

Toward Cognitive Usage of Binaural Displays



Yôiti Suzuki, Akio Honda, Yukio Iwaya, Makoto Ohuchi
and Shuichi Sakamoto

Abstract Based on acoustic input to their two ears, humans are able to collect rich spatial information. To explore their acoustic environment in more detail, they thereby move their bodies and heads to resolve ambiguities as might appear in static spatial hearing. This process is termed “*active listening*.” This chapter introduces new research regarding two specific aspects of active listening, namely, (i), facilitation of sound localization in the median plane and, (ii), augmentation of the discrimination angle for frontal auditory object. As active listening affects spatial hearing significantly, the design of systems for spatial-sound presentation requires substantial expertise in this field. In this context, a dynamic binaural display was developed that supports active listening. The display was applied to edutainment applications such as training the spatial-perception competence of visually impaired persons. Two examples were specifically investigated for this purpose, namely, a maze game and an action game. The former facilitates players’ ability to draw cognitive maps. The latter improves the sound-localization performance of players, their eye-contact frequency during conversation, and their ability to avoid approaching objects. Results suggest that binaural displays that support active listening are indeed capable of enhancing listener experience in reproduced and virtual auditory scenes.

Y. Suzuki (✉) · S. Sakamoto
Research Institute of Electrical Communication, Tohoku University,
2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan
e-mail: yoh@riec.tohoku.ac.jp

A. Honda
Faculty of Informatics, Shizuoka Institute of Science and Technology,
2200-2 Toyosawa, Fukuroi 437-8555, Japan

Y. Iwaya
Faculty of Engineering, Tohoku Gakuin University,
1-13-1 Chuo, Tagajo 985-8537, Japan

M. Ohuchi
Faculty of Comprehensive Management, Tohoku Fukushi University,
1-8-1 Kunimi, Aoba-ku, Sendai 981-8522, Japan

© Springer Nature Switzerland AG 2020
J. Blauert and J. Braasch (eds.), *The Technology of Binaural Understanding*,
Modern Acoustics and Signal Processing,
https://doi.org/10.1007/978-3-030-00386-9_22

1 Introduction

We humans are active creatures. It is therefore quite natural for us to explore environments by moving through them to collect accurate spatial information. These inherent movements are also true for spatial hearing (Blauert 1997). The process is not restricted to audition, but may include crossmodal cues, such as from the vestibular system. The process of gathering auditory information during exploratory head and body movements is known as “*active listening*” (Suzuki et al. 2012). Numerous studies show that active listening facilitates auditory spatial perception. Wallach (1939) demonstrated that these movements provide cues for elevation-angle assessment. Also, Thurlow and Runge (1967) showed that among horizontal head turning, nodding, and pivoting, horizontal head turning is critical for proper sound-source localization. Similar findings were obtained in virtual auditory spaces. For instance, Kawaura et al. (1989, 1991) examined sound localization of a virtual sound source, spatialized at a distance of 1.5 m from the center of listener’s heads by convolving acoustic signals with head-related impulse responses (HRIRs), the time domain representation of head-related transfer functions (HRTFs). They reported that front-back and distance judgments were markedly improved when horizontal head rotations were properly reflected in the binaural signals. Indeed, numerous reports describe the facilitation of sound localization by head movement—see Perrett and Noble (1997), Iwaya et al. (2003), Toshima and Aoki (2009), Brimijoin et al. (2013). Furthermore, active listening facilitates affective cognition (Iwaya et al. 2011). The latter authors demonstrated that head rotation enhances the *sense of presence* of listeners in virtual auditory spaces.

Moreover, in the last decade, a few studies have shown that sound-image-localization accuracy is reduced by head movement. For example, Cooper et al. (2008) presented a test sound while listeners were rotating their heads. Results showed reduction of sound-localization accuracy for sound stimuli presented during head rotation, compared with that of a static condition. Leung et al. (2008) examined auditory spatial perception during rapid head motion and reported that the perceived auditory space was compressed. Honda et al. (2016) measured movement detection for a virtual sound source during listener’s horizontal head rotation. Results showed that detection thresholds were higher (i.e. worse) when listeners rotated their heads. These results urge us to further investigate the active-listening process to draw a more complete picture.

Either way, since humans must take advantage of *active listening* to appropriately understand sound environments, knowledge of active listening is relevant and indispensable for optimal and effective design of three-dimensional (3D) sound-rendering systems. In this context, the authors are particularly interested in so-called *binaural displays*. These displays are a type of 3D auditory displays that render auditory spaces by controlling the sound signals directly at the ears of the listeners. Thereby it is important to consider the listeners’ movements in order to deliver appropriate signals to the two ears. This requires taking into account the actual listener positions with respect to the sound sources. This *head-related* rendering of sound signals

is in contrast to other types of 3D auditory displays, such as wave-field synthesis (WFS) (Berkhout et al. 1993), boundary-surface control (BoSC), (Ise 1997; Enomoto and Ise 2005), and high-order Ambisonics (HOA) (Poletti 2005), for which listeners' rotational and translational movements both are naturally reflected to their ear inputs if they remain inside the listening zone.

Following the notion of head-related sound-field rendering, a middleware module for binaural display, called *Simulation Environment for 3D Audio Software* (SiFASo) (Iwaya et al. 2005, 2011) was developed in the authors' laboratory. With SiFASo, education¹ applications have been built for the purpose of training spatial-perception competence, particularly that of visually impaired people (Honda et al. 2007, 2009, 2013; Ohuchi et al. 2006).

Binaural displays are generally applicable to virtual reality applications as assistive technology. Specifically, for visually impaired persons they are of relevance for enhancing the *quality of life*—compare (Afonso et al. 2005; Iwaya et al. 2009; Picinali et al. 2014; Seki et al. 2011; Seki 2016). Experiments in which the effect of the SiFASo system was assessed reveal promising potential of binaural displays to improve spatial perception and to take advantage of some transfer effects that are useful in daily life.

