Accessibility System for Capturing Distance Data and Converting them into Three-Dimensional Sound

Michael Barney[1], Jonathan Kilner[2], Aida Araújo[3], Gilmar Brito[3], Ioná Rameh[3], Ronaldo Aquino[4].
Federal Institute of Education, Science and Technology of Pernambuco (IFPE)

Recife, Brazil

[1]michaelbarneyjunior@gmail.com [2]jkss@cin.ufpe.br

[3]{gilmarbrito, aidaferreira, ionarameh}@recife.ifpe.edu.br [4]rrba@ufpe.br

ABSTRACT
This paper presents a developed system that captures distances from an array of sensors and converts them into 3D sound information that can be heard and positioned by its users. The generated binaural audio can be used for detecting and localizing physical objects and obstacles by utilizing the real time data inflow given by the sensors. It uses an Input and Output module with algorithms that receive the distances and generate their respective 3D sounds. Furthermore, this system was applied in an Accessibility device that was tested by visually impaired volunteers, allowing the verification of the system's functionality in real-world scenarios.

CCS Concepts
• **Accessibility Technologies**

• **Sound-based input/output**

Keywords
3D Sound; Visually Disabled; Sensory Glasses; Distance Conversion; Binaural Synthesis

1. INTRODUCTION
Three-dimensional sound has the capacity of audio source positioning in relation to a point, allowing a listener to describe with high accuracy the localization of any given audio source [1]. This effect is generated through various sound components, including: volume, frequency, and the delay that the audio waves have between the left and the right ears. This is shown in Figure 1.

This type of audio is usually recorded using specialized techniques, using two microphones placed inside of a head dummy with detailed ears. The captured sound-waves are in turn transformed into Head-Related Transfer Functions (HRTF), where the sound's data (such as volume and pitch) is stored. [2].

Conference'10, Month 1–2, 2010, City, State, Country.
W4A 2018, April 23-25, 2018, Lyon, France

DOI: http://dx.doi.org/10.1145/12345.67890

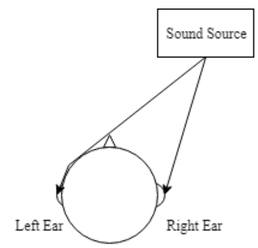

Figure 1 - Representation of Audio Source Positioning

The conversion of these functions into three-dimensional sound is called binaural synthesis. In this process the HRTF are conversed into audio that can be heard using simple devices such as common ear buds. [3]

Recorded binaural synthesis is used in various fields, from entertainment in movies and games until military purposes in battlefield simulations. [4]

Some applications for 3D sound, however, require the immediate generation of the audio source and not a recorded sound. Generated Binaural Synthesis uses input data (such as distances, color and object recognition) to output a representation using three-dimensional sound. [5]

The development of a system that accepts distance input to generate 3D audio can be applied in real time ambient generation, which can be used in many areas. This study's proposal is an approach to capture distances from an array of sensors and convert this data into 3D sound information which can be heard by users. The paper is organized as follows: In section 2.1 we provide an overview of the system for capturing distances and converting them into three-dimensional sound. Section 2.2 describes the Input Module, which is responsible for capturing the sensor's information. Section 2.3 details the Output Module, tasked with utilizing the distance data to generate and output binaural 3D audio. Section 3 describes the application of the developed system as an Accessibility device, with the aim to facilitate object detection by visually impaired users, along with its subsequent tests. Finally, section 4 presents the conclusions of the research and projects future works.

2. SYTSTEM DEVELOPMENT

2.1. OUTLINE

The developed system requires two modules: one for sensing 3D distances in real time and one for generating and outputting 3D binaural sound. This overview is represented in the following figure, which shows the Input Module measuring distances through a sensor and communicating with the Output Module which emits generated 3D sound:

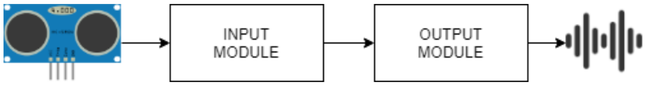

Figure 2- System Outline

2.2. Input Module

A circuit is necessary to receive distance information and send them to the Output Module. Devices such as ultrasonic sensors can be used to detect these distances. They return it's distance to an object by emitting a sound wave (sending a signal to the Trigger pin) and calculating the time that it takes for it to echo back (received through the Echo pin), as seen in Figure 3. Ultrasonic sensors are usually used when the distance measured is from 10cm to 10 meters.

