*Acoust. Sci. & Tech.* **27**, 6 (2006)

**INVITED REVIEW**

# Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears

Yukio Iwaya*

*Research Institute of Electrical Communication, Tohoku University, Katahira 2–1–1, Aoba-ku, Sendai, 980–8577 Japan*

**Abstract:** A virtual auditory display (VAD) is a system for generating spatialized sound to a listener. Commonly, VAD techniques are based on convolving head-related transfer functions (HRTFs) to a sound source. When HRTFs in a VAD are not fitted to a specific listener, the accuracy of localization is often low and produces large localization errors, typically appearing as frequent front-back confusion. However, the measurement of HRTFs for each listener for all sound-source directions requires a special measuring apparatus and a long measurement time with a listener's physical load. The author has therefore proposed an individualization method of HRTFs called the Determination method of OptimuM Impulse-response by Sound Orientation (DOMISO). In this paper, DOMISO and its effects are introduced.

**Keywords:** Virtual auditory display, Head-related transfer function, Individualization

**PACS number:** 3.60.Gk, 43.66.Qp [doi:10.1250/ast.27.340]

This article contains the supplementary media files. Underlined file names in the article correspond to the supplementary files. For more information, see http://www.asj.gr.jp/2006/data/ast2706.html.

## Supplementary files

Tounament.exe, HRTF1.wav-HRTF8.wav

## 1. INTRODUCTION

We can empirically localize a sound position/direction as one of the auditory abilities with head-related transfer functions (HRTFs) [1], where the HRTFs are transfer functions that are related to a sound propagation path from a sound source to listener's ears and are defined as the Fourier transforms of the ratio of the sound pressure at the ear entrance to that at the center of the head in a free sound field in the frequency domain [1]. As an application of this schema, a virtual auditory display (VAD) can be realized as a system for generating spatialized sound for a listener. Many researchers have studied VAD. Commonly, VAD techniques are based on convolving HRTFs to a sound source. The accuracy of the perceived sound position is the most important feature of the VAD. When HRTFs in a VAD are fitted to a specific listener, a high accuracy of localization can be obtained and low localization errors produced, typically as low-frequent front-back confusion [2,3]. However, the measurement of a listener's own HRTFs in all directions requires a special measuring apparatus similar to that shown in Fig. 1, along with a long measurement time with a subject's physical load. Therefore, the listener's own HRTFs are often unavailable. In such cases, a method of estimating the most individualized HRTFs is needed. Seeber and Fastl [4] have examined selection methods of HRTFs on the basis of listening tests. Although the conventional method uses frontal HRTFs only in fitting, a fitting method for all directions is needed in actual systems. Therefore, a new fitting method, called the Determination method of OptimuM Impulse-response by Sound Orientation (DOMISO), is proposed in this paper. In the proposed method, listeners are asked to choose their most suitable HRTFs from among many HRTF data based on tournament-style listening tests. The validity of the proposed method is revealed on the basis of sound localization performance.

## 2. INDIVIDUALIZATION METHOD

### 2.1. Construction of Corpus of Measured HRTFs

Before applying DOMISO, a corpus of measured HRTFs is constructed. Figure 1 shows a spherical speaker array which is used in the measurement of HRTFs for all directions in our laboratory. This spherical speaker array has the following specifications.

(1) Seventy loudspeakers (FE83E; Fostex Co.) are installed on ribs, 1.5 m away from the center for every 10 degrees in elevation and azimuth.
(2) Every loudspeaker can be controlled independently to

*e-mail: iwaya@fir.riec.tohoku.ac.jp

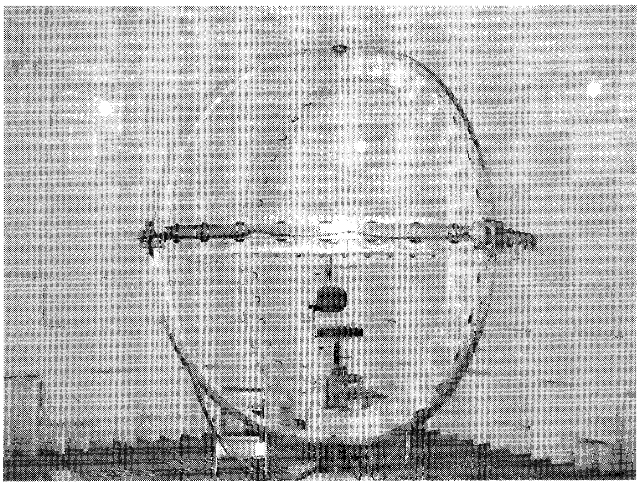


**Fig. 1** Spherical speaker array used in measurement of HRTFs.

generate arbitrary sounds.
(3) The spherical array can be rotated azimuthally with a precision of 0.1 degrees using a PC-controlled stepping motor.
(4) A chair is installed at the center of the sphere.

