USING PUBLISHED HRTFS WITH SLAB3D: METRIC-BASED DATABASE SELECTION AND PHENOMENA OBSERVED

Joel D. Miller, Martine Godfroy-Cooper

San Jose State University Foundation,
Advanced Controls and Displays Group,
NASA Ames Research Center,
Moffett Field, CA, 94035, USA
Joel.D.Miller@nasa.gov
Martine.Godfroy-1@nasa.gov

Elizabeth M. Wenzel

Advanced Controls and Displays Group,
NASA Ames Research Center,
Moffett Field, CA, 94035, USA
Elizabeth.M.Wenzel@nasa.gov

ABSTRACT

In this paper, two publicly available head-related transfer function (HRTF) database collections are analyzed for use with the open-source slab3d rendering system. After conversion to the slab3d HRTF database format (SLH), a set of visualization tools and a five-step metric-based process are used to select a subset of databases for general use. The goal is to select a limited subset least likely to contain anomalous behavior or measurement error. The described set of open-source tools can be applied to any HRTF database converted to the slab3d format.

1. INTRODUCTION

The quality of the spatial reconstruction of a virtual acoustic environment (VAE) is highly dependent on accurately measured and correctly reproduced binaural and monaural acoustic cues. Binaural cues arise through interaural time differences (ITDs) and interaural level differences (ILDs) between the two ears corresponding to the azimuthal component of a sound source direction relative to the head of the listener. The localization of a sound source in elevation is derived from monaural cues which consist of direction-dependent linear spectral filtering by the pinnae, head, and torso. The incident waveforms are reflected and diffracted in a complex and direction-dependent way yielding enhancement and attenuation at particular frequency bands [1], [2], [3], [4]. In VAEs, these physical phenomena are encapsulated in head-related transfer functions (HRTFs). Typically, the HRTFs are derived from head-related impulse responses (HRIRs) measured by inserting microphones in a subject's left and right ear canals for sound sources at a fixed radius from the head. The HRTFs are measured at different sound source azimuths (az) and elevations (el) that constitute the measurement grid. Since the process is quite laborious and prone to error, an attractive alternative is using publically available databases that provide the easiest access to so-called "generic", or non-individualized, HRTFs.

2. THE LISTEN AND CIPIC HRTF COLLECTIONS

The Listen (IRCAM [5]) and CIPIC (UC Davis [6]) collections are two widely used publicly available databases of high spatial resolution HRTFs. An interesting question is how comparable are these two collections given that they have been gathered using two different measurement systems and procedures (see documentation at [5], [6], [7] for details).

In particular, it can be seen from Table 1 that the two collections differ in terms of loudspeaker configuration, subject positioning, test signal, data formats, and spatial resolution (measurement grids). Note that the measurement radii are also different. Differing radii can conceivably introduce air absorption effects. However, at a radius difference of 0.9 m, these effects are minimal; for example, at 70° F and 50% humidity, there would be an attenuation of 0.005 dB/m at 1,000 Hz and 0.152 dB/m at 10,000 Hz [8].

Table 1: CIPIC and Listen database collection characteristics.

	CIPIC	LISTEN
Databases	45	51
Environment	Non-anechoic	Anechoic
Loudspeaker Configuration	5 Bose Acoustimass, Semi-fixed structure	1 TANNOY system 600, Mobile structure
Measurement Radius	1.0 m	1.9 m
Subject Positioning	Self head position monitoring, Head tracking	Rotating chair + head rest, Head tracking + visual feedback for az positioning
Microphone, Positioning	Etymotic Research ER-7C, Blocked meatus	Knowles FG3329, Blocked meatus
Test Signal	Golay-code	Logarithmic sweep
Coordinate System	Interaural-polar	Vertical-polar
Resolution Az	25 locations	24 locations
Resolution El	50 locations	10 locations
HRIRs	1250	187 (non-uniform)
HRIR Length	200	512
Data Format	Free-field compensated, no raw data	Raw and diffuse-field compensated
ITD Data	Yes	No
Sample Rate	44.1 kHz	44.1 kHz

Figure 1 illustrates the two measurement grids employed in the CIPIC and Listen measurement systems. It is clear from this figure that the density and the regularity of the grids vary significantly between the two databases.
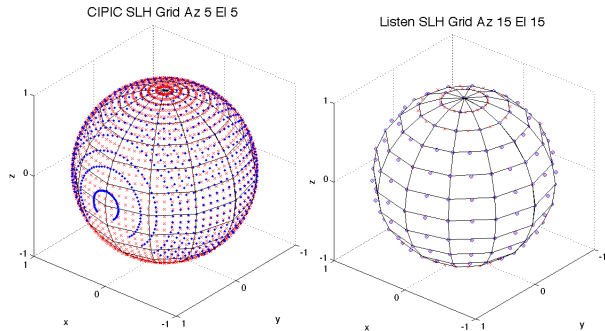
Figure 1: Measurement (blue) and SLH (red) grid patterns, CIPIC (left) and Listen (right). Note, CIPIC uses an interaural-polar coordinate system; Listen uses a vertical-polar coordinate system. Axes: slab3d coordinate system (vertical polar) with +x forward (0° az, 0° el).

