Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Effects of individualised headphone response equalization on front/back hemifield discrimination for virtual sources displayed on the horizontal plane

William L Martens, Abhishek Guru, and Doheon Lee

Faculty of Architecture, Design and Planning, The University of Sydney, NSW, Australia

PACS: 43.66.Qp , 43.60.Jn

ABSTRACT

In the most demanding virtual auditory display applications, in which individualised Head Related Transfer Functions (HRTFs) are used for the presentation of virtual sound sources via headphones, there is controversy regarding how important it may be for individualised Headphone Transfer Function (HpTF) measurements to be used in equalizing the headphone response for each listener. In order to test what impact the use of such individualized HpTF-based correction might have on directional judgments, filtered noise bursts were presented with and without such headphone correction during a test of front/back hemifield discrimination for virtual sound sources positioned on six sagittal planes offset from the median plane by 15^o, 30^o, and 45^o to either side. While perfect discrimination performance was observed given repeated two-interval forced choice discrimination trials in which a pair of short noise bursts were presented using individualised HRTFs, within-trial variation in the spectrum of the source submitted to HRTF-based processing made the task quite difficult, reducing performance to chance levels for 7 of the 17 listeners tested. For the remaining listeners who showed above-chance performance under all conditions tested, performance levels were well below the perfect performance that had been observed when the spectrum of the HRTF-processed source was held constant. Through inter-stimulus variation in source spectra, which functioned to remove the so-called “known-source-spectrum ceiling effect” associated with simple laboratory tests of virtual auditory display technology, it was possible to show that front/back discrimination performance was clearly affected when sources were processed using headphone correction filters that were based upon a each individual’s measured HpTF.

INTRODUCTION

In the practical application of binaural technology for presentation of virtual sound sources via headphones, there is some controversy surrounding the extent to which individual differences must be taken into account when processing audio using Head Related Transfer Functions (HRTFs). In the most demanding virtual auditory display (VAD) applications, such as those in avionics or other applications which emphasize directional accuracy of presented virtual sources, there is some consensus that using individualised HRTFs is to be recommended; however, it is not so well established how important it may be for the headphone responses for individual users to be equalized using measurements of their own individualised Headphone Transfer Function (HpTF). Although there are suggestions in the literature that such individual differences introduce undesirable variation in the presented signals that might be great enough to be a real concern in the deployment of binaural technology [1], there is relatively little evidence that the use of individualised HpTF-based equalization has a substantial impact on human responses in sound localization studies. The literature review done in preparation for the study reported herein revealed only one case of reduction in localization error with such equalization, and the reported reduction was quite small [2].

In response to the lack of objective data upon which a credible conclusion could be drawn, it seemed that there was a need to test the importance of including individualised headphone equalization in successfully deploying binaural technology, that is, applying correction based upon individually measured headphone responses. Placing this work in context requires an analysis of the relation between spatial hearing research results and realistic expectations regarding applications of binaural technology. First, it is assumed that most applications of HRTF-based processing for presentation of virtual sound sources via headphones will involve sources of unknown spectrum. While some VAD applications might involve the spatial positioning of a small number of audio signals, such as a warnings or alarms, these applications are regarded as a minor subset of those for which HRTF-based processing is likely to be employed. In fact, in applications using a single sound source, or small number of sound sources, designers would be well advised to verify empirically whether the directional perception of displayed sources, using repeated listening tests that allow for iteratively optimising the quality of the perceptual result. When this is not possible or practical, it will be important to know that related research results are being interpreted appropriately in considering how binaural technology is deployed.

So, to begin with, this introduction will discuss the testing paradigms most often used in spatial hearing research that is intended to verify or validate particular VAD applications of binaural technology. Shaw [3] described a fundamental problem in the interpretation of such applied spatial hearing research results, which he explained rests upon the potential for misunderstanding between two distinct testing paradigms, which he termed “sound localization” versus “auditory spatial perception.” In the context of directional hearing studies in particular, results of which have been driving how HRTF-based processing is deployed in practical VAD applications, this distinction can help to clarify why convolution of audio signals with HRTFs doesn’t always produce desired perceptual results. The task-oriented distinction that Shaw [3] introduced focuses on which of two questions is asked of listeners to whom an experimenter presents a spatial sound stimulus (i.e., a sound source in the space surrounding the listener). These two questions are:

“Where in physical space do you think the source was?”

“Where is the image of the source as perceived by you?”

The first question is targeted at determining how well listeners are in touch with the physical reality of sounds located in the space surrounding them, emanating from distal locations. Within this testing paradigm, for example, when asking for directional discrimination between sound sources in terms of their actual location in a listener’s front or rear hemifield, it is possible to give feedback to listeners after each response regarding whether their response was correct or not. In contrast, the second question is not about where the physical sound source actually is; rather, it is about the perceived location in a listener’s auditory space that is taken by the auditory image associated with a presented sound source. No feedback on response correctness is possible in this second case, as an auditory image that is heard to be in the rear hemifield is correctly reported to be in that hemifield if and only if it is heard to be there. If the physical sound source were in fact located in the front hemifield, but it was heard to be in the rear hemifield, the only correct response listeners can make here would be to report that the auditory image of the source was behind them, where it was heard to be.

