Visual-auditory substitution device for indoor navigation based on fast visual marker detection

Florian Scalvini, Camille Bordeau, Maxime Ambard, Cyrille Migniot, Stéphane Argon, Julien Dubois

Visual-auditory substitution device for indoor navigation based on fast visual marker detection

Florian SCALVINI
ImViA EA 7535
Univ. Bourgogne-Franche-Comté
Dijon, France
florian.scalvini@u-bourgogne.fr

Camille BORDEAU
LEAD CNRS UMR 5022
Univ. Bourgogne-Franche-Comté
Dijon, France
camille.bordeau@u-bourgogne.fr

Maxime AMBARD
LEAD CNRS UMR 5022
Univ. Bourgogne-Franche-Comté
Dijon, France
maxime.ambard@u-bourgogne.fr

Cyrille MIGNIOT
ImViA EA 7535
Univ. Bourgogne-Franche-Comté
Dijon, France
cyrille.migniot@u-bourgogne.fr

Stéphane ARGON
LEAD CNRS UMR 5022
Univ. Bourgogne-Franche-Comté
Dijon, France
stephane.argon@gmail.com

Julien DUBOIS
ImViA EA 7535
Univ. Bourgogne-Franche-Comté
Dijon, France
julien.dubois@u-bourgogne.fr

***Abstract*—This paper proposes a new navigation device to assist visually impaired people reach a defined destination safely in the indoor environment. This approach based on visual-auditory substitution provides the user a 2D spatial sound perception of the destination and of nearby and dangerous obstacles. Visual markers are placed at several relevant locations to create a mesh of the building where each marker is visually accessible from another marker. A graph representation of markers locations and their connection to each other defines by a way finding algorithm the shortest path reach to the wished position. The navigation task is achieved by moving from visual marker to visual marker until the desired destination is reached. These markers can be used independently of any other system or in addition to other solutions based on geolocalisation and/or a digital building model. Moreover, further information can be associated to the markers, and therefore verbalize to the user for instance a temporary hazards, a door presence or any other usual displayed information. The passive visual markers enables to deploy easily and quickly a scalable and low-cost solution to "signpost" the environment for users. Combined with our real-time implemented obstacle detection, their analysis enables the navigational abilities of visually impaired people to be improved.**

***Index Terms*—Auditory sensory substitution , Wearable assistive device, Navigation aid, Obstacle avoidance, Visual impairment, Sonification, Visual marker detection**

I. Introduction

A recent contribution to the WHO initiative, *VISION 2020: The Right to Sight* [1], estimate an increase of the number of people with visual impairment from 43.3 millions in 2020 to 61 millions in 2050 of blind people. This trend is worldwide and affects all groups of visually impaired people, from the low visually impairment to the totally blind. Indeed, this trend is caused by the increasing ageing and lack of treatment of visual degeneration such as glaucoma, presbyopia, etc. The increasing number of blind and severely visually impaired people, poses a challenge to enable them to lead an independent life while moving about safely. In fact, the variety of situations encountered in everyday navigation not only disrupts blind people's understanding of their destination but also, and above all, exposes them to risks such as the presence of a dangerous area such as a staircase, a static or mobile obstacle.

The ability to navigate in an environment is a fundamental element for an independent life. In daily life, a person is required to move frequently for various reasons such as work, leisure, etc. A person uses the spatial information acquired through their sensory modalities to orient themselves and move from one position to another. Although all sensory modalities are necessary to fully understand a scene, they do not provide the same level of spatial information. Vision is the most important sense for perceiving one's spatial environment. Indeed, it is easy to imagine how difficult it would be to navigate safely without visual perception in a familiar or unfamiliar environment.

A visually impaired person compensates by developing the spatial perception abilities of the other sensory modalities like the olfactory or the auditory sense, but some information is still not perceived. The degree of compensation varies according to the early age of onset of blindness. As the human body cannot fully compensate for this handicap, external means have been proposed to improve their navigation.

The classic methods used to remedy these problems are the white cane, the trained dog or the trained guide. Although these methods are widely used and effective for safe travel, they have limitations. The range of the white cane is limited to nearby objects, trained dogs and the guide require long and intensive training and are relatively expensive. In addition, due to the limitation of these means and the increase in the number of visually impaired people, the need for new technologies to enhance the spatial information perceived by a sighted person has increased. These devices or systems,

Thanks to the Conseil Régional de Bourgogne Franche-Comté, France and the Fond Européen de Développement Régional (FEDER) which are supporting financially this research.

called assistive technologies, must be ergonomic and have low latency to detect moving objects.

