ANALYSIS OF THE SPECTRAL VARIATIONS IN 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 discusses the range of spectral variations between HRTF sets measured on the same subjects. The analysis is done in a corpus of 40 HRTF datasets of 4 subjects (10 datasets per subject). Variations are observed as a function of frequency, distance of the ears to the sound source (ipsilateral or contralateral), and location. Assessments of the spatial quality of all datasets were made through a subjective study which confirmed that despite their spectral variations, all individually measured HRTF sets maintained a high degree of spatial realism. An understanding of the variability in HRTFs can offer new intuition on objective binaural filter evaluation, and has significance in the research fields of spatial audio reproduction and virtual auditory display.

1. INTRODUCTION

3D virtual auditory displays are created by applying the spatial cues that are characteristic of each sources intended location on monophonic sounds. Location-depended spatial cues are captured in the Head-Related Impulse Responses (HRIRs) or their frequency domain equivalent Head-Related Transfer Functions (HRTFs).

Due to the high variability in people's head shapes and especially pinnae, HRTFs vary vastly among individuals. The use of non-individualized HRTFs in binaural reproduction may lead to an unconvincing spatial impression, which can compromise the realism, quality, and accuracy of the delivered message. Consequently, various scientific approaches have emerged trying to overcome the need for personalized data, such as HRTF customization [1], [2], [3], database matching [4], [5], and HRTF modeling [6], [7], [8]. Also, in an attempt to create realistically dense auditory environments, much research has been done regarding high-accuracy HRTF interpolation methods [9], [10].

The accuracy of acquired or computed HRTFs can only be evaluated either perceptually through a user study, or objectively based on a defined metric, such as the mean squared error (MSE), the correlation distance, or the signal to distortion ratio (SDR). While in the first case a successful dataset is the one that conveys a convincing spatial image, in the latter it is the one that demonstrates the smallest variation from an originally measured set.

Even though objective evaluation processes can be quick, they are oversimplified as they assume uniformity in the perceptual weights of spectral variation across frequency. Nevertheless, the brain has a certain degree of tolerance in HRTF variations, as studies have shown that the human auditory system has the ability to successfully adapt to altered spectral cues, given time [11]. Hence, a more enhanced evaluation process would take into account prior knowledge of the expected variability within HRTF sets, in order to create perceptually driven thresholds.

Absolute variations in measured datasets have been reported before by Riederer [12], [13], who offered a detailed overview of the degree of error introduced to an HRTF pair 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. However, that work explored the effects of a single error-factor at a time, while in real-life situations multiple can contribute to the final measurement outcome concurrently. In a similar study Wersényi, and Illényi investigated the effect of near-the-head everyday objects like hair, caps, glasses and clothing, on HRTF datasets. The analysis was based on repeated measurements on a computer- controlled dummy head measurement system [14].

The goal for this work is to revisit this topic by discussing the spectral variability of binaural filters, through observation of the MARL database, a corpus of repeated HRTF measurements across multiple subjects. More specifically, this paper presents an assessment of the spectral changes that can be observed between repeated HRTF measurements on the same subject as a function of frequency, location (azimuth / elevation angle), and distance of the ear to the sound source (ipsilateral or contralateral filters).

2. THE DATABASE

2.1. Description of data

The database consists of repeated HRTF measurements on 4 subjects. It has a total of 40 datasets (10 sets per subject), all captured by the Music and Audio Research Laboratory, at New York University. 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. The measurements were reduced to 256-tap HRIRs to remove room reflections, and was free-field equalized to compensate for the spectral characteristics of the speaker setup.

The measurements took place at 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. For both cases 5 different Genelec 8030a speakers were positioned in a spiral configuration. The subjects were seated on a rotating stool with adjustable height, at the center of the spiral, at a distance of $1m$ from the speakers.

All data was measured using the blocked meatus technique with custom-made miniature binaural microphones using Sennheiser's $KE-4$ capsules. For 4 of the HRTF sets the positioning of each subject was monitored with the use of laser pointers, while for the remaining the Polhemus Liberty magnetic tracker was employed. The HRTF filters were captured in scanIR [15] with 2 different one-second excitation signals -a Maximum Length Sequence (MLS), and Golay codes-, sampled at 48000 Hz.

