Technical Report on the Development of a Journey Preparation Tool for People with Vision Impairment

Tuan D. Lai[1], Chris McCarthy[1]

[1]Swinburne University of Technology, Melbourne, Australia

Abstract - This report summarizes the development of a long-term scalable technology-enable solution to assist journey preparation for members of vision impaired community. The existing application allows users to familiarize themselves with a specific location in the CBD by letting them experience the immersive auditory-based simulator supported with state-of-the-art surround-sound technology called Ambisonic and dynamic interactive voice interface.

Keywords - *Cognitive Mapping; Assistive Technology; Ambisonic; Android Application, Web Application; Vision Impairment.*

I. INTRODUCTION

With almost one in five Australians experiencing some form of disability, a substantial proportion of the community faces challenges to actively participate in city life [1]. This project is developing a long-term scalable technology-enable solution to assist journey preparation for members of vision impaired community. A cross-platform application has been developed to simulate the sensory experience of Flinders Street Railway Station located in the Melbourne CBD. The simulator would allow users to rotate their body and experience the sounds change.

II. SURROUND-SOUND TECHNOLOGY

A. Ambisonic Technology

To achieve an "immersive" experience, Ambisonic Technology [2]: a full-sphere surround sound technique, in addition to the horizontal plane, it covers sound sources above and below the listener. 3D audio can be reproduced over many loudspeakers positioned all around the listener. For the scope of the project, the 3D audio is played through the headphone, the technology used is called Binaural Synthesis. It consists of simulating, at someone's eardrums, the same acoustical sound field produced as one which would have been produced by a real audio scene. This immersive 3D audio will assist people with vision impairment to visualize the surrounding areas in conjunction with improving their mental mapping skills.
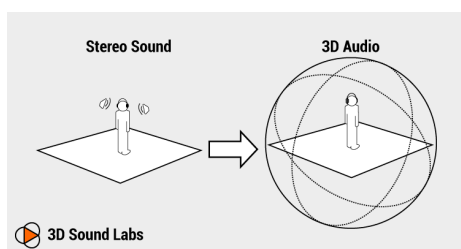

Fig. 1. Stereo sound and 3D audio comparison [3]

B. Recording Technique

Sounds are recorded by a tetrahedron microphone, which contains 4 capsules pointing toward 4 different directions (fig 2.1), to get 4-channel monophonic A-format. This is converted into 4-channel B-format, encoding the directional information of a given three-dimensional sound field to four channels called W, X, Y, Z using mathematical formula (fig 2.2)
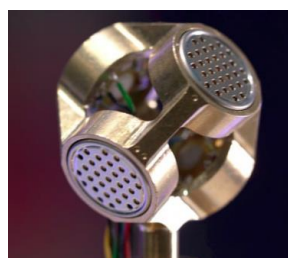
$$W = \frac{1}{k}\sum_{i=1}^{k} s_i\left[\frac{1}{\sqrt{(2)}}\right]$$

$$X = \frac{1}{k}\sum_{i=1}^{k} s_i\left[\cos\phi_i\cos\theta_i\right]$$

$$Y = \frac{1}{k}\sum_{i=1}^{k} s_i\left[\sin\phi_i\cos\theta_i\right]$$

$$Z = \frac{1}{k}\sum_{i=1}^{k} s_i\left[\sin\theta_i\right]$$

Fig. 2. (1) soundfield microphone

(2) B-format encoding formula [2]

Where s are mono audio signal recorded by each individual capsule we want to encode at the position Φ (horizontal angle (azimuth) phi), and Θ (vertical angle (elevation) theta) [2].

The microphone we utilized is called Sennheiser Ambeo VR Mic, an application is provided with the Mic with the capability of converting sound file from A to B. This software works as a VST2 or VST3 plugin, which can be used in any digital audio workstation (DAW) for editing purposes. Reaper software (a complete digital audio production application) is used within the project to generate a set of B-format audio files.


