

# Learnability in Audio-Spatial Working Memory

Sanchit Aggarwal, Koustav Ghosal, Pulkit Singhal, and Priyanka Srivastava

International Institute of Information Technology,  
Gachibowli, Hyderabad, India  
{Sanchit.Aggarwal@research.,Koustav.Ghosal@research.,  
Pulkit\_S@students.,Priyanka.Srivastava@}iiit.ac.in  
<http://iiit.ac.in/>

**Abstract.** Previous studies have shown that visual navigation can be effectively described by audio cues using pan and pitch to associate directions during navigation. However, to best of our knowledge no study has emphasized the importance of training method and its interaction with working memory for using sound representation for navigation. Our study is concerned with three aspects related to the above task. Firstly, we have tried to confirm how combination of pitch and pan can effectively represent directional sound cues. Secondly, we studied the capacity limit of Audio Spatial Working Memory. Thirdly, we exploited training methods to improve the capacity limit of audio stimuli. We conducted an experiment to measure the learnability by employing various training methods and its influence on audio-spatial working memory capacity by varying the set size from 2- 8. The experimental results are consistent with the previous findings showing that the sound cues attributed with pan and pitch can be instrumental in the construction of an efficient audio spatial map. In addition, the pilot results show that learnability increases the working memory capacity.

**Keywords:** Auditory Cognition, Visually Impaired, Working Memory, Learnability, Sonification, Navigation System

## 1 INTRODUCTION

We live in a world where every moment we come across numerous known and unknown pieces of visual information. We process this information and perform tasks like reading, playing soccer, navigation etc. The visually impaired people, however, use other modalities like audio, tactile etc. to perform the same tasks. One of the basic needs of the visually impaired is a robust navigational aid in unfamiliar indoor environments. The assistive tool should not only help in giving orientational information but should also have the least load in terms of learnability, where learnability refers to how easily the user can learn the system. Various approaches have surfaced over the years with an aim to design proficient navigation system but very few are able to set the tradeoff between the feasibility and learnability. Other limitations such as hardware and installation cost are associated with these systems [9, 10, 14–16]. The main question, which

has to be dealt with the design of any assistive system, is its efficiency in terms of learnability.

In general, the human visual system is much more sophisticated than the auditory system [13]. Consequently the same tasks require more effort when performed with auditory system alone than with both auditory and visual systems. In addition, Merat *et al.* [12] has established that accurate auditory localization can be achieved by the filtering effects of the pinnae and is due to differences in the timing and intensity of a sound source heard by each ear. This provides an understanding of using simple tonal sounds to represent direction for task like navigation. Again, Collignon *et al.* [17] found that in case of visual impairment the brain can rewire itself, showing an impressive range of cross-modal plasticity. They observed that blind individuals may demonstrate exceptional abilities in auditory spatial processing of sounds. Also the occurrence of cross-modal reorganization in the mature human brain has been studied by Kujala *et al.* [18]. Their experimental results suggested that plasticity between sensory modalities is possible even in adults and it was found that individuals blinded after childhood exhibited enhanced sensitivity to pitch-change discrimination. It can be inferred from these studies that the auditory system of visually impaired people becomes more sophisticated than those having normal vision due to cross-modal plasticity. Their sense of sound localization and parameter estimation is therefore, much better [17, 18]. Hence, a blindfolded person performs worse than a visually impaired person in tasks like navigation. Any assistive navigation system ought to consider this characteristic and it must not override the the natural ability of visually challenged to perceive the surroundings.

In this study we have tried to explore the possibility of using sound to convey direction to the visually challenged individuals. In a simple task like navigation, they have to deal with the fundamental challenges of understanding directional orientation and familiarization with the environment. Creation of a coherent spatial memory map of the environment is a challenging task for them. This calls for a robust system which can facilitate the user by providing visual information of the surrounding in an alternate modality.

Our work is organized as follows. In Section 1.1, and Section 1.2, we have given an account of the problem being addressed in this paper. In Section 1.3, we have mentioned some of the pioneering works in this area. In Section 2, 3, experiment and method have been described in detail. In Section 4, the results and the subsequent analysis have been carried out.

## 1.1 SONIFICATION

The existing *Assistive Navigation Systems* can be classified into two broad categories.

- Text to Speech (TTS) systems
- Sonified Systems

In both these systems, firstly, the data is obtained from GPS or GIS, infrared sensors, sonars or any other form of unimodal or multimodal sensing device [8–

10]. Then the data is either converted to speech or simple non-speech sounds, accordingly, and played to the user's ears. While both TTS and *Sonified Systems* have their own advantages and disadvantages [2–5], our study deals with the later. As described by Dingler *et al.* [1], the sonified systems can be further divided into three categories:

- *Auditory Icons* : These are symbolic sounds associated with different objects in a scene. In this case a dictionary of auditory symbols are created. E.g. sounds of tree branches rattling in the wind might represent trees.
- *Earcons* : These are the non speech sounds that are generated for some specific purpose like navigation. This is the area we will explore in the next sections.
- *Spearcons* : A mixture of speech and earcons, Spearcons have shown better results than auditory icons for recognizing objects.

