

Automated HRTF Individualisation Based on Localization Errors in 3D Space

Robin Yonge

August 2017

Background

The Importance of Spatial Audio

Spatial audio has never been more important. Though there has been a steady stream of interest in applications of spatial audio in fields like defence - primarily applied to Virtual Auditory Displays (VADs)(?) - virtual, augmented, and mixed reality form a large component of the current technological zeitgeist. One major stated goal of these technologies is that of immersion. This doesn't have to mean that the user feels as if they have been transported to somewhere completely new, they just have to believe in the virtual elements of the experience. No matter whether the intended application is entertainment, productivity, or assistance (unsure about this line), the end user must be deceived into believing in what they are experiencing.

Audio has parity with visuals here, as just small errors in either can irreparably break immersion(?).

[something about binaural audio here??? Explain meaning etc???]

Reproducing spatial audio convincingly involves a number of factors, including reflections and occlusion caused by the room and the objects in it. This project however, will focus entirely on the effect the anthropometry of the listener has on the audio signal - the attenuation of the sound caused by the various body parts that they sound waves come into contact with.

Implementing Spatial Audio

In most applications involving spatial audio, representing this attenuation is done using Head-Related Transfer Functions, or HRTFs, derived from their time-domain counterparts Head-Related Impulse Responses, or HRIRs. HRIR measurements are taken by placing microphones in the ears of a participant (human or mannequin) and measuring the impulse response resulting when a tone is played from a loudspeaker (?) (?). This measurement process should be repeated for as many positions/points of origin around the participant as possible in order to maximise coverage and provide the greatest amount of information when it comes to using the information to process audio. The number of source positions varies from database to database

[in the case of CIPIC it is every L degrees from N to M, in the case of ARI, it is every Y degree from X to Z, etc] [add diagrams!]

These databases may contain anything from data from a single mannequin in the case of the MIT KEMAR set (?), to the CIPIC database's 45 subjects (?), up to the 110-subjects-and-growing ARI HRTF database (?). This data is then commonly used in implementations of spatial audio - in the simplest sense the audio sample is convolved with the HRIR, producing audio that appears, through headphones, to come from the position in 3D space that the HRIR was originally measured from.

The problem with using this data in spatial audio implementations that are to be used in applications for the consumption of wide ranges of people, is that HRTF data is incredibly specific to the person the measurements have been taken from. Just small differences in the anthropometry of the measured participant and the end user can compromise the efficacy of the HRTF used(?).

However, when one tries to use average HRTF derived from a set of measurements, or uses a mannequin like the KEMAR(?) with very average features, the processed audio becomes too average and works for a very small subset of people - exactly the same as using an HRTF from a single human.

Literature Review

Generalised HRTFs

Models

Clustering

Frequency Scaling

Structural Models

Database Matching

Principal Components Analysis

Understanding PCA

PCA and HRTFs/HRIR

test citation (Hözl, 2012)

Search Methods

Simulated Annealing

Method

Analysis

Discussion

Bibliography

Josef Hözl. *An initial Investigation into HRTF Adaptation using PCA IEM Project Thesis*. PhD thesis, Graz University of Technology, 2012. URL <https://github.com/jhoelzl/HRTF-Individualization>.