Audio Engineering Society

Conference Paper

Presented at the Conference on
Headphone Technology
2016 Aug 24–26, Aalborg, Denmark

Interaural Distances In Existing HRIR Repositories

Chris Miller, Jordan Juras, Andrea Genovese, and Agnieszka Roginska

Music And Audio Research Laboratory, New York University

Correspondence should be addressed to Chris Miller (crm313@nyu.edu)

ABSTRACT

With the recent development of low-cost and efficient methods for generating individualized Interaural Time Differences (ITDs), this paper investigates the distribution of interaural distances among certain populations in order provide a framework for improving the performance of individualized binaural audio systems across a wider range of head morphologies. Interaural distances are extracted from the publicly available LISTEN and CIPIC spatial audio databases in order to generate distributions across subjects, and from the MARL-NYU database in order to investigate measurement stability across testing sessions. The interaural difference is shown to be a means to measure the magnitude of an individual's set of ITDs. Furthermore, the constraints introduced on the precision of measured ITDs by limited sampling rates across all three datasets are explored, and the authors motivate the use of higher sampling rates in the development of spatial audio databases.

1 Introduction

The Interaural Time Difference (ITD) is one of the primary localization cues in to the human auditory system, and as such is a crucial component in any spatial audio representation. Sound waves from an object at a location ρ, θ, ϕ (where ρ is the distance from the center of the head to the object, θ is the azimuthal angle and ϕ the polar angle or 'elevation' in spherical coordinates) reach the contralateral ear later than the ipsilateral ear, resulting in an Interaural Time Difference (ITD). In virtual surround and other binaural spatial audio systems, the ITD is encoded as part of the Head Related Impulse Response (HRIR) filter bank used for spectral localization cues.

Individualized HRIRs vastly improve the spatial and immersive quality of binaural audio when compared to generic HRIRs or those recorded from mannequin dummys, but obtaining individualized HRIRs is often expensive and impractical. Different head and pinnae morphologies affect not only the frequency response of the HRIR at a particular location, but the time delay (ITD) at that location as well.

Recently, a handful of modeling techniques for ITDs have been proposed using photographic means [1] [2]. Such techniques typically seek to measure the distance between the opening of the two ear canals, the so-called "interaural distance". This distance is the most crucial factor in determining the set of ITDs for an individual on the horizontal plane. In the spherical head model, this distance corresponds to the diameter of the head [3]. An individual's personalized ITD cues can be measured acoustically with the placement of microphones in the ear canals and a sweep of a sound source around the head, but this is again quite impractical for most.

With low-cost and efficient photographic interaural distance measurement systems in development, this paper takes a look at some of the more well-known existing HRIR repositories and extracts interaural distances from them. Armed with a clearer idea of the distributions of interaural distances among certain populations, binaural audio systems can be improved to serve a wider range of head morphologies.

2 Databases

Publicly available databases of HRIRs were gathered and studied for this paper. Optimal databases would have high temporal resolution (96 kHz), reliable and precise methodologies for acoustic measurements, good coverage over the horizontal plane (high azimuthal resolution), and a large amount of subjects. Our efforts were therefore primarily focused on the LISTEN database from IRCAM, and the CIPIC database from the University of California, Davis. Due to differences between databases in terms of azimuthal resolution and demographic characteristics, each database was analyzed separately.

Additionally, the MARL-NYU dataset was used to asses the validity of the interaural distance extraction method, as it contains only four subjects, but with each subject's HRIR measured ten times.

2.1 LISTEN

The LISTEN project began as a collaboration between IRCAM and AKG [4]. LISTEN contains the HRIRs of 51 subjects measured at a sampling rate of 44.1 kHz. The HRIR was measured at 187 different locations for each subject – 24 of which lie on the horizontal plane. HRIRs are therefore available at every 15° along the horizontal plane.

