Binaural Sensory Augmentation

N. Phipps O'Neill, MECE 2020/2021, Dr. N. Murphy

***Abstract*—Designing and developing a binaural navigational device has the potential to improve how we navigate the world. The present work studied binaural navigational technologies in virtual environments, but research suggests significant benefits for such a technology in a real environment. In this paper we describe the implementation of a device that uses fundamental binaural principles to spatialize audio for a human listener. A wearable prototype device is constructed using sensors to take in information about the world along with headphones to provide 3D-rendered auditory information back to the listener. This research confirms that such devices can be further developed using software libraries that are used mostly in gaming and virtual applications currently. The key concept is rendering spatialized audio within the read world for a navigational task. This research therefore asks to provide a solution to the research question: can a device be developed within a real environment, to improve a human's sense of direction by using spatialized sounds?**

***Index Terms*— Augmented reality, audio interfaces, headphones, navigation, wearable computers, wearable sensors, magnetic sensors, global navigation satellite system, global positioning system, body sensor networks, sensory aids.**

I. Introduction

Binaural technologies present ways in which humans can augment their sense of hearing through advances in modern electronics. A significant potential use for this augmentation is in the area of navigational technologies. This audio is represented in a virtual space by exploiting the binaural techniques described by [1]–[3] and in Section II. Augmentation of the human hearing experience for navigational circumstances could be of benefit to the visually impaired, hikers, travellers, pilots or any person navigating.

Currently, applications of spatial audio navigation in the real world are limited, as concluded from the research literature studied. Although there is a lack of real-world spatial audio navigation devices, there is a range of technologies and applications described in the literature. For example, [1] explores the future perspectives of binaural technologies in **electronic travel aids** (ETAs). The research presented in [4] and [5] focuses on implementing a purely virtual experience, while [6] describes a device that uses haptic feedback to augment people's sense of touch through a vibrating belt. [7] describes a wearable mobile device using global positioning and magnetic field sensing to place the user in the world, but spatial audio rendering is not used. Thus there is a gap for the work described here. The hardware and software components used to achieve this task are described in Section III of this paper. Analyzing the device's output shows that not only is the audio realistic in nature, but this device can be further developed to expand its capabilities in different navigational circumstances.

II. Existing Techniques for Binaural Sensory Augmentation

In this section the audiological fundamentals of binaural hearing are described, and existing attempts to use this binaural theory to augment a human's sense of hearing are presented.

A. Binaural Hearing and Head Related Transfer Functions

As mentioned in Appendix A, the position of a sound source relative to the human head can be described by an azimuth angle (θ), an elevation angle (ϕ) and a distance (r). A sound source can invoke an **interaural time difference** (ITD) and an **interaural intensity difference** (IID) at the ears which contributes to the localization of the sound source by the human brain [8]. The difference between the ITD and the **interaural phase difference** (IPD) for frequencies below 1500 Hz is described in (1), where the IPD is the phase difference between the sound waves reaching the ear, determined by the ITD and the period of the sound source tone T_{source} [8]. For frequencies above 1500 Hz, the IID plays a greater role in locating a sound source than the ITD or IPD since an acoustic shadow forms as the frequency of the sound increases [8]. Between 500 Hz and 1500 Hz this shadow is not as effective for the average human and the ITD and IPD are still the most prominent localization factors [8]. An additional essential factor that assists in localizing a sound source are Doppler shifts relative to the listener. This Doppler shift is only noticeable at high velocities between the sound source(s) and the listener(s).

$$IPD = \frac{ITD}{T_{source}} \times 360^{\circ} \tag{1}$$

Since the theory models these equations assuming a spherical human head, the reality of these ITD, IID and resulting IPD values differ based on the human individual. The Head Related Transfer Function (HRTF) is a common way to model how the head filters the sound source. This transfer function is the inverse Fourier transform of the Head Related Impulse Response (HRIR). The HRTF values are therefore obtained through modelling the individual's impulse response by placing microphones in each ear of a dummy head or a human head. A software library is described in Section II B and contains HRTF values as well

as the convolution operations that spatialize the audio in both left and right channels. These HRTF values are from a dummy head and originate from the KEMAR data set. (The acronym KEMAR is derived from a name of one of the researchers who was involved in designing the dummy head for the gathering of HRTF values.)

