PERCEIVED NATURALNESS OF SPEECH SOUNDS PRESENTED USING PERSONALIZED VERSUS NON-PERSONALIZED HRTFS

John Usher[1] and William L. Martens[2]

[1]Personics,
jusher@personicslabs.com
[2]McGill University,
wlm@music.mcgill.ca

Montréal, Canada

ABSTRACT

Speech sound sources were spatially processed using measured HRTF data that were obtained from nine individuals. The speech signals were auditioned via headphones by two groups of listeners via a paired comparison task in which listeners were asked to judge which of two stimuli sounded more natural. One group of listeners was composed of those whose HRTFs had been used to create subsets of the stimuli that were presented, while a second group of listeners were never presented with stimuli that were processed using their own HRTFs. Results from the first group showed that stimuli generated using an individual's own HRTFs will not necessarily be judged as more natural than those generated using HRTF data from other individuals. However, this was not because one set of HRTF data gave the most natural listening experience for all listeners, since the stimulus ranked highest differed between individuals. An analysis of the Interaural Level Difference (ILD) showed that the frequency dependence of ILD for an individual's HRTFs was quite similar to that of the HRTFs that produced for them an auditory image that was ranked as the most natural sounding. The results suggest that the interaural spectral difference presented via HRTF-based processing can affect perceived naturalness as strongly as the overall spectral shape that is related to source tone coloration.

[Keywords: HRTF, ILD, individualization, personalization, naturalness, auditory scene analysis]

1. INTRODUCTION

Headphone-presented virtual auditory scenes created using HRTF-based processing are increasingly used in audio systems for voice communication and music reproduction. The motivations for and benefits of presenting virtual auditory scene using binaural synthesis are many. Such processing can result in an increased sense of naturalness, realism, or presence within an auditory scene; manipulate the spatial sound aesthetic of music; facilitate detection of auditory warning signals; or increase intelligibility of speech, especially in a multi-talker scenario. It is the focus of this paper to investigate how different HRTF datasets affect perceived *naturalness* when used to process multiple voice recordings for headphone display.

Naturalness in audio is related to *sound fidelity*. The *fidelity* of a sound listening experience is a measure of how perceptually similar a listening experience is to another reference listening experience. The point of reference may either be a reproduced environment or a live venue [1]. For our experiment, we asked listeners to compare auditory scenes containing a mix of two spoken messages, one message containing male speech and the other containing female speech, both of which where processed using a sampling of 9 different HRTF datasets (from different subjects). The listeners were instructed to attend to the female speech, and respond in terms of how similar it sounded to what can be heard in natural everyday listening; i.e. the point of reference was provided by the listener's memory of the character of a spoken voice in a natural acoustic setting, rather than a spoken voice reproduced via loudspeakers or headphones.

Note that this study focused on *naturalness* rather than *preference*, as *preference choices* may be understood by experimental listeners in many different ways. Preference decisions may be based more upon hedonic *sentiments* than upon *judgments* of sound character: a judgment is an opinion that can be subject to veridical evaluation, but veridicality does not apply to sentiments [2]. In contrast, it is reasonable to assume that every normal-hearing individual has a similar notion of what spoken voices sound like in the natural world, and this should provide a clear reference for making the *naturalness choices* required in the current study. It should also be clear that conventional tests of impairment in speech sound quality for standard teleconferencing situations (as described in [3]) are not designed to reveal differences in detailed features such as those associated with spatial sound processing.

1.1. HRTF processing and Spatial Release from Masking

HRTF-based processing of audio signals can be used for affecting Auditory Spatial Imagery (ASI) [4] in binaural reproduction of virtual acoustic scenes (using headphones). Changes in ASI afforded by such binaural auditory display can be in terms of perceived source image direction and distance, but also in other characteristics such as image width [5]. In his now classic studies on the "Cocktail Party Effect," Cherry [6] found that increasing the perceived spatial separation between the sound from a "target" voice (that person we are talking with at the cocktail party) and a "masking" sound (e.g. someone else speaking, or the general hub-bub of the party) can increase the intelligibility of the speech signal received from a target person (to which the listener is selectively attending).