2.2. Variability in the data

The capturing procedure of the collection was designed to explore variability in HRTF sets while maintaining a high degree of accuracy in the measurement process. This was achieved by introducing changes in five otherwise controlled conditions: the space where the measurements took 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 2 moderators with different levels of expertise in the HRTF measurement process. A more detailed description of the database and its acquisition process can be found in [16].

3. SPECTRAL VARIATION IN HRTF SETS

3.1. Data post processing and variation estimation

The database contains 256-tap HRIRs in the form of minimum phase filters with incorporated ITD information. The responses are low-pass filtered with a cosine window at 20000 Hz to eliminate high frequency content. For the purposes of this analysis, the length of the 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, to eliminate the impact of overall amplitude differences between datasets on the intended analysis. All binaural filters were band-limited between 500 Hz and 16000 Hz using a rectangular window. In order to reduce the weight of high frequency content on the results, signals were averaged across 12^{th} octave bands. Ipsilateral and contralateral content data was stored independently, and analyzed separately.

The spectral variation between HRTF sets was calculated per 12^{th} octave band. When the differences between all possible sets, which belonged to the same subject, were calculated, the frequency-band-dependent results were sorted and averaged out, to indicate the average increase in dB magnitude difference across all repeated measurements. Changes in the average spectral variations were then studied as a function of frequency, proximity to the sound source (ipsilateral or contralateral filters), and location.

3.2. Spectral variation as a function of frequency

For this part of the analysis variation in dB magnitude across HRTF sets is studied exclusively as a function of frequency. Changes are observed across binaural filters from all available azimuth elevation locations, and all subjects. Results are reported at the median value, the 75^{th} percentile, at + 2.7 σ and at the maximum value. Figure 1 shows a graphical report of the results.

Below 1.5 kHz the median spectral variations are around 1.5 dB, the 75^{th} percentiles 2.5 to 3.5 dB, and the upper boundaries 5.0 to 7.0 dB. From 1.5 kHz to 6 kHz the medians are between 2.5 dB and 4.0 dB, the 75^{th} percentiles from 3.5 to 6.5 dB, and the upper boundaries from 8.0 to 16.0 dB. For frequencies above 6 kHz there is an even larger increase in spectral variation. Median

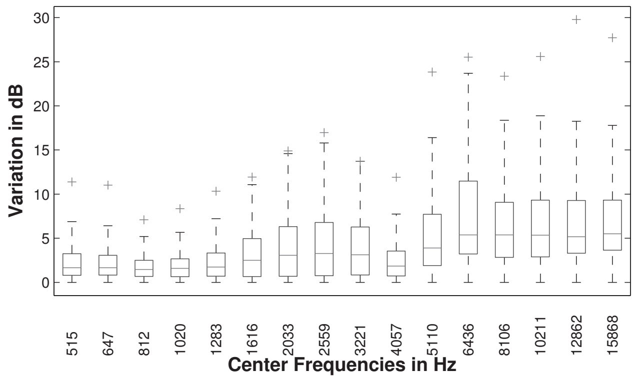

Figure 1: Box plot of the frequency dependent spectral variations. Box edges indicate the 25^{th} and 75^{th} percentiles and the red line marks the median. Variations are also marked at +2.7σ (black stem), and at the maximum value (red cross).

values are around 5.3 dB, 75^{th} percentiles are between 9.0 dB and 11.0 dB, and upper boundaries vary from 18.0 dB to 23.0 dB.

As far as maximum variations are concerned below 5 kHz they range between 7.0 dB and 16.9 dB, while for frequency bands above that they raise up to 30.0 dB. Deviations of up to 20 dB have been reported before [14] for dummy head measurements conducted in conditions where unwanted movements, body alignment, and microphone placements were well controlled. The increase of up to 10 dB in our results can be expected since data is collected on human subjects.

3.3. Spectral variation as a function of distance from the ears to the sound source

In this part spectral variations are analyzed separately for ipsilateral and contralateral filters across frequency. Results are presented at the same distribution points as above (Figure 2).