Since our main focus lies on navigation, in this study, we have used *earcons* and varied the two different parameters *pan* and *pitch* to represent *X* and *Y* directions respectively, in two dimensions. High and Low Pitch ( with both pan ) represents *+ve* and *-ve* *Y* directions respectively. Similarly, Left and Right pan ( medium pitch ) represents *-ve* and *+ve* *X* directions respectively. The mapping procedure has been elaborately explained in Section 3.4.

## 1.2 AUDIO-SPATIAL WORKING MEMORY

Our study is based on the assumption that we have a map of the path that we intend to represent using earcons. Now, in any system, if the beeps or cues are produced at every step then the user has to continuously concentrate on the cues. Consequently, the natural superiority of the visually challenged people, in terms of sound perception is overridden. Giving more information ( multiple cues ) at a time and having silent intervals between two successive stimuli, can be a solution to this problem. The user can exploit her natural ability to perceive the surroundings during this interval. The amount of information that the user is able to remember depends on her *audio spatial working memory*.

In our study, we have tried to find out a limit to the amount of information that can be conveyed at once to the user. We have tried to validate the possibility of an improvement of this capacity using proper training methods. Our results suggested a trend of increasing capacity supporting our claim. After observing this trend, we tried to look for the best method to train a user with the different sound parameters. The *goodness* of the method was measured in terms of response time and number of trials required .

## 1.3 RELATED WORK

There are many systems which have been built in recent times as assistive navigation systems.

Wilson *et al.* [5] developed, a System for Wearable Audio Navigation (SWAN) to

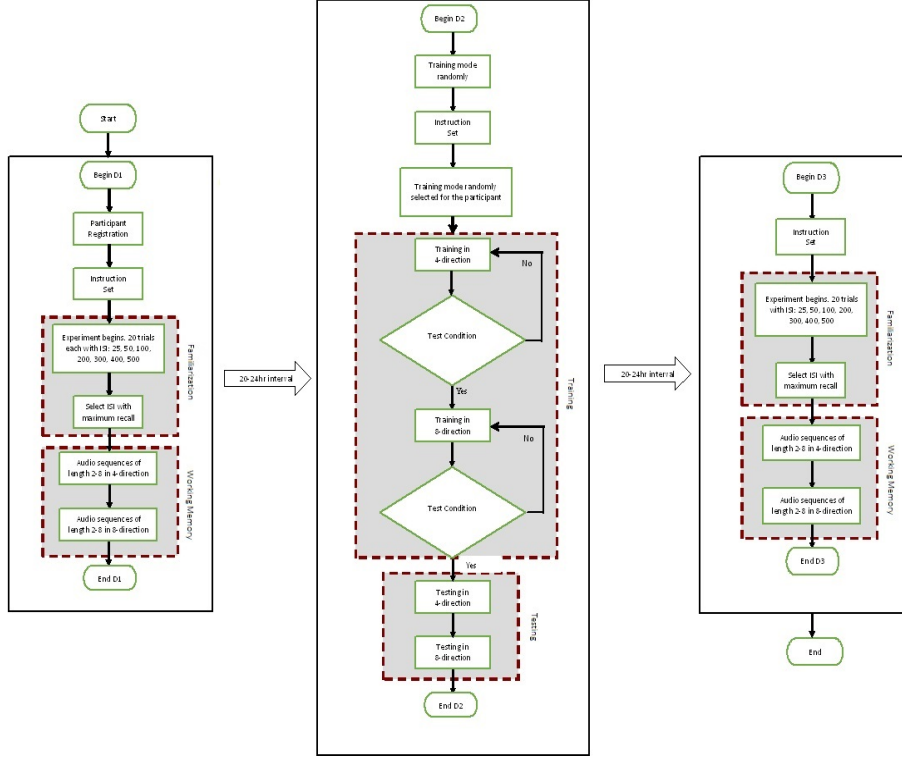
serve as a navigation and orientation aid for persons with vision temporarily or permanently impaired. SWAN is a wearable computer consisting of audio-only output and tactile input via a task-specific hand-held interface device. SWAN aids a user in safe pedestrian navigation and includes the ability for the user to author new GIS data relevant to their needs of way-finding, obstacle avoidance, and situational awareness support. Brock *et al.* [6] have a system designed to help blind people navigate around obstacles. Their system perceives the environment in front of the user using a depth camera (a Microsoft Kinect). The system identifies nearby structures from the depth map and uses sonification to convey obstacle information to the user. Tsubasa *et al.* [7] proposed a framework to aid visually impaired with the recognition of objects in an image by sonifying image edge features and distance-to-edge maps. Kurniawan *et al.* [10] designed a spatial audio system that is capable of modelling the acoustic response of a closed environment with varying sizes and textures. They evaluated the orientation and mobility of the system in a controlled experiment using seven distinct sounds in three environments. A infrared sensors based system which enables to identify user's trajectory and locate possible obstacles along the route in unfamiliar environments with the use of Augmented Objects Development Process methodology was presented by Guerrero *et al.* [9].

With the cumulative understanding that seems to transpire from the above systems, we have endeavoured to reflect the similar aspects of navigating in unfamiliar indoor environments. Along with the above, our understanding of the complexity that lies in the designs of these systems and the cost involved, directs us to the establishment of a requirement of a better system. Also the *learnability* is one aspect which has not been explored in the existing systems. We have attempted to address this problem differently, in the following way. Firstly, there have been various investigations marking differences between auditory and visual working memory. It has been found that humans have relatively low capacity limit of auditory information as compared to the visual information. Apparently, we haven't come across much work in the direction of improving the capacity limit of auditory working memory. Secondly, a comparative study among the different training methods have been conducted.

