INVESTIGATION OF ITD SYMMETRY ACROSS EXISTING DATABASES OF PERSONALIZED HRTFS

Andrea F. Genovese, Jordan Juras, Chris Miller, Agnieszka Roginska

New York University
Music and Audio Research Lab
35 West 4th Str. 10012, New York NY, United States
genovese@nyu.edu

ABSTRACT

The Interaural Time Difference is one of the primary localization cues for 3D sound. However, in the case of individualized HRIR measurements, the cues are measured on a variety of subjects with different anthropometric characteristics. This fact may cause an angle-dependent ITD asymmetry between locations mirrored across the left and right hemispheres. In fact, asymmetric ITDs might require compensation for applications that aim for a perceptually accurate binaural sound scene. This paper describes an exploratory study performed on a number of publicly available databases of individually measured HRIRs. The analysis was performed separately for each dataset due to varying measurement techniques between sets and, in some cases, different sample rates. Asymmetry in ITDs was found to be consistently more prominent in the rear-lateral angles (approximately between 90° and 130° azimuth) across all databases investigated. It was found that within this region of sensitivity the average difference between specular ITDs exceeds the just noticeable difference value for perceptual discrimination. These findings motivate further exploration of the perceptual impact of ITD asymmetries as well as their relation to anthropometric asymmetries of the head and the outer ear.

1. INTRODUCTION

One of the most crucial binaural cues in spatial sound is the *Interaural Time Difference* (ITD), which is, for a particular location in space, the difference of time-of-arrival between the two ears for a sound source's direct path. ITDs are an important factor in how the human brain localizes a sound source and it is the primary binaural cue for low frequencies [1]. The ITDs are contained into special filters called *Head-Related Impulse Responses* (HRIRs) used for transferring the spatial binaural cues of a position in space to any mono sound file [2].

It is generally accepted that a users' experience is significantly improved by the use of individual HRIRs rather than those recorded on mannequin dummy heads [2]. Different pinna and head morphologies affect both ITD and spectral cues individually for each user, and only those whose morphological characteristics are approximately close to the dummy head would experience a satisfactory binaural or transaural reproduction.

Recently, many modeling techniques for parameterizing individual HRTFs from pictures and scans came to the attention of the sound engineering community [3] [4]. One of such techniques [5] generates a measure of the user's head size through a photographic technique and a subsequent individualization of a general HRIR subset by insertion of head-size related ITDs calculated with a spherical head model (first developed by [6]).

However, these models are all based on the assumption of symmetry of sound perception between the left and right sides of the listener, across the median plane. This assumption of symmetry uses an equal ITD value for a source placed, for example, at an azimuthal location of -30° and its respective mirrored-equivalent at +30° azimuth in the opposite hemisphere. In literature, there has been an indication that this assumption of symmetry does not always hold true [7], motivating and justifying the need for more exploratory studies aimed at analyzing and assessing the available data in order to estimate the severity of asymmetry across measured subjects which is the scope of this paper. While Zhong et al. [7] have explored the asymmetry of HRTFs (the Fourier transformation of HRIRs), this paper will focus on the ITD only.

Significant presence of asymmetric ITD data would motivate a further exploration of correlations between morphological asymmetries and mirrored ITDs. This would also be of significant value to the future development of improved modeling techniques that can adjust or compensate individualized HRIRs that require asymmetric cues. This paper does not cover the issue of "perceptual asymmetry" [8] which confronts the issue of perceptual subjective compensation to the asymmetries of the physical signal. The goal of this paper is rather that to explore whether ITDs in personalized-HRTF databases are statistically asymmetric and whether there is a pattern in the asymmetry across angles.

2. DATABASES

For the scope of this paper, publicly available databases of individual HRIRs were collected. An optimal database would have properties such: high temporal resolution (sample rate, preferably 96 kHz), high azimuthal resolution and a reliable and precise methodology which limits asymmetries created by undesired head movements during measurements. Databases which included anthropometric data of the measured subjects were most desireable, although not all the analyzed sets provided this information. Four databases of individual HRIRs measurements were found to be suitable enough for this analysis. Due to the different characteristics of the databases, the analysis of ITD symmetry variability was performed separately and independently for each dataset. The four selected datasets are: LISTEN [9], CIPIC [10], FIU [11] and MARL [12].

