Round Robin Comparison of Inter-Laboratory HRTF Measurements – Assessment with an auditory model for elevation

Roberto Barumerli*
Department of Information
Engineering
University of Padova

Michele Geronazzo†
Dept. of Architecture,
Design, and Media
Technology
Aalborg University

Federico Avanzini‡
Dept. of Computer Science
Univeristy of Milano

Abstract

Repeatability of head-related transfer function (HRTF) measurements is a critical issue in intra- and inter- laboratory setups. In this paper, simulated perceptual variabilities of HRTFs are computed as an attempt to understand if different acquisition methods achieve similar results in terms of psychoacoustic features. We consider 12 HRTF independent measurement sets of a Neumann KU-100 dummy head from the international round-robin study Club Fritz. Our analysis of HRTF variabilities focuses on localization performance in elevation within the mid-sagittal plane. A round robin evaluation is performed by means of an auditory model which is able to predict elevation errors and front-back confusion for a given pair of target and template HRTF sets. Results report comparable localization performances between four HRTF databases, suggesting that these acquisition methods led to similar performances in providing elevation cues. Such findings further emphasize the intrinsic complexity and the sensitivity of the HRTF measurement process. The final aim of this study is to certify the quality and repeatability of a measurement process at perceptual level; this findings could be extended to the acquisition of human head acoustics.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Spatial hearing defines the perceptual ability to localize sound sources in space. In particular, mammals – and thus humans – continuously analyse the acoustic scene retrieving and monitoring surrounding source positions. This process is performed based on the two-channel sound stream which is filtered by subject physicality: sound waves diffract and interact with the torso, head and external ears, causing listener-dependent temporal and spectral transformations [6]. The resulting effects provide meaningful cues about sound source locations in an egocentric view [20]. Binaural information heavily influence azimuth and lateral localization that is evaluated mostly by means of *interaural time difference* (ITD), and *interaural level difference* (ILD).

On the other hand, spectral cues are primary cues for elevation perception, and *head related transfer function* (HRTF) contains such relevant information; HRTF measurements summarize the direction-dependent acoustic filtering of a free-field point source due to the head, torso, and pinna [7]. Knowledge of such a complex process is needed and used to develop accurate and realistic artificial sound spatialization in immersive virtual and augmented reality scenarios [5].

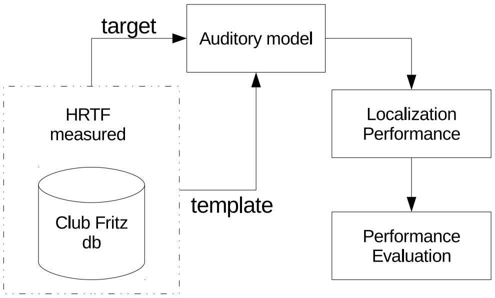


Figure 1: Schematic view of the proposed virtual experiment

HRTFs are usually measured over a discrete spatial grid with discrete frequency samples in anechoic chambers requiring a logistical complexity and delicate acquisition sessions.A post-processing step is then applied to these measurements obtaining space-dependent digital filters used for binaural rendering [14]. Due to the intrinsic complexity of the system setup, several studies have been conducted to understand the repeatability of HRTF measurements and have shown that different sources of variability are occurring on the acquired quantities [2, 11, 21].
Inter-laboratory differences is here investigate further, building on the work of Andreopoulou *et al.* [1]. Specifically, their approach compared measurements on a commercial dummy head: listener geometry and microphone type remained constant while the acquisition systems was laboratory dependent. The collected dataset, known as *ClubFritz*, resulted from an open project developed by Katz and Begault since 2004 [11]: HRTF measurements were performed on the same dummy head, and the resulting dataset is publicly available.

