QUANTITATIVELY VALIDATING SUBJECTIVELY SELECTED HRTFS FOR ELEVATION AND FRONT-BACK DISTINCTION

Ziqi Fan, Yunhao Wan, Kyla McMullen

University of Florida - SoundPad Lab
313A CSE Building
Gainesville, FL 32611
`[fanzq1991|yunhao|drkyla]@ufl.edu`

ABSTRACT

As 3D audio becomes more commonplace to enhance auditory environments, designers are faced with the challenge of choosing HRTFs for listeners that provide proper audio cues. Subjective selection is a low-cost alternative to expensive HRTF measurement, however little is known concerning whether the preferred HRTFs are similar or if users exhibit random behavior in this task. In addition, PCA (principal component analysis) can be used to decompose HRTFs in representative features, however little is known concerning whether the features have a relevant perceptual basis. 12 listeners completed a subjective selection experiment in which they judged the perceptual quality of 14 HRTFs in terms of elevation, and front-back distinction. PCA was used to decompose the HRTFs and create an HRTF similarity metric. The preferred HRTFs were significantly more similar to each other, the preferred and non-preferred HRTFs were significantly less similar to each other, and in the case of front-back distinction the non-preferred HRTFs were significantly more similar to each other.

1. INTRODUCTION

3D audio is used in many settings to augment a wide variety of tasks including improving immersion in virtual reality, enhancing speech intelligibility, improving video games, providing spatial cues for assistive technology for persons with visual impairments, enriching positional systems for air traffic controllers, and sonifying multidimensional data. In each of these scenarios, 3D audio minimizes cognitive load, provides spatial cues for degraded visual environments, and provides information redundancy, which significantly improves task performance [1, 2, 3].

3D audio cues are most effectively realized through the use of head-related transfer functions (HRTFs). Proper perceptual fidelity in virtual auditory environments requires the use of individualized or customized HRTFs. The use of a generic or non-individualized HRTF leads to poor elevation perception, decreased externalization, and increased front/back reversal errors [4, 5, 6, 7, 8].

The most accurate HRTFs are obtained by direct acoustic measurement, however the measurement process is very expensive in terms of time and resources [4, 9, 10, 11, 5]. By placing a lis-

tener in an anechoic chamber and positioning a loudspeaker at a known location, it is possible to measure the entire acoustic transformation of a sound by the listener's body. The frequency response is recorded at each ear with microphones placed in the ear canal. When used to synthesize virtual auditory sources, HRTFs are typically realized as the cascade of a minimum-phase FIR filter and an all-pass filter that accounts for the lag in the wavefront arrival time between the two ears [8, 12]. Directly measured HRTFs provide an individualistic 3D sound experience for each person, according to their specific anthropometic features. Although this direct measurement may produce the most accurate measurement, perhaps such costly and resource intensive measurements are not completely necessary to convey 3D sound to a listener.

As a solution, many researchers have proposed less costly methods to alleviate the need for individualization that have been met with varying levels of success. These methods include: HRTF approximation using theoretical computation [13, 14, 15, 16, 17], active sensory tuning [18, 19], machine learning [20, 21, 22, 23, 24], genetic algorithms [25, 26, 27], clustering [28, 29, 30, 31], generic models[4, 5, 32, 7], physical feature measurement[33, 34, 35, 36, 37, 38, 39], pre-measured [40, 41, 10], and subjective selection from pre-measured HRTF databases [42, 43, 44, 45, 46].

Many quantitative and qualitative metrics have been proposed to analyze HRTF similarity [47, 48, 49, 28, 23]. Principal Component Analysis or PCA arises as an objective method to use the common features of an HRTF to decompose it into features that can be varied. PCA is used to describe a data set by using only a few orthogonal components and corresponding weights. For example, Martens was one of the first researchers to show that variations in an HRTF's spectral energy distribution with changing azimuth could be adequately captured by four principal components, quantified in terms of spectral band weighting functions. This finding simplified HRTF analysis by providing a simple (4D) measure of global spectral variation which until then was otherwise difficult to quantify [50]. Following this work, PCA has been used my many researchers to decompose the HRTF.

Though many researchers have used PCA and other decomposition tools to define features and (sometimes) measure HRTF similarity, many if not all of these approaches neglect to address the perceptual validity of the features used to represent HRTFs. Even if the exact mathematical relationships had been discovered,is no perceptual basis to prove that the extracted features affect 3D audio perception. The most straightforward method to determine a listener's preference for HRTFs and how it impacts their 3D audio perception is through a subjective selection procedure in which a listener can choose useful HRTFs from a database of pre-measured

filters. This approach is an effective yet inexpensive method to obtain a listener's HRTF preference, however, little is known about the similarities of the subjectively selected HRTFs. For example, Schonstein and Katz [51] found large variances in judgments for listeners perceptually evaluating HRTFs. The next logical step in this line of research is to analyze the subjectively selected HRTFs chosen by a listener to discover if similarities exist that can be quantified.

