RAPID PSYCHOPHYSICAL CALIBRATION USING BISECTION SCALING FOR INDIVIDUALIZED CONTROL OF SOURCE ELEVATION IN AUDITORY DISPLAY

William L. Martens

Multimedia Systems Lab.
University of Aizu
Aizu-Wakamatsu 965-8580, Japan
`wlm@u-aizu.ac.jp`

ABSTRACT

In an effort to reduce problems stemming from individual differences in spatial hearing, a rapid method for customizing an interactive spatial auditory display for individual users was developed and tested. This paper describes how new users of a DSP-based spatial auditory display system perform a short series of psychophysical calibration tasks via realtime manipulation of the elevation of a virtual sound source removed from the median plane by a constant angle (on a "cone of confusion" centered on the interaural axis). The user first produces five settings indicating the point at which the perceived elevation of a virtual source matches their own internal standard for "ear-level" incidence. The median of these settings provides an anchoring stimulus for creating an individualized psychophysical scale for controlling source elevation as perceived by the user of the display system. The experimentally-derived anchoring stimulus is regarded as the origin for an angular bisection session that enables the rapid construction of a look up table (LUT) for the full range of elevations produced by the display for each individual user. In contrast to systems that base source elevation control upon individualized head-related transfer functions (HRTFs), the tested system uses a generic set of HRTFs, and manipulates only the values in the LUT for the elevations produced by each HRTF. The method does not attempt to find for each individual listener a single best frequency scaling for the generic set of HRTFs, but attempts to map the useful range of elevations produced by them. Though such a LUT for perceived elevation can be based upon angular estimates made for virtual sources created using each of many HRTFs, the bisection task presented here requires users to complete only a short listening session in which they adjust the elevation of a comparison stimulus to bisect the angle subtended by a pair of reference stimuli. In contrast to other rapid methods of customization, such as those based upon a user's subjective preferences, the current method is based upon active spatial manipulation of a virtual source. The adjustments are referenced to the user's internal standard for "ear-level" incidence, which is tangibly defined and quite easily explained to new users.

1. INTRODUCTION

Spatial auditory display systems typically rely on measured head-related transfer functions (HRTFs) to place virtual sound sources at desired positions in the space surrounding the system user. Due to differences that exist between HRTFs for different individuals, sound processing employing HRTFs measured for the individual user are often regarded as critical to the success of headphone-based applications requiring high accuracy in the spatial positioning of virtual sound sources; and this is especially true for display systems intended to distinguish front from rear, and up from down [1]. But individual differences in HRTFs are only one potential source of variation in the performance of spatial auditory display systems; there may also be variations due to individual differences in spatial perception and cognition. It is well established that individuals often differ in their abilities to perceive various attributes of sensory stimulation, including their abilities to notice similarities and differences between stimuli:

> It is difficult – perhaps impossible – to find any domain or aspect of human responsiveness and performance in which individual differences are so small as to be negligible. (Carroll [2], p. 1).

So 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 [3]. For applications requiring high accuracy in spatial positioning of virtual sound sources, properly equalized headphone reproduction of individualized HRTFs (i.e., HRTFs measured for the intended user's own ears) is probably the approach that engenders the most confidence in success. However, listeners using their own ears will not necessarily localize sounds accurately [4], and so perfectly accurate reproduction of binaural signals does not guarantee perfectly accurate source localization. It is also the case that most spatial auditory display systems do not provide the option of employing individualized HRTFs, but rather employ a single standard set of HRTFs that has been chosen according to a range of criteria (sometimes vaguely specified). Such systems can benefit from customizing the deployment of a set of generic HRTFs to the individual user via frequency scaling of those HRTFs, as shown in [5]. Such frequency scaling of non-individualized HRTFs has been shown to improve virtual source localization accuracy, most likely by minimizing the mismatch between spectral features existing in the listener's own HRTFs and the spectral features existing in the generic set [6].

There has long been a debate about whether listeners localize best using their own HRTFs or using some other subject's HRTFs that exhibit more systematic spectral variation as a sound source moves through space (see Butler, et al. [7] and Morimoto, et al. [8] for the beginnings of this debate, continuing now for more than 30 years). The issues underlying this debate can be clarified somewhat by recognizing two component arguments that potentially may be confused. First, there is the issue of individualization, which is most often associated with properly equalized headphone reproduction of the individual's own HRTFs; second, their is the issue of customization, which has come to be associated with the adjustment for individual use of any set of transfer functions, whether they are measured HRTFs [6], or synthesized HRTFs, constructed from either analysis and resynthesis of measured HRTFs [9] or from some structural model [10]. The methods presented in this paper address the latter issue by implementing a psychophysical calibration scheme designed for rapid completion by new users of any spatial auditory display system.