Our reference study [1] compared HRTF acquisitions showing variability in extracted ITD, spectral variations across commonly measured locations, and left/right spectral symmetry between each laboratory. Since the HRTF measurements are used by mean of spatialization, our work aims to evaluate if the underlined inter-laboratory numerical variations led to perceptually equivalent localization cues. More in detail the analysis quantifies statistically the elevation error in the mid-sagittal plane. This is a meaningful and relevant assessment, since reliable elevation cues are difficult to provide considering that HRTFs are highly influenced by anthropometry [13, 15, 16]. Moreover in the mid-sagittal plane, ITD and ILD are ambiguous [12]. In this paper, we employ the auditory model developed by Baumgartner *et al.* [4] in order to quantify the inter-laboratory variability. This auditory model simulates localization experiments following a *template-based* comparison, having as outcome the perceived elevation performances for a given spatialized sound stimuli in static conditions [15]. The resulting performances are affected by the sum of two main errors: the intrinsic perception uncertainty on elevation [16], and the extrinsic error between two different HRTF sets. As showed in Figure 1 this work used the model assigning, as target and template, two measurement sets coming

*e-mail: barumerli@dei.unipd.it
†e-mail: mge@create.aau.dk
‡e-mail: federico.avanzini@di.unimi.it

2018 IEEE 4th VR Workshop on Sonic Interactions for
Virtual Environments (SIVE)
18 March 2018, Reutlingen, Germany
978-1-5386-5713-3/18/$31.00 ©2018 IEEE

Index	Institution	Country	Year	Positions	f_s [Hz]	loudspeakers	Dist. [m]
1	Institut de Recherche et Coordination Acoustique/Musique (IRCAM)	France	2004	2016	44100	1	1.95
2	University of Maryland (UMD)	USA	2007	823	44100	6	0.90
3	Advanced Controls and Displays Group, NASA Ames Research Center	USA	2007	432	96000	12	0.90
4	Institut de Recherche et Coordination Acoustique/Musique (IRCAM)	France	2007	1944	44100	3	2.00
5	Institute of Technical Acoustics, RWTH Aachen	Germany	2009	2016	44100	1	2.00
6	Multisensory Cognition and Computation Laboratory, NICT	Japan	2009	1008	48000	1	1.00
7	RIEC, Tohoku University	Japan	2010	648	48000	35	1.50
8	Takeda Laboratory, Nagoya University	Japan	2009	2017	48000	1	1.00
9	Acoustics Research Institute, Austrian Academy of Sciences	Austria	2014	1550	48000	22	1.20

Table 1: Overview of each contributor to the open project "Club Fritz". The index of the first column will be use to indicate the each laboratory, hereafter.

from different laboratories. Finally, distributions of the localization predictions from the virtual experiments were statistically compared, by means of a paired Student t-test, in order to understand if two inter-laboratory HRTF sets are perceptively consistent.[1]

The main contributions of this work are the following: (i) by using an auditory model to address the perceptual relevance of HRTF sets, our procedure allows a systematic analysis without performing data acquisition with real subjects; (ii) our approach can be used to evaluate the quality and the repeatability of laboratory measurement protocols.

2 Materials and Methods

2.1 Dataset

A comparison between different measurement protocols is performed using the dataset released by the open project "Club Fritz" [1], which consists of twelve independent measurements of HRTF sets on the same commercial dummy head, and microphone system: "Fritz II" model KU-100 made by Neumann GmbH. The dataset is available in the open standard SOFA file format for HRTFs [17] from a public repository [2]. Each laboratory taking part on this project was required to acquire the HRTF set using their internal measurement protocol. Table 1 summarizes some significant differences between each protocol.

The comparison was conducted on the largest common set of available measurements for elevation points in the mid-sagittal plane, in order to avoid interpolation errors. Accordingly, the analysed dataset had nine HRTF sets with an elevation range of $[-30°, 70°] \cup [110°, 210°]$ with a 10°-step. The complementary interval, between $[110°, 210°]$ with a 10°-step, was used to compute the front-back confusion. The intersection between datasets allowed a uniform grid for the analysis. Other differences, such as the distance of the sound source or the sampling rate, had marginal influence on the predictions, because the adopted auditory model smoothed out such related issues (see the next session for details).

2.2 Model

The tool for the inter-dataset perceptual evaluation is based on the auditory model for sound localization in sagittal planes proposed by Baumgartner in [4]. This model is implemented into the *Auditory Modeling Toolbox* as `baumgartner2013`[3]. In particular, the model simulates virtual experiments providing a perceptual metric on localization for *stationary broadband* auditory stimulus and an internal template. The model was validated by comparing the algorithm results with a real subject set which had performed the same localization task simulated into the model [16].

