STUDIA INFORMATICA
Nr 2 (33) Systems and information technology 2025

Dariusz Mikułowski
ORCID: 0000-0002-4567-7846

University of Siedlce
Faculty of Exact and Natural Sciences
Institute of Computer Science
ul. 3 Maja 54, 08-110 Siedlce, Poland

`dariusz.mikulowski@uws.edu.pl`

Binaural Sound Scenes Creator Supporting The Development Of Spatial Orientation Skills of Blind Individuals

DOI: 10.34739/si.2025.33.02

Abstract. This article introduces an approach to automatic generation of binaural sound scenes BSS that describe in auditory form environments for visually impaired individuals. This audio, with descriptions of the city environment, can be beneficial for developing spatial orientation skills among Blind Individuals (BI), especially those who have recently lost their vision and must adapt to new, difficult living conditions. The idea to solve the problem is that artificially generated binaural scenes can simulate situations in which a BI navigates open spaces, such as a city, or large buildings like train stations, airports, and university campuses. Practising these skills in a simulated environment can be less stressful than real-world training. This work describes the development of a specialised application that facilitates the creation of binaural sound scenes from simple monophonic sounds. Several scenes were subsequently generated and then tested by three visually impaired individuals. This testing aimed to determine whether entirely artificially generated sound scenes are a sufficiently robust approach for constructing more complex simulation environments.

Keywords: Binaural sounds, Spatial orientation of the blind, Automatic sound generation

1 Introduction

The development of assistive technologies is an extensive research field aimed at helping people with disabilities, i.e., those who are blind or have low vision, overcome physical, social, and infrastructural barriers in everyday activities. Globally, at least 2.2 billion people have near or distance vision impairment. Total blindness is estimated to affect approximately 39 million people worldwide, and another 246 million have poor vision [14]. Vision loss negatively impacts the ability to navigate independently and be mobile, which is why efforts to develop Electronic Travel Aids are crucial [3]. The main goal of assistive technologies and specialised Orientation and Mobility training is to achieve equal access, enabling independent living, the freedom to choose a career path, and the ability to perform everyday activities such as navigation and obstacle detection [2]. However, real-world training can be stressful and requires significant resources, including trained instructors and specialised infrastructure like spatial orientation parks. [18]

To take a step towards overcoming this barrier, we propose a virtual simulation environment as an alternative, using **Virtual Reality** (VR) and **Artificial Intelligence** (AI). In this setup, VR can replace physical infrastructure, and AI can act as a virtual instructor. A key element of this simulation is modelling the user's sound environment. For this purpose, **binaural sounds** can be used to create a 360-degree auditory experience, enabling the simulation of complex soundscapes. These sounds can be generated from simple monophonic sounds using, for example, the **Head-Related Transfer Function** (HRTF). [13]. The **HRTF** filter imitates how sound waves are reflected by a person's head and body, altering sound frequency, volume, and arrival time to create the illusion of direction and movement. By processing individual monophonic sounds, we can create acoustic binaural scenes that realistically model situations a blind person might encounter. These scenes include walking along a street, crossing an intersection with traffic and a sound-enabled traffic light, and being inside a public vehicle. A proposed method also allows simulation of sounds such as passing cars, pedestrians' footsteps, and the tapping of a white cane.

The core of the work is an application that generates detailed binaural sound scenes from multiple monophonic inputs, arranging them in a chronological sequence to create immersive training scenarios. An application was developed as part of a project within the framework of Adrian Wysocki's engineering thesis [20].

The rest of the paper was organised as follows. In the next section, we present the current state of the art regarding binaural sounds, the problem of supporting spatial orientation skills, and the important phenomenon of echolocation, which affects the independent mobility of blind people. In the next section, we define a binaural scene and discuss the proposed solution for creating them. We then present the results and user validation of the created binaural scenes, and conclude with a summary.

2 The state of art

Research is underway to improve spatial orientation and navigation skills for individuals with visual impairments by combining auditory, haptic, and virtual reality (VR) technologies. In this context, prototypes of various devices, focusing on aspects supporting the independent mobility of blind people, were created. On the one hand, projects focus more on recognising obstacles and informing the user about them using sound, vibration or Braille. The second trend of prototypes is more focused on problems related to locating the user in space and guiding him to a selected point, similar to commonly used navigation programs, but with greater accuracy and a slightly different set of information provided to the user.

