Audio Engineering Society

Convention Paper

Presented at the 133rd Convention
2012 October 26–29 San Francisco, USA

Subjective Selection of Head-Related Transfer Functions (HRTFs) based on Spectral Coloration and Interaural Time Differences (ITD) Cues

Kyla McMullen[1], Agnieszka Roginska[2], and Gregory H. Wakefield[1]

[1] *Electrical Engineering and Computer Science Department, The University of Michigan, Ann Arbor, MI 48109*

[2] *Music and Audio Research Lab, New York University, 35 West 4th St, New York, NY 10012*

Correspondence should be addressed to Kyla McMullen (kyla@umich.edu)

ABSTRACT

The present study describes an HRTF subjective individualization procedure in which a listener selects from a database those HRTFs that pass several perceptual criteria. Earlier work has demonstrated that listeners are as likely to select a database HRTF as their own when judging externalization, elevation, and front/back discriminability. The procedure employed in this original study requires individually measured ITDs. The present study modifies the original procedure so that individually measured ITDs are unnecessary. Specifically, a standardized ITD is used, in place of the listener's ITD, to identify those database minimum-phase HRTFs with desirable perceptual properties. The selection procedure is then repeated for one of the preferred minimum-phase HRTFs and searches over a database of ITDs. Consistent with the original study, listeners prefer a small subset of HRTFs; in contrast, while individual listeners show clear preferences for some ITDs over others, no small subset of ITDs appears to satisfy all listeners.

1. INTRODUCTION

The utility of headphone-based spatial virtual auditory environments depends on the degree to which externalization, elevation, and front-back differentiation of sources in the soundscape are achieved by

the spatial-audio system. Useful perceptual cues are introduced naturally by convolving a source with a pair of head-related impulse responses (HRIRs). However, the use of standard HRIRs obtained from acoustic manikins is often associated with decreased externalization and degraded localization (1; 2; 3) presumably because of a mismatch between the cues and the those expected by the listener. Accordingly, some degree of individualization appears necessary.

The most accurate HRTFs are obtained by acoustic measurement (1; 2; 4; 5). From a practical standpoint, requiring acoustic measures of each user adds a potentially costly layer of complexity to any spatial audio system, including lengthy measurement sessions, specialized equipment, and the need for adapted facilities. Therefore, finding a means to provide some degree of user selection, while retaining spatial-audio fidelity, is important.

Individualized HRTFs can be obtained by alternative means. *Physical modeling* based on individually-specified anthropometry (6; 7) is less costly than acoustic measurement, but is subject to modeling and measurement error. Listener-guided search over multidimensional parameter spaces (8; 9) utilizes user feedback to correct for modeling/measurement error, but requires substantial amounts of time to tune even one spatial location, let alone 100's.

Among the methods for customizing HRTFs, *subjective selection* appears to be the easiest to implement while still providing reasonable spatialization. In this approach, a listener chooses an HRTF by searching over sets of sounds that are rendered by a database of HRTFs and selecting those with the best perceptual qualities. To implement this approach, the audio engineer requires a database of HRTFs, many of which are free and publicly-available (for example, (10)), a collection of test stimuli, and a set of perceptual attributes that best represent the demands of the spatial-audio application.

For example, in Seeber and Fastl (11), pulses of 30 ms white noise were generated sequentially at -40°, -20°, 0°, 20°, and 40° on the horizontal plane and were played over an electrostatic headphone. The change in azimuth over time created the perception of a moving noise source. Participants compared the quality of the moving source rendered by several pre-measured HRTFs according to externalization, front/back distinction, perceived direction, and source width. Seeber and Fastl (11) observed that localization improved when sources were rendered using an HRTF selected by the listener as opposed to rendering by rejected HRTFs.

In Roginska et al. (12), participants listened to sounds and chose HRTFs based on tournament-style listening tests. Listeners were asked to reject HRTFs in which pairs of sources were poorly discriminated on the basis of externalization, elevation, and front/back distinctiveness. Included among the pre-measured HRTFs for each listener was their own HRTF. The findings showed that listeners preferred a HRTFs belonging to a small subset of standardized HRTFs as often as they preferred their own HRTF.