2.1. LISTEN database

This database was recorded for the LISTEN project [9] initiated as a collaboration between IRCAM and AKG. This set was recorded on 51 subjects[1] at 44.1 kHz sampling rate. The set consists of 187 locations per subject, measured at different azimuth resolutions according to the respective elevation.

The strength of this set lies in its measurement technique which made use of a crane structure to precisely move a loudspeaker along a rig and a software-controlled rotating chair with headrest to rotate the subject to the desired azimuthal degree. Using a single loudspeaker avoided the issue of having to compensate for each loudspeaker position. In comparison with other datasets in this list, the method of LISTEN limited measurement errors caused by human misalignment with the target angle, which is likely to result in an ITD asymmetry. Furthermore, the capsule microphones used for recording allowed for blocked-meatus conditions to prevent resonance of the ear-canal.

The major drawback of the LISTEN database is the 44.1 kHz sampling rate which gives us a lower temporal resolution for estimating the degree of ITD asymmetry (distance between samples is $22\mu s$ as opposed to $10\mu s$ in the 96 kHz case).

2.2. CIPIC database

The CIPIC database [10] was compiled and made publicly available by UC Davis. This set consists of 25 azimuth locations recorded between -80° to +80° in azimuth and 50 different elevations from -45° to +230.625° (steps of 5.625°), for a total of 1250 locations recorded per subject (45 subjects). HRTFs were measured using Golay-code signals at 44.1 kHz.

One particular appealing aspect of this measurement set is the inclusion of a detailed collection of anthropometric data of head and pinna morphology parameters which allows for the possibility of a future correlational study between ITD and anthropomorphic asymmetries. The pinna measurements were performed on both sides, providing useful information about morphological asymmetry within subjects' ear characteristics. Unfortunately, also in this case, the dataset was constructed using a non-optimal sample rate (44.1 kHz) for the scope of detecting onsets with high precision.

2.3. FIU database

Florida International University DSP Lab measured the individual HRTFs of 15 different subjects[2] at 12 azimuthal locations (30° spacing) [3] and 6 different elevations [11]. Recordings were conducted at 96 kHz and were accompanied by anthropometric data measured via a 3D scan of the pinnae. The set includes 6 different elevations spaced at 18° apart. The measurements were conducted using the HeadZap system from AuSIM 3D using Golay-Code. The recording methodology relied on a rotating chair and a laser pointer to align the subject, and is thus not fully reliable as tiny head movements between measurements might give rise to ITD asymmetries.

This database could not be retrieved in its original form but only through the NYU MARL standardized repository described in [13], where the FIU data was downsampled to 44100. In the MARL repository the filters were translated to minimum phase filters creating problems in retrieving ITDs in a consistent way (see section 4).

2.4. MARL database

The last database to be analyzed was the NYU MARL (Music and Audio Research Lab) dataset collected by Andreopolou et al. [12] in 2013 and formatted to the MARL repository standard described in [13]. Four participating subjects had their personal HRTFs measured ten times each using different alignment methods (rotating stool with laser pointer or magnetic tracker) at 48 kHz sampling rate. MARL dataset thus allows for the study of HRTF features variability across measurements. HRTFs were measured for a resolution of 10° azimuth and 15° elevation angles (going from -30° to +30° of elevation) for a total of 180 filter pairs per set. 40 sets (10 sets per 4 subjects) were collected.The repository had already filtered out the corrupted sets, reducing the number of subjects to 32. HRTFs were recorded differently across repeated measurements using both Golay Codes, Maximum Length Sequences and Sinesweep technique.

For this analysis, each repeated measurement on the same subject was treated as if it were a different independent subject, thus pooling together different measurement techniques for the same dataset. In this approach, the anthropometric variability between measurements, and consequently the average ITD difference, should result smaller.

In the context of this paper, the shortcomings of this dataset regard the low-precision of the alignment techniques and non-uniform measurement technique, these factors make it hard to discern between the presence of asymmetry in ITD and simple misalignments due to head movements during recording. However a strong ITD asymmetry would stand out against the measurement error as the difference is averaged across subjects.