In the ipsilateral ear filter-banks, for frequencies below 1.5 kHz the median variations are around 1.0 dB, the 75^{th} percentiles around 2.0 dB, and the upper boundaries from 3.0 to 5.0 dB. Between 1.5 kHz and 6 kHz the medians vary between 1.5 dB and 2.9 dB, 75^{th} percentiles between 3.5 dB and 5.0 dB, and the upper boundaries between 5.0 dB and 11.0 dB. For frequencies above 6 kHz, median values are around 3.5 dB, 75^{th} percentiles are between 6.0 dB and 8.0 dB, and upper boundaries vary from 11.5 dB to 17.0 dB. Looking at the maximum variations, octave bands below 5 kHz reach maximum values between 5.0 and 12.75 dB, while high frequency content between 16.0 dB and 23.0 dB.

A comparison of the just reported ipsilateral content results to the contralateral spectral variations yields striking similarities between to two binaural filter groups. Both demonstrate almost identical variation patterns, with the contralateral ones being higher up to 0.5 dB for the vast majority of the times. Yet, this is not the case for the maximum variations. Especially for frequency bands above 5 kHz differences between the two groups increase by up to approximately 4.0 dB. However, given that the magnitude of contralateral filters is lower that that of ipsilateral, and the noise floor is significantly higher, it is questionable whether contralateral spectral variations can be equally perceived, and are of the same importance to human listeners.

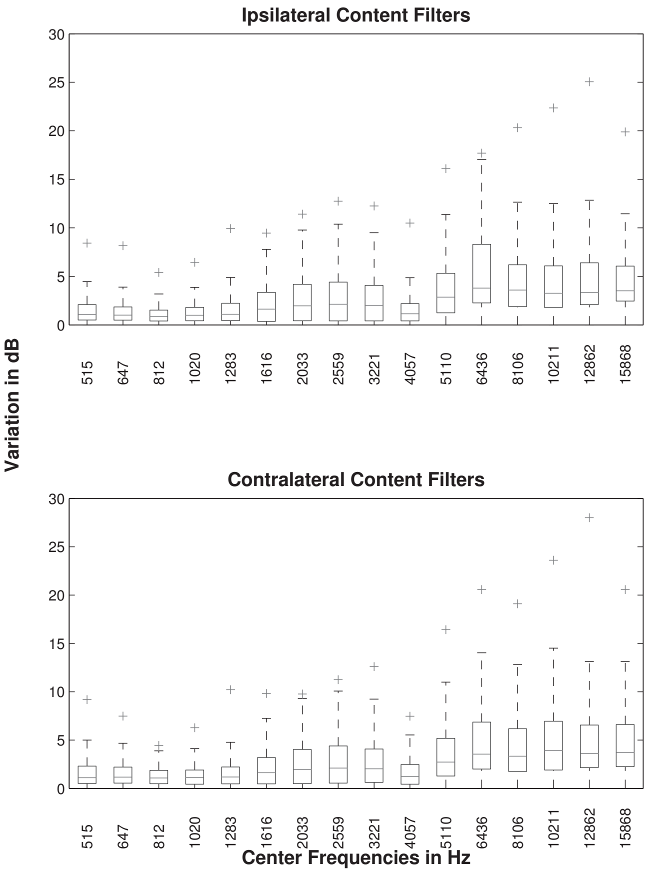

Figure 2: Box plots of the frequency dependent spectral variations across ipsilateral and contralateral HRTFs. Box edges indicate the 25^{th} and 75^{th} percentiles and the red line marks the median. Variations are also marked at +2.7σ (black stem), and at the maximum value (red cross).

3.4. Spectral variation as a function of location

In the last part of the analysis spectral variations are observed as a function of location. For every azimuth - elevation pair, average distribution curves are computed in the dB magnitude across all 12^{th} octave bands. Examples of the combined results for a single location can be viewed in Figure 4, where each of the subplots demonstrates the degree of change (from minimum to maximum) in the spectral variations across frequencies.

Results indicate that there is a relationship between the rate of increase in spectral variation and the elevation angle, for frequencies between 1 kHz and 5 kHz. In this frequency range high-elevation distribution vectors are relatively flat, and their maximum values stay within the boundaries of the 75^{th} percentiles, as reported above for ipsilateral content filters. However, as elevations decrease variations become less flat and the increase more drastic, reaching at $-30°$elevations the previously reported upper boundary rates of spectral variation rates. Similar observations have been reported by Wersényi in his analysis of repeated HRTF measurements on dummy head mannequins [17]. Results have shown that as the elevation of a sound source increases spectral variations become flatter, and re-measurement accuracy improves significantly, especially when the source moves outside of the head shadow (above $30°$ in elevation).

