ICAD 2013
6-10 JULY, 2013, ŁÓDŹ, POLAND
INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY

A DATABASE OF REPEATED HEAD-RELATED TRANSFER FUNCTION MEASUREMENTS

Areti Andreopoulou[1], Agnieszka Rogińska[2], Hariharan Mohanraj
Music and Audio Research Laboratory (MARL)
New York University
New York, NY
[1]aa1510@nyu.edu, [2]yroginska@nyu.edu

ABSTRACT

This paper describes a new HRTF collection, measured at the Music and Audio Research Laboratory (MARL), at NYU. This collection of 40 datasets, consists of repeated HRTF measurements on 4 subjects (10 HRTF sets per subject). Analysis of the data offers an understanding of the expected degree of variation in HRTF sets, which, when supported by subjective evaluation, can provide deeper insight for perceptually detected differences between binaural filters. Such knowledge has applications in various research fields and related signal processing tasks, such as: binaural auditory displays, HRTF modeling, and HRTF interpolation.

1. INTRODUCTION

As sound travels towards a listener, it undergoes a number of distortions, caused by the interaction of its waves with the environment, as well as the listener's body, head, and outer ears (pinnae). This source-to-receiver system can be characterized by pairs of Head-Related Impulse Responses (HRIRs), or in the frequency domain Head-Related Transfer Functions (HRTFs). Due to the high variability in people's head shapes and especially pinnae, HRTFs vary vastly among individuals, much like fingerprints. As a result the use of non-individualized HRTFs in binaural reproduction can lead to higher localization errors, lateralization effects, and unconvincing spatial impression in general. At the same time the need for personalized measurements makes it challenging to apply binaural technologies in fields like virtual auditory displays, mission critical applications, entertainment etc.

In the past decades, in an attempt to overcome this problem, a lot of research has been conducted in the fields of HRTF individualization [1], [2], [3], [4], [5] and database matching [6], [7], [8]. Such research relies thoroughly on the analysis and use of preacquired HRTF dataset from human subjects or dummy heads. The need for more data of high quality and accuracy has soon become apparent. Several laboratories have created their own databases to support their research, which led to a small number of publicly available HRTF collections. The following section offers an overview of the most known HRTF databases.

2. HRTF DATABASES OVERVIEW

In 1995 the Media Laboratory in Massachusetts Institute of Technology (MIT), released an HRTF database captured on a KEMAR mannequin [9]. A total of 710 locations were recorded at 14 different elevations from -40° to 90° in 10° vertical, and roughly 5° horizontal increments. The resulting Impulse Responses were minimum-phase filters at a sampling frequency of 44100 Hz. This database remains until nowadays one of the most popular options in binaural audio related research.

In 2002 IRCAM in collaboration with AKG released an HRTF database, as part of the Listen research project [10]. The set that consists of 51 subjects was captured at 10 elevations with an azimuthal resolution of 15°, and a sample rate of 44100 Hz. The Impulse Responses were made publicly available as minimum-phase filters. The LISTEN database is popular in binaural research, mainly due to its population, as it remains one of the longest available collections of personalized filters.

The CIPIC database was captured at the Center for Image Processing and Integrated Computing, University of California Davis [11]. The collection consists of 45 subjects, measured at 50 elevation and 25 azimuth locations in roughly 5° increments with a sampling rate of 44100 Hz. Each set is accompanied by a series of 27 anthropometric measurements. Due to the density of the measured locations, and the high resolution of the anthropometric properties, the CIPIC database is another commonly used pool in HRTF customization research.

In 2007 the Acoustics Laboratory, of South China University of Technology, published an article introducing a database of measurements from Chinese subjects, exclusively [12]. Filters were captured in 9 elevations at roughly 5° azimuthal increments with a sample rate of 44100 Hz. This was the first time that a database was created with targeted anthropometric properties, originating from similarities in people's broader ethnic characteristics. Through statistical analysis the authors presented their claim for the need of such targeted databases by proving that there is significant difference between the average ITD, head and pinnae sizes, of their filter collection and the CIPIC sets. This data collection hasn't been made publicly available yet.

The Florida International University DSP Lab released in 2010 a collection of 15 HRTF sets [13]. The measurements were done in 6 different elevations, at 30° horizontal increments, with a sample rate of 96000 Hz. 8 anthropometric measurements of each subject's pinnae, were also collected from high resolution, 3-dimensional laser scans. This high resolution recorded filters and scans provide a solid ground for HRTF individualization research.