3. ANTHROPOMETRIC ASYMMETRY

Table 1 illustrates a preliminary analysis on the symmetry of individual anthropometric data included for the CIPIC database. The provided data [10] includes a variety of pinnae features measurements for both the left and the right hemisphere. The table reports the values of the average difference across subjects, and standard deviation of the difference, between the left and right ear measure, for each feature. The most varied feature across subjects is the "Pinna Height", the average subject presented a difference in the left and right pinna height feature of about 0.3207 cm. The presence of these morphological asymmetries further supports the hypothesis that mirrored ITDs will statistically also be asymmetrical.

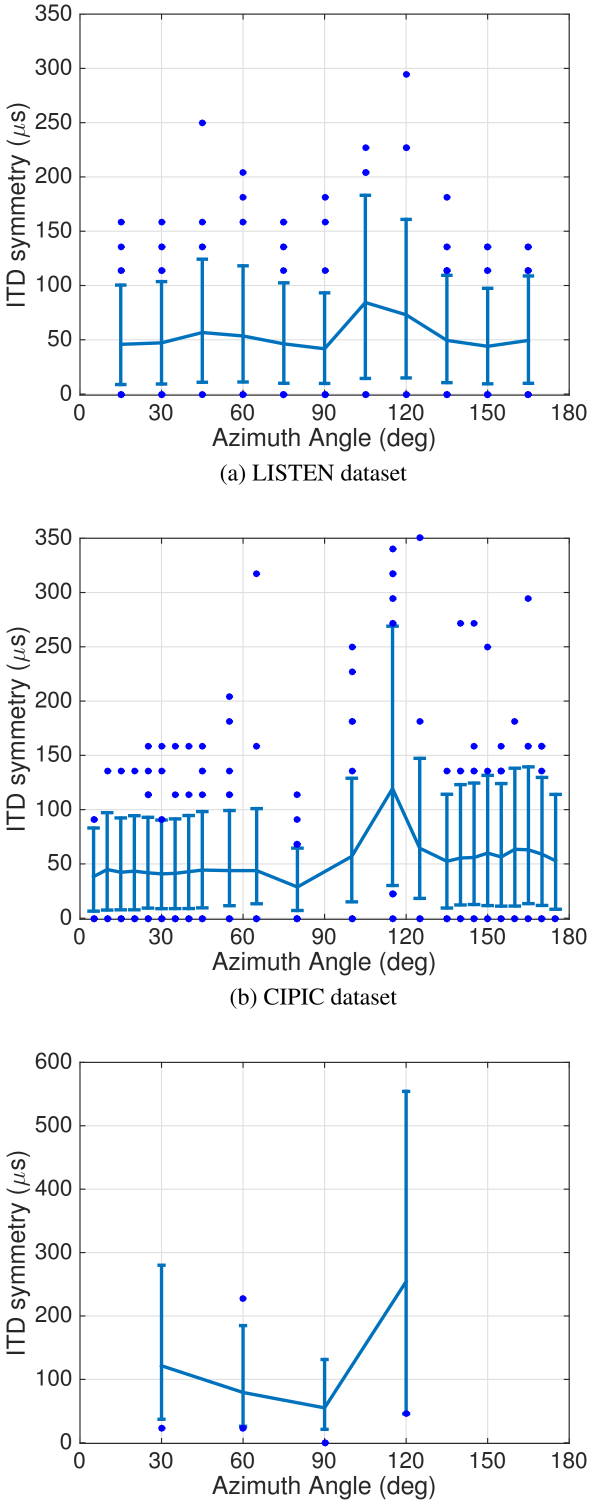
4. ANALYSIS OF SYMMETRY

All the available data was analyzed to produce a series of plots intended to illustrate the presence and severity of asymmetry specific to ITDs. Due to the mismatch in angle resolutions, sample

[1] "Subject 1034" was later removed from the set due to inconsistencies in measurement data. So, 50 subjects were ultimately considered.