LISTEN's methodology is its greatest asset, as measurements involved the use of a crane for precise sound source movement, and a software-controlled rotating chair with headrest to rotate the subject to the desired azimuthal location with respect to the sound source. Capsule microphones used for HRIR capture allowed for blocked-meatus conditions, preventing resonances in the ear-canal. HRIR capture with a 44.1 kHz sampling rate, however, is not ideal, as the distance between samples grows to $22\mu s$ as opposed to $10\mu s$ in the 96 kHz case.

2.2 CIPIC

The CIPIC database was assembled at the University of California, Davis, and is publicly available [5]. CIPIC contains the HRIRs of 45 different subjects measured at a sampling rate of 44.1 kHz. HRIRs was measured at 1250 locations for each subject: elevation coverage spans from $-45°$ to $230.625°$ in increments of $5.625°$, while azimuthal coverage sweeps out from each location along an equidistant path on the median plane from $-80°$ to $80°$. Azimuthal coverage contains samples at $-80°$, $-65°$, $-55°$, from $-45°$ to $45°$ in steps of $5°$, $55°$, $65°$, and $80°$.

Blocked-meatus microphones were used in CIPIC measurements, and subjects heads, while not restrained, were surrounded by a hoop of loudspeakers. HRIRs were measured using Golay-codes, and later free-field compensated to account for resonances of the transducers. As with the LISTEN dataset, the use of a 44.1 kHz sampling rate remains less than ideal for the purposes of accurate ITD capture.

2.3 MARL

The MARL HRIR repository was collected by Andreopoulou et al. in 2013 at New York University [6] and formatted to the MARL standard as described in [7]. Four subjects had their individualized HRIRs measured multiple times, allowing for the analysis of HRIR variability within subjects – resulting in a total of 32 sets once corrupted sets have been filtered out. Different alignment techniques were used for each measurement, such as magnetic trackers and rotating stools with laser pointers. Various excitation signals (sine sweeps, Golay codes, and Maximum Length Sequences) were also used. HRIRs were measured with a resolution of 10° azimuth along the horizontal plane, and were captured using blocked-meatus microphones at a sampling rate of 48 kHz.

3 Interaural Distance Extraction

Using the spherical head model, the radius of a spherical head r can be extracted from the ITD τ produced by a sound located at an azimuthal angle θ on the horizontal plane. Sound arrives to the head as plane waves, hitting the ipsilateral ear first and the contralateral ear a time τ later. This time difference τ arises due to a difference δ in the distance traveled by the plane wave to the contralateral ear with respect to the ipsilateral ear [3]. This understanding of binaural sound localization, often referred to as "duplex theory", was first proposed by Lord Rayleigh in 1907 [8].

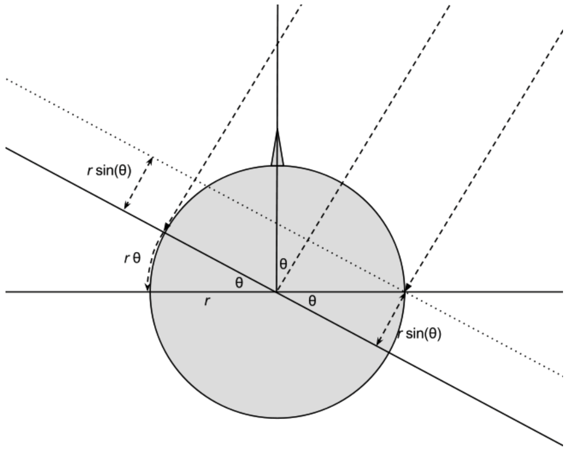

Fig. 1: Duplex theory for determining the path length distance δ between sounds incident to the ipsilateral and contralateral ears from a sound source at azimuthal location θ.