In the customization of HRTF datasets for individual listeners [6], three types of calibration have been identified:

1. Anthropometric calibration (based on human anatomy)
2. Acoustical calibration (based on binaural signals)
3. Psychophysical calibration (based on binaural perception)

A psychophysical calibration scheme presented by Middlebrooks, et al. [5] was based upon the assumption that there exists for each individual listener a single best frequency scaling for all HRTF measurements collected from another single subject (possibly a dummy-head system such as the historically popular source of HRTF data, KEMAR [11], the **K**nowles **E**lectronics **M**anikin for **A**coustic **R**esearch). This assumption has its foundation in overall summary statistics calculated for many source locations, but the assumption may not lead to the best HRTF scaling for specific virtual source locations. An alternative psychophysical calibration scheme is presented here that was designed to find the best HRTF for a number of key virtual source locations. It will be shown that the choices listeners make are inconsistent with the assumption of a single best frequency scaling for each individual listener.

2. METHODS

2.1. Stimulus Preparation Using Measured HRTFs

In this preliminary study of calibration of virtual source elevation, a set of measured HRTFs used in popular sound spatialization software was employed.[1] The 72 HRTFs employed in the current study were collected at 5° steps in interaural-polar (IP) elevation, but at a fixed IP azimuth angle of 50° (see Fig. 1). These HRTFs have been used in a number of other psychophysical calibration studies [13] [14], and their adequacy in creating useful variation in spatial auditory imagery has been generally confirmed. The magnitude responses of the ipsilateral HRTFs from this dataset are shown in Fig. 2. Each vertical strip in the image corresponds to an HRTF at the angle specified on the x-axis.

One obvious visual feature in the set of curves in Fig. 2 is the relatively deep notch that appears to migrate in frequency with changes in source elevation (identified as antiresonance f_v in [15]). Typically, this primary notch in HRTF magnitude migrates from lower frequencies for sources located below ear level toward higher

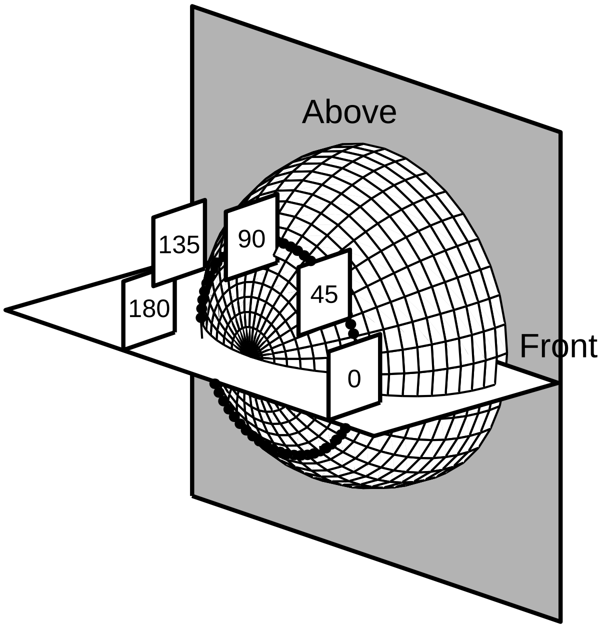

Figure 1: Illustration of the interaural-polar (IP) spherical coordinate system used in this paper to specify sound source directions on sagittal planes (i.e., on circles centered on the interaural axis, intersecting "cones of confusion"). This system was probably first introduced by Morimoto, et al. [12], and though not as often used as the vertical-polar (VP) spherical coordinate system, has certain advantages both in explaining errors in directional judgment [12], and in describing sets of HRTF measurements (as in [10]). In this figure, the dark circle centered on the interaural axis intersects a "cone of confusion" defined at a constant IP azimuth angle, describing all sources displaced 50° from the median plane. The IP elevation angles in the upper hemifield are labeled at 45° intervals, and vary around this circle starting at 0° for a source arriving from a frontal location at ear level. The angle increases to 90° for a source arriving from the highest point on the circle, and reaches 180° for ear-level incidence from the rear. The spherical coordinate system is viewed here from a 45° IP azimuth angle and a 20° IP elevation angle. More explanation of IP coordinates may help the unfamiliar reader at this point: Besides the usual radial coordinate corresponding to source range, this coordinate system includes an azimuth and an elevation angle, but contrary to the more conventional VP spherical coordinate system, the IP coordinates define individual saggital planes parallel to the median plane using the IP azimuth angle. All measurements used in the current study were made on a sphere of radius 1.5 m, and at an IP azimuth angle of 50°. Holding these two coordinates constant while varying a third angular coordinate defines the circle shown in the graphic, lying in a saggital plane that is displaced 50° relative to the median plane, which is shown in the figure as the dark vertical sheet that contains the points directly above and in front of the subject. Source locations within the upper half plane of the circle are labeled by their IP elevation angles, from 0° to 180°; and locations on the horizontal plane (shown in the figure as the lighter sheet) are specified either by a 0° IP elevation angle for the anterior portion, or by a 180° IP elevation angle for the posterior portion.

