Comparison of headphones and equalization for virtual auditory source localization

David Schonstein[a], Laurent Ferré[b] and Brian Katz[b]

[a]Arkamys, 5 rue Frédéric Bastiat, 75008 Paris, France
[b]LIMSI-CNRS, B.P. 133, 91403 Orsay, France
dschonstein@arkamys.com

This study investigates the variation in localization performance between different headphone types. Eight different headphones (including various in-ear, circumaural open and closed, and bone conduction headphones) were tested. In addition, the effect of headphone equalization (aiming to produce an approximately flat frequency response) was investigated. Localization was examined for 24 locations distributed on a sphere surrounding the listener. A single subject participated in the study using a single chosen non-individualized HRTF. Each location was repeated 6 times, resulting in a total of 144 localization reports. Overall, localization was relatively accurate for 3 out of the 8 headphones tested. For these 3 headphones, there was no significant difference in lateral angle error, whilst polar angle errors, associated with the cone of confusion, did vary significantly. The headphone equalization had varying effects on localization accuracy depending on the headphone. Globally, headphone equalization showed no significant effect on localization accuracy. These results serve as a preliminary investigation, demonstrating accurate localization for only a select group of headphones, tested for effective sound rendering in virtual auditory space. In addition, the results suggest that headphone equalization has a minimal influence on localization accuracy under these conditions.

1   Introduction

Humans are able to seamlessly localize sound sources in most auditory environments. It is understood that the auditory system is able to do this with the use of three auditory cues: the interaural time difference (ITD), the interaural level difference (ILD) and spectral information. The ITD, a difference in the time of arrival of a sound at each ear, and ILD, a difference in the level of a sound at each ear, provide information about the so-called "cone of confusion" [1] on which a sound source lies. Spectral information, the third localization cue, is used to resolve where on the cone of confusion a sound source is positioned. This spectral information is produced as a result of the free-field to eardrum transfer function, generated by the pinna, head, and torso, and acts as a directional acoustic filter [2-5]. These acoustic filters in effect produce location-dependent spectral patterns, known as head-related transfer functions (HRTFs). A virtual auditory space (VAS) can be created for listeners using HRTF measurements. This is achieved by digitally filtering audio content using left and right ear HRTF recordings for desired positions in space. The resulting binaural stimulus is then played to the listener over headphones and is perceived as in a realistic three-dimensional auditory environment, outside of the listener's head.

The effectiveness of a particular VAS rendering is often measured as a function of a listener's ability to accurately localize sound sources using the binaural stimulus. Studies have shown that when using HRTFs that were specifically measured using the listener's own ears (individualized HRTFs) binaural directional cues are correctly simulated and accurate localization occurs [6, 7]. Non-individualized HRTFs (*i.e.* HRTFs not recorded using a listener's own ears, but rather that of another person [8] or manikin [9]) have often been used for generating VAS. By using non-individualized HRTFs the somewhat tedious and expensive recording process is bypassed, however less accurate localization performance occurs [8, 10, 11] and thus a less realistic VAS is generated for the listener.

Another factor relating to the perception of a realistic VAS is the choice of headphone used for the binaural rendering. The two main issues related to headphone choice are: 1) some headphones have a poor frequency response (*i.e.* regions in the frequency domain where the headphone is unable to produce a signal) and/or a frequency response that is not flat, 2) when using circumaural headphones (*i.e.*

headphones that are not placed directly into the ear canal) the generated binaural signal will be filtered by the pinna (in a directionally independent manner). In both cases the spectral information, used for accurate localization in VAS, transmitted in the HRTF will be somewhat distorted.

With respect to a headphone's frequency response, some studies have looked at the effectiveness of different headphones as a means for the reproduction of binaural signals [12-14] highlighting significant differences in the style of headphone used. Bone conduction headphones have also been assessed for use in VAS systems [15]. For pinna distortions of binaural signals when using circumaural headphones, there is some general consensus in the literature that the effect is significant and that these distortions differ from listener to listener much like HRTFs vary between listeners [13, 14]. In such cases a headphone equalization using an individualized headphone transfer function (HpTF) is proposed [11].

Whilst these mentioned studies provide invaluable insights into the role headphone choice and HRTF selection in VAS rendering, they all draw their conclusions using individualized HRTFs or have only used one headphone type. It is the purpose of this study to assess the ability of a wide range of headphone types to render sounds in VAS using non-individualized HRTFs. In addition, the effectiveness of non-individualized headphone equalization is assessed.