2 Head Turning and Sound-Source Localization in the Median Plane

Many studies, including those described in this chapter, indicate that head movements facilitate sound localization, including front-back discrimination. One of these studies used a robot, the TeleHead (Toshima et al. 2003). “TeleHead” is an avatar robot that follows the head movements of a human in the following way. A listener in a location different from that of the robot listens to sound signals delivered from two microphones on the robot at ear positions. Listening with TeleHead improves horizontal-plane sound localization (Toshima and Aoki 2009). However, head turning improves sound localization in the median plane as well (Perrett and Noble 1997). Previous studies with the robot (Suzuki et al. 2012) revealed that it facilitates median-plane sound localization even if the rotation angles of the robot are smaller than those of human listeners.

However, it has not yet been clarified how horizontal head rotation of human listeners should be reflected to generate virtual auditory space by means of binaural displays. To clarify this issue, the effects of horizontal head rotation on sound localization were studied by the authors in greater detail. In this context the effects of horizontal rotation of the robot were investigated, whereby the turning was either in phase or in anti-phase compared to the turning of the human listener.

¹A portmanteau word composed from “education” and “entertainment”.

2.1 Experimental Procedure

Three young male adults with normal hearing participated as listeners. They were all well trained for sound-localization experiments.

The robot used was the same as that used by Suzuki et al. (2012), namely, a simplified TeleHead that follows only horizontal head turns.² This simplification was applied because horizontal head rotation (*head turning*) plays by far the most important role among the three possible rotational head movements (Thurlow and Runge 1967). A head simulator (dummy head) cast after each listeners' own head was set atop the robot. The dummy head can follow head turning up to 200%. The average system latency during operation was about 50 ms. This is slightly shorter than the detection threshold in binaural displays, and much shorter than the acceptable limit in listening tests (Yairi et al. 2007, 2008b).

Figure 1 shows a schema of the experimental setup. The TeleHead was positioned in an anechoic chamber at the center of a circular loudspeaker array set-up in the median plane. The distance between the center of the set-up and the surface of the loudspeakers was 1.5 m. The sound stimulus was a pink noise of 10 s duration including 6 ms rise and decay times using a raised cosine function, and was presented through one of the 16 loudspeakers in the median plane. The loudspeaker array was arranged with an elevation of -60° in front, climbing up to the zenith (elevation of 90°), and continuing to -60° in the rear, with 20° separation between the loudspeakers. Sound signals received at the TeleHead's two ears were reproduced by headphones (Sennheiser HDA-200) in real-time. The listeners were seated in a soundproof room next to the anechoic chamber. The sound pressure level of the stimulus was set to 70 dB for frontal sound incidence. The sound pressure levels were calibrated with an artificial ear conforming to the IEC 60318-1:2009 (B&K 4153) standard, with an adapter for circumaural earphones specified in the same IEC standard (B&K DB 0843). The force added to the adapter was set to 8 ± 1 N.

The experiment consisted of seven conditions. In one, the robot kept still, with the virtual sound source in front, that is, the static condition. In the other conditions, the robot moved in-phase or anti-phase with respect to the remote listener's horizontal head rotation. The amount of head turning of the robot with respect to the listener's head rotation was modified. The rotation ratio was selected among ± 0.05 , ± 0.1 , and ± 1.0 . Here, plus and minus signs respectively mean in-phase and anti-phase rotation in relation to that of the remote listener. Listeners were asked to move their heads freely, at least once during each trial, and to identify the loudspeaker direction from the 16 alternatives. The number of repetitions was five, and the total number of trials for each condition was 80 (16 directions \times 5 repetitions each).

²This simplified version is based on TeleHead's fourth version (Hirahara et al. 2011).

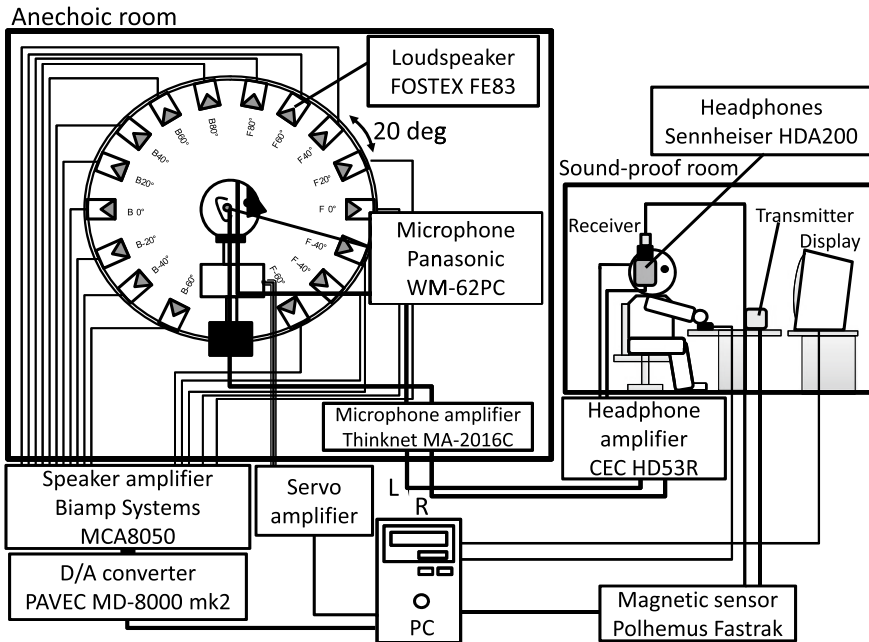


Fig. 1 Median-plane-localization setup with TeleHead

2.2 Results and Consideration

Figures 2, 3, and 4, respectively, show the localized direction as a function of physical direction in static, in-phase, and anti-phase conditions. Figure 5 shows the front-back error rate as a function of the rotation ratio for three in-phase and three anti-phase conditions compared with the static condition. Two one-way analyses of variance (ANOVA) were applied. However, Mauchly's sphericity tests indicated that both data shown in Fig. 5a, b exhibited significant departures from sphericity ($p < 0.01$). Therefore, the Greenhouse–Geisser correction factor was applied to the degrees of freedom of the ANOVA analysis. Results for in-phase rotation indicate that the effect of the rotation ratio is statistically significant ($F(1.42, 2.85) = 18.03, p < 0.05$). Multiple comparisons (Tukey's HSD, $p < 0.05$) indicate significant differences between the static and +0.1, static and +1.0, and +0.05 and +1.0 conditions. Results for anti-phase rotation indicate no significant effect ($F(1.56, 3.13) = 3.20, n.s.$).

