Audio Engineering Society

Conference Paper 56

Presented at the Conference on
Immersive and Interactive Audio
2019 March 27 – 29, York, UK



# The Effect of Transitioning Between Individualized and Generic HRTFs on Localization Performance in a Virtual Environment

Yun-Han Wu[1], Scott M. Murakami[1], and Agnieszka Roginska[1]

[1]*New York University*

Correspondence should be addressed to Yun-Han Wu (yhw261.nyu.edu)

**ABSTRACT**

The main purpose of this paper is to observe and analyze how people's localization performance changes as the HRTFs used transition from an individualized set to a non-individualized (generic) set. Two common HRTF interpolation techniques, one in the time domain and the other in the frequency domain, are used to create averaged sets of HRTFs which fit between the two extremes and helps to examine the trend of localization behavior change through this continuum.

## 1 Introduction

The recent boom in XR (virtual, augmented, and mixed reality) technology has brought an increased necessity to 3D audio developments for these applications, in particular tackling the problem of personalized Head-Related Transfer Functions (HRTFs) to create a more realistic and immersive sound environment.

In their research, Roginska, Gregory & Thomas (2010) pointed out that non-individualized HRTFs are capable of providing as good of a spatial image as individualized ones through a process of user preference selection of HRTFs based on externalization quality, elevation, and front/back discrimination [1]. As a result, some papers have discussed the use of machine learning techniques on HRTF databases to achieve preference selection [2].

Other research still stresses the importance of individualized HRTFs and focuses on finding more convenient ways to capture it, such as physical modeling [3] or incorporating the measurement process into real world environments that does not require the need for an anechoic chamber or special excitation signals such as a sine sweep [4].

However, there is relatively little research done so far in comparing differences between human's localization performances on individualized versus generalized HRTFs in the same environment, and how the localization performance changes as the HRTFs are shifted between these spectrums via averaging and interpolation methods of two sets. As a result, the main purpose of this paper is to observe how people's localization performance and behavior changes in response to HRTF sets generated by averaging general and individualized HRTFs and interpolating between the two.

In section 2 of this paper, the background knowledge of human sound localization behavior and the HRTF

interpolation techniques utilized in this research is presented.

In section 3 of this paper, the process of obtaining the averaged sets of HRTF from the individualized and generic sets with techniques in time domain and frequency domain respectively are described in detail. In addition, the VR application built in Unity to run the subjective testing and capture subjects' sound localization data is also detailed.

The subjective testing is also described below, where users were asked to localize 30 short bursts of broadband noise signals for four sets of HRTFs (generic, individual, averaged set generated from the time domain method, and averaged set generated from the frequency domain method). Subjects were only asked to localize sounds along the horizontal plane.

## 2 Literature Review

### 2.1 Head-Related Transfer Function

Humans use three important auditory cues to localize sounds and pinpoint where they are coming from. The three cues are known as the Interaural Level Difference (ILD), Interaural Time Difference (ITD) and Spectral cues. It is indicated that the ILD is "the amplitude difference of the audio signal perceived at the eardrums", while the ITD is defined as "the delay between the arrival time of a waveform at the first and second ear." ILD and ITD cues provide the brain information to locate sounds along the horizontal plane [5].

However, since sounds that fall on the median plane will result in no changes between the ITDs and ILDs between each ear, spectral cues from reflections off the pinna and other body parts such as the shoulders are necessary to localize sounds along this plane [6]. As a result, according to [7], this often results in a front/back or a up/down confusion on sound localization. In order to identify the elevation of the sound source during localization, more detailed spectral cues are needed.

As the sound reflects within the convolution of external ear, specifically the pinnae, spatial cues that disambiguate the potential locations of the source are created. To be more specific, the changes in the frequency spectrum of the sound source, which resulted from different reflection angles of each sound wave, provide the human brain with the necessary information for identifying the position of the source relative to the listener.