---

[1] The HRTF dataset employed was a high spatial resolution superset of the author's measurements made for one of the five human subjects chosen by Intel Corp. for use in their 3D-RSX software for sound spatialization. These data have been made available for non-commercial use via the following website: `http://www.u-aizu.ac.jp/~wlm/data/intel/`. The data comprise blocked-meatus, probe-tube measurements that did not use minimum-phase reconstruction of impulse responses, but rather left phase response intact for those spatial regions where the HRTFs exhibited non-minimum-phase behavior.

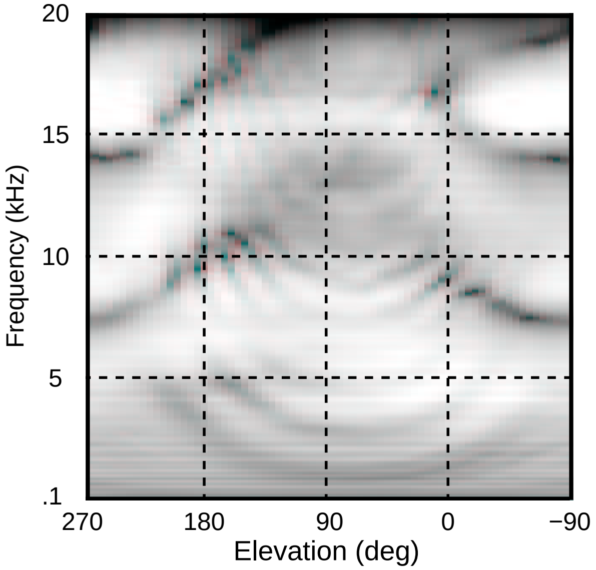

Figure 2: Image constructed from the magnitude response of 72 HRTFs, all measured at constant lateral angle, and varying in elevation in 5° steps. The elevation angle labels the vertical grid lines for cardinal directions (below is -90°, front is 0° above is 90°, rear is 180°, and 270° wraps the plot back to the initial -90° angle). A conventional gray-scale colormap was employed with black coding minimum HRTF gain

frequencies for sources arriving from above. The HRTFs employed in this study showed notches at roughly the same frequencies as those typically observed [15]. For frontal incidence, the first primary notch occurs at around 8 kHz, while for rearward incidence, the notch occurs at around 10 kHz. There is a great deal of evidence that migration of notches in frequency is associated with variation in perceived source elevation (see, for example [16] [7] [17] [18]), yet it is also expected that listeners will base elevation judgments upon spectral cues associated with variation in their own HRTFs [19]. Furthermore, for the lateralized incidence angles of the virtual sources presented in this study it has been established that spectral cues to source elevation exist at frequencies lower than 5 kHz [20]. Despite the great deal of research that has been done on the basis of elevation perception in the acoustical signals presented to the ears, the existence of significant individual differences have made it difficult to develop a robust prediction model for the performance of any given spatial auditory display system with regard to elevation perception. Suffice it to say that individually-oriented calibration is needed, and so for the purposes of evaluating the calibration methods proposed here, a standard approach to directional manipulation was adopted based upon measured human HRTFs.

Preparation of the HRTF data for convolution with the experimental sound source entailed some special treatment. The interaural time difference (ITD) present in the original measurements was removed from each ipsilateral and contralateral pair of HRTFs, and then a constant ITD was re-introduced so that the degree of lateralization would be relatively constant for the sequence. In this way, virtual source elevation should be modulated only by the spectral variation between the 72 pairs of HRTFs. In a preliminary exposure to the nature of the spatial auditory imagery created by the system under test, headphone listeners were presented with two sets of stimuli synthesized using these HRTFs. Always the well-lateralized sources progressed around a cone of confusion from lower-elevation HRTFs to higher-elevation HRTFs (always at a 50° IP azimuth angle). Whether the sources were intended to move through the frontward or the rearward hemifield, listeners reported a relatively smooth increase in perceived elevation of the virtual sound source. Furthermore, all listeners reported that the first stimulus in the presented sequence seemed to arrive from a location below ear level, and the last stimulus in the sequence reached a location well above ear level. Note, however, that despite the use of HRTFs measured for frontal incidence angles, several listeners reported difficulty perceiving a substantial frontward shift for these stimuli when compared to the those processed using HRTFs from the rearward hemifield.