Assistive technologies are electronic systems designed to transmit information acquired by an artificial sensor to the blind person. In a review of different assistive technologies, [2] categorized devices according to the technologies used: vision replacement, vision enhancement, and vision substitution. Vision replacement and vision enhancement have some limitations. The first category relies on displaying information directly to the visual cortex of the brain and requires medical intervention, while the second category excludes people with severe visual impairment. Unlike other categories, Visual Substitution Devices (VSD) can be worn without limitation of use. In fact, the process transcribes visual information to another unimpaired sensory modality with out-of-body communication.

The auditory and tactile sensory modalities are primarily employed to transcode relevant information. Tactile-auditory substitution devices emit haptic feedback corresponding to a visual event while visual-auditory substitution devices emit a sound with verbal or non-verbal information. The VSD assist visually impaired persons in many tasks or situations of daily life, such as avoiding obstacles in the environment, defining the path to a desired location, or determining the precise position of the user. These three functionalities allow the devices to be grouped into three subcategories referred to in the literature as: Electronic Travel Assistance (ETA), Electronic Orientation Assistance (EOA) and Position Location Devices (PLD).

Fig. 1. Schematic view of the navigation aid system.

In this paper, we propose a new indoor EOA device based on a spatialized visual-auditory substitution approach. A navigation aid for visually impaired people provides information on the current trajectory, but above all, it must consider the presence or absence of obstacles that obstruct the user's progress. In addition, the system must be responsive with fast data processing and short sound emission to allow smooth movement to the desired destination. In response to this constraint, we proposed a system designed to be fast, easily integrated into an existing building, and easily expandable without economic cost. We propose a navigation method based on a mesh of the building by printable visual markers where the user navigates by intermediate beacons to reach the desired destination. The figure 1 shows a schematic view of the system operation in which a person receives a louder sound to the right indicating the relative position of the marker. The obstacle detection required for any navigation method is extracted from the depth map provided by the RGB-D camera. The 3D positions of the path and nearby obstacles are sonified into separate short 2D spatialized sounds. We propose to merge these separate information into a single sound in order to reduce the sound emission time, but also to have simultaneously the information on the trajectory and the spatial environment.

II. Related work

Navigational aids for blind people operate mainly by tracking the user's movements in real-time and guiding with feedback to the desired destination. However, the diversity of the visual environment or specific task space makes it difficult to establish a universal method of navigational aid. Indeed, some outdoor systems are based on the use of GPS (Global Positioning System) signals [3], [4], but the accuracy of GPS is limited and does not allow to precisely define the user's position in a building. Moreover, the problem is amplified in an indoor environment with the degradation of the GPS signal. Although GPS cannot be used to locate accurately in an indoor environment, research on location methods based on other types of waves and protocols has been performed [5]. Some used passive radio frequency identification (RFID) tag to guide the user from tag to tag until the final destination is reached [6] [7]. The predefined path is obtained by a path finding algorithm. The short operating range of these beacons requires a relatively large number of beacons to cover an entire space, which increases the cost of installation. Similar systems replace the passive RFID tag with an active tag or a Bluetooth beacon [8] [9]. However, despite the increased information flow and range of the beacons compared to passive devices. The use of active transmitters and receivers requires permanent battery power and therefore maintenance. Other competing systems use ultra-wideband (UWB) sensors to track the person in the environment [10]. This technology is more robust and the operating range of UWB (50-60m) allows it to be deployed in large spaces, but the unit cost of UWB transceivers is relatively higher than that of an RFID system for small spaces.

Similarly, the camera-based system [11] [12] is less complex to implement in any building with only an electronic device carried by the visually impaired person. Indeed, a visual marker detection evenly distributed in the spatial environment substitutes the use of electronic beacon. Detecting markers in visual space is resource and time-consuming and therefore disrupts the smoothness of navigation. In addition, a large distance between the user and a marker reduces the ability

to detect markers. [13] proposes a hybrid system with a visual marker and an active RFID is used to detect possible distant markers. However, the size of the marker is one of the main reasons for this difficulty, and the use of markers of varying size depending on the distance solves this problem.
The information, provided by the sensors, is enriched by semantic information obtained from the building information model (BIM) [14]. Indeed, a BIM representation combine with sensor data allow to represent the current position in the building space. Moreover, the BIM provides semantic information about the building configuration and the presence of hazards such as stairs and emergency exits.
Furthermore, knowledge of a fixed or mobile obstacle or more generally of a danger must be taken into account by the assisted navigation system. An obstacle will require a deviation from the initially predefined trajectory by bypassing it through a clear area. ETA devices are designed to identify obstacles around users. Some ETA methods guide the path to follow to circumvent the obstacle by indicating the area of free visual space via the subdivision of the space into distinct zones [15] or by using dynamic path methods inspired by the field of robotics based for example on an optimization of ant colonies [16]. Other methods just transmit information on the positions of obstacles [3] in order to allow a better understanding of the spatial environment and better freedom of movement.
Visual-auditory sensory substitution systems implement a protocol for transcoding the desired trajectory into an understandable sound. Most navigation aids provide verbal information to effectively guide the user [12] [17]. These approaches do not require special training for the handling of the system and allow a wide range of semantic description of the environment. In contrast, other methods propose to spatialize the information by emitting short spatialized stereophonic sounds to guide the user [18], [19]. This results in a shorter transmission time and therefore a lower latency. The ability to localize a static or moving object using spatialized stereophonic sounds has been studied.
The localization and sound generation processing is complex and requires an appropriate processing unit in order to obtain low latency information to ease navigation and the detection of possible dangerous zones. Approaches propose to perform processing via cloud computing [12] [17] in order to obtain a device with better ergonomics, performance and autonomy instead of an offline platform. Nevertheless, this type of device is subject to communication disturbances linked to the presence of a white spot in a building.
We propose a dynamic aid device with embedded processing, accurate in its indications, robust to disconnection, easily deployable and costless in all existing buildings. For this purpose, our proposed visual-auditory navigation aid system is offline, based on a detection of visual markers with an emission of short and spatialized 2D sounds.

