International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Eliminating noise from a speech signal based on a pair of filters

Nilufar Niyozmatova [1, *], Kuanish Jalelov [2], Boymirzo Samijonov [3] and Mukhtaram Madrahimova [2]

[1] *Institute of Fundamental and Applied Research under the National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers", Uzbekistan.*
[2] *"Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University, Uzbekistan.*
[3] *Sejong University, South Korea, Neungdong-ro 209, Gwangjin-gu, Seoul, Korea.*

International Journal of Science and Research Archive, 2024, 13(02), 401–410

Publication history: Received on 22 September 2024; revised on 02 November 2024; accepted on 04 November 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.2.2058

Abstract

Today, the increase in the number of devices working based on speech interfaces increases the importance of speech quality. However, even a minimal noise level can seriously affect the accuracy of speech signal recognition and processing. Therefore, denoising speech signals is an important task in signal processing, and it also serves to improve speech quality in telecommunications, speech recognition systems, and other speech-related applications. Applying existing noise reduction filters separately may not always be effective. Therefore, in this research, a noise reduction approach based on the sequential application of filters is proposed. Based on literature analysis, filters such as low-pass, band-pass, Kalman, Butterworth, and elliptic filters were selected, and pairs were formed based on them. Pairs of filters were applied to speech signals with different levels of noise, and the resulting filtered speech signals were evaluated based on the PESQ evaluation criterion.

The purpose of this study is to determine the optimal pair of filters that can minimize the impact of noise on the quality of speech signals using the PESQ evaluation criterion. The results of the experiments showed that it is optimal to use a pair of band-pass and Butterworth filters at a low level of noise, a pair of low-pass and elliptic filters at a medium level, and a pair of band-pass and elliptic filters at a high level of noise. The results obtained are important and practical in the development of other hybrid methods of noise reduction.

Keywords: Noise reduction; Signal processing; Speech signals; Pair of filters; White noise; Indicator.

1. Introduction

Speech plays a central role in interpersonal communication as the main information carrier [1]. Currently, the share of information transmitted in the form of speech signals remains significant, and most of them are digital. The development of biometric systems that use speech signals for personal identification is gaining importance today [2]. In addition, speech signals are the main component of video signals. However, since speech contains both low-pass and high-pass components, it can be affected by various factors [3]. One of the main problems is the various interferences that occur in this environment, especially when it is transmitted through communication devices. Noise sources are diverse and include background noise such as street or engine noise, electrical noise caused by microphone and transmission channel imperfections, and noise such as reverberation and echo [4]. In this case, the types of noise affecting speech signals can be divided into the following several categories:

- white noise is common in speech signals, it is added to the original speech signals in a state that is evenly distributed across the frequency spectrum, which strongly affects the quality of speech;

* Corresponding author: Nilufar Niyozmatova

- pink noise – it is caused by slow changes in the characteristics of condensed materials in electronic devices. Although it is weaker than white noise, it significantly reduces the quality of speech;
- when several speech streams overlap, mixed noise is created and it complicates speech signal recognition.

The above-mentioned noises can not only reduce the quality of signals but also cause the loss of some information, which complicates the correct and complete perception of speech. Also, all the factors related to speech recognition and transmission reduce the recognition efficiency. Therefore, it is very important to minimize the noise in improving the quality of speech transmission and perception. All this allows us to conclude how relevant it is to conduct a lot of research on eliminating noise in speech signals and to gather knowledge. The main problem is the need for filtering to preserve useful information in speech signals and eliminate noise. Currently, many filters have been developed for noise reduction, each of them has its characteristics and is related to the range of practical problems being solved [5]. A single filter is usually not sufficient to achieve optimal results, and this gives rise to the idea of trying combinations of filters to reduce noise.

