terms of output SNR and perception quality.

Other conclusions stemming from [14] are the following. The threshold type (hard and combined [28] thresholds were studied in [14]) has a certain impact on processing. Parameter β, used in threshold setting, is also important, since maximum processing efficiency can be observed for different (optimal) β values. Filtering efficiency

characterized by, for example, SNR improvement is larger for smaller input SNR depending on signal properties as well. Meanwhile, the analysis in [14] was carried out for a single block size, N=32. Because of this, it is unclear if the aforementioned conclusions are valid for other values of N, and how filtering efficiency depends on N.

Presentation of the main material

To study the filtering efficiency, we need noise-free test signals. Here, we use the same option as in [14], i.e. employ several (five) recorded English language Harvard phrases which are often considered as standard [29, 30]. All records are 2 s long and the sampling rate for them is 16 kHz which is considered to be the standard for high quality speech records.

One condition in filter design is that the denoising techniques (algorithms) have to be applicable and efficient for a wide range of input SNRs. Different SNRs can be simulated in various ways. To vary input SNR, we used the fixed power of signal (P_{sig}) and different intensities (powers) of AWGN (P_{noise}) added artificially to obtain the signal/noise mixture. The following SNRs have been modelled: 0, 5, 10, 15, and 20 dB where SNR in decibels is expressed as:

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{\mathrm{P_{sig}}}{\mathrm{P_{noise}}}\right) \quad (1)$$

For SNR_{dB}≤20 dB, the noise is clearly audible in noisy signals.

Here, it is worth giving details concerning 1-D DCT-based denoising. Suppose S(i), i = 1,...,I is the noise-free signal that should be estimated having an observed realization $S_n(i)=S(i) + n(i)$, i=1,...,I of signal/noise mixture where(i is the sample index, I is the total number of samples, n(i), i=1,...,I denotes the AWGN having zero mean and variance σ^2 supposed to be known in advance or accurately pre-estimated. The estimation of information signal has to be done by means of obtaining such an estimate $S_f(i)$,i=1,...,I that has to be as close as possible to the noise-free signal S(i),i=1,...,I according to a chosen criterion. A good filter (estimator) should provide the output MSE considerably less than σ^2 or improvement of another metric (criterion) chosen for a task to be solved.

For the considered filter, an l-th block includes values $S_n^{bl}(l, j)=\{S_n(l + j - 1)\}$, j= 1,...,N where, for the fully overlapping block case, l = 1,...,I-N+1 (in other words, l is the index of the leftmost sample included in a given l-th block). For each block, a direct DCT is first carried out with obtaining the DCT coefficients D(k),k=1,...,N. Note that D(1) corresponds to the block mean and the thresholding operation is not applied to it. For the hard thresholding, one has:

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

In turn, the combined thresholding presumes that even small-amplitude DCT values might contain information and, because of this, their values have to be diminished but not assigned to zeroes:

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

where, in both cases, T denotes the threshold value, which is set as βσ. Here, β is a factor that can be set by a user or determined in some other way, e.g., adaptively. For the hard thresholding, it is recommended to set β=2.7 by default [31]; for the combined thresholding (3), the recommended values of β are about 4.3 [32].

Then, the inverse DCT is applied to $D_{thr}(k)$,k=1,...,N with obtaining the denoised filtered values for the given block $S_f^{bl}(l) = \{S_f(l + j - 1)$, j= 1,...,N. As one can see, there can be from one (for the first and last blocks) to N filtered values belonging to different overlapping blocks. Below, we concentrate on the simplest variant of their processing assuming simple averaging of the obtained estimates. Other options have not resulted in significantly better outcomes.

Filtering efficiency can be characterized [33] by MSE/σ^2 ratio where:

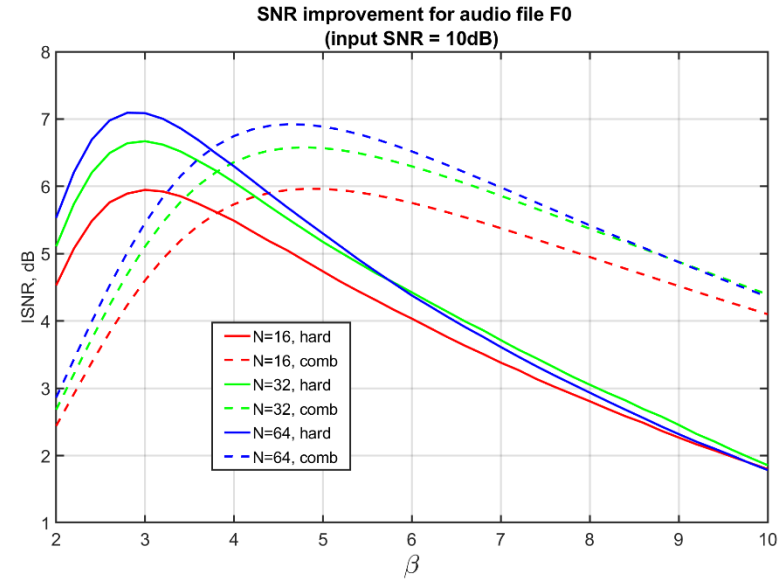
$$\mathrm{MSE} = \sum_{\mathrm{i=1}}^{\mathrm{I}} \frac{[\mathrm{S(i)} - \mathrm{S_f(i)}]^2}{\mathrm{I-1}}, \quad (4)$$

or, equivalently, by the improvement in the SNR due to filtering (expressed in dB):

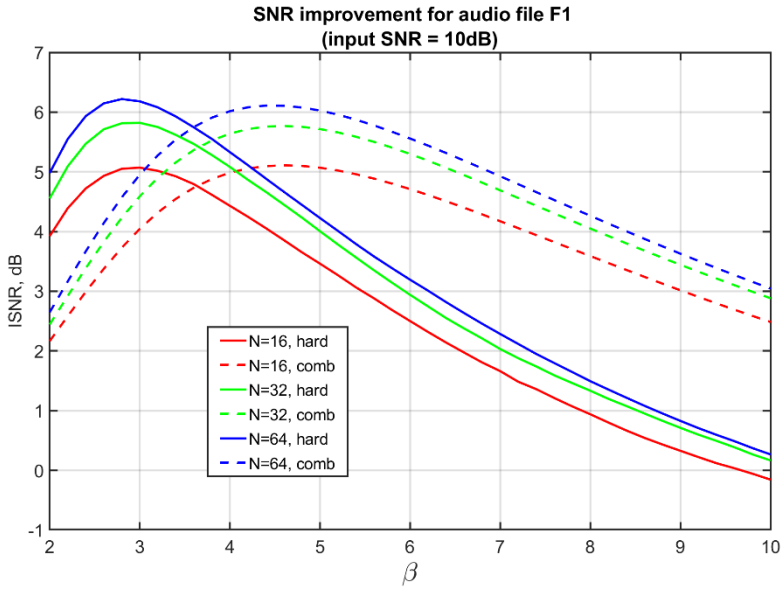
$$\mathrm{ISNR} = 10log_{10}\left(\frac{\sigma^2}{\mathrm{MSE}}\right) = SNR_{\mathrm{out}} - SNR_{\mathrm{inp}} \quad (5)$$

Since speech signals are subject to perception by humans, it is also worth using special speech quality metrics, e.g., the Perceptual Evaluation of Speech Quality (PESQ) [34] recommended by ITU-T (the standard ITU-T P.862) [35]. PESQ takes into account such aspects of speech quality as clarity, crispness and naturalness. Some details are given in [14, 34]. Here, we would like to mention the following. PESQ values are in the limits from -0.5 to 4.5 where the scale rearrangement to mean opinion score from 1 to 5 is possible. Then, the quality is considered bad for PESQ≤2.6 and for PESQ≥4.3 all listeners are very satisfied.

We have used five files with notations F0-F4. Figure 1,a shows the dependence of improvement in signal-to-noise ratio (ISNR) on the parameter β for the speech signal (file F0) for SNR=10 dB at the filter input. A set of six graphs is presented. The solid lines show the dependences for hard thresholding whilst the dashed lines – for the combined thresholding. As one can see, all dependences have maxima where they are observed for β_{opt}≈2.9 for hard thresholding and β_{opt}≈4.7 for combined thresholding. The optimal β has some tendency to increase for smaller N. Meanwhile, the results for the corresponding optimal β are always better for N=64. ISNR is large enough and reaches 7 dB for N=64 for both threshold types. The results for both threshold types are almost the same.



a)