## 2 EXPERIMENT

An experiment was designed to measure the learnability employing various training methods (Sec. 3.4) and its influence on audio-spatial working memory capacity. The experiment was conducted by controlling training methods (no training, non-supervised, supervised by varying visual and auditory feedback) and audio-spatial working memory capacity limit varying from 2 - 8. It was a mixed design: 4 (training method: no training, non-supervision, visual and audio supervision) as between group  $\times$  7 (working memory capacity limit varying from 2 - 8) as within group.

A three phase experiment (Sec. 3.4) was set up in order to study the effect of



**Fig. 1.** Illustrating the Three phase Experiment. D1 and D3 are done to see the Working memory before and after association of sounds with directions. D2 is the phase where participants are subjected with different modes of training. Test condition is achieved when participant scores a accuracy of 80% consecutively for three schemas.

learnability on audio-spatial working memory. The first phase (Sec. 3.4) aimed to check the audio-spatial working memory capacity before any *association* of the sound cues with the 8 directions. Association here refers to the construction of spatial mapping of different sounds with different directions in mind. Second phase (Sec. 3.4) employs different training methods to construct the spatial map. In addition, it was also done in order to investigate the fastest mode of training for association of the sound cues with 8 directions. Third phase (Sec. 3.4) aimed to check the increase in audio-spatial working memory capacity limit due to training or construcion of spatial map. The flow diagram of the Experiment is depicted in Fig 1. Generally, the presence of semantic association with the stimuli might require fewer cognitive resources than the absence of this association. Hence, there might be some difference in audio-spatial capacity before and after this association. Therefore, we hypothesize that association of sounds with different directions through training increases the capacity limit of audio-spatial

working memory. The experiment investigates the effect of construction of the semantic association for sound cue with 8 directions employing various training modes.

### 3 METHOD

#### 3.1 Ethic Statement

All participants were asked to fill the consent form before starting the experiment. They were briefed about the three days requirement for the study. They were allowed to refuse to complete the study at any point during the experiment.

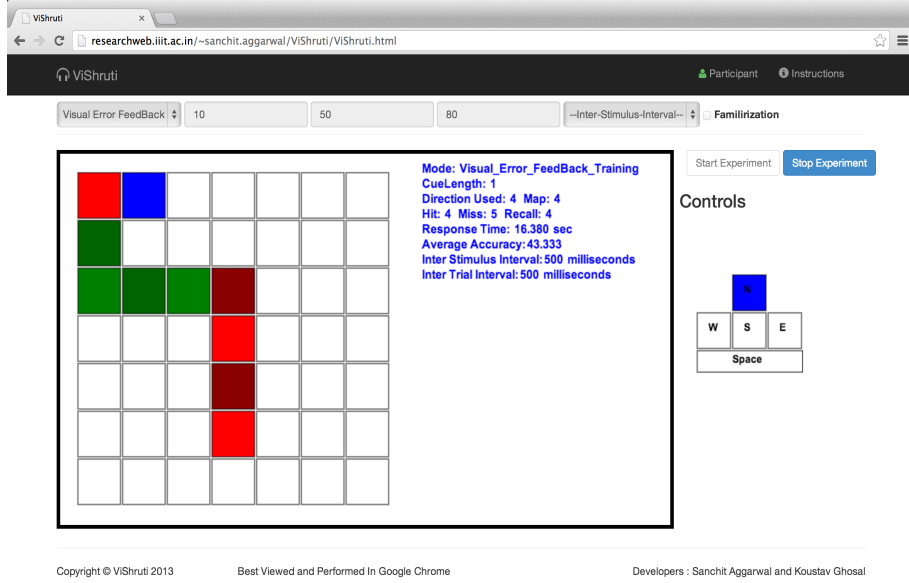
#### 3.2 Participants

A total of 24 participants volunteered (16 male and 8 female) for the experiments. Average age of all the participants was 23.5. The participants were randomly divided into four groups. They were called for the three phases after a gap of 20 to 24 hours between the phases. Each phase was of 15-25 minutes of duration. The participants were randomly assigned to one of the 4 groups i.e. Audio Error Feedback, Visual Error Feedback, Unsupervised and No-training. They were instructed about the task during each phase. All experiments were performed in experimental lab conditions with minimal noise intervention from the surroundings.

#### 3.3 Apparatus and Stimuli

**Apparatus:** The navigation task is simulated by the system named *ViShruti*. The system illustrated in Figure 2. is developed using Java Script, HTML5 and Riffle Sound Library. It is a simulation of the real world indoor environment where schema of the walk-able regions within the environment is known. The task is to navigate from source to destination with 4 or 8 directional sound cues. The system simulated navigation in 2-dimensions. The sound cues are represented by the combination of *pan* and *pitch*. Pan refers to the channel in which sound is being played and pitch is the frequency of the sound. To represent the 8 directions, three categories of pitch are generated and are combined with the two channels i.e. the ear in which they are played. A high pitch sound when heard from both the ears is mapped with north direction and when heard from left and right, it is mapped with northwest and northeast respectively. Similarly, a medium pitch sound when heard from left or right is mapped with west and east direction. The low pitch sound when heard from both the ears is mapped with south and when heard from left or right is mapped with south west and south east direction.