As pointed out by Begault and Erbe [7], studies investigating the binaural advantage of presenting information in an auditory display have focused on changes in the detectability of a sound signal (generally, a pure-tone) in the presence of noise; or in improvement in speech intelligibility. The improvement in either detection of a tone or the increase in intelligibility of a voice by spatial separation of target and marker sound images is the process called Spatial Release from Masking (SRM) [8]. The focus of this paper differs from the majority of work on HRTF audio processing by investigating sound quality rather than SRM effects.

1.2. Personalized and non-personalized HRTF processing

It might be intuitively reasoned that HRTF mixes made with an individual's own head are perceived as more natural-sounding than HRTF mixes made with other peoples' heads, as the spectral filtering due to the HRTF function is that which the person is used to. (Of course, this hypothesis does not consider the acoustic coupling of the headphones. Wenzel et al [9] found that when non-individualized HRTF processing was applied to headphone presented test-stimuli, the perceived location of the corresponding auditory image was consistently reported with a similar azimuth as to real (loudspeaker) sources around the listener. However, when it came to reporting the perceived *elevation* of the image, subjects frequently confused up with down for the HRTF processed, headphone-presented stimuli. This may be explained by the fact that whilst interaural level and time differences (ILDs and ITDs) may vary less between subjects, elevation cues are affected by the shape of the ridges in the pinna, and it is perhaps these pinna shapes which vary more between individuals whilst ILD and ITD cues vary less.

With regard to the effectiveness of individualized HRTFs, it is unclear whether such exacting reproduction is necessary for a perceptually adequate spatial auditory display. With regard to practicality, it is unclear whether such individualization can be successfully attained at a reasonable cost (in terms of both time and money). Therefore, some alternative approaches should be considered, four of which can be described quite simply in terms of differences in design goals for, and the targeted user(s) of, the filters to be used in binaural synthesis. Table 1 shows four varieties of filters for controlling source direction based upon these two distinctions. The left column identifies who is targeted, the rows of the table specifying whether filters are to be designed for an individual (one) subject or a sampled population of (many) subjects. The middle column of the table (*empirical*) specifies that the HRTF detail should be exactly reproduced, whether individualized or averaged across a sampled population of subjects. Entries in the right column of the table (*analytic*) result from some analysis of the HRTF data, performed before filter design is attempted. Such analytic methods may use anthropomorphic measurements from one or more individuals such as body size or the shape of the outer ear.

In the case of a foundation in ''Individualized HRTFs,'' a successful analysis will extract details that should allow for the synthesis of filters that are ''customized'' to the individual, and yet not necessarily matching the individual's measured HRTFs; giving a *personalized* HRTF set. *Individualized* HRTFs are contrasted with *personalized* HRTFs in that the former is an HRTF set made from empirical measurements of one individual's head, whereas the latter is an HRTF set tailored for one individual and may include other signal processing, such as combing subjective judgments or other measurements of an individual such as body size.

It may be more appropriate to refer to derived HRTF datasets according to their intended use in controlling direction, hence the term ''directional transfer functions'' (DTFs). When a single set of DTFs are created from many sets of HRTFs, and intended for a whole population of potential users, they are termed ''generalized'' DTFs. If the analysis and synthesis of such filters is successful, then the resulting set of generalized DTFs will capture global spectro-temporal features that provide sound positioning cues for many listeners, regardless of the details of any individual's HRTFs (see [10] for further discussion).

Target	**Empirical**	**Analytic**
One	Individualized HRTFs	Customized DTFs
Many	Averaged HRTFs	Generalized DTFs

Table 1: Four varieties of directional filters categorized according to distinctions between target users and measurement foundations of the filters for use in spatial auditory display. Whether the target is one user or many users, HRTFs exhibiting exact detail can be engineered empirically for binaural synthesis. But analysis of measured HRTF data can produce results that provide effective DTFs that can be customized for one user or generalized for many users.

2. METHODS

2.1. Subjects

Of the 20 subjects who took part in the experiment, 2 were from the group who had their HRTFs measured. The remaining subjects consisted of professional musicians studying at the Banff Centre (Alberta, Canada) on self-directed music programs. Data from all subjects who undertook the experiment are presented here (i.e. no data was rejected).