2   Methods

2.1   VAS stimuli

The audio stimulus used in the localization experiments was Guassian noise of duration 130-ms and was windowed by applying a 5-ms onset and offset Hanning ramp. HRTFs used in this experiment were measured in an anechoic chamber, using a "blocked-ear" recording technique. This technique involves embedding a small recording microphone in an earplug secured flush with the distal end of the ear canal [12]. For a detailed explanation of the procedure used to generate the HRTFs used in this study see [16].

This study used a non-individualized HRTF to render the noise stimulus in VAS. In order to select an HRTF that best rendered the noise stimulus in VAS, the subject was put through an HRTF selection test prior to the localization

experiments. The HRTF measurements (for 44 different subjects) used for the selection test were taken from an LISTEN project online database [16]. The subject listened to a broadband noise stimulus over headphones convolved with each of the 44 HRTFs. The subject judged how well the noise stimulus was rendered in VAS for a circular trajectory. Based on the subject's judgements, one particular HRTF was then selected to be used for the localization experiments. The noise stimulus was then rendered in VAS by convolution of the noise stimulus in the time domain with head-related impulse responses (HRIRs). The HRIRs used were the raw recordings, and had not been equalized or diffuse field compensated. There was an ITD adaptation performed on the noise stimulus after it had been convolved with the required HRTFs. This adaptation was used to accommodate for the non-individualized HRTF that introduces incorrect ITDs. The circumference of the subject's head was used as a parameter to generate the correct ITDs [17].

2.2 Headphones and headphone transfer functions

A total of 8 different headphones were compared in the current experiment. The headphones used were: an open and closed circumaural headphone (OC and CC respectively), an in-ear headphone (IE), two tube insert headphone models (varying significantly in their market value) which blocked the ear canal (TI-1: least expensive, TI-2: ER-2 reference headphones, most expensive), a tube in-ear headphone that left the ear canal open (TO), and two commercially available bone conduction headphones varying slightly in their market value (BC-1: least expensive, BC-2: most expensive).

The headphone transfer function (HpTF) of each headphone (except for the TI-2 reference) was obtained. An inverse filter was created out of the measured data to produce an approximately flat frequency response (non-individualized equalization). This procedure was not performed for the T1-2 as they are designed to have a flat frequency response up to 10 kHz at the human eardrum. Attention should be made with respect to the definition of the HpTF calculated in this experiment as it is different to similar definitions used by authors such as [12, 13].

In this experiment the HpTF was measured by placing the headphone on a metal plate with a flush mounted measurement microphone with approximately 2cm of foam padding between plate and headphone in order to mimic a human head. This varies from methods employed by [13] in which the headphone response was measured *in situ* with a probe-microphone system inside the ear canal for each individual subject, thus taking into account the spectral transformations of the headphone and individual subject's outer ear.

For the two bone conduction headphones, frequency responses were obtained directly from the manufacturer. For the remaining 5 headphones an impulse response was measured for each left and right ear. The impulse response stimulus was a 500-ms sweep (frequencies from 60 to 20000 Hz). An impulse response was recorded and then converted into the frequency domain, in which all proceeding manipulations were made. This process was repeated ten times; at each recording the headphone was

repositioned in order to introduce some variance into the measurements and accommodate for differences in the positioning of headphones on listeners' heads. The mean of the ten frequency-magnitude responses was then taken and used as the HpTF. An inverse filter was then created for each ear of every headphone measured (and the bone conduction headphones), over the mentioned frequency range. This was done by calculating the inverse magnitude of the HpTF and creating a recursive IIR filter using a warping factor and filter order (ranging between 24 and 32) that was manually adjusted to maximise the filter detail over a specific frequency range, depending on the frequency response of each headphone.

In the localization experiment the noise stimulus was convolved with the subject's selected HRTF, and then filtered by a series of biquad filters, calculated from the inverse filter of the HpTF, in order to produce an approximately flat frequency response from each of the specified headphones.