An example of this behavior from the MARL databse can be viewed in the right column of Figure 4. The top three plots depict the rate of change in spectral variation across frequency, for azimuth angle $90°$, at elevations $30°$, $0°$, and $-30°$. For frequencies from 1 kHz to 5 kHz one can see how the rate of increase in the spectral variation distribution changes as a function of elevation, being very low for high elevations and higher for upper elevations.

Another observation is that, irrespectively of elevation, frontal plane azimuth angles yield higher spectral variations across frequency content above 5 kHz, than interaural or rear plane ones. This can be seen by comparing the plots in the the left to those in the right column of Figure 4. The left column contains azimuth location $0°$, at elevations $30°$, $0°$, and $-30°$, and azimuth location $50°$ on the horizontal plane, while the right column contains azimuth locations $90°$, at elevations $30°$, $0°$, and $-30°$, and $130°$ on the horizontal plane. For example, by looking at frequencies above 5 kHz in just the bottom two plots, one can see a difference in the maximum variation between 4.0 dB and 8.0 dB.

4. SUBJECTIVE EVALUATION OF THE HRTF DATABASE

4.1. Experiment overview

The purpose of this experiment was to perceptually assess the spatial quality of the repeated measurements given three criteria: externalization perception, front / back discrimination, and up / down discrimination. This was necessary in order to verify whether, in spite of their spectral and ITD variations, individually measured HRTF sets maintained spatial accuracy. A similar experimental design was first introduced by Roginska et al. in [5].

A bank to 10 HRTF datasets was created for each of the 4 subjects in the study, which consisted of their personally measured sets. All binaural filters were reduced to 128-tap minimum-phase responses with their corresponding interaural time delays. The test signals used were 0.5 second pink noise bursts. An overview of the experimental procedure follows.

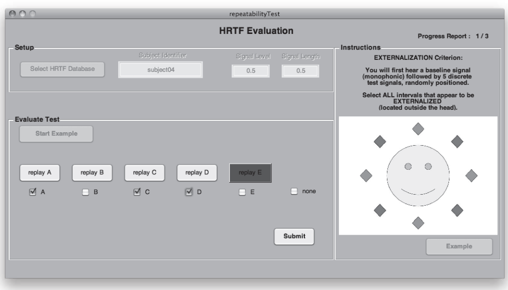

Figure 3: The MATLAB® interface used in the study

4.2. Experimental procedure

The experiment was conducted in the Spatial Auditory Research Lab at New York University. Stimuli were presented to all subjects through the Sennheiser HD650 open headphones. The graphical interphase used for playback and collection of the user responses was designed in MATLAB®2010 $_b$. (Figure 3).