In conclusion, the ITDs, ILDs, and spectral cues can be considered as filters by the physical structure of human torso, head and pinna. The combination of these auditory cues is used in developing head-related transfer functions (HRTFs) which describes the effects of the cues as a linear time-invariant system [8]. HRTFs represent head-related impulse responses (HRIRs) as a result of a transfer function from the time domain to the frequency domain, and can be identified by direct excitation with various approaches such as an exponential sine sweep, impulse of Gaussian white noise, and MLS codes.

### 2.2 HRTF Interpolation

In a time-variant sound field environment, updating a dense grid of HRTFs is necessary to get binaural audio. However, if the spatial resolution between each HRTF is too large, spatial artifacts will become present while switching directly from one HRTF point to another [9]. To prevent such spatial artifacts caused by a lack of resolution between HRTF points, different interpolation methods were proposed to interpolate points between different sets (individual sets that may lack in resolution and generic sets which have many points and high resolution) such as direct interpolation in the time domain [10] or frequency domain [11], to fulfill a smooth transition. For the purpose of this paper, two of the linear interpolation methods, one in the time-domain and one in the frequency domain, are chosen and modified to obtain an averaged set of HRTFs from individual and generic sets.

The two HRTF interpolation techniques used in this paper are based on [12] and [13] respectively.

One of the most commonly used interpolation algorithms is the linear time domain method due to its computational efficiency. However, the time delay difference between the HRIRs corresponding to locations near the desired spatial location sometimes results in a distorted magnitude spectrum on the interpolated HRIR. As a result, researchers came up with several different approaches to preprocess the data. For example, aligning two neighboring HRIRs on the time axis before interpolating them. It has been found that this method produces less absolute error between the logarithmic magnitude spectrum of the interpolated point and the corresponding original HRIR.

Despite the fact that a lot of the research approximates the phase component of HRTFs with onset delay, it has been stated that minimum-phase version of HRTF still contains some temporal components. However, if a set of HRTF is separated into phase and magnitude components, the temporal and spectral components can be processed separately. In addition, It is stated that the phase component of weighted averaged complex HRTFs is not exactly the same as the weighted average of phase response. Therefore, the method used in this paper was originally proposed in [13] to ensure that the complete phase response is included and accurately represented when interpolating HRTFs over the horizontal plane.

## 3 Methods

### 3.1 HRTF measurements

10 Subjects' individual HRIRs were measured in the NYU MARL research lab. The ScanIR tool [14] was used to generate the sine sweep, played back via a Genelec 8030A loudspeaker and recorded with the BACCH-BM Pro binaural in-ear microphones. In addition to generating the sine sweep, ScanIR also records the HRIRs. A Sound Devices USBPre 2 interface was used to send the sine sweep to the speaker and receive the incoming the microphone signals for recording at a sampling rate of 96kHz. Subjects were placed at a distance of one meter away from the speaker, with the high frequency driver at ear level. Individual HRIRs were recorded at 24 points across the horizontal plane, every 15°.

The generic set of HRTFs used were also recorded in the NYU MARL research lab using a Neumann KU-100 binaural dummy head attached to a stepping motor controlled by an Arduino UNO and MATLAB, using the Psychtoolbox [15] as well as the Arduino support package for MATLAB. A modified version of the ScanIR tool was used to generate the sine sweep and record the HRIRs; the tool was modified for use with a multi-loudspeaker configuration as well as the stepping Arduino motor. Six Genelec 8030A loudspeakers were arranged in a spiral configuration, with a spacing of 22.5° in azimuth and 18° in the median plane. The dummy head was placed with a one meter radial distance from the 0° azimuth speaker. The resulting set of HRTFs contains a total of 1200 measurement points, 200 along the azimuth and 6 elevations per step [6].

For the purpose of this paper, only the points along the horizontal plane were used.