2.2. HRTF acquisition

Complete HRTF datasets were measured for 9 individuals using a 128-microphone array at University of Maryland, College Park, using the reciprocity technique [11] (this is explained in Figure 1). 9 HRTFs were measured at 5^{o} intervals in the upper sphere of the listener, and 15^{o} below the horizontal plane (i.e. the equatorial plane which is parallel to the floor and has its reference at the level of the individuals ears). The resulting HRTFs were truncated to a length of 128 samples

(ITD's were maintained). Due to the small size of the transducers involved in the reciprocal HRTF measurement rig, the low-frequency component of the final HRTF is not measured by empirical acoustic means but is estimated using a mathematical model based on anthropomorphic measurements of the individual subject.

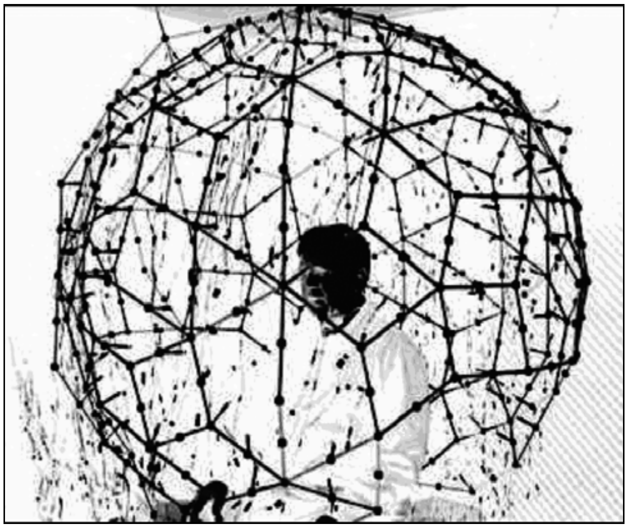

Figure 1. Photograph of an individual undergoing HRTF measurement using the reciprocity technique [11]. (Note that the image is contrast-reversed to improve visibility of the microphone array structure.) There are 128 miniature microphones on short (3 cm) inward-facing sticks attached at some of the nodes of the spherical mesh in which the subject is seated. The subject has miniature loudspeaker receivers mounted in an earplug in each ear to create sound at the occluded ear meatii. One loudspeaker emits 25 clicks and the last 20 clicks are recorded simultaneously using the 128 microphones and a computer. The process is then repeated with the other loudspeaker. The individual undertakes the measurement with their head facing 0^o, 90^0, 180^o, and 270^o azimuth.

2.3. Stimuli

Recordings of a male and a female voice were simultaneously presented. The single-channel male voice was taken from a recording of spoken poetry[1] in American-English, and the female voice was spoken text[2] in British-English. Both 44.1 kHz, 16-bit recordings were converted to mono, edited to 35 seconds, with pauses greater than 50 ms removed. The mono inputs were convolved with the 9 HRTF datasets using the "Panorama" software.[3] The male voice was positioned 10^o to the left of "straight-ahead" location), and the female voice 10^o to the right, with both voices at 0^o elevation. Standard iPod insert headphones were used, with no additional equalization.[4]

2.4. Procedure

The methods employed here were motivated by the following research question: Is the listening experience superior when we listen to an auditory scene created with our own, personalized HRTFs versus an auditory scene that is created with someone else's HRTF? Specifically, the question asked was: Is perceived *naturalness* of an auditory scene affected by using personalized versus non-personalized HRTF processing?

Given 9 HRTF datasets, there were 36 possible stimulus pairings. These were presented to the subject with earphones in a double-blind manner using the audio processing software PureData (PD). The subjects could freely select which of the two stimuli to audition by hitting a software button labeled A or B. The subjects were asked: *"In which scene does the female voice sound most natural?"* and responded by selecting stimulus A or B on the GUI and then hitting a select button. The stimuli would repeat until the subject had selected their response (there was no time limit on the experiment). The reproduction level was equal for all subjects (at a listening level of approximately 75 dB). The subjects were told to listen to both spectral and spatial aspects of the sound image, and to judge naturalness using their memory of what speech sounds like in the real world (i.e. not with reference to recorded speech). After each run of the 36 pairs, the subject could take a break of 5-10 minutes before proceeding to the next run (i.e. repeat). Between 2 and 6 runs were undertaken per subject, and their response time recorded (response time data is not reported in this paper).