The purpose of this research work is to evaluate the effectiveness of various filter combinations for noise reduction in speech signals. The main task is to determine the optimal pair of filters that provide the best speech quality after noise reduction. To achieve the set goals and objectives, the following was implemented in the work:

- analysis and selection of filters. Familiarity with theoretical data on noise reduction and analysis of industry literature is the key to the successful selection of filters.
- creating pairs of filters. All possible pairs of filters are created based on the filters selected through the literature review. A combination of filters involves their sequential application to a noisy speech signal:
- applying a combination of filters to noisy speech signals.
- evaluation of the quality of processed speech signals.
- determining the optimal pair of filters.

2. Literature analysis

The purpose of this section is to review the existing research on speech noise removal using various filters. Much research has been carried out in this field, and various approaches, methods, and algorithms have been proposed to improve the quality of speech, mainly in noisy environments. Among them, noise reduction filters such as Weiner, Median, elliptic, low and high-pass, Kalman, Butterworth, and Chebyshev are common and popular. Below is information about them, with a special emphasis on each of them.

A Weiner filter is an adaptive filter that removes additive noise by minimizing the root mean square error between the original and filtered signal. This is especially effective in filtering speech signals when noise statistics are known. However, the main drawback of this filter is that it depends on the accurate estimation of noise and signal statistical properties [6,7]. A Weiner filter can lose its efficiency under dynamic noise variations or due to extremely high computational intensity in real-time.

A least-mean-squares filter reduces noise in speech signals by updating the filter coefficients to minimize the error between the incoming and the desired signal. It was noted in [8] and [9] that this filter is effective in eliminating noise such as background noise under static conditions. However, high-pass noise can significantly reduce the efficiency of the filter, which limits its use in dynamic speech conditions. In addition, one of the main disadvantages of the LMS filter is that it has a low convergence rate, which makes it unsuitable for real-time applications.

The Savitsky-Golay filter is used to smooth signals and preserve their structure by reducing noise. It works not by averaging the points in the filter window, but by approximating the points in the shift window with a polynomial of a certain degree. In the research conducted by Savitsky and Golay, it was shown that this filter is effective in smoothing signals and eliminating low-amplitude noise in signals [10]. However, they work better with continuous signals such as time series rather than with complex and rapidly changing signals such as speech.

The Chebyshev filter is an effective filter for reducing background noise in a signal, and it uses the Chebyshev polynomial, named after the famous 19th-century Russian mathematician PL Chebyshev. The main drawback of the Chebyshev filter is presented in the research paper [11], where it is noted that the use of this filter in filtering high-pass components can lead to a deterioration of speech quality.

A band-pass filter plays an important role in reducing noise in signals. Unlike a low-pass filter, it allows you to select a certain frequency range corresponding to the speech signal and eliminate low- and high-pass noise. In this case, the use of a band-pass filter to separate the frequency range suitable for speech signals significantly improved speech intelligibility [12]. Also, the effectiveness of using a band-pass filter in noisy environments such as conferences or crowded places was presented in [13], where it was noted that the filter helped to isolate only the frequencies most important for speech data and significantly reduce the impact of noise.

The median filter is a traditional impulse noise reduction filter based on non-linear filtering that replaces each value of the signal with the median of the values in the filter window. The median filter can effectively remove the impulse noise without affecting the basic structure of the signal [14]. However, it is not suitable for filtering broadband noise such as white or pink noise, which significantly limits the use of this filter in speech-processing tasks. Also, when the filter window size is made relatively large, the median filter can create distortions in the signal.

The Butterworth filter is known for its smooth amplitude-frequency characteristic, which is one of the optimal filters for noise reduction in speech signals. One of its advantages is that it provides minimal signal distortion within the passband. This is especially important when working with speech signals, where each component contains valuable information. The use of the Butterworth filter in combination with other filters in improving the quality of the speech signal through noise reduction was seen in [15], where it was found that the combination of filters for improving the quality of the speech showed good results.

The Kalman filter was first proposed for noise reduction and optimal state estimation of systems under uncertainty. It is an iterative algorithm that predicts the state of the system and adjusts it based on observations. This filter is used in speech signals to filter random and dynamic noise. A modified version of the Kalman filter for removing acoustic noise in speech signals was proposed in [16], and one of its main advantages is its ability to adapt to changing noise characteristics, which increases its ability to work in dynamic environments such as mobile communication. However, the Kalman filter has high computational complexity compared to other filters.

