UDC 004.8

DOI: 10.18664/ikszt.v29i3.313606

Olesia Barkovska,
Anton Havrashenko.
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

RESEARCH OF THE IMPACT OF NOISE REDUCTION METHODS ON THE QUALITY OF AUDIO SIGNAL RECOVERY

Abstract.

The subject of the study is the analysis of various filtering algorithms for the quality of the resulting audio files. The importance of audio line filtering has grown significantly in recent years due to its key role in a variety of applications such as speech reduction and artificial intelligence. Taking into account the growing demand for solving problems related to speech recognition, the processing of audio series becomes important for determining the accuracy and efficiency of the obtained solution.

The purpose of the work is to study the impact of noise suppression methods on the quality of restoration of an audio signal, which was alternately noisy with one of five types of noise - white, pink, brown, impulse, Gaussian with different power. To achieve the goal, the following tasks were solved: an analysis of the types of noise was carried out and analysis of noise reduction and filtering methods. A generalized model of noise reduction and filtering was developed, and an experiment was planned depending on the type and power of noise. Simulation of the experiment was performed by comparing the parameters of the signal-to-noise ratio before and after the experiment and the peak signal-to-noise ratio in the processed file. The following methods are used: spectral subtraction, filtering based on frequency filters and wavelet transformation.

The following results were obtained: depending on the selected noises and algorithms, it was possible to achieve the lowest value of the peak signal-to-noise ratio of 21.52db, and the signal-to-noise ratio increased, which allowed further work with these audio files. The practical significance of this work is the increase in the number of available audio files for further work.

Conclusions: the analysis of the obtained results showed that filtering based on frequency filters only worsened the output signal, that is, not only noise, but also useful information is filtered. In all runs, the SNR deteriorates to -18dB. which is worse than no filtering. Algorithms of spectral subtraction and wavelet transformation improved SNR parameters and output audio files noisy with the most powerful noises in the range of 20dB, which can be considered acceptable for further processing. The results highlight the importance of using denoising and filtering for complex audio processing tasks, particularly neural network training tasks.

Key words: *noise suppression, filtering, audio, noise, SNR, PSNR, spectral subtraction, frequency filters, wavelet transform, experiment*

Introduction

In today's world, speech recognition is becoming increasingly important as a key technology in many aspects of our lives. From user interfaces to security systems, from audio and video transcription to interacting with electronic devices using voice commands, automated speech recognition is becoming a necessary element of our digital lives.

Language analytics covers a wide range of technologies and methods that allow processing and analysis of speech information. One of the key technologies in this field automatic speech recognition[1]. This technology converts spoken speech into text, which has numerous applications in various industries, from captioning to interactive voice assistants.

Text-to-Speech[2] is another important aspect of speech analytics. This technology allows you to convert text data into natural speech, which is used in the creation of audio books, interactive customer support systems, as well as in assistive technologies for people with visual impairments.

In addition, the analysis of emotions in the voice is becoming more and more popular. This technology uses machine learning algorithms to determine a person's emotional state based on their voice. This can be useful in the fields of psychology, healthcare, and customer service, where a customer's emotional state can affect the quality of service.

Language analytics also includes automatic conversation analysis. Such systems can provide important information to improve customer service and optimize business processes.

However, the accuracy of speech recognition can significantly depend on the quality of the input audio signal and the efficiency of signal processing algorithms before its analysis by neural networks. One of the main challenges in this context is managing the noise that may occur during audio recording (for example, noise from background music, conversations or the noise of household appliances).

Modern trends in the development of speech recognition technologies include the use of deep neural networks[3] and machine learning methods, which have significantly increased the accuracy and reliability of systems. However, even with the most advanced algorithms, noise in the input data remains a significant obstacle.

Thus, in this study, we focus our attention on investigating the impact of different audio preprocessing algorithms on the quality of speech recognition using neural networks. The choice of optimal signal processing methods before further analysis can significantly improve the effectiveness of automatic speech recognition systems in conditions of noise and interference. We will also look at current approaches to noise removal and their effectiveness in real-world use cases.

The purpose of our research is to develop and test pre-processing algorithms that will reduce the noise level in input audio signals and increase the accuracy of speech recognition. Tasks include comparing existing methods, developing new approaches, and evaluating their performance on different data sets.