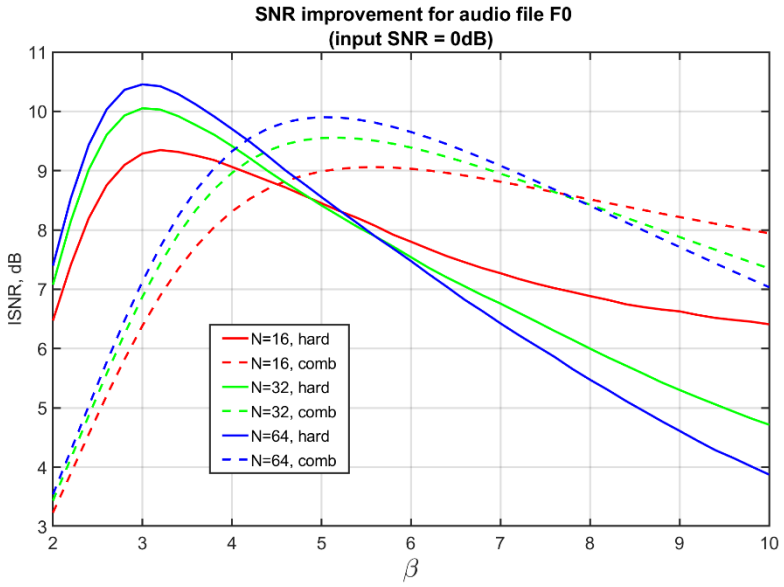
b)

Fig. 1. Dependences of ISNR on β for the files F0 (a) and F1 (b)

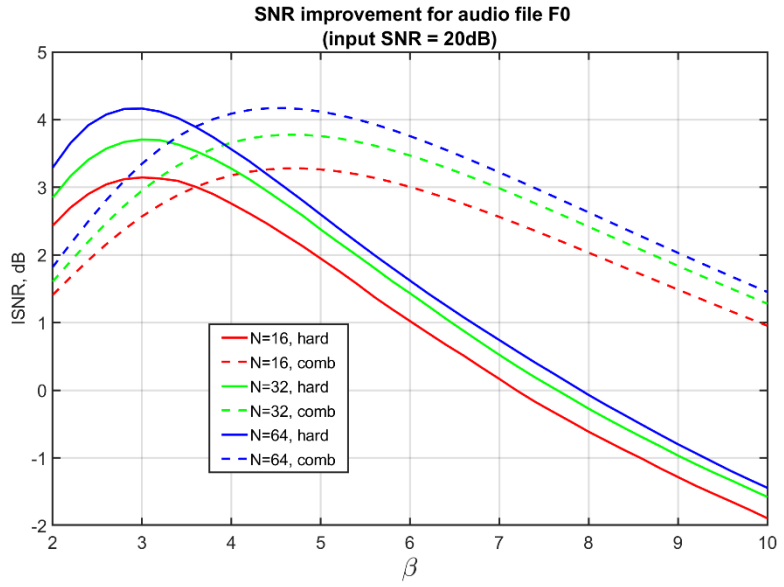
Let us check whether the results are similar for another test signal. For the audio file F1, the dependences are given in Fig. 1,b. The optimal β are slightly smaller and the maximal values of ISNR are slightly smaller too. We associate this with a more complex structure of the signal component for the file F1 compared to the file F0. For other files, the results are fall between the results for the files F0 and F1.

Consider now the results for other input SNRs. The dependences for input SNR equal to 0 dB are represented in Fig. 2,a. The corresponding optimal β are slightly larger than for the dependences in Fig. 1,a. The maximal ISNR values are larger too. The method based on the combined thresholding performs slightly worse. The dependences for the input SNR equal to 20 dB are shown in Fig. 2,b. Here, the optimal β are smaller and the maximal ISNR are smaller too. There is practically no difference in efficiency for both types of thresholding. In all cases, the results for N=64 are the best.

The dependences obtained for input SNRs equal to 0 and 20 dB for other test signals are in good agreement with the results in Fig. 2. Thus, according to ISNR, the conclusions are the following: 1) there is no significant difference what type of thresholding is applied; 2) it is reasonable to use N=64; 3) it is possible to recommend setting β about 3 for hard thresholding and about 4.8 for combined thresholding although adaptation to signal complexity and input SNR seem possible (this can be studied in the future).



a)



b)

Fig. 2. Dependences of ISNR on β for the file F0 for input SNRs equal to 0 dB (a) and 20 dB (b)

Consider now the PESQ metric. The obtained results are given in another way in Figures 3-5. First of all, horizontal solid lines of different color show PESQ values for input signals having five different input SNR: 0 dB (red), 5 dB (green), 10 dB (blue), 15 dB (black), and 20 dB (cyan). Dashed lines correspond to hard thresholding and dotted ones correspond to combined thresholding. The dependences for N=16 are presented in Fig. 3. Their analysis shows the following:

