

Methods of information systems synthesis

UDC 004.357:004.415:004.9

doi: <https://doi.org/10.20998/2522-9052.2026.4.03>

Olesia Barkovska, Denys Storchai, Andriy Kovalenko, Yuri Romanenkov, Vadym Trunov

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

HARDWARE-SOFTWARE PROTOTYPE OF A VOICE NAVIGATION SUBSYSTEM IN NOISY ENVIRONMENTS

Abstract. The work is devoted to the development of the hardware-software prototype of a voice navigation subsystem intended for assistive navigation under acoustic interference. The relevance of the topic is explained by the need to improve the accuracy of generating voice or haptic prompts regarding the recommended route for visually impaired users based on their voice-based destination requests. The accuracy of the recommended route depends on the accuracy of voice command recognition in noisy and crowded environments. **The goal** of this paper is to develop a hardware-software module for voice navigation within an assistive system that enables user positioning and command recognition in the presence of acoustic interference, with the goal of enhancing the independence and mobility of individuals with visual and/or musculoskeletal impairments. **The subject** of this study is the processes and methods for improving the accuracy and robustness of voice command recognition in a hardware-software navigation subsystem in the presence of stationary and non-stationary noise and a multi-voice acoustic environment. **The task** of this work is to develop a functional model of the voice navigation hardware-software module, to develop a comprehensive sound processing algorithm that accounts for various types of acoustic interference and voice overlap; to analyze the impact of various types of acoustic interference and voice overlap on the performance of the software-hardware system. **Methods** such as comparative analysis, experimental models, acoustic signal recognition were applied. **The results** demonstrate that the proposed preprocessing and speaker-oriented filtering improve speech command recognition quality in all considered cases. WER decreased from 15.4% to 13.2% in reference conditions, from 48.6% to 30.9% under non-stationary noise, from 68.7% to 15.4% under 50 dB stationary noise, and from 73.0% to 35.3% for overlapping speech of equal loudness. The obtained results confirm the practical suitability of the proposed subsystem for assistive navigation systems operating in acoustically challenging real-world environments. **Further research** will focus on integrating intelligent personalization techniques into navigation feedback.

Keywords: navigation; speech recognition; speaker verification; voiceprint; audio preprocessing; wake-word detection; vibrotactile feedback.

Introduction

Current developments in information technology designed to assist people with household and everyday tasks are increasingly moving toward a type of interaction in which voice commands and auditory responses are becoming widespread as the most natural and intuitive way to interact.

Despite the high appeal of speech recognition and synthesis technologies, which are key components of voice control, they are still susceptible to external acoustic interference and background noise, which can reduce their effectiveness in real-world conditions [1, 2].

Interference leads to a significant reduction in recognition accuracy or complete distortion of the query, which in assistive or navigation systems can have negative consequences for user safety and health [3, 4]. Therefore, the general goal of research in the field of voice control interfaces is to accurately separate the target speaker from other speech sources or noise, a task complicated by the potential need for frequent changes. This goal can be achieved through hardware and software solutions that offer both high reliability and flexibility in adapting to and overcoming dynamic interference.

Acoustic disturbances take various forms and can have fundamentally different origins depending on the conditions under which the system is used and the specific characteristics of each individual speaker; however, they can generally be divided into three main categories [5, 6]:

- stationary, i.e., constant or monotonous noise;
- non-stationary, i.e., sudden, one-time sounds;
- speech-related noise, i.e., speech defects or voice overlap.

Based on existing research, a general classification of acoustic noise was conducted, and the most important types were identified within the scope of the study's objectives (Table 1) [7, 8].

Voice overlap poses a particular challenge for speech recognition due to the difficulty of filtering and its significant impact on speech recognition mechanisms. Studies show that when voices overlap for 30 – 40% of the time, the error rate increases twofold or more depending on the system due to a decrease in the signal-to-noise ratio (SNR) and spectral distortion [9, 10].

Therefore, the primary task is noise suppression with improved SNR and voice isolation in multi-voice environments during the preprocessing stage of the input signal.

Given the aforementioned characteristics of technological development and the complexity of their practical application, a software-hardware complex for voice navigation capable of functioning effectively in the presence of acoustic interference and overlapping voices becomes essential. Such a system must ensure stable and accurate recognition of voice commands using efficient and fast audio signal processing algorithms, as well as include modules for user interaction via an audio response to a voice query.

Table 1 – Classification of types of hearing impairments

No.	Group	Characteristic	Examples
1	Stationary	Acoustic characteristics remain stable over time and are similar to white noise with a certain power level	Equipment hum, wind, road traffic, static noise
2	Non-stationary	Frequent temporal variations in acoustic characteristics over a wide range	Impacts, sirens, isolated shouts, keystrokes, sneezing, acoustic signals, speech defects
3	Reverberation	Temporal smearing of the useful signal, causing sound distortion and degradation of acoustic characteristics	Echo
4	Overlapping noises	Active sound sources that have a stationary nature but lower power and/or spatial localization.	Small operating devices, music players
5	Overlapping speech	Superposition of two or more voices in the time and frequency domains.	Background conversations, crowd noise

Literature review and problem statement

The key elements underpinning user interaction with voice navigation tools are speech recognition (Speech-to-Text, hereinafter STT) and speech synthesis (Text-to-Speech, hereinafter TTS) [11, 12]. When combined, these technologies form a voice interface that can be easily and naturally used by people, even those with certain physical limitations. Both technologies have undergone active development in recent years thanks to deep learning techniques, which have significantly expanded the capabilities for working with voice data and improved the quality of synthesis.

