Spatial audio simulation with the use of binaural audio and recommender systems

Author: David Bernhauer, **Supervisor**: Tomáš Skopal

Faculty of Information Technology, Czech Technical University

FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Virtual reality devices have been experiencing a boom among the public. Although they mostly specialize only in reproducing visual information. Visual information cannot be interpreted by **blind people** or in situations that do not support the interpretation of graphical information (dark room, phone call, ...). We need to use another sense, such as **hearing**.

The aim of the diploma thesis was to design and implement a model with the use of binaural audio and recommender systems. This model is enriched of analysis of methods of pointing to a sound source. The last part of diploma thesis focuses on testing and interpretation of results.

1 Binaural audio

People use their ears to perceive sound. Besides basic sound information about the source of sound (tone pitch, colour, length, ...) people are able to recognize the position of source or shape of the surroundings. Processing of signal takes place in the middle and inner ear. The **outer ear** works like a **sound filter** and serves to locate the sound. Because of the multiple reflections from the outer ear, there are typical signal modifications. These modifications are **unique** for every person (similarly to fingerprint).

Head-related transfer function (**HRTF**) is a function that represents the modification of source signal and corresponds with outer ears. The basic approach is to measure the HRTF for an individual what is the most precise method. The biggest disadvantage is longer measurement time, difficulty and discomfort. The commercial approach is to use one *pseudo-universal* HRTF for every person what is less precise. The basis of these methods is the usage of headphones.

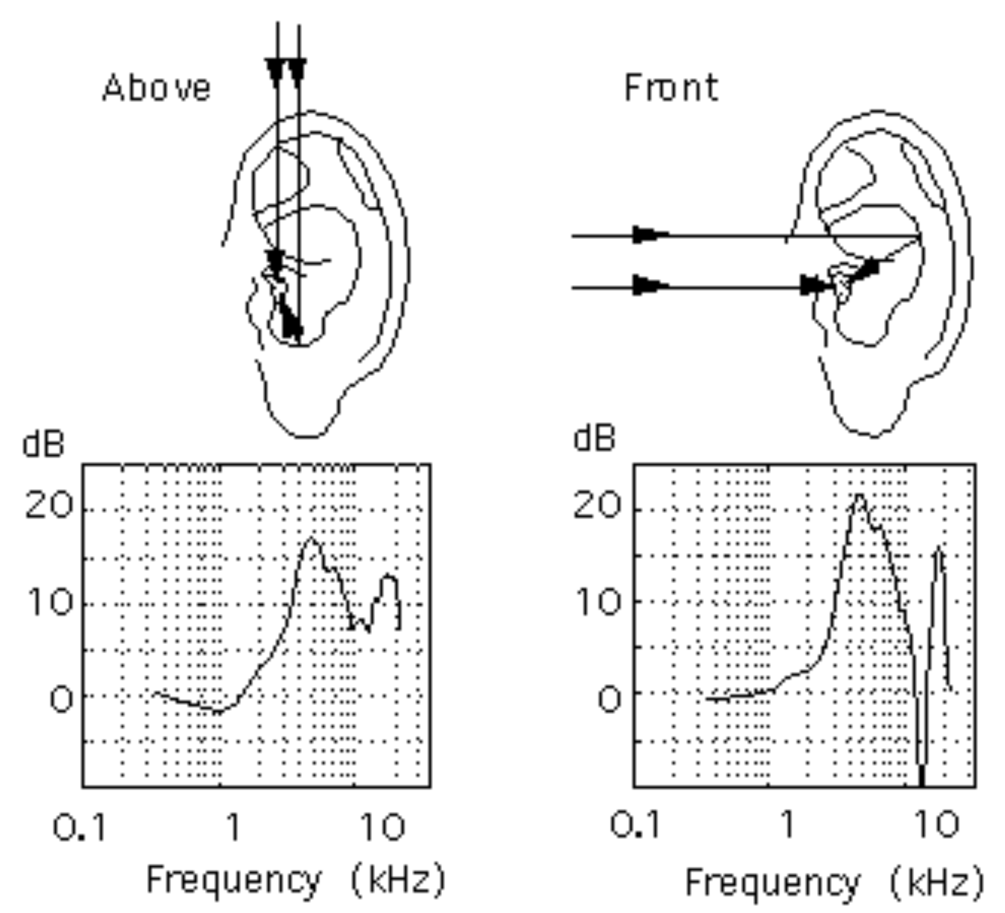

Figure 1: Pinna reflections originally from `http://interface.cipic.ucdavis.edu/sound/tutorial/psych.html`

2 Recommender systems

Many e-shops implement a simple recommender system. Customers **provide information** about their preferences, shopping carts, and ratings. They **receive recommendations** for other interesting and similar products. The similarity between products (or customers) is the core of recommendations. The **similarity** can be based on product attributes (size, weight, ...) or ratings. The idea is to find potential similar items.