3. SLAB3D CONVERSION

Since the Listen and CIPIC HRTF collections were being evaluated for use with the slab3d virtual acoustic environment rendering system [9], [10], they were converted to the slab3d HRTF database format (.slh files) before evaluation. After conversion, the files contain the HRIR and the ITD filter data used for real-time rendering. From an analysis perspective, conversion also provides a common format that accounts for differences in individual lab's data formats, coordinate systems, and measurement grids. As reported in Table 1, ITD data are not released with the Listen database collection. Thus, ITDs were extracted using a modified version of the Nam, Abel, and Smith algorithm described in [11] where ITDs are derived from HRTF group-delay [12], [13], [14]. This method of extraction requires uncompensated data. Being the CIPIC release does not include uncompensated data, the release ITDs were used. In the figures that follow, a positive ITD implies a left ear lag (source in right hemisphere, positive az) and a negative ITD implies a right ear lag (source in left hemisphere, negative az). For both Listen and CIPIC, the compensated data were converted to minimum-phase to reduce filter length and simplify real-time interpolation. Bi-harmonic spline interpolation [15] was used to resample the data to a uniform spherical grid (via the map2map() tool kindly provided by Jonathan Abel).

4. HRTF DATABASE SELECTION

The spatial accuracy of the rendered sound image is primarily determined by the quality of the HRTF measurement and how well the database matches the listener. To evaluate the quality of the conversion, and the HRTFs themselves, several tools were developed that are available publicly in the slab3d release [9] as part of the "slabtools" collection of MATLAB utilities. An HRTF listening utility named HeadMatch is also available but, at present, simply allows for the rapid comparison of HRTF databases.

The most straightforward form of HRTF evaluation is visualization. To this end, a variety of visualization tools were written that allow detection of sensible trends in ITDs,

magnitude responses, energies, and IIDs (interaural intensity differences). The initial purpose was to spot anomalies in the conversion process, but potential measurement issues were also revealed. Visualization has a few drawbacks such as laboriousness, human error, and subjectivity. In some cases, visual anomalies can be easily quantified into metrics. A variety of these were created to enable a statistical analysis approach to HRTF evaluation. For both visualization and metrics, midsagittal head symmetry was frequently employed as a reference.

Given that 96 total HRTF databases are available after slab3d conversion, it is desirable to select an optimal subset for general use. Thus, our approach was to select a subset least likely to contain conversion artifacts, anomalous behavior, and measurement error. It is important to note that the selection procedure and the corresponding results specifically relate to the use of the converted Listen and CIPIC HRTFs. Indeed, artifacts related to the conversion process itself could potentially influence the metrics and the outlier removal procedure. This is especially true in regions where spherical interpolation was used to fill in sparse locations in the measurement grid (e.g., low elevations).

4.1. HRTF Visualization

4.1.1. ITD

A useful ITD visualization is the simultaneous display of database and spherical head model (SHM) ITD overlaid in two continuous curves, with el and az on the horizontal axis and microseconds on the vertical axis (see Figures 2-4). Elevations are ordered starting at el 90° and proceed downwards. At each elevation, azimuths are ordered starting at 180° and proceed counter-clockwise. The Woodworth & Schlosberg SHM model (as described by Blauert in [16]) provides horizontal-plane ITD formulae for three distance thresholds, extreme near field, near field, and far field. The model can be extended to support elevation by rotating the horizontal plane about the interaural axis and modifying azimuth accordingly. For compatibility with a previous custom SHM, the near-field model was selected. The SHM parameters are azimuth, elevation, measurement radius, head radius, and speed of sound. The measurement radii were determined from the measurement systems (Listen 1.9m, CIPIC 1.0m) and the speed of sound from slab3d's default value of 346 m/s. The head radius was also taken from a slab3d default, 0.09m, a value consistent with published measurement averages [7], [17]. In general, ITDs tended to follow the spherical head model quite closely, with the main exception being the extreme values in the ±90° az region along the interaural axis. However, a few patterns were observed that suggested subject positioning errors might have occurred.

To determine the impact of subject positioning errors on visualization patterns, errors in x, y, z, yaw, pitch, roll were transformed to spherical-head model az, el, range and plotted with the ITD visualization. The results were ITD curve skew (x), bias (y), expansion-compression (+z), diagonal bias (roll, Figure 2), curve shift (yaw, Figure 3), and no effect (pitch). Some of these patterns were, in fact, observed in both the

Listen and CIPIC ITD data. Expansion-compression was frequently observed, but this could easily be due to head vertical asymmetry. Note, when converting databases to the slab3d format, ±90° elevation ITDs were set to 0 to reduce biharmonic-spline interpolation artifacts. Thus, these regions will differ from ITD subject-positioning error model predictions.

The three most common ITD patterns observed were *curve vertical bias*, *curve horizontal shift,* and *curve discontinuity*.