2.3 Testing procedure

One subject, a male aged 25 years old, first author of this paper, took part in the localization experiments. The subject was shown, prior to the localization tests, to have clinically normal hearing. The localization experiments were conducted in a sound dampened room using a hand pointing localization paradigm. All audio stimuli were played to the subject using one of the specified headphones or bone conduction headphones. Throughout the experiments the position of the subject's head and hand were continuously monitored using an electromagnetic tracking system (Flock of Birds) in which two receivers were used. The head receiver was attached to a headband worn by the subject, and the hand receiver, equipped with a tip to serve as a pointer, was held by the subject. The subject stood at a specified position in the room. Initial head orientation for each run was set and verified using visual feedback from a computer screen. Once at this calibrated position, the audio stimulus would play, and the subject would point to where the sound was thought to have originated, and register a response using a MIDI foot pedal. The subject's response was calculated as the position of the pointer relative to the centre of the subject's head at the time the noise stimulus was played.

There were a total of 24 positions used in the localization experiments. Using the hoop coordinate system, there were 4 elevations tested: -30, 0, 30, and 60 degrees. There were 5 azimuths tested: -135, -75, -15, 45, 105, and 165 degrees. This set of positions was chosen randomly from a subset of possible azimuths and elevations so that the subject did not know the exact positions of the target locations. All target locations where of equal distance from the listener describing positions on an imaginary sphere in VAS (distance equal to that of the measured HRTF data). Noise stimuli presentation level for each headphone was calibrated to approximately 50 dB sensation level (*i.e.* determined by adding 50 dB to the audible threshold of the noise stimulus at position 0 degrees azimuth and elevation).

For each localization test, in which one particular headphone was tested, the subject listened to a total of 72 noise stimuli (i.e. three repeats of each of the 24 positions tested). The position of the noise stimulus was randomly

generated. There were a total of 8 headphones. Each headphone (except for the TI-2 reference headphone) was tested with and without the inverse filter created as described above. Thus there were a total of 15 headphone conditions tested. Each localization test of 72 positions was selected randomly and repeated twice (144 observations for each headphone). This produced a total of 6 repeats of each of the 24 positions tested for each of the 15 headphone conditions. The tests were performed over two days with a pause between every test.
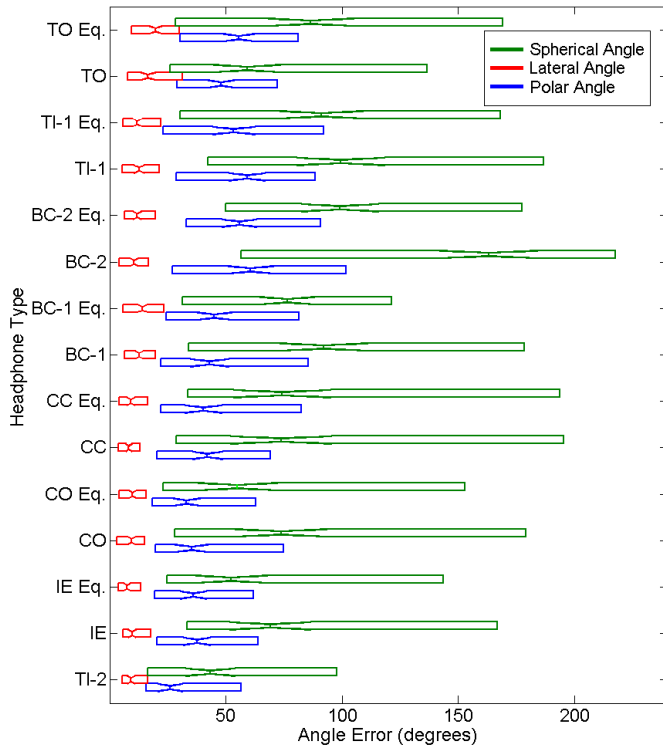

Fig 1 Boxplot of spherical, lateral, and polar angle error for all headphone conditions.

3   Results

In this experiment two coordinate systems were used to describe the positions of the target noise stimulus rendered in VAS. The first is what is known as 'hoop' coordinate system [18]; azimuth and elevation coordinates correspond to the standard single pole coordinate system where 0,0 is directly ahead and positions to the right and up are positive and to the left and down are negative. The second is known as a 'lateral/polar' coordinate system [19]; there is a single pole passing through the two ears, lateral angle being the horizontal angle away from the midline and polar angle describing the angle around the circle described by a particular lateral angle . Thus in the paper positions will be described in azimuth/elevation or lateral/polar angle.