III. Method

Our camera navigation system locates visual markers placed in the building. The markers are positioned at different points of interest such doors and in such a way another marker position is observable. In fact, the door crossing is an essential element for a useful navigation aid system in an indoor context. A spatial distribution of markers allows the user to move from marker to marker until he or she reaches the desired destination. Meshing the building with markers allows for a graphical representation of the environment with their associated connections and thus allows for the use of a graph-based pathfinding algorithm. However, the path finding algorithm requires that the nodes be identifiable and therefore the visual marker symbol used must be unique. The overall operation of our navigation's system is described in the figure 2 and below:

- **Initialization:** The path finding algorithm requires knowledge of the start node and the end node to initialize. The selection of the start node is performed by searching for a visual marker in the user's near space. After detecting the start node, the user is prompted for the desired destination.
- **User movement:** The user moves to the selected marker following a spatialized sound symbolizing the position in a sound space of the visual marker.
- **Marker reach:** When a marker is located within close enough to the user, the system checks if the marker is referenced as a point of interest and alerts the user if an action is required, such as opening a door. Then, if the marker reached is not the final destination, the user is prompted to scan the environment again to find the next marker.

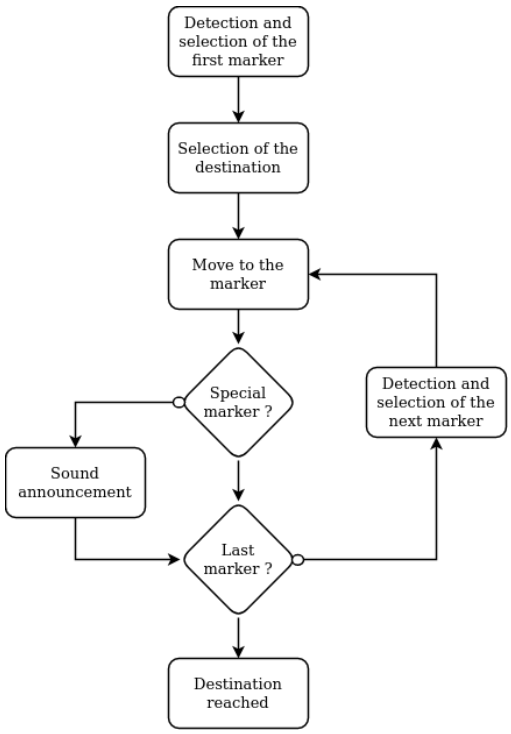

Fig. 2. Diagram of the navigation algorithm.

The detection and decoding of markers is a critical step in the operation of the system. The absence of detection prevents the user from continuing his journey, while the slow processing of visual information makes navigation less fluid or even jerky. Consequently, the choice of a marker with a unique and identifiable symbol as well as a robust detection method is necessary. We have chosen the STag fiduciary marker system [20] designed to be fast, stable, robust to detection of distant marker, to partial occlusion against difficult viewing angle conditions. Indeed, the detection of Stag markers is combined on an extraction of geometric features such as ellipse, corner and edge with a refinement of the homography to allow a real-time processing. In addition, the large number of individual markers in the libraries allows deployment in large indoor spaces. An example of three separate STags is provided in figure 3.

Fig. 3. Example of different STag visual markers [20]

Our undirected graph representation of the building mesh consists of nodes that are linked to each other without knowing the distance between them. An unweighted graph allows us to use a Depth First Search (DFS) or Breadth-First Search (BFS) brute force path finding algorithm to obtain the path to the destination. These algorithms are similar in their method with the search for the first occurrence of the arrival node by traversing the graph from the departure node. These algorithms are similar in their method with a graph traversal from the start node to the first occurrence of the end node. The main distinction is the traversal mode: DFS traverses the graph in depth while BFS traverses the graph in width. Although DFS is quicker and needs less memory than BFS to obtain a solution, the obtained solution is suboptimal in contrast to BFS. In our method, we have privileged the BFS algorithm in order to propose to the visually impaired person the shortest path to reach the destination.