Currently, there are a number of voice navigation solutions based on the aforementioned principles, ranging from purely software-based implementations in the form of applications or software services [13, 14] to full-fledged hardware-software systems [15, 16] that integrate sensors, computing modules, and interfaces. The vast majority of existing devices are designed to meet the needs of users with visual impairments, mobility limitations, or other disabilities. Current products may vary in implementation depending on the manufacturer, but in general, they can be divided into several categories:

- wearable environmental description devices that assist visually impaired people with navigation [17, 18];
- devices in the form of canes or bracelets with integrated artificial intelligence capabilities that alert users to obstacles or deviations [19];
- 2-way radio-style devices with button-adapted software and a voice interface;
- hybrid services that connect users to a cloud-based AI assistant or a live person for real-time assistance.

Most modern commercial products are expensive, with prices reaching thousands of dollars per device, making them inaccessible to many users – a situation that is particularly critical for older adults and those unable to work fully due to physical disabilities. In addition, a significant portion of these products are not self-sufficient, requiring a connection to a mobile app or cloud services via the internet, and their control functionality may be limited to only clearly defined commands, the recognition of which is not resistant to significant interference.

Based on the results of our review of existing solutions, we can conclude that a general model of voice navigation – which could serve as the foundation for future developments – can be described in several stages:

- data input – a speech recording session initiated by a user via an interface or a keyword;
- preprocessing – rapid processing to remove potential noise;
- recognition – converting speech to text or extracting keywords for processing;
- decision-making – logical analysis, system modifications, and decision-making based on conditions and the query;
- response generation – generating a text-based speech response to the user's query based on the decisions made;
- synthesis and waiting – speech synthesis and waiting for the initiation of a new data input session.

Aims and tasks

The aim of this work is to develop a hardware-software module for voice navigation within an assistive system that enables user positioning and command recognition in the presence of acoustic interference, with the goal of enhancing the independence and mobility of individuals with visual and/or musculoskeletal impairments. To achieve this goal, the following tasks must be addressed:

- development of a functional model of the voice navigation hardware-software module;
- development of a comprehensive sound processing algorithm that accounts for various types of acoustic interference and voice overlap;
- analyzing the impact of various types of acoustic interference and voice overlap on the performance of the software-hardware system.

The next step in the development of the system is the implementation of an advanced navigation system with extensive use of artificial intelligence to personalize prompts and integrate with other devices to provide a wider range of information.

Materials and methods

The primary focus in the development of a hardware-software system is the design and description of the system's logical structure, which will ensure the execution of all intended functions and integrate the software and hardware components. For the voice navigation prototype being developed under conditions of acoustic interference and voice overlap, a logical structure has been developed and described that outlines the system's functions, data flows within it, and the interaction between software and hardware modules. The functional model of the proposed system, based on the logical structure, is shown below (Fig. 1).