The present work uses PCA metrics to decompose a set of pre-measured HRTFs and uses clustering to group them. A given set of listeners perform a subjective selection task in which they indicate their HRTF preferences. Following this, a difference analysis is used to measure HRTF similarity of preferred HRTFs as compared to not-preferred HRTFs. If similarity is observed, this will achieve our goals of (1) demonstrating that the mathematical decomposition of HRTFs using PCA actually has perceptual validity and (2) validating subjective selection as an appropriate method for naive listeners to customize their listening experience. If this work proves to be successful, instead of using the complete HRTF, 3D audio designers could build simplified HRTF models based on the relevant features extracted.

2. BACKGROUND

When clustering a large multidimensional dataset, among the first factors to decide upon is an appropriate metric to use when representing HRTFs. Bondu et. al [28] analyzed several different criteria for clustering HRTFs and discovered that the Avendano criterion performed best both in terms of localization performance and clustering. This would be ideal for our usage, however, Avendano criterion is restricted to the frequency domain. This means that by using this metric, any time domain information would be omitted from the analysis. The present work employs a metric that combines both time and frequency domain information for clustering. The method used in the present work is described in the following sub-sections.

2.1. HRTF Decomposition & Clustering

2.1.1. Principal Component Analysis

PCA is a data reduction method that is used to describe the relevant features of the HRTF. In his seminal study, Martens [50] proposed a method of using spectral band energy to analyze the statistical features of HRTFs. In his study, the spectral area of an HRTF was decomposed into 24 sub-bands centered with 24 frequencies from low to high spectral area. Each HRTF was transformed into a 24 element vector. Then, each vector corresponding to each HRTF was grouped together into a matrix, whose rows corresponded to observations and columns corresponded to frequency elements. Lastly, PCA was performed on the formed matrix to decompose the observations into the combinations of different bases. Based on the study of the shape of the bases, Martens found that the energy bands of frequency should be regrouped into four sets, which are grouped in the following table.

2.1.2. HRTF Clustering

After decomposing an HRTF into relevant features, a *k-Means* clustering algorithm can be used to partitions data into k mutually exclusive clusters based on their distance to the centroid of a cluster. The algorithm forms groupings or clusters in such a way that data within a cluster have a higher measure of similarity than data in any other cluster. The measure of similarity on which the clusters are modeled is defined by Squared Euclidean metric. The *k-Means* algorithm treats each observation in the dataset as an object in a specific location in space. The partition found in the algorithm ensures that objects in the cluster are as close to each other as possible and as far from other objects in other clusters as possible. The k centers of the clusters are initialized through Arthur & Vassilvitskii's algorithm [52] and then an iterative algorithm is used to minimize the sum of the distances from each object to its cluster centroid, for all clusters. The algorithm moves objects between clusters until the sum cannot be further minimized. The algorithm runs as follows:

Given cluster number k and a set of n data points χ.

1. Randomly choose one center c_1 from χ

2. Take a new center c_i, choosing $x \in \chi$ with probability $\frac{D(x)^2}{\sum_{x \in \chi} D(x)^2}$

3. Repeat step 2 until k centers $C = \{c_1, \ldots, c_k\}$ have been taken together

4. For each $i, j \in \{1,...,k\}$, cluster $C_i \in \chi, x_n \in C_i$, if $\|x_n - c_i\| \leq \|x_n - c_j\|$, for all $j \neq i$.

5. For each $i \in \{1,...,k\}$, set c_i to be the center of mass of all points in C_i: $c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$

6. Iterate steps 4 and 5 until C no longer changes

2.2. HRTF Similarity

Once the HRTFs have been decomposed into their relevant features, an HRTF similarity metric can be employed. The similarity for different sets of HRTFs, i.e., preferred, non-preferred, is described by the distance between the average energy for all relevant HRTF azimuths and elevations, as represented in Equation 1,

$$distance_j = \frac{1}{N} \sum_{i=1}^{N} (E_{j1i} - E_{j2i})^T (E_{j1i} - E_{j2i}) \quad (1)$$

where the subscripts 1 and 2 denote 2 different sets of HRTFs to be compared, N denotes the total number of pairs of HRTFs from the total HRTF database, and i represents all of the HRTF directions for two different HRTF sets. Letter j can be defined to be either left or right ears, E represents the 4 element energy vector, derived from the PCA described in Section 2.1.1. Thus, Equation 1 provides a metric to quantify distance between different sets of HRTFs that can be generalized for any analysis.

It should be noted that since HRTFs consist of both left and right ear filters, we must first redefine the energy vector as the ratio between left and right ears:

$$E = log \frac{E_l}{E_r} = log E_l - log E_r \quad (2)$$

Group1	166,282,410,543,681,825,980,1158,1368,1616
Group2	1909,2255,2664,3146,3716,4390,5185
Group3	6125,7235,8545,10094,11923
Group4	14083,16634

Table 1: Frequency band grouping, from Martens [50]

After this step, Equation 1 was used to calculate HRTF similarity.