TABLE I
SYMBOL MEANINGS FOR BINAURAL HEARING AND HRTFS

Symbol	Meaning
θ	Azimuth angle to the sound source from the vertical plane.
ϕ	Elevation or altitude angle to the sound source from the horizontal plane.
r	Distance to the sound source from the origin i.e. the center of the head.
T_{source}	Period of the sound source.
a	Radius of the head.
c	Speed of sound in ambient room temperature (20℃).
ω	Frequency of the sound source.
P	Sound pressure at the ear.
P_0	Free-field sound pressure at the origin.

$$HRTF(r,\omega,\theta,\phi,a) = \frac{P(r,\omega,\theta,\phi,a)}{P_0(r,\omega))} \quad (2)$$

B. *Existing Research and Implementations for Binaural Cues in Electronic Travel Aids*

There is a lack of research literature in creating binaural devices that are designed to be worn regularly while improving navigation. A previous implementation in [7] demonstrates the development of an ETA for the visually impaired with a combination of sonification techniques and text-to-speech functionality. However, spatialized binaural cues are not developed in [7] since the sonification techniques are merely a confirmation of reaching a given beacon/waypoint through speech synthesis. Contrastingly, a thorough analysis of the ETA is tested with human volunteers in [7]. An extensive effort to design a database consisting of **points of interest** (POI) is also made in [7] on a handheld mobile device.

In [1] the author comes to a similar conclusion that spatial audio delivery is not included in the most popular ETAs and that future perspectives in the area exist. In comparison to that opinion, a noticeable implementation of spatialized binaural cues for navigation is conducted in a virtual environment [4]. This research shows that different sound source types affect the efficiency of a navigator's path. For example, [4] shows that humans exhibit a more accurate localization of the sound source (or beacon) when the sound is similar to engine noise (i.e. has a wide bandwidth) compared to a worse localization of the sound source when the beacon emits a sound similar to a pure tone or a sonar (i.e. has a narrow bandwidth). This research does not provide the necessary steps for developing an ETA in a real environment, instead [4] provides the necessary comparison of beacon sounds that ensure efficient navigation to waypoints using binaural cues.

In summary, the combination of existing binaural methods and theories allows the creation of a device which uses binaural cues to assist human navigation. A feedback of sensory information gives directional information to the user which is similar to [6], but differs in that this research endeavours to use auditory cues rather than tactile cues. A key concern is whether the output sound is accurate compared to the input of a human's head pose. A head pose includes the orientation and position/location of the head in the world.

III. TECHNICAL DESCRIPTION

A. *Device Hardware*

The binaural device prototype being developed in this research is described fully in Appendix D and is illustrated in Figure 1. Briefly, the prototype consists of an **embedded Linux device** (ELD) with a Raspbian **operating system** (OS) installed. This ELD, in combination with the OS, provides the necessary hardware-software abstractions to interface-with and power the two sensors and the headset. The sensors used to calculate the head pose within the world are a **magnetometer** and **global navigation satellite system** (GNSS) sensor (these are two typical sensors that are found in many handheld mobile devices). The sensors are mounted within a 3D printed sensor mount which was designed in a **computer aided design** software (CAD) and printed with **thermoplastic polyurethane** (TPU). This material was chosen for its flexibility and subsequent wearability. The sensors are powered by the ELD's **general purpose input-output** (GPIO) pins. Subsequently, the ELD is powered by a 22.4 Ampere hour battery and the headset is wired to the 3.5 millimeter auxiliary input jack of the ELD. The typical active power consumption of the setup is between 500 milli Amperes (mA) and 600mA. Therefore the system runs for approximately 37.33 hours assuming maximum current consumption on a fully charged battery.