Figure 1 shows the plane waves from a sound source at θ with respect to the listener. The path from the source to the ipsilateral ear is direct, but the path traveled by the plane wave to the contralateral ear is a further δ meters, where the path differential δ is

$$\delta(\theta) = r\sin(\theta) + r\theta = r(\sin(\theta) + \theta) \quad (1)$$

Dividing equation 1 by the speed of sound (c) in meters per second (340.29 m/s) gives τ at location θ:

$$\tau(\theta) = \frac{r}{c}(\theta + \sin(\theta)) \quad (2)$$

Rearranging equation 2 gives the radius of the spherical head, r, as a function of θ and $\tau(\theta)$. The interaural distance D, defined as the distance between the two ears, is simply $2r$, and therefore

$$D = \frac{2c\tau(\theta)}{(\theta + \sin(\theta))} \quad (3)$$

Since each subject has their HRIR measured at many azimuthal locations around the head, we calculate D at every location in the horizontal plane available and then average to smooth out measurement errors. ITDs are not symmetric about the head, so the calculated interaural distances at $\pm 120°$, for example, might be quite different due to ITD asymmetry [9].

As documented in the literature, the spherical head model is an idealized model of the head and not an accurate representation of typical head morphologies [3] [2]. The cross-section of the head in the horizontal plane is not a circle, and pinnae and ear canal openings often have some displacement with respect to the center of the head. The interaural distance is therefore less a physical measure of a subject's head size and more a measure of the magnitude of a subject's set of ITDs along the horizontal plane.

4 Results

Figures 2 and 3 show the distributions of interaural distances for the LISTEN and CIPIC datasets. Table 1 shows the mean and standard deviation for each dataset's distribution of interaural distances.

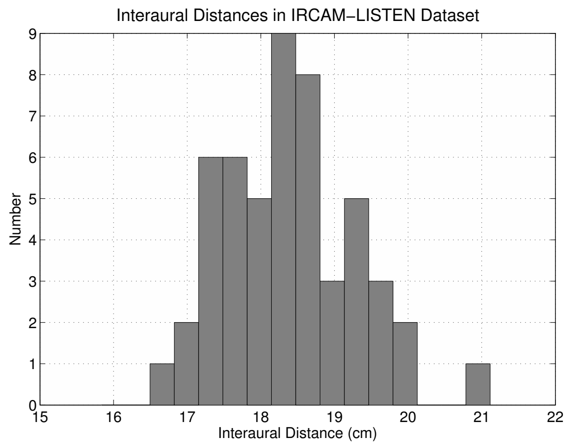

Fig. 2: Distribution of interaural distances in LISTEN dataset.

The primary goal of this paper was to investigate the distribution of interaural distances D across subjects in the aforementioned databases. As is depicted in Figures 2 and 3, the distribution of D for the LISTEN and CIPIC databases is not consistent. Despite using different techniques and experimental setups, the HRIRs in

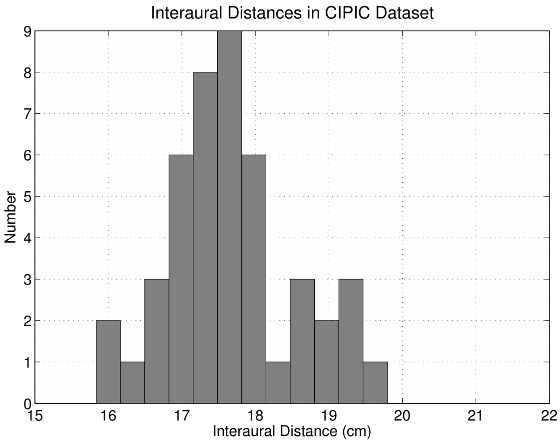

Fig. 3: Distribution of interaural distances in CIPIC dataset.

Database	μ_D (cm)	σ_D (cm)
LISTEN	18.39	0.90
CIPIC	17.67	0.86

Table 1: Mean μ and standard deviation σ of interaural distances for the CIPIC and LISTEN databases

Database	μ_x (cm)	σ_x (cm)
LISTEN	15.57	1.29
CIPIC	14.43	0.92

Table 2: Mean μ and standard deviation σ of measured head sizes for the CIPIC and LISTEN databases

both databases were captured using the same sampling rate of 44.1 kHz. The median value of D for the LISTEN database is 18.4 cm, compared to 17.6 cm for the CIPIC database. For both distributions, the mean μ_D and standard deviation σ_D are shown in Table 1.