Figure 2 shows that frequent front-back errors occurred in the static condition where TeleHead did not respond to the listener's head rotation. In contrast, Fig. 3 shows that front-back confusions were suppressed when the robot rotated in-phase to listeners' rotation, irrespective of the rotation ratio. The results of ANOVA confirm that the suppression is significant, not only when the ratio of TeleHead's rotation is 100%, but also when it is 10% of the listener's head rotation. Although the multiple comparison does not show any significant difference between the static and +0.05

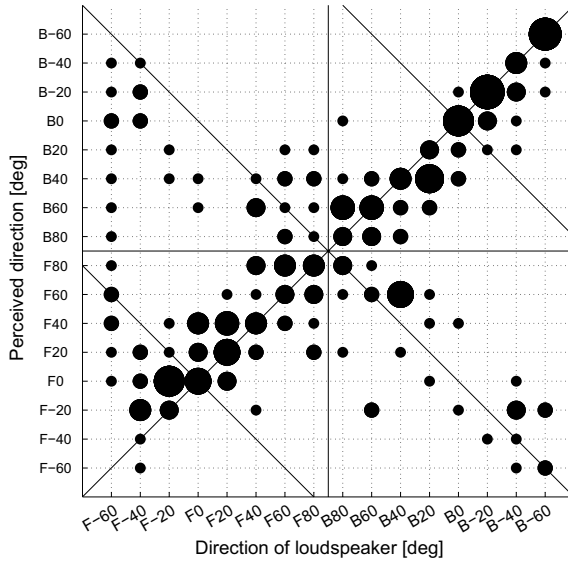


Fig. 2 Relationship between presented and localized elevation angles for the static condition ($r = 0.44$)

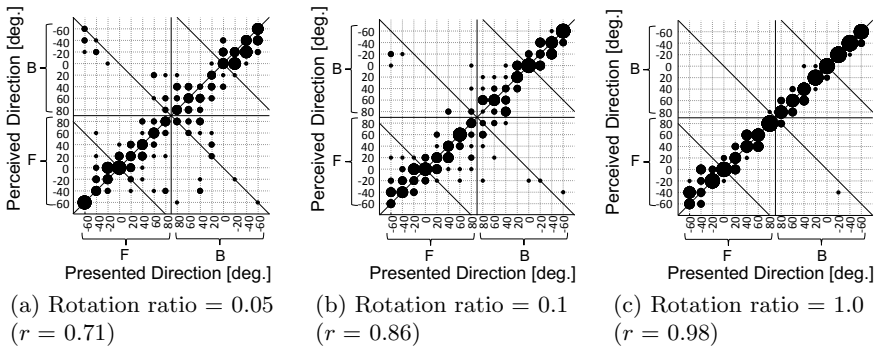


Fig. 3 Relationship between presented and localized elevation angles when TeleHead rotates in-phase

conditions, the differences in correlation coefficients between the static condition and those of the +0.05, +0.10, and +1.0 conditions are all statistically significant ($ps < 0.001$). This signifies that the in-phase feedback to the listeners results in significantly “sharper” distributions, which may imply fewer localization errors, including front-back confusions, even when the ratio is +0.05.

These results mean that an avatar robot can provide effective dynamic sound-localization cues when it rotates in-phase with the listener’s rotation, even with head turnings of the robot are as low as 5% that of the active listener.

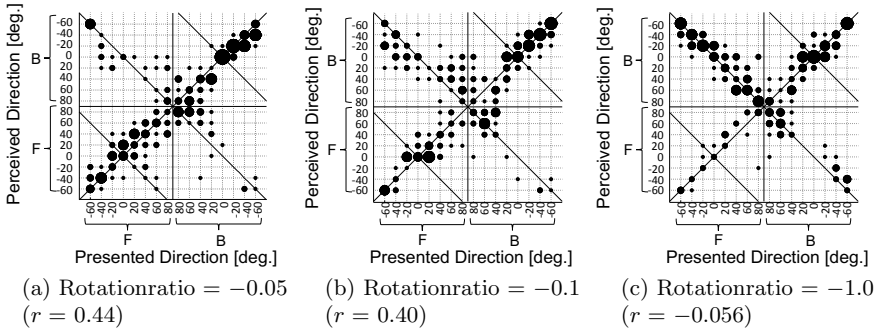


Fig. 4 Relationship between presented and localized elevation angles when TeleHead rotates anti-phase

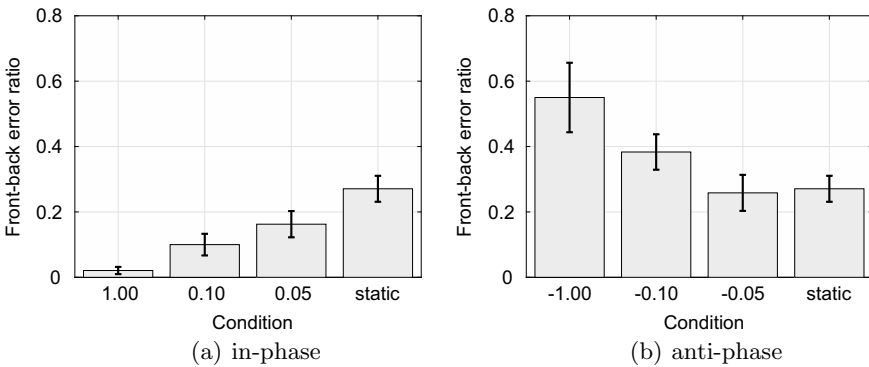


Fig. 5 Front-back error rates as a function of rotation ratio

It is noteworthy that all listeners reported that the dynamic control to reflect head rotation was unnoticeable when the ratio of the in-phase rotation was less than or equal to 10%, while they noticed the dynamic control when the ratio was 100%. Hirahara et al. (2013) showed similar results for horizontal sound localization in an experiment also using TeleHead. These results suggest that head movements can be implicitly utilized to stabilize sound localization, and that the direction of rotation is important, not just the head rotation itself.