The system generates a map that is similar to the schema of the indoor environment with known walk-able path. The whole path is divided into unit distance. For each unit distance user is presented with its directional sound cue and is



**Fig. 2.** *ViShruti*: Simulation of real world indoor environment. Each cell in the grid represent unit distance. The colored path represent the walkable region. The figure illustrate the 4-Direction Visual Error Feedback Mode. Blue cell is the starting position, for each colored cell Audio Stimuli for corresponding 4-Direction (N,E,W,S) was given. Participant had to respond by pressing corresponding Arrow Keys (Up,Right,Left,Down). Green Cells Represents the correct response. Red Cells represents incorrect response.

asked to press the corresponding arrow key or number-pad key on the keyboard for 4-Direction or 8-Direction mode respectively. In 4-Direction Mode, Arrow keys (Up,Down,Left,Right) correspond to Directions (N,S,E,W) respectively. In 8-Direction mode, Number-pad keys (7,8,9,4,5,6,1,2,3) corresponds to the directions (NW, N, NE, W, E, SW, S, SE).

**Stimuli:** The stimuli were presented using headphones. Three simple tones (sine waves) of frequency 440Hz, 880 Hz, and 1760 Hz were generated with a sample rate of 44100 per sec by Riffle Sound Library. Two adjacent tones had a frequency difference of 50%. These tones were combined with Pan i.e the 2-channels to produce 8 different audio tones. Lowest frequency of 440 Hz played in Left, Both and Right channels represented South West, South and South East respectively. Similarly Medium frequency of 880 Hz played in Left and Right channels represented West and East respectively. Highest frequency of 1760 Hz played in Left, Both and Right channels represented North West, North and North East respectively. Each tone was played for 500 ms. The tones were kept as far as possible (spanned about in three octaves) so that they are easy to discriminate and represent 8-directions. In addition, a silent tone of duration varying from

25ms to 500 ms is also generated and presented as a blank stimuli for Inter Trial and Inter Stimulus Intervals.

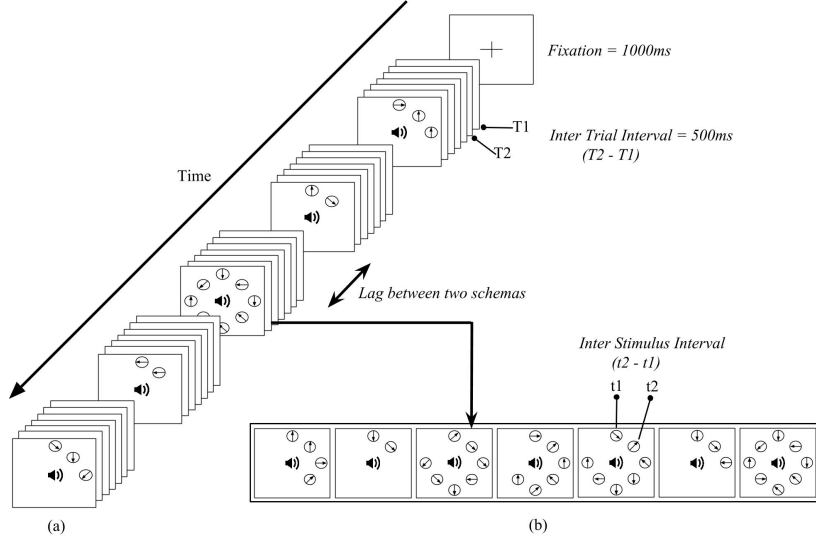
### 3.4 Design and Procedure

We conducted a three-day experiment as illustrated in Fig 1. to test our hypothesis. The first day (D1) was intended to check the audio spatial working memory limit without any association of the directional cues. This represented the capacity limit without any training or association of sound cues with direction. The second day (D2) aimed at training the participants and creating the spatial map of directional cues with different modes of learning. The third day (D3) was again directed towards the checking of the capacity limit of audio spatial working memory after establishing the association with the directional cues and to check which mode of training has caused the increase in the capacity limit.

The division of experiment into three days provides the opportunity to see the effect of priming on the working memory and to check the efficiency of sound attributed with pan and pitch to represent directional cues for the navigation task.

**Audio Spatial Working Memory before association (D1) :** This is being conducted to gauge the capacity limit of audio spatial working memory when the receiver doesn't have any association with the directional cues. The users were given multiple directional cues of length 2,3,4,5,6,7 and 8 in random order. The task was to remember the directional cues and recall them in the same sequence as they were presented by pressing the corresponding keys. The inter stimuli time between the sound cues for each trial of the sequence was determined individually for each participant in a practice session that preceded the main experimental session. In practice session, an informal staircasing procedure was used to find the maximal inter stimuli time. Participants were presented with 20 trials of varying length (2-3) sequences. the inter-stimulus time varied from 25ms, 50ms, 100ms, 200ms, 300ms, 400ms, 500ms during the staircase procedure. The time for which they were able to recall the maximum sequence was selected as the Inter-Stimuli time for the remaining experiment where they were presented with sequence length varying upto 8 cues. As illustrated in fig. 3, the experimental session consisted of a total of 35 trials of varied length (2-8) sequences with inter-trial time of 500 ms. The task was further divided between 4 and 8 directional sound cues. The participants were first asked to perform for 4 directional sound cues where only the sounds corresponding to East, West, North and South were presented with varying length from 2-8, and then they were asked to recall. Thereafter the complexity of working memory task is increased by introducing all the 8 directional sound cues i.e. North, South, East, West, North-East, North-west, South-east, South-west .