It is expected that the results of the research will make a significant contribution to the development of speech recognition technologies, which will improve the quality of user interaction with various digital systems and increase the overall effectiveness of these technologies in everyday life.

STATE OF THE ART

In works [4-6], a significant amount of research was conducted aimed at improving the quality of speech recognition using neural networks and the impact of various audio signal processing methods on recognition accuracy. The results show that the noise present in the audio sequences has a significant impact on the recognition accuracy. The classification of noise types is shown in (fig. 1).

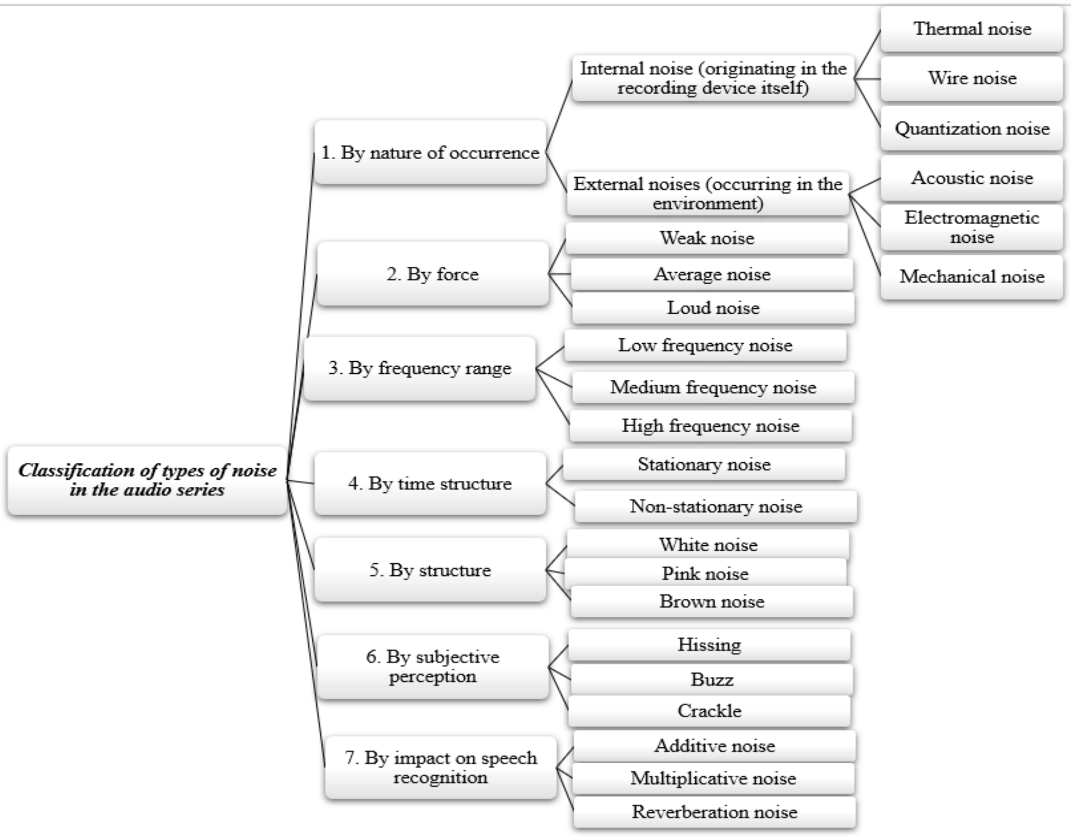


Figure 1 – Noise classification

Further studies analyze the effect of white, pink, brown, impulse, and Gaussian noise on recognition accuracy, because these types of noise were chosen due to their wide use in various fields of science and technology for modeling and testing systems under various noise exposure conditions.

Noise suppression [7] consists in actively reducing unwanted sounds or signals. It is applied in real time, using algorithms to identify and eliminate noise, leaving a useful signal. The primary purpose of noise reduction is to remove background noise or interference, such as hum, hiss, or extraneous sounds, to improve the intelligibility or clarity of the underlying signal, such as a voice in a telephone conversation or music.