Curve vertical bias (Figure 2): A variety of bias patterns were noticed, e.g., in high elevations only, in low elevations only, or positive at one end while negative at the other ("diagonal bias"). The *dbias* metric quantifies bias by taking the mean of the ITD values at each elevation and subtracting the min from the max.
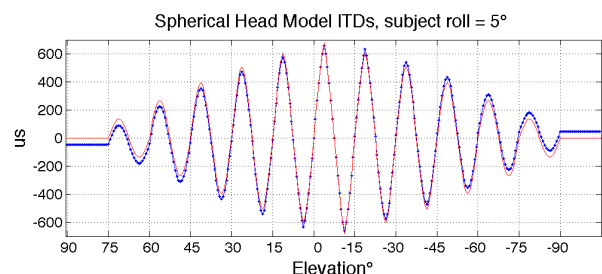

Figure 2: Subject-positioning error model for a 5° roll positioning error yielding a diagonal ITD bias (dotted line) relative to SHM ITD (solid line).

Curve horizontal shift (Figure 3) refers to the entire ITD curve being shifted left or right relative to the spherical ITDs. This may imply a subject-positioning error in yaw that remained constant throughout the measurement session. The *shift* metric is calculated by taking the cross correlation of the database ITD curve and the spherical ITD curve. The two curves can be upsampled to improve accuracy.
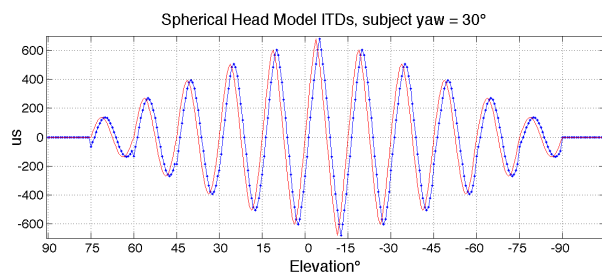

Figure 3: Subject-positioning error model for a 30° yaw positioning error yielding a curve shift (dotted line) relative to SHM ITD (solid line). Note that the yaw error has been exaggerated for illustration purposes.

Curve discontinuity (Figure 4) refers to unexpected curve slope changes and/or uneven value distribution. Likely causes include subject movement [17], subject positioning relative to multiple speakers, and ITD extraction artifacts. This behavior was more apparent in the CIPIC data than in the Listen data.

Since the ITD extraction algorithm for the Listen data was developed in-house, curve discontinuities due to the algorithm itself could be studied and eliminated. The resulting Listen ITD data were, in general, quite smooth. The max "discontinuity" metric, *mdis* (max ITD/SHM instantaneous slope difference, omitting ±90° az), was created to capture ITD curve discontinuity.
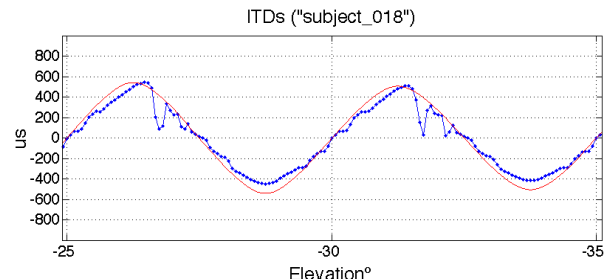

Figure 4: ITD visualization for a CIPIC HRTF database with ITD curve discontinuity (extreme case, subject_018, HRTF index 9). Dotted line: CIPIC ITDs. Solid line: SHM ITDs.

4.1.2. Magnitude Response, Energy, and IID

One expected behavior pattern in HRTF data sets is trend continuity in the magnitude response peaks and valleys as a function of space and frequency. The magnitude response symmetry visualization (Figure 5) provides two right ear displays on either side of the left ear display (center). Since the two dark vertical lines separating the ears serve as the symmetry reference, the images on either side should be roughly symmetrical. Note that an asymmetric high-frequency attenuation is observable as the vertical dark stripes located in the upper portion of each pane of Figure 5.
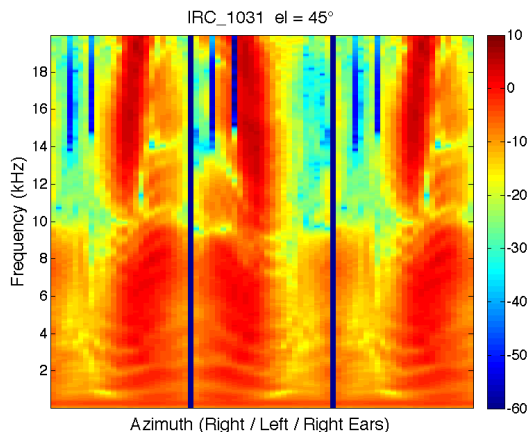

Figure 5: Magnitude response symmetry visualization for Listen IRC_1031 database (HRTF index 25), all azimuths at el 45°. Horizontal axis: right ear az -180° to 180°, left ear az -180° to 180°, right ear az -180° to 180°. Color axis in dB.