One of the latest example prototypes that can be included in the group related to obstacle recognition is a portable device consisting of a stereovision camera with structured lighting, a control unit, a haptic belt, speakers and a specially designed [16] keyboard. It uses haptic stimuli and auditory signals to convey environmental information and facilitate navigation. The solution uses depth-image processing, segmentation, and a mapping scheme that transforms the segmented image into a matrix of haptic actuators. The device provides the user with haptic feedback via a custom touch strip equipped with 20 vibration actuators arranged in a 4x5 grid, as well as two additional actuators for feedback on special object categories. The obstacle detection range in this prototype is limited to 2.5 m in a user-defined corridor 0.8–1.5 m wide. Dynamic tests showed 100Similar portable devices, however, in the form of special helmets with headphones and a camera, include the prototype systems described in [21] and [22].

Another approach focuses on auditory systems, especially those using binaural sound to create a Virtual Spatial Reality (VSR). Mikulowski and Szuba's work [12] on a VSR game utilises 3D binaural audio to enhance spatial perception, delivered via bone-conduction headphones, allowing users to hear both virtual and real-world sounds simultaneously. Earlier research by the same group [11] proposed an "Ontologically-Based Object Map" that uses binaural sounds to create a simplified audio map for navigation and pre-trip planning.

Binaural guidance is also central to an autonomous navigation system by Hajebi et al [5]. This system uses mobile devices to combine real-time positioning with 3D binaural audio, providing features for journey planning and real-time guidance. To avoid interfering with ambient sounds, Fujita et al. [4] developed a navigation device that uses bone conduction to deliver binaural sound, leaving the ear canals open. This method has been shown to reduce navigation time and improve localisation accuracy. The American Foundation for the Blind's paper, ”Exploring New Horizons: Binaural Soundscapes,” [1] highlights the importance of **Interaural Time Difference (ITD)** and **Interaural Level Difference (ILD)** in creating immersive 3D soundscapes. This research extends the use of binaural sound beyond navigation to education and recreation, emphasising the need to involve visually impaired individuals in the design process to ensure the technology is truly user-centred.

Another auspicious approach, due to its relatively good availability, is the use of smart glasses in VR systems [23]. The authors of this solution developed the Foveated Audio Device (FAD), created in cooperation with ARIA Research. FAD consists of NReal Light AR glasses and the OPPO Find X3 Pro smartphone The glasses used in this project have low weight (88g), support for computer vision, a 6-degree-of-freedom inertial measurement unit (IMU) and built-in binaural speakers. The authors used the Semi-Global Block Matching (SGBM) and stereo-matching algorithms to calculate a depth map and determine the distance of objects from the user. However, in the described study, the classification was limited to four everyday objects (book, bottle, bowl, cup), each with an appropriate sound assigned (e.g., book - the sound of turning pages, bottle - the sound of a glass bottle). The system was verified by 14 participants (7 blind or partially sighted and seven sighted individuals wearing blindfolds). It was found that the device effectively supported object recognition and reaching.

Another approach is the use of haptic feedback, which provides a tactile dimension to spatial orientation through vibrations and physical sensations. Kleinberg et al. [8] developed a portable haptic system that utilises computer vision, ultrasonic sensors, and machine learning to provide real-time tactile feedback, alerting users to obstacles and promoting safer navigation. Similarly, Kaplan et al. [6] investigated the use of standard gamepads with haptic feedback and audio descriptions of landmarks to support route learning and environmental preparation before physically entering a space.

An interesting aspect of natural movement techniques among people who are blind is echolocation. Echolocation is a natural ability that some blind people develop, in which they emit sounds (e.g., tongue clicks, called FlashSonar) and listen to the reflected echoes to determine the position, distance, size, shape, movement, and even, in any case, the texture of surrounding objects. This technique, while effective, requires significant training. Echolocation used by blind people, such as Daniel Kish [7], involves short sounds that bounce off surfaces of obstacles and return to the listener. Thanks to this, it receives spatial information similar to that which is provided by light for sighted people. Studies have shown that blind people using echolocation activate the primary visual cortex, just as sighted people do when processing visual stimuli [19]. It suggests that spatial information processing at the neuronal level is metamodal. Echolocation provides real functional benefits in daily life, such as greater mobility in unfamiliar places. An example diagram showing how the echolocation phenomenon works is presented in Figure 1.