In order to understand which noise reduction methods will be the most effective, it is necessary to understand their differences. The paper offers a comparative analysis of selected types of noise according to the following criteria: frequency spectrum, power spectral density, acoustic perception, application. The comparison is shown in the Table1.

Table 1. – *Noise comparison*
.

Comparison criteria	**White noise**	**Pink noise**	**Brown noise**	**Impulse noise**	**Gaussian noise**
Frequency spectrum	Even, all frequencies with equal intensity	The power spectral density drops by 3 dB per octave	The power spectral density drops by 6 dB per octave	Consists of short, intense bursts of sound that occur due to sudden changes in the signal	The amplitude is distributed according to a normal (Gaussian) distribution.
Power spectral density	Constant density at all frequencies	Density is inversely proportional to frequency (1/f)	Density is inversely proportional to the square of the frequency ($1/f^2$)	Uneven character, with peaks on pulses	Even frequency spectrum
Acoustic perception	"Sharp" and "noisy" sound	A more natural and soft sound, similar to rain	A deep and soft sound, similar to heavy sea waves or thunder	Sharp, intense sounds	A hiss that has no orderly structure or rhythm
Application	Audio equipment testing, sound masking, sleep aid, concentration improvement	Audio engineering, acoustics testing, relaxation, sleep aid	Sound masking, audiological experiments, relaxing background	Security systems to detect intrusions or other abnormal events	Modeling and testing of communication systems

Filtering is a signal processing process that allows or blocks certain frequencies or frequency ranges. It works on the principle of selecting the desired frequencies or reducing unwanted frequencies that can cause noise or distortion. Filtering can be done in a variety of ways, such as low-pass, high-pass, or band-pass filtering, and is applied not only to audio, but also to other types of signals, such as radio, images, or data.[8]

So, the main difference between noise suppression and filtering lies in their approaches and mechanisms:

- noise suppression focuses on active noise detection and removal;
- filtering adjusts the frequency composition of the signal, allowing or blocking certain frequencies.

Among the existing methods of noise filtering and noise suppression we can distinguish:

- wavelet denoising;
- homomorphic filtering;
- singular value decomposition filtering;
- neural network-based denoising;
- least mean squares filter;
- spectral subtraction;
- bilinear filtering;
- non-linear noise reduction;
- time-frequency domain filtering.

In this work, wavelet transform, spectral subtraction and filtering were selected as different methods that are applied to different types of noise, and their comparison will help to choose the best one for this task.

Spectral subtraction to remove noise

Spectral subtraction is a simple but effective method of removing noise from audio signals. It is based on the assumption, that the spectrum of the noise differs from the spectrum of the useful signal.[9]

First, the spectrum of both the noisy and the clean signal is calculated. This can be done using methods such as the Fourier transform. Then the noise spectrum is determined. This can be done in various ways, for example, using a noise profile obtained from a clean segment of the signal, or assuming that the noise is concentrated in certain frequency ranges. The noise spectrum is subtracted from the spectrum of the noisy signal. This is done component by component, that is, for each frequency. Finally, a reconstructed signal is obtained from the modified spectrum using the inverse Fourier transform.

Filtering based on frequency filters is a general technique for removing noise from signals that uses specialized filters to suppress unwanted frequency components. This method is flexible and powerful. It can be applied to a variety of signal types, including audio, images, and sensor data.

First, a suitable frequency filter is selected. The type of filter depends on the type of noise and the characteristics of the signal. For example, you can use a high-pass filter to remove low-frequency noise, and a low-pass filter to remove high-frequency noise.

The filter is then applied to the noisy signal. This results in the suppression of unwanted frequency components of the noise, leaving a useful signal.

There are several types of frequency filters used to suppress noise:

– FIR (Finite Impulse Response) Filters: These filters are simple to implement and computationally efficient.

– IIR (Infinite Impulse Response) filters: These filters can provide sharper noise suppression, but they are more difficult to implement.

Adaptive Filters: These filters can automatically adapt to the characteristics of the noise, making them useful for removing non-stationary noise.

Wavelet transform to remove noise

The wavelet transform is a powerful signal analysis and processing technique that can be used to remove noise from various types of data, including audio, images, and sensor data.[10]

Unlike traditional filtering techniques that work in the frequency domain, the wavelet transform uses time-localized functions called wavelets to analyze the signal at different scales. This allows it to effectively remove noise that has a local time structure without affecting the useful signal.