The method used by Roginska et al. (12) for selecting HRTFs paired a listener's ITD function with the minimum-phase forms of database HRTFs. The present study investigates whether listeners can select ITD functions as well. Specifically, the method is modified by substituting a KEMAR ITD (13) for the listener's to obtain a subset of minimum-phase HRTFs that satisfy the spatialization criteria. Using one of these selections, the method is repeated by searching over ITD functions from the same database to pair the selected minimum-phase HRTF with a selected ITD. As in the original study, the present study investigates whether a listener finds any set of HRTFs acceptable, and, if so, whether they also show a preference for particular ITD functions.

2. METHOD

2.1. Subjects

Four women and eleven men ranging from 20 to 43 years old participated in the study. All were undergraduate students at the University of Michigan. Each participant underwent audiometric screening to make sure their hearing thresholds were within normal range. Before participating, each listener read and signed a consent form.

2.2. Apparatus

The graphical user interface used in this study was developed using MATLAB. The interface was used to control all phases of the experiment. Before beginning the experiment, the HRTFs datasets were

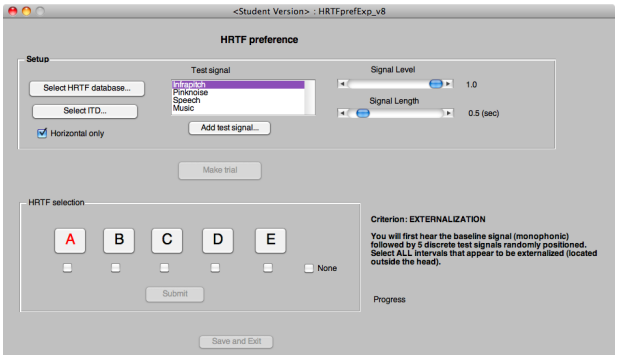

Fig. 1: User interface for the spectral coloration selection tasks. "A" is the interval currently playing during the externalization stage.

loaded into the system. The experiments were rendered on a 21" iMac system with an Apogee Duet audio interface. All stimuli were delivered using open circumaural Beyerdynamic DT 931 headphones at a level set by the participant. All experiments took place in a Tracoustics double-walled sound-proof booth.

2.3. Stimuli

A 500-ms infrapitch noise was used as the stimulus. The infrapitch noise was constructed by sampling 200 msec of a pink noise source and repeating the signal 2.5 times. With this period, listeners were able to hear the characteristic small-temporal spectral structures that are associated with infrapitch noise sources (14). A digital pink noise generator was randomly seeded each time an infrapitch noise was generated.

The HRTF datasets used in this study were the same pre-selected HRTFs used in (12). From the 97 sets of HRTFs, 27 were selected: 13 from the IRCAM database, 13 from the CIPIC database and the KEMAR dataset. For each trial, a given set of HRTFs was selected and impulse responses for the left and right ears were created by cascading the minimum-phase head-related impulse response (drawn from the given set) with the all-pass impulse responses for the KEMAR ITD.

2.4. Experiment One: Spectral Coloration Selection

The first experiment, *spectral coloration (SC) selection*, replicated Roginska et al. (12) with the exception that the individually measured ITD was replaced by the KEMAR ITD and the listener's HRTF was not included in the search space. As in the original study, listeners judged the quality of the spatial rendering according to a three-stage listening procedure during which, the perceptual spatial criteria of externalization, elevation differentiation and front/back differentiation were evaluated.

The user interface for controlling the experiment is shown in Figure 1. At the beginning of a session, the HRTF database, ITD, and test signal were selected. Following this selection, the listener began the first phase of the experiment in which they judged each rendering on the basis of externalization. Upon completion, the database was culled of sets that weren't selected two or more times (out of a total of 3 presentations), and the elevation phase was begun. Similarly, upon completion, the database was culled and the front/back phase was begun. To remind the listener, instructions on the screen included a definition of the criterion and a description of the listener's task.

Each phase of the experiment consisted of a set of trials, each trial of which presented examples of five different renderings. At the beginning of a trial, the listener heard each example in sequence. The visual display was used to cue the listener by highlighting the appropriate option button while the example was played. Before making their judgment, the listener could replay any one option by selecting the corresponding button. Check boxes below each option button were used to indicate which options provided adequate externalization (Phase One), elevation differentiation (Phase Two), or front/back differentiation (Phase Three).