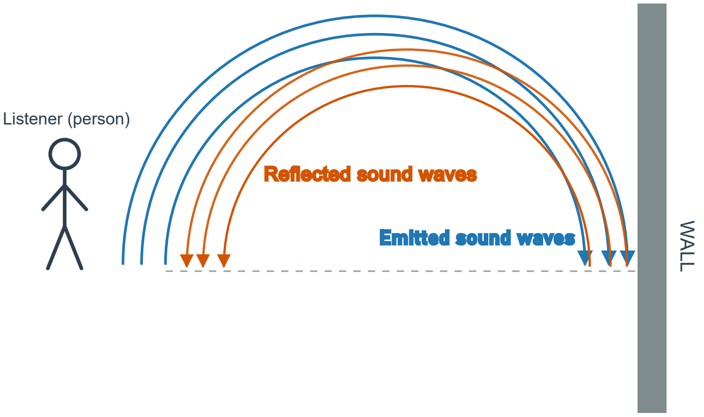


Figure 1: Echolocation ability of the man - schema. (Source: own study)

Since humans cannot naturally generate or hear ultrasound, technology can assist with this process. That is why devices called EBATs (Echolocation-Based Assistive Technologies) are being developed that emit ultrasonic signals and convert the reflected echo into sound audible to humans. Such an example device described in [17] consists of a headset (with an ultrasound emitter and stereo microphones with artificial auricles attached). It uses ultrasonic pulses reflected from an obstacle, which are recorded, then their frequency is lowered (time-stretched) to the range audible to humans and played back to the user. Tests showed that users could judge the size and distance of the obstacle without prior training. However, trained users also demonstrated the ability to judge height. EBAT's devices have also been integrated with smartphones for convenience and portability. Some EBATs use ultrasonic sensors for spatial mapping. These systems often achieve high accuracy, with 92.5–100In the case of echolocation-based devices, machine learning (ML) for object recognition can complement human abilities, reducing the required object detection time by 35.8% and significantly reducing mental workload.

VR environments and simulators provide a safe and controlled space for individuals with visual impairments to learn and practice spatial skills. Lahav's review [9] highlights the potential of VR as a tool for navigation planning, training, and as a diagnostic aid for Orientation and Mobility specialists. Several software libraries exist for creating realistic binaural sounds. **OpenAL** [10] is a cross-platform 3D audio API that models sound sources relative to a listener and supports various effects crucial for realistic binaural audio. **OpenTK** [15] is a C# binding that allows .NET programs to access OpenAL's functions. Another solution is a **Slab3D**, a real-time virtual environment rendering system designed explicitly for 3D sound processing and that includes Head-Related Transfer Function (HRTF) databases for accurate binaural rendering. Finally, the Unity Audio Spatializer SDK is also noted for its ability to generate binaural sound.

3 Designing binaural scenes

3.1 What is a binaural scene

Before we proceed to describe the solution proposed in this work, let us explain what a binaural scene is. A binaural scene is a comprehensive composition of binaural sounds, played simultaneously or sequentially, the most important feature of which is the creation of the impression of immersion in a three-dimensional acoustic space. Thanks to the appropriate selection and processing of signals, the listener has the impression that sounds are coming from specific directions around his head - from the front, back, sides, and even from above or below. In this way, the binaural scene allows for a realistic simulation of presence in a real acoustic environment. To achieve this effect, a binaural scene usually consists of many sound elements that reflect typical stimuli found in urban or natural spaces. These may include, among others, the sounds of passing cars, the footsteps of passers-by, the sound of wind, or the conversations of people passing by the user. The use of various sources allows for the representation of a rich, dynamic space, enabling the recipient to orient himself acoustically. A pretty apt comparison is the juxtaposition of a binaural scene with a musical piece in which several instruments play (each their own part) at different times, producing entirely different sounds. The entirety of such a musical piece constitutes a coherent work offering a specific artistic performance. It will be similar to a binaural scene, which consists of many different sounds in order to give the user a coherent impression of a small fragment of reality. An example of a schematically simplified binaural scene with a street and cars driving on it, a listener in its centre, passers-by and a traffic light with sound is shown in Figure 2.