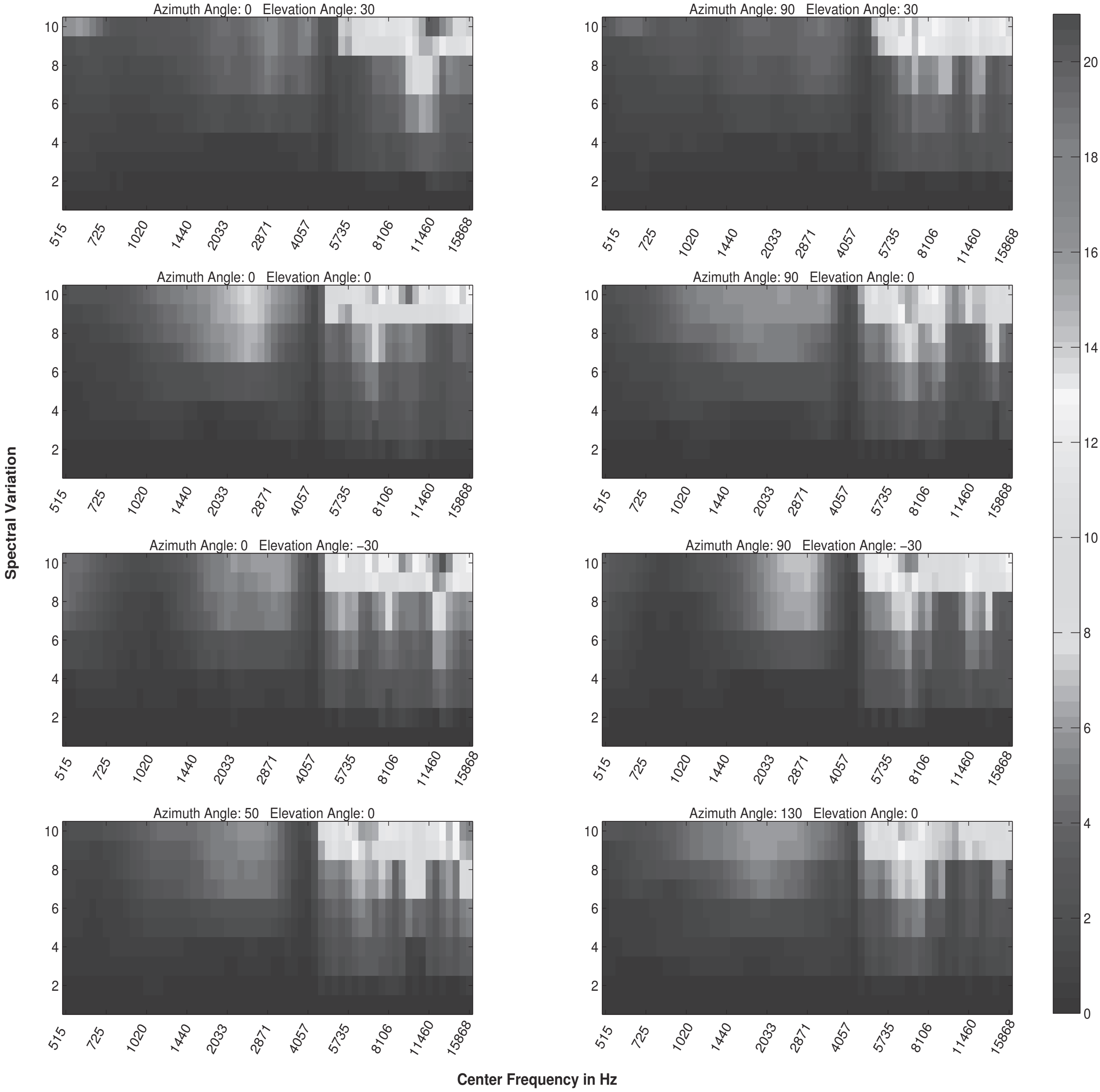

Figure 4: Rate of change in spectral variations as a function of location for frequencies 500 - 16000 Hz. The x-axis denotes center frequencies, the y-axis the indexes of the 10-point distribution vectors, and the color-coding the magnitudes of the spectral variations in dB. 8 different locations are depicted in the graph, azimuth angles $0°$ and $90°$, across elevations $\pm30°$ and $0°$, and azimuth angles $40°$ and $130°$ on the horizontal plane.

The procedure consisted of three stages designed to test the spatial quality of the involved HRTFs, rather than the perceived localization accuracy. In each of the stages the trials were created by randomly selected datasets convolved with the test signal. For each trial subjects were presented with 5 different intervals (A - E) and were instructed to select all that met the stage-specific criterion. HRTF sets were presented 5 times in each of the three experimental stages. Stages 2 and 3 (front / back and up / down discrimination) only used binaural filters that were selected at least 60% of the time in the externalization task.