The importance of this distinction is often overlooked in experimental tests designed to verify the performance of VAD systems using HRTF-based source processing for headphone reproduction. There is an abundance of performance tests that are based upon the sound-localization testing paradigm. Such tests can produce results showing that sound-localization performance nearly matches that observed in free-field listening, but these studies do not necessarily inform designers about where auditory images associated with those HRTF-processed sources will be heard to be. They only report on whether listeners are able to report the “correct” source locations, assuming that well-engineered HRTF-processing of arbitrary sources should produce an auditory image that matches the original location of the analytic sound signal that was used to measure a binaural pair of HRTFs (an assumption certainly worth questioning, as is well discussed in Blauert’s [4] seminal text on “Spatial Hearing”). The revealing observation regarding the problem identified here is that the sound-localization task can be performed successfully without hearing auditory images in the target locations.

Of course, sources that are incident from locations left of the median plane are indeed most often associated with auditory images that are heard as arriving form the left hemifield, and likewise for sources in the right hemifield. But what can be said about sources arriving from the front versus the rear hemifields? Strangely, when head motion is eliminated (via a “bite bar” or some other mean for immobilizing the head), listeners find that brief sounds presented via a loudspeaker located directly in front of them can reliably be identified as arriving from the front, even though those brief sounds are nonetheless heard to be arriving from the rear. This is a surprising realization for many listeners, since their experience in listening to actual sources with the naked ear seems to provide good agreement between auditory perception and physical location of sound sources. The discrepancy between sound localization and auditory spatial perception is not so easy to explain, but at least the sound localization performance can be explained in this case: Correct front/back discrimination performance can be enabled by attending to the difference in tone coloration observed for sources in the front and rear hemifields. The perception in the rear hemifield of auditory images associated with frontal incidence is more difficult to explain, but seems to be related to a strong bias in how human listeners resolve what is, in fact, an ambiguous spatial percept (which Von Békésy [5] likened to the perceptual reversals common when viewing a line drawing of a cube). This perceptual bias has been addressed via several alternative explanations, but for the present discussion it suffices to say that the resolution of front/back ambiguity towards the “in-back” percept is reliably observed in a majority of listeners, although there is a minority of listeners who reliably show the opposite bias towards the “in-front” percept (see Blauert [4] for a more thorough discussion of this bias).

Since the phenomenon is not typically observed when head movements are allowed [4], there is typically no opportunity for listeners to become aware of the operation of the perceptual bias discussed here. However, when virtual sources are presented via headphones, the consequences of this bias for the successful application of binaural technology are devastating. It is quite disappointing to most designers how frequently it is observed by listeners that headphone-presented sources that are intended to be arriving from locations in the front hemifield are nonetheless heard to be arriving from the rear. Years of reported research results have created an expectation that well-engineered VAD systems should support good auditory spatial imagery, and yet satisfying auditory images of frontal sources are almost never experienced by listeners when using VAD systems that do not employ head-tracking technology (that is, HRTF-based processing that is updated whenever the VAD system user’s head motion is registered by a head tracker). How VAD system users have been misled here regarding how well HRTF-based processing should work, it should be clear, is based upon reported excellent results in sound-localization tests. The results of these tests may seem to be confirming nearly perfect system performance, with accuracy measures showing nearly perfect correlation between original and reported locations. So, despite a VAD system’s failure to produce auditory images of sources that are heard to be arriving from intended directions in auditory space, sound-localization tests of that same VAD system suggest that the deployed set of measured HRTFs are working very well. It is no wonder in this regard that thwarted expectations are so frequently reported by first-time VAD system users.

One goal of this introduction has been to attempt to raise awareness of this dilemma for readers working within the auditory display field. A second goal is to propose an alternative testing method that might better serve to clarify the performance of a VAD system in terms of what directional discrimination can be well supported in its application using input sources with variable spectral content. Although typical use conditions might be quite different from laboratory test conditions, the tasks that are required of listeners in experimental tests can be designed to provide more sensitive indications of VAD system performance. The current paper presents one such testing method designed to address the problematic front/back directional distinction in VAD use.

One aspect of the proposed means for improving experimental sensitivity to variations in VAD system performance, at least with regard to the dilemma of front/back confusions, is to employ a two-interval testing paradigm that requires a report on which source seems to arrive from a more frontal direction. How this approach avoids one of the pitfalls of the "in-back" perceptual bias will be described in more detail below. Another aspect of the proposed means of experimentally examining front/back confusions associated with a VAD system is to vary the source spectra within each two-interval trial within the proposed testing paradigm. The idea here is to submit to the front and rear HRTFs under examination a variety of sources that differ in tone coloration in a manner that undermines the reliability of sound localization cues based upon the gross spectral differences between front and rear HRTFs (HRTF differences that can be quantified in terms of the spectral centre of gravity known as the spectral centroid). This was done in a recent study by the authors [6], producing the result that listeners who had performed perfectly on front/back discrimination with a known source signal (a brief burst of white noise) showed much poorer performance when the spectral centroid of the input source signal was allowed to vary between presentations. A good number of the listeners tested in that study, however, were able to maintain better than chance performance, perhaps basing their front/back discrimination responses on the more subtle binaural cues remaining after a broad spectral-tilting manipulation of the sources submitted to HRTF-based processing.