1) For any filtering (except the case of input SNR equal to 20 dB for $\beta>6$), the hard threshold DCT-based denoising leads to PESQ improvement; the combined threshold DCT based denoising improves PESQ for entire considered range of $2<\beta<10$;

2) However, the maxima for different input SNRs are observed for considerably different β; for hard thresholding, $\beta_{opt}\approx 2.8$ for input SNR equal to 10 dB or larger although for $SNR_{inp}<10$ dB the dependences might have several maxima and the global one corresponds to β_{opt} considerably larger than 2.8; for combined thresholding, optimal β also vary and they are significantly larger than according to ISNR (see the plots in Figures 1 and 2); it is possible to recommend setting $\beta\approx 7$ but adaptive setting seems reasonable as well (this can be the direction of further studies);

3) Even being filtered, the speech signals remain of poor quality for $SNR_{inp}\leq 10$ dB;

4) According to the PESQ metric, the DCT-based denoising with combined thresholding is preferable under condition that β is set properly; thus, filtering with the best ISNR is not the best solution if one wants to provide the best perception of the speech signal after denoising.

5) Note that the conclusions given above for the file F0 are in perfect agreement with conclusions for other four files.

The results for N=32 are presented in Fig.4. Here, the analysis leads to the same conclusions with the only exception. For SNR_{inp} equal to 0 and 5 dB, the combined thresholding does not lead to better results than hard thresholding. Meanwhile, despite PESQ improvement, the speech signal quality remains poor. Compared to N=16 (Fig. 3), the results for N=32 have improved.

Finally, Fig. 5 gives the dependences for N=64. For them, the conclusions are the same as above. Meanwhile, the filtering efficiency is better than for N=16 and N=32. Consider a particular case of SNR_{inp}=15 dB. For the hard thresholding with β=3, one has PESQ=2.60 for N=16, 2.73 for N=32, and 2.89 for N=64. For the combined thresholding with β=7, we have PESQ=2.81 for N=16, 3.15 for N=32, and 3.22 for N=64.

To partly prove the conclusions, Fig. 6 presents the dependences for the file F1 for N=64. Here the advantages of filtering with combined thresholding are even more obvious.

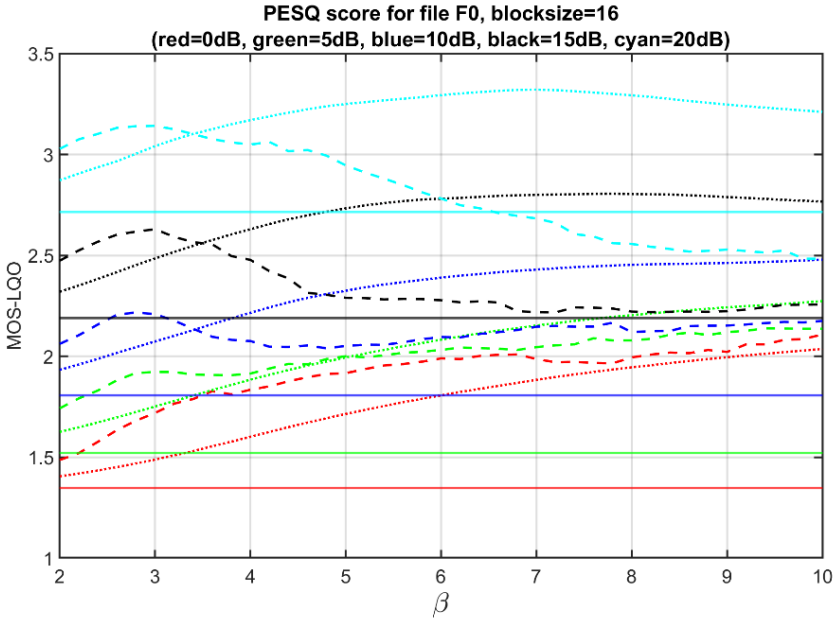


Fig. 3. Dependences of PESQ on β (file F0) for five input SNRs and two threshold types, N=16

