Head Related Transfer Functions (HRTFs)

By Joey Cirone, ECE '17

Introduction

A key attribution to the way humans hear is the ability to detect where a sound is coming from, often referred to as localization. There are many methods to achieve localization, but Head Related Transfer Functions (HRTFs) allow a more accurate and exact effect. Team Purple's project uses HRTFs to add an immersive and interactive listening experience to music.

Background

Virtual Reality

Immersive audio is used by Virtual Reality (VR) gaming companies like Oculus Rift, HTC VIVE, and Google Cardboard. They create a VR where a user can be "transported" into a fictional world and see it right before their eyes. However, visuals account for only one aspect of a true VR experience. Without the correct audio cues that move in three-dimensional space with the visuals, the VR illusion is broken and a true immersive VR experience cannot be realized.

Spatial Audio

With the increased technology in low latency, faster data transfer, head tracking, auditory displays can now update the spatial sound field fast enough to compensate for listener head movements, which is critical for how humans localize sounds (1). However, head tracking is only one aspect of spatialized audio. Problems with capturing the full audio experience exist because of how every person hears. The shape of someone's head can affect how long it takes a sound to come from a source to each ear, called the Interaural Time Difference ITD (2). This time difference effects the phase (the position of a waveform in a point in time) and amplitude (intensity) of the source signal. The brain uses these differences to figure out where the sound came from. Although somebody can distinguish where the sound is coming from with headphones, they only hear it as if the sound is coming from inside their head, and not spatially around them.

Why HRTFs

Spatial audio does not only capture localized sound, but externalizes it to perceive distance, as well as location. Spatial audio is complicated to capture, so the individual HRTFs are used. Unfortunately, conventional methods of personalized HRTF capturing is time consuming and impractical (3). Although there exists surround sound systems such as Dolby 5.1 that create spatialized sound, this is only practical for a physical space, with fixed speakers. Only with personalized HRTFs can you place virtual speakers in a VR environment and hear it back with headphones. With the additional effects of how the sound is filtered based on the size and shape of the listener, spatial audio using HRTFs creates a realistic, spatialized, and immersive auditory experience.

What are HRTFs?

The public database

HRTFs capture sound localization cues created by how sound reflects, diffracts, and is generally filtered by the geometry of a person's head, face, and pinna (external part of the ear) before entering the ear canal (4,5). Even though general geometric models can capture most of the filtering effects for the HRTFs, small variations of the pinna can produce large changes in the HRTFs (4). The Center for Image Processing and Integrated Computing (CIPIC) public database contains many HRTF responses with varied head and ear measurements that can be used to match an individual.

The Captured Data

Specifically, HRTFs consist of three parameters in spherical coordinates; azimuth (measured clockwise) φ, elevation θ, and either time or frequency (6). Using these parameters, you can apply the specific HRTF filter for a sound at the specific azimuth, elevation, and frequency to make the sound appear as it is coming from that location in space. The following figure illustrates these three parameters in a 3D space.

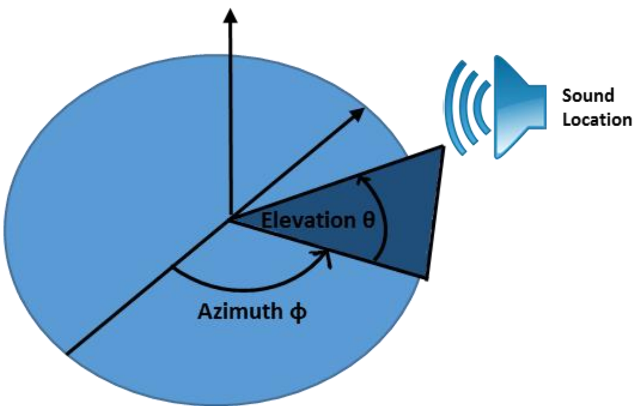

Figure 1: Azimuth, Elevation, and Sound Location Parameters of HRTF

Measuring HRTF's

As stated earlier, personalized HRTFs are a difficult and time consuming measurement to capture. It is measured as a transfer function (relationship between any output and input) from the sound source (input) to the microphones placed inside the ears (output). The individual subject sits in an anechoic chamber (an echo-free room designed to isolate sound reflection from the individual) with microphones in both ears. Speakers rotate around the subject, playing a reference sound at specific angles and recording the received sound from the microphones in the ear, until all angles are captured. These relationships between the reference sound, and the sound being recorded after it reflects around the subject's head and ear at various angles make up the HRTF. The following picture is of an HRTF measurements system at Microsoft Research (7).