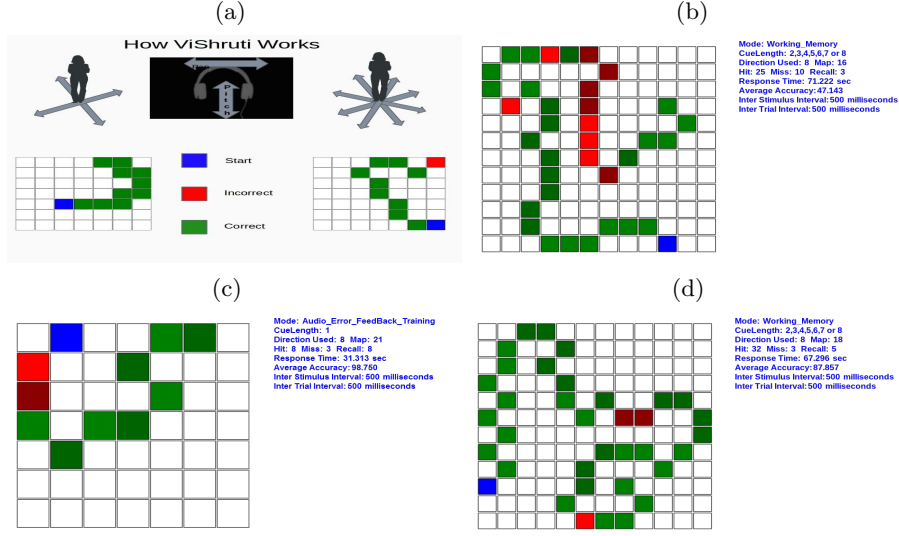


**Fig. 3.** Illustration of Working memory phase described in Sec. 3.4 and Sec. 3.4. (a) 5 schemas were presented to each participant. (b) 7 trials presented per schema in Working Memory with each sequence of length (2-8) occurring once per schema in random fashion. Inter Stimulus interval was determined from the staircasing procedure for each participant. Inter trial interval was set to 500 ms. Total 35 trials spanning over 5 schemas were presented for both 4 and 8 directions respectively.

**Association of Sound Cues with Direction by various training modes (D2) :** This is done with the intention to determine the effectiveness of pan and pitch as combination to represent 8 directional sound cues. It also helps in creating the spatial map among the participants. To achieve the above we used three training modes. All the participants were divided into 4 groups. Participants from 3 out of 4 groups were provided with supervised or unsupervised mode of training while those from fourth group were not subjected to any training. Following are the modes of training which were randomly assigned to the participants.

**Audio Error Feedback Mode:** In this mode, the participants were given audio stimuli and whenever they made a mistake, a feedback in form of audio buzz was generated. They were provided with a single directional cues per trial and were asked to respond by pressing the corresponding key. If they pressed the correct key, the next cue was presented. In case of wrong key, an error sound (i.e. the buzz) followed.

**Visual Error Feedback Mode:** Here the participants were given audio cues but the error feedback was given visually. The empty schema with hidden path was shown to the participant. If they pressed the wrong key for a given cue, the



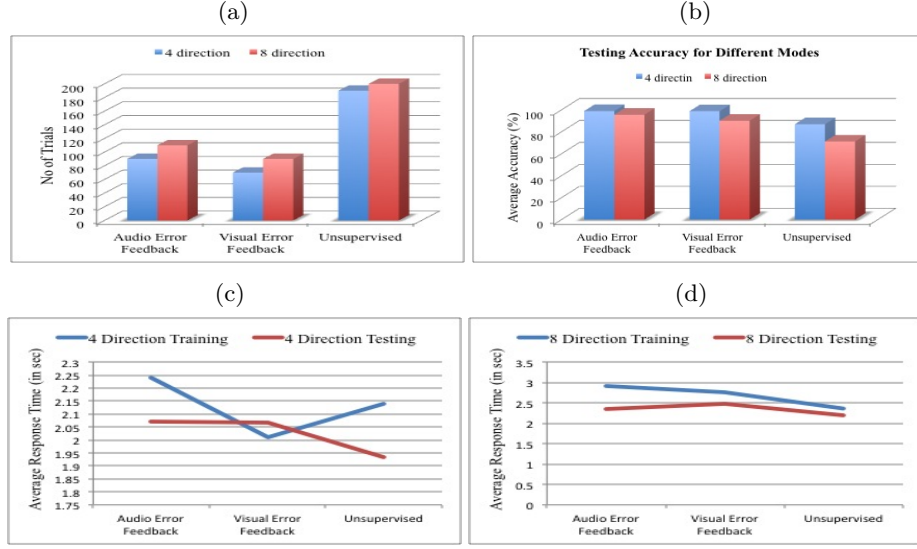
**Fig. 4.** Illustrating the working of ViShruti and data recorded for analysis a) *Pitch and Pan representation for Directions.* b) Schema recorded while performing Working Memory Capacity before association, D1 (Sec. 3.4) by a participant. c) *Schema recorded while performing audio error feedback training for 8 directions (Sec. 3.4).* d) Schema recorded while performing Working Memory after association, D3Sec. 3.4) respectively.

current unit step became red and the next cue was played. If the response was right, the unit step became green which was followed by the next cue.