However, a linear navigation is impossible in complex environment with various multiple static or moving obstacles such as an indoor space. We have integrated in our EOA method informations about nearby obstacle of the user in order to increase the scene understanding and to avoid the obstacle. Moreover, a better scene understanding allow more freedom of movement in the navigation task. Nearby obstacles are extracted from the depth video stream of the RGB-D camera with a threshold of elements located in the visual scene within one meter of the user.

Our sonification method is based on a low latency auditory substitution approach [21] where each visual position is associated with a short spatialized stereophonic audio pixel. A spatialize sound allows a precise localization of an area of interest without emitting a long verbal expression which increases the delay between two successive information and thus increases the danger of navigation. In the interest of limiting the computational delay, we have pre-computed for each pixel position the corresponding audio pixel. The spatialization of a given pixel into an audio pixel is based on the convolution of a monophonic sound with the impulse response associated with the corresponding azimuthal position in the head-related impulse response (HRIR) data set. A HRIR is a impulse response with mimic natural deformation made by body of the listener mainly the head on the sound to know its origin. The coordinate of the pixel is mapped into a spherical coordinate according to expression (1). With *fov* the azimuthal field of view of the camera, *pos* the position of the pixel and *size* the width of the image in pixel.

$$angle = \frac{2 * pos - size}{size * fov} \tag{1}$$

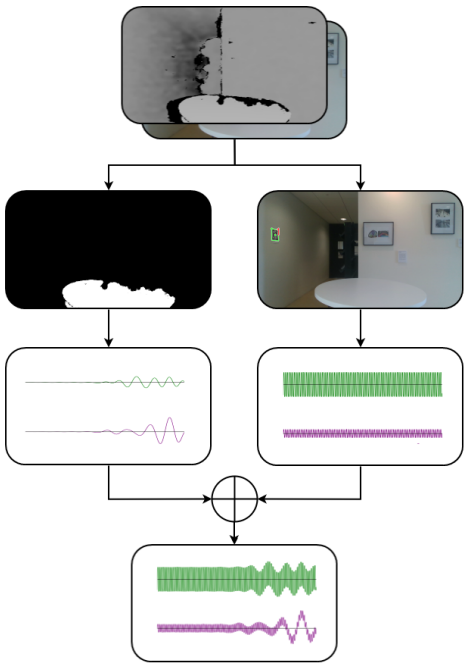

Fig. 4. Video processing and sonification pipeline for obstacle detection (left column) and marker detection (right column). The green and purple colours symbolise the left and right channels of stereo sound respectively.

The visual data is a combination of path and obstacle information. However, both information must be easily understandable and differentiable by the user. In addition, all information must be provided to the user at the same time, without alternation, to enable safe and smooth navigation. The combination of visual information into a unique understandable auditory signal is the cornerstone of our EOA method. To each of these visuals information, we associate a unique and brief monophonic sound, easily identifiable. Then we spatialized these sounds using a two-metre HRIR. Finally, we add the resulting sounds together to obtain an overall audio sound. The figure 4 summarizes

our visual processing and sonification method. The obstacle detection (left) and the path direction (right) information are sonified into two distinct spatialized sounds. The green colour and the purple colour symbolise respectively the left and right channel of the stereophonics sound. Obstacle detection (left) and path direction (right) information are sonified into two distinct spatialized sounds. The green and purple colours symbolize the left and right channels of a stereophonic sound. The last graph represents the sum of these sounds in an overall sound.The difference in amplitude between the left and right channels on the navigation sound indicates that a marker is located on the right. On the obstacle sound, the left and right channels are similar and symbolize an obstacle in the centre and at the bottom. Two distinct monophonic sounds are used to spatialize and distinguish auditory information. The obstacle monophonic sound have a lower frequency than the navigation monophonic sound.

In addition to these spatialized sounds and due to the multiple steps to reach the desired destination. We have added verbal information about the current stage but also information if a door has to be passed or a lift has to be used. Semantic sounds are pre-recorded and are played at the beginning or end of a step.

IV. System

The pipeline of our system is similar to that of other sensory substitution devices, with an acquisition step performed by a camera, followed by processing of the visual data by a computational unit, and then transcription and output into auditory information by an audio device. The videos are acquired by an Intel RealSense D435i stereoscopic camera with a depth field of view (FOV) of 87° × 58° and a colour camera FOV of 69° × 42°. The colour and depth image were synchronously captured and realigned with a resolution of 1280 × 720 pixels at 30 frames per second. The camera module is placed at eye level on electronic glasses to allow the blind person to scan the scene with a head movement. The processing unit used is a standard laptop in a backpack with the Ubuntu 20.04 operating system, an Intel Core i7-6700HQ processor (4 Cores - 8 Threads with a frequency of 2.60 GHz), 8 GB of RAM, and a Nvidia GTX 1070 mobile graphics card. Bluetooth's headphones are used to transmit auditory information.