The most recent, publicly available HRTF database was released by the Quality and Usability Lab, Deutsche Telekom Laboratories, TU Berlin [14]. It includes HRTF sets of the KEMAR mannequin at four different distances from the sound source ($3m$, $2m$, $1m$, and $0.5m$). All measurements were done on the horizontal plane at 1° azimuth increments, and a sample rate of 44100 Hz.

This data allows for a detailed analysis on the variations between the far and near field binaural filters.

3. NEED FOR STUDY

When one acquires a pre-measured HRTF database, mainly they have access to is a collection of filters along with their corresponding metadata, namely per-location azimuth and elevation coordinates, distance from the sound source, and optional information regarding the subject's anthropometry. Moreover, there is little that can be done to verify the accuracy of the reported metadata and filters' ability to fully represent the spatial cues of a certain subject, and much less to confidently adjust any of their discrepancies without introducing further variations.

HRTFs are measured at a small set of finite locations, which offers a low resolution representation of a subject's spatial cues to begin with. The space that the measurements take place in and the equipment used can potentially affect the quality of the measurements. For example, loudspeakers are far from being omnidirectional point sources, and their physical size may introduce unwanted reflections and directivity patterns to the system. The placement of the miniature microphones in the subject's ears, is also not a trivial process as it has been found to significantly affect high frequency content [15].

The absolute variations in HRTF measurements has been reported before by Riederer [16], who offered a detailed overview of the degree of error introduced to an HRTF filter-set by a number of factors, such as the background noise, reflections from the measurement system, accidental movements or misalignment of the subject, the use of ear plugs, accurate placement of the miniature microphones etc. The study explored the effects of a single error factor at a time, while in real-life situations more than one of them can occur during the HRTF measurement process.

Another similar study was done by Wersényi, and Illényi [17], who have carried out repeated measurements on a computer-controlled dummy head measurement system. The study was exploring the effect of near-the-head everyday objects like hair, caps, glasses and clothing, on the captured binaural filters. Unfortunately, the measurement data of either of the two studies was never publicly released, in order to allow for further experimentation, or reproduction of the reported results

This paper describes a publicly available HRTF collection, which consists of repeated impulse response measurements across 5 different elevations, for multiple subjects. The process was designed to maximize variation between sets. The data was measured in 2 different sound treated spaces, using 2 different tracking systems of the subject, a laser pointer and a magnetic tracker. The speaker setup was positioned twice during the data collection process, the miniature microphones were removed and replaced between HRTF measurements, 2 different excitation signals were used (MLS and Golay codes), and different rounds of the experiment were monitored by two different supervisors.

Such data will allow for further studies on the variability of HRTF measurements, and reproduction of the previously reported observations. Knowledge of the expected variations in spatial filters can offer an insight on the range of change in directional cues and its affect on a participant's spatial listening experience. When combined with subjective experiments, it will also describe the perceptually relevant changes in binaural filters.

Having identified the boundaries of expected variations in HRTF sets, operations like binaural filter tuning may be done more confidently to adjust cues, or even drastically to correct them, achieving in that way optimum personalized immersive experiences for people. For these reasons, the creation of a publicly available database of repeated HFRT measurements is very relevant to the current state-of-the-art research in binaural audio reproduction, with direct application in fields like auditory displays, virtual auditory environments, HRTF personalization etc.