The noisy signal is decomposed into wavelet components using the wavelet transform. This gives an idea of the signal at different time and frequency scales.

Wavelet components likely to correspond to noise are identified. This can be done using various methods such as thresholding or statistical analysis.

The determined noise components of the wavelets are removed or modified.

A cleaned signal is recovered from the modified wavelet components using the inverse wavelet transform.

A comparative analysis of selected methods of noise filtering in audio sequences is given in Table 2.

Table 2. – *Comparison of filtering and noise suppression algorithms*

	Spectral subtraction	Filtering based on frequency filters	Wavelet transform
Principles of work	First, the noise spectrum is estimated, which is then subtracted from the signal spectrum to reduce the noise.	Uses filters to select certain frequency components of the signal. Band-pass, high-pass, or low-pass filters are often used to remove noise.	Breaks the signal into components of different scales or levels. These components can be analyzed and modified to remove noise.
Advantages	Simplicity: Easy to implement and understand. Performance for stationary noise: Works well for removing continuous, stationary noise. Removes noise without significantly changing the spectral structure of the	Ease of implementation: Widely used and easily implemented using DSP (Digital Signal Processing) libraries. Performance: Works well for removing noise in specific frequency ranges, such as low-frequency hum or high-frequency noise.	Allows you to analyze the signal at different scales, which helps remove both high-frequency and low-frequency noise. Works well with non-stationary signals: Effective for processing signals with variable frequency characteristics.

	signal.	Frequency Band Control: Allows you to fine-tune the frequency bands to be kept or deleted.	Wavelets provide good locality in time and frequency, which helps preserve important details of the signal.
Disadvantages	Residual Noise: May leave artifacts known as musical tones. Sensitivity to noise estimation: Incorrect estimation of the noise spectrum can degrade the quality of the reconstructed signal. Not good for non-stationary noise: Does not work well with time-varying noise.	Useful signal losses: Can remove useful frequencies along with noise, especially if the noise and useful signal frequencies overlap. Poor performance for broadband noise: Limited performance for noise covering a wide frequency range. Can cause phase distortion: Incorrect filter settings can cause phase distortion in the signal.	Complexity: More complex to implement compared to simple methods such as filtering. Wavelet selection: Requires correct selection of wavelet type and decomposition level, which may not be obvious. Computational cost: May require significant computational resources, especially for large signals or high levels of decomposition.

Each method has its advantages and disadvantages, making them suitable for different signal types and noise conditions. Spectral subtraction is effective for frequency noise isolation, filtering is useful for basic noise reduction in simple conditions, while wavelet transform provides the best signal quality in complex and non-uniform noise environments.

AIMS AND TASKS OF THE WORK

The main goal of the article is to study the impact of noise suppression methods on the quality of restoration of an audio signal that was alternately noisy with one of five types of noise - white, pink, brown, impulse, Gaussian with different power.

To achieve the set goal, the following tasks must be solved:

- comparative analysis of types of noise in an audio file, noise reduction methods and filtering methods;
- creation of a working dataset for further research;
- development of the methodology of the experiment;
- performing a study of the effect of spectral subtraction, frequency filtering and wavelet transformation on the quality of audio file recovery;
- analysis of the obtained results.

The conducted experiments are the basis for further research on the influence of filtering and noise suppression methods on the accuracy of speech and voice recognition based on neural network models.

RESULTS AND DISCUSSION

The methodology for conducting the experiment, necessary to achieve the goal, is as follows - first, a working dataset was prepared by noise-free input audio files. Each of the noises had two different variants - powerful and not powerful. To determine the noise power, we will use the value of the signal-to-noise ratio (SNR). Not powerful noise, this is the kind of noise with an SNR value close to 50db. It can vary depending on the type of noise. Powerful noise is noise with a negative SNR value, i.e. noise with a power slightly greater than the useful signal. This is not done for impulse noise, due to its peculiarities, when the noise power increases, the SNR value does not increase significantly.

The created working dataset for the research has the following structure:

- 25 noise-free audio files;
- 250 noisy audio files (one of five types of noise - white, pink, brown, impulse, Gaussian with different power) is applied to each of the noiseless ones.