3. RESULTS

3.1. Analysis of HRTF Spectral Differences

The HRTF datasets (magnitude only) were submitted to Principal Components Analysis (PCA) with the 9 binaural HRTFs as the cases (one pair for each measured subject), using as the variates the response within 30 frequency bins for each pair of ears (i.e., ipsilateral and contralateral magnitude response curves). This matrix with 60 columns of correlated magnitude values was reduced via PCA to just two Principal Components (PC) vectors. For more information about such analysis of HRTF spectral differences, see the second author's 1987 paper that first presented the use of PCA in this context as a means to reduce a large set of HRTFs to a smaller set of spectral basis functions [12]. Figure 2 plots the PC scores on the resulting two-component space for each of the 9 HRTF datasets. These first two PCs accounted for more than 40% of the total variance in the HRTF datasets, and examination of the knee in the associated scree plot supported the exclusion of other PCs from further consideration here. Especially since resynthesis of the HRTF data was not at issue here, it was most appropriate to find just a few spectral basis functions that might explain simple differences between HRTFs that could also predict their naturalness rankings by human listeners.The eigenvectors (orthonormal bases, or weights) associated with these two PCs are plotted over

[1] Allen Ginsberg reading Jack Karoac's "Brooklyn Bridge Blues".
[2] Diana Deutsch; track 17 from her CD "Phantom words and other curiosities". The audio stimuli can be heard at http://www.JAR-lab.com/ICAD07
[3] Manufactured by Wave Arts, version 5.
[4] The same headphone set was used for all listening tests. The insert iPod earphones were as shipped with N. American iPods in December 2006.

frequency in Figure 3, with 30 ipsilateral bins (blue solid lines) and 30 contralateral bins (red dashed lines).

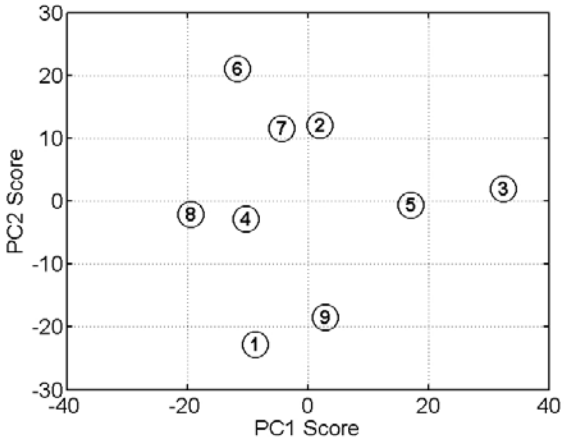

Figure 2. Scores on the first two Principal Components (PCs) resulting from analysis of HRTF magnitude measured for nine human subjects. The number in each plotting symbol indicates the subject number (i.e. HRTF set number).

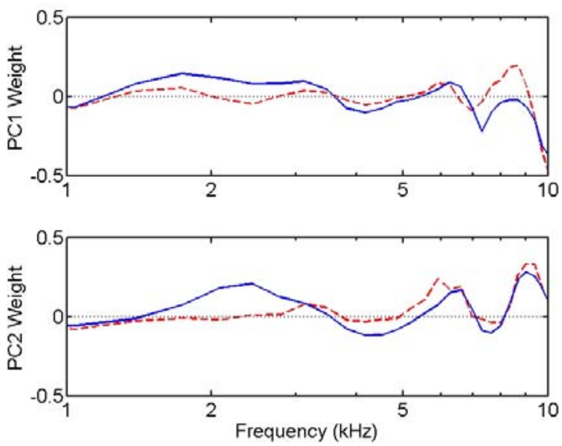

Figure 3. Principal Component (PC) weights resulting from analysis of HRTF magnitude measured for nine human subjects. Blue solid lines for ipsilateral magnitude, and red dashed lines for contralateral magnitude. The eigenvectors (orthonormal bases, or weights) associated with these two PCs are plotted over frequency in Figure 3 (with 30 ipsilateral bins and 30 contralateral bins distinguished by plotting-line style; blue solid lines for ipsilateral magnitude, and red dashed lines for contralateral magnitude). The scores plotted in Figure 2 were obtained by summing the observed magnitude values after weighting them with the values plotted in Figure 3. These two sets of scores provide the foundation in HRTF spectral differences for predicting the naturalness values calculated for each binaural HRTF.

The scores plotted in Figure 2 were obtained by summing the observed magnitude values after weighting them with the values plotted in Figure 3. These two sets of scores provide the foundation in HRTF spectral differences for predicting the naturalness values calculated for each binaural HRTF.