The perceptual metric, introduced by Langendijk [15], compares the *target* sound, processed to obtain an internal representation, with an internal *template*, resulting in a probabilistic prediction of polar angle responses. The template is assumed to be created by means of learning the correspondence between the spectral features and the direction of an acoustic event based on feedback acquired by training [13]. In this work, we use the model predictions in order to understand if the amount of elevation error is comparable by switching a pair of HRTF sets from target to template and vice-versa. A graphical structure representation of the experiment is depicted in Fig. 1.

2.2.1 Internal Representation

The adopted auditory model computed perceptual metrics on a difference between target and template internal representation [4]. Before processing the HRTF sets, the relative head related impulse response, HRIR, were re-sampled at $48kHz$ in order to compute the comparisons. For both template and target the first elaboration step has consisted in converting the HRTF model into a *directional transfer function* (DTF) [15]. Using the DTF instead of the HRTF set increased the robustness to the environment variability between different laboratory setups [18]. The DTFs were subsequently filtered with a gammatone filter bank with a frequency spacing of one equivalent rectangular bandwidth, ERB. Each frequency band was processed with a half-wave rectifier and a low-pass filter in order to simulate the inner hair cells. Finally each band was averaged in time by the mean of root-mean-square (RMS) amplitude resulting in a internal representation of the sound [4, 16].

2.2.2 Distance metric

In this model, the main distance metric is represented by the *spectral standard deviation*, SSD, of the inter-spectral differences between the internal representation for each combination of the target and the template elevation angle. A probabilistic approach was introduced for mapping the distance metric to the predicted response probability: for each target angle, template angle, and ear, the SSD was translated into a *similarity indices* (SI) using a Gaussian function: the SI represented, in degrees, the response probability for the response angle. The Gaussian function, which had zero mean and the standard called *uncertainty* (U), modelled the loss of precision due to perceptual process of the listener [3].

Furthermore, the contribution of both ears was taken into account, by a binaural weighted sum, obtaining a binaural SI. Finally, for each target angle, the binaural SI was computed for each template angle, then these were normalized to one in order to be interpreted as a probability mass vector (PMV) describing the listener's response probability as a function of the response angle for a given incoming sound.

2.3 Perceptual metric

In order to evaluate the perceptual error, the metrics considered the difference between the target angles and the response angles, leading us to define two metrics which were firstly introduced by Middlebrooks in [19] and further formalized in [10]: *Polar Error* (PE) and *Quadrant Error* (QE) averaged for the same template angle

[1] A similar methodology for HRTF evaluation was also proposed in our previous works (see [8–10] for further insights on this approach).