The procedure used to measure HRTFs for all directions is as follows: a listener is seated on the chair and a miniature microphone (FG3329; Knowles Electronics LLC) is placed at the entrance of ear canals of each listener. The listener's ear canal is blocked with clay [5]. Then, a time stretched pulse (TSP) is generated from one loudspeaker on the sphere for every elevation angle. This process is repeated by rotating the stepping motor for all azimuths at 5-degree intervals. Consequently, the HRTFs of a listener for every 5 degrees in azimuth and for every 10 degrees in elevation are obtained. In the corpus, 120 sets of HRTFs are installed in all.

### 2.2. DOMISO Procedure

Figure 2 shows a schematic of DOMISO. The proposed fitting method is described as follows.

(1) From among all the sets of HRTFs, 32 sets of HRTFs in the corpus are selected randomly.
(2) An orbit consisting of 13 virtual sound images (one-second-long pink noise, 16 bits, $f_s = 48\,\text{kHz}$) for every 30 degrees counterclockwise from the front of a listener in the horizontal plane is prepared using the selected 32 sets of HRTFs.
(3) The outline of this orbit of the virtual sound images is shown to listeners as an illustration before the listening test.
(4) Tournament matches are scheduled for 32 orbits rendered by the 32 sets of HRTFs.
(5) In a session (or a match), the listeners select one of two orbits that better resembles the instructed orbit. The selected orbit is the winner. The winner then proceeds to subsequent matches.
(6) Finally, one set of HRTFs wins the tournament; it is selected as the fitted set of HRTFs.

In this study, a Swiss-style tournament is used with slight modification. In a Swiss-style tournament, a match is scheduled between same-time winners. This style avoids a

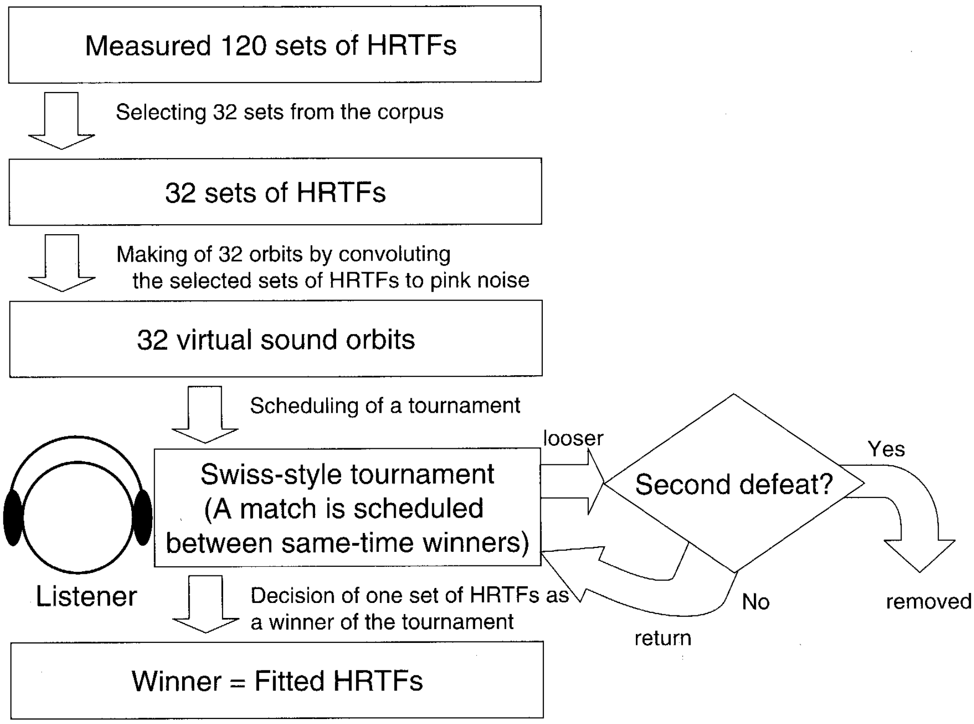



**Fig. 2** Schematic of DOMISO procedure.

defeat of two strong sets of HRTFs at early stages of the tournament. At the end of the tournament, the rankings for all the sets of HRTFs are obtainable by the original method of a Swiss-style tournament. However, because the author needs merely to evaluate strength, not weakness, a special rule was applied: a set that lost twice is removed from the tournament.

### 2.3. Advantages of DOMISO

Some advantages of DOMISO exist as follows: DOMISO can be carried out with a VAD system. A large-scale measurement of HRTFs is not required.