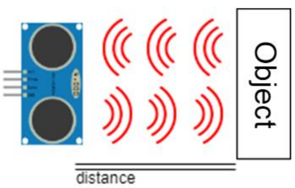

Figure 3- Ultrasonic sensor functioning representation, as the sound waves are emitted from the trigger, echoed by the object and received by the sensor.

To receive the 3D distance, a sensor array should be connected to a microcontroller and placed in a circle manner, covering the desired radius (Figure 4). The ratio of the number of sensors per radius is one of the values that determine the precision of the Input Module, in conjunction with the hardware characteristics of the sensor (such as maximum distance and operation angle).

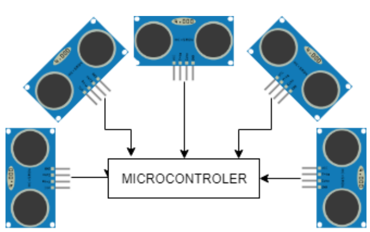

Figure 4- Sensor Connection Representation

The algorithm on the microcontroller calculates the measured distances of each sensor of the array, fully expansible though programmatically setting the trigger and echo pins. Applying individual moving average filters, it corrects eventual mismatches on distance readings. The stored distances will be used by the following Output Module. The following outline summarizes the pseudo-code of the microcontroller:

```
Data: numSensors
Result: distances[numSensors]
for i ← 1 to numSensors do
    trigPin ← numSensors * 2;
    eschoPin ← numSensors * 2 + 1;
    % Distance ;
    write(trigPin, LOW);
    delayMicroseconds(2);
    write(trigPin, HIGH);
    delayMicroseconds(10);
    write(trigPin, LOW);
    duration ← pulseIn(echoPin, HIGH);
    distance ← duration * 0.034/2;
    distances[i] ← movingAverage(i, distance);
end
```

Algorithm1-Microcontroller Pseudo-Code

2.2. OUTPUT MODULE

The Output Module uses the stored distances to generate and output three-dimensional sound. This module can be based upon the same microcontroller as the Input Module or it can be deployed on another device (given that the two modules need to communicate between each other).

Many attempts of generating 3D Sound were made during the development of the system, going from the use of pre-recorded samples until the wave-manipulation of the final result.

The development started by using the Amphiotik Synthesis computer software [6] (which uses audio samples to apply generated binaural synthesis) to test how software controlled audio positioning of an audio sample compares to recorded binaural synthesis.

With positive results, the next step was to implement a system that connected to the Input Module and reproduced recordings from the Amphiotik Synthesis. With 5 sensors and 5 audio samples, the volumes of each one were inversely proportional to the distances of each correspondent sensor (set at a maximum of 4 meters). In this manner, if the left sensor detected a distance of 2 meters, the audio sample that represents an object from the left would play at 50% of its volume.

The second prototype made the switch between using audio samples and started generating the 3D sound, allowing the ease of expansion and control of the system. Focusing only on volume control, the developed software generated an audio wave and reproduced it through left and right audio channels with volumes dictated by the following algorithm:

```
Data: numSensors, currentSensor, distances[numSensors],
      maxDistance
Result: rightVolume, leftVolume
step ← 100/(numSensors − 1);
proportion ← 1 − distances[currentSensor]/maxDistance;
rightVolume ← proportion * step * currentSensor;
leftVolume ← proportion * (100 − step * currentSensor);
```

Algorithm 2- First audio generation Pseudo-Code

This prototype showed a linear variation of the volume from 0 to 1 as it received a distance from 0 to a maximum distance (having already applied the logarithmic correction of audio intensity relative to human hearing).

The Output Module spends a programmable number of milliseconds representing the distance of each sensor, giving the user the sensation of an audio-source going around his head, jumping from one position to another. As can be seen in the following graph:

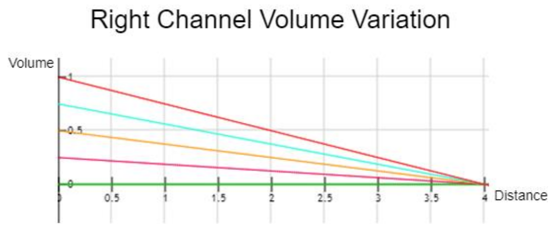

Figure 5- Right channel of system with 5 sensors and a maximum distance of 4 meters.