The software is developed in C++ using the LibRealsense2 , OpenCV libraries for video acquisition and processing and Advanced Linux Sound Architecture libraries (ALSA) for writing the sound driver. Visual marker detection is performed on a colour image converted to grey scale required by the STag detection algorithm, and with the highest possible input resolution provided by the camera in order to identify the marker pattern over a large distance. The audio pixel sonification of nearby objects is performed at a resolution of 160 x 120 pixels. The resolution is limited by the HRIR dataset used to spatialize a sound, but mostly by the inability of the human ear to distinguish a small angular variation of the sound emission.

A realistic daily navigation aid system for visually impaired people must meet response time constraints. Indeed, a significant processing delay between the acquisition and the associated sound emission can disturb the user, or even cause serious injuries if the information of an obstacle is not received in time by the user. The limitation of delays is therefore a necessity, for which we have optimized our method with preloaded sound and multi-threaded approaches. The program is split into 4 threads where each one performs the following function:

- **Acquisition thread:** Acquisition of the depth and colour image, the alignment the depth map to the colour image.
- **Video processing thread:** Detection and localization of the STag markers in the RGB image and of the nearby objects on the depth map.
- **Manager thread:** Based on our navigation method, this thread generates an audio sound according to a visual information given by the video processing thread or a verbal sound expression.
- **Sound thread:** Loading the generated sound or generated speech to be transmitted to the user according to the navigation manager.

Our multithreaded approach shown in figure 5 is based on the sequential operation of sensory substitution systems. The navigation manager is the central node of the system that links the video processing stream and the audio output stream. The separation of the audio and video processes reduces the system delay, so that while the audio is being transmitted, the next visual information is already being processed. In addition, this separation allows continuous sound to be output without disturbing the user with an absence of sound. A lack of sound may be due to a slowdown in the video stream or to images being lost during acquisition. The sound is emitted in such a way that if new visual information has not yet arrived, then information from nearby objects is re-emitted.

The table I summarises the impact of the different video and audio processes during a navigation between two markers. The processing time for the sonification of nearby objects and the detection of STag markers varies slightly with the number of nearby elements or markers in the visual scene. Our measurement of processing times was realized in a scene with a marker displayed on the wall at 60 cm. The detection time of an environment is given for the detection of a single STag marker, which represents the general use case of the system with markers evenly distributed in the spatial environment. The overall system time of approximately 78 ms allows satisfactory navigation in an indoor space. Furthermore, our EOA is a CPU-only implementation and does not take advantage of the fact that most of the algorithms used are sequential and allow GPU optimization. An implementation of the segment detection algorithm used for marker detection with CUDA shows a speed-up of $\times 12$ [22] and an overall detection speed-up estimated at $\times 4.4$ on older GPUs [20]. GPU optimization limits CPU usage, reduces overall system time, and could enable real-time hardware implementation on a low-power

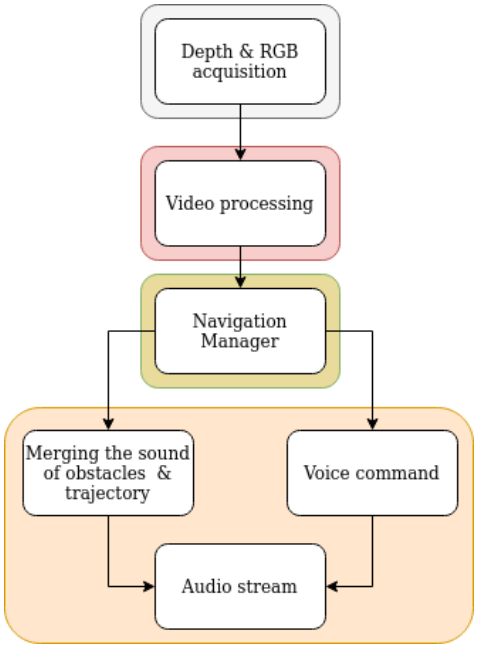

Fig. 5. Implementation's diagram

embedded target with a GPU module such as the Nvidia Jetson cards.