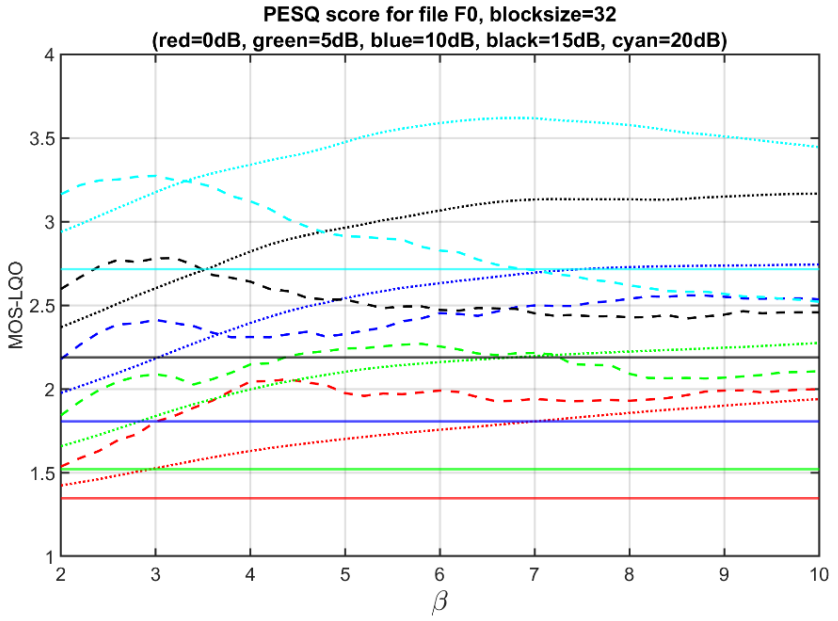


Fig. 4. Dependences of PESQ on β (file F0) for five input SNRs and two threshold types, N=32

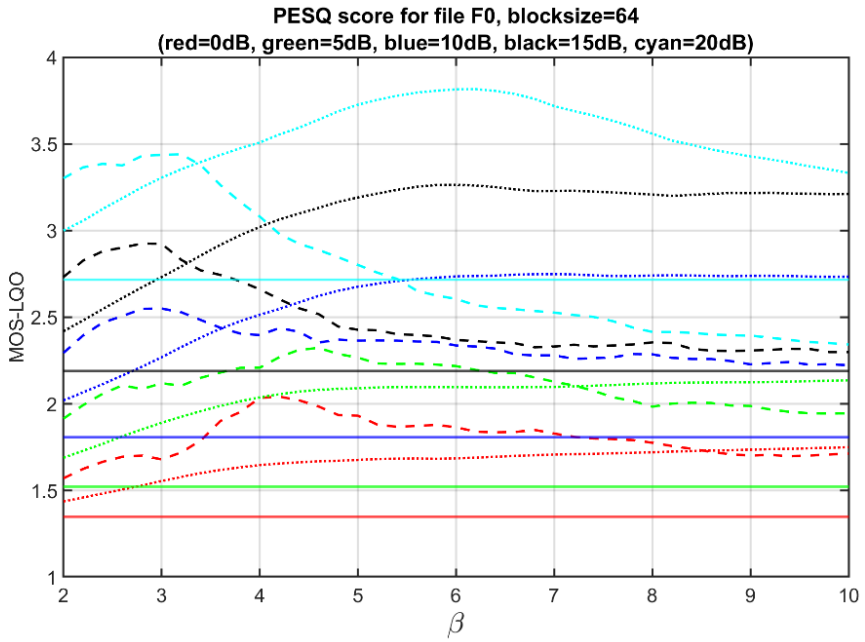


Fig. 5. Dependences of PESQ on β (file F0) for five input SNRs and two threshold types, N=64