The **inter-integrated circuit** protocol (I²C) is used to communicate with the sensors at relatively low speeds for receiving head pose data. Those sensors are placed in the 3D printed holder and attached to the headset. It is important to

note that the sensors are at fixed positions relative to the headset to ensure the *hard-iron distortion* is mitigated from the effects of the headphones.

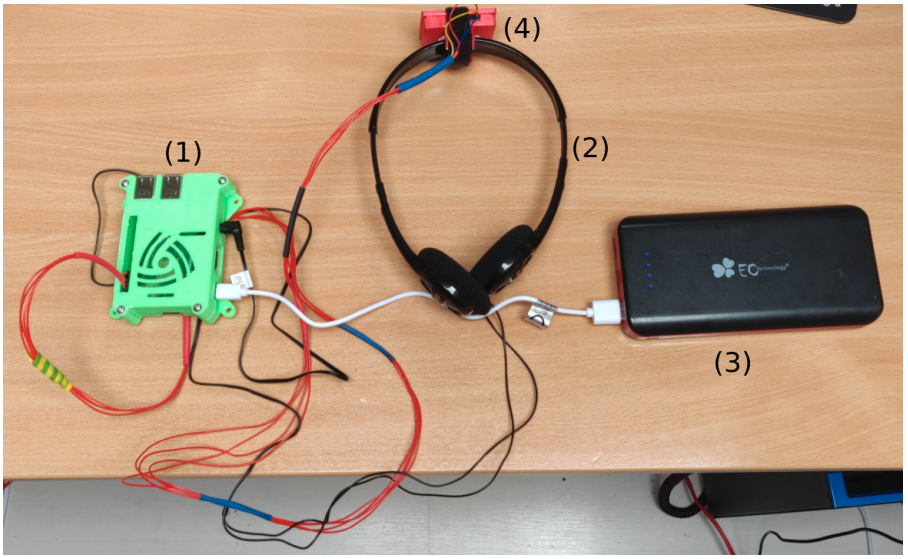

Figure 1. System setup consisting of a Raspberry Pi Model 3 B+ (1), a KOSS KPH7 headset (2), an EC Technology® 22400mAh Powerbank (3), an LSM303AGR (magnetometer and accelerometer) and a PA1010D (GPS) on a custom 3D printed sensor mount (4).

B. Device Software

C++ is the language of choice for the custom software implementation of the binaural device. This language is widely known for its low-level bit manipulation capabilities and is therefore ideal for interacting with the chosen sensor technologies. Custom software is developed to interface to the sensors through the ELD which incorporates bit masking, bit shifting and other techniques to obtain the appropriate sensor values. The raw magnetic values of the magnetometer are read into the input buffer and multiplied by the sensor's sensitivity value to obtain the real magnetic field values measured in Gauss. The hard-iron distortion is mitigated by subtracting the average values of the extrema in each axis (*x, y* and *z*) from the current measured value. (2) shows an example of this assigned update on the *x* axis, and would be an identical update in the *y* and *z* axes.

$$x_{new} := x_{old} - \frac{x_{max} + x_{min}}{2} \tag{2}$$

$$\theta = \arctan\left(\frac{y}{x}\right) \times \frac{180}{\pi} \tag{3}$$

Using the hard-iron calibrated values, the azimuth θ angle is calculated with respect to magnetic north[1]. This is achieved by computing the arctangent of the ratio between the two coordinate vectors which outputs an angle in radians with respect to the axis belonging to the denominator vector. (3) shows how the azimuth is calculated with the arctangent. That angle in radians is converted to degrees in the interval [-180°, 180°]. 0° azimuth is defined as magnetic north although true north is possible to implement since the PA1010D provides magnetic variation information associated with the world location using the RMC sentence format (consisting of a latitude, a longitude and an altitude). Assigning 0° azimuth to magnetic north acts as a simple compass, meaning that the sound position is placed statically at magnetic north as a fundamental navigational reference for the listener on the horizontal plane.