TABLE I
SYSTEM'S PROCESSING TIME ON THE LAPTOP TARGET (TIME IN MILLISECOND)

Video processing		**Sonification**		
STag	Nearby Obs.	STag	Nearby Obs.	Sound fusion
57	1.4	0.02	17.5	1.9

V. EXPERIMENTATION

We measured the capabilities offered by our navigation aid system with a short sound emission combining a path direction and an obstacle information. As a proof-of-concept, we asked a blindfolded person with no specific training with a visual-auditory substitution device to reach an unknown destination with our system within a marked building. The destination is entered by a supervisor without the user being informed of this information. Before the start of the experiment, a brief explanation of the system and of the two sound information emitted was given. For this purpose, we created a realistic scene within a marked building with obstacles placed. Indeed, a realistic experimentation of an indoor navigation system requires an environment that represents the daily life of visually impaired people.

The experimental space is a part of the 3rd floor of the I3M building (ImViA & LEAD Laboratories, Dijon). This space contains several static obstacles (chairs, desks, tables, ...) with 6 markers printed on a A4 paper and placed at relevant position in order to create a mesh the building. The figure 6 shows the graph representation of the mesh used to compute the way

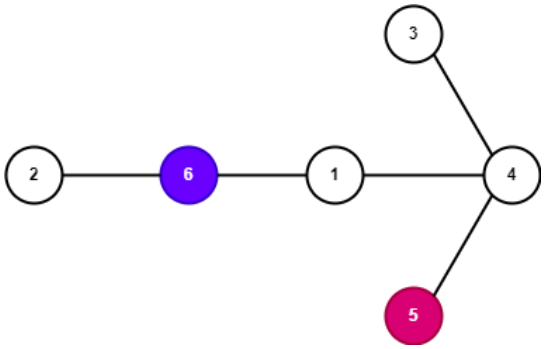

Fig. 6. Graph representation of the indoor space used for the experimentation. The possible navigation between two nodes is symbolized by a link. The purple node represents a close door and the pink node a lift.

finding algorithm to reach the desired destination. A door is indicated with a purple node.

The starting position of the blindfolded person is aligned with the marker 2 in the room and the unknown position is the floor's lift indicated by the marker 5. The path predefined by the way finding algorithm to reach this destination from the initial position is a transit through the clues $2 \rightarrow 6 \rightarrow 1 \rightarrow 4 \rightarrow 5$. The user had to pass through an open door to reach marker 2 and open the door indicated by marker 6 to reach marker 1. Finally, it had to avoid various static obstacles positioned between markers 1 - 4 and 4 - 5. Figure 7 represents a top view of the floor building captured by multiple Lidar (laser imaging detection and ranging) scans with information about the positions of the visual markers (green and purple arrows), the participant's starting (red dot) and destination (pink dot) positions, and the path followed (white line).

The path taken to reach marker 5 was recorded in order to analyse the viability of our indoor navigation aid system and specifically the user's behaviour when encountering an obstacle. The relative position of the person in the building was recorded using an Oculus Quest 2 headset in parallel with our system. The oscillations of the trajectory in the path followed by the user are due to the swaying of the human being when walking but also to the movements of the user's head necessary to scan the environment and possible obstacles. Indeed, the inertial measurement unit sensor (IMU) is positioned in front of the user's eyes (Oculus Quest 2 computing unit) and is therefore sensitive to head movements. The maximum distance between the current marker and the person before moving to the next marker is 80 cm. The path followed by the blindfolded person represents the difficulties of the route. Indeed, the person backs up slightly to locate himself before opening the door, then detects and avoids the tables to reach markers 4 and 5. The experiment can be successfully reproduced and the results shows that a user could navigate in an outdoor environment by avoiding static obstacles (including the wall) or mobile obstacles (i.e anyone walking past) and passing through a door to reach the desired destination or passing through a door to reach the desired destination using our visual to auditory sensory substitution device. In addition,