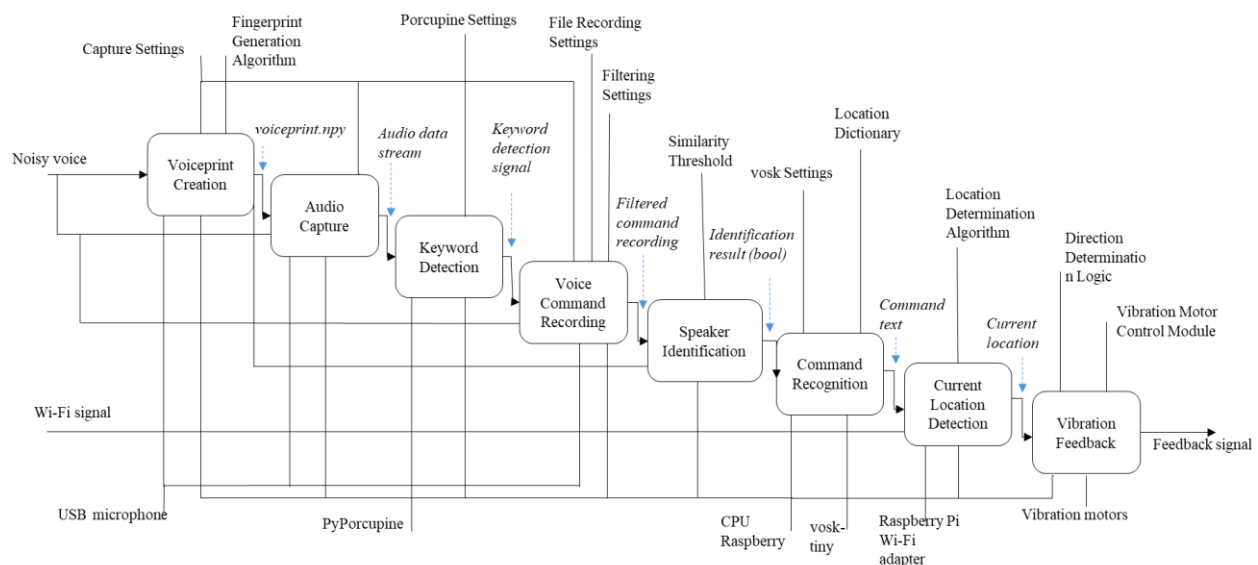


Fig. 1. Functional model of the proposed system in the IDEF0 standard

At the first level of the proposed model, A0, a voiceprint of the user is generated, which is a crucial element of the system's operation in scenarios involving overlapping voices. Based on short phrases recorded via a microphone, a file is created that contains the distinctive characteristics of the user's voice.

The subsequent levels of the model describe the logical structure of the system's main operating cycle. At level A1, an unfiltered audio stream is captured from the system's environment, which contains both useful information in the form of voices and acoustic noise. The captured audio stream is transmitted to level A2, where a small Porcupine model operates, configured exclusively to detect the keyword. Data regarding the keyword detection event is transmitted to the next level of the model.

At level A3, provided the keyword is detected, the main process of recognizing the user's command begins. The command is recorded via a microphone into an audio file and processed according to a developed and configured noise filtering algorithm, after which the filtered command recording is transmitted to the next level. At level A4, the filtered recording is evaluated based on its acoustic characteristics and compared to the user's voiceprint created at stage A0. Based on the results of the acoustic comparison, a determination is made as to whether the voice belongs to the user, and the data is transmitted to the next level of the model.

At level A5, provided that the voice in the command recording is identified as the user's voice, speech recognition occurs and the speech is converted to text using the small STT vosk model. Additionally, the recognized text is compared against a location dictionary for identification. If a word from the dictionary is found in the text, the recognized text is passed to the next level of the model. At level A6, if the command text is successfully received, navigation guidance is provided based on the user's current location. The current location is determined based on signal strength data from known Wi-Fi sources and by comparing this data with a predefined location table. Once the current location is determined, its value is passed to the final level of the model.

At the final level of the model, A7, based on the current location data and the user's command text (recognized at level A5), a decision is made regarding the direction in which the user should move. Based on this decision and the settings of the vibration motor control module, a GPIO signal is generated that activates the necessary vibration motors, providing feedback to the user (A8).

The proposed solution is based on a loop for monitoring the audio channel, divided into specific stages. The elements of the loop include:

- initialization of functional modules;
- continuous monitoring of the audio channel;
- keyword recognition;
- recording of the voice command;
- transmission of the recording to the analyzer;
- checking for the presence of the command in the recording;
- comparison of the voice with the template;
- interpretation of the command and transmission of the control signal to peripheral devices;
- handling of external interrupts.

The hardware component of the proposed solution directly affects both the device's performance and parameters such as modularity, energy efficiency, scalability, and so on. The hardware configuration of the system includes (Fig. 2):

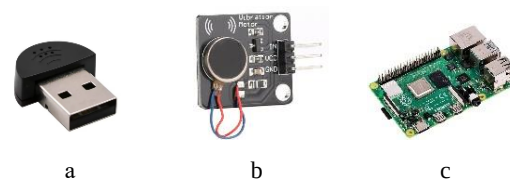


Fig. 2. Hardware stack required for testing:
a – USB microphone, b – vibration module, c – computer

- a single-board computer serving as a programmable processor;
- a USB microphone;
- vibration modules;

- auxiliary components;
- a power cable.

A prototype of the device in its assembled form, with the vibration modules and USB microphone connected, is shown below (Fig. 3).

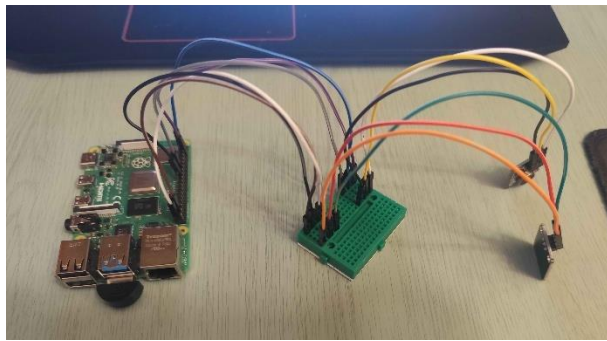


Fig. 3. Photo of the device without the power cable connected

As part of the software component of the hardware-software prototype for voice navigation, a step-by-step processing of the audio recording is proposed to minimize the impact of acoustic noise, thereby preparing the speech for subsequent recognition by the language model. The pipeline is executed entirely on the computing device and consists of the stages shown in Fig. 4:

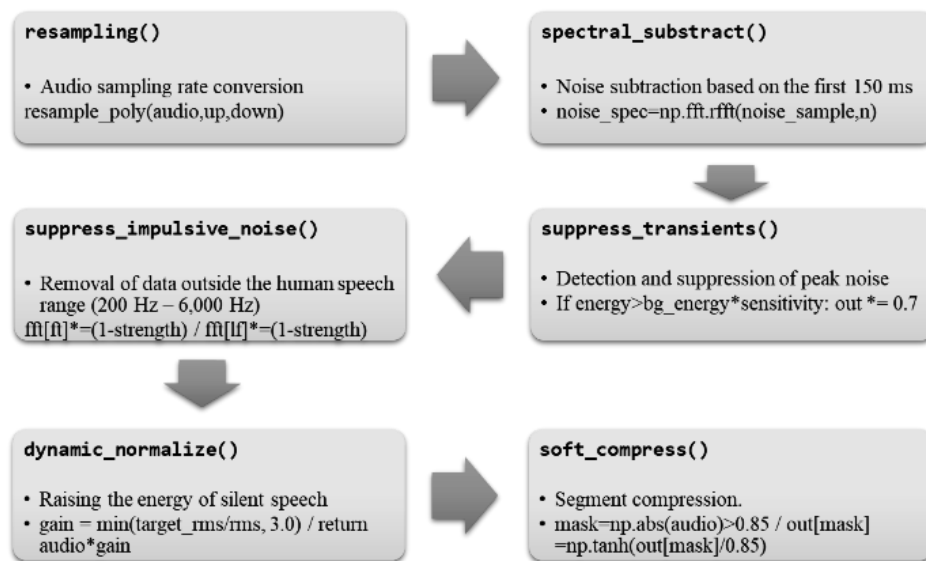


Fig. 4. Diagram of the audio processing pipeline in noisy environments

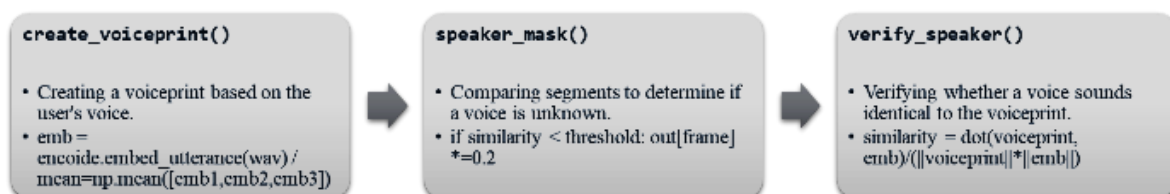


Fig. 5. Diagram of the audio processing pipeline for voice overlay

To counteract the influence of extraneous voices, a pipeline is proposed that includes the preliminary creation of a reference user voiceprint, segment-based masking of fragments with low similarity to this reference, and final speaker verification based on the cosine similarity measure. The reference voiceprint is formed as an averaged

- recording of a noisy command into an audio file;
- subtraction of constant noise;
- detection and suppression of peak noise;
- attenuation of the frequency range outside the speech frequency range;
- volume normalization;
- light audio compression;
- recognition of the processed signal by the language model.

The recorded voice command undergoes preprocessing through a sequential pipeline that includes resampling to 16 kHz, steady-state noise suppression, impulse noise attenuation, signal spectrum band-limiting, volume normalization, and mild compression. This approach reduces the impact of acoustic noise and stabilizes the speech signal parameters prior to recognition. Afterward, the processed recording is sent to the Kaldi Recognizer as part of the Vosk model vosk-model-small-uk-v3-small, which enables command recognition in Ukrainian on portable devices with limited computational resources.

To improve system robustness in environments with overlapping voices, a protection pipeline has been additionally implemented, focused on speaker verification and reducing the impact of background noise on the recognition quality of the target command (Figure 5).

embedding obtained from three samples of the user's speech, preprocessed using the same noise suppression pipeline as the system commands. After speaker verification and command recognition, the user's current location is determined using Wi-Fi signals. To do this, SSID-RSSI pairs are analyzed, and the current position is identified as

the location with the highest signal level among known access points. The result is used to select the direction of the navigation prompt (Fig. 6).

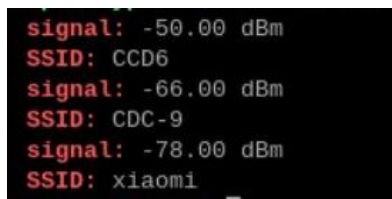


Fig. 6. Example of data from available WiFi sources received by the system

The current location determined based on the WiFi source is compared with the location identifier present in the recognized voice command. Depending on whether the locations match or differ, a decision is made to generate the corresponding GPIO signal to control the vibration motors.

To further evaluate the recognition quality and, consequently, the effectiveness of the proposed system pipelines under conditions of acoustic noise and voice overlap, the following evaluation metrics have been defined:

- Word Error Rate (WER) – the word error rate, which determines the proportion of errors out of the total number of recognized commands. WER is defined as the ratio of the number of incorrectly recognized words to the total number of words in the command;

- Speaker Verification Confidence (SVC) – the degree of confidence in the user. A value close to 1 indicates high confidence that the voice belongs to the profile owner.