### 3.2 Time Domain HRTF Interpolation

Since each HRIR can be represented as a minimum-phase system integrated with a positional dependent ITD, the first step of the time domain interpolation method is to get the ITDs from both individualized and generic set of HRTF. In order to achieve that, a cross-correlation is performed on each pair of the left-ear and right-ear HRIRs. In other words, ITD is estimated as the difference of the left and right HRTF arrival time as in Equation 1.

$$ITD_{\theta,\phi} = argmax \sum h_{l,\theta,\phi}(n) h_{r,\theta\phi}(n-t) \quad (1)$$

According to Smith (2011), any filter transfer function can be made minimum-phase by completely factoring it and reflecting all the zeros of the filter inside the unit circle [16]. However, factoring such a large polynomial can be impractical. An approximate, non-parametric method based on the property of complex cepstrum, which is that each minimum-phase and maximum-phase zero in the spectrum gives rise to a causal and anti-causal exponential in the cepstrum respectively. As a result, by converting the latter to the former, the corresponding spectrum is transformed non-parametrically into its minimum-phase version [12].

After the minimum-phase HRTF is obtained, the individualized and generic frequency spectrums are summed and averaged to get the third set of HRTF based on these two sets. In the original paper, the time related component, such as the ITD, is also averaged between HRTFs at different angles. For example, the HRTF at 15° azimuth, 0° elevation is interpolated with the one at 10° azimuth, 0° elevation and 20° azimuth, 0° elevation. However, in our paper the individualized ITD is inserted back onto the averaged set of HRTFs for the reason that in the previous literature, HRTF interpolation is only utilized on the data captured from the same person.

However, the same technique is used to average different sets of HRTFs, which in the case of this paper is a subject's individualized HRTFs and a generic set of HRTFs. According to previous research, ITD alone provides a critical piece of information for sound localization. The complete process of obtaining the averaged HRTF is demonstrated above in Figure 1.

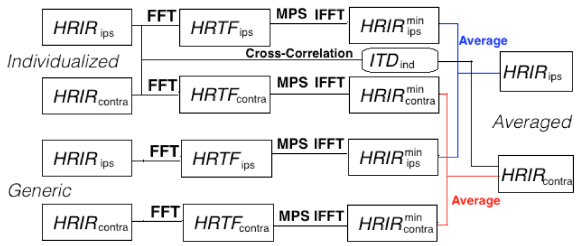


**Fig. 1:** The pipeline for time-domain interpolation method

### 3.3 Frequency Domain HRTF Interpolation

The first step of the frequency domain interpolation method is to perform FFTs on both the generic and individualized set of HRIRs to obtain the phase and magnitude components of its frequency domain representation, HRTFs. Only the magnitude components of both sets of HRTFs are averaged to acquire the third set of HRIRs, while the phase components from the individualized HRTF is kept and used for the third set to make sure we have the right temporal information.

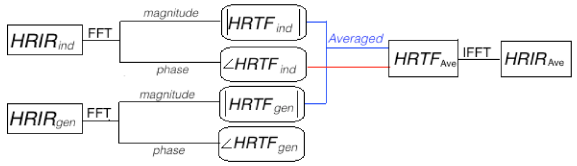


**Fig. 2:** The pipeline for frequency-domain interpolation method

$$HRTF_{ave} = |HRTF_{ave}|e^{j\angle HRTF_{ave}} \tag{2}$$

The complete process of obtaining the averaged HRTF is demonstrated below in Figure 2. In addition, the last step is also represented mathematically in Equation 2.

### 3.4 VR application in Unity

A first-person VR environment is built in the Unity application to assess the localization abilities of 10 subjects using different sets of HRTFs, individualized, generic, and the two interpolated sets. All of the data captured for a subject, including information such as the perceived and correct location of each sound stimuli, response time, angle of rotation, every 100 ms, are stored in a single comma-separated values (CSV) file.

The Unity application was run on an Alienware 13 laptop with an Oculus Rift VR headset and a pair of Sennheiser HD 650 open back headphones. The use of customized sets of HRTFs in Unity is achieved with a framework provided by SteamAudio and an additional script written by Jim Mattingly, a master graduate from the Music Technology program at NYU. A four-nearest neighbor interpolation method, weighted by each point's Euclidean distances from the target point, is also used to get a smooth transition for binaural rendering in the virtual environment.