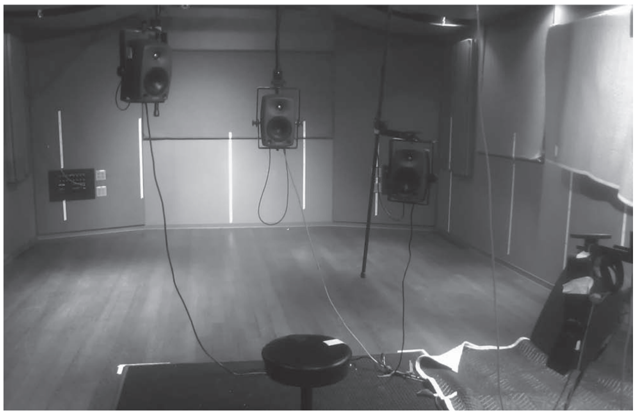

Figure 2: HRTF measurement speaker setup in the Dolan Studio live room

4. THE MARL HRTF DATABASE

4.1. Measurement process

The measurements were conducted at the facilities of the Music and Audio Research Laboratory, at New York University. The recordings took place in two different locations: the Spatial Auditory Research Lab, a 4.5 by 3.5 by 2.5 m semi-anechoic room, and the Dolan Studio live-room, a 9 by 4.6 by 3 m sound treated space. In both cases five Genelec 8030a speakers were setup in a spiral configuration at 15° vertical increments from -30° to $+30^\circ$ in elevation. Subjects were seated on a rotating stool with adjustable height, at the center of the spiral, at a distance of $1m$ from the speakers. The stool had no back or arm rests, and the location of one's head was not fixed to a head rest. All speakers were calibrated using pink noise at 87 dB SPL as measured from the position of the subjects. Impulse responses from all speakers were measured from the center of the spiral, to be used in post processing for elimination of the playback equipment responses from the HRIR filters. The data was recorded at the blocked entrances of the subjects' ear canals, using foam earplugs and custom made miniature binaural microphones with Sennheiser's $KE-4$ capsules.

For four of the HRTF sets the alignment of the subjects was monitored using laser pointers. Marks were put on the room walls and floor to help subjects adjust their rotation angle and head position. Two laser pointers were used for alignment, adjusted on the subject's head and stool. For the remaining six measurements the Polhemus Liberty magnetic tracker was employed. A sensor was adjusted on the subject's head, that continuously monitored their rotation angle. Subjects adjusted their position based on verbal instructions by the experiment moderators.

The HRTF filters were captured using the scanIR application [18] with 2 different one-second excitation signals -a Maximum Length Sequence (MLS), and Golay codes-, sampled at 48000 Hz,

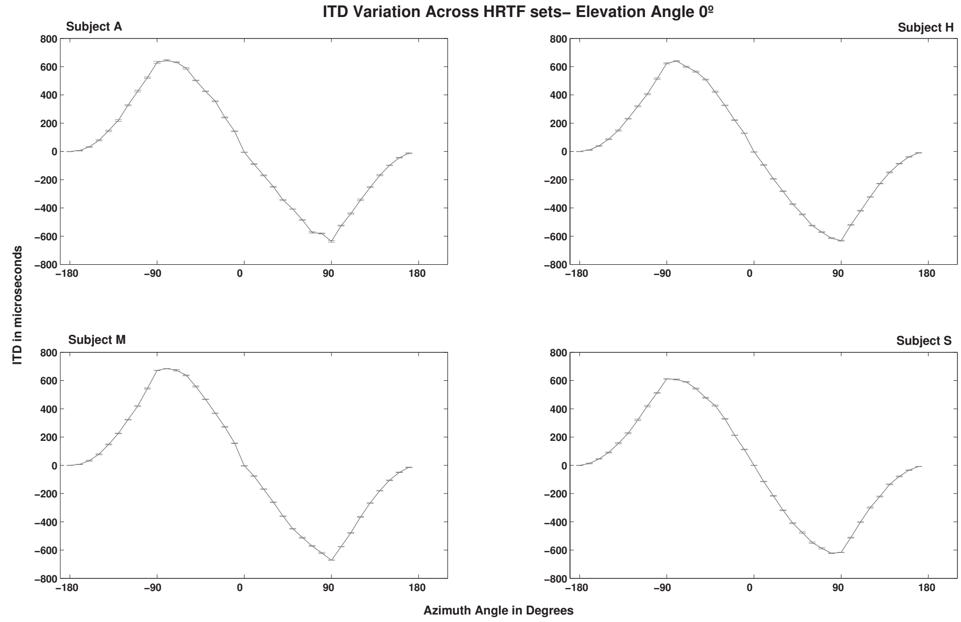

Figure 1: ITD Variation across HRTF sets on the horizontal plane, for all subjects. The curve shows the average ITD values, while the error bars depict the offset range. The x-axis stands for azimuth angles in $+10^\circ$, increments, and the y-axis for ITD offsets in μsecs.

24 bits. A total of 10 different HRTF datasets were measured on each of the 4 subjects who volunteered for this study. During any measurement round each subject had 2 complete datasets captured before having their earplugs and microphones removed, and standing up from the stool. Every measurement round lasted for approximately 35 minutes, and was monitored by one of two moderators.

4.2. The HRTF data

The spatial resolution of the database was designed to be uniform at 10° horizontal and 15° vertical increments from -30° to $+30^\circ$ in elevation, resulting in a total of 180 binaural filters per set. The raw data was reduced to 256-tap HRIRs in order to remove room reflections, and was free-field equalized to compensate for the spectral characteristics of the speaker setup.