Wavelet-based filters are used to eliminate noise that occurs at different frequencies. They can effectively reduce noise in speech signals while preserving important information, but they have disadvantages such as setting filtering parameters and computational complexity [17]. Also, if the filtering parameters are selected incorrectly, distortions can occur in the signal.

Elliptic filters are effective in dramatically reducing noise in speech signals. Unlike Butterworth or Chebyshev filters, elliptic filters provide maximum frequency selectivity, which is especially important in tasks where it is necessary to separate the speech signal from strong noise. It was reported in [18] that elliptic filters show significantly better results in noise reduction compared to other types of filters.

Low-pass filters are traditionally used to eliminate high-pass noise in signal processing systems. Studies have shown that speech signals mainly contain information in the range of up to 4 kHz, and a significant part of high-pass noise is located outside this range. A study on speech extraction using a low-pass filter was carried out in [19], where this type of filter is particularly effective in removing high-pass noise such as hissing or electrical noise. However, the limited capability of low-pass filters in cases where the noise spreads over a wide frequency range or the speech signal contains significant high-pass information indicates the need to develop more sophisticated approaches to noise reduction. In particular, as noted in the research paper [20], there is a demand for the use of combined filters. Also, combining several filters allows for a more flexible approach to the task of noise reduction and speech quality improvement. In this study [21] and [22], the approach of sequential application of filters is described, and the experiments conducted in them conclude that the combined application of filters shows better results than the application of a single filter.

One of the standard indicators of speech quality evaluation after applying noise reduction filters is PESQ (perceptual evaluation of speech quality), which was proposed in [23] as the most accurate objective evaluation indicator related to the subjective perception of speech quality. Also, many studies in the field of noise reduction have used PESQ to evaluate the effectiveness of filters [24]. In this study, the PESQ indicator was chosen to evaluate pairs of filters. It provides an objective assessment of the quality of a speech signal compared to a clean signal, and the range of values it accepts is from -0.5 to 4.5, with a higher value indicating better speech quality.

Based on the analysis of the literature, low-pass, band-pass, Kalman, Butterworth, and elliptic filters have many advantages over other filters, and these five filters were selected for use in this research. The following table shows the results of comparing these filters with their main characteristics such as frequency range, complexity, and application.

Table 1 Results of the filter comparison

Filter name	Frequency range	Complexity	Application
Low-pass	Low-pass	Low	High-pass noise reduction
Band-pass	Specified range	Average	Just keep the range of interest
Kalman	Adaptive	High	Real-time noise filtering
Butterworth	Fixed Range	Medium	Smoothing without waves
Elliptic	Specified range	High	Highly effective noise reduction

3. Methodology

Suppose a filter set $F = \{ f_1, f_2, f_3, f_4, f_5 \}$ is given. It includes $f_1 -$ low-pass filter, $f_2 -$ band-pass filter, $f_3 -$ Kalman filter, $f_4 -$ Butterworth filter, $f_5 -$ Elliptic filter. Also, $x_{org}(t) -$ represents the noise added to the incoming speech signal and $n(t) -$. $x(t)$ is a noisy speech signal, which is generated as follows:

$$x(t) = x_{org}(t) + n(t)$$

The result of applying two filter sequences f_i and f_j to denoise a noisy speech signal $x(t)$ is denoted as $y_{ij}(t)$. And $PESQ\left(x_{org}, y_{ij}\right)$ is a speech quality evaluation function that returns $PESQ$ value for $x_{org}(t)$ clean and $y_{ij}(t)$ filtered signal.