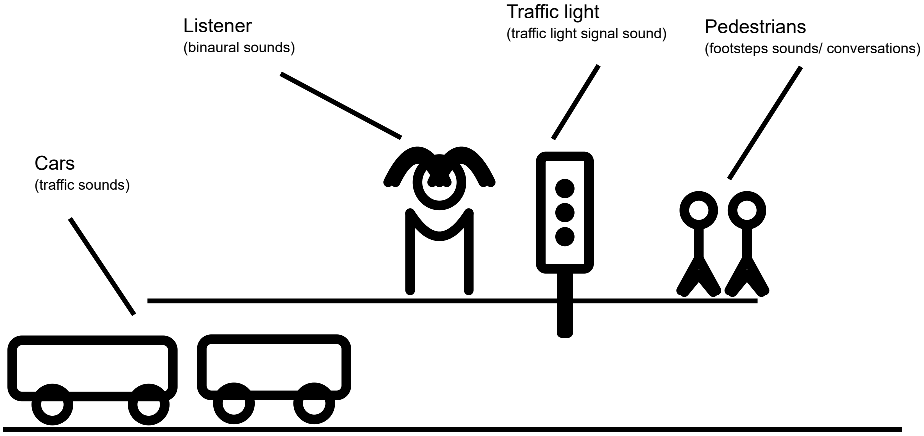


Figure 2: The binaural scene schema. (Source: own study).

An important component of the binaural scene used in spatial orientation training for blind people is voice commentary generated by speech synthesis. This comment acts as a virtual guide - it informs the user about the surroundings and provides tips on the appropriate response in a given situation, e.g.: "You are approaching a traffic light at an intersection. Stop and wait until the light turns green." Thanks to this, the user can practice spatial orientation and movement in conditions similar to those of real life, but in a much safer and more comfortable environment.

Sounds used in binaural scenes can be acquired directly from the environment using special binaural microphones or generated synthetically. In the latter case, signal processing techniques, such as HRTF (Head-Related Transfer Function) filters, are used to map sound directionality to the listener’s head position realistically. In addition, the arrangement and movement of sounds can be dynamically modified, which increases immersion and allows simulating situations in which elements of the environment move relative to the user. Binaural scenes can therefore be a valuable tool for supporting the development of spatial orientation and mobility skills among blind people, enabling safe training and skill development in navigating a complex acoustic environment without exposing them to real threats.

3.2 Binaural scene generator

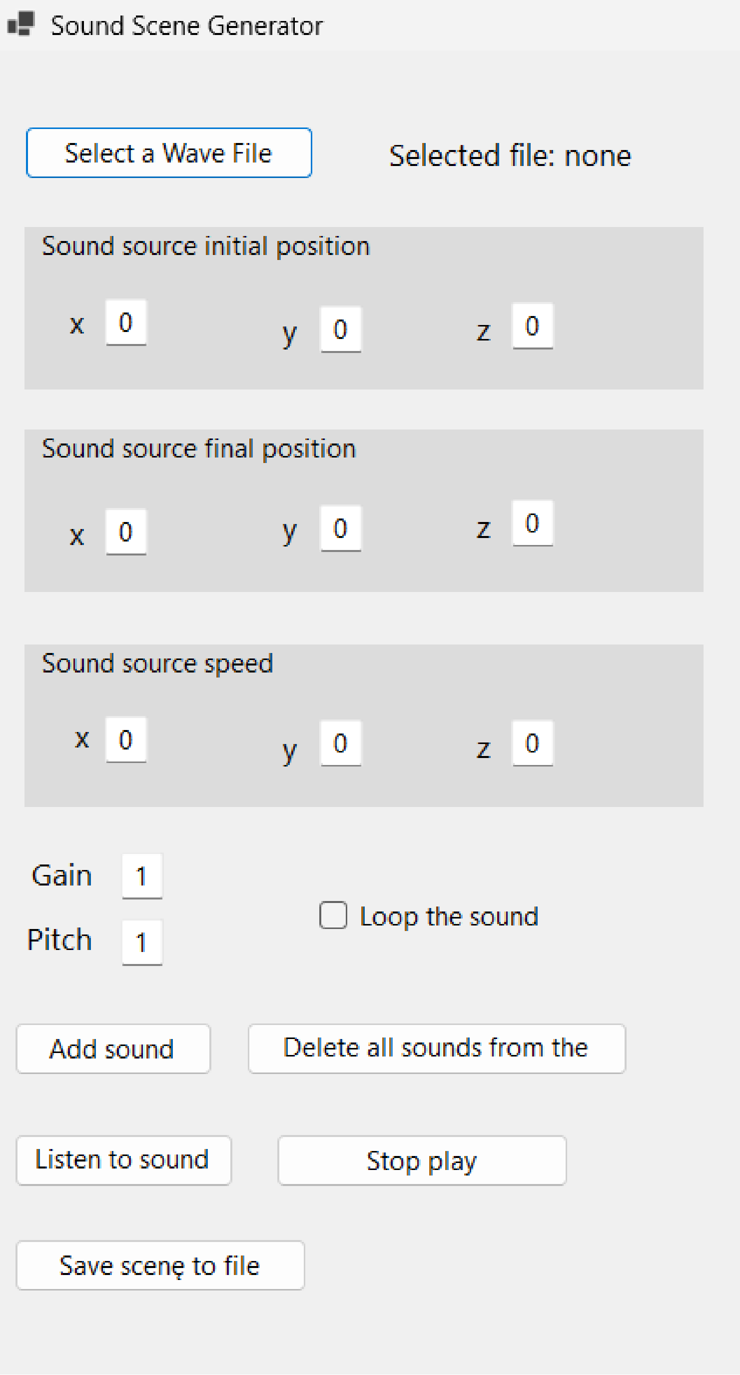
The system proposed in this work was implemented as a monolithic desktop application in .NET and C#. The program features a simple graphical interface that enables the creation of binaural sound scenes from input monophonic sounds. OpenTK and OpenAL libraries were used to work with the sound content.