2.3. Spectral Notch Analysis

The locations of spectral notches within HRTFs convey 3D audio cues that allow listeners to distinguish sounds rendered in the front from sounds rendered in the back and provides and provides elevation cues that allow a listener to determine the height of a spatialized sound source [53, 54]. Notches are characterized by the minimal value points in HRTF magnitude response. To determine the locations of spectral notches within an HRTF, a fast Fourier transform is performed on the head-related impulse response (HRIR). Then, every point in the magnitude response is compared with its adjacent points.

3. METHOD

3.1. Subjective Selection Experiment

The subjective selection experiment methodology used in the present study is thoroughly outlined in in Wan et al. (2015) [55] and summarized in this section.

3.1.1. Subjects

12 subjects (5 male and 7 female) with normal hearing were recruited to participate in the study. Participants were either full-time students or undergraduate summer research students at Clemson University. It should be noted that 3 subjects' performance classified them as outliers. This was because they either (1) preferred all of the HRTFs in the dataset all of the time or (2) Never picked an HRTF consistently. Accordingly, their data was removed from the analysis.

3.1.2. Stimuli

The stimulus used in the study was a 500-ms infrapitch noise, which was constructed by creating pink noise with 200-ms sampling that was repeated 2.5 times.

14 HRTF datasets were used. 13 were from the CIPIC database[9], and one KEMAR dataset. For each trial, the program created impulse responses for the left and right ears by cascading the minimum-phase head-related impulse responses, which were drawn from a given set of HRTFs, with the all-pass impulse responses for the KEMAR ITD.

3.1.3. Procedure

All participants were randomly divided into two groups, *A* and *B*, which performed the experiment in different order. In each session, subjects in *Group A* went through a three-stage listening procedure in which they evaluated the perceptual criteria in the following order: externalization, elevation discrimination and front/back discrimination. Listeners in *Group B* evaluated the perceptual criteria in the following order: externalization, front/back discrimination, elevation discrimination. In each stage, the participants judged each HRTF's ability to render the given perceptual cue. Each three stage procedure constituted an experimental session. Each subject completed 3 experimental sessions, occurring between 1 and 2 days apart. To eliminate any potential stimuli judgment bias, subjects were unknowingly deceived by being told
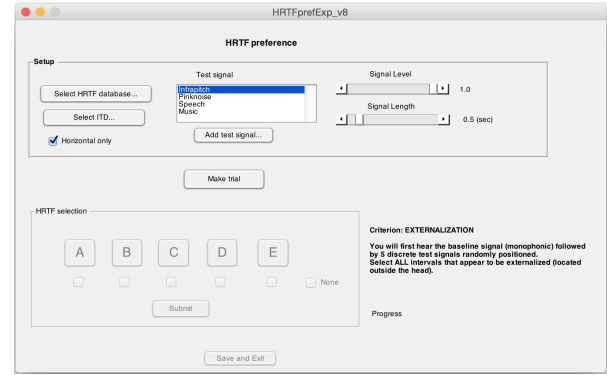

Figure 1: User interface for the spectral coloration selection tasks.

that they would be judging 3 different sets of sounds on each day, when in fact the same HRTFs were used for each session.

The experiment's user interface is shown in **Figure 1**. In each stage, the HRTFs were randomly ordered and presented 4 times per HRTF. In each interval (as shown in Figure 1), 5 HRTFs (marked as letters A, B, C, D, and E) rendered 3D sound. The HRTF corresponding to each letter was highlighted as it was used. The participant could replay any option by clicking its letter. There was a check-box below each HRTF letter that allowed the listener to select whether the HRTF that was used provided adequate cues for the given stage.

In the case of *front/back discrimination*, each trial started with an unspatialized (monaural) reference signal, which was generated by processing the test signal with the HRTF at 0° azimuth, 0° elevation and cross-summing the left and right channels. The purpose for this step was to avoid spectral coloration variability between the raw and the processed test signal. This signal was used as an in-the-head reference, to which the HRTF rendered sounds could be compared. Following the reference signal, the listener heard five consecutive sounds, generated from randomly selected HRTFs, at randomly selected azimuths ($\pm 150°, \pm 120°, \pm 60°, \pm 30°$) on the horizontal (ear level) plane. All of the intervals in each trial used the same sequence of azimuths. If none of the intervals gave the perception of externalization, the listener selected the "None" check-box. After submitting the selections, the results were saved, and the listener proceeded to next trial. The *elevation discrimination* phase proceeded almost identically to the *front/back discrimination* phase. The only difference was that each interval was rendered using a selected HRTF at a random azimuth from: $\pm 150°, \pm 120°, \pm 90°, \pm 60°, \pm 30°$; at elevations of $\pm 36°$. The *externalization* phase required the listener to judge the externalization of each interval.

4. RESULTS

4.1. Evaluation Metrics

As described in Section 2.1.1, PCA was performed on each of the CIPIC HRTFs. After band pass filtering the HRTF with center frequencies given in the 1, the energy in the four groups were calculated thus forming a four elements feature vector for each HRTF. PCA is then repeated on the newly derived energy vectors to analyze the energy of the HRTFs.