Figure 4 shows that perceived elevation angles hardly correlate with physical sound-source direction when the rotation ratio is -1.0 (Fig. 4c). Anti-phase rotation with a ratio of 1.0 provides reversal dynamic cues in terms of front-back confusion. In fact, all three listeners reported that they often experienced such reversals, resulting in small correlation of perceived and physical directions. In contrast, localization seems hardly affected by anti-phase rotation with rotation ratios of -0.05 and -0.1 (Fig. 4a, b). This may result in robustness against unusual disturbances of static spectral cues in median-plane localization. The results of this study confirm

that horizontal rotation of the heads of listeners provides dynamic cues for proper elevation-angle localization. This suggests that this kind of rotation should be taken into account in when designing high-definition dynamic binaural displays.

3 Localization Accuracy of the Subjective-Straight-Ahead During Active Head Rotation

In most of the studies that show facilitation of sound localization by listener movements, listeners were asked to estimate over-all sound image positions after presentation of sounds and listener movements had ended. However, *localization accuracy* can be defined as a measure of the deviation of the position of a perceived auditory object from the physical position of the respective sound source. Following this notion, there have been a few studies that observed deteriorated sound-localization accuracy during listeners' head rotations when dealing with real sound sources (Cooper et al. 2008; Leung et al. 2008) as well as virtual sound sources as rendered by a binaural display (Honda et al. 2016).

There is obviously a need for more knowledge regarding sound-localization accuracy during listener motion. Thus, sound-localization accuracy regarding the *subjective-straight-ahead* was investigated in the horizontal plane precisely at the moment when listeners actively rotated their heads. The findings were compared with the static case (static condition).

3.1 Auditory Subjective-Straight-Ahead in Static Condition

In the static condition, the listeners were sitting still on a chair, but their heads were not mechanically fixed.

Experimental Procedures

Eight males and one female with normal hearing (22–40 years of age) participated. The sound stimuli consisted of 1/3-octave-noise bursts ($f_c = 1$ kHz, SPL: 65 dB when presented continuously) of 15, 30, 80, 150, and 300 ms, including 5-ms rise and decay times. An arc array of 35 loudspeakers arranged with 2.5° separation at a distance of 1.1 m from the listener (see Fig. 6) was set up. A sound stimulus was presented from one of seven loudspeakers located within $\pm 7.5^\circ$. An LED was mounted on the loudspeaker at 0° . For each trial, the LED lit first for 1 s, and then a sound stimulus followed. The experiment was conducted in an anechoic chamber which was kept dark during sessions, so that loudspeaker positions were not visible. Listeners were asked to judge whether a test stimulus was located to left or right of their subjective straight ahead exactly at the time of presentation (two-alternative forced choice). The method of constant stimuli was used, and the number of repetitions for each

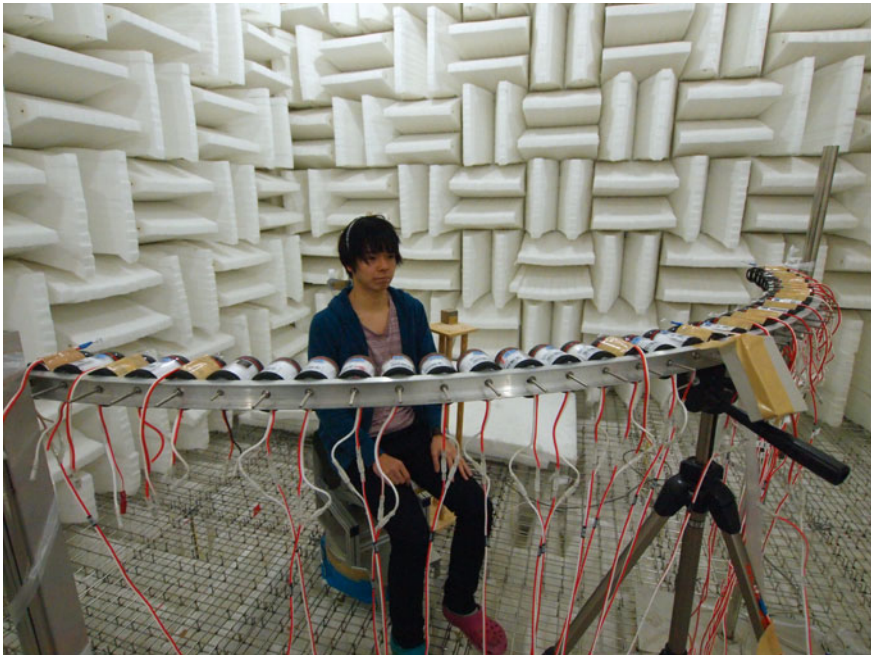


Fig. 6 Experimental setup

direction was 20. Thus, the total number of trials was 700 (5 durations \times 7 directions \times 20 repetitions).

Results

The cumulative normal distribution function was fitted to the ratios of the right-of-subjective-straight-ahead judgments as a function of stimulus direction for determining the point of subjective-straight-ahead (PSSA), and its just-noticeable difference (JND). Here, PSSA is the direction of the subjective-straight-ahead relative to the physical front.

The PSSA and its JND were respectively estimated as the mean and $0.675 \sigma^3$ of the fitted cumulative normal distribution function. Figure 7a, b respectively show PSSAs and JNDs as a function of the stimulus duration. One-way analysis of variance (ANOVA) indicates no significant difference for the direction, while the effect of the duration on JNDs are significant ($F(4, 32) = 5.29, p < 0.05$). Multiple comparisons (Tukey's HSD test, $p < 0.05$) indicate significant differences between 15 and 150 ms, and between 15 and 300 ms.