The audio signals will then be pre-processed using various denoising and filtering algorithms, such as spectral subtraction, wavelet transform and band-pass filtering. After that, the processed signals will be analyzed, and the noise removal quality will be measured. The results of the experiments will be evaluated using the signal-to-noise ratio and peak signal-to-noise ratio (PSNR) metrics. For each of the noises, non-powerful noise (with a high positive SNR in the uncleaned file) and powerful noise (with a negative SNR value) will be taken. The step-by-step model is visualized in the (fig. 2).

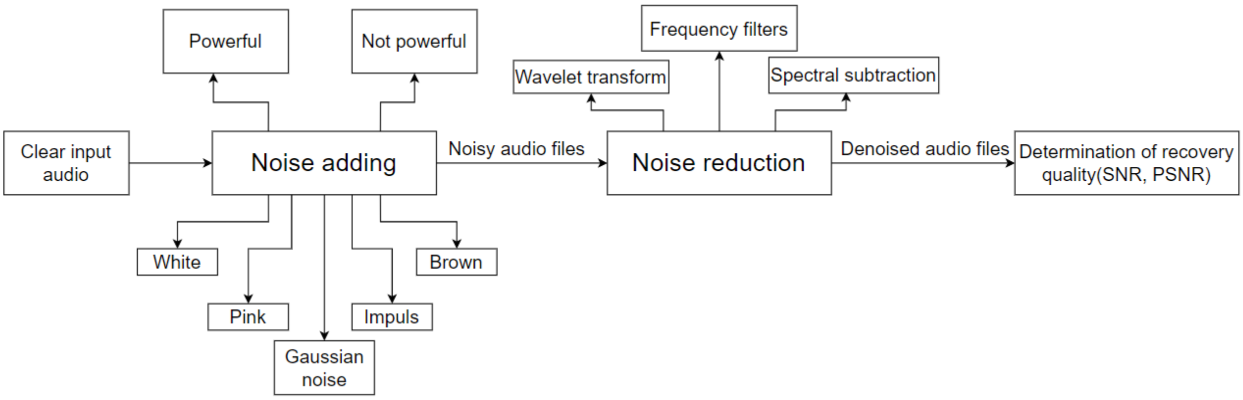


Figure 2 – Functional diagram of the proposed system

SNR, or signal-to-noise ratio, is a metric used to compare the level of a useful signal to the level of background noise. This metric is usually measured in decibels. A higher SNR value indicates that the signal is much stronger compared to the noise, resulting in better signal reception and processing. It depends on the ratio of signal power to noise.[11]

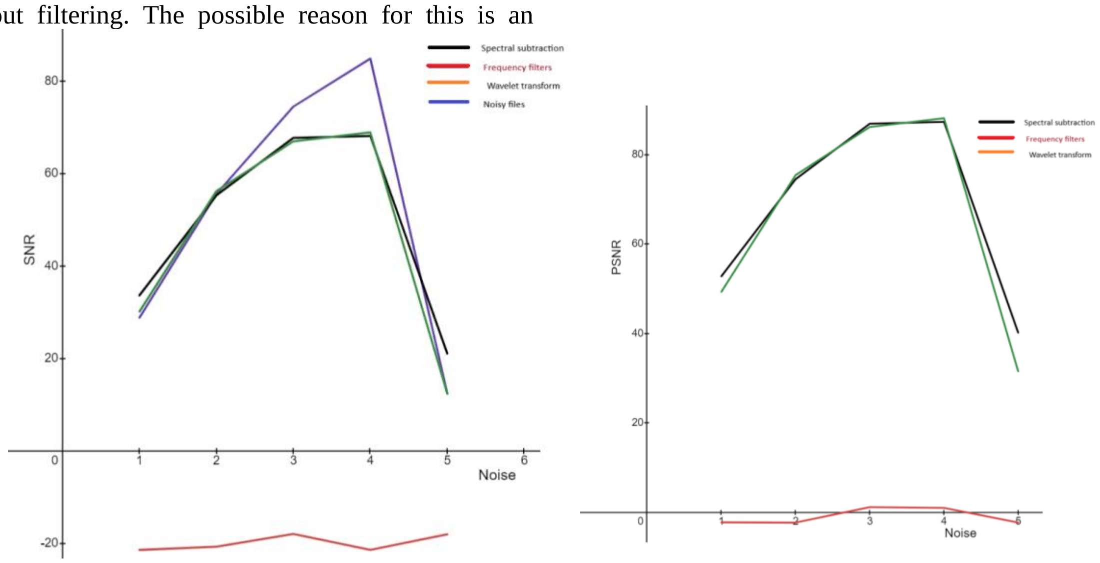
PSNR is commonly used in audio, image and video signal processing. In the context of audio processing, PSNR helps determine the degree to which audio is distorted after certain processes, such as compression or passing through noisy channels. It depends on the square of the maximum value of the amplitude and the root mean square error between the original and the processed signal. A normal PSNR level for further audio work is generally considered to be between 20 and 50 dB, depending on the specific application and audio quality requirements. At 20dB, the sound may contain noticeable distortion, but can still be understood and used in less demanding applications. These are standards for telephone communication where some level of noise and distortion is acceptable. Results are shown in table 3.