Energy and IID visualizations provide useful tools for spotting spatial anomalies in azimuth and elevation (Figures 6-8). To make the data more manageable and perceptually relevant, critical-band processing can be performed to reduce

the frequencies analyzed to 20 critical bands (critical-band processing is performed using the Auditory Toolbox [18]). Viewing response energy on a sphere can help detect speaker-related issues. In multi-speaker systems, specific speakers are frequently mapped to specific sets of azimuths or elevations. Issues can appear as constant azimuth or elevation trends in interaural-polar (CIPIC) or vertical-polar coordinates (Listen). For example, some CIPIC databases exhibit a sagittal pattern when viewing database energy on a sphere (Figures 6 and 7, vertical striations). The CIPIC measurement system uses a multi-speaker approach where specific speakers are mapped to specific sagittal bands. Viewing the energy by critical band (Figure 7 top) reveals an attenuation pattern across multiple critical bands. Since the impact of this phenomenon differed for the left and right ears, the pattern appeared in the symmetric-head test as well (Figure 7 bottom). The symmetric-head test calculates the RMS energy difference across all critical bands for a single-ear response (e.g., az -30°, left) and its symmetric counterpart (e.g., az 30°, right). This test and the corresponding metrics (Table 2) are useful for detecting anomalous behavior.
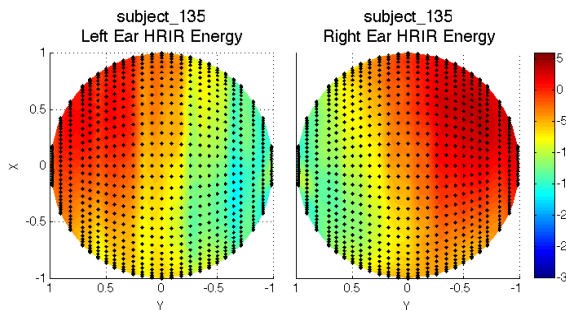

Figure 6: CIPIC HRTF energy display exhibiting a sagittal energy pattern (subject_135, HRTF index 33). Color axis in dB. The overlaid points are the CIPIC measurement grid where each sagittal line corresponds to the same speaker.

Viewing IID on a sphere can help detect subject-positioning issues. Indeed, the ITD visualization bias and curve shift patterns have counterparts in the IID visualization. ITD and IID are related in that ITDs are L/R-relative sound propagation delays and IIDs are L/R-relative energies. The key is to focus on the approximate 0dB IID region on the median plane (akin to the 0μs axis in the ITD display). The IID visualization typically spans 54dB and captures the entire IID span of the CIPIC and Listen HRTF collections (-27dB to 27dB). This span can be narrowed to 2dB (-1dB to 1dB) to highlight the region of ~0dB IID. High ITD *shift* metrics should be observable as IID sphere rotations, while high *dbias* values should yield IID y translations. The upper and lower hemispheres shifted in the same direction would indicate a y-axis subject-positioning error while opposite directions would indicate a roll error. In Figure 8 left, one can see the yaw positioning error maintained throughout the measurement since the IID pattern appears rotated by a 7.5° angle. In Figure 8 right, the IID pattern appears shifted towards the left in the upper hemisphere and shifted towards the right in the lower hemisphere. This would be consistent with a roll positioning error.
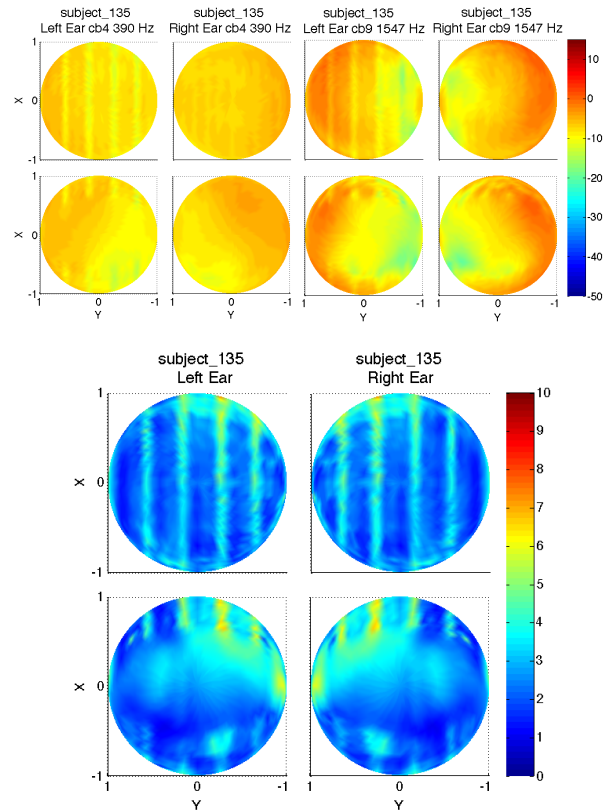

Figure 7: Top Plots: CIPIC critical-band energy display for subject_135, critical band 4 (390 Hz; left) and critical band 9 (1547 Hz; right). Bottom Plots: Symmetric-head RMS energy difference across all critical bands. The upper hemispheres are shown above the lower hemispheres. Color axes in dB.
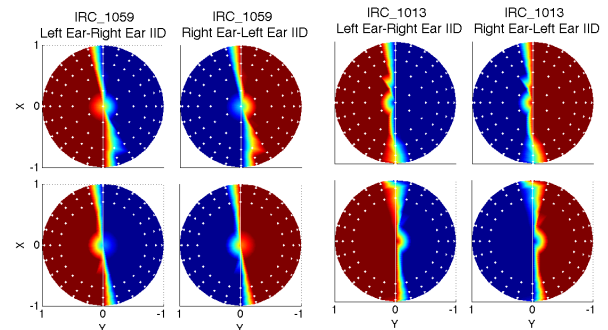

Figure 8: Listen IID visualizations exhibiting subject positioning-error patterns. Left Plot: IRC_1059 (HRTF index 51), *shift* metric = 7.5°. Right Plot: IRC_1013 (HRTF index 10), *dbias* metric = 73.8μs (*shift* 3.75°). The overlaid points are the Listen measurement grid with outer ring points 15° apart. The upper hemispheres are shown above the lower hemispheres.