2.2. Bisection Scaling

Reports of bisection scaling may be found as early as 1872, in Plateau's [21] search for the lightness of a gray stimulus midway between black and white. If bisections are made for more than one pair of end points, then the result may be termed an equisection scaling, such as the scaling using eight equal steps that was constructed for the lightness of grays by Munsell, et al. [22]. As opposed to psychophysical scaling methods based upon magnitude estimation for a given set of stimuli varying along some perceptual dimension, the bisection and equisection scaling methods are examples of magnitude production, in which the subject is given control over an adjustable stimulus, and must produce specified differences in perception. The listener's task in this study was to produce, via adjustment of a variable-elevation comparison stimulus, an elevation percept that was midway between two reference stimuli. In effect, the listener was to bisect the angle subtended by the reference stimuli, which were intended to differ primarily in terms of their elevation (though some variation in perceived virtual source azimuth angle might be experienced).

Before bisecting elevation angles using two fixed standards and a variable comparison, all listeners first bisected the 180° angle subtended by two imaginary points, one directly over head, the other directly under the listener's feet.[2] For this first listening session, the HRTF-processed sound source was a staccato trombone tone, with attack and decay amplitude slightly shaped to make the duration 100 ms. This short sound stimulus was convolved with the ipsilateral and contralateral pairs of HRTFs measured at IP 72 elevation angles and at a 50° IP azimuth angle to create a sequence of virtual sources that were heard to vary progressively in elevation angle (confirmed unanimously by all listeners upon initial exposure).

For subsequent bisection scaling sessions, three short trumpet tones of differing pitch were processed (the musical notes being G, C, and E). The sound source that was processed by an HRTF pair measured at the lowest elevation of the three was always presented at the lowest pitch (G); the highest-elevation sound source was always presented at the highest pitch (E). These two extreme-elevation stimuli were presented as fixed-elevation standard stim-

---

[2] In an initial evaluation, the elevation adjustment sessions were followed by a more intensive psychophysical investigation of elevation using the method of constant stimuli. The goal here was to confirm that the rapid adjustment sessions produced reliable and valid estimates of psychometric functions near the "ear-level" response, the target of the initial adjustment.

Figure 3: The graphical user interface (GUI) that was used for the listener's interactive control of stimulus elevation in the adjustment task. Though this image of the window was grabbed from the SGI workstation on which the task was performed, it appears only slightly different when the Matlab-based GUI is displayed on a PC. Note that the title bar reminds the listener of the meaning of the button labels "Lower" and "Higher," and so is specific to the task requiring adjustment to "ear-level" incidence. For the elevation matching task, the title bar displayed the fragment "to match first, make second..."

uli, and a third stimulus of variable elevation was presented in the temporal interval between the two standard stimuli, and at a pitch (C) in between the pitches of the standard stimuli. As higher-pitched sources suggest incidence at higher elevation angles, it was hoped that the pitch variation would create a bias that would aid inexperienced listeners in quickly understanding the bisection task. Of course, only the elevation of the comparison stimulus could be adjusted by the listener, and not its pitch, so the function of the bias was primarily to remind the listener that virtual source elevation angle was intended to be increasing from standard, to comparison, to standard. The experimental sessions included several varieties of the general bisection method, but a few additional procedures were completed as well during the development of the calibration tasks proposed here. The specific details of all the procedures employed in the current study are not described in this methods section, but rather are presented along with results in section 3.

2.3. Subjects

In addition to the subject whose measured HRTFs were used in stimulus preparation, 25 students (both graduate and undergraduate) at the University of Aizu served as listeners in the experiments. All reported normal hearing, but were not audiologically screened. The students had no previous training as subjects in experiments on human spatial hearing, but a few had participated in one previous perceptual study of musical timbre.

3. CALIBRATION TASKS AND RESULTS

This section reports on the results of preliminary user testing with only 25 listeners, but also explains the tasks that the listeners were required to complete in the various experimental sessions. Additional tasks were completed by one subject to provide a deeper check of the reliability and validity of the bisection scaling methods employed in the calibration tasks presented here. An additional detail that allowed for a validity check using this one subject was that this subject was listening through his "own ears" (i.e., this was the subject whose measured HRTFs were used to vary the apparent elevation of the stimuli). It is most informative to begin with an examination of the performance of this special-case listener.