By forcing listeners to focus on these more subtle cues during inter-stimulus variation in source spectral centroid, it was possible to remove the so-called "known-source-spectrum ceiling effect" associated with simple laboratory tests using white noise bursts. Indeed, for the 9 out of the 21 tested listeners who maintained above chance performance on the more difficult task, the virtual sources presented using individualised headphone correction filters supported significantly better front/back discrimination rates than did virtual sources presented under the same conditions, but without correction to headphone responses. These experimental testing conditions revealed the potential importance of correcting for an individual listener's measured HpTF, since the aforementioned gross spectral differences between front and rear HRTFs were undermined as potential cues to the measurement's hemifield. It was concluded in that study [6] that these gross cues, associated with easily noticed shifts in the auditory attribute typically termed "sharpness," had kept discrimination performance nearly perfect, producing a "ceiling effect" that made the simpler white-noise test insensitive to the presence or absence of correction for individualised HpTFs). An important qualification of those previous results, however, is that the tests were run only for sources located on the median plane of the listener, and in fact, only for HRTFs measured under anechoic conditions at a distance of 2m directly in front and in back of the 21 listeners on the horizontal plane.

Naturally, it is difficult to generalize from tests run under such stimulus constraints. So, what was not tested in that previous study, and is therefore proposed as the focus of the current study, is a test to determine whether similar evidence for the effects of individualised headphone correction will be found for sources shifted laterally away from the listener's median plane, to positions lying upon six sagittal planes offset from the median plane by ±15°, ±30°, and ±45° (i.e., equal angular offsets to either side). Experimental sound sources presented at equal lateral angles to the right and left of the median plane can be considered together, and so the matched pairs of front/back oriented sources that targets will take in the current study can be pictured as positioned at just the three lateral angles shown in Figure 1.

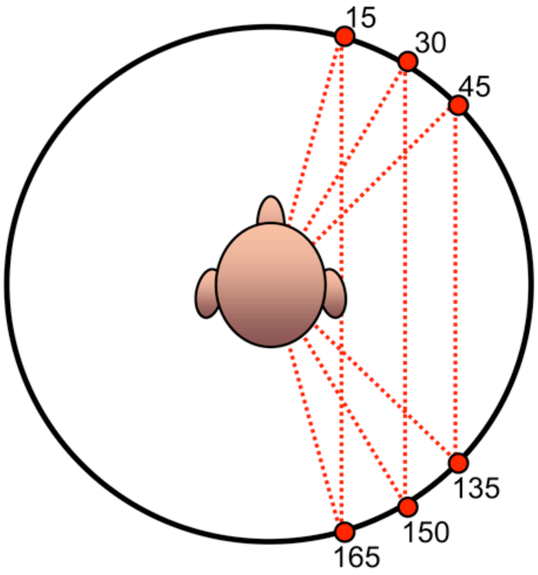

Figure 1: Target positions for virtual sources lying on three sagittal planes offset from the median plane by 15, 30, and 45 deg., illustrating complementary points within the front and rear hemifields, reflected with respect to the interaural axis. The dashed lines indicate the labelled azimuth angles of sources relative to a front-centre location, and also connect the front and back source angles that share a common lateral angle (offset form the median plane) and therefore lie on a common sagittal plane.

METHOD

Stimuli

Stimuli were generated by convolving brief bursts of white noise with 256-point Head Related Impulse Responses (HRIRs) that were generated using Farina's logarithmic sine sweep technique [7]. The following is a summary of the employed transfer function measurement procedures, which are the same as those used in the authors' previously reported study [6], and described in much more detail in [8].

HRTF and HpTF measurements were made for each of the 21 listeners participating in this study using a pair of DPA 4060 miniature microphones that were embedded in a layer of Anti-Nois™ ear plug wax and positioned approximately at the centre of the entrance to both of a listener's ear canals. Binaural HRTFs were captured for just two spatial positions on the listener's horizontal plane, one directly in front and one directly in back of the listener, corresponding to 0˚ and 180˚ azimuth angles respectively. These were not free-field HRTFs, but were transfer functions that included the response of the Bose Acoustimass Cube speaker that produced the analytic signal at a distance of 2 metres from the centre of the listener's head. For reproduction of the stimuli via Sennheiser HD600 headphones, repeated measurements of the headphone response (coupled to each listener's head) were made with each of the DPA 4060 microphones in the same position as that used for the HRTF measurements (i.e., the HpTFs were made during the same session as the HRTF measurements, and the microphones were not disturbed between measurements, so that HRTF and HpTF were strictly comparable for sake of optimal headphone response correction for the HRTF-based processing of audio sources).

The intention of this measurement approach, at least with regard to the authors' previously reported study [6], was to enable reproduction of signals via the Sennheiser HD600 headphones that nearly matched the signals from the speakers that were captured at the same microphone positions. In that study, nearly identical signals were created at the blocked ear canal by headphone and loudspeaker reproduction for two different incidence angles on the horizontal plane. For the current study, in order to create virtual sound sources at twelve new azimuth angles (six of which are pictured in Figure 1), the previously generated median plane stimuli were further processed to introduce appropriate interaural differences for the six sagittal planes offset by $\pm 15^o$, $\pm 30^o$, and $\pm 45^o$ from the median plane, following the method reported by Morimoto, et al. [9]. Of course, this means that the measured HRTFs were not those for the target azimuth angles, but were algorithmically modified versions of the HRTFs measured in the previous study at 0 and 180˚ azimuth. The algorithm that was employed to introduce these interaural differences is completely specified by a Matlab script that is freely available at the website associated with the book entitled DAFX - Digital Audio Effects [10]. The Matlab code was authored by David Roccesso and is well explained within his contributed Chapter 3 of that book, which he entitled "Spatial Effects" [11]. The employed audio signal processing will be summarized only briefly here.