In this research work, after adopting the above definitions, the following steps are recommended to eliminate noise in speech signals:

- **Step 1.** A noisy speech signal is generated, that is, for each $x_{org}(t)$ original speech signal, a noisy speech signal is generated as follows:

$$x(t) = x_{org}(t) + n(t)$$

- **Step 2.** Filters are applied sequentially. In this case, each (f_i, f_j) pair of filters is used to denoise the $x(t)$ noisy speech signal. As a result, a total of 5x5=25 combinations for the 5 selected filters. The sequential application of two filters f_i and f_j to $x(t)$ can be written as follows:

$$y_{ij}(t) = (x * h_i)(t) * h_j(t)$$

where $h_i(t)$ and $h_j(t) -$ are the impulse characteristics of the filters f_i and f_j, respectively.

- **Step 3.** The quality of the processed speech is evaluated. Here, the quality score for each filtered $y_{ij}(t)$ signal is calculated by comparing the value of the $PESQ$ function with the original clean $x_{org}(t)$ value:

$$PESQ_{ij} = PESQ\left(x_{org}, y_{ij}\right)$$

where $PESQ_{ij} -$ is the evaluation value of $PESQ$ for the pair of filters f_i and f_j.

- **Step 4**. The maximum value of $PESQ$ is determined.

After all pairs of filters are evaluated, 25 $PESQ_{ij}$ values are generated. The task to be performed at this step is to determine the maximum value among the possible combinations, that is:

$$(i^*, j^*) = \arg\max_{i,j} PESQ_{ij},$$

where $(i^*, j^*) -$ $PESQ_{i^* j^*}$ are the indices of filters with maximum value.

- **Step 5**. Determination of the optimal pair of filters is carried out. In this case, the optimal pair of filters is f_{i^*} and f_{j^*}, which provides the best speech quality according to the function $PESQ$, that is, the optimal pair is (f_{i^*}, f_{j^*}).

4. Result

Computational experiments used the LibriSpeech speech database, which contains clean speech signals, to apply pairs of filters to noisy speech signals and evaluate their performance. It is one of the standard datasets for various filter evaluation, speech recognition and signal processing tasks. Below is information about this database:

Data volume: about 1000 hours of high-quality speech recordings are available;
Sampling rate: 16 kHz;
Format: .wav format with 16-bit depth;
Data structure: male and female voice recordings.
To simulate noisy speech signals, this work uses 125 clean speech files from the LibriSpeech database, which are mainly 2-5 second speech signals and added different levels of noise, including 1%, 5%, 10%, 15% and 20%.

Various sequences of f_1 – low-pass, f_2 – band-pass, f_3 – Kalman, f_4 – Butterworth and f_5 – Elliptic filters selected on the basis of literature analysis were used in the computational experiment. In this case, the parameters of the filters are configured with the values listed in the table below.

Table 2 Noise reduction filters parameter values

Filter	Parameter name	Parameter value
Low-pass	Filter order	4
	Cutoff frequency	4000 Hz
Band-pass	Filter order	4
	Frequency range	500-4000 Hz
Kalman	Initial state	[0;0]
	Covariance coefficient	1.0
Butterworth	Filter order	4
	Cutoff frequency	1000 Hz
Elliptic	Filter order	4
	Cutoff frequency	1000 Hz

After applying the filters individually to different levels of noisy speech, they were evaluated with PESQ and the results are shown in the table below.

Table 3 Average PESQ values of the filters

Filter name	Average values of PESQ				
	1%	5%	10%	15%	20%
Low-pass	1,927433	1,137477	1,065398	1,050538	1,044547
Band-pass	1,840276	1,169966	1,080178	1,060205	1,055247
Kalman	1,873715	1,131599	1,060947	1,044301	1,038500
Butterworth	1,988550	1,215471	1,253763	1,170208	1,110495
Elliptic	2,001324	1,32467	1,346334	1,245065	1,171756

The table below shows the average PESQ values of pairs of filters for different noise levels.