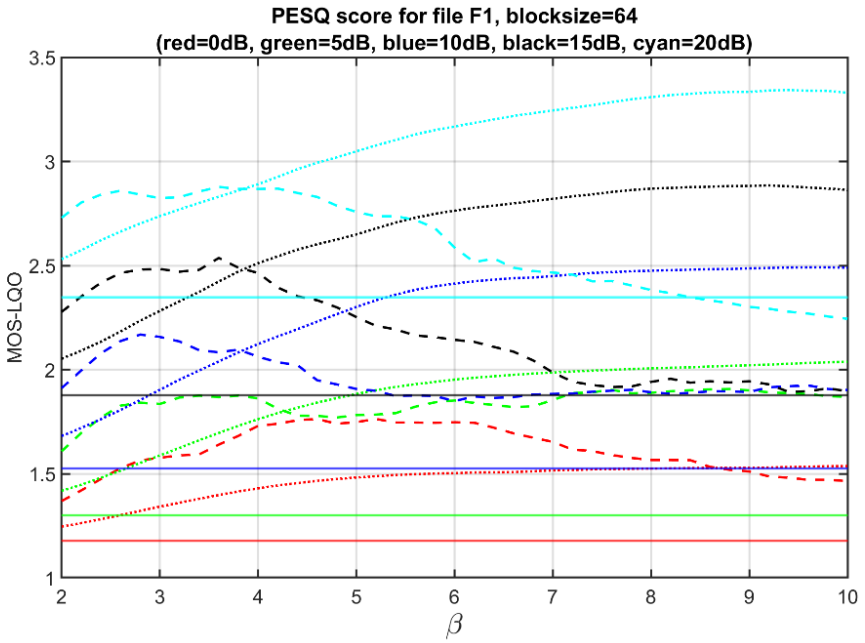


Fig. 6. Dependences of PESQ on β (file F1) for five input SNRs and two threshold types, N=64

Let us also give some filtering examples. Fig. 7 presents a 0.5 s fragment of the test signal (noise-free and noisy, SNR_{inp}=10 dB) as well as its filtered versions for N=32 and N=64 with hard and combined thresholding with optimal β. As seen, filtering is quite efficient in the sense of noise removal and signal preservation. However, the filtering results for N=64 seem to be the best.