3.2 Naturalness Choice Analysis

The naturalness choice data were analyzed using Thurstonian scaling for binary paired comparisons [12]. For each listener, a 9x9 matrix was created for the 9 stimuli which contained pairwise choice data for the paired comparisons. For example, if HRTF dataset 3 was considered more natural than HRTF dataset 7 for 3 out of the 4 pair presentations, then the matrix value at column 3, row 7 would be 3, and the value at column 7, row 3 would be -3 (alternatively, these could be converted to a proportion; i.e. 0.75 and 0.25, respectively). By summing the columns, a stimulus merit scale value could be obtained, and these scale values are those that are plotted for each listener horizontally in Figure 4. Note that the result based upon the pooled responses of all listeners appears in the lowest row of the plot, labeled "ALL."

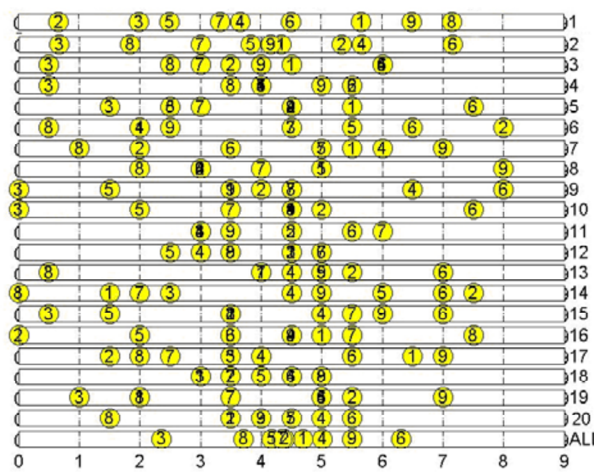

Figure 4. Naturalness scale values calculated for each of the 9 auditory scenes by each of 20 listeners (plus the result based upon the pooled responses of "ALL"). A marker with a value of 9 indicates this auditory scene was chosen as being more natural-sounding every time it was presented; a marker with a value of 4.5 indicates this auditory scene was selected as providing an auditory image more natural than other HRTF mixes 50% of the time; and a marker with a value of 0 indicates this HRTF dataset was never selected as being more natural-sounding. Listeners were presented the 36 stimuli in all pair-wise comparisons at least 2 times (with some of the listeners completing up to 6 separate runs). Note that subjects no. 1 and no. 2 were those for whom HRTFs were available.

It was of great interest to analyze the naturalness scale values for the two individual subjects for whom HRTF datasets were available, which HRTFs were used to create a subset of the stimuli presented to all listeners. In addition, it was of interest to attempt to predict the naturalness scale value for each from the PCA results. These results are plotted in Figure 5 as a function of the predicted naturalness scale value (upper and lower panels for the two listeners, respectively). The predicted values resulted from multiple regression analysis using the PC scores calculated for the 9 HRTF datasets as the predictor variables. Note that the symbol corresponding to the HRTFs measured for each of these two subjects is plotted as a filled symbol since this data-point was produced using the individual's own HRTFs.

4. DISCUSSION AND CONCLUSION

The results show that the HRTF dataset which gives the most natural auditory scene is not necessarily that auditory scene created with the listeners own HRTF; as can be seen in figure 4, where the subject no. 1's own HRTF dataset was ranked 3rd in terms of perceived naturalness. The stimulus produced using that HRTF dataset was ranked as less natural than two other datasets (no. 8 and 9) that came from people with very different body size. An analogous result obtained for subject number 2, whose own HRTF was also ranked 3rd behind two other datasets (no. 4 and 6). Note that 10 out of 20 listeners chose HRTF dataset No. 6 as the more natural sounding across a dominant number of comparisons. The HRTF magnitude for set no. 6 is shown in Figure 6.

The results show that individuals might consider auditory scenes made with HRTF datasets measured for other subjects (i.e., other than their own) can be chosen as more natural-sounding. This general conclusion is in contrast to the findings regarding localization performance. For example, Møller, et al [5] found that accuracy in spatial localization of sound source images, individualized HRTF-based processing provided superior localization compared to non-individualized HRTF-based processing (i.e. more accurate and consistent localization, with less front-back confusion, etc.). Of course, naturalness is not often used as a criterion for evaluating the quality of HRTF datasets, though it might be considered an important aspect in some application contexts.