Having established in the authors' previous studies that the HRTFs for each individual listener's front and back source directions were adequately well engineered to support good sound localization performance, a new set of stimuli were generated for 17 of the original 21 listeners beginning with the same stimuli that had been presented to them in that previous study. A simple filtering model was employed to producing these new virtual sources that simulated a head-shadow-based higher-frequency boost for the ipsilateral signals and a head-shadow-based higher-frequency attenuation for the contralateral signals. Typical interaural time differences corresponding to the observed average variation between HRIRs for such source directions were imposed as well, so that the lateral angles associated with the output signals would shift by $\pm 15^o$, $\pm 30^o$, and $\pm 45^o$ from the staring azimuths of 0^o and 180^o (as illustrated in Figure 1). Such an approach to producing non-median plane virtual sources from median plane virtual sources has a strong precedent in the literature, and has been well established to produce desired directional percepts despite the deviation from the more accurate transformations that would result from using measured HRTFs for each of the target directions (see, for example, the paper by Morimoto, et al. [9] entitled "Upper hemisphere localization using head-related transfer functions in the median plane and interaural differences").

Just as in the authors' previous related study [6] the input stimuli submitted to the above extended HRTF-based processing was varied in spectral centroid using the MATLAB routine developed by D. Cabrera (makenoise.m). The spectral exponent was given one of three values ($x = 0, 0.5$ or 1), where these values satisfied the following relationship between M, the Power magnitude of the signal as function of frequency, in Watts, and f (frequency), in Hz:

$$\frac{\partial M}{\partial f} = f^{x}$$

A 140 ms source signal was gated on and off with a rise/fall time of 20 ms, and spectra were adjusted by the above method, but just in the case of the "back" stimuli, to have three spectral slopes, such that the spectral centroids of the "back" stimuli progressively increased to approach the higher value typical of a "front" stimulus. This was done to make it difficult to use the above-discussed gross variation in tone coloration as a cue for making front/back discriminations. This modification of the "back" stimuli produced a series of stimuli varying in gain over frequency. The starting point was a white noise signal, which, though spectrally flat, increases in Power magnitude by 3 dB every octave. The other two stimuli were given steeper spectra, gaining 4.5 dB and 6 dB every octave, respectively (examples of which are shown in Figure 2).

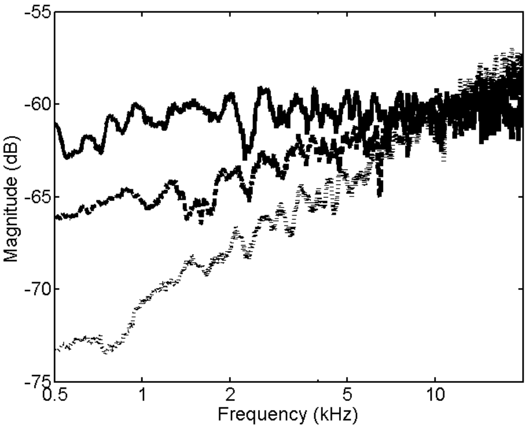

Figure 2: Spectra of noise input signals submitted for HRTF-based processing, specifically and only for stimuli submitted to the individually-measured "back" HRTFs, giving them increasing spectral slope that progressively increased to have the rear stimuli approach the higher spectral centroid value typical of a "front" stimulus.

Procedure

On each trial, listeners were presented sequentially with a pair of stimuli matched in their targeted lateral angles, but one was processed using a "front" HRTF and the other using a "back" HRTF. This is commonly known as a 2-Interval Forced Choice (2IFC) task, where one of the two stimuli is presented in each of the two temporal intervals here (as opposed to simultaneous presentation of stimuli). Listeners were "forced" after each presentation to respond either 'F' or 'B' to indicate whether the net displacement in direction from one sound to the next was frontward or backward. It was decided to present a pair or noise bursts in sequence for this discrimination rather than an individual noise burst in one of two locations, as it was generally found that in case of a single noise burst presented to a listener, there was a tendency to report only "back" percepts, irrespective of which HRTF was used.

In each session 144 such pairs of noise bursts were presented for this modified "sound-localization" judgement. While each of the trials consisted of one front-HRTF and one back-HRTF processed noise burst, they could be ordered as back-to-front (defined as "frontward") or front-to-back (defined as "backward"). This order manipulation defines the first factor in the experimental design, which included five other factors. A second factor was whether the sources were presented from the left or the right side. These first two factors were treated as "nuisance" variables, across which response data were summed (which preliminary analysis justified). The remaining factors were manipulations more suitable for subsequent analysis, as these "analytic" factors were expected to have more substantial effects on discrimination performance.

These four analytic factors are listed in Table 1, which gives the number of levels associated with each, and identifies the levels with brief descriptive labels.