**Un-Supervised Mode:** In this training mode, there was neither visual nor audio feedback and only one audio stimulus is presented per trial. The user had to give the response and learn implicitly the differences among the three categories of sound pitch with combination of pan for 8 directions. The implicit learning occurred by repeated trials and participant had to fix a baseline for each direction.

**No-Training Mode:** In this mode, no feedback was given. It was assumed that implicit training happens even when trials were presented to participants. Therefore participants were directly tested.

Each mode of training was conducted for minimum of 5 schemas having random paths. Each schema had 10 trials of single audio stimulus. After crossing minimum number of schemas, the participants were trained until they achieved accuracy greater than 80% per schema for consecutive three schemas, i.e 8 out of 10 trials were responded correctly. The training was given with 4-directions followed by 8 directions. To validate whether the association has been developed by the participant, they were subjected to testing followed by training after a brief rest of 5 minutes. There were 3 schemas with 50 trials each during testing for 4



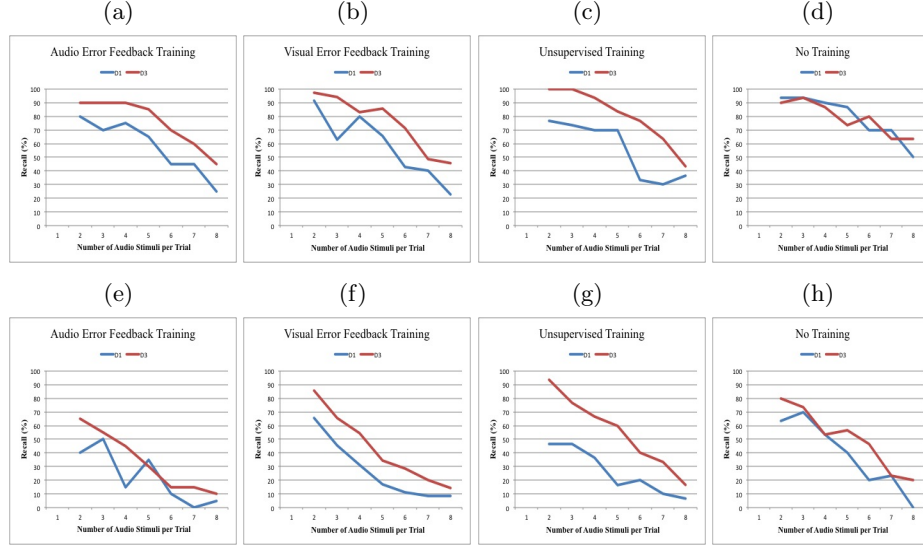
**Fig. 5.** D2 Results. a) *No of Training trials vs Training Mode.* b) *Average Testing Accuracy vs Training Mode.* c) *Average Response Time vs Training Mode for 4-Direction Training and Testing.* d) *Average Response Time vs Training Mode for 8-Direction Training and Testing.* As observed from the data the Visual Error Feedback training requires least number of training trial followed by Audio error feedback and unsupervised for both 4 direction and 8 directions respectively.

and 8 direction respectively. For both training and testing modes the schemas were saved and metrics were recorded.

**Effect of Training on capacity limit of Audio Spatial Working Memory (D3) :** This is being conducted to ascertain whether there is effect of training on the capacity limit of audio spatial working memory after the spatial map is constructed. The experimental session task was similar to that of D1 (Sec. 3.4)

### 3.5 Recordings

The image of the path traversed, hits (number of correct responses), response time and recall (in case of working memory) was recorded and analyzed. Figure 4. illustrates the data recorded in the three experimental sessions and working of ViShruti.



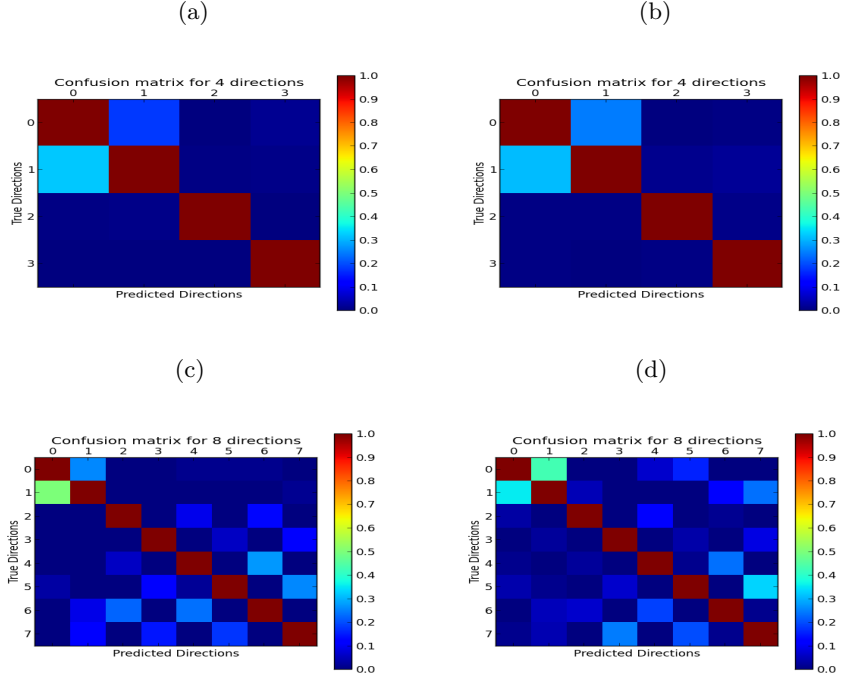
**Fig. 6.** Recall vs No. of Audio Stimuli per trial curves for D1 and D3 with different training modes averaged for all participants. *Row 1) 4- Direction:* (a) Audio Error Feedback, (b) Visual Error feedback, (c) Unsupervised, (d) No-training. *Row 2) 8-Direction:* (e) Audio Error Feedback, (f) Visual Error feedback, (g) Unsupervised, (h) No-training. The influence on the working memory capacity can be seen from difference between the D1 and D3 curves in 3 categories of training 3.4 i.e Audio error feedback, Visual Error feedback and unsupervised training for both complexity of 4 and 8 directions. No-Training didn't influence much as can be seen in d) and h)

