19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

SOUND LOCALIZATION WITH SCALED DUMMY-HEADS ON A *TELEHEAD*

PACS: 43.66.Qp, 43.66.Yw, 43.66.Pn

Tatsuya Hirahara[1]; Hiroyuki Sagara[2]; Makoto Otani[1]
[1]Fuculty of Engineering, Toyama Prefectural University; 5180 Kurokawa, Imizu, Toyama, 939-0398, Japan; hirahara@pu-toyama.ac.jp
[2]Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology; 4259 Nagatsuda. Midori, Yokohama, 226-8502, Japan

ABSTRACT

Accurate head-related transfer functions, appropriate compensation of headphones and external-ear canal transfer functions are thought to be necessary for the binaural system to reproduce accurate three-dimensional sound image. One area of uncertainty is the extent to which the head movement can contribute to reducing the requirement for the binaural signals acoustical accuracy to hold the accuracy of three-dimensional sound image representation. This study of the issue asked listeners to localize sound image positions of real sources through full-, 70%- and 50%-scaled dummy-heads put on a *TeleHead*, which is a steerable dummy-head system that tracks three-dimensional human head movement quietly and quickly.

When head movements are allowed while listening, sound localization is nearly perfect with a full-scale dummy-head without compensating earphones and external-ear transfer functions. Most sound images are localized externally and correctly even with a 70%-scaled dummy-head. Moreover, some sound localization is possible even with a 50%-scaled dummy-head. Results suggest that listener's head movements strongly contribute to sound localization and that accurate binaural signal, which would be required for head stationary conditions, are not necessary when listeners can move their heads freely.

INTRODUCTION

Binaural technology requires binaural impulse responses. The best binaural impulse responses are head-related impulse responses (HRIRs) accurately measured with a listener's own head. The use of a dummy head precisely replicated from a listener's head also provides good binaural impulse responses. The use of other's HRIRs or other's dummy head provide somewhat different binaural impulse responses from those of the listeners own, thereby generating distorted binaural signals for the listener and thus producing degraded 3D sound. Moreover, it is known that appropriate compensation of transfer functions of both headphones and external-ear canal is necessary for reproducing accurate 3D sound.

Accurate and reproducible measurement of HRIRs with human subjects, however, is difficult. Physical factors, such as temperature and humidity of the measuring room, instability of the HRIR measurement system, a subject's head motions during the measurements, significantly affect HRIR measurements. In fact, the human subjects' heads moved in all directions during our 90 minutes HRIR measurements. Heads moved with a maximum pitch motion of ±20 degrees and head positions at the beginning and end of HRIR measurement differed more than 6 degrees. Without aid for head support, it is difficult to maintain a fixed head position. With aid for head support, a head support affects HRIRs.

Precise replication of a human head including pinnae is not as easy as someone thinks. We have made several user-like dummy heads by two methods [9,10]. One is a real life-cast dummy head of an FRP frame covered with soft silicone skin. This dummy head is made by moulding a human head shape using impression material. Another is a hard resin dummy head made by a rapid prototyping system using 3D human head shape data measured by MRI scan. Due to the gravity, inaccuracy of 3D shape measurement and expansion/shrinkage factor of materials, dimensional difference among real heads and dummy heads reach 10 mm or more.

These facts suggest that accurate individual HRIRs which differ among listeners are hard to obtain. Accordingly, inaccurate HRIRs or non-individual HRIRs are being used to reproduce binaural signals, resulting in a generation of distorted virtual 3D sound. A number of efforts have been paid to overcome these issues: adaptation of non-individual HRIRs to a listener [1]; selection of the best-fit HRIRs for a listener from an existing database [2]; and generation of individual HRIRs from listener's head geometry [3]. It should, however, be noted that most of these works have been evaluated under the head stationary condition; listeners are NOT allowed to move their heads during listening tests. It is natural to take listener's head and body movements into account in reproducing a virtual 3D sound, because, when we hear sound, there is often some accompanying movement of the head and body.