To evaluate the effectiveness of the proposed pipeline under noisy conditions, five test scenarios were selected, for which five series of recognition tests, each consisting of fifteen tests, were conducted. As part of the tests, after the system detected the keyword, a three-word command was issued that included a word from the location vocabulary: “go to the bedroom.” The command varied slightly in intonation and tempo to simulate real-world variability and evaluate the performance of the SVC estimation mechanism. The distance to the speaker and the noise source was 20 cm in all cases. Based on the command recognition accuracy and the system’s confidence score in matching the voice to the voiceprint, the WER and SVC metrics were determined and analyzed. Scenarios in which the system was tested:

- recognition under reference conditions without acoustic noise and voice superposition, serving as a baseline and for evaluating the impact of intrinsic noise;
- recognition under conditions of non-stationary noise with direct superposition on useful information (commands);
- recognition under conditions of stationary noise at three volume levels.

Results of the study. Discussion of the results

During the experiments, the voice volume was approximately 37 dB, while the noise level varied from 40 to 60 dB in 10-dB increments; the distance from the microphone was fixed (approximately 20 cm).

In the first experiment, the quality of voice command recognition was evaluated under reference conditions, i.e.,

without additional acoustic interference or voice overlap. The aim of the experiment was to determine the system’s baseline accuracy and assess the impact of the proposed preprocessing pipeline even in a favorable acoustic environment. The WER and SVC metrics were used for analysis. SVC values in subsequent experiments are presented as percentages, although they were originally recorded on a scale from 0 to 1000 (Table 2).

Table 2 – Recognition results under standard conditions

Condition	WER, %	SVC, %
Without filtering	15.4	77.0
With filtering	13.2	78.6

The results showed that, under standard conditions, the system delivers acceptable recognition quality; however, the use of a preprocessing pipeline further improves performance: the WER decreases from 15.4% to 13.2%, and the SVC increases from 77.0% to 78.6%. This indicates the advisability of using filtering even in the absence of significant external interference, which can be explained by the suppression of the system’s inherent noise and the stabilization of the audio signal characteristics. In this case, the keyword from the vocabulary was recognized consistently, while the majority of errors occurred with short command service words.

The second experiment investigated the quality of voice command recognition under conditions of non-stationary noise superimposed on the useful signal. The aim of the experiment was to assess the impact of dynamic acoustic interference on the system’s performance and to verify the effectiveness of the proposed preprocessing pipeline under such conditions. The WER and SVC metrics were used for evaluation (Table 3).

Table 3 – Recognition results for non-stationary noise

Condition	WER, %	SVC, %
Without filtering	48.6	70.7
With filtering	30.9	73.7

The results showed that, in the presence of non-stationary noise and without preprocessing, recognition quality deteriorates significantly: the average WER increases to 48.6%, while the SVC decreases to 70.7%. The use of the proposed pipeline allowed us to reduce the WER to 30.9% and increase the SVC to 73.7%. Thus, the use of preprocessing reduces the recognition error by approximately 37.7% relative to the baseline scenario under the same conditions, confirming its effectiveness for the system’s operation in a dynamic acoustic environment.

In the third experiment, the system’s robustness to stationary acoustic noise was investigated using white noise as an example. Testing was conducted for three noise intensity levels – 40, 50, and 60 dB – which allowed us to assess changes in recognition quality as speech was increasingly masked by noise. The WER and SVC metrics were used for evaluation (Table 4).

The results showed that, under steady-state noise conditions without preprocessing, recognition quality rapidly deteriorates as noise intensity increases: the WER rises from 39.7% at 40 dB to 86.6% at 60 dB.

Table 4 – Recognition results for stationary noise with an intensity of 40 – 60 dB

Background noise level	WER without filtering, %	WER with filtering, %	SVC without filtering, %	SVC with filtering, %
40 dB	39.7	15.4	77.5	75.0
50 dB	68.7	15.4	74.3	69.7
60 dB	86.6	33.1	65.1	66.4

The use of the proposed pipeline ensures a significant reduction in recognition error at all studied noise levels: down to 15.4% at 40 and 50 dB and down to 33.1% at 60 dB. At the same time, SVC decreases moderately, indicating that speaker verification robustness is partially preserved even when speech is significantly masked by noise. Thus, the proposed approach allows recognition performance to be maintained at a level close to the baseline up to and including 50 dB of noise, while at 60 dB a predictable deterioration is observed, yet the system remains operational. In the fourth experiment, the system's performance in a multi-speaker environment was

investigated. To this end, three scenarios were considered: superimposition of a voice with lower volume relative to the target speaker, superimposition of a voice with equal volume, and command delivery by an intruder only.

The aim of the experiment was to assess the extent to which the use of voiceprint-oriented processing, specifically speaker masking and final speaker verification, increases the system's robustness against background speech.

The metrics WER and SVC were used for analysis, and for the scenario with a voice of equal loudness, FAR and FRR were also used (Table 5).