3.1. Adjustment to "Ear-Level" for One Listener

The first task that was required of listeners in order to begin to do elevation calibration for the spatial auditory display system under examination was the adjustment of stimulus elevation to each listener's internal standard for "ear-level" incidence. Listeners heard a warning announcement that the task was about to begin, and then the first stimulus was presented at an elevation randomly selected within a frontward or rearward hemifield from one of 37 possible nominal elevation angles ranging from the lowest to the highest measured HRTF elevation angles. The task was to respond whether the elevation of the virtual sound source needed to be higher or lower in order to move it toward "ear-level." The response was indicated via interaction with the graphical user interface (GUI) depicted in Fig. 3. If a listener judged the source too low, then that choice was indicated by clicking the appropriate button (i.e., the completed statement made on the GUI was "*To reach ear level, make source* ...`Higher`"). A single staircase tracked the 50% point for judging the stimulus higher than "ear-level." If a listener was unsure how to respond, the stimulus could be heard again by clicking on the "Replay" button. When listeners were satisfied that the stimulus had reached "ear-level," the next staircase could be initiated by clicking on the "Next" button. After listeners completed five such staircase runs, an ending message was presented and the GUI was removed. The median of the ending points of the five runs was taken as the estimate of the stimulus closest to "ear-level."

Whereas this adjustment task was completed by all listeners in the manner just described, the "own-ear" listener also participated in a much longer experimental session in which the method of constant stimuli was employed, rather than adjustment of the stimulus to the constant "ear-level" response. The primary difference between these two methods is that in the constant stimuli session the listener has no control over which stimulus will be presented next. Otherwise, the two tasks are practically the same, since an identical judgment needed to be made on each trial for each task. Many more constant stimulus trials were completed in order to provide the best estimate of the probabilities of making one response or the other. This alternative method was employed in order to assess the validity of the less time consuming method of adjustment. The psychometric curves resulting from both methods, and for two spatial regions, are shown in Fig. 4.

The upper panel of Fig. 4 shows the percentage of "Higher" responses that listeners made when the virtual sound source arrived at a frontal incidence angle (consistent with the 50° IP azimuth angle of HRTF measurement). The filled symbols (with dashed connecting lines) correspond to the percentage of the adjustment

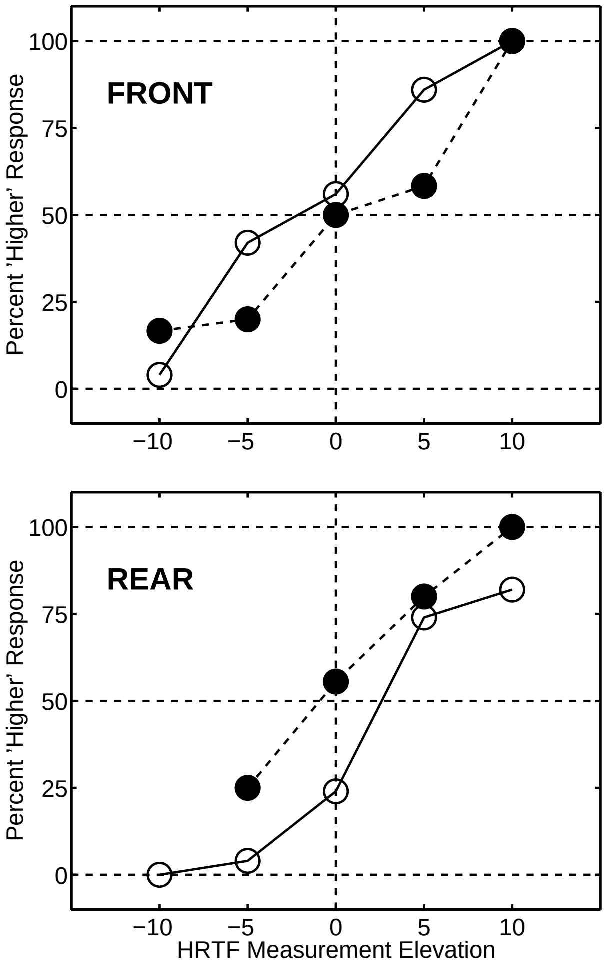

Figure 4: Comparison of the percentage of "Higher" responses obtained using two psychophysical methods, and for virtual sources located in two spatial regions (FRONT vs. REAR). In both upper and lower panels, open symbols (with solid connecting lines) correspond to results obtained using the method of constant stimuli, while filled symbols (with dashed connecting lines) correspond to results obtained when the subject was asked to adjust source elevation to match ear level (i.e., the method of constant response). Note that the missing datapoint for the -10° HRTF in the lower panel corresponds to a constant response category in which the stimulus happened to be presented on less than 5 trials, and was therefore regarded as providing a poor estimate of response probability.

trials on which listeners actively increased source elevation to attempt to bring it to ear level. The open symbols (with solid connecting lines) correspond to the percentage of "Higher" responses given when listeners had no control over source elevation, but were rather presented with a source at an elevation chosen randomly by the computer within a range of 20°. The range of source elevations presented was determined by the previously completed adjustment session for each listener. Although the percentages obtained using the method of constant response (adjustment) were based upon only a few trials (typically around 10 trials per elevation angle), they estimate well the psychometric function based upon the method of constant stimuli with 50 trials per elevation angle. Note that for this listener, the stimulus processed by an HRTF measured at 0° elevation corresponded well to the 50% point on the psychometric function, meaning that the internal standard for "ear-level" matched the externally-defined horizontal plane reference.