Factor (N Levels)	Levels
SOURCE SPECTRUM (3)	Original, Brighter, Brightest
EQUALISATION (2)	With ECF, without ECF
LATERALIZATION (3)	15, 30, 45 degrees
CORRECTNESS (2)	Correct, Incorrect (see text)

Table 1: Analytic Factors

The factor termed SOURCE SPECTRUM corresponds to the variation in the spectrum of the noise sources submitted to processing using the “back” HRTF signals, with sources labelled as the “original” (flat, or white), one that was “brighter,” and one that was “brightest” of all (in which case the measured spectral centroid values approached very near to the spectral centroid of the front-HRTF processed stimuli). Inputs were in turn either equalized for headphone reproduction using a monaural Earphone Correction Filter (ECF) or not filtered (presented with no ECF), and this analytic factor was termed EQUALIZATION. The third analytic factor was LATERALIZATION, corresponding to the lateral angles associated with the output signals (which were shifted by $\pm15^o$, $\pm30^o$, and $\pm45^o$ from the median-plane azimuths of 0^o and 180^o (as illustrated in Figure 1). The final factor is the outcome variable, which was the CORRECTNESS of response on the 2IFC task. A response was defined as correct if it matched correctly the order of stimuli presented, where these stimuli are identified by the original physical hemifield in which the employed HRTFs were measured. Again, it should be stressed here that the task did not require any indication of perceived direction of a given presented sound source, but only a response regarding the relative directional offset between stimuli presented in the two temporal intervals: F (for Frontward) or B (for Backward) based upon a best estimate regarding sound (measurement) locations.

RESULTS

For easy interpretation of results and subsequent discussion, the following means for treating the correctness of responses should first be summarized. The response “Frontward” refers to a judgement of displacement between two sounds presented sequentially in an overall frontward direction; while the word “Backward” refers to a displacement between a paired set of sounds in an overall backward direction; the proportion of correct responses was formed as a ratio of correct to overall responses for a given set of conditions. Again it is stated for clarity here, a “correct” response was explicitly defined here to be a response that matches correctly the order of stimuli presented, treating the offset between actual physical locations at which HRTFs were measured as the ground truth that listeners were attempting to discover while listening to the pairs of stimuli.

The likelihood of observing by chance alone 80 correct responses on 144 trials is less than 1 in 100. So at a risk of error set to a probability of $p<.01$, any listeners who got more than 56% responses correct over all 144 trials have been regarded as discriminating frontward from rearward motion at a rate greater than that expected by chance alone. Following this criterion for selecting listeners who were successful in the frontward/rearward discrimination performance, only 6 of the 17 listeners tested were included in a group termed “Normal Discriminators” for the subsequent analysis of response frequencies. A second group of 4 listeners were placed in a group termed “Reverse Discriminators” because they made significantly more “front to back” responses when presented with stimuli moving in the reverse direction, and vice versa for their “back to front” responses (i.e., they were “wrong” in detecting source displacement according to which presentation order of stimuli that had been processed and presented using their own individualized HRTFs, but nonetheless were able to make consistent “wrong” discriminations with percent correct below 44%). The remaining 7 listeners were labelled as “Chance-discriminators”, since their discrimination performance was between 44% and 56% correct responses. The upper panel of Figure 3 shows the results for the Normal Discriminators, expressed in terms of the absolute value of z-transformed front/back response proportions, and the lower panel of Figure 3 shows the results for the Reversed Discriminators. The ordinate uses these values rather than using percent correct data, since percent correct is more difficult to compare between Normal and Reversed discriminators. Thus a good comparison of results can be made for performance as plotted over the target lateral angle of the processed virtual sound source.

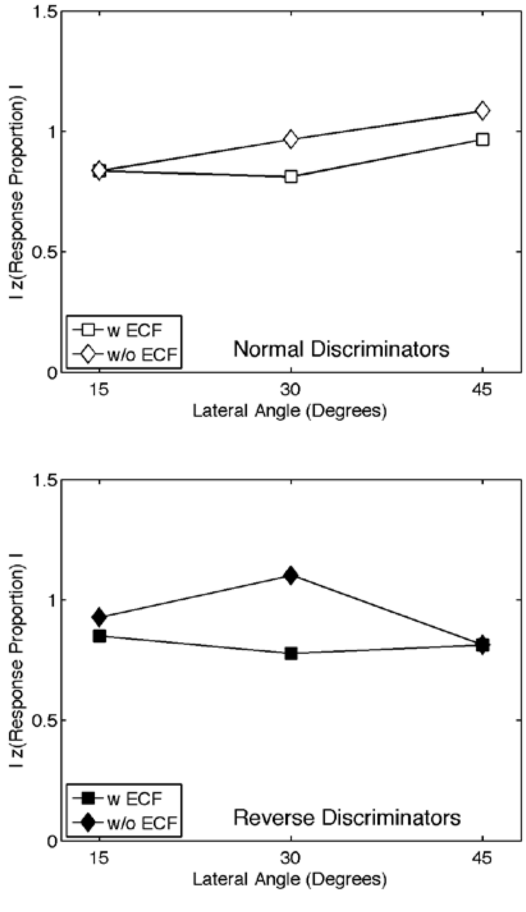

Figure 3: Absolute value of z-transformed front/back response proportions, plotted as a function of the target lateral angle of the processed virtual sound source. The upper panel shows the result for Normal Discriminators (responding ‘front-to-back motion’ for sources processed using individualized HRTFs for front followed by rear speaker locations); the lower panel shows the result for discriminators making the reverse indication (‘back-to-front motion’ reported for sources processed using individualized HRTFs for front followed by rear speaker locations). Square symbols were used to plot the outcome when stimuli were presented using an Earphone Correction Filter (ECF), while diamond symbols were used to plot the outcome when stimuli were presented without ECF.