When a head is stationary, our brain computes the sound source position in a 3D space based on the static binaural acoustic information governed by the head and pinnae shape. When a head moves, the sound source position relative to the ears varies with head movements. The movement alters the HRIRs, thus deforms the binaural acoustic information dynamically. The brain then computes the sound source position in a 3D space based on the combination of the head movement information and the dynamic binaural acoustic information. Effects of head movement on the sound localization have been widely discussed. In general, head movement can reduce the front–back confusions as well as the sound localization error [4-10]. Further, it has also been suggested that accurate individual HRIRs are not necessarily required to reproduce a virtual 3D sound image when a listener is allowed to move his/her head while listening to the sound [10].

This study examined the extent to which the head movement can contribute to reducing the requirement for the binaural signals acoustical accuracy to hold the accuracy of 3D sound representation through sound localization experiments, in which listeners were asked to localize sound image positions of real sources through full-, 70%- and 50%-scaled dummy-heads in the head stationary and in the head movement conditions.

DUMMY HEADS

Dummy heads used in this study wee hard resin dummy heads made by rapid prototyping systems. 3D shapes of real heads were measured by a 1.5 T MRI (magnetic resonance imaging) system. The spatial resolution of the MRI image was 1×1×1 mm, where the image size was 256×256 pixel, fields of view was 256×256 mm and slice thickness was 1 mm. MRI images in DICOM format were transformed to 5 mm thick hollow 3D head models in STL format. During the format transformation, back of the external-ear canal and nostrils were shut, head surface was smoothed. Rapid Meister 6000 (Cmet) was used to make full-scale dummy heads and Dimension (Stratasys) was used to make 70%- and 50%-scaled down dummy heads. The photopolymer materials used in Rapid Meister 6000 was the TRS-821 epoxy-based resin and that in Dimension was ABS (acrylonitrile butadiene styrene) resin. It took 90 hours to make a full-scale dummy head.

HEAD RELATED IMPULSE RESPONSES OF DUMMY HEADS

HRIRs of the real heads and the dummy heads were measured in an anechoic room. The distance from the head center to the sound source was 1.2 m. The range of the measurement was from zero to 360 degrees in the direction of the azimuth. It was from -40 to 90 degrees in the direction of elevation. HRIRs were measured at 143 positions for a head. Each measured position in a median plane and the horizontal plane was set at intervals of 10 degrees. Other measured positions were set at intervals of less than 20 degrees between adjoined measured position s in the direction of the elevation and the azimuth. The external-ear canal was closed with individual earplug when HRIRs were measured. Small condenser microphones (Panasonic, WM62-AT102) imbedded in earplugs made of the silicon impression material were placed in the vicinity of the left and right external-ear canal entrance of subjects. Before each measurement began, we confirmed that the both positions of tragus and tip of the nose were calibrated using laser pointers. The sound source signal was the time-stretched pulses (TSPs) [11] of 32768 points with a sampling frequency of 48 kHz. Each HRIR was obtained by averaging ten TSPs.

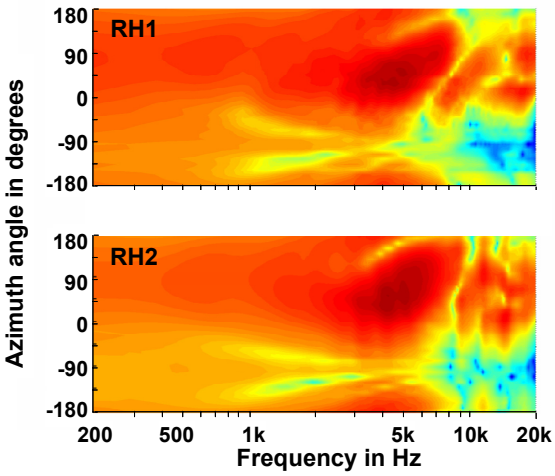
A head related transfer function (HRTF) is Fourier transform of the HRIR and is usually represented as an amplitude spectrum of the HRIR. Figure 1 shows HRTFs for horizontal plane of two real heads RH1 and RH2 and full-scale, 70% scaled and 50%-scaled dummy heads of

RH1. The principle in scale modelling is that all physical dimensions including the wavelengths are reduced by the scale factor. Namely, HRTF spectra of a 50%-sclaed dummy head should be that of a full-scale dummy head with 200% expanded frequency axis. Spectral dips in the HRTF of the full-scale dummy head shift upward as reducing the size of the dummy head. The amount of spectrum shift, however, is not exactly those expected from scale factors. Degraded accuracy of pinnae shape of scaled dummy heads, larger microphone size compared with concha size of scaled dummy heads, and reflections form a dummy head locking device might cause the deviation from the principal.