Table 5 – Recognition results under conditions of overlapping voices at different volumes

Condition	WER without filtering, %	WER with filtering, %	SVC without filtering, %	SVC with filtering, %
Overlaying a quieter voice	41.8	22.0	76.2	73.7
Overlaying a steady voice	73.0	35.3	73.9	72.3
A third-party voice without a designated speaker	17.6	22.0	61.5	57.6

The results showed that, under conditions of background speech superimposition, the proposed pipeline reduces the average WER from 41.8% to 22.0% with a moderate decrease in SVC from 76.2% to 73.7%, indicating the effectiveness of background speech suppression without a significant loss in the verification quality of the target speaker.

In a more complex scenario involving voices of equal loudness, the WER decreases from 73.0% to 35.3%, and the FAR decreases from 26% to 6% while maintaining an FRR of 13%. This confirms that the use of speaker masking and voiceprint verification significantly reduces the probability of falsely accepting a foreign command even in the case of strong acoustic overlap.

In a scenario where the command is spoken only by an out-of-band speaker, SVC drops to 61.5% without processing and to 57.6% after processing, indicating that the characteristics of the out-of-band voice are being separated from the target user's voiceprint.

Although WER is not the primary target metric in this case, the results confirm the system's ability to suppress responses to non-target speech (Fig. 7). Overall, this demonstrates the effectiveness of using a voiceprint-oriented pipeline to enhance the system's robustness in a multi-voice environment.

Conclusions

The results obtained experimentally confirm that combining a speaker mask with voiceprint verification improves the system's robustness not only against noise but also against multi-voice acoustic environments, thereby reducing the number of false acceptances of extraneous commands.

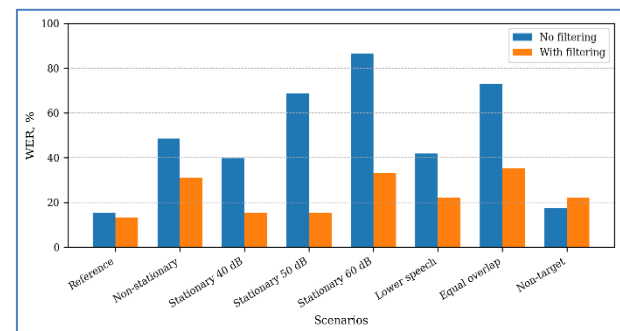


Fig. 7. Summary comparison of voice command recognition quality across different acoustic scenarios

This provides grounds for considering the proposed approach as a means of enhancing the robustness of voice interaction in assistive systems designed for real-world, rather than laboratory, conditions. The practical value of the results lies in the fact that the system remains operational even in the presence of background voices and reduces the risk of erroneous execution of extraneous commands, which is critically important for assistive navigation solutions. This work addresses the pressing challenge of improving the effectiveness of voice interaction within an assistive navigation system for users with visual and/or musculoskeletal impairments in the presence of acoustic interference.

In accordance with the stated objective, a functional model of a hardware-software voice navigation module was developed, covering the stages of audio signal acquisition, keyword detection, command preprocessing, speaker verification, speech recognition, current location determination, and the generation of a control signal for feedback devices.

A comprehensive audio processing algorithm has been developed that combines resampling, suppression of stationary and non-stationary noise, bandpass filtering, normalization, soft compression, as well as voiceprint-based speaker masking and speaker verification procedures under voice-over conditions.

Experimental results show that using the proposed pipeline reduces the WER in all scenarios considered. Under reference conditions, the WER decreased from 15.4% to 13.2%; with non-stationary noise, from 48.6% to 30.9%; with 50 dB stationary noise—from 68.7% to 15.4%, and with voice overlap of equal loudness—from 73.0% to 35.3%.

It has been shown that for the voice-overlay scenario, the use of a voiceprint-oriented approach not only improves recognition quality but also reduces the frequency of false acceptance of foreign commands: the FAR decreased from 26% to 6% while the FRR remained unchanged at 13%. This confirms the increased robustness of the system in a multi-voice environment.

The results obtained demonstrate the practical suitability of the proposed module for use in assistive

navigation systems designed to operate in real-world acoustically complex environments.

Further research will focus on integrating intelligent personalization techniques into navigation feedback.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

Acknowledgements

The study was funded by the Ministry of Education and Science of Ukraine in the framework of the research project 0126U002094 on the topic “Multimodal System for Indoor Navigation and Assistive Support for Users with Special Needs”..