Normal Discriminators were able to correctly identify 'front-to-back motion' for sources processed using individualized HRTFs measured at front versus rear speaker locations, despite the difficulty introduced by the variation in the rear stimulus spectrum, but showed little dependence upon the ECF. In contrast, the performance of the Reversed Discriminators did show a dependence on the presence of the ECF, noteably at ±30^o from the median plane.

Overall, there appears to be some dependence in this outcome upon the target lateral angle of the front/back HRTF-processed sources, but this effect is quite small. There also appears to be some dependence upon the presence or absence of an ECF in the processing, but again, this effect is quite small. Of particular interest here is that the influence of the ECF on discrimination performance, though small, appears to be in an unexpected direction (and would contrast with the results found in [6] on the median plane). In order to determine whether the variation in outcome was significantly dependent upon the two factors, the summed forward/backward response frequencies across all Normal Discriminators were submitted to a multi-way contingency table analysis. This analysis allows for a test of the association between factors influencing discrimination performance. A preliminary Hierarchical Loglinear Analysis (HLA) showed that the response frequencies could, without loss of information, be collapsed across two of the manipulated factors, that being the order of front/back presentation, and the switching between left and ride side source locations (allowing six lateral angles to be treated as just three lateral angles).

By treating left and right side stimulus presentation as having no noteworthy effect upon discrimination performance (this having been apparently symmetrical on either side of the median plane), and not dependent upon presentation order, the designed HLA model could be tested using the summed response frequencies expressed in a four-way contingency table according to the analytic factors listed in Table 1. Given the obtained response frequencies for each of the two groups of listeners, Normal and Reversed, HLA was run using backward elimination of associations, a process that begins with a saturated model that includes all associations between all factors. In each case, only two associations were required to fit the data from these two groups, with Log-Likelihood-Ratio Chi-square values (df=28) of 14.176 and 4.679 for the normal discriminators and the reversed discriminators, respectively. The associations found to be statistically significant for these two groups of listeners will be explained below after some explication of the analysis method.

Hierarchical Loglinear Analysis (HLA) can be run in a "model selection" mode, to determine how many association terms are required for a good fit of the loglinear model to the observed frequencies in each cell of the four-way contingency table. Using the first letters of the factors listed in Table 1 to identify each in reporting results, the saturated model that includes all associations between all factors would be identified by the label S*E*L*C for the multi-way association between factors SOURCE, EQUALIZATION, LATERALIZATION, and CORRECTNESS. Of course, the finding of associations between CORRECTNESS and the other three is the primary interest here. It also simplifies interpretation if the high-order associations are not required to fit the data. The following is a brief summary of how the HLA method works, and the results that were obtained using it to analyse the four-way contingency data.

At each step in the backward elimination process, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than .05. When deleting an additional effect produces no significant change in the fit of the loglinear model to the observed cell frequencies, the elimination is terminated. In the case of the Normal Discriminators, the only significant effects on cell frequencies were those due to the associations labelled E*C and S*C, i.e., the association between correctness of response and the manipulated factors of source Spectrum and headphone Equalization. In the case of the Reversed Discriminators, the only significant effects on cell frequencies were those due to the associations labelled E*C and L*C. So, again the association between correctness of response and headphone Equalization was significant for this group of listeners, but in contrast to the Normal Discriminators, the responses of the Reversed Discriminators showed no dependence on source Spectrum; rather, the model fit to their cell frequencies required the inclusion of the association between correctness of response and source Lateralization. If the dependence on headphone Equalization were to be removed from the loglinear model, the goodness of fit for the reversed discriminators would drop from .999 down to .289. If the comparable dependence on headphone Equalization were to be removed from the model for the normal discriminators, the goodness of fit would drop from .986 down to .376. These observations lead to the natural conclusion that headphone Equalization has a significant impact on front/back discrimination performance, regardless of whether listener responses are influenced mostly by spectral variation across stimuli, or differences due to lateralization of stimuli, and also regardless of whether listener responses are normal or reversed with respect to the individually measured front/back differences in HRTFs.

DISCUSSION

Before discussing the results more generally, it would be best to look at the difference observed in performance of listeners showing normal versus reversed discrimination. An informal debriefing suggested that different auditory attributes of the stimuli were employed in attempting to make consistent "correct" responses. In fact, there was greater reported externalisation of the stimuli presented in this study with lateralized source directions than that experienced for the median plane stimuli presented in the authors' previous study [6] involving the same listeners. However, it was difficult to relate the subjective reports of these listeners to their normal versus reversed grouping. Nonetheless, it was clear that attention to the apparent "spatial extent" of the auditory image between front and back hemifield stimuli could have aided in making the discrimination, regardless of whether listeners were in the normal versus reversed group. The question also arises regarding the presence of some obvious physical difference between the acoustical responses measured for listeners in the normal versus reversed discrimination groups. Such comparisons can be made in two domains for these groups. Since EQUALIZATION had a significant effect upon discrimination in both groups, this is the first domain to be examined. Figure 4 shows the averaged ECFs for the two groups of listeners.