Ideally, the results of HRTF analyses should be consistent across the various domains and their permutations: ITDs, frequency responses, energies, critical-band energies, visualizations, metrics, and listening. Indeed, phenomena observed in one domain are often observable in another, e.g., a shifted ITD visualization also appearing in the ITD shift

metric, a rotated IID visualization, and a shifted audio image. Of course, cross-domain consistency is also useful for verifying the metrics and tools. In addition, domain corroboration can help determine the underlying cause of the phenomenon observed, e.g., subject positioning error.

4.1.3. HRTF Metrics

As various phenomena were discovered using the visualization tools, it became clear that a more systematic approach to anomaly detection was required. As mentioned in the visualization section, some visual patterns can be quantified into metrics. Also, assuming a midsagittal-symmetrical head can yield a set of metrics quantifying asymmetry, or, potential anomaly.

Table 2 summarizes the ITD and energy metrics used in the database selection process. All energy values were calculated in the frequency domain. Frequencies outside the audible range were omitted. No scaling or frequency warping was applied. It is important to keep in mind that metrics can't capture all anomalous behavior and HRTF visualization remains a necessary and important part of the evaluation process.

Table 2: ITD and Energy Metrics.

ITD Metrics
ITD Magnitude Metrics (µs) *reject statistical outliers* **max, min**: max/min ITD values **over, under**: max/min measured-SHM magnitude difference
ITD Threshold Metrics *always positive, 0 desired, reject over threshold* **dbias** (µs), **shift** (°): see section 4.1.1
ITD Threshold Metrics *always positive, low mean expected, reject over threshold.* **mdis** (µs/°): see section 4.1.1 **masym** (µs): max symmetric-head ITD difference

Energy Metrics (dB)
Energy Magnitude Metrics *reject statistical outliers* **maxL, minL, meanL**: left-ear energy stats **maxR, minR, meanR**: right-ear energy stats **total**: mean of all energies **maxIID**: maximum database IID, absolute value
Energy Global Asymmetry Threshold Metrics *always positive, low mean expected, reject over threshold* **maxD, minD, meanD**: maxL - maxR, minL - minR, meanL - meanR, absolute value
Energy Local Asymmetry Threshold Metrics *always positive, some mean expected, reject over threshold* **maxF, minF, meanF**: symmetric-head critical-band energy RMS difference **maxE, minE, meanE**: symmetric-head energy difference

One can make physical, psychoacoustic, or statistical arguments for the various HRTF database metric rejection thresholds. In reality, all metrics should have physically based limits given morphological constraints, but the constraints and the limits can be difficult to determine. Some metrics have "expected" or "desired" values, e.g., ITD shift 0º, ITD bias

0 µs, and L-R average database energy difference near 0 dB. For this first pass at metric-based analysis, a statistical approach was used for all metric thresholds.

The following 5-step metric-based rejection procedure was developed:
1. Statistical Outlier Rejection of ITD Metrics
2. Statistical Outlier Rejection of Energy Metrics
3. ITD Thresholds
4. Energy Thresholds Omitting ITD Outliers
5. Energy Thresholds Using Minimum Collection Thresholds
The metric means (µ) and standard deviations (σ) at the conclusion of each step appear in Tables 3 and 4.

Steps 1&2: Statistical Outlier Rejection
The rejection of statistical outliers proceeded as follows:
- Each collection (Listen, CIPIC) was analyzed independently with each database initially classified as "selected" (versus "rejected").
- The means and standard deviations for the selected database metrics were determined for all metrics in Table 2.
- All databases with metrics beyond µ ± 3σ thresholds were rejected and the process repeated until the selected databases for the current iteration matched the selected databases for the previous iteration. For the energy threshold metrics, a 1dB default value was assigned if the threshold was below 1dB.

Table 3: ITD Metrics µ and σ after outlier rejection for the Listen and the CIPIC databases, steps (S) 1 and 3.

ITD Metrics	S1 Listen µ (σ)	S1 CIPIC µ (σ)	S3 Listen / CIPIC µ (σ)
max (µs)	639.7 (30.9)	645.0 (33.3)	
min (µs)	-639.2 (29.4)	-644.3 (25.0)	
over (µs)	83.4 (83.4)	83.5 (22.9)	
under (µs)	-91.6 (24.5)	-95.5 (24.1)	
shift (°)	2.5 (2.2)	1.9 (1.5)	1.2 (1.0)
dbias (µs)	43.0 (19.1)	29.6 (14.0)	26.3 (8.6)
masym (µs)	48.4 (16.0)	54.1 (18.7)	41.5 (9.5)
mdis (µs/°)	4.6 (0.8)	10.5 (4.2)	5.0 (1.1)
Outliers	28	6,9	*

*Listen: 5,7,8,10,12,13,15,19,23,24,26,27,28,29,31,32,37,38, 41,45,47,49,51.
*CIPIC: 2,3,4,6,7,9,10,11,13,15,16,17,18,19,20,22,23,25,26, 27,29,30,33,34,35,36,38,39,41,42,43,44.