The question that begs to be asked here is whether there is something about the HRTF datasets that were chosen most often as natural sounding that could predict choices for individuals. Figure 5 showed that the naturalness scale values for the two individual subjects for whom HRTF datasets were available could be predicted for each from the PCA results, but further analysis is required to find what spectral features of the individual's HRTFs might predict those choices. The overall spectral shape, such as that shown in Figure 6 did not produce the best results in this regard. Rather, it was the frequency-dependent Interaural Level Difference (ILD) that produced the best explanation for individual choice.

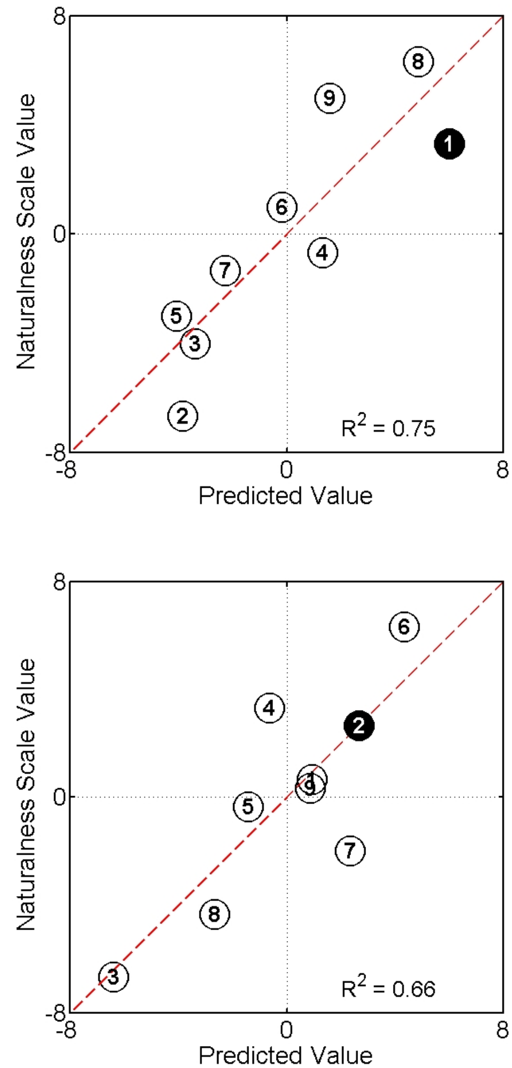

Figure 5. Upper Panel. Naturalness scale values calculated for each of the nine tested HRTFs to quantify the perceived naturalness of each according to the rankings obtained from listener number 1 (i.e. with HRTF dataset 1),. These scale values are plotted as a function of the predicted scale value that resulted from regressing them on the PC scores that were plotted in Figure 2. The number in each plotting symbol indicates the number of the subject providing the tested HRTFs. Here, open symbols are used for all subjects except for subject number 1, who was listening via his own HRTFs in this case. **Lower Panel.** Naturalness scale values calculated for the nine stimuli for listener number 2 (i.e. with HRTF dataset 2). In both figures, the coefficient of determination (R^2) shows the goodness of fit for the prediction of naturalness scale values from the PC scores.

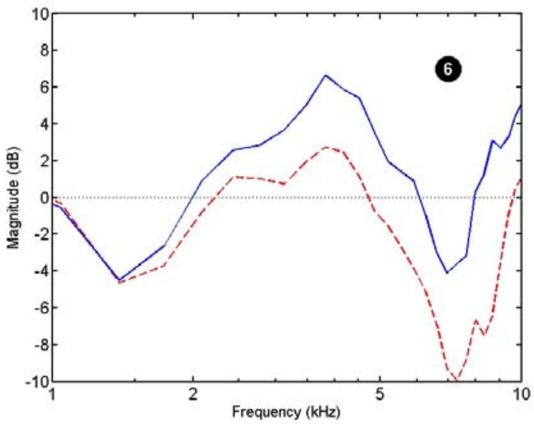

Figure 6, Binaural HRTF magnitude plotted as a function of frequency for subject 6 (solid lines for ipsilateral magnitude, dashed lines for contralateral magnitude).