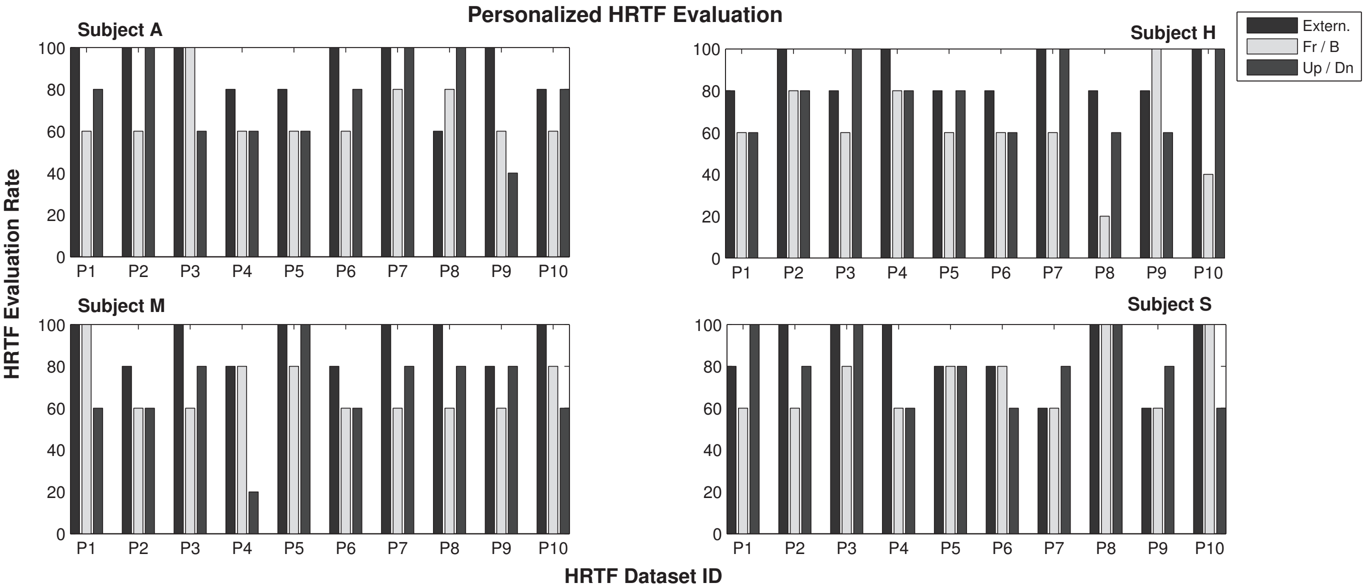

Figure 5: Evaulation of the individually measured HRTF sets of the MARL database, for each criterion separately, across all 4 subjects. Bar colors represent the 3 criteria of the study, and identifiers P1 - P10 the 10 individually measured HRTFs for each subject.

More specifically, stage 1 of the experiment tested the ability of each HRTF set to convey externalized images. Each interval consisted of a monophonic reference signal, followed by a series of 5 signals spatialized at various locations on 5 elevations from $-30°$ to $+30°$. The azimuth locations used were $\pm30°$, $\pm60°$, $\pm90°$, $\pm120°$, and $\pm150°$.

In stage 2 subjects were asked to select intervals for which they could discriminate front from back. Each interval consisted of a monophonic reference signal, followed by 3 pairs of signals, alternating on the horizontal plane on azimuths locations at opposite sides of the cone of confusion. The utilized azimuthal angles were $0°$, $\pm30°$, $\pm60°$, $\pm120°$, $\pm150°$, and $180°$,

Stage 3 of the experiment presented subjects with an up / down discrimination task. Stimuli consisted of a monophonic reference signal, followed by 3 pairs of signals alternating between $\pm30°$ elevations. Users were instructed to select all intervals for which they perceived a change in elevation. The presented azimuth locations were $0°$, $\pm30°$, $\pm60°$, and $\pm90°$.

4.3. Results

Figure 5 presents the subjects' assessments of their personalized sets, for every one of the three experimental criteria. In terms of externalization perception, all HRTFs were rated very highly in their ability to create convincing auditory images outside of the subjects' heads, receiving a selection rate of 80 - 100%.

These rates, however, dropped by 20 - 40% for the front / back discrimination case. By looking at the graph, we see that front / back discrimination was the weakest rated task across all subjects. This behavior was rather expected, as the experiment consisted only of static (not following head movement) auditory cues, which are not strong enough to contradict visual components reporting no potential sound sources in front of the subjects. However, despite this drop, all filter sets remained at or above the 60% evaluation margin, with the exception of measurement 8 in subject H which received 20% preference only on the given task.

Individually measured HRTFs received a 20% increase in the ratings of the up / down discrimination task, compared to the front / back stage. However, these scores were still most of the times lower than in externalization criterion case. This behavior was unanimous across subjects and for all datasets, with very few exceptions, like HRTF 4 in subject M, or HRTF 10 in subject S.

In general, all subjects indicated a strong preference towards individually measured HRTFs, by selecting them at least 60% of the time in most cases. All binaural filters were consistently evaluated above chance as successful means for realistic sound spatialization, regardless of their potential variations. Such findings support the known ability of the human auditory system to accept / disregard variations in HRTF datasets within certain boundaries.