As described in section 2.1.2, HRTF clustering was also performed to observe any HRTF groupings. Each HRTF's left and right ear energy vectors for all directions (Table 1) were used for clustering. HRTF similarity was quantified using the metrics described in 2.2.

The metric described in Section 2.3 was used to determine the locations of the spectral notches to evaluate its position and magnitude in each band group. The manner in which to locate the spectral notches was slightly revised due to the fact that CIPIC's own post-processing of the raw data uses a Hanning window to remove any room reflections [9]. This processing is problematic because the high-frequency components of the HRTF are filtered away. It is for this reason that the notch in the last energy band group was purposefully omitted.

In order to analyze the time dependent HRTF features, the ITD information was extracted and plotted with the notch position to further assess similarity. In these analyses, the notch positions and ITDs are scaled such that their values are between -1 and 1 inclusive. All of the aforementioned metrics in this section were used to assess whether the preferred HRTFs for elevation and front/back distinction are related to their energy features, spectral notch location, or time-dependent features.

4.2. Subjective Selection Preferences

Figure 2 summarizes the results of the experiment described in Wan et al. (2015) [55]. Figures 3 through 7 show the results of the present study. It should be noted that 3 of the original twelve subjects were considered outliers because they did not demonstrate selective behavior. These subjects either did not pick any HRTFs consistently, or they picked almost all of the HRTFs all of the time. Thus, their data was not chosen to be included in the analysis. In addition, the present analysis focuses on the *front/back* and *elevation* distinction stages of the experiment since these stages rely heavily on the listener's ability to discriminate HRTF spectral features.

4.3. Spectral Similarity of Chosen HRTFs

4.3.1. Elevation Distinction

Figure 3 shows the clustering of the HRTFs used during the elevation distinction stage. The vectors for clustering are the spectral band energy based on the band group in table 1. Each axis represents a frequency band. In the analysis, band groups 2, 3 and 4 are used for clustering. The 4 asterisks in the plot are the centroid of four clusters. The various colors indicate the cluster in which a particular HRTF belongs. The text in the plot represents each of the HRTF datasets used in the analysis. The symbol 'k' represents the KEMAR dataset and the numbers '1' through '13' represent the CIPIC HRTF datasets that were used in the experiment, in numeric order. In this figure, the HRTFs preferred in all 3 sessions by Subject 2 are displayed more prominently than the HRTFs that were not preferred.

Table 2 displays the similarity between the preferred and non-preferred HRTFs chosen in the elevation distinction stage (as compared to the average HRTF similarity) for each subject as calculated according to the metric described in Section 2.2. A small value indicates a higher degree of similarity and a larger value indicates dissimilarity. A dash ('-') symbol in the table indicates that a listener did not have a specific preference. An '0' in the
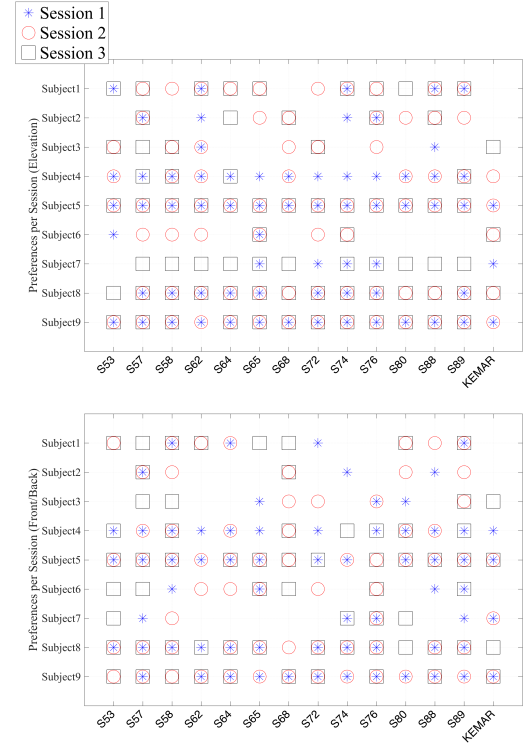

Figure 2: HRTFs chosen by each subject at the end of each session. Along the abscissa is the HRTF identifier and along the ordinate is thee subject identifier

table occurs if the listener only consistently preferred one HRTF, and therefore the difference between the HRTFs is 0.

The aggregated results are summarized in Figure 4.

Figure 4 compares the mean distances between preferred HRTFs, non-preferred HRTFs, and the distances between those chosen and not chosen, as compared to the average distance between HRTFs, displayed by the horizontal line in the plot. The results presented were assessed with an ANOVA. Error bars in all figures indicate 95 % confidence intervals. Overall, the preferred HRTFs were significantly more similar than the non-preferred HRTFs. Furthermore, the preferred and non-preferred HRTFs were significantly less similar than the average HRTF similarity $[F_{3,113} = 3.36, p < 0.05]$. The similarity score for the average HRTF group is significantly higher than that for the PvsNP group, suggesting that, for the elevation dimension, PvsNP were more similar than the average HRTFs.