The MARL database consists of 40 HRTF sets (10 datasets for each of the 4 subjects) which are available in two different filter types to facilitate various kinds of binaural research; first as a collection of minimum phase filters with the corresponding ITD information, and the second as collection of fixed filters with the ITD data incorporated. The database along with further information can be found at: `http://marl.smusic.nyu.edu/projects/database-of-repeated-hrtf-measurements`.

4.3. File format

All HRTF data is stored according to the marl-nyu file format specifications [19]. For each of the 40 datasets, the measured filters and metadata are organized in two different components: the *data* array of structures and the *specs* structure. *Data* holds all the location-specific information of the measurements, while *specs* holds all the general information of an HRTF set.

More specifically, each element in the *data* array is a structure that encapsulates all the location-specific information, namely: the azimuth and elevation positions, the distance from the sound source to the subject, the left and right ear Head-Related-Impulse-Responses, or Head-Related-Transfer-Functions, and the corresponding ITD value. Every consecutive measurement is stored as a separate structure and is appended to the original array.

All information that is particular to a whole set of measurements (dataset) is stored in *specs*. This information is most likely to remain unchanged throughout the measurement process. The *specs* structure also contains all the dataset identifiers, such as the subject and database names, the type of the HRIR filters etc.

5. ANALYSIS

In this section we offer a preliminary analysis of the collected data, focusing on the spectral content of the filters, and the extracted ITD information. A full report on the variability of the data is beyond the scope of this paper, and is subject to future research.

The HRTF measurement procedure was designed to explore variability in the collected data, while at the same time maintaining a high degree of accuracy in the capturing process. This was achieved by introducing changes in five otherwise controlled conditions, which have been found to affect binaural filters [16]. These factors were: the space where the measurements have taken place, the speaker setup, the removal and readjustment of the miniature microphones between measurements of different datasets, the degree of precision in monitoring the alignment of the subjects, and the involvement of two moderators with different levels of expertise in HRTF measurements.

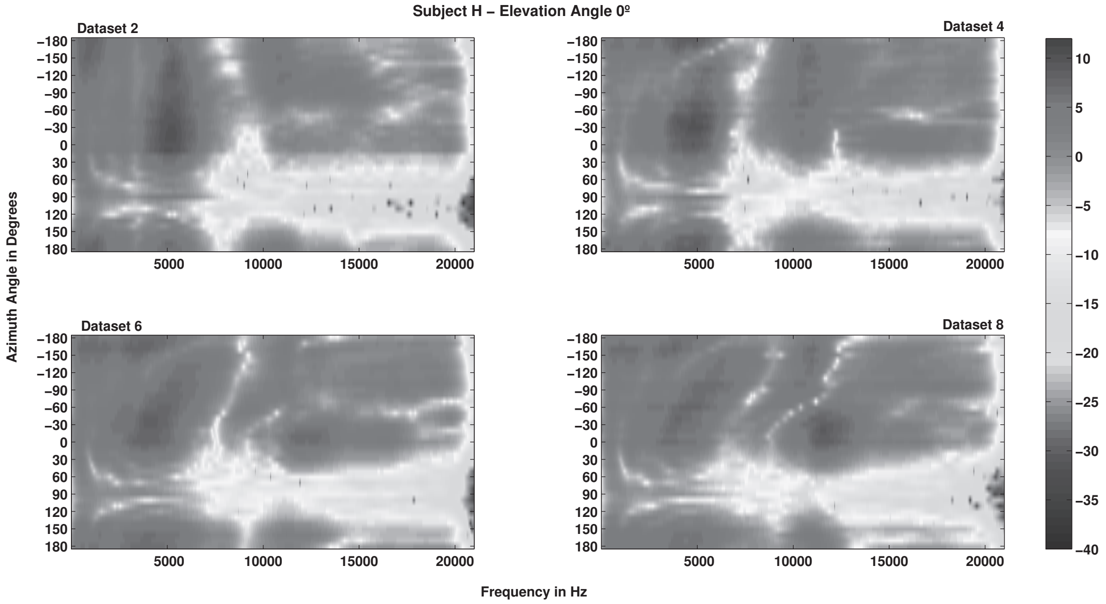

Figure 3: Spectral content variation across 4 HRTF sets of subject H, at elevation angle 0°. The x-axis denotes frequency in Hz and the y-axis azimuth angles. Azimuth angles from $-180°$ to $0°$ contain ipsilateral content filters. The color bar indicates dB magnitude values.