[2] http://sofacoustics.org/data/database/clubfritz/

[3] http://amtoolbox.sourceforge.net/models.php

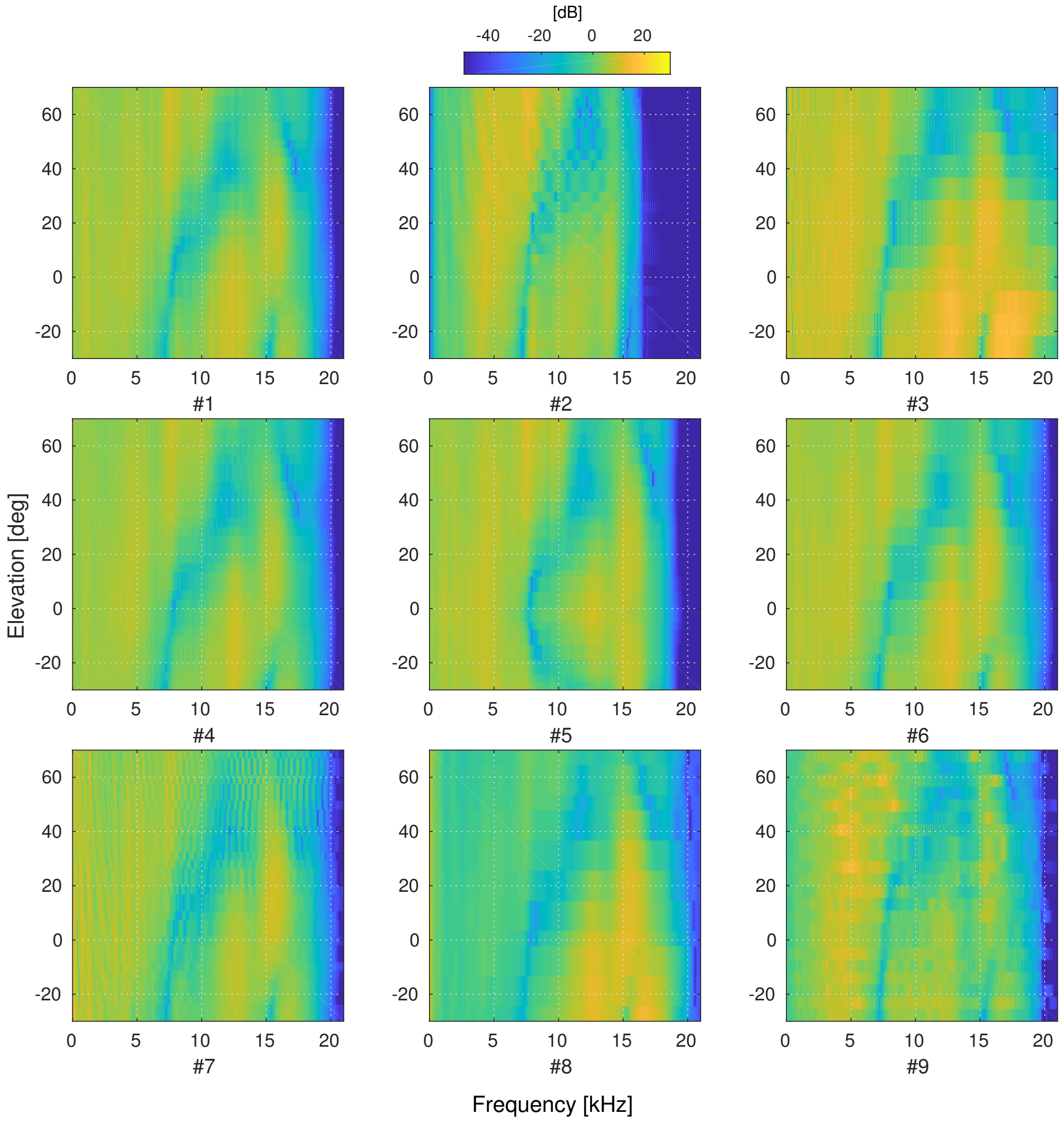


Figure 2: HRTF magnitudes vs. frequency for each analysed dataset. The id number refers to Table 1. Normalization was computed for each HRTF set based on the intra-subject maximum.

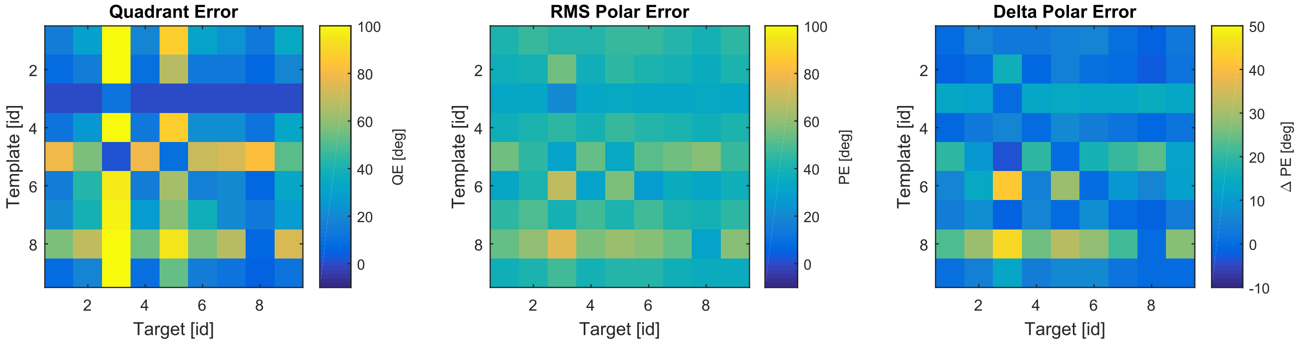


Figure 3: Resulting matrices, from left to right: Quadrant Error, Polar Error and Delta Polar Error.

on all available target angles. The PE is defined for every j-th elevation response close to the target position:

$$PE_j = \sqrt{\frac{\sum_{i \in L} (\phi_i - \phi_j)^2 p_j[\phi_j]}{\sum_{i \in L} p_j[\phi_i]}} \quad (1)$$

with $L = \{i \in N : 1 \leq i \leq N_\phi, |\phi_i - \phi_j| \, mod \, 180^\circ < 90^\circ\}$, where ϕ_i, ϕ_j represent the local response and the target position respectively and $p_j[\phi_i]$ denotes the probability mass vector.