The distribution of interaural distances in the LISTEN database skews nearly 1 centimeter larger than those in CIPIC. LISTEN does contain only half the amount of locations on the horizontal plane than CIPIC, but while CIPIC has higher azimuthal resolution, LISTEN's methodology (as detailed in Section 2.1) allows for more precise head-to-sound-source alignment.

4.1 Correlation with Measured Head Size

The simplest explanation for this skew would be that the distribution of actual head sizes x in the LISTEN database is larger than those in CIPIC. Luckily, both datasets provide head morphology data for most of their subjects (LISTEN provides morphology for 94% of their subjects and CIPIC 82%). Head width, the distance between the opening of the ear canals, is extracted from high-resolution photographs in both LISTEN and CIPIC. The mean μ_x and standard deviation σ_x for the distribution of measured head widths in LISTEN and CIPIC is given in Table 2, and follows the same pattern as the distribution of interaural distances (Table 1). The distribution of measured head widths x in LISTEN skews about a centimeter larger, with a slightly larger standard deviation σ_x in LISTEN as well.

The interaural distance measurement does not overlap directly with the physically measured head sizes — the mean interaural distance μ_D is around 3 centimeters larger than the mean head size μ_x in both datasets. As discussed in Section 3 this is not surprising, as the cross-section of the head in the horizontal plane is typically more elliptical than circular. Our calculation of the path length difference δ, then, systematically overestimates the radius of the head, which, realistically, is not constant. However, the interaural distance does track the trends of head sizes in databases of acoustically measured HRIRs, and thus provides a good measure of the magnitude of the set of ITDs for a given individual.

5 Discussion

5.1 Variability of Interaural Distance Across Subjects

A sampling rate of 44.1 kHz corresponds to a temporal resolution per sample of 0.02 milliseconds, and a physical resolution of 0.77 centimeters assuming a sound wave is traveling at 340.29 meters per second. Considering that the standard deviations σ_D for the LISTEN and CIPIC datasets (see Table 1) are both only slightly larger than a single sample's physical resolution, it is hard to make a confident claim regarding the accuracy of these distributions. The ITDs for all measurements are generated using a cross-correlation between the discrete ipsilateral and contralateral HRIR signals, which returns the number of samples separating the points of maximum correlation between the two waveforms. As such, the generated ITDs can only provide measurements with a precision on the order of one sample, limiting the precision of the calculated interaural distances to approximately 0.77 centimeters at a sampling rate of 44.1 kHz. In comparison, a sampling rate of 96 kHz corresponds to a physical resolution of 0.35 centimeters.

5.2 Variability of Interaural Distance Across Measurement Trials (MARL Database)

The MARL dataset differs from the LISTEN and CIPIC databases by studying only 4 subjects across 10 different sessions. For subjects B and D, data corruption resulted in 8 and 4 recorded sessions, respectively.

Analyzing the variability of D across sessions for a single subject further illustrates the limitations introduced by a low sampling rate. Unlike the LISTEN and CIPIC databases, the MARL ITDs were captured with a sampling rate of 48 kHz, corresponding to a physical resolution of 0.71 centimeters. Figure 4 depicts a bee swarm plot showing the variability in D (difference of calculated D for a given trial to average $\bar{D}$ across trials) for each subject. As with the inter-subject variability discussed for the LISTEN and CIPIC datasets,

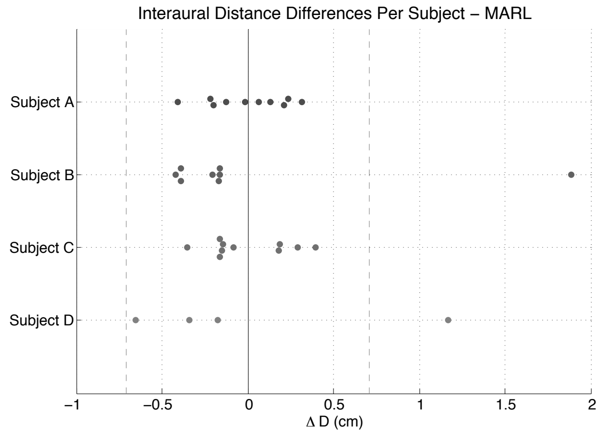

Fig. 4: Difference per measurement trial of interaural distance D to subject's mean interaural distance $\bar{D}$. Dashed vertical lines at ± 0.71 cm show physical resolution around mean due to sampling rate of 48 kHz.