On the next prototype, the team expanded this concept by adding wave frequency manipulation according to the calculated volume. This means that the pitch of the audio was lower the farther an object was relative to the ultrasonic sensor (in conjunction to de volume control of the previous prototype).

Multiple sensor simulation was also added, giving the sensation that a larger number of sensors are being used, but with the same accuracy. This removed the sensation of the audio source jumping from one place to another and replaced it with a gradual movement of the source.

The final research added to the generation algorithm by changing the linear progression of volume, relative to the distances, into a rational one, as seen in Figure 6. This modification was made due to the perception that any change of proximity in closer distances was shown to need of larger feedback to the user.

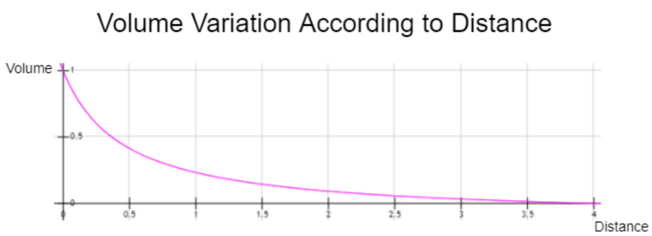

Figure 6- Audio variation according to detected distances.

The equations for the complete sound generation system are as follow:

(1) $volumeProportion = \frac{variation}{distance - \max + \sqrt{c}} + 1 + \frac{variation \times 2}{\max - \sqrt{c}}$

With:

$$c = max^2 + 4 \times var \times max$$

Where:

- "variation" is a constant number between 0 and 1 that determines how fast the function grows.
- "max" is the constant maximum distance that device will measure
- "distance" is the variable measured distance

(2) $rightChannel = \frac{sensor}{numSensors-1} \times volumeProportion$

(3) $leftChannel = \frac{(numSensors - 1 - sensor)}{numSensors-1} \times volumeProportion$

Where:

- "numSensors" is the constant number that represents the number of sensors that the system is measuring.
- "sensor" is the variable number that represents the current sensor that's being converted, starting from 0 as the leftmost and ending in numSensors – 1 for the rightmost.

(4) $frequency = volumeProportion \times 1.8$

3. APPLICATION

3.1 DEVICE OVERVIEW

A following research was conducted to apply this system, allowing the verification of the effectiveness of the audio generation system in a real world environment. The developed device was a pair of Sensory Glasses for the Visually Impaired [7]. It uses an Arduino and five ultrasonic sensors as the Input Module and an Android device with a custom app as the Output Module, connecting to the microcontroller through USB-OTG or Bluetooth (Figure 7).

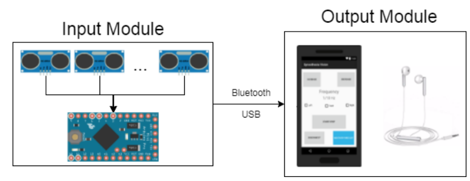

Figure 7- System modules on the developed device

The Microcontroller sweeps each sensor and immediately communicates with the output module the raw data. In turn, the App applies all of the described algorithms to generate 3D Binaural Sound that gives the user positioned data for each sensor, jumping from one to the other after a specified period of time.

A Smartphone was chosen as the core of the Output Module due to its fast prototyping and natural interface for the visually impaired users, due to the current accessibility ecosystem available in their most common operating systems.

With the "Internet of Things" capabilities of the Smartphone available, features were added to the system in order to supply additional ambient information, such as the current weather prediction being read by the app when a physical button is pressed.

Figure 8- Functioning Prototype

The glasses itself is very simplistic, containing only the ultrasonic sensors, a small Microcontroller and a Bluetooth module. The powering of the device is done though the USB connected to the Smartphone, which can also be used for communication, replacing the Bluetooth. The final prototype (Figure 8) was 3D printed to fit all of the sensors in a more discrete and lightweight manner, approximating the look of standard sunglasses.

3.2 TESTS AND EVALUATIONS

In total, the system was tested with 4 visually-disabled volunteers that agreed to fully use the device, with one of them fulfilling multiple tests during the entire development of the project, as seen in Figure 9. Visually disables volunteer were prioritized, since they are more sensitive to binaural cues [8].