Step 3: ITD Thresholds
Using the visualization tools, the observed patterns appear to indicate that the variability in ITD behavior may be due to subject positioning, subject movement, and/or ITD extraction instead of morphological variability, i.e., extreme ITD metric values may be the result of measurement and/or post-processing error. Analyses of variance (ANOVAs) were performed to compare the ITD Magnitude Metrics for the Listen and CIPIC HRTFs. Because these differences were not significant (*max*: $F_{1,92}$=.64, *p*=.42; *min*: $F_{1,92}$=.77, *p*=.38; *over*: $F_{1,92}$=.38, *p*=.53; *under*: $F_{1,92}$=.64, *p*=.42; *p* value significant at *p*<.05), the databases were combined to further refine selection based on the ITD Threshold Metrics. For each of the

ITD Threshold Metrics, a value near 0 is expected and/or desired. Thus, the databases can be sorted low-to-high for each of the four ITD Threshold Metrics. The 64 "best" databases (i.e., 2/3 of the databases) for each of the four ITD Threshold Metrics were then statistically analyzed. All databases above threshold were rejected. This algorithm excluded all rejected databases using the ITD-metric statistical outlier removal method and several additional databases. A quick visual inspection was performed to verify that the 64 "best" databases did indeed yield threshold values that caught anomalous behavior.

Step 4: Energy Thresholds Omitting ITD Outliers

The Listen and CIPIC collections exhibited somewhat unique behavior in regards to the energy metrics, e.g., the CIPIC collection exhibited significantly more variability than the Listen collection and some means differed by more than 2 dB, as seen in Table 4 (*meanL*: $F_{1,84}$=153.33, p<.0001; *meanR*: $F_{1,84}$=302.36, p<.0001; *total*: $F_{1,84}$=268.84, p<.0001; *maxIID*: $F_{1,84}$=81.27, p<.0001; p value significant at p<.05). For this reason, the merged collection approach used with the

ITD metrics was not used with the energy metrics. The energy metrics also lacked clear expected and/or desired values. If one assumes the ITD analysis in Step 3 correctly determines the databases with the least potential measurement and post-processing error, the rejected ITD databases can be used to further refine the post-outlier-rejection energy metric means and standard deviations (i.e., re-running the iterative statistical rejection process of Step 2, but combining the Step 3 ITD rejects with the energy rejects).

Step 5: Energy Thresholds Using Min Collection Thresholds

An unknown at this point was the amount of metric variability due to subject variability (e.g., morphology, hairstyle) and the amount due to potential measurement and post-processing error. Thus, the collection with the minimum variability, Listen, was used to determine the Asymmetry Metric thresholds. For Listen, these thresholds were already applied during the previous step. For CIPIC, the Listen thresholds replaced CIPIC's Asymmetry thresholds. The statistical rejection process was then repeated.

Table 4: Energy Metrics µ and σ after outlier rejection for the Listen and the CIPIC databases, steps (S) 2, 4, and 5.

Energy Metrics (dB)	S2 Listen µ (σ)	S2 CIPIC µ (σ)	S4 Listen µ (σ)	S4 CIPIC µ (σ)	S5 CIPIC µ (σ)
maxL	1.3 (0.5)	3.3 (1.0)	1.3 (0.6)	3.1 (1.1)	3.0 (0.7)
minL	-19.4 (0.9)	-19.6 (1.6)	-19.4 (0.9)	-19.7 (2.0)	-19.2 (1.6)
meanL	-6.0 (0.5)	-3.8 (1.1)	-5.9 (0.6)	-3.8 (1.1)	-3.8 (1.0)
maxR	1.2 (0.5)	3.6 (0.6)	1.2 (0.6)	3.4 (0.4)	3.2 (0.3)
minR	-19.3 (0.9)	-19.7 (1.6)	-19.2 (0.9)	-19.9 (2.1)	-19.3 (2.2)
meanR	-6.0 (0.5)	-3.4 (0.9)	-5.9 (0.6)	-3.4 (0.9)	-3.5 (0.9)
total	-6.0 (0.5)	-3.6 (0.9)	-5.9 (0.6)	-3.6 (0.9)	-3.6 (1.0)
maxIID	19.1 (0.9)	21.1 (1.2)	19.0 (0.9)	21.2 (1.2)	20.8 (1.3)
maxD	0.4 (0.3)	0.8 (0.7)	0.4 (0.3)	0.8 (0.8)	*0.4 (0.4)*
minD	0.3 (0.2)	0.7 (0.5)	0.3 (0.2)	0.7 (0.3)	*0.6 (0.2)*
meanD	0.1 (0.1)	0.8 (0.6)	0.1 (0.1)	0.6 (0.6)	*0.3 (0.3)*
maxF	5.1 (1.0)	5.9 (1.2)	4.6 (0.7)	5.1 (1.0)	*4.7 (0.9)*
minF	0.6 (0.2)	0.8 (0.2)	0.6 (0.2)	0.8 (0.2)	*0.7 (0.2)*
meanF	2.0 (0.3)	2.4 (0.4)	1.8 (0.3)	2.2 (0.4)	*2.0 (0.3)*
maxE	3.7 (1.1)	4.6 (1.3)	3.3 (0.9)	3.7 (0.9)	*3.1 (0.7)*
minE	0.006 (0.006)	0.001 (0.001)	0.005 (0.007)	0.001 (0.001)	*0.001 (0.001)*
meanE	1.0 (0.3)	1.3 (0.5)	0.9 (0.2)	1.0 (0.4)	*0.7 (0.2)*
Outliers	25,28,43	3,6,11,19,23,34,42,44	5,13,20,24,25,28,34,38,45	2,3,6,7,10,11,19,20,22,23,26,27,34,42,44	*