D varies between trials by an amount smaller than the physical resolution of the experiment. Two positive outliers (contained within subjects B and D) skew the mean value of these subjects' plots significantly in the positive direction.

The MARL database was generated using the experimental setup described in Section 2. The technique involved rotating the subject through 360° in azimuth with respect to a sound source guided by a laser that was not fixed to either the stool or the subject. Without a fixed physical reference for a subject's head (as was used, for example, when creating the LISTEN database), ITD measurement errors on the order of a few samples are not surprising. These measurement errors are further exacerbated by a limited sampling frequency, which limits the precision of calculated interaural distances to increments of 0.71 centimeters. By calculating the interaural distance at all points around the head and averaging, we smooth out some of these errors and are able to generate values within one sample accuracy, as seen in Figure 4, but more precise measurements before averaging would improve precision downstream.

6 Conclusion

Using duplex theory and the spherical head model, distributions of interaural distances for two of the most popular existing repositories of HRIRs — LISTEN and CIPIC — were generated. The interaural distance, calculated as a function of an azimuthal location on the horizontal plane θ and the acoustic ITD at that location $\tau(\theta)$, tends to overestimate the physical head width by roughly 3 centimeters due to the spherical head model's idealizations. However, the interaural distance does seem to scale accordingly, as the distributions of both calculated interaural distances and measured physical head widths exhibit the same trends. As such, the interaural distance provides a way to extract a measure of the magnitude of an individual's set of ITDs that is very closely related to their morphological attributes.

The existing HRIR repositories analyzed in this study date back over a decade: the LISTEN dataset was assembled from 2002-2003 and CIPIC in 2001. While the use of a sampling rate of 44.1 kilohertz is reasonable for the capture of the most significant spectral cues in the frequency domain version of the HRIR (the Head Related Transfer Function, or HRTF), it has been shown that such low sampling rates put a less than ideal constraint on ITD precision, and thus the extracted interaural distance. HRIR capture at higher sampling rates would increase temporal resolution and allow for more precise extraction of head morphology from acoustic measurements.

References

[1] Dellepiane, M., Pietroni, N., Tsingos, N., Asselot, M., and Scopigno, R., "Reconstructing head models from photographs for individualized 3D-audio processing," in *Computer Graphics Forum*, volume 27, pp. 1719–1727, Wiley Online Library, 2008.

[2] Juras, J., Miller, C., and Roginska, A., "Modeling ITDs Based On Photographic Head Information," in *Audio Engineering Society Convention 139*, 2015.

[3] Algazi, V. R., Avendano, C., and Duda, R. O., "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, 49(6), pp. 472–479, 2001.

[4] Warufsel, O., "LISTEN HRTF Database, IRCAM," 2002.

[5] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., "The CIPIC HRTF Database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 99–102, IEEE, 2001.

[6] Andreopoulou, A., Roginska, A., and Mohanraj, H., "A Database of Repeated Head-Related Transfer Function Measurements," 2013.

[7] Andreopoulou, A. and Roginska, A., "Towards the Creation of a Standardized HRTF Repository," in *Audio Engineering Society Convention 131*, 2011.

[8] Strutt, J. W., "XII. On our perception of sound direction," *Philosophical Magazine Series 6*, 13(74), pp. 214–232, 1907.

[9] Genovese, A., Juras, J., Miller, C., and Roginska, A., "Investigation Of ITD Symmetry Across Existing Databases Of Personalized HRTFs," *Manuscript submitted for publication*, 2016.