5. DISCUSSION AND CONCLUSIONS

This study investigated the range of spectral variation between HRTF datasets originating from the same subject, across 12^{th} octave bands from 500 to 16000 Hz. Variations, observed as a function of frequency, distance of the ears to the sound source, and location, are reported at the median values, the 75^{th} percentile, and at + 2.7 standard deviations from the mean. Maximum values are also provided.

It is an extension of Riederer's [12] work on the repeatability of HRTF measurements. That research presented a case for the most accurate method of capturing binaural filters, by identifying the impact of different controlled conditions on the measurement outcome, and the appropriate experimental adjustments that could minimize it. The significance of modifying each of the control conditions, quantified as the change in binaural filters, was analyzed separately. However in this work the impact of all controlled conditions is studied collectively, as in regular HRTF measurement routines more than one of them may occur at a time, affecting the data.

The analysis is based on a collection of 40 HRTF datasets on 4 subjects (10 sets per subject), on a 10° azimuthal grid, across 5 elevations from -30° to $+30^\circ$. It is expected that the more complete data collection of this work will give a broader overview of the variability in the HRTF spectrum across a wider range of locations, given a greater number of repeated measurements.

Observations of inter-HRTF variations across frequency, ear (ipsilateral or contralateral), and azimuth-elevation location, can be used to create more dynamic ways of objectively evaluating the quality of non acoustically measured binaural filters. For example, they can be applied to the definition of thresholds for the distance functions used to evaluate HRTFs. Even though such thresholding will still be not cognitively driven, we believe that it will offer a better approximation of the perceptual HRTF evaluation process, than the normally used metrics, such as the mean squared error (MSE), which assume uniformity in the perceptual weights of spectral variations across frequency.

Knowledge of the expected range of variation in measured HRTF sets can offer new intuition on objective binaural filter evaluation. Several spatial audio related research fields, such as HRTF customization and HRTF modeling, which currently primarily rely on perceptual evaluation tasks, could benefit from that, by gaining a more robust method of assessing the quality of their designs. This research can also lead to more effective metrics for computing HRTF similarity, which can be directly applied to database matching tasks. Finally, assessments of the accuracy of different HRTF interpolation algorithms will become more effective as filter evaluations will be done according to perceptual boundaries, rather than strict mathematical models.

Future work will focus on studying the database distributions. This analysis will lead to a new representation of HRTFs that will favor discrimination between sets that belong to dissimilar people, while at the same time promoting groupings of more similar ones in the same class. Such a representation will have to be supported by further subjective evaluation studies, which will assure that the computed models do not contradict cognitive assessments.

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

[3] 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.

[4] 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.

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

[6] E. A. Durant, S. Member, and G. H. Wakefield, “Efficient Model Fitting Using a Genetic Algorithm : Pole-Zero Approximations of HRTFs,” *Audio*, vol. 10, no. 1, pp. 18–27, 2002.

[7] A. Kulkarni and H. S. Colburn, “Infinite-impulse-response models of the head-related transfer function,” *The Journal of the Acoustical Society of America*, vol. 115, no. 4, pp. 1714 – 1728, 2004.

[8] H. Hu, L. Chen, and Z. Y. Wu, “The Estimation of Personalized HRTFs in Individual VAS,” in *4th International Conference on Natural Computation*, vol. 1. Ieee, 2008, pp. 203–207.

[9] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, “Interpolation of head related transfer functions considering acoustics,” *118th Audio Engineering Society Convention*, 2005.

[10] Z. Xiao-Li and X. Bo-Sun, “Spatial characteristics of head-related transfer function,” *Chinese Physics Letters*, vol. 22, no. 5, pp. 1166 – 1169, 2005.

[11] P. M. Hofman, J. G. Van Riswick, and A. J. Van Opstal, “Relearning sound localization with new ears.” *Nature neuroscience*, vol. 1, no. 5, pp. 417–421, September 1998.

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

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

[14] 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

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

[16] A. Andreopoulou, A. Roginska, and H. Mohanraj, “A Database of Repeated Head-Related Transfer Function Measurements,” in *19th International Conference on Auditory Display (ICAD 2013)*, Lodz, Poland, July 6-10 2013.

[17] G. Wersényi, “Directional Properties of the Dummy-Head in Measurement Techniques based on Binaural Evaluation,” *Journal of Engineering, Computing & Architecture*, vol. 1, no. 2, pp. 1–15, 2007.