The lower panel of Fig. 4 shows the analogous result for virtual sound source arriving from the rear (again, consistent with the rearward angle of HRTF measurement). The match is not quite as good in the rear hemifield, as the results obtained using the method of constant stimuli show a 50% point that is shifted to a slightly higher elevation. This result underscores an important difference between constant stimuli and constant response methods, revealing a potential advantage of the constant response method. The shift of the psychometric function in the constant stimuli case appears to be due to above/below reversals rather than an actual small shift in perceived elevation. These reversals are not as common when the listener is allowed to adjust source elevation actively, and so a natural conclusion here might be that a perceptual "hysteresis" effect is operating to keep elevation unambiguous in the adjustment task, avoiding above/below reversals that could bias the estimate of the 50% point in the constant stimulus case.

3.2. Adjustment to "Ear-Level" for 25 Listeners

Each listener completed the calibration procedure within a 10 minute period. All 25 listeners produced adjustment data that were consistent within subject, though the stimulus associated with "ear-level" response differed between subjects as expected (since no customizing was performed, such as frequency scaling [5]). In fact, quite extreme variation was observed between subjects regarding which HRTF-processed stimulus was selected as the best choice for an "ear-level" response. Fig. 5 reveals that the measurement angles of selected HRTFs ranged from $-60°$ to $+30°$ elevation. Although the most common response for both frontward and rearward incidence was that corresponding to the 0° measurement angle, three out of 25 listeners chose the stimulus processed by the frontal HRTF that was measured at $-60°$ IP elevation. Bearing in mind that these directions are lateralized from the median plane at 50° IP azimuth, and therefore these IP elevation angles are not as extreme as their corresponding VP elevation angles[3], there still is perhaps a surprising amount of variation. The variation greatly exceeds the between subject variation in frequency scaling reported by Middlebrooks, et al. [5]. A reasonable conclusion might be that there are significant individual differences in where listeners imagine their own "ear level" must be located when projected to a location outside the head. Nonetheless, there

---

[3] The 90° maximum IP elevation angle at a 50° IP azimuth corresponds to a VP elevation angle of only 50° .

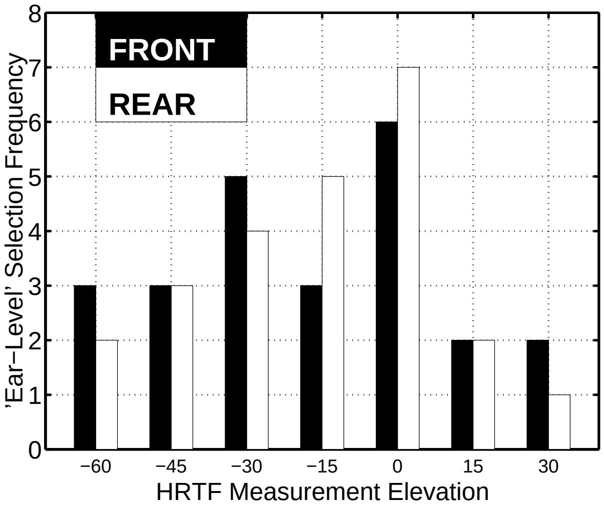

Figure 5: Histogram of “ear-level” selection frequencies for 25 listeners plotted for groups of HRTFs spaced at 15° in elevation measurement angles, comparing results for frontward (black bars) vs. rearward (white bars) spatial regions.

is certainly a concern about reliability of these judgments and validity of these estimates of “ear-level” when listeners employ an internally-defined standard of reference. Therefore, an additional check was made on the results for five of the 25 listeners using an alternative task. The alternative task used an externally-defined standard, rather than asking the listener to imagine an internal reference level while listening to a single stimulus.

3.3. Matching Frontward to Rearward “Ear-Level” Stimuli

Each of five listeners completed an elevation matching task within a 3 minute period. On each trial two stimuli were presented: First, a fixed-elevation standard stimulus was presented at the rearward direction that each listener indicated in previous sessions was closest to “ear-level.” The second stimulus was presented at a selected elevation from a frontward direction, and the listener had to indicate whether it seemed to be higher or lower in elevation than the rearward standard stimulus. Just as in the previous single-stimulus adjustment task, five such runs were completed. The median ending points of the five runs was taken as the estimate of the stimulus that matched best the elevation of the standard stimulus.