4.3.2. Front/Back Distinction

In a similar fashion as Figure 3, Figure 5 shows the HRTF clustering for the HRTFs used during the front/back distinction stage. In this figure, the HRTFs preferred in all 3 sessions by Subject 1 are displayed more prominently than the HRTFs that were not preferred.

In a similar fashion as Table 2, Table 3 displays the similarity between the preferred and non-preferred HRTFs chosen in the front/back distinction stage (as compared to the average HRTF
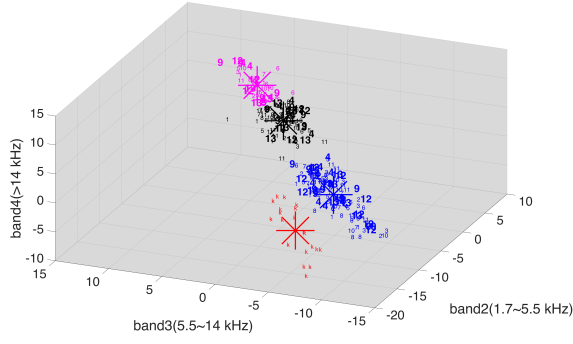
Figure 3: Clustering of the HRTFs used in the elevation distinction stage. Along each axis is the frequency band group. The preferred HRTFs chosen by Subject 2 are highlighted for comparison. The four different colors represents 4 HRTF clusters.

	Preferred	*Non-preferred*	*Average*	*PvsNP*
Subject1	2.01	2.48	4.91	6.03
Subject2	0	4.01	4.91	4.66
Subject3	-	4.70	4.91	4.93
Subject4	1.48	-	4.91	-
Subject5	4.35	-	4.91	-
Subject6	0	3.88	4.91	6.71
Subject7	0	4.03	4.91	7.20
Subject8	5.0	-	4.91	-
Subject9	4.41	-	4.91	-

Table 2: Similarity of preferred and non-preferred HRTFs by subject in elevation distinction stage.

Figure 5: Clustering of the HRTFs used in the front/back distinction stage. Along each axis is the frequency band group. The preferred HRTFs chosen by Subject 1 are highlighted for comparison.

	Preferred	*Non-preferred*	*Average*	*PvsNP*
Subject1	1.75	1.76	4.52	3.73
Subject2	0	3.51	4.53	7.46
Subject3	-	2.52	4.53	3.81
Subject4	2.02	-	4.53	-
Subject5	3.99	-	4.53	-
Subject6	0	1.99	4.53	4.44
Subject7	0	4.28	4.53	4.70
Subject8	4.22	-	4.53	-
Subject9	4.92	-	4.53	-

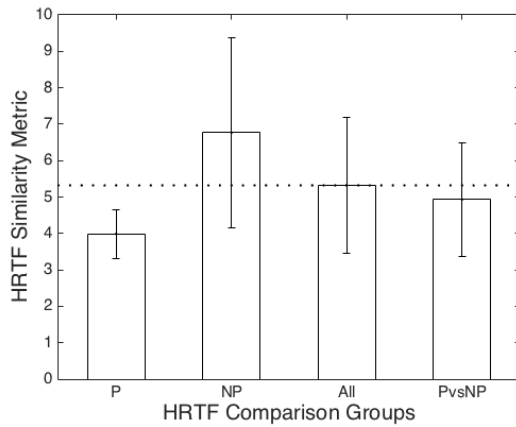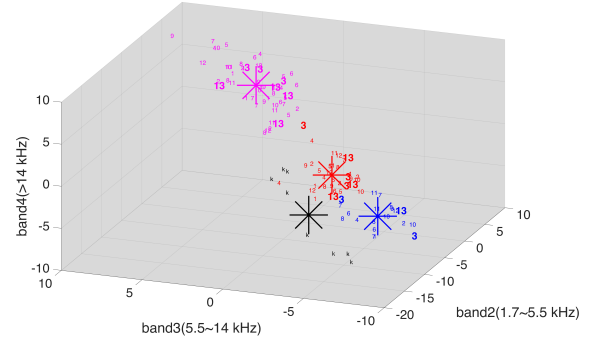Table 3: Similarity of preferred and non-preferred HRTFs by subject in front/back distinction stage.

Figure 4: Similarity of preferred (P), non-preferred (NP), All, and preferred to non-preferred (P vs NP) HRTFs used for elevation distinction. Along the abscissa are the comparison groups and along the ordinate is the similarity metric score. Lower values indicate more similarity.
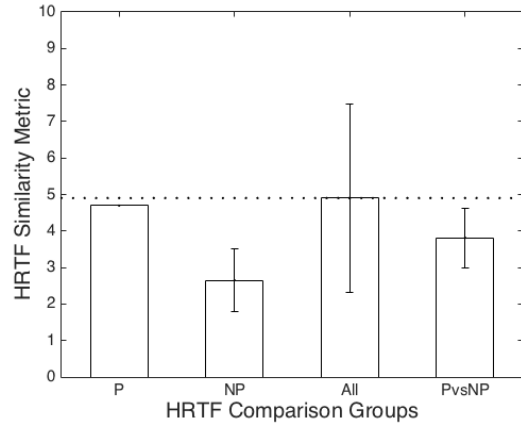