The audio is rendered using an open source software library called **OpenAL** (Open Audio Library) [9]. Recall that this library enables rapid rendering of 3D audio using inbuilt HRTF values and convolution functions. Other effects that influence the localization of sound such as doppler shift can also be manipulated using OpenAL in the case of fast moving objects. It is important to note that a rotation of 90° about the *x* axis is needed between the LSM303AGR magnetometer sensor coordinate system and the OpenAL 3D audio rendering coordinate system. This is because the sensor is mounted parallel to the ground on the *xy* plane whereas OpenAL considers the *xz* plane to be parallel to the ground. The listener orientation in yaw and pitch is converted from a spherical coordinate system to a cartesian coordinate system for the audio rendering process. This is implemented by

[1]The magnetometer only provides a compass heading (i.e. yaw).

calculating the sine and cosine of the azimuth which will output a value on the unit circle of x (4) and z (5).

$$x_{AL} := \sin(\theta \times \tfrac{\pi}{180}) \tag{4}$$

$$z_{AL} := \cos(\theta \times \tfrac{\pi}{180}) \tag{5}$$

To incorporate the distance between the listener and sound source, the position of each object is required. The GNSS module outputs a character byte-string in differing sentence formats. The **Satellites in View** (GSV) sentence format is used for acquiring initial position, and the **Recommend Minimum Navigation Information** (RMC) sentence format is used in tracking or navigating. This is because GSV contains information about the received signal (such as the number of satellites in view and the signal-to-noise ratio for each satellite signal) while RMC contains positional information such as latitude and longitude, magnetic variation and speed over ground. This information is parsed to extract the position of the listener. The sound source is resultingly defined within the listeners frame of reference but may be defined within the world frame when the system is using waypoint coordinates.

The main algorithm pseudocode for initial calibration and repeated spatial audio rendering (based on the sensor readings) is shown in Table II. This calibration stage is necessary to calibrate the hard-iron distortion before any sound is played back to the listener. Hence, the sound source appears more stable and more localized or stationary in space.

TABLE II
MAIN ALGORITHM PSEUDOCODE

Line #	
1	Set listener and source poses
2	Open sound playback device context
3	Calibrate for 30 iterations: Read the sensor values Adjust hard-iron distortion
4	Load the beacon sound to the sources buffer
5	Play the beacon sound
6	Forever: Read sensor values Update the listener pose Render the sound

The sound source selected as a beacon sound is a non-copyright sound of birds and machinery in an outdoor setting. As mentioned previously in Section II B, noisey beacons are one of the most effective beacon sounds for path traversal compared to pure tone and sonar-like sounds as proven by [4]. Care is also taken to choose the beacon sound so as to not distract the listener with repetitive noise.

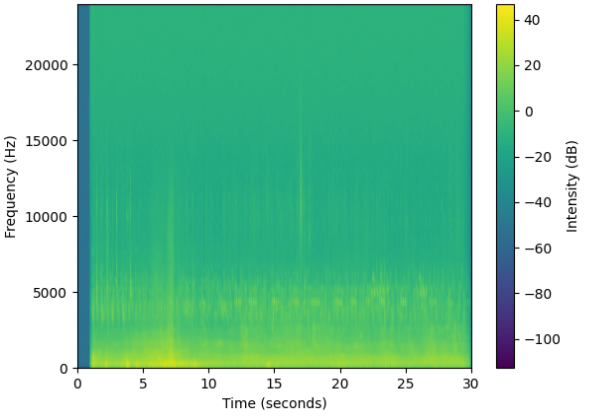

Figure 2. Spectrogram of the beacon sounds showing constant engine sounds below 2.5KHz and intermittent bird chirping sound around approximately 5KHz and above.

IV. DEVICE RESULTS

The results obtained from testing the binaural device are described in this section. Firstly, the offset cancellation results are shown in Figure 3 where a comparison of the magnetic values with and without offset mitigation is shown.