3 Model & Implementation

Our model extends the one *pseudo-universal* HRTF approach to **multiple different HRTF profiles**. Recommender system (**collaborative filtering**) serves as an oracle assigning the best possible HRTF profile to every person. In contrast to e-shop customer, the user of our application does not explicitly rate the HRTF profiles. The user just points to the sound source and the system calculates the rating as the difference between the actual position and assumed position.

We have created **web based application** implemented in WebGL and WebAudio technologies. The application is controlled using the mouse. This **provides access to a large number of people** regardless of the operating system. In addition, pointing the position directly in the application allows the user to have closed eyes all the time and concentrate. The implementation does not require any expensive virtual reality devices. However, the diploma thesis also describes approaches using other technologies (without a computer or with VR device) and other models of recommendation systems.

4 Testing

The observed statistic was the **angle** between the **tested** and the **assumed** position. We have expected that the inaccuracy will decrease in time compared to the non-personalised HRTF profile which was simulated by random recommendations in the first phase. Users used the application for about 2x15 minutes in two days. We used the recommender system only in the second part of testing.

model	**average inaccuracy** [°]
withour RS	29.7 ± 44.2
with RS	13.9 ± 17.3
with RS (at the end)	6.6 ± 1.5

Conclusions

The user's inaccuracy during the training session (visible sound source) was 5.5 degrees on average. At the end of the second part of the test, we achieved an inaccuracy of 6.5 degrees on average. **Improvements** are considerable.

We have proved that the combination of binaural audio with simple collaborative filtering can **increase the accuracy** of audio virtual reality in short time. Easier access to HRTF and growth of the HRTF databases can lead to the **expansion of assistive devices using spatial audio** and a new segment of the entertainment sector (games, movies, ...).

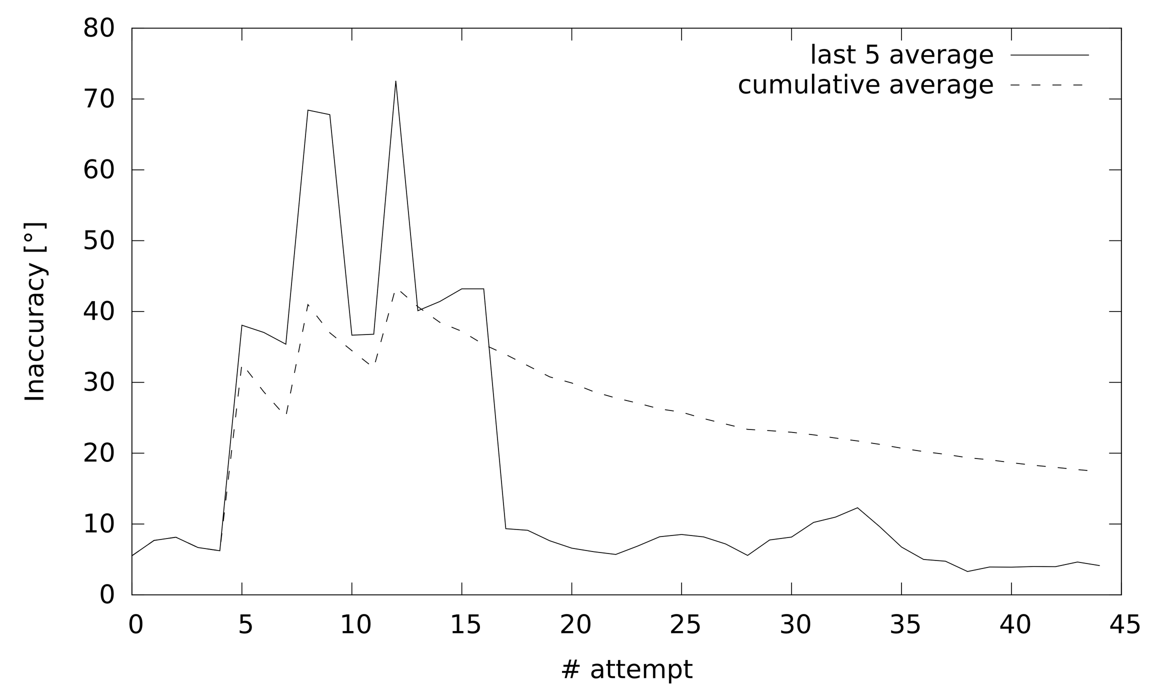

Figure 2: Inaccuracy of one user