Figure 1 is a boxplot display of the spherical angle error, lateral angle error and polar angle error for each headphone before any front-back confusions processing. The position of the notch represents the median, and lower quartile upper quartile values are represented by the extremities of each bar.

In order to gauge the overall performance of the subject's performance in each of the 15 headphone conditions, the spherical correlation coefficient [20] (SCC) was calculated. The SCC describes the degree of correlation between the

centroids (the mean directional cosine for a number of positions) of target and response locations (for a detailed description see [18]). The SCC ranges from +1 for complete positive correlation (where the set of centroid responses can be transformed to the set of centroid targets by a rotation about the vertical axis) to -1 for a complete negative correlation (where the set of centroid responses can be transformed to the set of centroid targets by a reflection about the horizontal axis) [21].

In order to further investigate the pattern of localization errors, the data were analysed in terms of lateral/polar angle. The lateral/polar coordinate system is most appropriate for the error analysis since it mirrors the physiology of how humans locate sounds in space. In general there are two types of localization errors. The first type is referred to as a local error in which the subject makes a judgement of the perceived location of the noise stimulus within approximately 20° of the actual target location [21]. The second type is referred to as front-back confusion error, in which the subject will correctly locate the noise stimulus in elevation (*i.e.* with respect to the median plane) but confuse the hemisphere of the target location (*i.e.* make a large error in azimuth about the interaural axis). There is a large qualitative difference between these two types of errors, and for this reason front-back errors were removed before further error analysis. In identification of front-back errors to be removed, target locations 10 degrees around the interaural axis were ignored, and responses were allowed to cross the midline by 10 degrees (see [18] for a detailed explanation and source code used). Mean absolute lateral and polar angle errors were calculated from the data for each headphone condition. In addition, mean spherical angle error (shortest distance in degrees, along an imaginary sphere on which sound sources are located, between target and response position) was calculated. The proportion of front-back errors that were removed was also calculated.

Headphone	SCC	% Front-Back Error	Mean Spherical Angle Error	Mean Lateral Angle Error	Mean Polar Angle Error
TI-2	0.77	37.50	23.14	10.58	24.38
IE	0.79	44.45	30.15	11.57	35.88
IE Eq.	0.69	40.28	30.92	8.79	37.02
CO	0.67	43.75	29.63	8.94	34.08
CO Eq.	0.75	38.20	27.98	10.18	30.34
CC	0.30	44.45	32.67	9.17	37.70
CC Eq.	0.20	41.67	35.87	10.43	43.58
BC-1	0.38	47.92	32.76	11.21	39.87
BC-1 Eq.	0.30	40.98	34.37	15.38	43.99
BC-2	-0.45	52.09	47.62	10.20	60.73
BC-2 Eq.	-0.26	41.67	45.38	12.55	61.47
TI-1	-0.07	39.59	42.87	13.77	53.81
TI-1 Eq.	0.13	40.98	40.24	16.47	50.28
TO	0.55	38.20	37.23	20.27	40.49
TO Eq.	0.52	50.00	36.90	19.27	41.34

Table 1 Table of SCC, front-back, spherical angle, lateral angle and polar angle errors.

4   Discussion

Table I shows the SCC, proportion of front-back errors, mean spherical angle error, mean lateral angle error and mean polar angle error for each headphone condition. Note that detected front-back error responses have been removed before this error analysis. The SCC was highest for the IE without equalization condition, at 0.79, with the TI-2, CO with equalization conditions, having values of 0.77 and 0.75 respectively. SCC values for a noise stimulus using individualized HRTFs are generally reported to be greater than 0.9 in the literature [7, 21-23]. The IE with equalization, and the CO headphone without equalization conditions, had SCC values of 0.69 and 0.67 respectively. The CC, BC-1, and TI-1 (with and without equalization) conditions had relatively low SCC values demonstrating a relatively poor correlation between target and response centroids. The BC-2 with and without equalization conditions demonstrated low negative SCC values, which represents a slight correlation between reflected target and response centroids. This can probably be attributed to judgements in elevation that are negative when the target elevation is positive. Wenzel *et al.* [8] found, under similar conditions to this study (i.e. non-individualized HRTFs using a non-individualized equalization), SCC values in the range of 0.52 to 0.76 for 16 different subjects using a circumaural closed model. However in this study, front-back confusions were resolved, rather than extracted from the analysis as in the current study, which means they were coded as if the response was in the right hemisphere. This may have introduced data into the error analysis that produced higher SCC as compared to this study. The CC headphone used in this study demonstrated a significantly lower SCC for both the conditions with and without equalization (0.30 and 0.20 respectively). This could be due to the fact that different circumaural headphone models were used, and that only one subject was tested in this study.