*CIPIC: 1,2,3,4,6,7,8,9,10,11,14,17,18,19,20,21,22,23,24,25,26,27,30,31,33,34,36,37,42,43,44.
Italics: The corresponding thresholds were initially overridden by Listen thresholds.

5. COLLECTION RESULTS

ITD visualization patterns indicative of subject-positioning error (*shift, dbias,* Table 3) were noticed in several databases. In fact, this motivated the development of the ITD metrics and the metric outlier removal process. Note, simple within-collection outlier removal (step 1) did not eliminate these databases and the use of the ITD Thresholds (step 3, Table 3) was necessary to detect and remove the most pronounced ITD patterns.

Three additional CIPIC behaviors were observed, ITD curve discontinuity, a sagittal energy pattern, and increased subject, location, and ear variability. With the ITD visualization, one notices some discontinuity in the data points, an uneven distribution or "bunching" of values relative to the Listen data (*mdis*, Table 3). Since the converted ITDs were spherically interpolated from the CIPIC az,el data grid to the slab3d grid, the ITDs were examined in CIPIC's native format. The discontinuities were noticed in the native ITD data as well. When examining CIPIC HRTF energy visualizations, some databases exhibited a sagittal energy pattern (Figures 6 and 7). Given that specific speakers were mapped to specific sagittal plane measurements, this phenomenon may be due to multi-speaker issues, e.g., speaker compensation.

Before threshold-based removal, several CIPIC databases showed symmetric-head RMS differences and total energy differences that were high relative to the Listen metrics (F and E metrics, Table 4). When examining the ITD and Energy metrics, one also notices more between-subjects metric variability than in the Listen collection (σ values, Table 4). Further, the Energy Magnitude Metrics for the CIPIC data exhibited more left-ear/right-ear energy differences than the Listen data (D metrics, Table 4). The subject, location, and ear metric variability was reflected in the ITD and energy visualizations as well.

These tools were also applied to slab3d's default HRTF database, jdm.slh, measured using the Snapshot HRTF measurement system. The Snapshot software extracts ITD as part of the measurement process. When this was originally performed, discontinuous ITDs were noticed that were hand-corrected using symmetric-head values. Viewing the ITD visualization, the jdm.slh ITDs appear to be somewhat large relative to the spherical-head model (JDM is an average head, r ≈ 0.09m). In fact, jdm.slh would be a *max, min,* and *over* outlier using the Listen and CIPIC ITD Magnitude Metric thresholds (step 1) and a *masym* outlier using the ITD Thresholds (step 3). jdm.slh would be a *minL, minR,* and *maxIID* outlier using the CIPIC energy thresholds (step 5).

6. CONCLUSIONS AND FUTURE DIRECTIONS

The Listen and CIPIC collections are excellent resources for HRTF-based spatial rendering. This paper presents the conversion of the Listen and CIPIC collections to the slab3d SLH format, visualization and metric methods used to detect conversion artifacts, and the application of these methods to the selection of a homogenous subset of databases for general use. It should be noted that the 5-step rejection procedure

used in this analysis has not yet been perceptually validated and could easily be considered overly critical. It is a first pass at the proposed method. By design, it errs on the side of caution. The goal was to select a limited consistent subset least likely to contain conversion artifacts, anomalous behavior, and measurement error.

The visualizations and metrics appear to indicate the presence of subject-positioning artifacts in both the Listen and CIPIC data. This can be expected given the great difficulty in keeping a subject stationary and properly aligned during a measurement session. In addition, the converted CIPIC databases revealed a sagittal energy pattern that may correspond to the use of multiple speakers. Relative to Listen, the CIPIC databases exhibited greater ITD discontinuity and subject, location, and ear variability, but, at this point, the causes and perceptual implications are unknown. These phenomena did, however, impact the selection procedure, resulting in fewer CIPIC databases selected than Listen. Though potentially too limited a subset, the following Listen and CIPIC HRTF databases passed metric-based selection and should yield good results with slab3d: Listen (25 of 51), HRTF indices: 1, 2, 3, 4, 6, 9, 11, 14, 16, 17, 18, 21, 22, 30, 33, 34, 35, 36, 39, 40, 42, 44, 46, 48, 50; CIPIC (6 of 45), HRTF indices: 5, 12, 28, 32, 40, 45. All Listen and CIPIC databases in the SLH format, as well as the corresponding conversion, visualization, and analysis tools, are available online [9].