Figure 6: Similarity of preferred (P), non-preferred (NP), All, and preferred to non-preferred (P vs NP) HRTFs for front/back distinction. Along the abscissa are the comparison groups and along the ordinate is the similarity metric score. Lower values indicate more similarity.
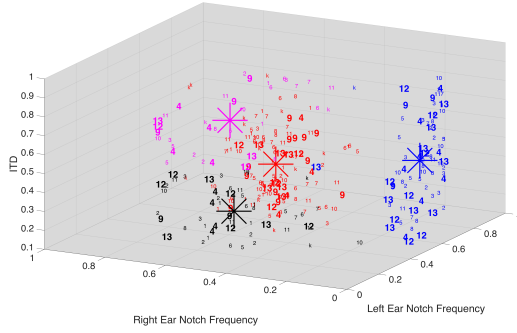
Figure 7: Clustering of HRTF notch locations and ITD for the elevation distinction stage. Along the x-axis is the location of the left ear notch frequency, along the y-axis is the location of the right ear notch frequency, and along the z-axis is ITD. The preferred HRTFs chosen by Subject 2 are highlighted for comparison.
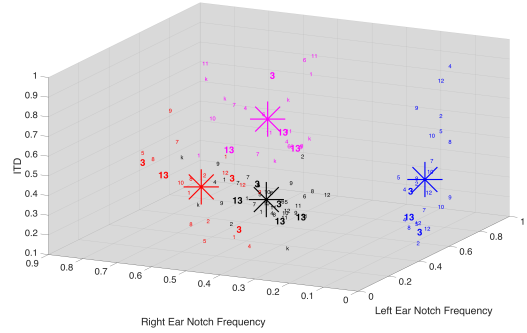

Figure 8: Clustering of HRTF notch locations and ITD for the front/back distinction stage. Along the x-axis is the location of the left ear notch frequency, along the y-axis is the location of the right ear notch frequency, and along the z-axis is ITD. The preferred HRTFs chosen by Subject 2 are highlighted for comparison.

similarity) for each subject as calculated according to the metric described in Section 2.2.

Figure 6 compares the mean distances between preferred HRTFs, non-preferred HRTFs, and the distances between those chosen and not chosen, as compared to the average distance between HRTFs, displayed by the horizontal line in the plot. The results presented were assessed with an ANOVA. Error bars in all figures indicate 95 % confidence intervals. Similar to the results displayed in Figure 4, for each subject the preferred HRTFs were significantly more similar than the average similarity. The non-preferred HRTFs were also found to be significantly more similar than the average similarity [$F_{3,98} = 3.09, p < 0.05$]. What's more, with a comparison between non-preferred and preferred, we find that the non-preferred is more similar than the preferred HRTFs.

4.4. Time Dependent Similarity of Chosen HRTFs

4.4.1. Elevation Distinction

Figure 7 shows the results of the HRTF notches and ITD clustering for the HRTFs used during the elevation distinction stage. The axes on horizontal(xy) plane are locations of HRTF notches for left(x-axis) and right(y-axis) ears. The frequency locations of all notch points have been normalized based on a maximum frequency of 22,050 Hz. On the vertical axis(z axis), the ITD is normalized to scale [0,1]. In this figure, Subject 2's preferred HRTFs are highlighted. HRTF similarity was assessed using the ITD and notch features and no significant differences were found. [$F_{3,113} = 0.18, p = 0.91$]

4.4.2. Front/Back Distinction

Similarly, Figure 8 shows the results of the HRTF notches and ITD clustering for the HRTFs used during the elevation distinction stage. In this figure, Subject 2's preferred HRTFs are highlighted. HRTF similarity was assessed using the ITD and notch features and no significant differences were found. F statistics and p value are given as[$F_{3,98} = 1.72, p = 0.17$]

5. DISCUSSION & FUTURE WORK

The goal of the present work was to determine if the PCA features on which HRTFs are clustered have a perceptual linkage and to quantitatively validate the HRTFs chosen in a subjective selection procedure.

In the work we found that HRTFs that were picked were similar in terms of distance from each other. This suggests that PCA decomposition is indeed a valid tool that has a perceptual significance when analyzing HRTFs. This work further validates the subjective selection methodology by showing that there is similarity between chosen HRTFs and novice listeners are capable of using spectral cues to discriminate between features.

Furthermore, the present work shows that HRTFs that were not selected also share similar qualities as they are typically grouped near each other.

The present work highlights the challenge of accurately decomposing an HRTF such that it can be represented as a set of points in space. In creating the current work, many solutions were tried, however it was found that the most informative HRTF representation centered around band energy. This finding suggests that the clusters can be better separated by only considering one direction.

In addition, preferred HRTFs were not significantly more similar than non-preferred HRTFs or as compared to average HRTF similarity. This suggests that the time-dependent features of the HRTF are not as critical in elevation and front/back distinction as the spectral features.