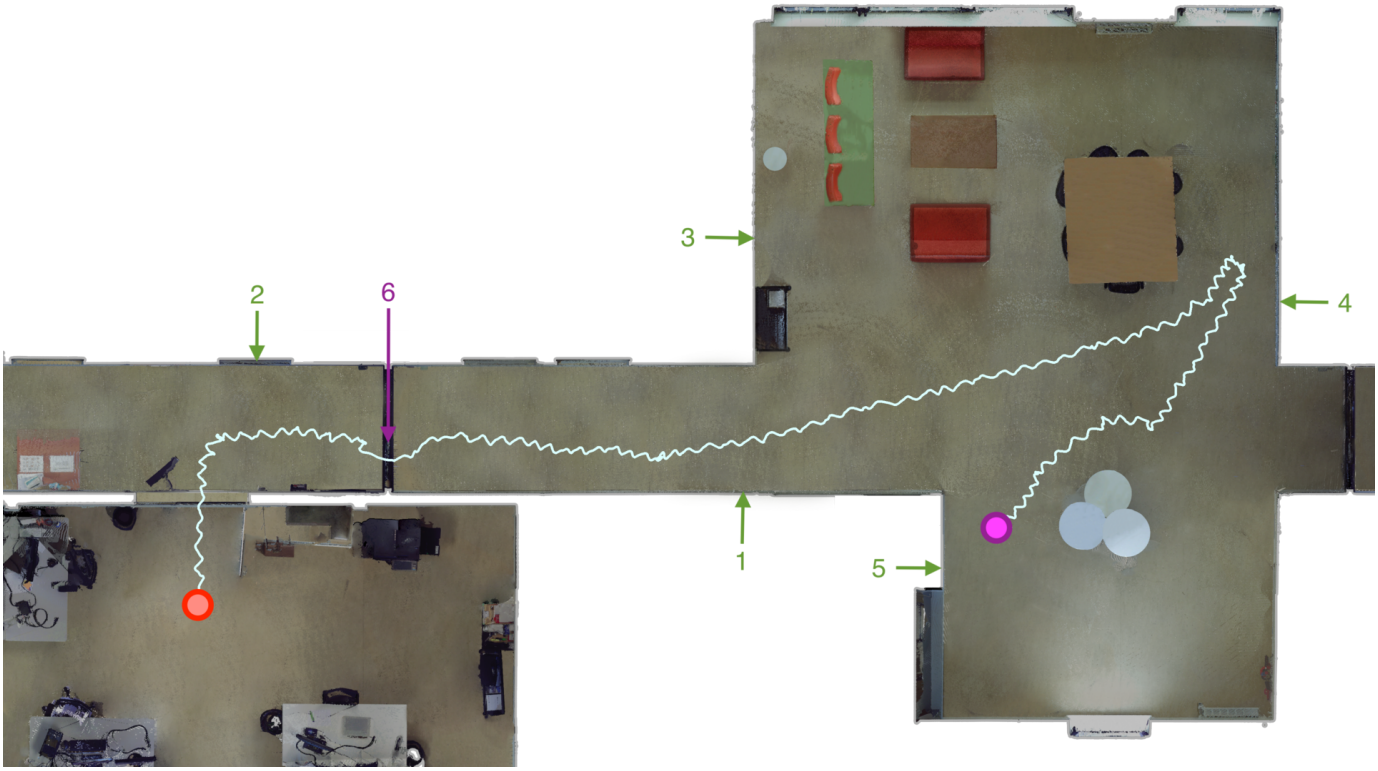

Fig. 7. Top view of the user's movements to reach the desired destination. The red dot and pink points indicate the starting position and the destination respectively. The arrows represent the position of the markers in the building space and the purple arrow indicates that the marker is placed on a door. The number associated to an arrow represents the tag number.

additional information can be included in any visual marker to be verbalized.

VI. Conclusion & Future work

Our proposed indoor navigation system allows a blind person to reach a desired destination in a building by guiding 2D spatialized sounds based on the recognition of unique visual markers. In addition to trajectory information, the presence of nearby obstacles or key points, such as a door, is provided. Our method can be adapted to any building or room and does not require remote transmission to achieve real-time performance. However, this system have some limitations. Indeed, the detection distance of the marker is limited in terms of distance, and moreover is sensitive to marker occlusion by dynamic obstacles. In addition, the position of the camera at the eye level cause a issue with the detection of small obstacle on the floor. In fact, reason is the height between the feet and the eyes combined with a low vertical camera's FOV makes it impossible to detect a obstacle. This problem could be solved by looking down from time-to-time to check whether an object is present or not. The use of a conventional white cane in addition to our method could also solve this problem. In fact, a white cane is the opposite of our obstacle detection with an excellent low obstacle but ineffective for high obstacles. The design of our EOA is not ergonomic for the user, with excessive weight and power consumption due to the use of a laptop as a computing unit. An optimization of our method on a low-power device with a GPU implementation is possible. Furthermore, the obtrusive nature of the ear canal with headphones could interfere with the understanding of the natural auditory scene. However, a replacement with bone conduction earphones avoids the auditory obstruction. The navigation method can be enhanced with an integration of voice command processing to interact with the system in order to obtain some additional information about the spatial scene. The semantic information could be a brief description of the person's location, or a description of the type of obstacle on the path if the user requests it. Moreover, a fusion with the building information model could provide information on the spatial organisation of the building and the presence of hazardous areas such as stairs. Finally, a cognitive psychology study could be considered to investigate the user's behaviour with our sonification method.

References

[1] Bourne and al., "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study," *The Lancet Global Health*, vol. 9, no. 2, pp. 130–143, Feb. 2021.

[2] W. Elmannai and K. Elleithy, "Sensor-Based Assistive Devices for Visually-Impaired People: Current Status, Challenges, and Future Directions," *Sensors*, vol. 17, no. 3, p. 565, Mar. 2017.

[3] M. Poggi and S. Mattoccia, "A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning," in *IEEE Symposium on Computers and Communication (ISCC)*. Messina, Italy: IEEE, Jun. 2016, pp. 208–213.