After starting the application, a form is presented to add individual sounds to the newly created binaural scene. In addition to the possibility of adding the sound file itself, it is necessary to provide many parameters that describe the entire binaural scene, including its component files and time dependencies. The idea is to enable playing sounds at a specific time, in sequence or in parallel, and to determine the direction of sound movement towards the listener. Therefore, in the main application form, some parameters define the behaviour and features of the selected sound in the scene, such as its movement speed or looping. The following fields are available here:

- The initial position of the sound source relative to the listener, determined by coordinates x, y and Z The listener is located at point 0, 0, 0 in space, and it is directed in the direction of negative values on the Z axis.
- The final position of the sound source relative to the listener, determined by coordinates x, y and Z
- the speed vector of the source movement in the three-dimensional space, determined by the value x, y and z. The values in the speed vector determine how fast sound moves along each axis. It is very important that the parameters in the speed vector should match the parameters of the initial and final position of the sound so that the direction of the velocity is the same as the direction from the starting point to the end point.
- GAIN - sound volume, the default value is 1 - representing 100% of sound volume,
- sound pitch, directly proportional to the speed of its playback, the default value is 1 - representing 100% of the pitch. As sound reproduction improves, its height increases.
- the time of starting sound playback, expressed in seconds from the beginning of the scene, the default value is 0. If we want sound to join the scene at a specific second, we can set this parameter to a positive value.
- looping sound - a double selection button that determines whether sound playback is to be repeated.

The main form of the application is presented in figure 3.

With the method described above, by completing subsequent forms and choosing files, many sounds can be added to the scene. However, the situation may arise when the user does not want to enter the parameters for each sound manually. In this case, he can create a CSV file, for example, in Excel. The user can thus enter parameters for all sounds at once according to a set scheme, and then load the data from this file into the program. The CSV file structure corresponds to the fields mentioned above in the form. Thus, in the subsequent columns, user can give: the path to the file, three coordinates of the starting point of the sound, three



Selected file: none

Sound source initial position

x 0 y 0 z 0

Sound source final position

x 0 y 0 z 0

Sound source speed

x 0 y 0 z 0

Gain 1

Pitch 1

☐ Loop the sound

Figure 3: The graphical interface of sound scene generator.
(Source: own study).

coordinates of the sound direction, three coordinates of the sound speed vector, sound volume, pitch sound, the time of its launch expressed in decimal parts of a second and the value of 1 OUB 0 indicating whether the sound is to be looped.

After creating the whole scene, the user can listen to it. It is recommended to listen to the scenes with headphones, as this allows the user to experience the full quality of the binaural scene. To listen to the scene or stop it, use the appropriate buttons. The last button saves the entire scene to a separate sound file recorded using the binaural technique.