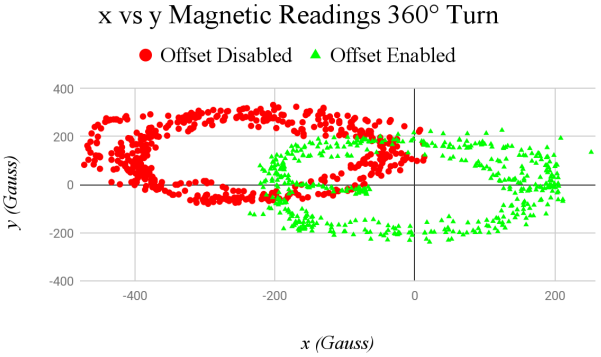

Figure 3. x versus y magnetic axes with and without offset. These values are obtained by rotating 360° yaw.

This hard-iron distortion mitigation contributes to the rendering of the audio. An object is placed 1 meter north relative to the listener. As the magnetometer mounted on the headphones rotates to the right from 0° to -180° the output audio is rendered in space and plotted over time as seen in Figure 4. A close-up view of the impulse response of the headphone output is shown in Figure 5 to inspect whether the ITD is present after the convolution of the HRTF values with the beacon sound.

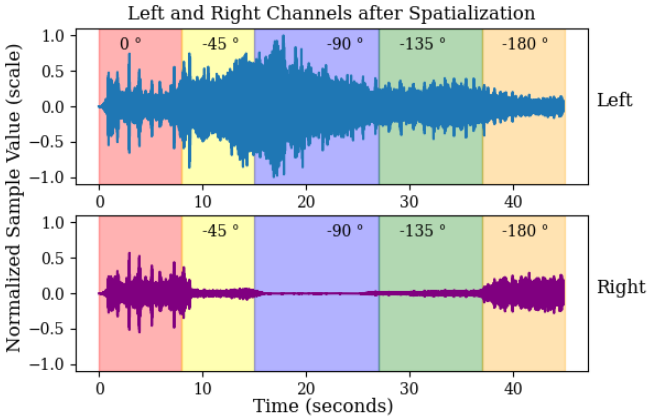

Figure 4. Stereo sound wave amplitudes over time rendered during a 180° rotation to the right.

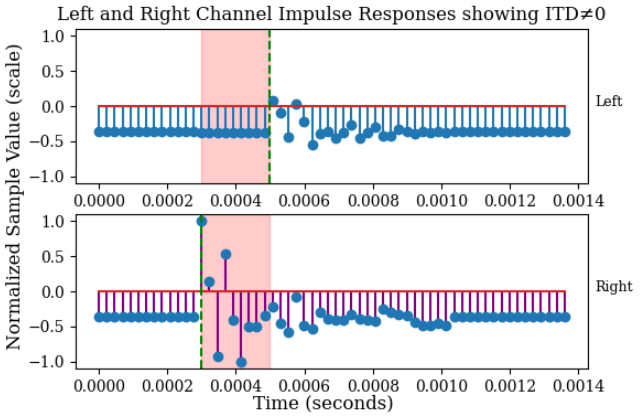

Figure 5. Impulse response of the headphone output with a 20° rotation to the left. The ITD exists between the two peaks indicated by dashed green lines.

These resulting figures exhibit the expected behaviour of the device as the listener rotates their head around the horizontal plane. Functions are debugged to ensure that the angles of rotation stay within the bounds specified. The calculation of coordinates for the listener pose are also checked with a debugger, especially at boundary conditions such as at 0° and ±180°.

V. Device Analysis

Figure 3 verifies the efficacy of the hard-iron offset algorithm since the distortion is mitigated by centering the magnetic values around the origin. The hard-iron offset value can be put directly into the hard-iron offset register of the magnetometer, although this is not necessary since the value would need to be adjusted every time a new pair of headphones is used or the sensors are moved slightly. Therefore the dynamic adjustment of the hard-iron offset is key for prototyping where factors such as headset type or sensor position are changeable. It is important to note that the magnetic field of the environment has an effect on the listener orientation due to the magnetometer; therefore this particular device is not suitable for circumstances where strong and moving magnetic forces are at play relative to the magnetometer.