(1) **Easy Task**: A listener judges the equality of the perceived orbit. Therefore, the task for a listener is simple and little physical restriction exists.

(2) **Speedy**: The time cost for individualization is very short (about 15 min), whereas the measurement time for a set of HRTFs to construct the above-mentioned corpus required two hours per listener.

(3) **Simple**: A compensation for the characteristics of headphones is needed [5] for a high-quality reproduction when a sound image is displayed with headphones. However, because DOMISO uses the VAD system to be individualized, an additional signal compensation is not required.

## 3. PERFORMANCE EVALUATION

The selected set of HRTFs by DOMISO is evaluated in matches from the perspective of the agreement of the virtual sound orbit. The accuracy of the performance of sound localization, however, should be examined using another experiment. Therefore, the following experiment is carried out for this purpose.

### 3.1. Method

Seven adults with normal hearing ability participated. The participants carried out DOMISO in a soundproof room. Each virtual sound orbit was synthesized using personal computer (PC); sound was presented from headphones (SR-202 by STAX, with driver unit SRM-212) through a TDT System III. Each evaluation in DOMISO is answered through a graphical user interface (GUI) on a PC display. A set of HRTFs for each listener was also measured with the spherical loudspeaker array mentioned above.

Then, an experiment on the sound localization of virtual sound images is performed to evaluate the performance of DOMISO. A sound image of pink noise (2 s, 16 bits, $f_s = 48$ kHz) is synthesized with a specific direction of HRTFs and is displayed through the same experimental setup. The directions displayed in the experiment are from 0 degrees to 330 degrees in the horizontal plane with 30-degree intervals, where 0 degrees indicates the front of the listener. Each participant answered his or her perceived direction with a resolution of 10 degrees. Each direction is displayed five times. In all, 60 stimuli are presented in a random sequence. In the experiment, the following three conditions are examined.

- **Own condition**: a set of the listener's own HRTFs is used.
- **Individualized condition**: a set of HRTFs that is individualized by DOMISO is used.
- **Away condition**: a set of HRTFs that could not win at all in the DOMISO procedure is used.

### 3.2. Results and Discussion

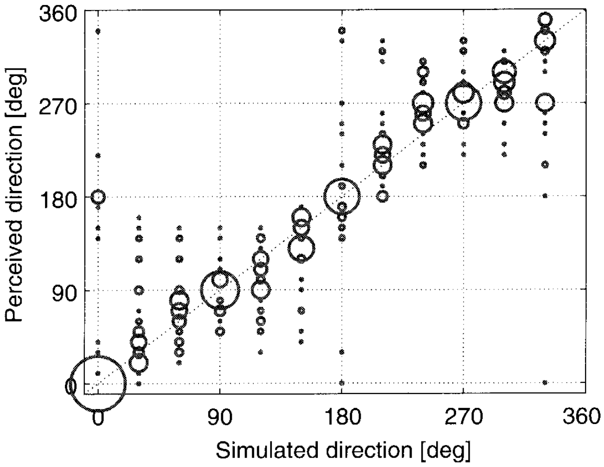
Figure 3 shows the results of the experiment. The horizontal and vertical axes respectively show the simulated and the perceived directions, where 0 degrees indicates the front of the listener. Panels (a), (b) and (c) respectively indicate the Own, Individualized, and Away conditions. The area of each circle in these figures is proportional to the frequencies of the perceived direction. Cases of front-back confusion are observed under all conditions. Therefore, the front-back confusion rates are calculated (Fig. 4). When a presented direction in the frontal semi sphere is perceived in the rear semi sphere, the judgement of the direction is treated as a front-back confusion and vice versa. In the count of front-back error rate, presented directions of 90 degrees and 270 degrees are excluded. From Fig. 4, the relationship of the front-back confusion rates among the three conditions can be expressed as *Away* > *Own* > *Individualized*. An analysis of variance (ANOVA) is performed, where the condition of HRTFs is treated as a factor. Results show that the main effect is significant ($F(2, 18) = 4.1$, $p < 0.05$). Therefore, Tukey's multiple comparison test (HSD) is performed. The difference between the Individualized condition and the Away condition is significant. On the other hand, the difference between the Individualized condition and the Own condition is not significant.

Similar ANOVA for the accuracy of the perceived direction was also performed after excluding the Front-Back errors. However, no significant difference among the three conditions is apparent.