[2] Later reduced to 14 due to a formatting problem

[3] One of the angles (±150°) had to be dropped due to corrupted data



(a) LISTEN dataset

(b) CIPIC dataset

(c) FIU dataset

Figure 1: Average ITD symmetry and standard deviation across subjects for the available datasets calculated using (1). The azimuthal resolution depended on the set. All the ITDs were taken at zero elevation.

Anthropometric Feature	Mean	σ
Cavum Concha Height (cm)	0.1213	0.1077
Cymba Concha Height (cm)	0.0910	0.0642
Cavum Concha Width (cm)	0.1357	0.0883
Fossa Height (cm)	0.1551	0.1421
Pinna Height (cm)	0.3207	0.2737
Pinna Width (cm)	0.1613	0.1251
Intertragal Incisure Width (cm)	0.0715	0.0551
Cavum Concha Depth (cm)	0.1295	0.1152
Pinna Rotation Angle (deg°)	0.0957	0.0415
Pinna Flare Angle (deg°)	0.0813	0.0638

Table 1: Mean and Standard deviation of the left-right difference of pinnae features measurements for the subjects in CIPIC

rate, and reliability of measurement techniques, the analysis was not pooled but rather conducted separately for each database.

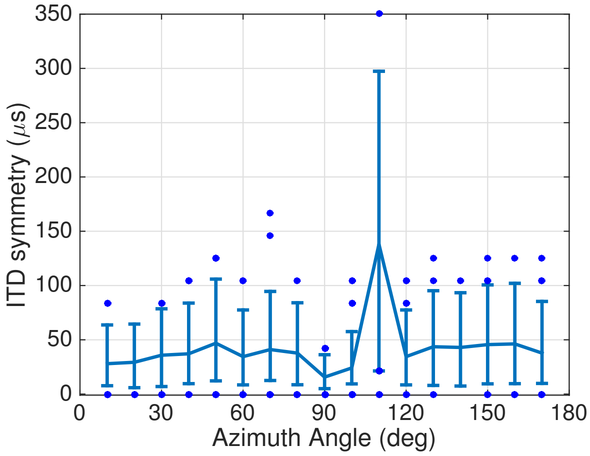
In the context of this document, the ITD was calculated as a measure in samples (then translated into seconds) between the two points of maximum time-domain cross-correlation, between the left and right ear HRIR signals. For consistency, the ITD cross-correlation calculation function was applied to each database, including those who already provided ITD data. This was not possible, however, for the FIU database as the HRIR measurements were only available in a minimum phase format. Hence, for the FIU set, time-domain cross-correlation was not a compatible option for ITD extraction, so the provided ITD values were used.

The ITD symmetry for every subject was calculated as the absolute value of the difference of the magnitude of the ITDs with the magnitude of their respective mirrored counterpart. A difference of 0 would indicate complete symmetry between hemispheres, while, a higher absolute difference would point to a higher the level of asymmetry. The values were averaged across each set of N subjects for every available angle $\theta \in [0; 180]$ within each dataset:

$$\bar{S}(\theta) = \frac{1}{N} \sum_{n \in N} ||ITD_n(\theta)| - |ITD_n(360^\circ - \theta)|| \quad (1)$$

The mean and standard deviation of the absolute asymmetry across subjects was measured for those angles that possessed an ITD value at both hemispheres. Thus, the 0° and 180° angles were excluded from the analysis as they are lacking a mirror correspondent across hemispheres.

Figure 1 illustrates the average ITD difference across subjects between the left and right hemispheres on the horizontal plane (0



(d) MARL dataset

Dataset	$N_{subjects}$	θ_{peak}	$\bar{S}(\theta_{peak})$
LISTEN (a)	50	105°	84μs
CIPIC (b)	45	115°	119μs
FIU (c)	14	120°	254μs
MARL (d)	32	110 °	138μs

Table 2: Peak ITD asymmetry values across examined databases

° elevation). The error bars represent the 95% confidence interval boundaries for each angle. In table 2, the peaks of the ITD average curves are reported, along with the associated azimuth angles. It is worth to notice the proximity of the peak azimuths, placed around $110° \pm 10°$.