Instead the QE error is formalized as:

$$QE_j = \sum_{i \in NL} p_j[\phi_i] \quad (2)$$

with $NL = \{i \in N : 1 \leq i \leq N_\phi, |\phi_i - \phi_j| \mod 180^\circ \geq 90^\circ\}$, and for the j-th elevation response.

The PE metric accounts for localization judgements occurring within a local range, i.e. L, of the response angle, thus being an estimate for precision. The QE metric accounts all the performance affected by the front-back error and the responses where the absolute polar error exceeded 90°. Finally, the metrics were averaged among angles in the evaluated interval for each comparison between datasets.

An additional metric was derived from the polar error, named as *Delta Polar Error*, ΔPE, which is defined as follow:

$$\Delta PE_{ij} = PE_{ij} - PE_{ii} \; with \; i \neq j$$

where i and j were the HRTF sets with index indicated in Table 1. This metric were used to scale individual performances resulted from introducing uncertainty that simulates individual localization abilities for a virtual listener. Uncertainty was fixed at $U = 2$ which is an average value providing a good fit with experimental data from real listeners [16].

An extra step was performed to evaluate the matching between measurement protocols: distributions of polar error of inter-HRTF set and intra-HRTF set was compared with a *paired t-test*. The assumed *null hypothesis* was: two different HRTF sets led to a comparable perception error. The α-value was set to 0.001.

Finally a binary diversity matrix, D, was built where the ones correspond to a p-value above the threshold, zero otherwise. Two HRTF sets were considered comparable from a perception point of view, if $D(i,j) = D(j,i)$. Then the clusters of perception's comparable measurement protocols were consequently built. In the following, we define the rules that were used to create these clusters:

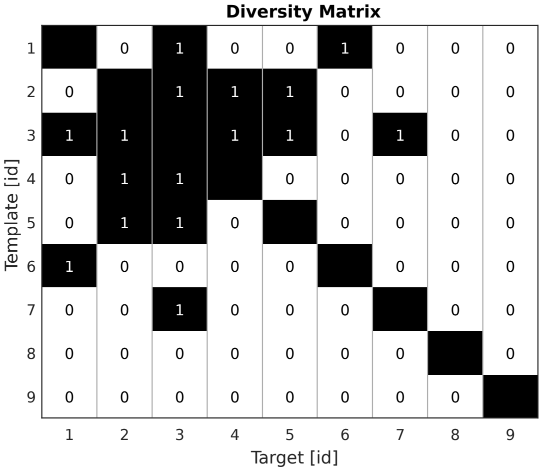
1. unity cluster $C = (i,j) \Leftrightarrow D_{(i,j)} = 1 \wedge D_{(j,i)} = 1$
2. $i \in C \Leftrightarrow \forall j \in C \,|\, (D_{(i,j)} = 1 \wedge D_{(j,i)} = 1)$

3 Results and Discussion

An overview of the nine HRTF magnitudes is displayed in Fig. 2. It is worthwhile to notice that there were common spectral features between each HRTF sets. However, several differences could be extracted after a simple visual inspection. For instance example: set #2 does not held any frequency above $16kHz$, set #5 has a different trend from the others near the point $(10kHz, -20^\circ)$, set #7 exhibits a diffraction-like pattern in the whole spectrum, set #8 demonstrates higher values in the frequency range below $10kHz$ and set #9 reports more discontinuities between adjacent spatial locations.

The analysis was performed on these datasets and Fig. 3 depicts the resulted metrics.

Local errors PE did not exhibit any noticeable patterns hence the ΔPE was evaluated instead.