Table 4 Average values of PESQ

Filter pair	Average values of PESQ				
	1%	5 %	10%	15%	20 %
(f_1, f_1)	1,721695	1,136433	1,076374	1,063907	1,058721
(f_1, f_2)	1,743703	1,17873	1,102569	1,083341	1,075333
(f_1, f_3)	1,855774	1,18122	1,090062	1,068486	1,05962
(f_1, f_4)	1,984431	1,424011	1,232756	1,16138	1,12629
(f_1, f_5)	1,959423	1,538009	1,341449	1,250039	1,198521
(f_2, f_1)	1,743703	1,17873	1,102569	1,083341	1,075333
(f_2, f_2)	1,486045	1,143342	1,087838	1,073582	1,07213
(f_2, f_3)	1,888117	1,230259	1,125497	1,096062	1,083632
(f_2, f_4)	2,05072	1,488074	1,339752	1,275773	1,236873
(f_2, f_5)	1,991748	1,521206	1,380035	1,315569	1,274075
(f_3, f_1)	1,650582	1,110866	1,062156	1,050282	1,044775
(f_3, f_2)	1,650372	1,110863	1,062156	1,050282	1,044775
(f_3, f_3)	1,952921	1,212948	1,100128	1,071696	1,05998
(f_3, f_4)	1,241436	1,096901	1,060689	1,049831	1,044554
(f_3, f_5)	1,637927	1,11073	1,062146	1,050281	1,044774
(f_4, f_1)	1,984431	1,424011	1,232756	1,16138	1,12629

(f_4, f_2)	2,05072	1,488074	1,339752	1,275773	1,236873
(f_4, f_3)	1,943939	1,430058	1,239757	1,16532	1,128575
(f_4, f_4)	1,792492	1,428798	1,255606	1,179127	1,13925
(f_4, f_5)	1,691798	1,428315	1,284209	1,211944	1,168787
(f_5, f_1)	1,959423	1,538009	1,341449	1,250039	1,198521
(f_5, f_2)	1,991748	1,521206	1,380035	1,315569	1,274075
(f_5, f_3)	1,912985	1,526712	1,335389	1,244773	1,19328
(f_5, f_4)	1,691798	1,428315	1,284209	1,211944	1,168787
(f_5, f_5)	1,670179	1,420391	1,293947	1,231625	1,191065

As a result of the analysis of the above table, it was found that the pair (f_2, f_4) and (f_4, f_2) are optimal for 1% noise, the pair (f_1, f_5) and (f_5, f_1) for 5%, and the pair (f_2, f_5) and (f_5, f_2) are optimal for 10%, 15% and 20% noise. The figure below shows the spectrograms of the original speech, noisy and filtered speech signals for each %.