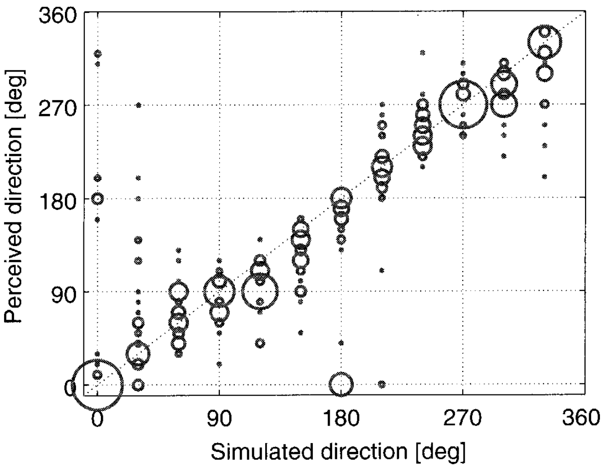
From these results, we conclude that the selected set of HRTFs with the proposed procedure, DOMISO, has better performance in the direction of perception than in the Away condition. For this reason, DOMISO might be an effective method for the individualization of HRTFs.

## 4. CONCLUSION

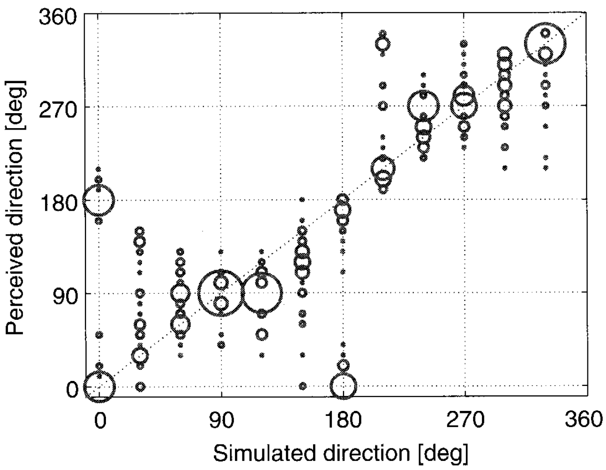
In this paper, a new method for the individualization of a set of HRTFs using actual VADs was proposed. The proposed method requires no large measurement environ-




(a) Own Condition




(b) Individualized Condition




(c) Away Condition

**Fig. 3** Results of experiment.

ment or effort; it requires only a VAD environment and a short time. An evaluation experiment of the proposed method showed that the sound localization performance of the proposed method is better than that of a non fitted set of HRTFs.

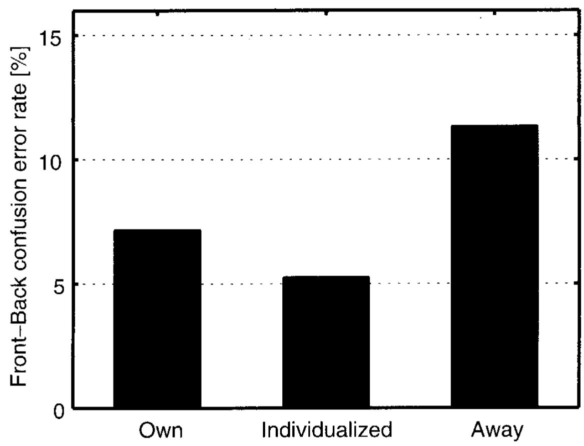



**Fig. 4** Front-back error rate among conditions.

Although the difference between the Own and the Individualized conditions are not significant, the front-back confusion rate in the Individualized condition is slightly smaller than that in the Own. When we wear glasses, we can see everything around us. In the auditory case, a set of HRTFs which enable us to localize a sound position clearly may exist.

## ACKNOWLEDGEMENTS

This study was supported in part by the Consortium R&D Projects for Regional Revitalization (15G2025) of the Ministry of Economy, Trade and Industry, and by a Grant-in-Aid for Young Scientists (B) of the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government (17700090).

## REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Revised ed. (The MIT Press, Cambridge, Mass., 1996).

[2] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, **1**, 167–174 (1980).

[3] Y. Iwaya, Y. Suzuki and S. Takane, "Effect of listener's head movement on the accuracy of sound localization in virtual environment," *Proc. ICA 2004*, pp. 997–1000 (2004).

[4] B. Seeber and H. Fastl, "Subjective selection of non-individual head-related transfer functions," *Proc. ICAD 2003*, pp. 259–262 (2003).

[5] H. Møller, "Fundamental of binaural technology," *Appl. Acoust.*, **36**, 171–218 (1992).

## APPENDIX

A program (Tournament.exe) can execute a simple tournament-style listening test for the individualization of HRTFs. In this program, eight wave files (HRTF1.wav, HRTF2.wav, HRTF3.wav, HRTF4.wav, HRTF5.wav, HRTF6.wav, HRTF7.wav, HRTF8.wav,) are used. You can listen to this program with headphones.