[4] A. Brilhault, S. Kammoun, O. Gutierrez, P. Truillet, and C. Jouffrais, "Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning," in *IFIP International Conference on New Technologies, Mobility and Security*. Paris: IEEE, Feb 2011, pp. 1–5.

[5] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A Survey on Wireless Indoor Localization from the Device Perspective," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–31, Nov. 2016.

[6] C. Tsirmpas, A. Rompas, O. Fokou, and D. Koutsouris, "An indoor navigation system for visually impaired and elderly people based on Radio Frequency Identification (RFID)," *Information Sciences*, vol. 320, pp. 288–305, Nov. 2015.

[7] R. Ivanov, "Indoor navigation system for visually impaired," in *International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies - CompSysTech '10*. Sofia, Bulgaria: ACM Press, 2010, p. 143.

[8] S. A. Cheraghi, V. Namboodiri, and L. Walker, "GuideBeacon: Beacon-based indoor wayfinding for the blind, visually impaired, and disoriented," in *IEEE International Conference on Pervasive Computing and Communications (PerCom)*. Kona, Big Island, HI, USA: IEEE, Mar. 2017, pp. 121–130.

[9] D. Ahmetovic, C. Gleason, C. Ruan, K. Kitani, H. Takagi, and C. Asakawa, "NavCog: a navigational cognitive assistant for the blind," in *International Conference on Human-Computer Interaction with Mobile Devices and Services*. Florence Italy: ACM, Sep. 2016, pp. 90–99.

[10] A. Martinez-Sala, F. Losilla, J. Sánchez-Aarnoutse, and J. García-Haro, "Design, Implementation and Evaluation of an Indoor Navigation System for Visually Impaired People," *Sensors*, vol. 15, no. 12, pp. 32 168–32 187, Dec. 2015.

[11] G. E. Legge, P. J. Beckmann, B. S. Tjan, G. Havey, K. Kramer, D. Rolkosky, R. Gage, M. Chen, S. Puchakayala, and A. Rangarajan, "Indoor Navigation by People with Visual Impairment Using a Digital Sign System," *PLoS ONE*, vol. 8, no. 10, p. 76783, Oct. 2013.

[12] Y.-J. Chang, S.-K. Tsai, and T.-Y. Wang, "A context aware handheld wayfinding system for individuals with cognitive impairments," in *International ACM SIGACCESS conference on Computers and accessibility - Assets '08*. Halifax, Nova Scotia, Canada: ACM Press, 2008, p. 27.

[13] S. Alghamdi, R. van Schyndel, and A. Alahmadi, "Indoor navigational aid using active RFID and QR-code for sighted and blind people," in *IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. Melbourne, VIC: IEEE, Apr. 2013, pp. 18–22.

[14] R. Marroquin, J. Dubois, and C. Nicolle, "Ontology for a panoptes building: Exploiting contextual information and a smart camera network," in *Semantic Web*, Aug 2018, vol. 9, p. 803 – 828.

[15] W. M. Elmannai and K. M. Elleithy, "A highly accurate and reliable data fusion framework for guiding the visually impaired," *IEEE Access*, vol. 6, pp. 33 029–33 054, Mar 2018.

[16] A. Benabid Najjar, A. Rashed Al-Issa, and M. Hosny, "Dynamic indoor path planning for the visually impaired," *Journal of King Saud University - Computer and Information Sciences*, p. S1319157822000751, Mar. 2022.

[17] J. Bai, D. Liu, G. Su, and Z. Fu, "A Cloud and Vision-based Navigation System Used for Blind People," in *Proceedings of the 2017 International Conference on Artificial Intelligence, Automation and Control Technologies - AIACT '17*. Wuhan, China: ACM Press, 2017, pp. 1–6.

[18] M. J. Proulx, P. Stoerig, E. Ludowig, and I. Knoll, "Seeing 'where' through the ears: Effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution," *PLOS ONE*, vol. 3, no. 3, pp. 1–8, Mar 2008.

[19] F. Scalvini, C. Bordeau, M. Ambard, C. Migniot, and J. Dubois, "Low-latency human-computer auditory interface based on real-time vision analysis," in *ICASSP - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 36–40.

[20] B. Benligiray, C. Topal, and C. Akinlar, "Stag: A stable fiducial marker system," *Image and Vision Computing*, vol. 89, pp. 158–169, Sep 2019.

[21] M. Ambard, "Software Design for Low-Latency Visuo-Auditory Sensory Substitution on Mobile Devices," *Computer and Information Science*, vol. 10, no. 2, p. 1, 2017.

[22] O. Ozsen, C. Topal, and C. Akinlar, "Parallelizing edge drawing algorithm on CUDA," in *IEEE International Conference on Emerging Signal Processing Applications*. Las Vegas, NV: IEEE, Jan. 2012, pp. 79–82.