The difference in HRTFs between two heads was evaluated by a mean spectrum difference SD as (1);

$$SD = \frac{1}{M}\sum_{d}\sqrt{\frac{1}{N}\sum_{\omega}\left(20\log_{10}\frac{H_j(\omega,d)}{H_i(\omega,d)}\right)^2}\ [\mathrm{dB}] \qquad (1)$$

where *d* is measurement direction determined by azimuth angle θ and elevation angle φ, M is total number of directions, ω is angular frequency and N is number of frequency bin. The HRTF for direction *d* with head *i* is shown as $H_i(\omega,d)$. The SDs for all 36 directions in the horizontal plane between the real head (RH1) and the full-scale dummy head (DH) was 8.05 dB, that between RH1 and the 70%-scaled DH was 9.10 dB, and that between RH1 and the 50%-scaled DH was 7.45 dB. Those between the full-scale DH and the 70%-scaled and 50%-scaled DHs were 9.48 dB and 7.37 dB, respectively. That between RH1 and RH2 was 6.75 dB. SD value does not necessarily correspond to head shape/size differences.



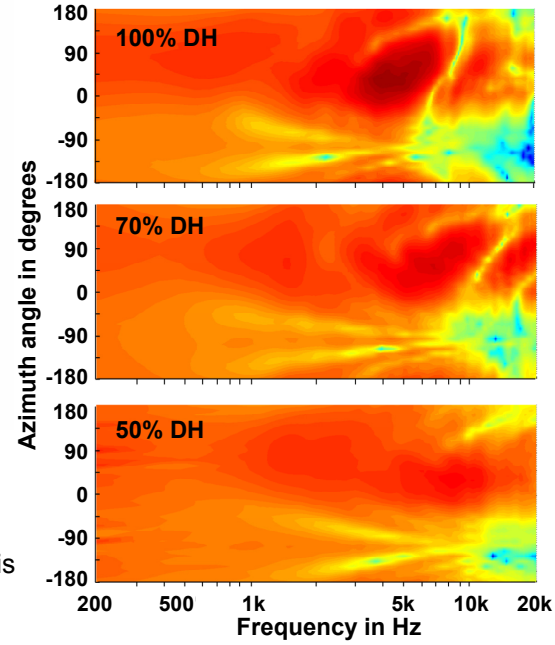


Fig.1. HRTF pattern of real heads (left panels) and dummy heads (right panels) in the horizontal plane. The real head 1 (RH1) is the model of dummy heads.

SOUND LOCALIZATION EXPERIMENT

Method

Subjects were asked to localize sound image positions of real sources on the horizontal plane through full-, 70%- and 50%-scaled dummy-heads put on *TeleHead I*, which is a steerable dummy-head system that tracks three-dimensional human head movement quietly and quickly [9]. As the task is the sound localization on the horizontal plane, the *TeleHead I* of which pitch and roll motors were disabled was used in the experiment. *TeleHead I* was set in an anechoic room, and 12 loudspeakers (VIFA, MG10SD0908) were placed around *TeleHead I* at intervals of 30 degrees in the horizontal plane. The distance from the loudspeakers to *TeleHead I* was 1.2 m. Sound stimuli were presented at the loudspeakers, collected by microphones in the dummy head put on *TeleHead I*, and transmitted to the subjects through headphones (Sennheiser, HDA200). Fig.2 illustrates experimental setups.

Stimulus was 5 seconds white Gaussian noise generated independently for each trial. Each stimulus was D/A converted (sampling frequency 48 kHz, 16 bits) and presented randomly in

directions with interval of 5 seconds. The stimulus level was 70 dB SPL. Each session consisted of 60 trials; stimulus was presented 5 times in random order from each of the 12 directions in the horizontal plane. One experiment consists of 4 sessions, resulting in responses of 20 trials from each of 12 directions.

Two normal hearing adult male subjects participated in the experiment. In a real head condition, subjects listen to the stimuli directly with their own head and ears in an anechoic room. In a dummy head condition, subjects listen to the stimuli over headphones in a soundproof room located 10 m from the anechoic room. Full-scale, 70%-scaled and 50%-scaled dummy heads of the subject 1 were used. In each condition, subjects were asked to keep their head as still as possible or were allowed to move their head freely.