REFERENCES

1. Deshmukh A. M., and Chalmeta, R. (2024), “User experience and usability of voice user interfaces: A systematic literature review”, *Information*, vol. 15, no. 9, art. 579, doi: <https://doi.org/10.3390/info15090579>
2. Chandollikar, N., Joshi, C., Roy, P., Gawas A., and Vishwakarma, M. (2022), “Voice Recognition: A Comprehensive Survey”, *2022 Int. Mobile and Embedded Techn. Conf.*, pp. 45–51, doi: <https://doi.org/10.1109/MECON53876.2022.9751903>
3. Yarovy, V., Barkovska, O., Maksymov, D., Ni, Y., and Raptanov, D. (2026), “Methods for recognising hand gestures using computer vision to assist people with movement or speech impairments”, *Control, Navigation and Communication Systems*, vol. 1 (83), pp. 135–141, doi: <https://doi.org/10.26906/SUNZ.2026.1.135>
4. Barkovska, O., Nechitailo, O., Serdechnyi, V., and Baturin, O. (2025), “Software Application for Assisting Visually Impaired Persons at Public Transport Stops”, *Information and control systems at railway transport*, vol. 30, no. 3, doi: 38–48, doi: <https://doi.org/10.18664/iksz.v30i3.351322>
5. Kumar, C., Ganguly, B., Dey, D., and Chatterjee, S. (2025), “Recent trends and open challenges in acoustic partial discharge signal denoising techniques: A review”, *Electric Power Systems Res.*, vol. 248, doi: <https://doi.org/10.1016/j.epsr.2025.111931>
6. Barkovska, O., Havrashenko, A., and Botmar, P. (2025), “The influence of reverberation, equalization and compression methods on speaker recognition”, *2025 IEEE 6th KhPI Week on Advanced Technology (KhPIWeek)*, IEEE, pp. 1–5, doi: <https://doi.org/10.1109/KhPIWeek61436.2025.11288718>
7. Opochninsky, R., Moradi, M., and Gannot, S. (2025), “Single-microphone speaker separation and voice activity detection in noisy and reverberant environments”, *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 18 (2025), doi: <https://doi.org/10.1186/s13636-025-00404-7>
8. Sadeghi, M. E., Sheikhzadeh, H., and Emadi, M. J. (2024), “A proposed method to improve the WER of an ASR system in the noisy reverberant room”, *Journal of the Franklin Institute*, vol. 361, no. 1, Elsevier, pp. 99–109, ISSN 0016-0032, doi: <https://doi.org/10.1016/j.jfranklin.2023.11.039>
9. Raghav, S., Vikhyath, B.M., Karthikeya, R., and Lalitha, S. (2025), “Multi-speaker Speech Processing in Noisy Environments: A Hybrid Model for Source Separation and Summarization”, *Advancements in Signal, Image and Video Processing, Lecture Notes in Electrical Engineering*, vol 1449, Springer, Singapore, doi: https://doi.org/10.1007/978-981-96-9967-4_12
10. Kim, M., and Jang, G.-J. (2024), “Speaker-Attributed Training for Multi-Speaker Speech Recognition Using Multi-Stage Encoders and Attention-Weighted Speaker Embedding”, *Applied Sciences*, vol. 14, is. 18, article number 8138, doi: <https://doi.org/10.3390/app14188138>
11. Kshama, N., Andanur, P. Udupa, E. P. P., Airani, S. S., and Dattathreya, (2025), “AI-Powered Voice Assistants for Accessible E-Commerce: A Review of ASR, RAG, LLMs and TTS”, *2025 5th International Conference on Evolutionary Computing and Mobile Sustainable Networks*, Coimbatore, India, pp. 707–712, doi: <https://doi.org/10.1109/ICECMSN68058.2025.11383381>
12. Dixit, M., Landage, R., and Raut, P. (2026), “Comparative Analysis of Deep Learning Models for Speech-to-Text and Text-to-Speech Conversion”, *Smart Trends in Computing and Communications*, SmartCom 2025, *Lecture Notes in Networks and Systems*, vol 1465, Springer, Singapore, doi: https://doi.org/10.1007/978-981-96-7517-3_5
13. Patel, D., and Verma, T. (2022), “Application of voice assistant using machine learning: A comprehensive review”, *Advances and Applications in Mathematical Sciences*, vol. 21, no. 9, Mili Publications, India, pp. 5063–5073, doi: https://www.mililink.com/upload/article/1856195715aams_vol_219_july_2022_a18_p5063-5073_deepika_patel_and_toran_verma.pdf
14. Barkovska, O., Holovchenko, O., Storchai, D., Kostin, A., and Lehezin, N. (2025), “Investigation of computer vision techniques for indoor navigation systems”, *Innovative technologies and scientific solutions for industries*, vol. 2(32), pp. 5–15, doi: <https://doi.org/10.30837/2522-9818.2025.2.005>
15. Malokhvii, E., Kuchuk, H., Korobko, A., Zhovnir, N., and Oryshchenko, S. (2026), “Multi-objective framework for end-device processing and offloading in industrial IOT”, *Advanced Information Systems*, vol. 10, no. 2, pp. 117–137, doi: <https://doi.org/10.20998/2522-9052.2026.2.12>