Fig. 3. Ambeo® A to B converter [4]

III. DEVELOPMENT

A. Web Application

The Ambisonic sound capsules have been encoded in 360 videos, recorded by Garmin Virb 360 Video Camera, using Facebook 360 Encoder (image on the right) and deployed on YouTube. To upload 360 videos with 3D audio on YouTube, the audio must be a 4 channels B-format with first-order ambiX, ACN ordering, and SN3D normalization. This will make the video compatible with VR mode, allowing users to use Google Cardboard or Oculus Rift to navigate and change direction.
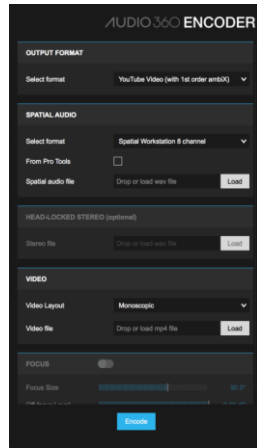
The video can also be rotated by sliding the video using a mouse or using WASD keys on the keyboard or using the controller on the top left of the video.

Users can interact with the interface using either voice or text commands and then narration will be played back upon user's request. This function is based on Web Speech API [6], a JavaScript API provided by Google Inc to enable web developers to incorporate speech recognition and synthesis into web pages. Here are current support commands that users can interact with the web interface via microphone.

- Help – provide a list of available commands for users
- What is this – describe the project and website
- Where am I – describe the place where the video was recorded
- Describe – describe the surrounding environment
- When recorded – show the time when the video was recorded
- Next location – play the next video
- Replay – replay the current video
- Hi/ Hello – guide user to help command.
- Start Exploring – play the first location (outside Flinders Street Station entrance)
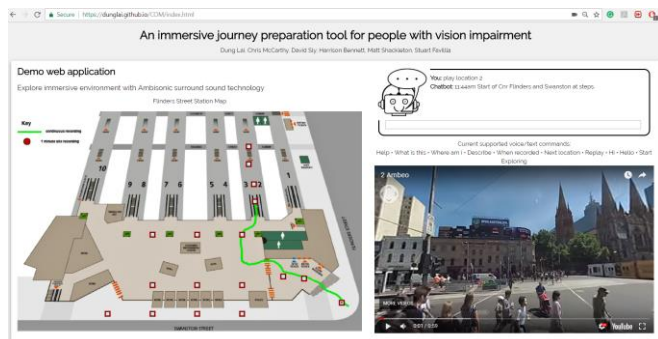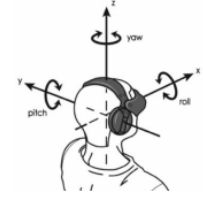

Fig. 4. Web application interface

B. Desktop Application

A Desktop application is also being developed using Max/MSP, a visual programming language, to explore the use of headtracking, allowing the tracking of yaw-pitch-roll value that determine the orientation of the head/body. This value is used to change sound field and generate directional narration.

Thanks to Stuart Favilla, the Hedrot head tracker is successfully implemented and the integration is shown in figure 4. By separate the Yaw value from the head tracker into 4 quarters representing front, left, right and back, the surround area will be described so that description is on-request and dynamically adjusted based on head direction.
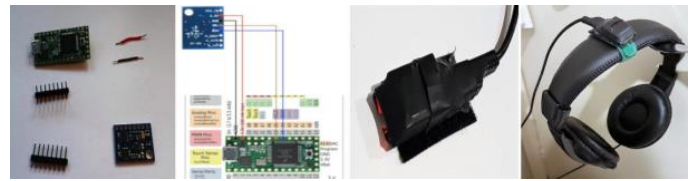

Fig. 5. Teensy board parts – Schematics – Build head tracker – Integration on the headphone

Fig. 6. Desktop application interface

C. Mobile Application

In the process of building a proof of concept in a variety of platforms, an Android application was developed, allowing users to experience immersive spatial audio recorded in various locations on Flinders Street Station. The application is capable of changing soundfield using hardware-based sensors which are physical components built into a handset or tablet device. Specifically, data is extracted from the geomagnetic field sensor in combination with the accelerometer using android sensor API [8] to determine a device's position relative to the magnetic north pole, this orientation data in quaternion format is then passed through Google VR audio engine to rotate Ambisonic soundfield.