Results confirmed the validity of the previous estimates, in that all five listeners produced nearly the same median when matching frontward to rearward stimuli as when judging frontward elevation in isolation. It should be noted as well that matching to an external standard was reported to be easier by all listeners than matching to an internal standard had been. Furthermore, the runs took less time and required fewer trials to complete (a single “ear-level” stimulus adjustment run required an average of around 19 trials, while the matching task required an average of only around 17 trials). It was concluded that the best approach for the elevation calibration procedure proposed here would be to always first complete a single-stimulus adjustment task (rearward incidence was preferred for this initial task), and then follow that with an elevation matching task (with the rearward virtual source as the standard stimulus).

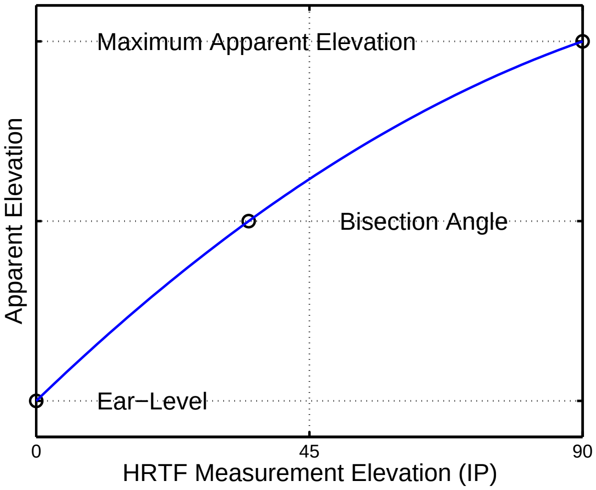

Figure 6: Look up function constructed from the bisection scaling results for one listener. The three circular symbols mark the angles at which HRTFs were measured that produced three apparent elevations which are plotted here as a function of HRTF measurement elevation (specified in interaural-polar coordinates). The solid line plots a curve from values of the calculated LUT for elevation.

3.4. Bisection Scaling of Elevation in the Upper Hemifield

Each of five listeners completed a bisection scaling task in which their own selected rearward stimulus provided the lower of two standard stimuli (i.e., this standard was presented at the rearward direction that each listener indicated in previous sessions was closest to “ear-level”). The second standard stimulus was presented at an IP elevation of 90° and an IP azimuth of 50° (the topmost point on the cone of confusion at this degree of lateralization). A variable comparison stimulus was presented during the temporal interval between the two standard stimuli, and at an HRTF measurement elevation that was between that of the two standard stimuli. The listener always heard the lower-elevation standard first. The task was to adjust the elevation of the comparison stimulus until it seemed to lie midway between the elevations of the two reference stimuli.

Fig. 6 shows bisection scaling results for one listener, and also displays a smooth, curvilinear function for finding the measurement angle of the appropriate HRTF for producing a desired apparent elevation. The stimulus that was judged closest to “ear-level” for the listener was that produced by the 0° HRTF measurement elevation, and the HRTF that was selected as that producing an apparent elevation midway between this lower reference point and maximum apparent elevation was the HRTF measured at 35° IP elevation. These preliminary results suggest that the median setting obtained listeners can be used to create an effective look up table (LUT) for the full range of upper-hemifield elevations produced by the display for each individual user, as discussed below.

3.5. Constructing a LUT for Virtual Source Elevation

Though only 25 listeners have participated in this preliminary study, and only 5 listeners have completed all the tasks that might be regarded as required for proper calibration of displayed elevation, a suggested application of the results can be offered with some confidence. The application is to base the means of HRTF deployment on data that can be collected in roughly ten minutes from listeners who receive minimal instruction.

From only two "ear-level" adjustments, and a single bisection adjustment, a LUT for HRTF selection and interpolation could be constructed. The solid line shows a 2^{nd} order polynomial fit to the three points, providing a smooth look up function for controlling HRTF selection and interpolation. What should be noticed here is the relative compression of the apparent elevation function at higher elevations. This compression in the function is due to the fact that only 7 of the 19 upper-hemifield HRTFs are used below the elevation bisection angle, while 12 are used above that angle. This implementation is inconsistent with the assumption of a single best frequency scaling for an individual listener, since that would be represented as a simple change in slope of the look up function. What is captured using bisection is only slightly more complex, but very useful, since it effectively splits into two equal portions the perceived angular subtense spanning from "ear-level" incidence to the maximum elevation observed for that individual listener.

It might be asked whether such an approach is the natural approach, since it equalizes the spatial subtense of two regions that might normally and naturally be perceived as unequal. The argument might be made that equalization of angular subtense would degrade elevation discrimination performance where it is most needed, for virtual sources localized just above "ear-level," since fewer HRTFs would be used below the elevation bisection angle. If an application calls for very few sources to be localized above the elevation bisection angle, then maintaining discriminability for the highest elevation angles is not useful. Conversely applications expected to use the full range of variation in IP elevation within the upper hemifield will benefit. Suffice it to say that sensitivity to application requirements must always be maintained.