In Figure 4 a changing IID is demonstrated. For example, the transition between 0° and -45° azimuth shows an appropriate decrease in amplitude in the right ear. The peaks within the 0° segment in Figure 4 are not present within the -180° segment because those peaks are the chirping sounds of the birds which do not play after approximately 37 seconds to the end of the plot. Within the -90° segment, the peak represents a perfect -90° orientation and is then slowly rotated to -135° after the peak. This transition is plotted to demonstrate that the device renders the audio amplitude in a continuous manner (i.e. has a high angular resolution) compared to the discrete intervals shown in Figure 4.

The ITD is shown to be non-zero in Figure 5 marked between the two dashed lines. This ITD equates to approximately 0.2 milliseconds and contributes to the spatialization of the rendered audio.

VI. Conclusions

From the materials presented in this paper, it is certain that the technology exists to develop an ETA with spatialized auditory cues. Although the GNSS information and proper roll and pitch orientations could not be integrated into the device, the yaw orientation on the horizontal plane provided insight into capabilities of such a device acting as an *audio compass* of sorts. This audio compass is similar to the artefact in [6] but different in that the research in [6] used vibrating motors. Incorporating the accelerometer in future research would include pitch and roll values. The coordinate systems used in this paper are simply defining the azimuth of the beacon from north and the distance to that object. However, this coordinate system assumes the human head is stationary within its own frame of reference. If research was to be continued in the area then a global coordinate system would have to be defined which standardizes the pose of objects in space, such as the coordinate system built into the OpenAL framework.

The research presented here shows that the position and orientation metrics of both listener and sound sources may be corrected to further develop an ETA with spatialized auditory cues. One could imagine sensors, without the large breakout boards, integrated within the headset. One could also imagine the ELD substituting for an *already* widely used mobile device. As a result and to conclude, the insights that have been documented here from the development of a binaural sensory augmentation device informs the next generation of binaural technologies in navigational settings.

References

[1] S. Spagnol *et al.*, ‘Current Use and Future Perspectives of Spatial Audio Technologies in Electronic Travel Aids’, *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–17, 2018, doi: 10.1155/2018/3918284.

[2] G. J. Brown and D. Wang, ‘Binaural Sound Localization’, in *Computational Auditory Scene Analysis.*, John Wiley & Sons, Inc., 2005. [Online]. Available: https://www.cs.cmu.edu/~robust/Papers/SternWangBrownChapter.pdf

[3] M. Gröhn and CSC Tieteellinen laskenta, ‘Application of spatial sound reproduction in virtual environments: experiments in localization, navigation and orientation’, CSC - Scientific Computing, Espoo, 2006.

[4] B. N. Walker, ‘Effect of beacon sounds on navigation performance in a virtual reality environment’, presented at the 2003 International Conference on Auditory Display, Atlanta, GA 30332-0170, Jul. 2003. Accessed: Jan. 23, 2021. [Online]. Available: https://www.academia.edu/153811/Effect_of_b_beacon_b_sounds_on_navigation_performance_in_a_virtual_reality_environment

[5] S. Xiaowei, Z. Shutao, H. Junwei, and H. Minghui, ‘A 3D Sound Simulation System for Flight Simulator’, in *2010 Second International Conference on Computer Modeling and Simulation*, Jan. 2010, vol. 1, pp. 44–48. doi: 10.1109/ICCMS.2010.127.

[6] S. Nagel, C. Carl, T. Kringe, R. Märtin, and P. König, ‘Beyond sensory substitution – Learning the sixth sense’, *J. Neural Eng.*, vol. 2, pp. R13-26, Jan. 2006, doi: 10.1088/1741-2560/2/4/R02.

[7] P. Skulimowski, P. Korbel, and P. Wawrzyniak, ‘POI Explorer – A Sonified Mobile Application Aiding the Visually Impaired in Urban Navigation’, Sep. 2014, pp. 969–976. doi: 10.15439/2014F293.

[8] Fred H. Bess and Larry Humes, ‘Audiology - The Fundamentals’, 4th ed., Philadelphia, PA: Lippincott Williams & Wilkins, 2008, pp. 95–97.

[9] kcat, *OpenAL soft*. 2021. Accessed: Aug. 16, 2021. [Online]. Available: https://github.com/kcat/openal-soft