Figure 2: https://www.engadget.com/2016/11/02/microsoft-exclusive-hololens-spatial-sound/, 4:54

Generic & Personalized HRTFs

Generic HRTFs are created by averaging existing HFTFs or by measuring the data from a dummy in Generic HRTFs are created by averaging existing HFTFs or by measuring the data from a dummy in place of an individual. However, generic HRTFs may not sound realistic to all individuals because of the variation of pinna geometries. Personalized HRTFs refer to using existing HRTF measurements, found in public databases such as CIPIC, and matching geometric characteristics to a subject with similar dimensions such as the shape of the pinna (8). This method can be practically used by taking pictures of the subject and matching it to a database of HRTFs (9). The use of an existing HRTF database eliminates the need to do the time-consuming measurement on every new individual who would like to experience localization.

Using Spatial Audio

The HRTF process chain

Because of the way HRTFs are measured, the audio must have added room effects such as reverb and early reflections (sounds that arrive to the listener after reflecting once or twice from walls, ceilings, and floors) to make it sound realistic. The output sound from the HRTF would be what it would sound like if sound was playing in an echo-less room (anechoic chamber). You would be able to tell where the location of the sound was, but it would not sound realistic. This is because the HRTF does not capture the effect of the rooms reverb and other characteristics that define how the sound reverberates from the environment in which it was produced. The HRTF is simply applied to a mono sound source (same audio in both left and right ears) as a filter at the specific location (azimuth and elevation) to where you want the sound to come from.

Other Techniques

There are various other techniques to capture similar spatialized audio. The simplest technique that many VR companies use is the amplitude panning method (6). Amplitude levels of the left and right channels of stereo audio (corresponding to left and right earbuds or headphone cups) are varied to suggest a sound source that is localized toward the left or right. However, this approach to spatialization only captures localization cues in a 2D plane and cannot replicate if a sound is in the front or back. Another method is recording sounds using two microphones inside a dummy head. This technique is called binaural audio and is best show by this virtual barbershop video. Another method, called Ambisonics, is a recording technique that captures spatialized sound in a 3D area which can be formatted for headphones, 5.1, and 7.1 surround sound. The problem with this method, is that you would need to have Ambisonics recordings for all sounds which is not practical. This demo from RealSpace 3D Audio uses a combination of all these techniques to create a realistic sound experience.

Use in Project

For our project, HRTFs are used to enhance the spatialization for the music. By adding HRTFs to our audio processing chain, we can make the sound seem as if it is in front of you, referred to as externalization, instead of coming from inside you head. Combined with our head tracking module, we use the HRTF algorithms to make the audio sound as if it is in a 3D space when the user moves their head. In the future, instead of using a generic HRTF algorithm, we can pick a more personalized HRTF based on a simple measurement, such as head width, and match the HRTF from the CIPIC database. The use of more personalized HRTFs in our project is the next step to re-imagining music for an immersive, interactive listening experience.

References

1. Brungart, D. S., Kordik, A. J., & Simpson, B. D. (2006, January). Effects of Headtracker Latency in Virtual Audio Displays

2. Virtual Acoustics and Audio Engineering. Retrieved October 12, 2016, from http://resource.isvr.soton.ac.uk/FDAG/VAP/index.htm

3. Rund F (2012). Alternatives to HRTF Measurement. IEEE

4. Algazi V. R. (2001). The CIPIC HRTF Database. IEFE Workshop on Applications of Signal Processing to Audio and Acoustics 2001.

5. Mokhtari P (2008). Computer Simulation of HRTFs for Personalization of 3D Audio. 2008 Second International Symposium on Universal Communication.

6. Schissler C (2016). Efficient HRTF-based Spatial Audio for Area and Volumetric Sources. IEEE Transactions on Visualization and Computer Graphics. 22(4), 1356-1366.

7. https://www.engadget.com/2016/11/02/microsoft-exclusive-hololens-spatial-sound/

8. Rund F (2012). Alternatives to HRTF Measurement. IEEE

9. Torres-Gallegos E (2015). Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. Applied Acoustics. 97, 84-95.