³The value of 0.675 corresponds to the z score where the cumulative normal distribution reaches 0.75, meaning an estimated correct answer rate of 75%.

3.2 Auditory Subjective-Straight-Ahead During Active Head Rotation

Sound-localization accuracy of the subjective-straight-ahead in the horizontal plane was determined with listeners sitting on a chair rotating their heads by themselves, both slowly and rapidly—that is, in active slow and active fast conditions.

Experimental Procedures

Eight of the previously tested nine listeners (seven males and one female) participated in this experiment. Examining the results of the static condition (Fig. 7), the duration of sound stimuli was set to 30 ms. This was determined because 15 ms seemed to be too short, since the JND is significantly larger for the short duration than for the longer one. However, the length should be as short as possible to minimize directional deviation during stimulus presentation. Otherwise, the experimental setup was the same as used for the static condition. Also, the method of constant stimuli was used as well. However, the method of stimulus presentation was modified to match the listeners' rotations as follows. For each trial, a guiding sound was presented from a loudspeaker located at either -45 or $+45^\circ$ for 100 ms to indicate the direction towards which the listener should rotate the head. Listeners were instructed to rotate their heads either quickly or slowly toward the direction of the guidance sound. The actual speed of rotation was observed with a motion sensors on the listeners' heads (Polhemus, Fastrak). A sound stimulus was presented when the listeners rotated their head by at least 15° . For clockwise or counterclockwise rotation, the stimulus was presented via one out of of 13 loudspeakers ranging from 0 to $+30$ or -30° . The total number of trials was 520 (2 rotational directions \times 13 stimulus directions \times 20 repetitions each).

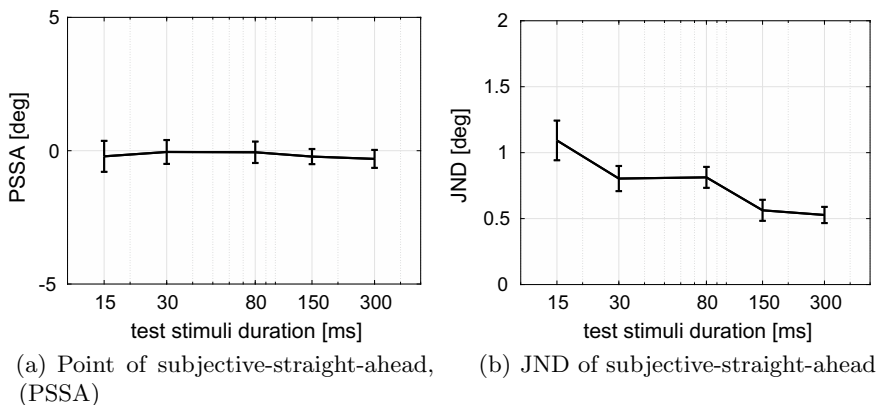


Fig. 7 Direction and JND of subjective-straight-ahead as a function of sound duration for the static condition

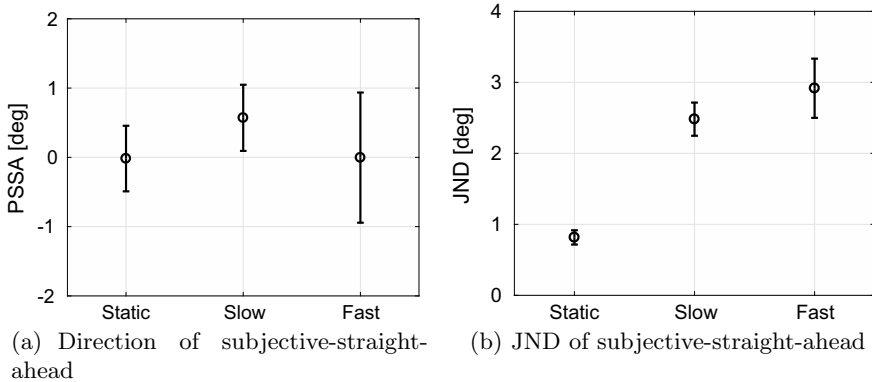


Fig. 8 Direction and JND of subjective-straight-ahead for static, active slow, and active fast conditions

Results

The means, m , and standard deviations, σ , for “slow” and “fast” head rotation speeds are $m = 12.2$, $\sigma = 4.0^\circ/s$ and $m = 166.9$, $\sigma = 69.5^\circ/s$, respectively. Paired t -test shows a significant difference between the two conditions ($t(7) = 6.07$, $p < 0.01$). This shows that the listeners could well control the two speeds. In the calculation of the PSSA and its JND, the data for clockwise and counterclockwise rotations were pooled by treating the sign of the direction toward rotation as positive, considering the symmetry of the experimental scheme.

Figure 8 depicts the PSSAs and their JNDs for the active fast and slow conditions, along with the results for the eight participants in the static condition. One-way ANOVA indicates no significance for PSSA. Alternatively, the effect of JND was significant ($F(2, 14) = 19.80$, $p < 0.01$). Multiple comparisons (Tukey’s HSD test, $p < 0.05$) indicate significant differences between the static condition and the two conditions with rotation.

3.3 Discussion

Figure 7 shows that auditory JNDs of the subjective-straight-ahead for the static condition are well below 1° when the sound duration is at least 30 ms, and are almost 0.5° when it is 150 ms or longer. This value is smaller than the minimum audible angle (MAA) in front (Mills 1958). While MAA is the difference limen of two sound images at a certain incident angle, this JND is the detection threshold for the deviation of a perceived sound object from a reference defined as directly in front. Considering this difference, the auditory subjective-straight-ahead can be regarded as very stable.

The experimental results plotted in Fig. 8b indicate that the auditory JND of the subjective-straight-ahead is significantly larger during head rotation than when lis-

teners are sitting still, irrespective of the rotation speed. Since the stimulus duration was 30 ms, the head rotated 0.36 and 5.0°, respectively, for the observed mean rotation speeds of 12 and 167°/s).