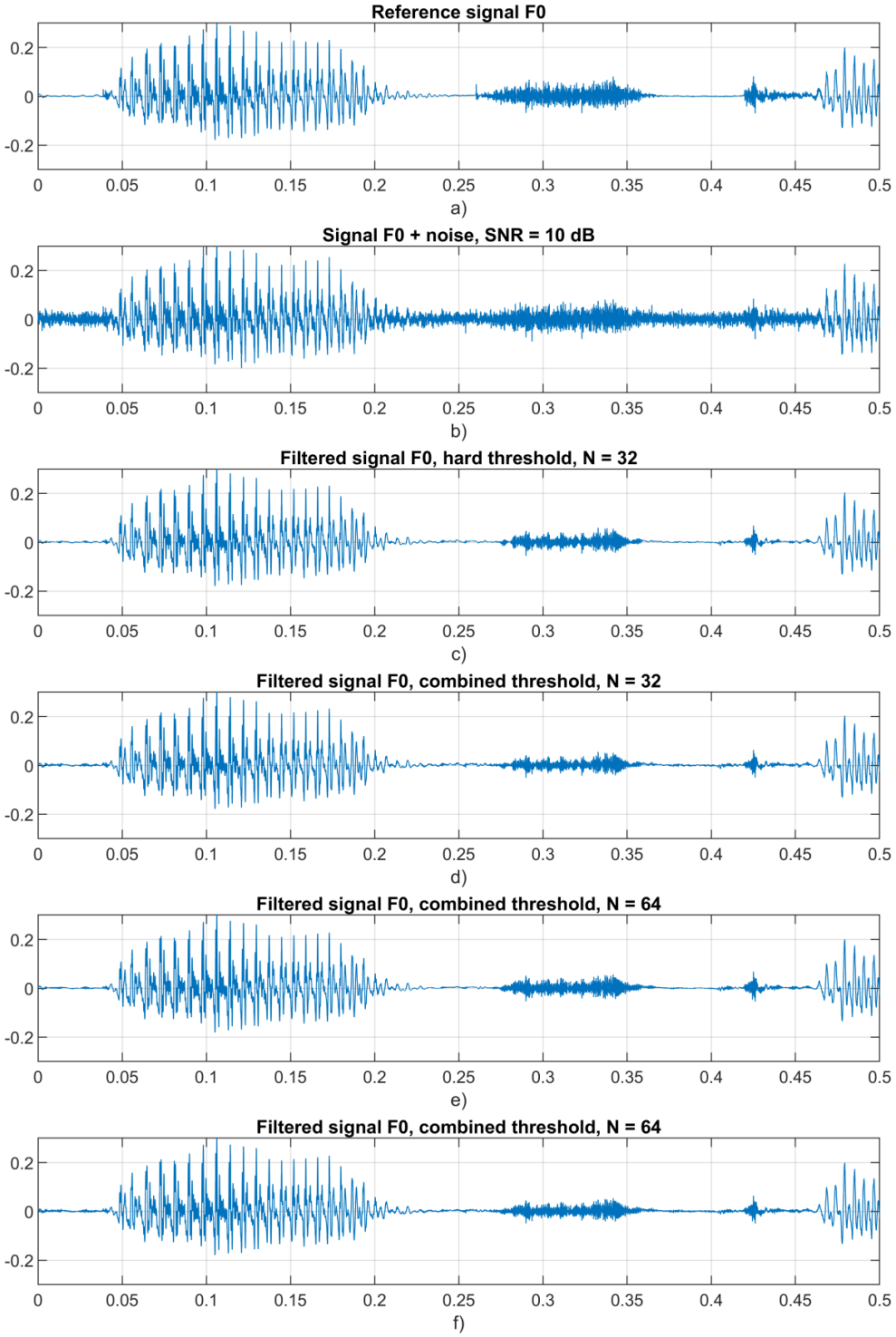


Fig. 7. Time diagrams of the F0 signal processing. a) - reference, b) - signal + noise (SNR =10dB), c) - filtered signal (N = 32, hard threshold, β = 3.0), d) - filtered signal (N = 32, combined threshold, β = 4.8), e) - filtered signal (N = 64, hard threshold, β = 2.8), f) - filtered signal (N = 64, combined threshold, β = 4.6)

Conclusions

In this paper, we have considered the task of denoising the speech audio signals contaminated by AWGN. Different versions of 1-D DCT-based filtering with fully overlapping blocks have been studied. It is shown that the use of the block size N=64 is preferable. According to the PESQ metric, the combined thresholding with β≈7 is preferable. Meanwhile, according to ISNR, both filtering approaches produce approximately the same results where the optimal β for the combined thresholding is smaller (about 5). This opens the room for adaptation to input SNR and signal complexity. Besides, other perception quality metrics are worth considering.

References

1. Hu, Y., & Loizou, P. C. (2008). Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech and Language Processing, 16*(1), 229–238. https://doi.org/10.1109/TASL.2007.911054
2. Muthu, R., & Bharath, P. (2020). Denoising of speech signal using empirical mode decomposition and Kalman filter. *International Journal of Innovative Technology and Exploring Engineering, 9*. https://doi.org/10.35940/ijitee.H6313.069820
3. Xie, X. H., & Wang, W. C. (2023). An improved LMS adaptive filtering speech enhancement algorithm. In *2023 5th International Conference on Natural Language Processing (ICNLP)*, Guangzhou, China (pp. 146–150). https://doi.org/10.1109/ICNLP58431.2023.00033
4. Vishwakarma, D. K., Kapoor, R., Dhiman, A., Goyal, A., & Jamil, D. (2015). De-noising of audio signal using heavy tailed distribution and comparison of wavelets and thresholding techniques. In *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India (pp. 755–760).
5. Ali, M. A., & Shemi, P. M. (2015). An improved method of audio denoising based on wavelet transform. In *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Thrissur, India (pp. 1–6). https://doi.org/10.1109/PICC.2015.7455802
6. Dogra, M., Borwankar, S., & Domala, J. (2021). Noise removal from audio using CNN and denoiser. In A. Biswas, E. Wennekes, T. P. Hong, & A. Wieczorkowska (Eds.), *Advances in Speech and Music Technology. Advances in Intelligent Systems and Computing* (Vol. 1320, pp. 1–12).
7. Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J. (2018). Characteristics of real-world signal to noise ratios and speech listening situations of older adults with mild to moderate hearing loss. *Ear and Hearing, 39*(2), 293–304. https://doi.org/10.1097/AUD.0000000000000486
8. Vergoz, J., Cansi, Y., Cano, Y., et al. (2021). Analysis of hydroacoustic signals associated to the loss of the Argentinian, ARA San Juan submarine. *Pure and Applied Geophysics, 178*, 2527–2556.
9. Ondusko, R., Marbach, M., Ramachandran, R., & Head, L. (2017). Blind signal-to-noise ratio estimation of speech based on vector quantizer classifiers and decision level fusion. *Journal of Signal Processing Systems, 89*. https://doi.org/10.1007/s11265-016-1200-z
10. Alexandre, D., Gabriel, S., & Yossi, A. (2020). Real time speech enhancement in the waveform domain. In *Proceedings of ISCA*.
11. Biscainho, L. W. P., Freelanci, F. P., Esquef, P. A. A., & Diniz, P. S. R. (2000). Wavelet shrinkage denoising applied to real audio signals under perceptual evaluation. In *2000 10th European Signal Processing Conference*, Tampere, Finland (pp. 1–4).
12. Beerends, J., Rix, A., & Hollier, M. (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*.
13. Lukin, V. V., Fevralev, D. V., Abramov, S. K., Peltonen, S., & Astola, J. (2008). Adaptive DCT-based 1-D filtering of Poisson and mixed Poisson and impulsive noise. In *Proceedings of LNLA*, Switzerland (p. 8).
14. Brysin, P. V., & Lukin, V. V. (2024). DCT-based denoising of speech signals. *Herald of Khmelnytskyi National University: Technical sciences*, 301–309. https://doi.org/10.31891/2307-5732-2024-339-4-48
15. Polyakova, M., Witenberg, A., & Cariow, A. (2024). The design of fast type-V discrete cosine transform algorithms for short-length input sequences. *Electronics, 13*, 4165. https://doi.org/10.3390/electronics13214165
16. Hong, H., He, M., Wang, K., & Wu, L. (2022). An image denoising method for real scene based on pixel-level noise estimation. In *Proceedings of the ACM Conference* (pp. 306–311). https://doi.org/10.1145/3569966.3570058
17. Rubel, O., & Lukin, V. (2014). Improved prediction of DCT-based filter performance using regression analysis. *Information and Telecommunication Sciences, 5*(1), 30–41. https://doi.org/10.20535/2411-2976.12014.30-41
18. Li, F., Krivenko, S., & Lukin, V. (2020). An approach to better portable graphics (BPG) compression with providing a desired quality. In *2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT)*, Kyiv, Ukraine (pp. 13–17). https://doi.org/10.1109/ATIT50783.2020.9349289
19. George, L. (2014). Audio compression based on discrete cosine transform, run length and high order shift encoding. *International Journal of Engineering and Innovative Technology (IJEIT), 4*, 45–51.
20. Abramov, S., Abramova, V., Lukin, V., & Egiazarian, K. (2019). Prediction of signal denoising efficiency for DCT-based filter. *Telecommunications and Radio Engineering, 78*(13), 1129–1142. https://doi.org/10.1615/TelecomRadEng.v78.i13.10
21. Donoho, D. L. (1995). Denoising by soft-thresholding. *IEEE Transactions on Information Theory, 41*, 613–627.
22. Jakati, J. S. (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. https://doi.org/10.30534/ijeter/2020/43862020