Subjects are placed in a virtual environment throughout the whole experiment. Twenty four audio sources surrounding the subjects along a circle on the horizontal plane represent all the possible locations of sound stimuli. They are placed at eye level, determined by calibration of the headset for users, and in fifteen degree intervals. On a slightly higher plane, two spheres are positioned on top of the $0^\circ$ and $180^\circ$ spheres, so that the subjects can identify their front-back orientation easily. An 800ms burst of Gaussian noise sampled at 44.1 kHz is used as the sound stimuli for testing in this experiment.

Each subject's localization ability is tested for 4 sets of HRTFs (individualized set, generic set, time-domain averaged set, and frequency domain averaged set) with 30 trials for each set. The sets of HRTFs were tested in random order for each subject. As the first step in each trial, subjects have to focus their gaze onto the front sphere above the $0^\circ$ sphere and push the A button on the Oculus touch controller to trigger a sound stimuli. Then, subjects are asked to localize the sounds by turning their head to the source and reporting the perceived position of the sound source as quick and accurate as possible by pressing the same button while facing the point at which the perceived target is.

## 4 Results

### 4.1 ITD Analysis

The results of the ITD analysis are presented as a time delay curve on the horizontal plane in Fig. 3. The ITDs were calculated as the maximum interaural cross-correlation, which was output as the index of the maximum coherence before converting into the final value as milliseconds.

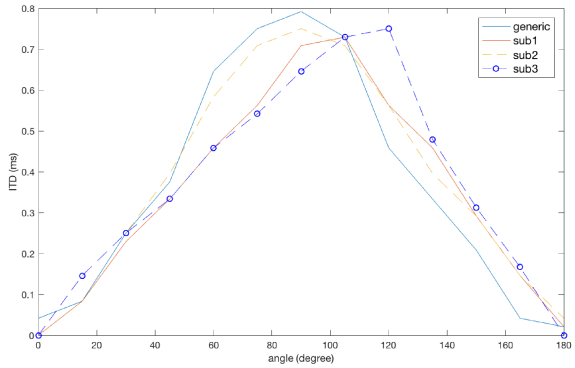


**Fig. 3:** The ITDs of the generic and three subject's individualized HRTF from 0°to 180°.

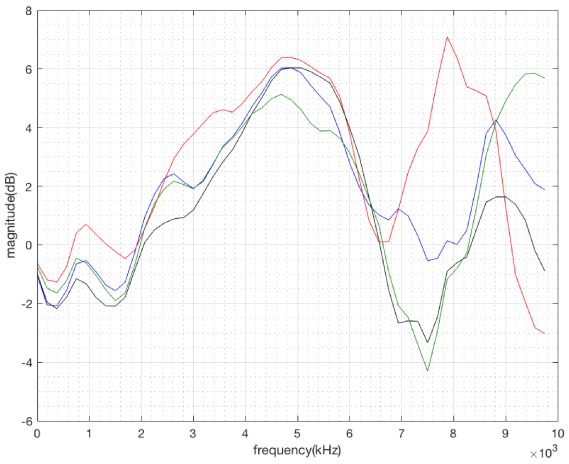


**Fig. 4:** Four subjects' smoothed right-ear HRTFDs at 0 azimuth degree.

As we can see from the plot in Fig. 3, there is a noticeable difference in the ITDs between some of the subjects' individualized sets and the generic data. Some of the subjects presented an evident divergent behaviour starting from 45°to 120°with a peak ITD value at distinct degrees. The largest difference found between the generic and all individualized ITDs is around 0.3ms occurring at $\phi$ = 120°(subject 3 in Fig. 3). On the other hand, there are also people like subject 2 who has an ITD similar to the dummy head.