In the case of *externalization*, the beginning of each trial was preceded by an unspatialized reference signal. Each interval was comprised of a series of sounds. The first sound in the interval was an unspatialized reference signal. The reference signal was generated by processing the test signal with the HRTF at 0° azimuth, 0° elevation and cross-summing the left and right channels. This was done in order to avoid spectral coloration variability between the raw and the processed test signal. This signal served as an in-the-head reference, to which the externalized sounds could be compared. Following the reference signal, the listener heard a se-

ries of five spatialized signals that were generated from randomly selected HRTFs at randomly selected azimuths on the horizontal plane: ±150° , ±120°, ±90°, ±60°, ±30°. The same sequence of azimuths was used for all intervals in each trial. The listener checked the boxes of the intervals in which externalization was perceived. In some cases, sound sources in the same interval were perceived as externalized and others were not. The listener was instructed to select the cases in which a majority (three or more) of the sounds was perceived as externalized. If none of the intervals were perceived as externalized, the listener checked the "None" checkbox. After submitting their selections, the results were saved, and the listener continued the externalization differentiation task for a new set of five intervals. Each HRTF in the database appeared in three intervals. Only the preferred HRTFs (selected at least two out of three times) were used in the elevation differentiation task.

The *elevation differentiation* phase, proceeded along lines similar to the externalization phase. Each interval was comprised of five pairs of stimuli. Each pair was spatialized using a randomly selected HRTF at a randomly-selected azimuth: ± 150° , ±120°, ±90°, ±60°, ±30° at ±36° elevation. The same sequence of azimuths was used for all intervals in each trial. The listener checked the boxes of the intervals in which they could discriminate the high and low signals in each pair of stimuli. The listener was instructed to select the intervals in which a majority (three or more) of the signal pairs had discriminable elevation differences. If none of the intervals contained pairs in which elevation could be discriminated, the listener checked the "None" checkbox. After submitting their selections, the results were saved and the overall completion progress was displayed. The trials continued in this manner until each set of HRTFs in the database had been evaluated three times. The database was then culled and only those sets of HRTFs that were selected two or more times were advanced to the front/back phase.

The *front/back differentiation* phase required the listener to judge the front/back distinctiveness of two sounds presented at locations along a common cone of confusion. Each interval was comprised of five pairs of stimuli presented on the horizontal plane. The pairs were spatialized using randomly selected HRTFs from the following azimuths: ± 150°

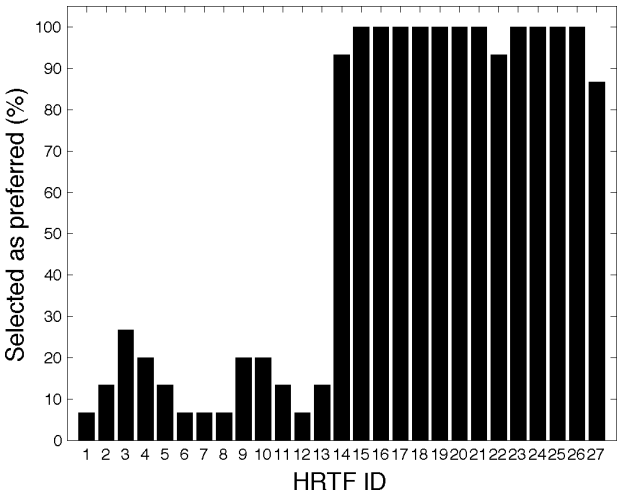

Fig. 2: Percentage of subjects that selected each HRTF during the externalization differentiation stage of HRTF spectral coloration selection. The abscissa indicates HRTF ID and the ordinate indicates the percentage of participants that perceived each HRTF's externalization cues.

, ±120°, ±60°, ±30°. The same sequence of azimuths was used for all intervals in each trial. Upon completion of the third phase, the sets of HRTFs that remain after culling of the database have passed the listener's judgment with respect to the spatial quality judgment of externalization, up/down and front/back differentiation.

Each of the 27 HRTFs were ranked with a number that indicated the number of subjects that preferred the HRTF's spectral coloration. The highest-ranking HRTF coloration within a subject's set of preferred colorations was selected as their preferred coloration. In the event of a tie, a spectral coloration was randomly selected from the listener's highest-ranking colorations.

2.4.1. Results

In the figures that follow, HRTFs #1 - #13 come from the IRCAM database, #14 - #26 come from the CIPIC database, and #27 was the KEMAR HRTF measurement.