Table 3. – *Data subsets*

Experiment	SNR before (Db.)	Spectral subtraction		Frequency filters		Wavelet transform	
		SNR after (Db.)	PSNR(Db.)	SNR after (Db.)	PSNR(Db.)	SNR after (Db.)	PSNR(Db.)
White not powerful	28.86	33.67	52.82	-21.35	-2.19	30.18	49.34
White powerful	-3.16	5.35	24.50	-21.41	-2.25	3.91	23.07
Pink not powerful	55.61	55.35	74.50	-20.72	-1.56	56.25	75.42
Pink powerful	-4.78	3.50	22.64	-21.41	-2.25	-2.95	16.21
Gauss not powerful	74.53	67.77	86.92	-17.95	1.21	67.01	86.17
Gauss powerful	-0.67	7.70	26.85	-21.40	-2.24	6.12	25.28

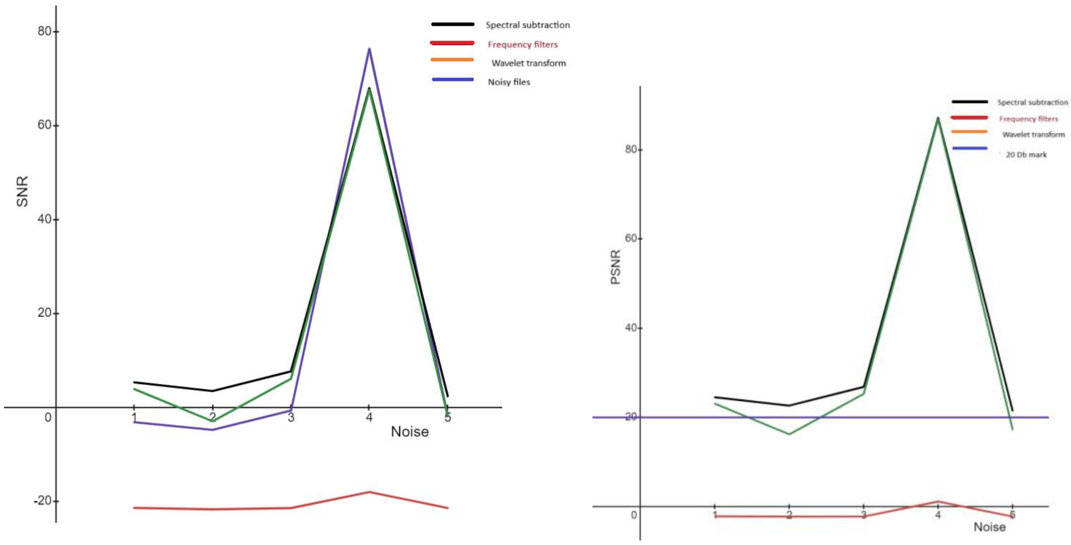
Impulse not powerful	84.93	68.18	87.32	-18.13	1.03	68.96	88.13
Impulse powerful	76.45	68.04	87.19	-18.04	1.12	67.68	86.84
Brown not powerful	12.42	21.07	40.22	-21.40	-2.24	12.43	31.59
Brown powerful	-1.86	2.38	21.52	-21.44	-2.28	-1.85	17.31