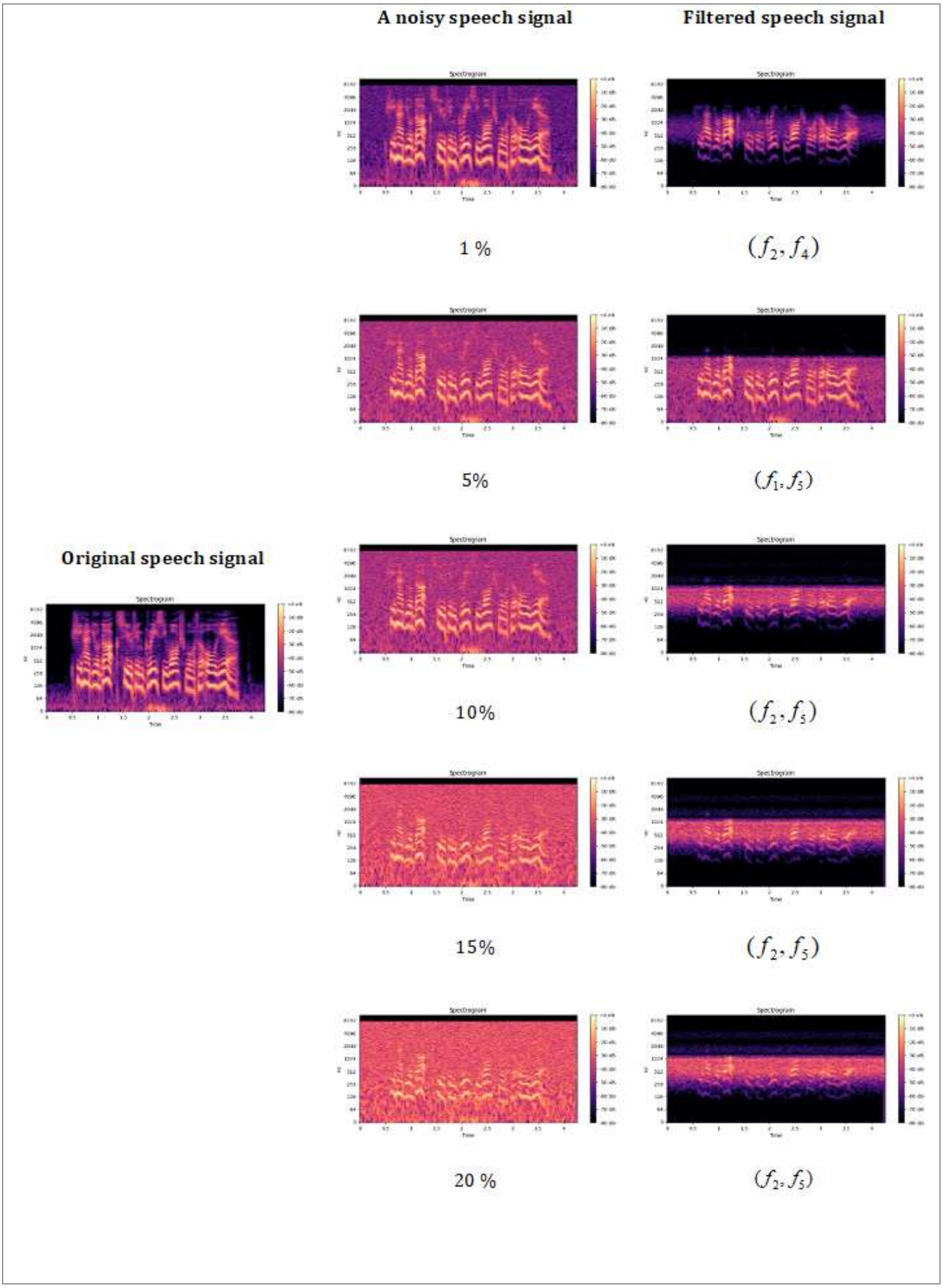

Figure 1 Spectrograms of speech signals obtained by applying a pair of filters to noisy speech signals

5. Conclusion

In this article, an approach was proposed to create pairs of noise reduction filters and use them to reduce the noise of speech signals. Five filters, namely low-pass, band-pass, Kalman, Butterworth and elliptic filters, were selected by analyzing the existing literature and pairs were formed by applying them two by one. Applying these pairs of filters to speech signals affected by different levels of noise was carried out. The PESQ indicator was used to evaluate the quality of the filtered speech signals. As a result of this research, the following conclusions were formed:

- application of noise reduction filters to noisy speech signals alone is ineffective compared to their simultaneous application;
- it was found that there is no difference in the sequential application of other filters except the Kalman filter. That is, in the pairs organized together with the $(f_{i,j}) = (f_{j,i}),\ (i, j = 1, 2, 4, 5)$ Kalman filter, the order of the filters is important and they are different from each other;
- it was found that the level of noise significantly affects the quality of the filtered signal. Because the average PESQ values of filter pairs decreased as the noise level increased;
- it was found that the sequential application of band-pass and Butterworth filters is effective at a low noise level of 1%;
- the pair of elliptic and low-pass filters was found to be effective under conditions of 5% average noise;
- It was determined that a pair of elliptic and band filters is optimal in the range of 10-20%, that is, at a high level of noise;
- Overall, this research shows that the proper selection of pairs of noise reduction filters can significantly improve the quality of speech signals in noisy environments

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Mamatov, N., Niyozmatova, N., & Samijonov, A. (2021). Software for preprocessing voice signals. International Journal of Applied Science and Engineering, 18(1), 1-8.