Figure 9- Volunteer Testing First Prototype

Tests were conducted by first explaining to the users the simplified functioning of the device, covering 3D Sound and echolocation. Next, a simple test is done with the user sitting down and objects are placed near individual sensors with varying distances as the volunteer describes what he is hearing and how he processes the audio as positioned distances. In the final step, the volunteer traverses a simple "obstacle course" while using the glasses and his standard white-cane, sharing his opinions afterward.

Initial data detected that in 30 minutes the guests were able to understand the functioning of the device and associate the approximation of obstacles through each sensor, individually and standing still, on the first and second part of the test.

Subsequent information verified that the system usually needs a 2 hour first contact experience on the first training session, for optimal results on the entire test. This is due to the need of association between audio and positioned distances combined with the necessity to keep constant attention on 5 information streams, one for each sensor.

As predicted, further away obstacles were shown more difficult to detect, due to the rational function of the volume proportion, while objects in the 2 meter range where easily detected with accurate positioning, as shown in the following figure:

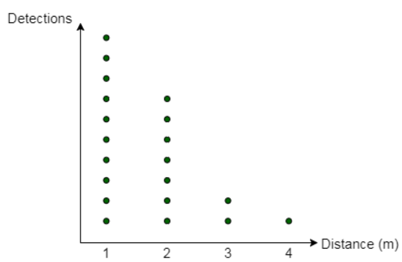

Figure 10- Detections per Distance

Opinions from the volunteers indicated that the device was great for detecting objects that are large or located in the upper torso area and gives more security when combined with the white cane. However, the continuous stream of information from 5 sensors proved difficult to maintain analyzed when multitasked with activities such as walking.

4. RESULTS AND CONCLUSIONS

To solve the problems encountered by the volunteers, the team developed a "cane" mode for the device. In this functioning, the system only presents to the user the most significant data and variations, normally the one with the shortest distance. In this manner, the user only has to analyze one continuous information stream, greatly simplifying the device usage without losing the Binaural Sound advantages. In this mode, data is sent only when important variations are detected, lowering the amount of data to be processed by the user at any given time. This new functioning mode also facilitated the assimilation between sound and distances, due to its continuous nature.

While multitasking was initially proven difficult when using the device, due to the competing audio signals in other tasks, the volunteers helped to determine the least intrusive generated audio for the device, preferring lower volumes and deeper frequencies.

Following research shall focus on expanding the tests to a larger number of subjects, with the aim to fine tune the algorithms. New connections from the device to other systems should also be developed, furthering the IoT capabilities of the system.

5. ACKNOWLEDGEMENTS

The Federal Institute of Science and Technology of Pernambuco for financing the Researchers and granting the laboratory space for the project development.

6. REFERENCES

[1] W. G. Gardner, "3D audio and acoustic environment modeling." Wave Arts Inc, 1999, 109 pp.

[2] Dumidu S. Talagala, Wen Zhang. 2014. “Binaural sound source localization using the frequency diversity of the head-related transfer function”. The Journal of the Acoustic Society of America 2014, 135:3, 1207-1217. DOI: https://doi.org/10.1121/1.4864304

[3] Cobos, M., Lopez, J. & Spors, S. 2010. “A Sparsity-Based Approach to 3D Binaural Sound Synthesis Using Time-Frequency Array Processing” EURASIP J. Adv. Signal Process. (2010): 415840. DOI: https://doi.org/10.1155/2010/415840

[4] Arrabito, G.R. 2000. “AN EVALUATION OF THREE-DIMENSIONAL AUDIO DISPLAYS FOR USE IN MILITARY ENVIRONMENT“;Canadian Acoustics, vol 28, no 4, 2000, p 5-14

[5] Henrik MÃ¸ller, Fundamentals of binaural technology, Applied Acoustics, Volume 36, Issues 3â€“4, 1992, Pages 171-218, ISSN 0003-682X,. DOI: https://doi.org/10.1016/0003-682X(92)90046-U.

[6] Tsakostas, C. “Binaural Simulation applied to standard stereo audio signals aiming to the enhancement of the listening experience”. Book of Proceedings, HELINA (2004)

[7] Michael Barney, Jonathan Kilner. 2017. Sensory Glasses for the Visually Impaired. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work* (W4A '17). ACM, New York, NY, USA, Article 27, 2 pages. DOI: http://doi.org/10.1145/3058555.3058584

[8] M. E. Nilsson and B. N. Schenkman, “Blind people are more sensitive than sighted people to binaural sound-location cues, particularly inter-aural level differences”, Hearing Research, Vol. 332, February 2016, pp 223–232.