Template [id] \ Target [id]	1	2	3	4	5	6	7	8	9
1		0	1	0	0	1	0	0	0
2	0		1	1	1	0	0	0	0
3	1	1		1	1	0	1	0	0
4	0	1	1		0	0	0	0	0
5	0	1	1	0		0	0	0	0
6	1	0	0	0	0		0	0	0
7	0	0	1	0	0	0		0	0
8	0	0	0	0	0	0	0		0
9	0	0	0	0	0	0	0	0	

Figure 4: Diversity matrix. White pixels show statistically significant diversity between HRTF sets.

Confusion errors, mapped by QE, showed a wide variability between HRTF sets. Set #3 had QE values almost near 0° when used as template while it returned an error above 90° when selected as target; the eighth set showed a complementary behaviour. Set #5 recorded the most high error in both cases, as template and as target, with an average of $\overline{QE}_5 = 62.30^\circ$.

The statistics on ΔPE revealed an inter-dataset overall average of 8.54° with st.d. of 10.46°. Here, the trends reported that set #8, as template, had the largest inter-dataset averaged error, with 25.44°, but, as target, it reached the lower bound with 3.91°; this outcome followed the same trend which was previously reported for QE.

Figure 4 shows the diversity matrix from which not-significant diversity clusters were created. The resulted clusters are listed in Table 2. The final evaluation reported that there were two clusters with three HRTF sets each. Considering the assumption that each laboratory performed their measurements independently, a low probability of systematic errors in HRTF sets belonging to these clusters could be attested; at least these laboratories were incurring into the same acquisition errors. On the other hand, the remaining clusters were too small to infer relevant conclusions: they show a bias in the results on set #3, probably because it returned lowest errors when considered as a template. Further investigations will be required in order to understand mechanisms of the auditory processing in the adopted model and HRTFs (e.g. sampling frequency, contributions in critical bands, etc.). Remaining sets #8 and #9 did not demonstrate any statistically significant evidence: they were not included in any cluster; this might be related to presence of introduced artefacts in the measurement process as already pointed in Fig. 2.

Length	Clusters
3	(2,3,4), (2,3,5)
1	(1,3), (1,6), (3,7)

Table 2: Similarity clusters

The two resulting major clusters suggested that the perceptual uncertainty, which is exhibited by a realistic subject, could return in a benefit to overcome intrinsic differences in the acquisition protocols. It is worthwhile to notice that set #2 belonged to both clusters, even if it has no information above 16 kHz; this result further confirmed that this frequency range is not relevant for vertical localization [10], and thus our evaluation on HRTF measurements might be bounded to a restricted frequency range.

Protocols which were not included into the biggest clusters were not considered similar by means of our method. Interestingly, these HRTF acquisition processes could lead to statistically different localization performances and to a degradation of localization cues in virtual audio environments.

4 Conclusions

In this paper, we proposed a method to analyse the repeatability of HRTF measurement protocols by means of a statistical metric built on elevation perception. Furthermore, this work further supports the criticality of measurement variations in HRTF sets acquired with different protocols. The study made use of an auditory model based on template-based comparison.

The evaluation of the *Club Fritz* database was made possible thanks to the recently introduced SOFA format, which allowed a smooth implementation into the Auditory Modelling Toolbox.

Results remarked that repeatability of the acquisitions on the same subject are difficult to achieve due to the high sensibility of HRTF measurements even while considering the perceptual domain. The proposed statistical method underlined that perceptive uncertainty can mitigate the intra-subject measurement errors in some cases: in fact some HRTF sets resulted comparable.

Further studies are required to determine the variability on perceptual uncertainty and the reliability of an auditory model for such kind of research methodology. Finally, it has to be stressed that the proposed method can be applied with HTRF measurements on human subjects, i.e. same subject in different setups. Related findings could be crucial in order to combine and merge HRTF databases from different laboratories and/or companies. Accordingly, performing new analyses on heterogeneous but perceptually-comparable data will be relevant for HRTF selection and personalization procedure [9, 10].

Acknowledgments

This study was supported by the 2016-2021 strategic program "Knowledge for the World" awarded by Aalborg University to Michele Geronazzo.

References

[1] A. Andreopoulou, D. R. Begault, and B. F. G. Katz. Inter-Laboratory Round Robin HRTF Measurement Comparison. *IEEE Journal of Selected Topics in Signal Processing*, 9(5):895–906, Aug. 2015. doi: 10.1109/JSTSP.2015.2400417

[2] A. Andreopoulou, A. Rogiska, and H. Mohanraj. Analysis Of The Spectral Variations In Repeated Head-Related Transfer Function Measurements. pp. 213–218. Proc. Int. Conf. Auditory Display (ICAD), July 2013.