Figure 7 shows an analysis of the ILD as a function of frequency for two pairs of HRTF datasets. In the top panel, the ILD for subject 1's own HRTF (solid line) is compared to the HRTF dataset (dashed line) that subject 1 chose as more natural sounding than his own: These two ILD functions are remarkably similar. It is particularly interesting to note that the contralateral ear has a higher level than the ipsilateral ear at about 2 kHz for both the subject's own HRTF and the most-natural-sounding HRTF. This trend is analogous to that for subject 2, in that subject 2's own ILD function was much more similar to that of his chosen most-natural-sounding HRTF dataset. Furthermore, the spectra of the personalized HRTF of subject 1 and subject 2 are very different (e.g. there is no contralateral boost in ILD for subject 2), which may explain why subject 1 did not find the HRTF dataset from subject 2 as natural sounding (curiously, subject 2 did not find the HRTF dataset of subject 1 quite as un-natural sounding).

Of course, there are a number of obvious limitations to the method used in the current experiment that could limit the extent to which the conclusions presented here can be generalized. Besides the small number of subjects tested with personalized HRTFs (2), the stimulus set was limited, and the coloration to sound image quality by the iPod insert headphones also may have affected which HRTF seemed most natural-sounding. Furthermore, individuals may differ both in terms of their anatomical size (therefore differing naturally in their measured HRTFs), and in terms of their perceptual responses, as discussed in [14]). They may also differ in their measured earphone transfer functions, which were not measured, and therefore not corrected in the current study. This may be less important, however, given that the overall shifts in peak frequency of earphone correction filters are difficult to detect for deviations less than 20% [15].

An important challenge in the optimal deployment of HRTF-based spatial auditory display systems is to identify the determinants of significant variation between individuals, and to determine how best to reduce problems associated with this variation. *Personalized* headphone-based display systems are likely to become more and more common, especially with current advances in mobile telephone technology. Perhaps the most important conclusion of this study might be that individual differences in both perceptual responses and anatomical size can be taken into account through the use of customized HRTFs that can selected on the basis of psychophysical calibration (as taught in [14]). It should be stressed, however, that all of the research reported here utilized dry binaural sources, and this is another factor that limits the generality of the results, since more and more spatial auditory display systems feature some simulated indirect sound. Indeed, aspects of display performance that are related to perception of indirect sound were not addressed here.

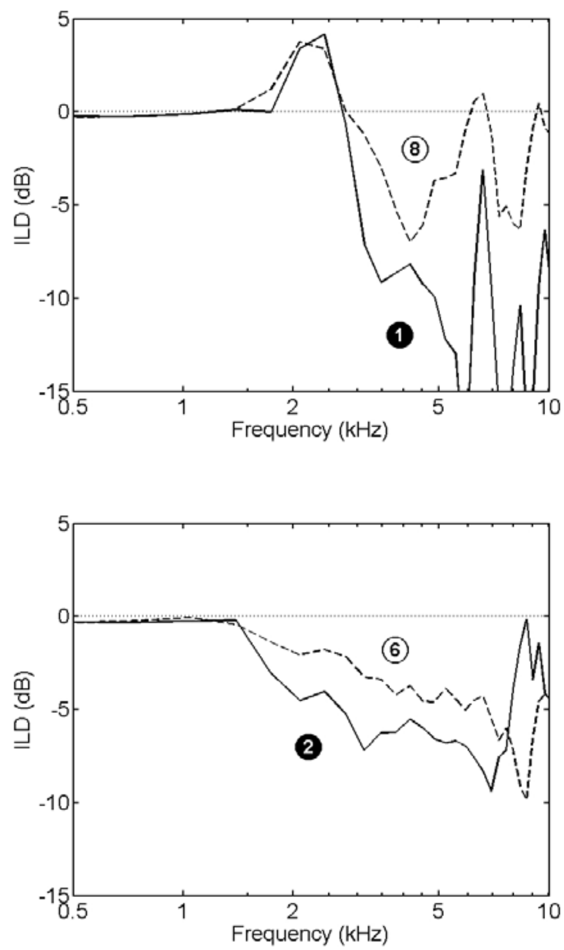

Figure 7, Upper Panel. Interaural Level Difference (ILD) as a function of frequency measured for subject 1 (solid line) compared to the ILD curve (dashed line) from another subject's HRTF dataset (from subject 8) that subject 1 chose as more natural sounding than his own. **Lower Panel.** The ILD function measured for subject 2 (solid line) compared to the ILD function (dashed line) from the HRTF dataset that subject 2 chose as more natural sounding than his own (from subject 6).