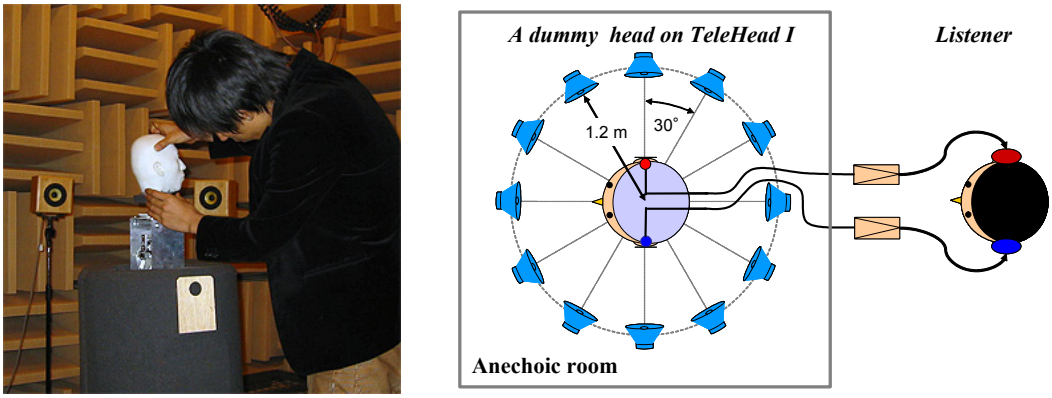


Fig. 2 Mounting a 70%-scaled dummy head on *TeleHead I* in an anechoic room (left). Setups for the sound localization experiment with *TeleHead I* (right).

Results

In the real-head condition, both subjects show almost perfect sound localization performance. There is no significant difference in performances between head stationary and moving-head conditions. All stimuli are, of course, localized externally.

Figure 3(a) shows the results of the subject 1 in three dummy head conditions. The upper middle and lower panels show the results for full-scale, 70%-scaled and 50%-scaled dummy head, respectively. The left panels show the results of the head stationary condition, the right panels show those of the head movement condition. In each panel, area of a blue filled circle is proportional to correct localization rate and that of a red filled circle is proportional to correct externally localization rate. When the full-scale own dummy head is used: there are some localization errors and most of stimuli presented at front position are localized inside the head in the head stationary condition. Localization error decreases and all stimuli are localized externally in the head movement condition. When the 70%-scaled own dummy head is used: localization error increases, front-back confusion occurs, and most stimuli presented at front and back positions are localized inside the head in the head stationary condition. Localization error decreases and most stimuli except from front position are localized externally in the head movement condition. When the 50%-scaled own dummy head is used: most stimuli are localized at rear position and all stimuli are localized inside the head in the head stationary condition. A number of stimuli are localized correctly and externally in the head movement condition. Subject 1 reported that stimuli are localized on the horizontal plane at his ear level with the full-scale dummy head but they are on the upper horizontal plane with scaled dummy heads.

.

Figure 3(b) shows the results of the subject 2 in three dummy head conditions. For the subject 2, the dummy head is not his own. When the full-scale others dummy head is used: there are typical front back confusions and most stimuli presented at front and back position are localized inside the head in the head stationary condition. Localization error considerably decreases and all stimuli except front position are localized externally in the head movement condition. When the 70%-scaled others dummy head is used: front-back confusion occurs, and most of stimuli presented at front and back position are localized inside the head. Total localization error, however, is almost the same as that with full-scale dummy head in the head stationary condition.

Localization error considerably decreases and most stimuli except front position are localized externally in the head movement condition. When the 50%-scaled others dummy head is used: some stimuli are still correctly localized in the head stationary condition. Some front back confusion remains but a number of stimuli are localized correctly and externally in the head movement condition. Subject 2 reported that the elevation of externally localized sound images moved upwards as reducing the dummy head size. He also reported that he felt as if his ears position moved upward to the top of head when 50%-scaled dummy head was used.