The Google VR audio engine [7] allows the user to spatialize sound sources in 3D space, including distance and height cues. The GvrAudioEngine is capable of playing back spatial sound in two separate ways: Sound Object rendering, allows the user

to create a virtual sound source in 3D space, the second method allows the user to play back Ambisonic soundfield. The first method was used to create the artificial directional audio narration of surrounding environment and objects, the second method was used to playback Ambisonic audio files recorded in real-time using Sennheiser Ambeo VR Microphone.
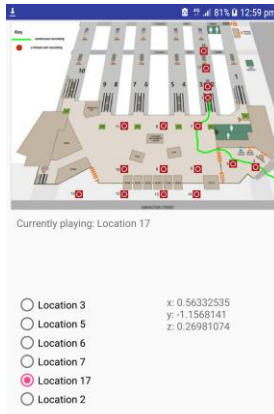

Fig. 7. Android application interface

IV. USER EXPERIENCE TESTING

The information being delivered by the app is refined constantly during the project to ensure that users will be benefited by the given information. A small consultation has been made with Lil Deverell – an orientation and mobility specialist with the aim of improving the quality of the overall system. It is suggested to add doppler effect in the platform, showing when and where trains are going, the transition between locations, the sound of cane, GPS sequence instructions in linear order. These factors will help users develop mental mapping skills. Furthermore, videos should be captured at different time of day and ideally, in crowded and pressure environment, crossroad route should also be recorded. This will prepare them with the real pressure of navigating around the city on their own and reduce their stress.

Improvement could also be done to existing directional description. The information provided need to have learned elements. The users should be notified when their orientation is perpendicular to the train lines or aligned with the train direction. Information such as tactile pavers and objects they can hear in real-time is also beneficial in making educated guesses. Information is not restricted to visual, non-visual cue such as smells, materials, temperature, wind, light pulse (for people with low vision), lift, entrance point, coffee shop sounds, barrier, reflect sound, consistent sound, and toilet. Additionally, information about the structure of the area will help users to get the big picture. Another significant detail that help reducing stress is preparing them to find convenient spots to wait for someone, this could be corners, spots where low vision people can lean on, solid objects that they can touch,

and balance surface, this prevents them from standing on the way of other people. Lastly, the source of assistance is also as important as other factors.

V. TECHNICAL CHALLENGES

Currently, the existing feature of the system has many technical aspects that need to be investigated. The Ambisonic playback on mobile application produces some form of white noise when the sound fields are being changed, audio processing techniques can be applied to reduce the level of white noise (the mixture of different frequency). Voice recognition and gesture recognition are useful features of the mobile application, it helps people with vision impairment interact with the application without other sources of assistance. Each Ambisonic B-format file contains 4 channels with high fidelity sound quality, this leads to the increase in memory if files are encoded in the Android app itself. However, the Android application has limited heap space, other approaches such as uploading sound files to the cloud and let users download them individually can solve the problem. Currently, users must put their Android device horizontally to correctly rotate the sound field, this is not effective and unusual for users. The application should allow users to point the device toward the directions that they want to rotate into.

REFERENCES

[1] Open Innovation Competition on Accessibility [http://www.melbourne.vic.gov.au/about-melbourne/melbourne-profile/smart-city/Pages/innovation-competition-city-accessibility.aspx]

[2] An Introduction to Higher Order Ambisonic by Florian Hollerweger, 2008 [http://flo.mur.at/writings/HOA-intro.pdf]

[3] 3D Audio, 3D sound lab [http://pro.3dsoundlabs.com/category/intro-to-3d-audio/]

[4] Ambeo software for virtual Reality [https://en-us.sennheiser.com/ambeo-blueprints-virtual-reality]

[5] Add Spatial Audio to a Video For YouTube [https://fb360spatialworkstation.zendesk.com/hc/en-us/articles/207888189-Add-Spatial-Audio-to-a-Video-For-YouTube]

[6] Web Speech API [https://w3c.github.io/speech-api/speechapi.html]

[7] Google VR Audio Engine for Android [https://developers.google.com/vr/reference/android/com/google/vr/sdk/audio/GvrAudioEngine]

[8] Android Sensor API [https://developer.android.com/guide/topics/sensors/sensors_overview.html]

[9] Hedrot headtracker [https://abaskind.github.io/hedrot/]