2.4.2. Externalization

Figure 2 shows the percentage of times that each HRTF of the 27 HRTFs made it through the externalization judgment phase. The CIPIC datasets

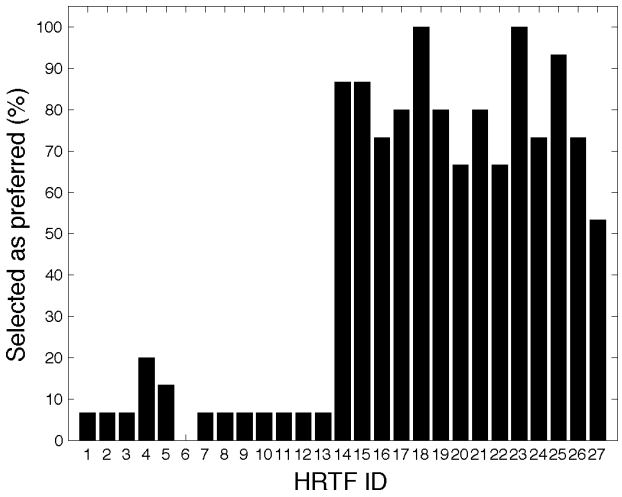

Fig. 3: Percentage of subjects that selected each HRTF during the elevation differentiation stage of HRTF spectral coloration selection. The abscissa indicates HRTF ID and the ordinate indicates the percentage of participants that perceived each HRTF's elevation cues.

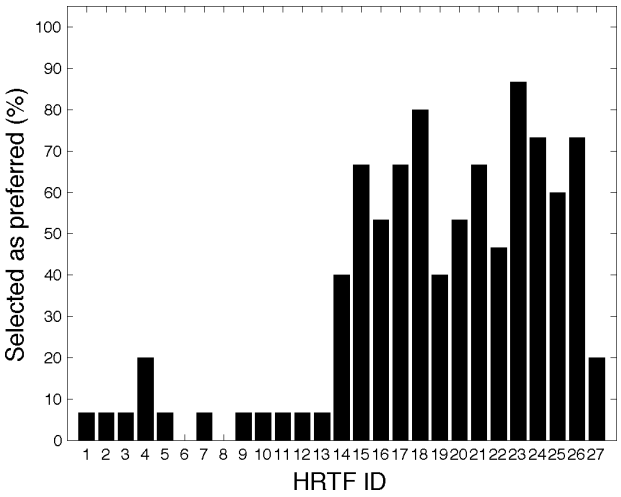

Fig. 4: Percentage of subjects that selected each HRTF during the front/back differentiation stage of HRTF spectral coloration selection. Along the abscissa is the HRTF ID and along the ordinate is the percentage of participants that perceived each HRTF's front/back cues.

and the KEMAR dataset were chosen by 86-100% of subjects for their externalization quality. In sharp contrast, the IRCAM datasets were selected by 6-26% of subjects. The selection results follow the trend observed in Roginska et al. (12). The CIPIC datasets and the KEMAR dataset were chosen by 70-80% of subjects for their externalization quality. In contrast, the IRCAM datasets were selected by 10-40% of subjects.

2.4.3. Elevation

In the second stage of SC selection, subjects were asked to discriminate between examples that were rendered at elevations above and below the horizontal plane. Figure 3 shows the percentage of subjects that chose each of the 27 HRTFs presented during the elevation differentiation task. Only HRTFs that were chosen in at least 67% of the presentations continued to the next stage . The CIPIC datasets and the KEMAR dataset were chosen by 53-100% of subjects for their elevation quality. In sharp contrast, the IRCAM datasets were selected by 0-20% of subjects. Among the subjects that judged #6 to have good externalization cues, none of them perceived it as having good elevation differentiation.

Our results follow the trend observed in the previous work in that a higher percentage of subjects preferred the spectral colorations of the CIPIC and KEMAR HRTFs. A smaller percentage of subjects preferred the IRCAM HRTFs. In the present work, HRTF #6 was eliminated at this stage, as it was not preferred by any listeners. In the previous work, the same HRTF was also eliminated at this stage.

2.4.4. Front/Back

Figure 4 shows the percentage of subjects that chose each of the 27 HRTFs presented during the front/back phase. The CIPIC datasets and the KEMAR dataset were chosen by 20-86% of subjects for their front/back discernibility. The IRCAM datasets were selected by only 0-20% of subjects. An additional HRTF (#8) from the public datasets has been eliminated at this stage.