16. Bhaganagare, S., Chavan, S., Gavali, S., and Godase, V. V. (2025), "Voice-Controlled Home Automation with ESP32: A Systematic Review of IoT-Based Solutions", *Journal of Microprocessor and Microcontroller Research*, vol. 2, is. 3, pp. 1–13, available at: <https://matjournals.net/engineering/index.php/JoMMR/article/view/2389>
17. Hersh, M. (2022), "Wearable Travel Aids for Blind and Partially Sighted People: A Review with a Focus on Design Issues", *Sensors*, vol. 22, no. 14, art. 5454, doi: <https://doi.org/10.3390/s22145454>
18. Serdechnyi, V., Barkovska, O., Kovalenko, A., Martovytskyi, V., and Havrashenko, A. (2025), "Knowledge systematization about the characteristics of existing technological means for assisting people with visual impairments", *Advanced Information Systems*, vol. 9, no. 1, pp. 13–23, doi: <https://doi.org/10.20998/2522-9052.2025.1.02>
19. Mai, C., Xie, D., Zeng, L., Li, Z., Li, Z., Qiao, Z., Qu, Y., Liu, G., and Li, L. (2023), "Laser Sensing and Vision Sensing Smart Blind Cane: A Review", *Sensors*, vol. 23, no. 2, art. 869, doi: <https://doi.org/10.3390/s23020869>

Received (Надійшла) 23.03.2026

Accepted for publication (Прийнята до друку) 12.08.2026

ABOUT THE AUTHORS / ВІДОМОСТІ ПРО АВТОРІВ

Барковська Олесь Юрївна – кандидат технічних наук, доцент, доцент кафедри Електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Olesia Barkovska – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;

e-mail: olesia.barkovska@nure.ua; ORCID Author ID: <https://orcid.org/0000-0001-7496-4353>;

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=24482907700>.

Сторчай Денис Геннадійович – магістрант кафедри Електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Denys Storchai – Master's Student of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,

e-mail: denys.storchai@nure.ua; ORCID Author ID: <https://orcid.org/0009-0006-0527-0095>.

Коваленко Андрій Анатолійович – доктор технічних наук, професор, завідувач кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Andriy Kovalenko – Doctor of Technical Sciences, Professor, Head of Department of the Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;

e-mail: andriy.kovalenko@nure.ua; ORCID Author ID: <https://orcid.org/0000-0002-2817-9036>;

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=56423229200>.

Романенков Юрій Олександрович – доктор технічних наук, професор, професор кафедри Електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

Yuri Romanenkov – Doctor of Technical Sciences, Professor, Professor of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,

e-mail: yuri.romanenkov@nure.ua , ORCID Author ID: <https://orcid.org/0000-0002-6544-5348>;

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57189377960>.

Трунов Вадим Олександрович – інженер, Харківський національний університет радіоелектроніки, Харків, Україна

Vadym Trunov – Engineer, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;

e-mail: vadym.trunov@nure.ua; ORCID Author ID: <https://orcid.org/0009-0009-0758-5228>.

Апаратно-програмний прототип підсистеми голосової навігації в умовах сильного шуму

О. Ю. Барковська, Д. Г. Сторчай, А. А. Коваленко, Ю. О. Романенков, В. О. Трунов

Анотація. Робота присвячена розробці апаратно-програмного прототипу підсистеми голосової навігації, призначеної для допоміжної навігації в умовах акустичних перешкод. **Актуальність** теми пояснюється необхідністю підвищення точності формування голосових або тактильних підказок щодо рекомендованого маршруту для користувачів із вадами зору на основі їхніх голосових запитів щодо пункту призначення. Точність рекомендованого маршруту залежить від точності розпізнавання голосових команд у шумних та багатолюдних середовищах. **Метою** даної роботи є розробка апаратно-програмного модуля голосової навігації в рамках допоміжної системи, що забезпечує визначення місцезнаходження користувача та розпізнавання команд за наявності акустичних перешкод, з метою підвищення незалежності та мобільності осіб з порушеннями зору та/або опорно-рухового апарату. **Предметом** даного дослідження є процеси та методи підвищення точності та стійкості розпізнавання голосових команд у апаратно-програмній навігаційній підсистемі за наявності стаціонарного та нестаціонарного шуму та багатоголосового акустичного середовища. **Завданням** даної роботи є розробка функціональної моделі апаратно-програмного модуля голосової навігації, розробка комплексного алгоритму обробки звуку, що враховує різні типи акустичних перешкод та перекриття голосів; аналіз впливу різних типів акустичних перешкод та перекриття голосів на продуктивність апаратно-програмної системи. Були застосовані такі методи, як порівняльний аналіз, експериментальні моделі та розпізнавання акустичних сигналів. **Результати** показують, що запропонована попередня обробка та фільтрація з урахуванням особливостей мовця покращують якість розпізнавання голосових команд у всіх розглянутих випадках. Показник WER знизився з 15,4 до 13,2% в еталонних умовах, з 48,6 до 30,9% в умовах нестаціонарного шуму, з 68,7 до 15,4% при стаціонарному шумі 50 дБ та з 73,0 до 35,3% при перекритті мовлення однакової гучності. Отримані результати підтверджують практичну придатність запропонованої підсистеми для допоміжних навігаційних систем, що працюють в акустично складних реальних середовищах. **Подальші дослідження** будуть зосереджені на інтеграції інтелектуальних методів персоналізації в навігаційний зворотний зв'язок.

Ключові слова: навігація; розпізнавання мови; верифікація мовця; голосовий відбиток; попередня обробка аудіо; виявлення слова-активатора; вібротактильний зворотний зв'язок.