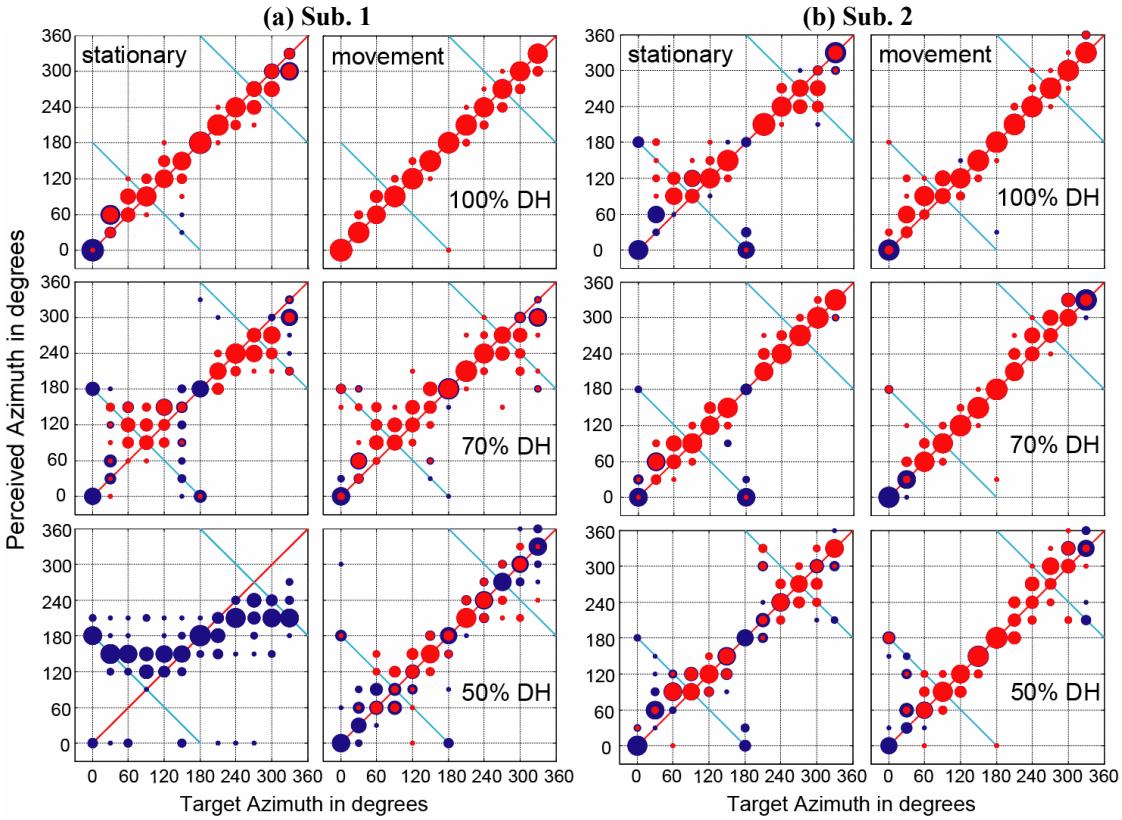


Fig.3 Sound localization results with dummy heads. Area of a blue filled circle is proportional to correct localization rate and that of a red filled circle is proportional to correct externally localization rate.

DISCUSSION

According to binaural theory, spatial localization performance relies on high-fidelity reproduction of binaural acoustic signal for the both ears. The use of others HRIRs or dummy heads often deteriorates localization performance. Compensation of listeners' external-ear canal transfer characteristics is mandatory when other's HRIRs or dummy head is used. Headphone transfer characteristics should also be compensated and the use of FEC (free air equivalent coupling to the air) headphones [12] is preferable. These strict acoustical requirements for binaural system are indispensable in the head stationary condition.

The experimental results, however, indicate that listeners can localize sound fairly well with others dummy head and even with 70%-scaled dummy head in the head movement condition. Neither headphone transfer characteristics nor listeners' external-ear canal transfer characteristics are equalized in the TeleHead system. Further, the headphones used in the experiment is HDA-200 [13], which are closed-type circumaural headphones and do not have FEC characteristics. Acoustical conditions needed in the head stationary condition are not necessarily indispensable in the head movement condition. It appears that head movements should be viewed as a natural and important factor of the sound localization process.