[2] Mamatov, N., & Dusanov, X. (2023). Shaxsni ovozi asosida tanib olish usullari. Digital transformation and artificial intelligence, 1(2), 90-95.

[3] Narzillo, M., Parakhat, N., Nilufar, N., & Abdurashid, S. (2019). Karakalpak speech recognition with CMU sphinx. International Journal of Innovative Technology and Exploring Engineering, 8(10), 2446-2448.

[4] Niyozmatova, N., Mamatov, N., Nurimov, P., Samijonov, A., & Samijonov, B. (2021). Development Software for Preprocessing Voice Signals. Research Transcripts in Computer, Electrical and Electronics Engineering, 2, 53-66.

[5] Маматов, Н. С., & Юлдошев, Ю. Ш. (1992). НУТҚНИ ТАНИБ ОЛИШ ТИЗИМЛАРИ. 3· 2018_, 38.

[6] Norbert Wiener, "The Wiener RMS (Root Mean Square) Error Criterion in Filter Design and Prediction," in Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications , MIT Press, 1964, pp.129-148.

[7] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 3, pp. 197-210, June 1978.

[8] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in 1960 IRE WESCON Convention Record, vol. 4, pp. 96-104, 1960.

[9] S. Haykin, Adaptive Filter Theory, 4th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[10] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," Analytical Chemistry, vol. 36, no. 8, pp. 1627-1639, July 1964.

[11] J. H. McClellan and T. W. Parks, "A unified approach to the design of FIR digital filters," IEEE Transactions on Audio and Electroacoustics, vol. 21, no. 2, pp. 140-146, June 1973.

[12] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no. 2, pp. 113-120, Apr. 1979.

[13] G. Lathoud and I. A. McCowan, "A sector-based approach for sound source localization with microphone arrays," in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 361-364, Apr. 2003.

[14] C. C. Huang and S. M. Yang, "A robust algorithm for estimating pitch period of speech," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, no. 1, pp. 60-65, Feb. 1979.

[15] Y. Li and Y. Zhao, "Speech enhancement based on deep neural networks and non-negative matrix factorization," IEEE Transactions on Audio, Speech, and Language Processing, vol. 24, no. 8, pp. 1441-1454, Aug. 2016.

[16] S. Chouksey and P. Rajesh, "A comparative analysis of speech enhancement techniques in non-stationary noise environment," International Journal of Computer Applications, vol. 98, no. 17, pp. 24-30, July 2014.

[17] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, pp. 674-693, July 1989.

[18] K. Park and S. Lee, "Speech enhancement based on the spectral subtraction method and a minimum mean square error criterion," IEEE Transactions on Speech and Audio Processing, vol. 7, no. 3, pp. 313-322, May 1999.

[19] L. R. Rabiner and R. W. Schafer, Theory and Applications of Digital Speech Processing. Upper Saddle River, NJ, USA: Pearson, 2010.

[20] M. H. Hayes, Statistical Digital Signal Processing and Modeling. Hoboken, NJ, USA: Wiley, 2014.

[21] J. Kim and R. M. Stern, "A study on the efficiency of various noise reduction techniques for robust speech recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 1, pp. 77-87, Jan. 2008.

[22] W. Xu and Y. Zhao, "A new approach to speech enhancement based on wavelet thresholding," IEEE Transactions on Audio, Speech, and Language Processing, vol. 23, no. 9, pp. 1468-1477, Sept. 2015.

[23] A. Rix, J. E. M. E. de Lathauwer, and J. W. Wouters, "Perceptual evaluation of speech quality (PESQ): A new method for speech quality assessment," IEEE Transactions on Audio, Speech, and Language Processing, vol. 11, no. 1, pp. 180-192, Jan. 2001.

[24] Shahhoud, F., Deeb, A. A., & Terekhov, V. I. (2024, February). PESQ enhancement for decoded speech audio signals using complex convolutional recurrent neural network. In 2024 6th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE) (pp. 1-6). IEEE.