3.3 Audio scene development

The binaural scene generator was used to produce three examples of scenes representing simple situations that a BI may encounter in the city. These were the following scenes:

- Scene S_1: The user goes along the street. She/He may listen to the sounds of cars driving from the back to the front and from the front to the back on her/his left side. Additionally, the sound of passers-by in front of the user may sometimes be emitted. The user also hears the sound of his own steps and the sound of hitting a white cane on the pavement.
- Scene S_2: The user is approaching the street bend. On his left and in front of him, the user hears the sound of driving cars. Cars move from the back to the front of the left side of the listener, then turn and move from left to right in front of the listener. Other cars move from right to left in front of the listener and then change direction and move from the front to the back on the left side of the listener. As in a previous scene, the user also hears the sound of his own steps and the sound of hitting a white cane against the pavement.
- Scene S_3: The user is in front of the pedestrian crossing through the intersection. As in the S_1 scene, the user hears the sounds of cars driving on his left and from the back to the front. In addition,/he hears the sound of traffic lights (quick taps) in front of him, meaning that a green light is currently shining, which allows him to cross the pedestrian crossing. In the case of this scene, no sounds of footsteps are heard.

4 Evaluation

The main goal of the evaluation of the created solution was to answer the following questions:

- Do binaural sound scenes give the impression of a real environment so close to realism?
- Can such scenes be used for constructing simulation environments supporting the training of mobility skills and spatial orientation of the blind individuals?

To answer these questions, we have presented ready-made sound scenes for three blind users. The blind testers were experienced people who, daily, themselves move in the opening and have trained skills in spatial orientation. They could listen to each stage five times and then answer a short survey with the following questions.

- ”What particular sounds do you hear within the presented scene?”
- ”What situation from the real world did you listen to in this scene?”
- ”Please rate on a scale of 1 to 10, how much the scene presented is, in your opinion, similar to the real situation in the environment?”

- ”What do you suggest to improve in the presented scene to make it closer to realism?”

Additionally, users could provide general comments on the survey.

All users easily recognised the types of sounds in the scene, such as the sound of a car driving, footsteps, or a cane hitting the pavement. They also had no problem recognising the situation presented by a given scene, although two testers had comments on the lifts and the light signalling. Users rated the similarity of such automatically generated scenes to reality slightly worse (usually 4 out of 10). Such scenes are rather suitable for training basic skills in recognising the environment before moving on to scenes closer to the realism of those recorded by nature.

5 Discussion and conclusions

In this work, a solution for automatically creating binaural sound scenes is proposed. These scenes can then be used to construct simulated environments that support the development of spatial orientation skills in individuals with visual impairments. Our specific approach focused on creating simple binaural scenes from basic monophonic sounds. To achieve this, an application was developed as part of Adrian Wysocki’s engineering thesis [20]. Using this application, several binaural scenes were generated and tested by three blind users.

The results of the experiments indicate that automatically generated specific sounds in these scenes are more distinct and less easily drowned out by ambient noise or additional sounds, which can be an issue with scenes recorded from real environments. For a beginner user who is just starting to learn spatial recognition through hearing, this approach can serve as an excellent foundational tool. Therefore, this technique is well-suited for creating initial sound scenes until users are ready for more advanced scenes created through recording. We are aware that the application is in a very early prototype state, which means it does not allow us to create more sophisticated, complex binaural scenes that could be tested in conditions more similar to real-world conditions. Also, the fact that the scenes were tested by only three blind users does not provide a broader range of responses or test results. Therefore, the proposed experiment should be developed in the future by firstly developing a more advanced application for building binaural scenes on the one hand. On the other hand, more binaural scenes describing different contexts should be created. Such scenes should be tested by a larger number of users to make the research results more robust.

Another conclusion is that, in the case of generating binaural scenes with an application, the process proved time-consuming. This activity requires many attempts, tuning the parameters of each sound until the desired result is achieved. Recording scenes is, by nature, a faster method.

However, in the future, we plan to enhance the application to test more scenes. We also aim to explore separate applications where this method might be more suitable than other solutions, such as recording scenes from an environment or using libraries designed for computer games.