### 4.2 HRTFD Analysis

The HRTF Difference (HRTFD) has been defined as the quotient of magnitude spectra of two HRTFs from the same direction and has been found to appropriately characterize differences between HRTFs [17]. This value is suitable for identifying the deviation in the frequency magnitude response of individualized and generic HRTFs. It can be calculated using the equation below:

$$HRTFD(\phi,\delta) = 20\log_{10}\left\{\frac{|HRTF_{ind}(\phi,\delta)|}{|HRTF_{gen}(\phi,\delta)|}\right\} \quad (3)$$

In the equation, |HRTF($\phi$,$\delta$)| is the frequency magnitude response of the HRTF located at azimuth $\phi$ and elevation $\delta$. HRTF$_{gen}$ is the free-field dummy-head HRTF measurement, while the HRTF$_{ind}$ is either the individualized HRTF for each subject.

From the horizontal plane analysis of the right-ear HRTFDs at 0°azimuth in Fig. 4, it can be observed that the differences between generic and each individualized HRTF becomes more obvious above 2 kHz, and the peak magnitude difference for this case can be found at around 8 kHz, up to 7 dB of difference. Considering the scope of this paper, only a preliminary analysis at the 0°azimuth is provided, so the readers can have a rough idea on the baseline difference of how the sound is received at each subject's ears, directly from the source. All the individualized HRTFs at 0°for different subjects demonstrates a relatively similar pattern under 6.5 kHz. An interesting observation to point out is that an opposite pattern can be recognized between some of the subjects' HRTFs between 7k Hz to 9 kHz, where some subject's have a drop in magnitude between these frequencies where some have a drastic increase in magnitude.

### 4.3 Localization Performance Analysis

#### 4.3.1 Data Preparation and Removing Outliers

The large amount of data gathered in this experiment requires some organization before the analysis of said data. In order to find out how each subject's localization performance changes between different sets of HRTFs, the analysis will be separated into four sections, which are the accuracy, speed, head movement onset time and head rotation pattern. First, the localization accuracy of each subject can be represented with the error angle

calculated as the absolute difference between the azimuth degree of the target and of the reported location. Second, the speed at which each subject localizes the sounds is directly reflected by their response time on each trial. Lastly, the head rotation data, recorded every 50 ms, is processed to find the onset of each subject's head movement and in which way they initially turn their head to search for the target.

Given that this experiment provided subjects adequate time to localize the sound and the fact that every subject is instructed to prioritize the accuracy of performing the task over speed, any trial with an error larger than 45°, which is significantly larger than the highest localization blur of humans on the azimuth plane and is commonly used as the boundary for rough direction pointing, is considered invalid and is removed from the results. These severe errors over 45°are not what we are trying to analyze in our research, because the localization performance difference between generic and individualized HRTF should be smaller than localization blur, and huge errors will also skew the data.

#### 4.3.2 Accuracy

As a first impression of the data, the mean and standard deviation of all subjects' error with each of the four sets of HRTFs is shown in Table 1, and the mean of ranking of all subject's error, the lower the error value the higher the ranking, with all sets of HRTFs is shown in Table 2. As can be seen from the tables, the individualized HRTFs produced the best results, while an increase in the subject's errors can be seen when tested using the generic HRTFs. One thing to notice is that the standard deviation of subject's error values with the use of the generic HRTFs is significantly higher in comparison to the other three sets.

| Error(degree) | Gen | Freq | Time | Ind |
|---|---|---|---|---|
| Mean | 11.74 | 7.03 | 8.06 | 6.80 |
| Std | 9.3 | 2.7 | 2.65 | 2.69 |

**Table 1:** The mean and standard deviation of all subjects error on four sets of HRTFs

| Error(ranking) | Gen | Freq | Time | Ind |
|---|---|---|---|---|
| Mean | 3.1 | 2.2 | 2.8 | 1.7 |

**Table 2:** The mean of ranking of all subjects error on four sets of HRTFs

#### 4.3.3 Speed

The mean and standard deviation of all subjects with each set of HRTFs are demonstrated in Table 3. Based on the data, there is no clear trend on how any set of HRTFs significantly influences subject's performance.

| Resp Time(secs) | Gen | Freq | Time | Ind |
|---|---|---|---|---|
| Mean | 8.82 | 7.55 | 9.33 | 5.18 |
| Std | 1.55 | 1.39 | 2.30 | 1.79 |

**Table 3:** The mean and standard deviation of all subjects response time on four sets of HRTF

#### 4.3.4 Head Movement Onset

In order to find the onset for each trial, the velocity of a subject's head movement, which is calculated as the derivative of the head rotation data every 50ms, is extracted from the raw data and sorted in ascending order. The greatest five values are considered but only the first in time is determined as the onset.