23. Aggarwal, R., Singh, J., Gupta, V., Rathore, S., Tiwari, M., & Khare, A. (2011). Noise reduction of speech signal using wavelet transform with modified universal threshold. *International Journal of Computer Applications, 20*, 14–19. https://doi.org/10.5120/2431-3269

24. Hidayat, R., Bejo, A., Sumaryono, S., & Winursito, A. (2018). Denoising speech for MFCC feature extraction using wavelet transformation in speech recognition system. In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Bali, Indonesia (pp. 280–284). https://doi.org/10.1109/ICITEED.2018.8534807

25. Makovoz, D. (2006). Noise variance estimation in signal processing. In *Proceedings of ISSPIT* (pp. 364–369). https://doi.org/10.1109/ISSPIT.2006.270827

26. Kharkov, A., Oliinyk, V., Lukin, V. V., & Krivenko, S. S. (2020). Blind estimation of noise variance for 1D signal denoising. *Telecommunications and Radio Engineering, 79*(7), 567–581.

27. Pogrebnyak, O., & Lukin, V. (2012). Wiener DCT based image filtering. *Journal of Electronic Imaging*(4), 14 p.

28. Fevralev, D. V., Krivenko, S. S., Lukin, V. V., Marques, R., & Medeiros, F. (2013). Combining level sets and orthogonal transform for despeckling SAR images. *Aerospace Engineering and Technology, 2*(99), 103–112.

29. IEEE Subcommittee on Subjective Measurements. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, AU-17*(3), 225–246.

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

31. Gotchev, A., Nikolaev, N., & Egiazarian, K. (2001). Improving the transform domain ECG denoising performance by applying interbeat and intra-beat decorrelating transforms. In *Proceedings of ISCAS 2001*, Sydney, NSW (Vol. 2, pp. 17–20). https://doi.org/10.1109/ISCAS.2001.920995

32. Lukin, V. (2010). Speeding up DCT-based filtering of images. In *Proceedings of TCSET'2010*, Lviv–Slavske, Ukraine (pp. 23–27).

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

34. ITU-T. (2001). 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. (2011). G.107: The E-model, a computational model for use in transmission planning. https://www.itu.int/rec/T-REC-G.107