In examining the average difference in ECF applied to the stimuli to correct for the HpTFs measured individually for members of these two groups of listeners, there is no obvious difference that could account for the performance differences between the groups. Only a slight 8 and 9 kHz shift in a peak in the ECF gain curves plotted in Figure 4 would seem to provide a potentially noticeable difference here, but there is evidence that such small shifts in ECFs are quite difficult to detect [12].

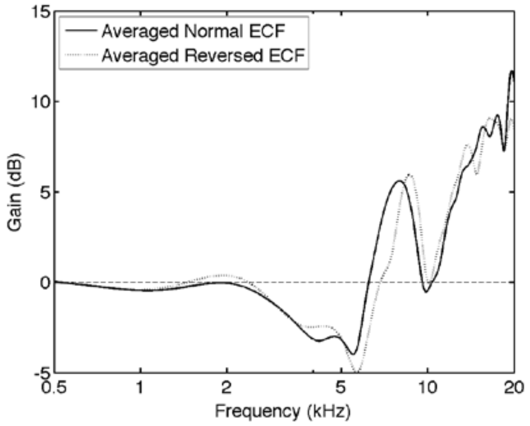

Figure 4: Average gain curve for two groups of listeners showing the slight difference in the Earphone Correction Filters (ECFs) that were derived for each listener from analysis of their individually measured Headphone Transfer Functions (HpTFs).

A second domain to investigate would be the differences between HRTFs measured when speakers were located in front versus in back of these 17 listeners. Figure 5 shows the averaged HRTFs measured for these two groups for frontward versus rearward incidence.

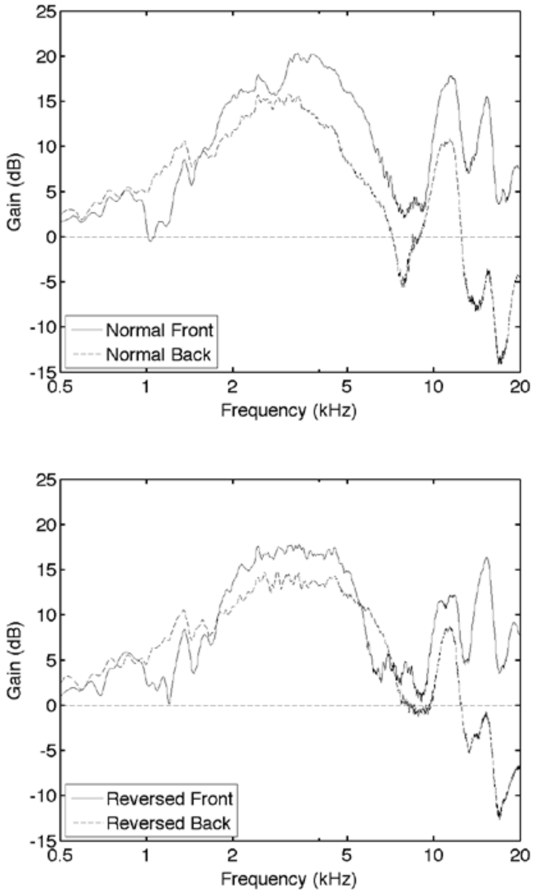

Figure 5: Gain of measured HRTFs corresponding to front and back speaker locations averaged over measurements for the 6 Normal Discriminators (upper panel) and for the 4 Reversed Discriminators (lower panel); solid lines show average gain for front speaker locations, while dashed lines show average gain for back speaker locations.

By looking at the difference between the shapes of the HRTF gain curves plotted in Figure 5, it seems that there is a difference between the two groups of listeners in the 2-5 kHz region. The Normal Discriminators have HRTF gain curves that appear more rounded in this frequency region than do those of the Reversed Discriminators. It is difficult to determine whether this difference could explain the greater sensitivity to the SOURCE SPECTRUM factor observed in the case of the Normal Discriminators. Suffice it to say that the differences in averaged HRTFs between the two groups do not suggest any obvious means that would allow subsequent listeners to be sorted into these two behaviourally-defined categories of Normal and Reversed Discriminators.

So, given that it is generally difficult to see any physical differences that might help to understand how individual listeners could be categorized as either a Normal Discriminator or a Reversed Discriminator, it remains to be determined whether there is some explanation in the perceptual domain. To begin with, it should be remembered that both groups give “correct” response frequencies that are modulated by the presence or absence of Earphone Correction Filters (ECFs), but that the detected dependence is not the same for each of the two groups, and is not so simple to describe. This complexity is underscored by the finding that frequencies summed over all Normal Discriminators reveal a single additional dependence on source spectrum, while those summed over all Reversed Discriminators reveal a single additional dependence on source lateralization.