As compared to the results of our previous work, fewer HRTFs were eliminated at this stage of the listening test. In the present study, HRTF #8 was eliminated, as also seen at this stage in the previous work.

The previous figures have shown data averaged across subjects. Figure 5 summarizes the selec-

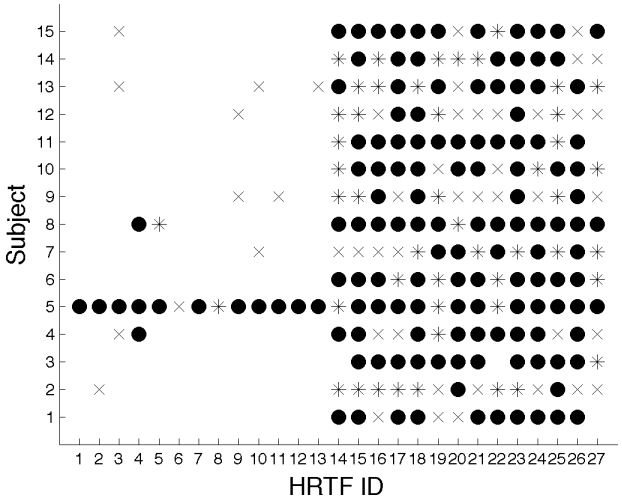

Fig. 5: HRTF spectral coloration preference for each subject, for the 3 judgment tasks: externalization (x), elevation differentiation (*), and front/back differentiation (filled circles). Along the abscissa is the HRTF ID and along the ordinate is the Subject ID.

tion process for each subject (ordinate) for each HRTF set (abscissa). The absence of a symbol indicates that the subject did not select that HRTF set two or more times during the externalization phase. The 'x' marker represents a selection during the externalization stage only. HRTFs selected after the externalization and elevation differentiation tasks are represented by the '*' marker and final winners are presented by filled circles. Selection results are very similar across subjects. All of the subjects had a strong preference for HRTFs from the CIPIC database and the KEMAR measurements. In addition, subjects #4, #5 and #8 preferred at least one HRTF from the IRCAM database. In the previous work, only two listeners preferred the HRTFs of the IRCAM database; however, in the present study, none of the listeners preferred the IRCAM HRTFs.

2.5. Experiment Two: ITD Selection

In the second experiment, *ITD selection*, the listener chose their preferred ITDs by judging the spatial qualities of sounds delivered using different ITDs from a database of HRTFs. The preferred spectral colorations, as determined in Experiment 1 (SC selection), were used to deliver the spatialized sounds.

To pick the preferred spectral coloration, the HRTFs were given a numerical score according to the number of subjects that preferred it in the final stage. Of each listener's final set of colorations, the highest scoring spectral coloration was chosen as the listener's preferred spectral coloration. In the case of a tie, one of the highest scoring HRTFs was randomly chosen. In the same manner, a preferred ITD was determined for each participant of the present experiment. Each listener's preferred spectral coloration and preferred ITD were combined to create the set of HRTFs for that listener throughout the rest of the experiments.

In ITD selection, each listener's preferred ITD was selected from among the ITDs within the HRTF database. At the beginning of the ITD selection experiment, the experimenter began by loading the database of HRTFs into the system. Next, the experimenter loaded each listener's preferred spectral coloration into the system. Afterwards the listener pressed a button to begin the procedure in which they completed the 3 stages of differentiation tasks in the same manner as described in Experiment 1.

2.5.1. Results - ITD selections

Figures 6-8 show the percentage of subjects that chose each of the 27 HRTF sets for the externalization, elevation and front/back stages of the selection process, respectively. In contrast to the choice of spectral coloration, there is relatively little clustering in selection around particular collections of HRTF sets. Some selectivity is achieved, and this selectivity improves with each subsequent phase. The average selection rate for externalization was 57%. This decreased to 42% after the elevation phase and was further reduced to 26% upon completion of the third phase. Thus, listeners are able to reject options at each stage, but their individual choices do not agree with those of the entire group.

When broken out by individual subject, the same differences are noted between the ITD selections and the spectral coloration selections. As shown in Figure 9, the 'x' marker represents items that passed externalization before being rejected, the '*' marker represents items that passed both externalization and elevation before being rejected, and the filled circle represents items that passed all three criteria. Selection results are very similar across subjects.