The relative drift of the direction of the sound sources, and thus that of the perceived auditory objects, can account for the increase of the JNDs of the fast condition to some extent, but this is not sufficient to also explain those of the slow condition. This may imply that the degradation is not only attributable to ambiguities of the ear-input signals induced by the movement but also to possible change of the binaural information processing in the brain for static and dynamic binaural inputs. To examine this possible difference in the mechanism of spatial hearing, further experiments for lower rotation speeds and with passive rotation should be performed.

The fact that the effects of listener's movement are not uniform is puzzling. Head motion often facilitates sound localization but may deteriorate in its accuracy, as shown here and in Sect. 2. A phenomenological explanation might be the difference in the way localization judgments occur. That is, facilitation seems to occur when overall sound localization is requested after presentation of sound stimuli and listener movements have ended, whereas deterioration is observed when instantaneous sound localization is reported during presentation while the listeners are in the course of moving. In other words, the former would resolve ambiguity in this ill-posed problem caused by scarcity of hearing inputs in only two channels (i.e. two ears), while the latter would stabilize auditory spatial perception during ear-input changes. Thus, this phenomenon can be compared to saccadic suppression in vision. Moreover, this phenomenon may be useful to design efficient 3D auditory displays, including dynamic binaural displays, because fewer computational resources can be assigned while listeners (or sound sources) are in motion.

4 A Binaural-Display Middleware, SiFASo

4.1 Binaural Displays

A binaural display is an architecture for 3D auditory display that synthesizes or reproduces the input sound signals at the listeners' ears. To realize this, sound-source signals are typically convolved with the impulse responses of the sound propagation paths from a sound source to listeners' ears. The frequency domain representation of the impulse responses of the paths can be expressed as a cascade of HRTFs and the room transfer functions (RTFs).

Psychoacoustic performance is generally good despite the simple signal processing. Moreover, as described earlier, rendering performance of binaural displays can be greatly improved by appropriately reflecting listener movements in the ear-input signals. Following the approach used in head-mounted displays (HMDs), such processing is indispensable for high-performance binaural displays to properly support active listening.

Since Morimoto and Ando (1980) first realized basic binaural displays digitally, special hardware such as digital signal processors (DSPs) had been necessary for many years to implement them (Takane et al. 1997; Blauert et al. 2000; Suzuki et al. 2002; Iwaya et al. 2002; Begault et al. 2010). However, by the 21st century it became possible to build simple binaural displays with CPUs of ordinary personal computers (Savioja et al. 1999; Wenzel et al. 2000; Miller 2001; Lokki and Järveläinen 2001; Yairi et al. 2006; Zhang and Xie 2013). In the authors' laboratory a high-performance dynamic binaural display supporting active listening as middleware, named SiFASO (Iwaya et al. 2005, 2011), has been developed. Note that so far binaural displays supporting active listening, including SiFASO, are compatible with active listening in terms of taking advantage of facilitation by listeners' movements. Designs of future dynamic binaural displays will certainly leverage current knowledge such as the suppression of sound-localization accuracy during listener movements, as discussed in Sect. 3.

4.2 *Outline of SiFASO*

SiFASO was developed based on experience with simple but low-latency (i.e. <12 ms) implementations, including the latency of position sensors (see also Yairi et al. 2006, 2008a). SiFASO can render a 3D auditory space including presentation of multiple sound sources by convolving source signals with proper individualized head-related impulse responses (HRIRs). Further, Doppler-effect (Iwaya and Suzuki 2007), 1st-order reflections, and reverberation processing are implemented. The HRIRs are interpolated to achieve smooth head and sound-source movements. Total system latency of SiFASO is about 30ms, including the head-tracker latency (Iwaya et al. 2011). Exploiting these advantages, SiFASO realizes stable, precise, and natural positioning of rendered sound images, even for moving sounds. The class diagram of the main part of SiFASO is presented in Fig. 9. SiFASO was developed as a dynamic-link library (DLL), so that it can be easily invoked from various applications. SiFASO runs under MS Windows on the CPU of a personal computer.

4.3 *Edutainment Welfare Applications for the Visually Impaired*

SiFASO was primarily developed for welfare systems to train spatial perception, particularly for visually impaired people, that is, those who must recognize spaces without having visual cues at their disposal. They are known to have better spatial hearing capabilities than sighted people. However, this sensory compensation varies with the etiology and extent of vision impairment (Paré et al. 1998), and with the age at which blindness occurs (Gougoux et al. 2004). Therefore, early support to

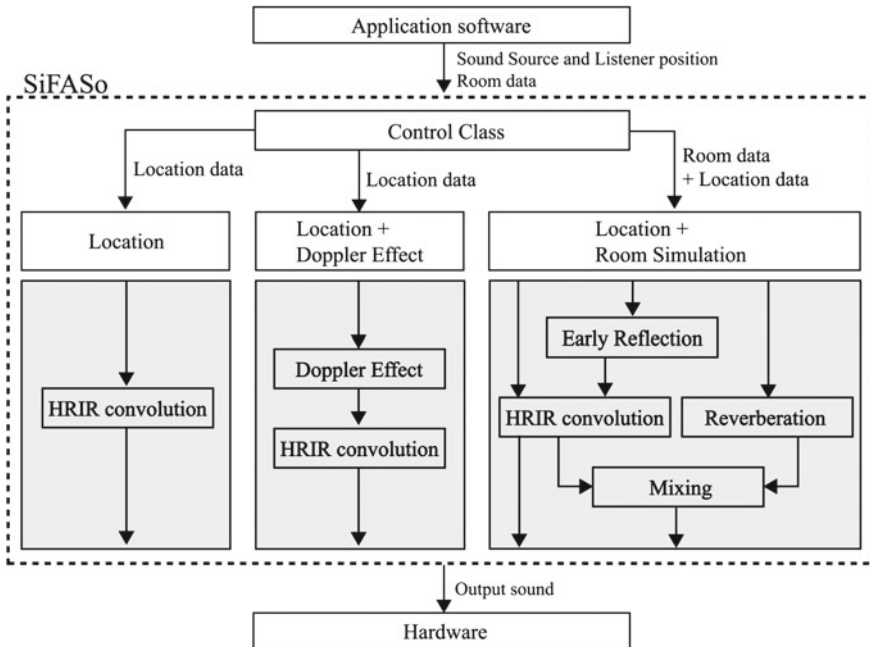


Fig. 9 Class diagram of the main part of SiFASO

improve auditory skills is extremely important, especially for people with low vision or late blindness (Afonso et al. 2005).