To minimize the chance of selecting an incorrect onset time frame due to subject's continuous head movement while searching for the target, the five highest velocity values found at all time points are picked as candidates rather than choosing the highest one directly. Finally, the output value is the index of the onset time frame and needs to be converted into the onset time in milliseconds.

| Onset(frames) | Gen | Freq | Time | Ind |
|---|---|---|---|---|
| Mean | 30.01 | 21.88 | 28.31 | 21.65 |
| Std | 12.94 | 8.48 | 10.63 | 9.18 |

**Table 4:** The mean and standard deviation of all subjects onset on four sets of HRTF

The mean and standard deviation of all subject's onset time with each four sets of HRTF is shown in Table 4. As can be observed from the data, the data distribution of onset time between different sets of HRTFs follows a trend similar to the one for error value, which will be discussed in detail in the section 4.

#### 4.3.5 Head Rotation Pattern

Head rotation pattern is relatively problematic to analyze due to the difficulty of converting it into statistically processable data. However, a systematic observation can be made on the plots generated for demonstrating each subject's head rotation changes in time.

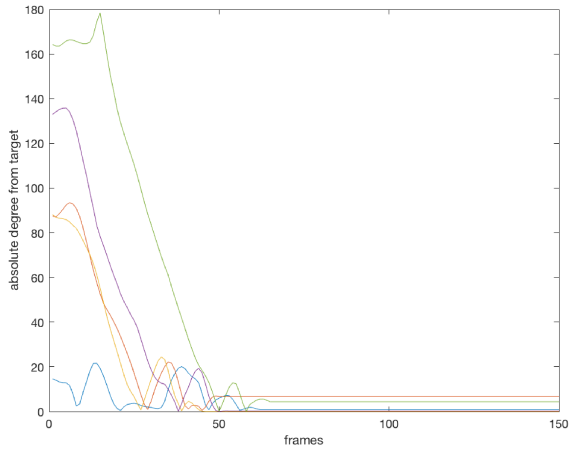


**Fig. 5:** Five Head Rotation data from one subject performing localization task with individualized HRTF.

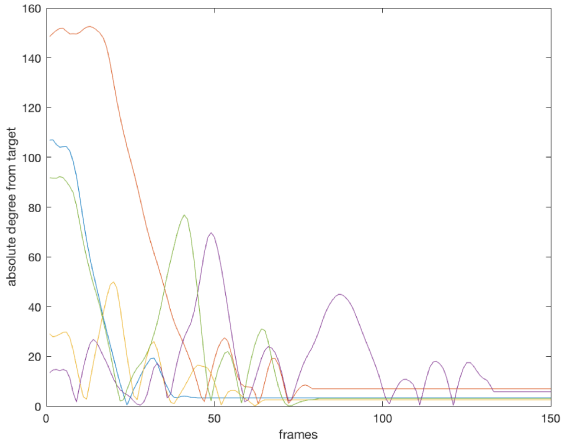


**Fig. 6:** Five Head Rotation data from one subject performing localization task with averaged HRTF

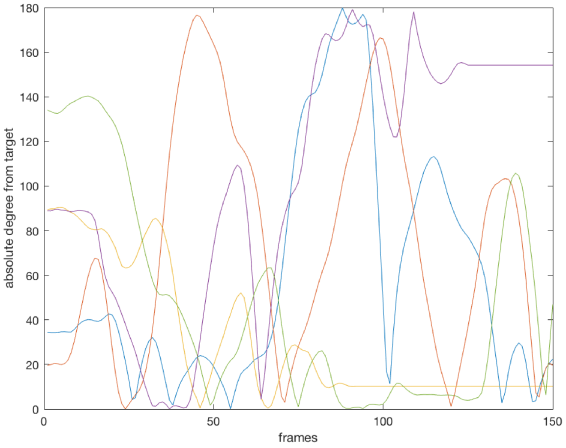


**Fig. 7:** Five Head Rotation data from one subject performing localization task with generic HRTF

In fact, analyzing head rotation pattern is a convenient approach to see whether subjects are having front back confusions or are struggling to identify the exact position of a sound stimuli at the final stage of the task.

Figures 5-7 demonstrate three head movement patterns seen frequently with subject's head rotation patterns on the individualized, averaged and generic sets of HRTFs respectively. The lines represent the angle difference from target of where the subject is facing in each frame.

## 5 Discussion and Conclusion

In this section, two major topics will be discussed based on the results shown in the previous part. First, we would like to compare subject's performance on four sets of HRTFs in all aspects: accuracy, speed, onset, and head rotation pattern; to find a trend on how it changes accordingly. Second, we take a closer look at the results on averaged HRTFs created using time and frequency domain techniques respectively to see which one produced better results and what might be the reason/cause behind it.