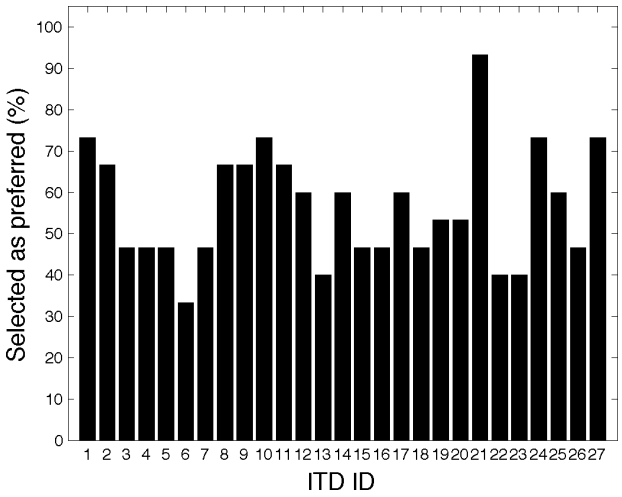

Fig. 6: Percentage of subjects that selected each ITD during the externalization differentiation. Along the abscissa are the HRTF IDs and along the ordinate is the percentage of participants that perceived each HRTF's externalization cues.

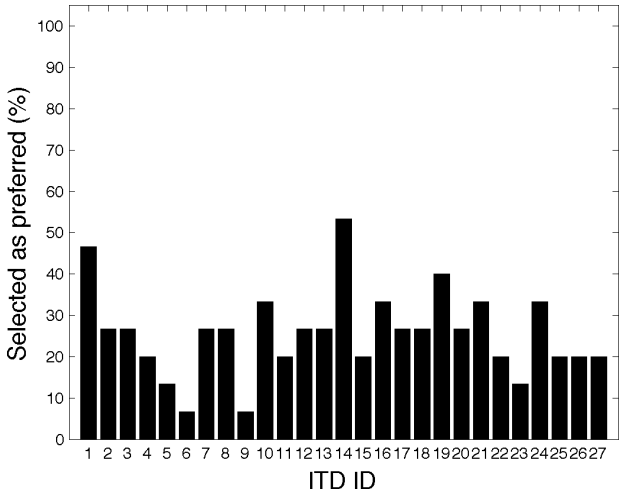

Fig. 8: Percentage of subjects that selected each HRTF during the front/back differentiation stage of ITD selection. Along the abscissa are the HRTF IDs and along the ordinate is the percentage of participants that perceived each HRTF's front/back cues.

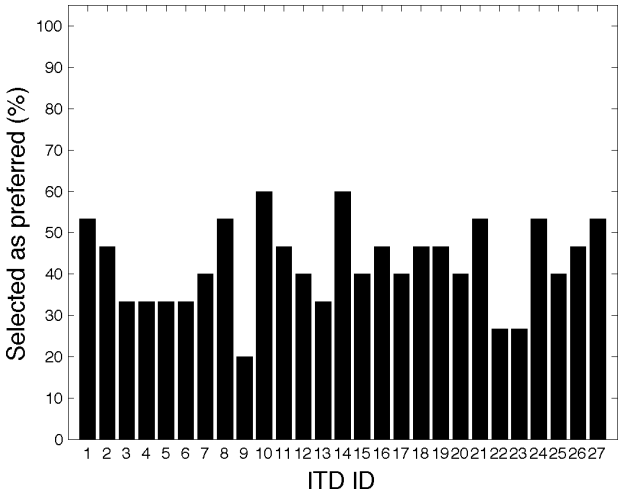

Fig. 7: Percentage of subjects that selected each HRTF during the elevation differentiation stage of ITD selection. Along the abscissa are the HRTF IDs and along the ordinate is the percentage of participants that perceived each HRTF's elevation cues.