When a listener moves his/her head while listening to a stationary sound source, the interaural time difference, interaural intensity difference and spectral profiles change in accordance with the head movements. In addition, head movements provide additional non-acoustic information to the listener via proprioceptive feedback from the neck muscles. These

dynamic binaural information in accordance with head movements contributed in improving sound localization accuracies regardless of the dummy head size.

Head movements greatly reduce front-back confusions. This result is in line with experimental results using virtual sound sources generated by a dynamic auditory display. Listeners must have used changes in ITD, ILD and spectral cues accompanied by a head movement to avoid the front-back ambiguity. Surprisingly, front-back confusions are reduced even with scaled dummy heads in the head-movement condition. Head movements also facilitate sound images to be externalized. Hartmann *et al.* showed that externalization requires realistic spectral profiles in both ears [14]. Listeners must have used real dynamic changes in spectral profiles accompanied by a head movement. Surprisingly, sound images are better externalized even with scaled dummy heads when head moves while listening. The spectral profiles in both ears are not realistic with scaled dummy heads while dynamic changes in spectral profiles are realistic, suggesting that dynamic spectral profiles rather than static spectral profiles might play important role in externalization.

It is interesting that both subjects reported that apparent sound elevation was lifted up above the horizontal plane at the ear level as a dummy head size decreases. HRTFs at some elevation angle might have similar characteristics with those at horizontal plane of a scaled dummy head. Elevation judgement, however, was not the task of the experiments and HRTFs at detailed elevation angle of listeners were not measured. Interpretation of this phenomenon is an issue in the future.

CONCLUSIONS

This study investigates sound localization of real sources through full-, 70%- and 50%-scaled dummy-heads in the head movement condition. When head movements are allowed while listening, listeners can localize sound image well even with another's full-scale dummy-head without compensating earphones and external-ear transfer functions. Most sound image is localized externally and correctly even with a 70%-scaled dummy-head. Moreover, some sound localization is possible even with a 50%-scaled dummy-head. Results suggest that listener's head movements strongly contribute to the sound localization and that accurate binaural signals required for the head stationary condition are not necessary when listeners can move their heads freely.

References:

[1] J. Middlebrooks: "Individual difference in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, 106 (3), 1480-1492. (1999)

[2] Y. Iwaya: "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoustical Science and Technology,* 27 (6), 340-343. (2006)

[3] M. Otani, S. Ise: "Fast calculation system specialized for head-related transfer function based on boundary element method," *J. Acoust. Soc. Am.* 119 (5), 2589-2598. (2006)

[4] H. Wallach: "The role of head movements and vestibular and visual cues in sound localization," *J. Exp. Psychol.*, 27, 339-368 (1940).

[5] S. Perrett and W. Noble: "The contribution of head motion cues to localization of low-pass noise," *Percept. Psychophys.*, 59, 1018-1026 (1997).

[6] F. Wightman, D. Kistler: "Resolution of front-back ambiguity in spatial hearing by listener and source movement," *J.Acoust. Soc. Am.*, 105, 2841-2835 (1999)

[7] M. Wenzel: "Localization in virtual acoustic displays," *Presence*, 1, 80-107 (1992).

[8] D. Begault, E. Wenzel, M. Anderson: "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source", *J. Audio Eng. Soc.*, 49, (10), 904-916. (2001)

[9] I. Toshima, H. Uematsu, T. Hirahara: "A steerable dummy head that tracks three-dimensional head movement: TeleHead," *Acoustical Science and Technology,* 24 (5), 327-329. (2003)

[10] I. Toshima, S. Aoki, T. Hirahara: "Advance version of steerable dummy head that tracks human head movement in real timet: TeleHead II," *J .Acoust. Soc Japans,* 62 (3), 244-254. (2006)

[11] Y. Suzuki, F. Asano, H-Y, Kim, and T. Sone: "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, 97, 1119-1123. (1995)

[12] H. Moller: "Binaural Technique: Do We Need Individual Recordings?" *J.Audio Eng.Soc.,* 44(6), 451-469 (1996)

[13] T. Hirahara: "Physical characteristics of headphones used in psychophysical experiments," *Acoustical Science and Technology,* 25 (4), 276-285. (2004)

[14] W. Hartmann, A Wittenberg: "On the externalization of sound images," *J. Acoust. Soc. Am.* 99 (6), 3678-88. (1996)