As we can see from Table 1, the mean error value decreases in order of generic, time, frequency and individualized sets, which matches our hypothesis that subjects will reach their best performance on individualized HRTFs and get worse when the HRTFs shifts closer towards the generic set. However, it is also obvious that there is a significantly larger gap between the error value of generic HRTFs and the other sets. We believe that this error can be explained by the fact that the majority of time-sensitive information in individualized HRTFs, such as ITD and phase, were preserved when creating both of the averaged sets, because the time information is regarded to be more important than spectral components for sound localization on the horizontal plane, as stated in previous literature. On the other hand, the relatively small differences between the individualized HRTFs and the two averaged sets indicates that frequency contents alone plays a minor role in horizontal sound localization.

It has been pointed out in the results section that the standard deviation of subject's error value on generic

HRTFs is noticeably higher than others. With a closer look of the data, we found that there are two subjects who have much larger mean errors while performing the localization task with generic HRTFs. As a result, a ranking chart is also provided in Table 2 to help us view the data without outliers. Despite the fact that the general pattern stays the same, the mean ranking is more evenly distributed comparing to directly looking at the error value.

In terms of the response time, no trend can be identified just by looking at the mean and standard deviation on Table 3. But this information helps us to interpret the onset time data. Usually, the onset time somehow follows the trend of response time, since people tend to have a shorter response time in finding the stimulus if they start moving their head earlier. Nevertheless, the pattern observed on the mean data in Table 4 is actually more similar to what we've seen in Table 1, which charts the error values. As a result, the head movement onset, which represents how easy it is for the subjects to localize the sound source at an instance, is positively correlated with amount of error that subjects made in our experiment.

Finally, the head rotation plot provided good representations of the kinds of movement people do and how efficient they are in finding the sound source with different sets of HRTFs. As we can see from Fig. 5, the subject turns directly to face the target in one smooth movement before doing minor adjustments on the specific angle. This pattern appears more commonly when subjects are tested on individualized HRTFs. On the other hand, we can see a very clear indication of front back confusion in Fig. 7 where the subjects rotated to the completely opposite side of the target, but eventually resolve the issue using the relative movement of the sound to their head. And matching with our hypothesis, this pattern of front/back confusion can be seen more often when subjects are performing the task with generic HRTFs. Last but not least, Fig. 6 demonstrates a head movement pattern that fits in between the above two categories, which happens quite evenly in the results from all four sets of HRTFs. An interesting finding for this section is that how frequently these patterns were observed in each set of HRTFs is not reflected in the response time analysis. Since it makes more sense to assume that the response time will be noticeably higher if the subjects turn back and forth to search for the sound source. This is something that will be researched further in the future work.