Future work will further delve into the findings and perform a subjective selection experiment in which the user hears a broadband sound coming from a known location, and HRTF features such as notch locations. spectral band energy, and ITDs are manipulated in real-time to interactively tune the HRTF. This experiment will allow us to narrow down on the exact cues that are relevant for 3D audio perception.

Currently, statistical methods of HRTF analysis are frequently limited by the capacity of data set. Up to now, the largest existing publicly available database is CIPIC, which consists of 45 HRTF data sets. Although this dataset is sufficient for many applications, the number of measured HRTFs are not sufficient enough

to perform a thorough analysis of HRTF statistical patterns. Thus, future work will involve a larger HRTF database that is formed by exploring the acoustical basis of HRTFs (as spherical waves) and using acoustic models to create more HRTFs and unify all existing public domain HRTF databases.

6. REFERENCES

[1] R. Bastide, D. Navarre, P. Palanque, A. Schyn, and P. Dragicevic, "A model-based approach for real-time embedded multimodal systems in military aircrafts," in *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 2004, pp. 243–250.

[2] J.-O. Nilsson, C. Schuldt, and P. Handel, "Voice radio communication, pedestrian localization, and the tactical use of 3d audio," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on*. IEEE, 2013, pp. 1–6.

[3] V. Y. Nguyen, "Audible assistance on escaping to an emergency exit: A comparison of a manipulated 2d and a 3d audio model," 2004.

[4] A. W. Bronkhorst, "Localization of real and virtual sound sources," *The Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2542–2553, 1995.

[5] H. Moller, M. F. Sorensen, C. B. Jensen, and D. Hammershoi, "Binaural technique: do we need individual recordings," *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

[6] S. Weinrich, "The problem of front-back localization in binaural hearing." *Scandinavian Audiology. Supplementum*, vol. 15, pp. 135–145, 1981.

[7] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using non-individualized head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[8] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited." *Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9035397

[9] 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.

[10] W. G. Gardner and K. D. Martin, "Hrtf measurements of a kemar," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3907–3908, 1995.

[11] D. Hammershoi, H. Moller, M. F. Sorensen, and K. Larsen, "Head–related transfer functions: Measurements on 24 subjects," in *92nd Audio Engineering Society Convention*, 1992.

[12] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. part 1: Stimulus synthesis," *Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 858–867, February 1989.

[13] V. Algazi, R. Duda, and D. Thompson, "Use of head and torso methods for improved spatial sound synthesis," *Proceeding of AES 113th Convention*, 2002.

[14] J. Huopaniemi, N. Zacharov, and M. Karjalainen, "Objective and subjective evaluation of head-related transfer function filter design," *Journal of the Audio Engineering Society*, vol. 47, no. 4, pp. 218–239, 1999.

[15] A. Meshram, R. Mehra, and D. Manocha, "Efficient hrtf computation using adaptive rectangular decomposition," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Aug 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17366

[16] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Frahm, and D. Manocha, "P-hrtf: Efficient personalized hrtf computation for high-fidelity spatial sound," *IEEE International Symposium on Mixed and Augmented Reality*, 2014.

[17] K. Terai and I. Kakuhari, "Hrtf calculation with less influence from 3-d modeling error: Making a physical human head model from geometric 3-d data," *Acoustical Science and Technology*, vol. 24, no. 5, pp. 333–334, 2003.

[18] P. Runkle, A. Yendiki, and G. H. Wakefield, "Active sensory tuning for immersive spatialized audio," *Proceeding of International Conference on Auditory Display*, 2000.

[19] A. Silzle, "Selection and tuning of hrtfs," in *Audio Engineering Society Convention 112*. Audio Engineering Society, 2002.

[20] H. Hu, L. Zhou, H. Ma, and Z. Wu, "Hrtf personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2, pp. 163–172, 2008.

[21] Y. Luo, "Fast numerical and machine learning algorithms for spatial audio reproduction," 2014.

[22] S. Morioka, I. Nambu, S. Yano, H. Hokari, and Y. Wada, "Adaptive modeling of hrtfs based on reinforcement learning," in *Neural Information Processing*. Springer, 2012, pp. 423–430.

[23] E. S. Schwenker and G. D. Romigh, "An evolutionary algorithm approach to customization of non-individualized head related transfer functions," in *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[24] M. Washizu, S. Morioka, I. Nambu, S. Yano, H. Hokari, and Y. Wada, "Improving the localization accuracy of virtual sound source through reinforcement learning," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4378–4383.

[25] N.-M. Cheung and S. Trautman, "Genetic algorithm approach to head-related transfer functions modeling in 3-d sound system," in *Multimedia Signal Processing, 1997., IEEE First Workshop on*. IEEE, 1997, pp. 83–88.

[26] E. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: pole-zero approximations of hrtfs," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 1, pp. 18–27, 2002.

[27] X. Hu, Y. Tang, and J. Cai, "Optimal approximation of head-related transfor function based on adaptive genetic algorithm," in *Neural Networks and Signal Processing, 2008 International Conference on*. IEEE, 2008, pp. 129–132.