## 4 RESULTS AND DISCUSSIONS

### 4.1 Effect of Training in increase in Working Memory Capacity Limit

From Figure 5, we can see that for participant of Audio Error Feedback, Visual Error Feedback, and Unsupervised groups, there is a significant increase in recall for different sequence length (2-8) from phase D1 to phase D3. While the participant who were subjected with no training doesn't show such a significant increase. Also the difference between the increase in 4-Direction and 8-Direction is due to the complexity involved in two levels. In case of 4-Direction, there are only 4 different sound which were presented and has to be recalled.

Hence the increase is much more in case of 4-Direction for different trainings as compared to 8-Directions, where there were 8 different stimuli presented. This significant difference illustrates that the association has been occurred within participants. As evident from the data, the training helped in construction of the spatial map of different sounds with different directions thus resulting in



**Fig. 7.** Confusion Matrices : Unsupervised Training and Testing. Directions 0,1,2,3,4,5,6 and 7 are North ,South,East, West,North-East,North-West,Soth-East and Soth-West respectively

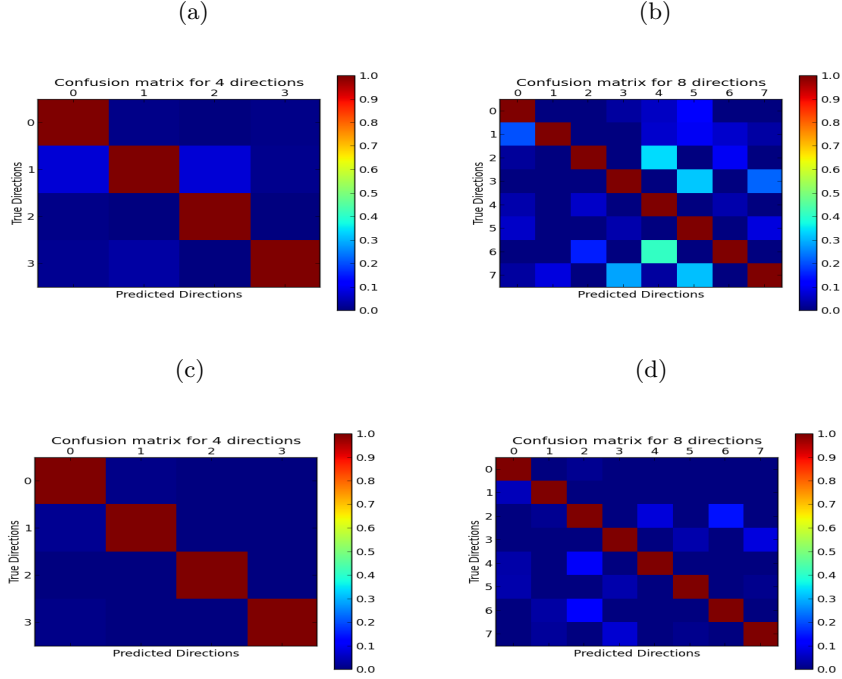
semantic association of sounds.This resulted in increase in recall for sequence length upto 8.

#### 4.2 Goodness of Training Mode

As observed from Figure 6. the number of trials needed to qualify for testing condition (80% accuracy in consecutive three schema) is in order, Visual Error Feedback training < Audio Error Feedback Training < Unsupervised Training, for both 4-directions and 8- directions respectively. In addition to this the Testing Accuracy for the three modes of training is also in same order i.e. Visual Error Feedback training == Audio Error Feedback Training < Unsupervised Training. This shows that Visual Error Feedback Training is able to construct the spatial map very fast. This is in accordance with the dependence of human being on visual modality as compared to audio modality.

#### 4.3 Effectiveness of Pan and Pitch to represent Directions

It can also be observed from Figure 6. that testing response time for all the training modes (Visual Error, Audio Error and Unsupervised) is less than the



**Fig. 8.** Confusion Matrices : Audio Error Training and Testing, Visual Error Training and Testing. Directions 0,1,2,3,4,5,6 and 7 are North ,South,East, West,North-East,North-West,Soth-East and Soth-West respectively

training response time. Also, the number of maps required to complete training, has been compared in Figures 5(a) and 5(b). It is evident from the figures that Visual Error Feedback Mode of training is the fastest mode of training.