In terms of whether the time- and frequency- domain HRTF interpolation techniques are effective in keeping components that are critical for auditory localization, we can interpret the results in several different ways. Looking at the statistical analysis on error values and head movement onset times, the results from the time-domain method is worse and is a lot closer to the data resulted from the use of generic HRTFs. In addition, the response time data and the feedback from several of our subjects reveals that stimuli is harder to localize and location is ambiguous during head movements. It is stated in previous literature that the time-domain method is mainly used for real-time systems that aim to minimize the latency, but that it also produces less stable results. The results from our subjective testing supports the literature in that the time domain averaging method produces less stable results in localization accuracy.

## 6 Acknowledgement

We thank Jim Mattingly for proving the framework of using customized HRTFs in Unity with the SteamAudio APIs, and Andrea Genovese for recording and sharing the generic HRTF data captured using the Neumann KU-100 binaural microphone.

## References

[1] Roginska, A., Wakefield, G. H., and Santoro, T. S., "Stimulus-dependent HRTF Preference," in *129th Audio Engineering Society Convention*, 2010.

[2] Chun, C. J., Moon, J. M., and Lee, G. W., "Deep Neural Network Based HRTF Personalization Using Anthropometric Measurements," in *143rd Audio Engineering Society Convention*, 2017.

[3] Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z., "Approximating the head-related transfer function using simple geometric models of the head and torso," *Journal of Acoustic Society in America*, 112(5), pp. 503–516, 2002.

[4] Diepold, K., Durkovic, M., and Sagstetter, F., "HRTF Measurements with Recorded Reference Signal," in *129th Audio Engineering Society Convention*, 2010.

[5] Moldoveanu, F., Moldoveanu, A., and Balan, O. M., "Training system for improving spatial sound localization." in *In The International Scientific Conference eLearning and Software for Education*, volume 4, p. 79, 2002.

[6] Genovese, A., Zalles, G., Reardon, G., and Roginska, A., "Acoustic perturbations in HRTFs measured on Mixed Reality Headsets," in *In Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, 2018.

[7] Toledo, D. and Møller, H., "The role of spectral features in sound localization," in *In Audio Engineering Society Convention*, 2008.

[8] Baumgartner, R., Majdak, P., and Laback, B., "Modeling sound-source localization in sagittal planes for human listeners," *Journal of Acoustic Society in America*, 136(2), pp. 791–802, 2014.

[9] Minnaar, P., Plogsties, J., and Christensen, F., "Directional resolution of head-related transfer functions required in binaural synthesis," *Journal of the Audio Engineering Society*, 53(10), pp. 919–929, 2005.

[10] Wenzel, E. M. and Foster, S. H., "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in *In 1993 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993.

[11] Sodnik, J., Sušnik, R., Štular, M., and Tomazic, S., "Spatial sound resolution of an interpolated HRIR library," *Applied Acoustics*, 66, pp. 1219–1234, 2005.

[12] Chen, L., Hu, H., and Wu, Z., "Head-related impulse response interpolation in virtual sound system," *In Natural Computation Fourth International Conference of IEEE*, 6, pp. 162–166, 2008.

[13] Reddy, C. S. and Hegde, R. M., "Horizontal plane HRTF interpolation using linear phase constraint for rendering spatial audio," *In 24th Signal Processing Conference of IEEE*, 6, pp. 1668–1672, 2016.

[14] Boren, B. and Roginska, A., "Multichannel Impulse Response Measurement in MATLAB," in *Audio Engineering Society Convention 131*, 2011.

[15] Kleiner, M., Brainard, D., Pelli, E., Ingling, A., Murray, R., and Broussard, C., "What's new in Psychtoolbox-3," in *Perception 36, no. 14*, 2007.

[16] Smith, J. O., "Minimum-Phase Filter Design," in *Spectral Audio Signal Processing*, 2011.

[17] Wersényi, G., "HRTFs in human localization: measurement, spectral evaluation and practical use in virtual audio environment," 2002.