Research is under way to begin to characterize the variation in auditory spatial imagery associated with headphone-based presentation of simulated virtual environments, and furthermore, to begin to determine what factors in virtual acoustic simulation lead users to prefer one simulated environment over another within defined binaural synthesis applications (see, for example, a report on the influence of spatial distribution of simulated reflections on auditory quality and character [16]). Whether the current, perhaps controversial, results will be supported by the results of further studies presenting more comprehensive virtual acoustic simulations remains to be seen; however, the inclusion of indirect sound should function to make the details of the HRTF processing of the direct sound less critical rather than more critical. It is therefore concluded that HRTFs that are personalized through a customization procedure, rather than through exacting acoustical measurements, may be quite adequate to producing acceptable results for many if not most applications,

5. ACKNOWLEDGEMENTS

The HRTF measurements used in preparing the stimuli for this study were made at the department of computer science at the University of Maryland Institute of Advanced Computer Studies (UMIACS) thanks to the kind assistance from Adam O'Donovan, Dr. D. Zotkin and Dr. R. Duraiswami.

Part of the listening tests were undertaken at the Banff Centre, thanks to the assistance of Theresa Leonard, Steve Bellamy, and all the musicians and engineers who undertook the listening tests.

6. REFERENCES

[1] Begault, D.R. (2006). “Preference versus Reference: Listeners as participants in sound reproduction”. Spatial audio and sensory evaluation techniques, Guildford, UK.

[2] Nunally, J. and Bernstein, I. (1994). Psychometric Theory. McGraw-Hill,

[3] ITU-T, Recommendation P.85, “A method for subjective performance assessment of the quality of speech voice output devices,” Int. Telecommunications Union, Telecommunications Standardization Sector, 1994.

[4] Usher, J. and Woszczyk, W. (2004). “Visualizing Auditory Spatial Imagery of Multi-channel Audio,” Presented at AES 116th Convention. Paper 6054.

[5] Møller, H., Sørensen, M. F., Jensen, C. B. and. Hammershøi, D. (1996). “Binaural Technique: Do We Need Individual Recordings?” J. Audio Eng. Soc., 44(6):451-469.

[6] Cherry, C. (1953). “Some experiments on the recognition of speech, with one and two ears.” J. Acoustical Society of America, 25:554–559.

[7] Begault, D. R. and Erbe, T. (1994). “Multichannel spatial auditory display for speech communications.” J. Audio Engineering Society, 42:819–826.

[8] Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. Y. (2001). “Spatial unmasking of nearby speech sources in a simulated anechoic environment,” J. Acoustical Society of America, 110:118–1129.

[9] Wenzel, E., Arruda, M., Kistler, D.J, & Wightman, F.L. (1993) “Localization using non-individualized head-related transfer functions,” J. Acoustical Society of America, 94(1):111-123.

[10] Martens, W. L. (2003) “Perceptual evaluation of filters controlling source direction: Customized and generalized HRTFs for binaural synthesis,” Acoustical Science and Technology, 24(5), 220-232.

[11] Zotkin, D. N., Duraiswami, R., Grassi, E. and Gumerov, N. A. (2006) ”Fast head-related transfer function measurement via reciprocity,” J. Acoustical Society of America 120(4):2202-2214.

[12] Martens, W. L. (1987) “Principal components analysis and resynthesis of spectral cues to perceived direction,” Proc. Int. Computer Music Conf., Champaine-Urbana, IL, Sept.

[13] Thurstone, L. L. (1927). “A law of comparative judgment,” Psychological Review, 34:273-286.

[14] Martens, W. L. (2002) “Rapid psychophysical calibration using bisection scaling for individualized control of source elevation in auditory display,” Proc. Int. Conf. on Auditory Display, Kyoto, Japan, ICAD, pp. 199–206.

[15] Martens, W. L. (2003) “Individualized and generalized earphone correction filters for spatial sound reproduction.” In: Proceedings of the 9th International Convention on Auditory Display, Boston University, USA, July, pp. 263-266.

[16] Martens, W. L. and Zacharov, N. (2003) “Spatial distribution of reflections affects auditory quality and character of speech sounds located in a virtual acoustic environment,” Proc. 1st Workshop on the Auditory Quality of Systems, Rühr Univ. Bochum, Germany, Apr., pp. 91–96.