#### 4.4 Common Confusions

In Figure 7, we have given the confusion matrices of training and testing process on second day of the experiment. It can be seen in Figure 6(a). that while pan was properly mapped as *X* direction and there were very few confusions, the participants had problem in identifying North and South directions using pitch. In 8-directions, the participants had difficulty in differentiating between NW-SW and NE-SE pairs. This observation leads to the conclusion that pitch has to be carefully adjusted to represent *Y* direction. In unsupervised mode, the baseline for *high* and *low* are fixed only after repeated trials with both the directions. But in other modes in Figure 8, due to the error feedbacks, the baseline was set at the very initial stage and this reduced the number of confusions.

## 5 CONCLUSION

The experiments were performed only with 24 participants. It is hard to generalize anything based on the data from such a small sample size. But a trend can easily be observed. Based on what has been observed, we came up with three main trends. a) The methods of training mentioned previously are indeed helps in constructiong the spatial map, which facilitated the increase in audio spatial working memory. b) Visual Error Feedback and Audio Error Feedback trainig methods are the best ways to train the users of this system. c) Combination of Pan and Pitch can be used effectively for representing the different directions and user can adapt to tese sounds easily with proper training.

## 6 FUTURE WORK

There are lots of other areas that can be explored from here that involve working memory. One such area is to study the effect of sound cues in presence of noise. This can be useful in real scenarios where users are subjected to surrounding noise while navigating. We plan to build a prototype of *ViShruti*, where the visual information about the schema of the indoor environment would come from a mounted camera and it will be processed to segment out walkable regions using machine learning and computer vision. The idea is to sonify the resultant information and present it to the user. Our main goal is to build a stable system which can provide user with both scene understanding and navigational aid using audio descriptors.

## References

1. Dingler, Tilman, Jeffrey Lindsay, and Bruce N. Walker. "Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech." In Proceedings of the 14th International Conference on Auditory Display, Paris, France, pp. 1-6. 2008.
2. Tran, Tuyen V., Tomasz Letowski, and Kim S. Abouchacra. "Evaluation of acoustic beacon characteristics for navigation tasks." *Ergonomics* 43, no. 6 (2000): 807-827.
3. Mowbray, GEORGE H. "Simultaneous vision and audition: the comprehension of prose passages with varying levels of difficulty." *Journal of Experimental Psychology* 46, no. 5 (1953): 365.
4. Wickens, Christopher D. *Engineering psychology and human performance*. Harper-Collins Publishers, 1992.
5. Wilson, Jeff, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. "Swan: System for wearable audio navigation." In *Wearable Computers, 2007 11th IEEE International Symposium on*, pp. 91-98. IEEE, 2007.
6. Brock, Michael, and Per Ola Kristensson. "Supporting blind navigation using depth sensing and sonification." In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pp. 255-258. ACM, 2013.
7. Yoshida, Tsubasa, Kris M. Kitani, Hideki Koike, Serge Belongie, and Kevin Schlei. "EdgeSonic: image feature sonification for the visually impaired." In *Proceedings of the 2nd Augmented Human International Conference*, p. 11. ACM, 2011.

8. Ivanov, Rosen. "Indoor navigation system for visually impaired." In Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies, pp. 143-149. ACM, 2010.
9. Guerrero, Luis A., Francisco Vasquez, and Sergio F. Ochoa. "An indoor navigation system for the visually impaired." *Sensors* 12, no. 6 (2012): 8236-8258.
10. Kurniawan, S. H., A. Sporka, V. Nemec, and P. Slavik. "Design and user evaluation of a spatial audio system for blind users." In Proceedings of the 5th International Conference on Disability, Virtual Reality & Assoc. Tech, pp. 175-182. 2004.
11. Colzato, Lorenza S., Michiel MA Spap, Merel M. Pannebakker, and Bernhard Hommel. "Working memory and the attentional blink: Blink size is predicted by individual differences in operation span." *Psychonomic Bulletin & Review* 14, no. 6 (2007): 1051-1057.
12. Merat, N., and J. A. Groeger. "Working memory and auditory localization: demand for central resources impairs performance." *The Quarterly Journal of Experimental Psychology Section A* 56, no. 3 (2003): 531-549.
13. Bigelow, James, and Amy Poremba. "Achilles Ear? Inferior Human Short-Term and Recognition Memory in the Auditory Modality." *PloS one* 9, no. 2 (2014): e89914.
14. Snchez, Jaime. "Mobile audio navigation interfaces for the blind." In *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, pp. 402-411. Springer Berlin Heidelberg, 2009.
15. Ran, Lisa, Sumi Helal, and Steve Moore. "Drishti: an integrated indoor/outdoor blind navigation system and service." In *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*, pp. 23-30. IEEE, 2004.
16. Jos, Joo, Miguel Farrajota, Joo MF Rodrigues, and JM Hans du Buf. "The SmartVision local navigation aid for blind and visually impaired persons." (2011).
17. Collignon, Olivier, Patrice Voss, Maryse Lassonde, and Franco Lepore. "Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects." *Experimental brain research* 192, no. 3 (2009): 343-358.
18. Kujala, T., K. Alho, M. Huottilainen, R. J. Ilmoniemi, A. Lehtokoski, A. Leinonen, T. Rinne et al. "Electrophysiological evidence for crossmodal plasticity in humans with early and late onset blindness." *Psychophysiology* 34, no. 2 (1997): 213-216.