For all the headphone conditions a notably high rate of front-back errors was observed; a mean of 43 percent was calculated across all headphone conditions. These results are somewhat similar to the findings of previous studies. Wenzel *et al.* [8] found a mean rate of front-back errors to be 31 percent using non-individualized HRTFs. This value is slightly lower, yet comparable to front-back error rates in this study for similar headphone types, namely the CO and CC. Typical rates of front-back errors using individualised HRTFs are in the range of 5 percent [23, 24]. However, Wightman and Kistler [7] reported a mean front-back error rate of 11 percent for 8 subjects, when using individualized HRTFs.

As can be seen in Table I, mean spherical angle error was the lowest for the TI-2 with 23 degrees, the IE with and without equalization, CO with and without equalization conditions also had relatively low values in the range of 28 to 31 degrees. Other headphone conditions had values in the range of 33 to 48 degrees. A one-way ANOVA demonstrated that difference in spherical angle error was significant across all headphone conditions (p-value = 0.00). Wenzel *et al.* [8] found similar values for inexperienced VAS listeners using non-individualized HRTFs, however their data was characterized by large individual differences between subjects, and included the resolved front-back data. Wightman and Kistler [7]

reported slightly lower mean spherical angle errors for individualized HRTFs and somewhat more experienced listeners. Experienced VAS listeners using individualized HRTFs in another study demonstrated a mean spherical angle error of 15 degrees [23]. A one-way ANOVA performed across the 5 headphone conditions in this study with relatively high SCC values (namely the TI-2, IE with and without equalization, CO with and without equalization conditions) confirm a significant difference in spherical angle error (p-value = 0.01).

Mean lateral angle errors were consistent for 9 out of the 15 headphone conditions, as can be seen in Table I. For these 9 headphone conditions mean lateral angle errors were about 10 degrees, which is consistent with values in the literature for localization tests using individualized HRTFs [22, 23]. For the remaining 6 headphone conditions, mean lateral angle errors were significantly larger. Mean polar angle error values varied across the different headphone conditions as can be seen in Table I. The TI-2 headphone condition demonstrated a markedly lower polar angle error than the rest of the headphones at 24 degrees. The CO with equalization also had a relatively low value at 30 degrees. The IE with and without equalization, CO without equalization, and CC without equalization conditions all had comparable mean polar angle errors in the range of 34 to 38 degrees. The remaining 9 headphone conditions had relatively high mean polar angle errors. Lateral and polar angle errors were significantly different across all headphone conditions (p-value = 0). Within the subset of 5 headphone conditions that had higher SCC values, lateral angle error did not vary significantly (p-value = 0.14), however polar angle error did prove to vary significantly (p-value = 0.02).

The headphone equalization had varying effects on localization accuracy. Globally, the headphone equalization did not have a significant effect on localization performance. A one-way ANOVA testing the significance of the effect of equalization on spherical angle error, lateral angle error, and polar angle error were 0.91, 0.21, and 0.67 respectively. To our knowledge the only psychophysical validation of headphone equalization (despite some general consensus of opinions in favour of its significance) in the literature is an informal test using an individualized HpTF [14] where it was suggested that the equalization was of importance.

5   Conclusion

The aim of this study was to test a wide variety of headphones and their ability to effectively render sound sources in VAS (with and without headphone equalization) using localization tests. It is important to note that a customized (using a personalized selection of best HRTF set from a given database of HRTF sets using an individual ITD estimation correction) HRTF was used in this study. The HRTF was nevertheless a non-individualized HRTF.

Of the 15 headphone conditions tested in this study only 5 conditions demonstrated consistent results, corresponding to two of the in-ear headphones (TI-2 and IE) and the CO. The CC, TI-1, and bone conduction headphone models did not produce accurate localization in VAS under these test conditions. The headphone equalization had little to no effect on localization accuracy in this study. Whilst it has

been argued in the literature that an *individualized* headphone equalization is "required for adequate VAS synthesis" [14], or that the "use of non-individualized HpTFs [headphone equalization] … could partly account for the increased incidence of mislocalizations reported in previous virtual auditory space localization experiments" [13], this study has shown that a *non-individualized* headphone equalization is most probably not effective.