5.1. ITD analysis

The analysis of the ITD content of the database yields some interesting information regarding the repeatability of the measurement process. The maximum offset across the six-HRTF subsets for which the position of the subjects was monitored using a magnetic tracker was no more than 40 μsecs. The remaining four HRTF sets, for which their position was observed using laser pointers was in some instances up to 80 μsecs.

It was also found that the $-30°$ elevation in the 4 datasets that were captured in the Spatial Auditory Research Lab had ITD values that were consistently longer by almost 20 μsecs across all azimuth locations for all 4 subjects, compared to the corresponding measurements from the Dolan Studio live-room. We believe that this is due to the speaker setup for this specific location. Finally, there were a few instances, where larger ITD differences were observed for some HRTF sets. In our analysis we treated those as outliers, and associate their presence with accidental minor movements or misalignment of the subjects during the measurement process.

Figure 1 displays an example of the variation from the average ITD curve across all azimuths at an elevation angle of $0°$ for all 4 subjects of the MARL database. As it can be seen most of the offsets are in the range of 20 - 40 μsecs, while there are some instances that the difference goes up to 63 μsecs. For subject A at azimuth $-90°$ one can also see an example of a longer ITD offset. A comparison of this data with a collection of repeated measurements on a dummy-head mannequin, where misalignment and movement factors are ruled out, can prove very informative in our analysis, and is part of our immediate plans.

5.2. Spectral analysis

For the purposes of this analysis, the length of the 256-tap HRIRs was shortened to 1.5ms, to include only the pinnae responses, and the amplitude of each filter set was normalized to the left-right ear maximum. The responses were reduced to minimum phase, and were low-pass filtered with a cosine window at 20000 Hz to eliminate high frequency content.

Spectral variations between HRTF sets that belong to the same subject are present, as expected, in the mid and high frequency range, and can be the result of the five factors, which, as described earlier, were alternating between measurements. For frequencies between 1000 and 10000 Hz the spectral differences range from 0 dB to 6 dB, with a few instances where they extend up to 10 dB.

For frequencies above 10000 Hz the maximum spectral variation reaches up to 10 dB. Such differences, when consistent across adjacent locations between HRTF sets, can be the result of the placement of the binaural microphones in the subject's ears, which has been found to affect high frequency content [16]. There are a few scattered cases where the magnitude differences of high frequencies are greater than 10 dB, the cause for that needs to be investigated though further in-depth analysis of the data.

Differences are present in both ipsilateral and contralateral locations. Figure 3 offers a comparison of the left ear dB magnitudes across 4 HRTF datasets for subject H, at elevation angle $0°$. The x-axis denotes frequency in Hz and the y-axis changes in azimuth angles, in increments of $10°$. An example of spectral variations can be observed for frequencies around 5000 Hz in azimuths from $0°$ to $-90°$. In the magnitude plots of datasets 2 and 4 there is a peak varying between 8 dB and 10 dB. However, in the plot of dataset 6 the spectral energy at the same positions is reduced by 2 dB, and is almost completely absent in the last plot of dataset 8.

6. DISCUSSION AND FUTURE WORK

This database was created to support exploration of the expected degree of variability in binaural filters, and their ability to fully describe the characteristics of a subject. As mentioned earlier, when one acquires a pre-measured HRTF dataset they have to rely on the accuracy of the reported metadata (such as azimuth and elevation coordinates, distance from the sound source etc.) for a complete description of the binaural filter pairs. Furthermore, the number of post-processing operations that can confidently be applied to binaural filter data, in order to adjust any discrepancies without introducing unwanted artifacts is very limited.

At the same time, the conditions in which HRTF sets are collected are far from ideal. Even in anechoic spaces the presence of the necessary reproduction and monitoring equipment introduces unwanted reflections in the data. Speakers are not omnidirectional point sources and can impose their own directivity patterns to the signal, while the placement of the miniature microphones at the blocked entrance of the subject's ear canals is hard to replicate even for expert experimenters, and can affect high frequency content. Finally, a high degree of precision in the placement of the subject with respect to the sound source, as well as the minimization of accidental head and body motion are key components to the measurement process, as they can introduce non-linear spectral and phase inaccuracies in the data [15].