By inspection of figure 2, it is possible to identify a common azimuth region of higher magnitude of ITD asymmetry. For all the examined databases, the mentioned region ranges roughly in the lower bound around 90° and 130° in the upper bound. However, the precision of this statement is undermined by the absence of coherent resolution and sample rates between databases. The FIU database presents in this case a lower azimuthal resolution which results in the lack of many azimuth points covered by CIPIC and LISTEN, making a direct comparison not always possible. Furthermore, all the data beyond 120°, had to be discarded due to missing measurements for the left hemisphere ITDs at 150°. However the main point of interest here lies in the similarity of the general shape of the asymmetry curves across databases (figure 2).

Another interesting feature is the consistent minimum ITD asymmetry at 90° azimuth. This can be explained by the fact that at that angle, the ear canal is less protected by the outer ear. Therefore it is unlikely that an asymmetric ear morphology would significantly interfere with the direct signal path to the timpanus.

4.1. Summarization table

Table 3 illustrates an attempt to summarize and characterize the nature of the ITD asymmetry curve of each database and parametrize its distribution across the horizontal plane. The table shows values approximated to the nearest microsecond for the grand mean, standard error, skewness and kurtosis. The standard error for the FIU database is much greater because of the smaller number of subjects.

Such a table could prove useful as a quick reference on the severity of the asymmetry in a database of interest. The skewness value would indicate the offset of the sensitivity region, while the kurtosis would indicate its width.

5. DISCUSSION

The interesting result of this analysis is the presence of a common "region of sensitivity" located between 90° and 130° azimuth. This common curve shape suggest that there might be a case of significant statistical asymmetry between all human subjects ($N_{total} = 141$). But is the magnitude of the ITD asymmetry large enough to cause perceptual asymmetry? According to a well known study by Klumpp and Eady [14], the ITD's JND is highly frequency-dependent. The study found that for a 1 kHz pure tone the JND was on average about $11\mu s$, but for a 90 Hz tone the JND increased to $75\mu s$. For band-limited random noise, the average JND drops to $9\mu s$. As depicted in figure 2, the peak values of the asymmetry (plotted in microseconds μs) range around $100\mu s$. According to the results published here, the localization error between two mirrored locations can be significant enough to be noticeable. Fig 1 (a) and (b) both present a peak higher than most JNDs. The results of fig 1 (c) are currently deemed less reliable due to the low number of subjects (14) and the error-prone methodology, but nevertheless the curve is indicative of the same range of sensitivity between 90° and 130°.

This study also highlights the difficulties related in relating differing formats of public HRIR databases and repositories. Different angle resolutions prevented the pooling of data while measurement errors led to the removal of a significant amount of subject data. It seems appropriate to reinforce the idea of creating a universal, well defined HRTF file format to be shared by the spatial audio community. NYU MARL made an attempt to encourage a repository standard in order to address differences of sample rates, HRIR lengths and angle resolutions [13] (Interpolated angles were not used for this study). The MARL repository was used to retrieve the FIU HRIR data used for this analysis, however, the data was reformatted to fit the defined repository standard, meaning that the HRIRs were downsampled from the original 96 kHz to 44.1 kHz and transformed to minimum phase filters. This standardization necessitated a loss of time resolution and created incompatibility with the cross-correlation measure of ITDs, thus compromising consistency.

Among datasets, only the LISTEN database made use of a very accurate and precise measurement structure in an anechoic chamber. According to the available documentation, the other sets allow for higher chances of measurement error. Even so, these problems did not significantly impact the main objective of this paper which is the exploration of ITD asymmetry patterns, on the horizontal plane, across public datasets of individual HRIRs.