This study has demonstrated some interesting findings based on headphone choice and equalization for VAS applications. However it is essential that it be seen as a preliminary study (with no implication on the quality of any of the tested headphone models) based on the limited number of subjects tested, and subsequently be used as a platform for future studies.

Acknowledgments

References

[1] Mills, A.W., "Auditory Localization", in *Foundations of Modern Auditory Theory*. 1972, Academic Press: New York. p. 303-348.

[2] Blauert, J., "Sound Localization in Median Plane", *Acustica* 22, 205-213 (1969)

[3] Carlile, S., R. Martin, and K. McAnally, "Spectral information in sound localization", *Auditory Spectral Processing* 70, 399-434 (2005)

[4] Carlile, S. and D. Pralong, "The Location-Dependent Nature of Perceptually Salient Features of the Human Head-Related Transfer-Functions", *Journal of the Acoustical Society of America* 95, 3445-3459 (1994)

[5] Mehrgardt, S. and V. Mellert, "Transformation characteristics of the external human ear", *Journal of the Acoustical Society of America* 61, 1567-1576 (1977)

[6] Kulkarni, A. and H.S. Colburn, "Role of spectral detail in sound-source localization", *Nature* 396, 747-749 (1998)

[7] Wightman, F.L. and D.J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation", *Journal of the Acoustical Society of America* 85, 868-878 (1989)

[8] Wenzel, E.M., et al., "Localization Using Nonindividualized Head-Related Transfer-Functions", *Journal of the Acoustical Society of America* 94, 111-123 (1993)

[9] Gardner, W.G. and K.D. Martin, "HRTF measurements of a KEMAR", *Journal of the Acoustical Society of America* 97, 3907-3908 (1995)

[10] Middlebrooks, J.C., "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency", *Journal of the Acoustical Society of America* 106, 1493-1510 (1999)

[11] Moller, H., et al., "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society* 44, 451-469 (1996)

[12] Moller, H., et al., "Transfer Characteristics of Headphones Measured on Human Ears", *Journal of the Audio Engineering Society* 43, 203-217 (1995)

[13] Pralong, D. and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space", *Journal of the Acoustical Society of America* 100, 3785-3793 (1996)

[14] Wightman, F. and D. Kistler, "Measurement and validation of human HRTFs for use in hearing research", *Acta Acustica United with Acustica* 91, 429-439 (2005)

[15] Walker, B.N., et al. "High fidelity modeling and experimental evaluation of binural bone conduction communication devices", in *Proceedings of the 19th International Congress on Acoustics* (2007)

[16] IRCAM. *LISTEN HRTF database*. [cited; Available from: http://recherche.ircam.fr/equipes/salles/listen/.

[17] Katz, B.F.G., G. Vandernoot, and O. Warusfel, "Individual adaptation", Deliverable D4.3 projet Européen LISTEN IST-1999-20646, (2003)

[18] Leong, P. and S. Carlile, "Methods for spherical data analysis and visualization", *Journal of Neuroscience Methods* 80, 191-200 (1998)

[19] Middlebrooks, J.C., "Individual differences in external-ear transfer functions reduced by scaling in frequency", *Journal of the Acoustical Society of America* 106, 1480-1492 (1999)

[20] Fisher, N.I., T. Lewis, and B.J.J. Embleton, "Statistical Analysis of Spherical Data". Cambridge: Cambridge University Press (1987)

[21] Carlile, S., P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners", *Hearing Research* 114, 179-196 (1997)

[22] Best, V., et al., "The role of high frequencies in speech localization", *Journal of the Acoustical Society of America* 118, 353-363 (2005)

[23] Jin, C., et al., "Contrasting monaural and interaural spectral cues for human sound localization", *Journal of the Acoustical Society of America* 115, 3124-3141 (2004)

[24] Langendijk, E.H.A. and A.W. Bronkhorst, "Contribution of spectral cues to human sound localization", *Journal of the Acoustical Society of America* 112, 1583-1596 (2002)