[3] G. Andol, E. A. Macpherson, and A. T. Sabin. Sound localization in noise and sensitivity to spectral shape. *Hearing Research*, 304:20–27, Oct. 2013. doi: 10.1016/j.heares.2013.06.001

[4] R. Baumgartner, P. Majdak, and B. Laback. Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications. In *The Technology of Binaural Listening*, Modern Acoustics and Signal Processing, pp. 93–119. Springer, Berlin, Heidelberg, 2013. DOI: 10.1007/978-3-642-37762-4_4.

[5] D. R. T. Begault. 3-D Sound for Virtual Reality and Multimedia. Technical report, Aug. 2000.

[6] J. Blauert and R. A. Butler. Spatial Hearing: The Psychophysics of Human Sound Localization by Jens Blauert. *The Journal of the Acoustical Society of America*, 77(1):334–335, Jan. 1985. doi: 10.1121/1.392109

[7] C. I. Cheng and G. H. Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. In *Audio Engineering Society Convention 107*, Sep 1999.

[8] M. Geronazzo, A. Carraro, and F. Avanzini. Evaluating vertical localization performance of 3d sound rendering models with a perceptual metric. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–5. IEEE Computer Society, Arles, France, Mar. 2015. doi: 10.1109/SIVE.2015.7361293

[9] M. Geronazzo, E. Peruch, F. Prandoni, and F. Avanzini. Improving elevation perception with a tool for image-guided head-related transfer function selection. In *Proc. of the 20th Int. Conference on Digital Audio Effects (DAFx-17)*, pp. 397–404. Edinburgh, UK, Sept. 2017.

[10] M. Geronazzo, S. Spagnol, and F. Avanzini. Do we need individual head-related transfer functions for vertical localization? The case study of a spectral notch distance metric. *IEEE/ACM Trans. Speech Audio Process. - accepted for publication*, 2018.

[11] R. Greff and B. F. G. Katz. Round robin comparison of hrtf simulation systems: Preliminary results. In *Audio Engineering Society Convention 123*, Oct 2007.

[12] J. Hebrank and D. Wright. Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America*, 56(6):1829–1834, Dec. 1974. doi: 10.1121/1.1903520

[13] P. M. Hofman, A. J. Van Opstal, and J. G. A. Van Riswick. Relearning sound localization with new ears. *The Journal of the Acoustical Society of America*, 105(2):1035–1035, Jan. 1999. doi: 10.1121/1.424942

[14] A. Kohlrausch, J. Braasch, D. Kolossa, and J. Blauert. An Introduction to Binaural Processing. In *The Technology of Binaural Listening*, Modern Acoustics and Signal Processing, pp. 1–32. Springer, Berlin, Heidelberg, 2013. DOI: 10.1007/978-3-642-37762-4_1.

[15] E. H. A. Langendijk and A. W. Bronkhorst. Contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America*, 112(4):1583–1596, Sept. 2002. doi: 10.1121/1.1501901

[16] P. Majdak, R. Baumgartner, and B. Laback. Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization. *Front. Psychol.*, 5, 2014. doi: 10.3389/fpsyg.2014.00319

[17] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig. Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions. In *Audio Engineering Society Convention 134*, May 2013.

[18] J. C. Middlebrooks. Individual differences in external-ear transfer functions reduced by scaling in frequency. *The Journal of the Acoustical Society of America*, 106(3):1480–1492, Aug. 1999. doi: 10.1121/1.427176

[19] J. C. Middlebrooks. Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *The Journal of the Acoustical Society of America*, 106(3):1493–1510, Aug. 1999. doi: 10.1121/1.427147

[20] J. C. Middlebrooks and D. M. Green. Sound Localization by Human Listeners. *Annual Review of Psychology*, 42(1):135–159, 1991. doi: 10.1146/annurev.ps.42.020191.001031

[21] K. A. J. Riederer. Repeatability analysis of head-related transfer function measurements. In *Audio Engineering Society Convention 105*, Sep 1998.