It is thus clear that these two groups are attending to disparate attributes of the stimuli. What could be proposed here is that the spectral variation imposed upon the rear stimulus must influence the Normal Discriminators more because they are focussing upon gross spectral features more than are the Reversed Discriminators. Perhaps this could mean that their observed sensitivity to differences in EQUALIZATION depends upon gross spectral features. In contrast, the observed sensitivity to differences in EQUALIZATION for the Reversed Discriminators might be due to these listeners hearing unexpected variation in interaural spectral differences introduced through the experimental stimulus generation that employed a simplified model of HRTF-based variation with lateral angle. Further investigation of this and other associated auditory attributes may be useful here, and it is therefore suggested that future studies might examine whether there is any basis in attribute identification data for these speculations.

A more general discussion of the implications of the current findings in the context of VAD applications is also warranted here. The first most striking aspect of the results reported here is that they appear inconsistent with the authors’ previously reported results [6] that were observed for sources located on the listener’s median plane. Indeed, whereas the earlier results suggested that using individualised HpTFs along with individualised HRTFs gave improved front/back discrimination performance, the inclusion of individualised HpTFs for virtual source display shifted away from the median plane had a negative, albeit small, impact upon front/back discrimination performance. It did seem to make a difference to the pattern of this impact whether listeners were biased toward normal versus reversed discrimination, since reversed discriminators showed a significant decrease in performance only at one of the three lateral angles tested. Perhaps these results are idiosyncratic to the interaural spectral differences to which the two groups might have been accustomed, as was found to influence judgments of naturalness reported for listeners using HRTFs measured for other subjects [13].

An important caveat that should be stressed here at the end of this discussion is that the difficulty associated with front/back confusion of virtual sources is almost completely eliminated when head-tracking VAD technology is used (as has been long known [14]). The fact that front/back discrimination performance is so poor without head tracking, especially under conditions in which source spectra are varied, underscores the importance of dynamic cues associated with head motion, even when a VAD system employs individually measured HRTFs and HpTFs.

CONCLUSION

This study explored a technical detail that by itself might seem relatively unimportant, but may indeed have a small yet significant impact upon of the performance of VAD systems designed to position virtual sound sources throughout the space surrounding the listener. It is natural to consider further exploration of the role of system calibration using individually measured HpTFs on other attributes of system performance such as the perceived naturalness of the spatial auditory imagery that can be produced (as in [13]). Chief among these opportunities might be a study of the applications of such a system to represent more realistic sound sources, such as speech sound sources, that function spatially, temporally and timbrally in a manner as experienced in day-to-day life. The current study of front-back discrimination extended the authors' previous study [6] to include virtual sound sources that were offset from the exact front or back centre position to range in lateral angles between 15º and 45º. While it is still difficult to generalize from the current results to more comprehensive simulations and applications, the results certainly suggest that individualised headphone response equalization can have an effect on front/back hemifield discrimination for virtual sources displayed on the horizontal plane.

REFERENCES

[1] D. Pralong and S. Carlile, “The Role of Individualized headphone calibration for the generation of high-fidelity virtual auditory space,” *J. Acoust. Soc. Am.*, vol. 100, no. 6, pp.3785—3793 (1993).

[2] H. Møller, C.B. Jensen, D. Hammershøi and M.F. Sørensen “Using a Typical Human Subject for Binaural Recording” *AES 100th Convention,* (1996).

[3] E. A. G. Shaw, “1979 Rayleigh Medal Lecture: The Ellusive Connection.” In Gatehouse, R. (Ed.), *Localization of Sound: Theory and Applications*. pp. 1-32, Groton: Amphora Press (1982).

[4] J. Blauert, *Spatial Hearing– The Psychophysics of Human Sound Localization*. Boston: MIT Press (1997).

[5] G. Von Békésy, “*Experiments in Hearing*” (E.G. Wever, *Trans.*), New York: McGraw-Hill Book Company (1960).

[6] A. Guru, W. L. Martens, and D. Lee “Effects of Individualized Headphone Equalization on HRTF-based Front/Back Discrimination of Virtual Sound Sources,” *Proc. 159th Meeting of the Acoustical Society of America*, Baltimore, Maryland, USA (2010).

[7] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept sine technique,” in *Proc. AES 108th Convention* (2000).

[8] A. Guru, W. L. Martens, and D. Lee “Effects of Individualised Headphone Correction on Front / Back Discrimination of Virtual Sound Sources Displayed using Individualised Head Related Transfer Functions." In: *AES 40th International Conference, Tokyo*, Japan (2010).

[9] M. Morimoto, K. Iida, and M. Itoh, "Upper hemisphere localization using head-related transfer functions in the median plane and interaural differences," *Acoustical Science and Technology*, 24, 267-275 (2003).

[10] Website for “DAFX – Digital Audio Effects,” http://www2.hsu-hh.de/ant/dafx2002/DAFX_Book_Page/

[11] U. Zölzer (Ed.), “DAFX – Digital Audio Effects,” John Wiley & Sons (2002).

[12] W. L. Martens, “Individualized and Generalized Earphone Correction Filters for Spatial Sound Reproduction,” in *Proc. Int. Conf. Auditory Display*, Boston, MA (2003).

[13] J. Usher and W. L. Martens, “Perceived Naturalness of Speech Sounds Presented Using Personalized Versus Non-Personalized HRTFs,” in *Proc. Int. Conf. Auditory Display,* Montréal, Quebéc (2007).

[14] J. Kawaura, Y. Suzuki, F. Asano, T. Sone. Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear. *J. Acoust. Soc. Jpn*.(E), **12**, 203-216 (1991).