Analysis of the obtained results showed that filtering based on frequency filters only worsened the output signal, that is, not only noise was filtered, but also useful information. In all runs, the SNR deteriorated to -18 - -22dB, which is worse than without filtering. The possible reason for this is an error in the selection of the filter, but to change this, you need to use other algorithms, such as adaptive filters, which will help preserve the useful information of the audio file. A comparison of solutions is shown in (fig. 3) and (fig 4).



a) b)

Figure 3 – Graphs of changes in values for weak noise a) SNR, b) PSNR



a) b)

Figure 4 – Graphs of changes in values for powerful noise a) SNR, b) PSNR

Comparing the SNR values for low-power noise, we can see that the algorithms give approximately non-variable data. They find in the range from -20% to +15% depending on the type of noise. When analyzing the values for powerful noises, we see a change from a negative value to a positive one, which means that the useful signal began to dominate the noise. The average value changed by 280% (not including the impulse noise, because its SNR values were not negative).

Comparing spectral subtraction and wavelet transformation, we can conclude that both algorithms improve audio quality and increase SNR. In addition, even with powerful noises, they have a fairly high PSNR. However, we can see that with the powerful brown and powerful pink wavelet, the transformation failed to raise the PSNR level to the recommended 20dB, so we will consider spectral subtraction more suitable for further work.

Conclusions

In this paper a comprehensive analysis of the impact of the use of noise reduction and filtering on the quality of processing noisy audio files was carried out. As part of the research, it was possible to achieve the set goals and solve the set tasks.

A detailed study and comparison of filtering, spectral subtraction, and wavelet transform algorithms was conducted. This made it possible to determine which techniques are best suited for noisy audio data.

Different types of noise have been classified as white noise, pink noise, brown noise, Gaussian noise and impulse noise.

As a result, we found that filtering based on frequency filters only worsened the output signal, that is, not only noise is filtered, but also useful information. In all runs, the SNR deteriorates to -18 - -22dB, which is worse than without filtering.

Algorithms of spectral subtraction and wavelet transformation improved SNR parameters and output audio files noisy with the most powerful noises in the range of 20dB, which can be considered acceptable for further processing. The results highlight the importance of using denoising and filtering for complex audio processing tasks, particularly neural network training tasks.

Considering the obtained results, it is possible to recommend spectral subtraction as the most effective tool for solving the given problem, especially when the appropriate computing resources are available. However, wavelet transforms remain a reliable option.

This study made it possible to use for training neural networks not only audio files without noise, but also with them. This will increase the sample of available input and test data for further research.

The scientific value of the work lies in deepening the understanding of the mechanisms underlying noise reduction and filtering and their impact on audio data processing. The presented conclusions can serve as a basis for further research in the field of machine learning and the development of intelligent systems, which will contribute to progress in the field of artificial intelligence.

References

1. MALIK, Mishaim, et al. Automatic speech recognition: a survey. Multimedia Tools and Applications, 2021, 80: 9411-9457.
2. KUMAR, Yogesh; KOUL, Apeksha; SINGH, Chamkaur. A deep learning approaches in text-to-speech system: a systematic review and recent research perspective. Multimedia Tools and Applications, 2023, 82.10: 15171-15197.
3. HAN, Wei, et al. Contextnet: Improving convolutional neural networks for automatic speech recognition with global context. arXiv preprint arXiv:2005.03191, 2020.
4. Barkovska O., Havrashenko A. Analysis of the influence of selected audio pre-processing stages on accuracy of speaker language recognition // Innovative technologies and scientific solutions for industries. – 2023. – №. 4 (26). – С. 16-23. https://doi.org/10.30837/ITSSI.2023.26.016
5. Havrashenko A., Barkovska O. Analysis of text augmentation algorithms in artificial language machine translation systems //Advanced Information Systems. – 2023. – Т. 7. – №. 1. – С. 47-53. https://doi.org/10.20998/2522-9052.2023.1.08
6. Barkovska O., Havrashenko A. Analysis of word search algorithms in the dictionaries of machine translation systems for artificial languages //Computer systems and information technologies. – 2023. – №. 2. – С. 17-24. https://doi.org/10.31891/csit-2023-2-2
7. DUBEY, Harishchandra, et al. Icassp 2023 deep noise suppression challenge. IEEE Open Journal of Signal Processing, 2024.
8. PATIDAR, Pawan, et al. Image de-noising by various filters for different noise. International journal of computer applications, 2010, 9.4: 45-50.
9. KARAM, Marc, et al. Noise removal in speech processing using spectral subtraction. Journal of Signal and Information Processing, 2014, 2014.
10. KUMAR, Ashish, et al. Stationary wavelet transform based ECG signal denoising method. ISA transactions, 2021, 114: 251-262.
11. DARRAS, Kevin FA, et al. High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. PeerJ, 2020, 8: e9955.