The proposed metrics are quite simple and more sophisticated metrics should be possible. For example, a metric quantifying frequency and spatial trend discontinuity would be quite useful (this is presently verified using visualization). Some of these techniques could also be integrated into the HRTF measurement process itself as a form of real-time feedback to the measurer, e.g., using real-time ITD extraction and spherical-head model comparison for subject-positioning verification.

6.1. HRTF Matching and Perceptual Validation

This paper has focused on the HRIR and ITD data rather than the end application of providing spatial cues to a listener. The 5-step selection procedure, the metrics, the thresholds, and the resultant selected databases would all greatly benefit from perceptual validation and iterative refinement. For example, in a perceptual analogue to the present work, Katz and Parseihian [19] have described a subjective ranking method to select an optimal subset of HRTFs from the Listen database, followed by perceptual validation using localization performance. Similar perceptually-based approaches have been developed by Roginska and colleagues [20], [21].

Once one has a subset of preferred HRTF databases, their use can further be narrowed by selecting the best database for a given subject. There have been a variety of HRTF matching methods proposed, e.g., two-step [22], tournament [23], cepstrum reduction [24], and anthropometric [25], [26]. The slab3d utility HeadMatch allows for the rapid comparison of different HRTF databases, but it does not presently have a selection process where a large number of databases are reduced to a best-fit database in a sequence of steps. Future work will develop an HRTF matching procedure that incorporates perceptual validation.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] E. A. G. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," *J. Acoustical Soc. Am.*, 56, 1848, 1974.

[2] J. C. Middlebrooks, J. C. Makous, and D. M. Green, "Directional sensitivity of sound pressure levels in the human ear canal," *J. Acoustical Soc. Am.*, vol. 86, no. 89, 1989.

[3] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoustical Soc. Am.*, vol. 100, no. 3248, 1996.

[4] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum - phase reconstruction," *J. Acoustical Soc. Am.*, vol. 91, no. 1637, 1992.

[5] http://recherche.ircam.fr/equipes/salles/listen/index.html

[6] http://interface.ece.ucdavis.edu/sound/hrtf.html

[7] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics,* Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001, pp. 99-102.

[8] http://www.sengpielaudio.com/calculator-air.htm

[9] slab3d release, slab3d User Manual, slabtools User Manual, http://slab3d.sonisphere.com

[10] J. D. Miller and E. M. Wenzel, "Recent developments in SLAB: A software-based system for interactive spatial sound synthesis," *Int. Conf. Aud. Display,* Kyoto, Japan, 2002.

[11] J. Nam, J. Abel, and J. Smith, "A method for estimating interaural time difference for binaural synthesis," *Aud. Eng. Soc. 125th Convention* 2008, San Francisco, CA.

[12] P. Minnaar, J. Plogisties, S. K. Olesen, F. Christensen, and H. Moller, "The interaural time difference in binaural synthesis," *Aud. Eng. Soc. 108th Convention*, Paris, 2000, Preprint 5133.

[13] J. M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," *Aud. Eng. Soc. 98th Convention*, February 1995.

[14] J. Huopaniemi and J. O. Smith, "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters," *16th Int. Conf. Aud. Eng.*, Rovaniemi, Finland, April 1999.

[15] D. T. Sandwell, "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophys. Res. Letters,* 2, 1987, pp. 139-142.

[16] J. Blauert, *Spatial Hearing*, MIT press, Cambridge MA, 1974.

[17] R. Algazi, C. Avendano, and R. Duda, "Estimation of a spherical-head model from anthropometry," *J. Aud. Eng. Soc*. Vol. 49, no. 6, pp. 472-479, June 2001.

[18] Auditory Toolbox, https://engineering.purdue.edu/~malcolm/interval/1998-010

[19] B. F. G. Katz and G. Parseihian, "Perceptually based head-related transfer function database optimation," *J. Acoust. Soc. Am.* 131(2), February 2012.

[20] A. Roginska, G. H. Wakefield, and T. S. Santoro, "User selected HRTFs: Reduced complexity and improved perception," *Tec. Rep., Undersea Human Systems Integration Symposium,* 2010.

[21] K. McMullen, A. Roginska, and G. H. Wakefield, "Subjective selection of head-related transfer function (HRTFs) based on spectral coloration and interaural time differences (ITD) cues," *Aud. Eng. Soc. 133rd Convention* October 2012, San Francisco, CA, Preprint 8770.

[22] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," *Int. Conf. Aud. Display,* Boston, MA, 2003.

[23] Y. Iwaya, "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoust. Sci. & Tech.* vol. 27, no. 6, 2006.

[24] A. Honda, H. Shibata, J. Gyoba, K. Saitou, Y. Iwaya, and Y. Suzuki, "Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game," *Applied Acoustics,* vol. 68, 2007.

[25] D. Zotkin and R. Duraiswami, "Rendering localized spatial audio in a virtual auditory space," *IEEE Trans. Multimedia*, Vol. 6, No. 4, August 2004.

[26] D. Schönstein and B. F. G. Katz, "HRTF selection for binaural synthesis from a database using morphological parameters," in *Proc. of 20th Int. Cong. on Acoust.,* Sydney, Australia, August 2010, pp. 23-27.