4. CONCLUSION

In contrast to a time-consuming comprehensive angular estimation session, requiring multiple angular estimates for stimuli created using each of many HRTFs, the bisection task presented here requires only an indication of which comparison stimulus produces an elevation midway between the elevations of a pair of reference stimuli (either actual or imaginal). The question addressed here was not whether a listener can accurately report actual source incidence angles, but rather focussed upon which stimulus presented via earphones was associated with an experience of an externalized spatial auditory image located at a particular angle, such as the 0° elevation angle termed "ear level." Preliminary results show that listeners are able to complete the calibration task quickly and with only a few instructions. Also, the relative compression of the apparent elevation function at high elevations compared to low elevations is inconsistent with the assumption of a single best frequency scaling for each individual listener.

5. REFERENCES

[1] B. G. Shinn-Cunningham, "Spatial auditory displays," in *International Encyclopedia of Ergonomics and Human Factors*, W. Karwowski, Ed. Taylor and Francis, Ltd., Atlanta, GA, (in press).

[2] J. B. Carroll, "1: Studying individual differences in cognitive abilities: Though and beyond factor analysis," in *Individual differences in cognition*, R. F. Dillon and R. R. Schmeck, Eds., vol. 1, pp. 1–33. Academic Press, Inc., 1983.

[3] W. L. Martens and A. Yoshida, "Augmenting spoken telecommunication via spatial audio transformation," *The Journal of Three Dimensional Images*, vol. 14, no. 4, pp. 169–175, 2000.

[4] S. R. Oldfield and S. P. A. Parker, "Acuity of sound localisation: a topography of auditory space I. Normal hearing conditions," *Perception*, vol. 13, pp. 601–617, 1984.

[5] J. C. Middlebrooks, E. A. Macpherson, and Z. A. Onsan, "Psychophysical customization of directional transfer functions for virtual sound localization," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 108, pp. 3088–3091, 2000.

[6] J. C. Middlebrooks, "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 106, pp. 1493–1510, 1999.

[7] R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 61, pp. 1264–1269, 1977.

[8] M. Morimoto and Y. Ando, "On the simulation of sound localization," **J**. **A***cous.* **S***oc.* **J***apan (E)*, vol. 1, pp. 167–174, 1980.

[9] W. L. Martens, "Principal components analysis and resynthesis of spectral cues to perceived direction," in *Proc. Int. Computer Music Conf.*, Champaine-Urbana, IL, Sept. 1987.

[10] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 109, no. 3, pp. 1110–1122, 1999.

[11] M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 58, pp. 214–222, 1975.

[12] M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere," **J**. **A***cous.* **S***oc.* **J***apan (E)*, vol. 5, pp. 236–240, 1984.

[13] W. L. Martens, "Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display," in *Proc. Int. Conf. on Auditory Display*, Espoo, Finland, 2001, ICAD, pp. 197–207.

[14] W. L. Martens and N. Zacharov, "Multidimensional perceptual unfolding of spatially processed speech I: Deriving stimulus space using INDSCAL," in *Proc. 109th Conv. of the Audio Engineering Society*, Los Angeles, Sept. 2000, Preprint 5224.

[15] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point-source," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 44, pp. 240–249, 1968.

[16] P. J. Bloom, "Creating source elevation illusions by spectral manipulation," **J**. **A***udio* **E***ng.* **S***oc.*, vol. 25, pp. 560–565, 1977.

[17] J. H. Hebrank and D. Wright, "Spectral cues utilized in the localization of sound on the median plane," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 56, pp. 1829–1834, 1975.

[18] A. J. Watkins, "Psychoacoustic aspects of synthesized vertical locale cues," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 63, pp. 1152–1165, 1978.

[19] J. C. Middlebrooks, "Narrowband sound localization related to acoustical cues," **J**. **A***cous.* **S***oc.* **A***mer.*, vol. 92, pp. 2607–2624, 1992.

[20] W. L. Martens, *Directional hearing on the frontal plane: Necessary and sufficient spectral cues*, Ph.D. thesis, Northwestern University, Evanston, Illinois, 1991.

[21] J. A. F. Plateau, "Sur la mesure des sensations physiques, et sur la loi qui lie l'intensie' de ces sensations a' l'intensie' de la cause excitante [On the measure of physical sensations, and on the law binding the intensity of these sensations with the intensity of their causal excitations]," *Bulletins de l' Academie Royale des Sciences, des Lettres, et des Beaux-Arts de Belgique*, vol. 33, pp. 376–388, 1872.

[22] A. E. Munsell, L. L. Sloan, and I. H. Godlove, "Neutral value scales: I. Munsell neutral value scale," **J**. **O***ptic.* **S***oc.* **A***mer.*, vol. 23, pp. 394–411, 1933.