These results show that there is a statistical case of asymmetrical ITD in individual HRIR measurements. Therefore, for the angles in the range where the asymmetry is more pronounced, localization accuracy may degrade when using standardized non-

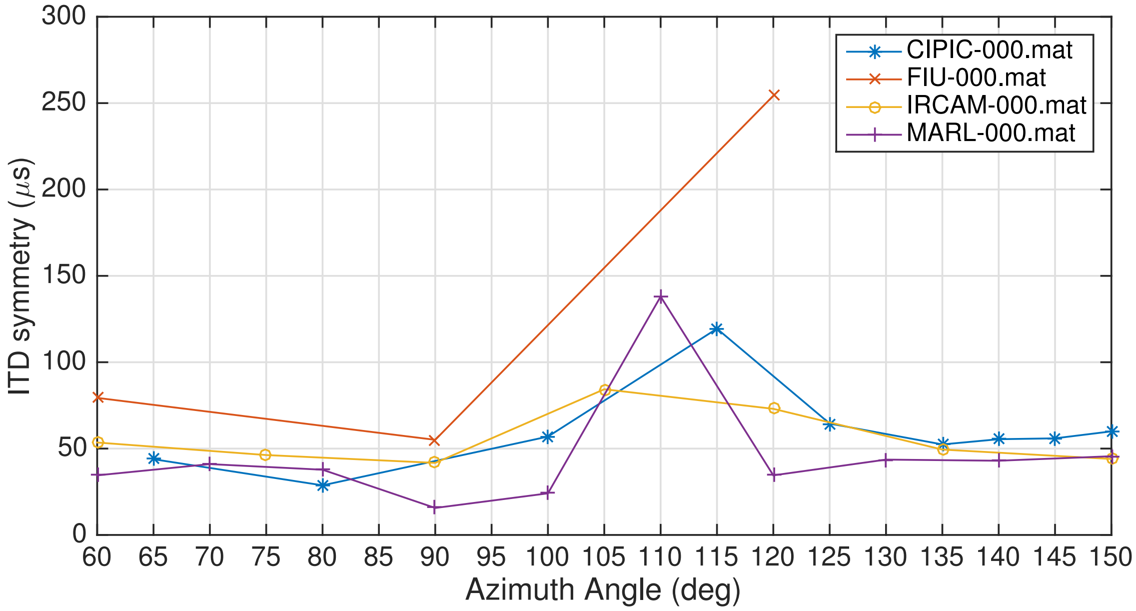


Figure 2: Close-up superposition of all mean ITD asymmetry curves across 60° and 150°. An identified region of sensitivity spans between 90° and 130°

Dataset	N_{subj}	Mean	Median	SE	Skewness	Kurtosis
LISTEN (a)	50	$54\mu s$	$50\mu s$	$2\mu s$	1.43	3.756
CIPIC (b)	45	$52\mu s$	$49\mu s$	$3\mu s$	2.524	11.087
FIU (c)	14	$128\mu s$	$100\mu s$	$24\mu s$	0.847	2.077
MARL (d)	32	$42\mu s$	$38\mu s$	$5\mu s$	3.082	12.2068

Table 3: Summarization of symmetry curves in each database

individual HRIRs or spherical-head modeling. Depending on the application, localization accuracy might be more or less important for the correct delivery of the intended user experience. The relative importance of these results is therefore highly context-dependent. Applications aimed for entertainment are often unlikely to necessitate precise localization and may prefer a simpler processing environment which assumes a spherical head model, but for perceptual studies and localization information scenarios (e.g. blind navigation) then there might be a need for asymmetry compensation.

6. CONCLUSIONS

This study used a number of publicly available personalized HRIR databases to investigate the presence and severity of asymmetric ITD measurement data between the left and right hemisphere. Although a number of these asymmetries are likely to be generated by measurement error, a consistent azimuth range of higher sensitivity to asymmetry was found across the datasets. The identified range approximately spans from ±90° to ±130° but due to poor resolution the bounds of this range are only approximate. This finding confirms the statistical presence of asymmetric ITDs in individually measured subjects caused by morphological differences across hemispheres.

By looking at the peak values on table 2, it is possible to compare the symmetry value with the perceptual Just Noticeable Difference (JND) for ITDs found by [14]. In fact, the JND for pure tones were found to range between $11\mu s$ for high-frequencies and $75\mu s$ for low-frequencies. The asymmetry maximum in the LISTEN database is the lowest peak within analyzed databases and it measures $84\mu s$. Following these results, at least for the examined datasets, the average asymmetry in the sensitivity region is likely to be severe enough to be noticeable.