Three edutainment applications have been developed with SiFASO in the form of auditory virtual reality games (Iwaya et al. 2011), based on the experience of developing a similar but simpler edutainment application with dynamic binaural display using a DSP (Ohuchi et al. 2005). It is expected that they will not only be useful for training purposes but also for improving the *quality of life* of visually impaired people, who can scarcely enjoy TV games. The three edutainment applications, which are all played only by auditory information, are as follows.

- **BBBeat** An action-game-type application, where players knock out bees like a whack-a-mole game by locating their position based on a humming sound.
- **Mentalmapper** A maze-game-type application with a maze editor. Players navigate mazes rendered by spatial sounds to reach sounding landmarks assigned within the mazes.
- **SoundFormular** A racing-game-type application. Players drive vehicles and compete with a computer-controlled vehicle. Both vehicles and the motor course are rendered by spatial sounds.

All three applications were very well accepted by pupils of a municipal special-needs education school for the visually impaired. They found them to be great fun to play. Sighted pupils enjoyed SoundFormular less than BBBeat and Mentalmapper. The

courses, as rendered by SoundFormular, may have been too simple to elicit fun for those who are familiar with commercial driving games, featuring complex courses and competitors.

5 Cognitive-Map Forming in an Auditory-Maze Game

MentalMapper, an auditory-maze game (see above), was specifically developed for training and evaluating the performance of visually impaired people regarding formation of environmental cognitive maps. This section provides an outline of the game and reviews experimental results collected with it (Ohuchi et al. 2006).

5.1 *Outline of the MentalMapper*

The MentalMapper consists of two subsystems, namely, a maze editor and an auditory maze navigator. With the editor, mazes are drawn by connecting 1-m cube cells.⁴ For each cell, eight types of different absorption coefficients can be specified for walls, ceiling, and floor (e.g., concrete, wood, fully absorptive, and solid). Acoustic landmarks can be assigned to specified cells. These landmarks involve animal cries and environmental sounds from cars, railway crossings, etc.

With the navigator, users navigate through mazes rendered with virtual spatial sound. Navigation is performed with a game controller to move forward or backward (Fig. 10). Alternatively, users employ body rotation to turn. Further, verbal confirmations are given after each movement, such as “You have faced north.” Users hear footsteps when they move one cell forward or backward. Direct sounds and 1st-order reflections are rendered. Both auditory and tactile (vibrational) feedback are given when a user accidentally hits a wall.

5.2 *Experiment 1: Evaluation of Cognitive Maps Formed via Tactile Maps*

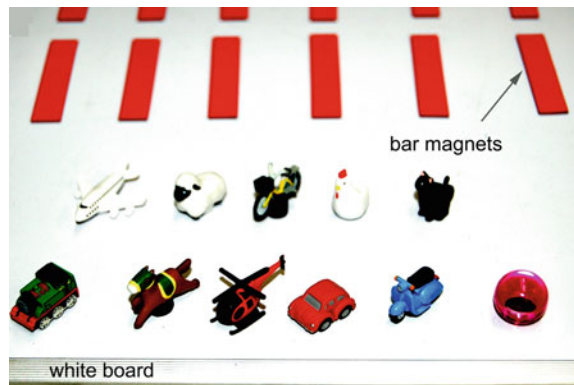
This experiment aims at examining the MentalMapper as an assistive technology for the formation of cognitive maps. Participants produce tactile maps after having navigated through virtual auditory sound mazes.

⁴This dimension was determined by technical limitations of SiFASo.

Fig. 10 Navigating a maze rendered with virtual spatial sound



Fig. 11 Whiteboard, bar magnets, and magnetic figures used in Experiment 1



Participants

Four congenitally blind adults (CMR, CSM, CTH, CKR) and four blindfolded sighted adults (SYN, SGK, SMS, SOM) participated in this experiment. Among them, one blind and all sighted participants were female.

Tactile Map

Soft and thin bar magnets were placed on a whiteboard on a desk for drawing tactile maps. Additionally, small magnetic figures (1–2 cm × 1–2 cm × 1–2 cm) were used to represent acoustic landmarks such as animals or cars in mazes (Fig. 11).

Pilot Experiment

Prior to Experiment 1, participants joined a pilot experiment. Participants were asked to navigate two auditory virtual mazes and then locate the landmarks on a blank

Fig. 12 An example of the mazes used in the pilot experiment. The cell with a star indicates the start position, and cells with circles denote landmarks

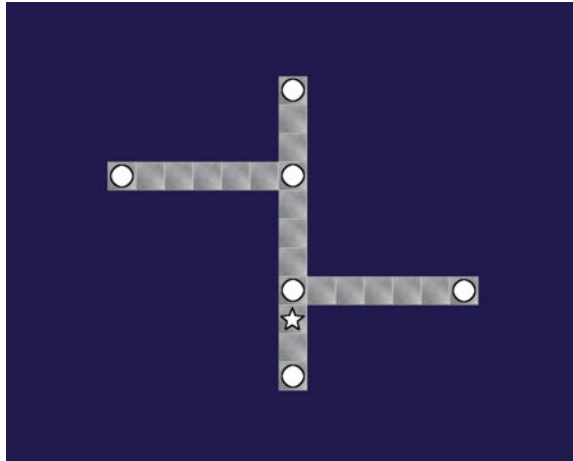


Fig. 13 Tactile-map drawing by a participant



tactile map. An example maze and a scene where a participant was drawing its map are respectively shown in Figs. 12 and 13. This was the very first experience for all of participants of moving through virtual auditory spaces by means of a game controller. While one sighted participant confused the locations of two landmarks, the others made no mistakes at all. Some of the congenitally blind participants reported that they had difficulties in creating a mental spatial image of the route, probably due to a lack of experience with such tasks.