References

1. American Foundation for the Blind: Exploring new horizons: Binaural soundscapes. American Foundation for the Blind Blog Post (Jan 2024)
2. Brunet, L., Darses, F., Auvray, M.: Strategies and needs of blind pedestrians during urban navigation. Le travail humain **81**(2), 141–171 (2018)
3. Dakopoulos, D., Bourbakis, N.G.: Wearable obstacle avoidance electronic travel aids for blind: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **40**(1), 25–35 (2010)
4. Fujita, K., Sato, T., Inooka, T., Minato, T.: Bone conduction auditory navigation device for blind people. Applied Sciences **11**(8), 3356 (2021). `https://doi.org/10.3390/app11083356`, `https://www.mdpi.com/2076-3417/11/8/3356`
5. Hajebi, A., Asghari, F., Fereydounian, A.: Autonomous navigation based on binaural guidance for people with visual impairment. ResearchGate (Pre-print/Conference)
6. Kaplan, H., Pyayt, A.: Fully digital audio haptic maps for individuals with blindness. Disabilities (2024)
7. Kish, D.: Daniel kish: How i use sonar to navigate the world (ted talk transcript) (2015)
8. Kleinberg, D., Yozevitch, R., Abekasis, I., Israel, Y., Holdengreber, E.: A haptic feedback system for spatial orientation in the visually impaired: A comprehensive approach. IEEE Sensors Letters **7**, 1–4 (2023)
9. Lahav, O.: Virtual reality systems as an orientation aid for people who are blind to acquire new spatial information. Sensors (Basel, Switzerland) **22** (2022)
10. Madmoni, L., Ben-Hur, Z., Donley, J., Tourbabin, V., Rafaely, B.: Design and analysis of binaural signal matching with arbitrary microphone arrays and listener head rotations. EURASIP J. Audio Speech Music. Process. **2025**, 11 (2024)
11. Mikułowski, D., Pilski, M.: An approach for supporting space orientation of the blind using ontologically-based object map. In: Proceedings of the International Conference on Human Aspects of IT for the Aged Population
12. Mikułowski, D., Szuba, B.: Supporting the development of spatial orientation skills of blind people using binaural sounds in VSR game. Studia Informatica. System and Information Technology
13. Miller, J., Godfroy-Cooper, M., Wenzel, E.: Using Published HRTFS With SLAB3D: Metric-Based Database Selection and Phenomena Observed (2014). `https://doi.org/10.13140/2.1.1379.6489`
14. Organisation, W.H.: Blindness and vision impairment, `https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment`, online; [Access: 28 Now. 2025]
15. Ratnarajah, A., Manocha, D.: Listen2scene: Interactive material-aware binaural sound propagation for reconstructed 3d scenes. 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) pp. 254–264 (2023)
16. Skulimowski, P., et al.: Haptic and auditory cues: a study on independent navigation for visually impaired individuals. Journal on Multimodal User Interfaces (2025). `https://doi.org/10.1007/s12193-025-00463-2`
17. Sohl-Dickstein, J., et al.: A device for human ultrasonic echolocation. IEEE Transactions on Biomedical Engineering **62**(6), 1526–1533 (2015)
18. Specjalny Ośrodek Szkolno Wychowawczy dla Dzieci Niewidomych w Owińskach: Park orientacji przestrzennej, `https://www.niewidomi.edu.pl/park-orientacji-przestrzennej/`, online; [Access: 03 Sep. 2025]
19. Thaler, L.: Echolocation may have real-life advantages for blind people: an analysis of survey data. Frontiers in Physiology **4**, 98 (2013)

20. Wysocki, A.S.: Projekt i implementacja aplikacji do tworzenia dźwiękowych scen 3D dla potrzeb wspomagania nauki orientacji przestrzennej osób niewidomych zrealizowana z użyciem biblioteki OpenAL. Master thesis (2025)
21. Xu, J., et al.: Multimodal navigation and virtual companion system: A wearable device assisting blind people in independent travel. Sensors **25**, 4223 (2025)
22. Ye, Y., et al.: A wearable vision-to-audio sensory substitution device for blind assistance and the correlated neural substrates. In: Journal of Physics: Conference Series. vol. 1229, pp. 12026–12026 (2019)
23. Zhu, H.Y., et al.: An investigation into the effectiveness of using acoustic touch to assist people who are blind. PLoS ONE **18**(10), e0290431 (2023)