Nevertheless, there is currently no indication that a physical ITD asymmetry necessarily entails a perceptual asymmetry. Further study needs to assess the impact of the ITD asymmetry from a perceptual standpoint.

6.1. Future Work

The importance of these results has to be assessed in light of the fact that physical asymmetry may not coincide with perceptual asymmetry of spatial sound, internal-delay neural compensation has been previously hypothesized [15]. Further testing is required on whether there is a direct correlation between the two and whether people could benefit from an additional layer of signal compensation in the context of individualization of binaural audio.

Further analysis on the morphological characteristics that give rise to the asymmetry could highlight a possible role of the outer ear pinnae in modifyng the direct path from the source to the ear canal. It would be interesting to run a regression test and look for correlations between physical ITD asymmetry and morphological asymmetry for a pool of subjects representative of the general population. As showns in table 1, the Pinna Height could be the right parameter to focus on. If the asymmetry found in the HRIR measurements were to be correlated to morphological asymmetries, the mismatch between hemispheres could be easily parametrized and accounted for by using a more comprehensive individualization technique (i.e. photographic information) once the relevant parameters are identified with confidence. This is currently a challenge due to the poor availability of morphological data across public datasets.

It would also be interesting to explore the role of other variables, such as elevation and measurement technique, and their relationship with ITD asymmetry more in depth. For example a measurement method which produces a smaller symmetry error could then be identified and prefered for applications where symmetry is a factor to be considered. The summarization table proposed in 4.1 could be improved to represent the symmetry characteristics of a database, given its variables.

7. REFERENCES

[1] F. L. Wightman and D. J. Kistler, “The dominant role of low-frequency interaural time differences in sound localization,” *The Journal of the Acoustical Society of America*, vol. 91, no. 3, pp. 1648–1661, 1992.

[2] M. Morimoto and Y. Ando, “On the simulation of sound localization.” *Journal of the Acoustical Society of Japan (E)*, vol. 1, no. 3, pp. 167–174, 1980.

[3] N. Gupta, A. Barreto, and M. Choudhury, “Modeling head-related transfer functions based on pinna anthropometry,” in *Proc. of the Second International Latin American and Caribbean Conference for Engineering and Technology (LACCEI)*, 2004.

[4] M. Dellepiane, N. Pietroni, N. Tsingos, M. Asselot, and R. Scopigno, “Reconstructing head models from photographs for individualized 3d-audio processing,” in *Computer Graphics Forum*, vol. 27, no. 7. Wiley Online Library, 2008, pp. 1719–1727.

[5] J. Juras, C. Miller, and A. Roginska, “Modeling itds based on photographic head information,” in *Audio Engineering Society Convention 139*, Oct 2015.

[6] V. R. Algazi, C. Avendano, and R. O. Duda, “Estimation of a spherical-head model from anthropometry,” *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–479, 2001.

[7] X.-L. Zhong, F.-c. Zhang, and B.-S. Xie, “On the spatial symmetry of head-related transfer functions,” *Applied Acoustics*, vol. 74, no. 6, pp. 856–864, 2013.

[8] S. E. Boehnke and D. P. Phillips, “Azimuthal tuning of human perceptual channels for sound location,” *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1948–1955, 1999.

[9] O. Warufsel, “Listen hrtf database, ircam,” 2002. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/

[10] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, “The cipic hrtf database,” in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 99–102.

[11] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo, “Hrtf database at fiu dsp lab,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 169–172.

[12] A. Andreopoulou, A. Rogińska, and H. Mohanraj, “A database of repeated head-related transfer function measurements,” 2013.

[13] A. Andreopoulou and A. Roginska, “Towards the creation of a standardized hrtf repository,” in *Audio Engineering Society Convention 131*, Oct 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16096

[14] R. Klumpp and H. Eady, “Some measurements of interaural time difference thresholds,” *The Journal of the Acoustical Society of America*, vol. 28, no. 5, pp. 859–860, 1956.

[15] P. Joris and T. C. Yin, “A matter of time: internal delays in binaural processing,” *Trends in neurosciences*, vol. 30, no. 2, pp. 70–78, 2007.