An analysis of the repeated measurements in the MARL database can offer an understanding of the expected degree of variation in a given HRTF set across all frequency bands. When combined with subjective experiments, it may help with operations like fine tuning binaural spectral cues in order to achieve optimum immersive experiences. Knowledge of the perceptually relevant changes in binaural filters and their impact on spatial impression can be applicable to various research fields and related signal processing tasks, such as: binaural auditory displays, HRTF modeling, database matching, and HRTF interpolation.

Related HRTF research could significantly benefit from a more complete data collection in the number and density of the locations measured, the size of the subject pool, and the number of measurement repetitions in various spaces. Future work includes designing a less logistically-complicated and time-consuming way of acquiring similar data in order to create a longer and more complete data collection for research applications. A full exploration of the collected data, and the design of a set of perceptual tasks that can evaluate any potential findings is also necessary, as part of near future work.

7. REFERENCES

[1] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function," *Journal of the Acoustical Society of America*, vol. 110, no. 5, pp. 2440–2455, 2001.

[2] S. Xu, Z. Li, and G. Salvendy, "Individualization of head-related transfer function for three-dimensional virtual auditory display: a review," in *Proceedings of the 2nd international conference on Virtual reality*. Springer-Verlag, 2007, pp. 397–407.

[3] ——, "Improved method to individualize head-related transfer function using anthropometric measurements," *Acoustical Science and Technology*, vol. 29, no. 6, pp. 388–390, 2008.

[4] S. Hwang and Y. Park, "Interpretations on principal components analysis of head-related impulse responses in the median plane," *Journal of the Acoustical Society of America*, vol. 123, no. 4, pp. EL65–EL71, 2008.

[5] K. J. Faller, A. Barreto, and M. Adjouadi, "Decomposition of Head-Related Transfer Functions into Multiple Damped and Delayed Sinusoidals," in *Advanced Techniques in Computing Sciences and Software Engineering*, K. Elleithy, Ed. Springer Netherlands, 2010, pp. 273–278.

[6] D. Zotkin, J. Hwang, R. Duraiswaini, and L. Davis, "HRTF personalization using anthropometric measurements," in *IEEE Workshop on.Applications of Signal Processing to Audio and Acoustics*, Inst. for Adv. Comput. Studies, Maryland Univ. College Park, MD, USA: IEEE, October 2003, pp. 157–160.

[7] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the 2003 International Conference on Auditory Display*. Boston, MA, USA, 2003, pp. 1–4.

[8] A. Roginska, T. Santoro, and G. Wakefield, "Stimulus-dependent HRTF preference," in *129th Audio Engineering Society Convention*, San Francisco, CA, USA, 2010.

[9] B. Gardner and K. D. Martin, "HRTF Measurements of a KEMAR," *Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, June 1995.

[10] O. Warusfel, "http://recherche.ircam.fr/equipes/salles/listen/," 2003.

[11] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, NY, October 2001, pp. 99–102.

[12] B. Xie, X. Zhong, D. Rao, and Z. Liang, "Head-related transfer function database and its analyses," *Science in China Series G: Physics, Mechanics and Astronomy*, vol. 50, no. 3, pp. 267–280, June 2007.

[13] N. Gupta, A. Barreto, M. Joshi, and J. Agudelo, "HRTF database at FIU DSP lab," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. Dallas, TX: IEEE, March 2010, pp. 169–172.

[14] H. Wierstorf, M. Geier, A. Raake, and S. Spors, "A Free Database of Head-Related Impulse Response Measurements in the Horizontal Plane with Multiple Distances," in *130th Audio Engineering Society Convention*, London, UK, 5 2011.

[15] K. A. J. Riederer, "Repeatability analysis of head-related transfer function measurements," in *105 thAudio Engineering Society Convention*, 9 1998.

[16] ——, "Head-related transfer function measurements, master thesis,," Ph.D. dissertation, Helsinki University of Technology, 1998.

[17] G. Wersényi and A. Illényi, "Differences in dummy-head HRTFs caused by the acoustical environment near the head," *Electronic Journal of Technical Acoustics*, vol. 1, pp. 1–15, 2005. [Online]. Available: http://www.ejta.org

[18] B. Boren and A. Roginska, "Multichannel Impulse Response Measurement in MATLAB," in *131st Audio Engineering Society Convention*, New York, NY, October 2011.

[19] A. Andreopoulou and A. Roginska, "Towards the creation of a standardized HRTF repository," in *131st Audio Engineering Society Convention*, New York, NY, 2011.