[28] A. Bondu, S. Busson, V. Lemaire, and R. Nicol, "Looking for a relevant similarity criterion for hrtf clustering: a comparative study," in *Audio Engineering Society Convention 120*. Audio Engineering Society, 2006.

[29] C.-S. Fahn and Y.-C. Lo, "On the clustering of head-related transfer functions used for 3-d sound localization," *J. Inf. Sci. Eng.*, vol. 19, no. 1, pp. 141–157, 2003.

[30] M. Neal and M. C. Vigeant, "Use of k-means clustering analysis to select representative head related transfer functions for use in subjective studies," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 2366–2366, 2014.

[31] S. Shimada, N. Hayashi, and S. Hayashi, "A clustering method for sound localization transfer functions," *Journal of the Audio Engineering Society*, vol. 42, no. 7/8, pp. 577–584, 1994.

[32] D. Schonstein and B. F. Katz, "Hrtf selection for binaural synthesis from a database using morphological parameters," in *International Congress on Acoustics (ICA)*, 2010.

[33] N. A. Gumerov, R. Duraiswami, and Z. Tang, "Numerical study of the influence of the torso on the hrtf," *Acoustics Speech and Signal Processing 2002 Proceedings ICASSP 02 IEEE International Conference on*, vol. 2, pp. 1965–1968, 2002. [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=1006155

[34] N. Inoue, T. Kimura, T. Nishino, K. Itou, and K. Takeda, "Evaluation of hrtfs estimated using physical features," *Acoustical Science And Technology*, vol. 26, no. 5, pp. 453–455, 2005. [Online]. Available: http://joi.jlc.jst.go.jp/JST.JSTAGE/ast/26.453?from=CrossRef

[35] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile, *Enabling individualized virtual auditory space using morphological measurements*. IEEE PacificRim Conference on Multimedia, 2000, pp. 235–238. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.5338&rep=rep1&type=pdf

[36] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of hrtfs on the horizontal plane using physical features," *Applied Acoustics*, vol. 68, no. 8, pp. 897–908, 2007.

[37] M. Rothbucher, T. Habigt, J. Habigt, T. Riedmaier, and K. Diepold, "Measuring anthropometric data for hrtf personalization," in *Signal-Image Technology and Internet-Based Systems (SITIS), 2010 Sixth International Conference on*. IEEE, 2010, pp. 102–106.

[38] D. Zotkin, R. Duraiswami, L. Davis, A. Mohan, and V. Raykar, "Virtual audio system customization using visual matching of ear parameters," *Pattern Recognition 2002 Proceedings 16th International Conference on*, vol. 3, pp. 1003 — 1006 vol.3, 2002. [Online]. Available: http://dx.doi.org/10.1109/ICPR.2002.1048207

[39] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, *HRTF personalization using anthropometric measurements*. Ieee, 2003, pp. 157–160. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1285855

[40] V. Algazi and R. Duda, "Estimation of a spherical-head model from anthropometry," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 472–478, 2001.

[41] E. Blanco-Martin, S. Merino Saez-Miera, J. J. Gomez-Alfageme, and L. I. Ortiz-Berenguer, "Repeatability of localization cues in hrtf data bases," in *Audio Engineering Society Convention 130*, May 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15888

[42] K. McMullen, A. Roginska, and G. H. Wakefield, "Subjective selection of head-related transfer functions (hrtf) based on spectral coloration and interaural time differences (itd) cues," in *Audio Engineering Society Convention 133*. Audio Engineering Society, 2012.

[43] *User Selected HRTFs: Reduced Complexity and Improved Perception*. Undersea Human Systems Integration, 2010.

[44] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," *Proceeding of ICAD 2003*, pp. 259–262, 2003.

[45] Y. Wan, A. Zare, and K. McMullen, "Evaluating the consistency of subjectively selected head-related transfer functions (hrtfs) over time," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Aug 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17349

[46] B. Xie, C. Zhang, and X. Zhong, "A cluster and subjective selection-based hrtf customization scheme for improving binaural reproduction of 5.1 channel surround sound," in *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[47] A. Andreopoulou, A. Rogińska, and H. Mohanraj, "Analysis of the spectral variations in repeated head-related transfer function measurements," 2013.

[48] A. Andreopoulou and A. Roginska, *Evaluating HRTF Similarity through Subjective Assessments: Factors that can Affect Judgment*. Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2014.

[49] A. Andreopoulou and B. F. Katz, "On the use of subjective htrf evaluations for creating global perceptual similarity metrics of assessors and assessees," in *Proceedings of the 21st International Conference on Auditory Display (ICAD 2015)*, 2015.

[50] W. L. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," 1987.

[51] D. Schönstein and B. F. Katz, "Variability in perceptual evaluation of hrtfs," *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 783–793, 2012.

[52] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[53] V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of hrtfs," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 103–106.

[54] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[55] Y. Wan, Z. Fan, and K. A. McMullen, "Temporal reliability of subjectively selected head-related transfer functions (hrtfs) in a non-eliminating discrimination task," in *Audio Engineering Society 138th Convention*, 2015.