Дослідження впливу методів шумопригнічення на якість відновлення аудіосигналів

Барковська Олеся, Гаврашенко Антон

Анотація.

Предметом дослідження є аналіз різних алгоритмів фільтрації на якість результуючих аудіо файлів. Значущість фільтрації аудіо ряду помітно зросла в останні роки завдяки її ключовій ролі в різноманітних застосуваннях, таких як зменшення розпізнавання мови та штучний інтелект. З урахуванням зростаючого попиту на рішення задач пов’язаних з розпізнаванням мови, обробки аудіо ряду стає важливою для визначення точності та ефективності отриманого рішення.

Метою роботи є дослідження впливу методів шумопригнічення на якість відновлення аудіосигналу, що попередньо був зашумлений одним із п’яти видів шумів - білий, рожевий, коричневий, імпульсний, гаусівський із різною потужністю. Для досягнення поставленої мети були вирішені наступні **завдання**: було проведено аналіз типів шумів та аналіз методів шумоподавлення та фільтрування, а також було розроблено узагальнену модель шумоподавлення та фільтрування і сплановано експеримент у залежності від типу та потужності шуму. Моделювання експерименту виконано шляхом порівняння параметрів співвідношення сигнал/шум до та після експерименту та пікове співвідношення сигналу до шуму в обробленому файлі . Використані такі **методи**: спектральне віднімання , фільтрація на основі частотних фільтрів та вейвлет-перетворення.

Отримані наступні **результати**: у залежності від обраних шумів та алгоритмів, вдалося досягти найнижчого значення пікове співвідношення сигналу до шуму у 21.52дб, та збільшувало співвідношення сигнал/шум що дозволило подальшу роботу з цими аудіофайлами. Практичною значущістю даної роботи є збільшення кількості доступних аудіо файлів для подальшої роботи .

Висновки:

Аналіз отриманих результатів показав, що фільтрація на основі частотних фільтрів лише погіршувала вихідний сигнал, тобто фільтрувася не лише шум, а і корисна інформація. У всіх запусках SNR погіршувадася до -18дб. що гірше ніж без фільтрування. Алгоритми спектрального віднімання та вейвлет перетворення покращили параметри SNR та вивели аудіофайли зашумлені найпотужнішими шумами у діапазон від 20дБ, що може вважатися сприйнятним для подальшої обробки. Результати підкреслюють важливість використання шумоподавлення та фільтрування для складних завдань обробки аудіо, зокрема у задачах навчання нейромереж.

Ключові слова: шумопригнічення, фільтрація, аудіо, шум, SNR, PSNR, спектральне віднімання, частотні фільтри, вейвлет перетворення, експеримент.

ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

Барковська Олеся Юріївна - кандидат технічних наук, доцент кафедри “Електронно обчислювальних машин”, Національний університет радіоелектроніки «ХНУРЕ», Харків, Україна;

Гаврашенко Антон Олегович – асистент кафедри “Електронно обчислювальних машин”, Національний університет радіоелектроніки «ХНУРЕ», Харків, Україна;

Barkovska O. – Ph.D (Engineering Sciences), Docent, Associate Professor Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. e-mail: olesia.barkovska@nure.ua; ORCID ID: https://orcid.org/0000-0001-7496-4353.

Havrashenko A. – Professor Assistant at Department of Electronic Computer, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.
e-mail:anton.havrashenko@nure.ua; ORCID ID: . https://orcid.org/0000-0002-8802-0529