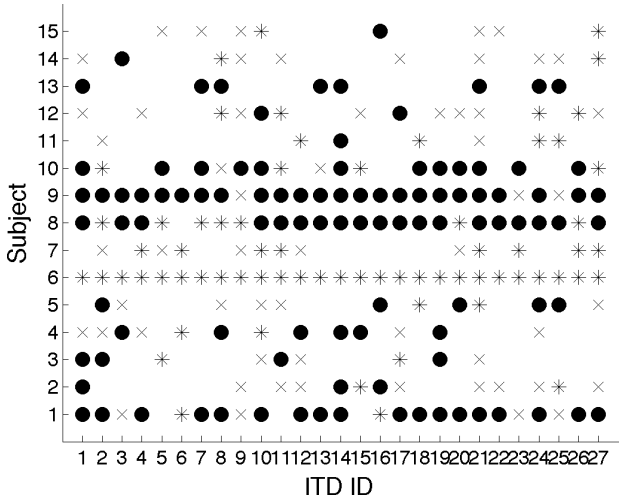

Fig. 9: ITD preference for each subject, for the 3 judgment tasks: externalization (x), elevation differentiation (*), and front/back differentiation (filled circles). Along the abscissa are the HRTF IDs and along the ordinate are the Subject IDs.

None of the subjects exhibited a strong preference for a specific ITD or set of ITDs from the HRTF database. There is no single ITD that is rejected or accepted by most subjects. Furthermore, most subjects find a large number of ITDs to be acceptable under these criteria. The IRCAM ITDs were selected as often as the CIPIC ITDs.

2.5.2. Results - Customized HRTFs

The HRTFs created in the two experiments are shown in Table 1. As in the SC selection procedure, the listener's preferred ITD was identified. Of the 15 customized HRTFs that were created, 12 were unique. One participant (#14) selected a spectral coloration and ITD from the same HRTF in the database.

Table 1: HRTF IDs of the spectral coloration and ITD used to create each customized HRTF

Participant	Spectrum	ITD
1	21	14
2	24	1
3	21	1
4	21	14
5	23	16
6	23	14
7	23	26
8	23	14
9	24	14
10	21	19
11	26	14
12	21	10
13	21	14
14	23	23
15	21	16

3. DISCUSSION

The results of the present work replicate the earlier findings of Roginska et al. (12) without the need for an individually-measured ITD. Repeating the selection procedure for spectral coloration suggests that listeners tend to reject previously selected options rather than accept ones they had previously rejected.

Additionally, the results affirm that there are discriminable spectral cues that most listeners prefer over another, even when listening using standardized ITDs. Similar to the observations in Roginska et al. (12), there was a group of listeners that preferred the spectral colorations of the HRTFs from the CIPIC and KEMAR databases. None of the listeners preferred the IRCAM HRTFs. Furthermore, we were also able to identify common HRTFs that did not provide elevation and front/back distinction to any listener in the present study and Roginska et al. (12). These conclusions are important with respect to the use of spatial audio in the field. By establishing that the same pre-measured HRTFs are selected using a pre-measured ITD, we have shown that it is not necessary to individually measure the ITD for the listener in a practical customization procedure. Furthermore, we have shown that listeners do prefer some ITD sets to others, and that they can refine their preferences through the same three stages of evaluation.

The results indicated that each ITD had about an equal likelihood of being preferred in the differentiation tasks. Listeners, as a group, did not show preference for any particular subset of possible ITDs. It should be noted that participants informally reported that the ITD selection task was more challenging than the SC selection task, as there were smaller differences in the spatial cues.

Bibliography

[1] A. Bronkhorst, "Localization of real and virtual sound sources,"Journal of the Acoustical Society of America, 98(5), 2542–2553., 1995

[2] H. Moller, M.F. Sorensen, C.B. Jensen, and D. Hammershoi, "Binaural technique: Do we need individual recordings?," Journal of the Audio Engineering Society, 44(6), 451–469, 1996.

[3] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using non-individualized head-related transfer functions," Journal of the Acoustical Society of America, 94(1), 111–123, 1993.

[4] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. Part 1: Stimulus synthesis," Journal of the Acoustical Society of America, 85(2), 858–867, 1989.

[5] 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.

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

[7] 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, 24(5), 333–334, 2003.

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

[9] A. Silzle, (2002), "Selection and tuning of HRTFs," AES, pp. 1–14, 2002.

[10] V. Algazi and R. Duda, "Estimation of a spherical-head model from anthropometry," Journal of the Audio Engineering Society, 49(6), 472–478, 2001.

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

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

[13] C.I. Cheng and G.H. Wakefield, "Moving sound source synthesis for binaural electroacoustic music using interpolated head-related transfer functions (HRTFs)," Computer Music Journal, 25(4), 57–80, 2001.

[14] R.M. Warren and J.A. Bashford, Jr. "Perception of acoustic iterance: Pitch and infrapitch," Perception & Psychophysics, 29, 1981.