Tasks

In Experiment 1, participants were first asked to navigate virtual auditory mazes with several landmarks and to then draw the map of the auditory maze as a tactile map. Figure 14 shows two mazes used in the task. The participants freely traversed the virtual auditory maze back and forth and then drew the tactile map. The time for the task was unrestricted. The landmark sounds were set to audible only when the user

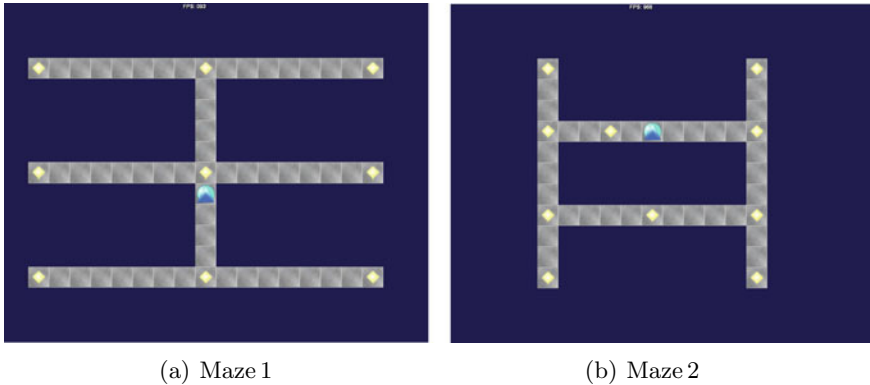


Fig. 14 Two mazes used in Experiment 1

entered the respective cell. The absorption coefficient of concrete was used for all cell boundaries.

Results and Consideration

The times required to complete the task are shown in Table 1. The blind participants spent 9' 53" and 24' 03" on average to complete Mazes 1 and 2, respectively, while the blindfolded sighted participants spent 15' 33" and 28' 01" on average. A two-way ANOVA was performed on the times to complete the task considering the group (blind or blindfolded sighted) and the maze (Maze 1 or 2) as factors. The results indicate no significant differences between blind and sighted participants ($F(1, 6) = 0.53$), although the blind group spent less time on average than the blindfolded sighted group for both mazes.

Moreover, both groups took significantly longer time on Maze 2 than on Maze 1 ($F(1, 6) = 11.61, p < 0.05$). This is probably due to the redundant structure of Route 2, namely that the passage of this maze causes some confusion because it forms a square walking path. Figure 15 shows examples of the maps drawn by congenitally blind and blindfolded sighted participants. All except CTH and SYN drew geometrically accurate maps. CTH and SYN could not draw complete maze shapes, but the geometry of the drawn parts was accurate. The drawn maps were then evaluated quantitatively by calculating bi-dimensional correlation coefficients between the shapes of the virtual mazes and digitized shapes drawn by participants (Tobler 1977). For the incomplete mazes of CTH and SYN, only the completed parts were analyzed. Table 2 shows the correlation coefficient for each participant on each route. All participants showed correlation coefficients greater than 0.85, suggesting that virtual auditory navigation of mazes is indeed effective in assisting formation of cognitive maps.

Table 1 Time required to complete task of Experiment 1

	Congenitally blind					Blindfolded sighted				
	CMR	CSM	CTH	CKR	Av.	SYN	SGK	SMS	SCM	Av.
Maze 1	7' 32"	15' 37"	9' 50"	6' 32"	9' 53"	18' 23"	20' 55"	10' 55"	11' 58"	15' 33"
Maze 2	12' 27"	45' 34"	24' 40"	13' 32"	24' 03"	23' 59"	47' 31"	25' 51"	14' 43"	28' 01"

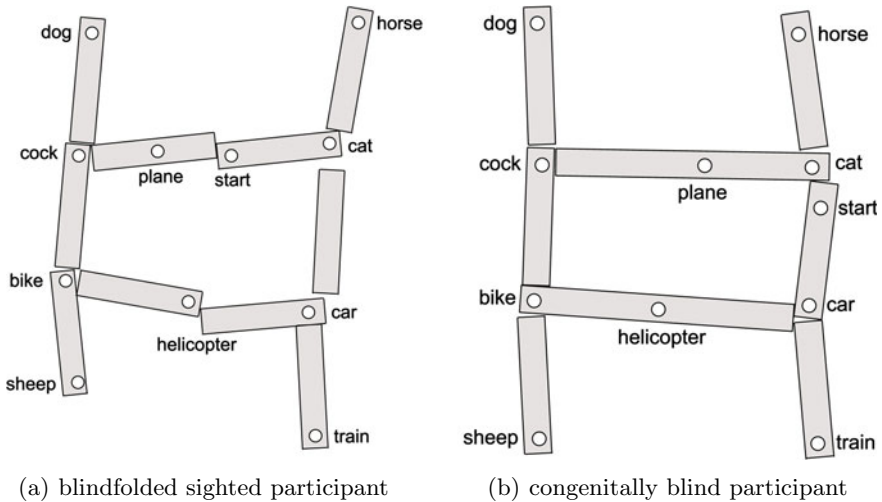


Fig. 15 Example of the maps drawn by two participants (Maze 2)

Table 2 Bidimensional correlation coefficient for each participant in each maze

	Congenitally blind				Blindfolded sighted			
	CMR	CSM	CTH	CKR	SYN	SGK	SMS	SCM
Maze 1	0.861	0.948	0.927	0.997	0.994	0.990	0.982	0.981
Maze 2	0.944	0.983	0.947	0.995	0.966	0.985	0.974	0.967

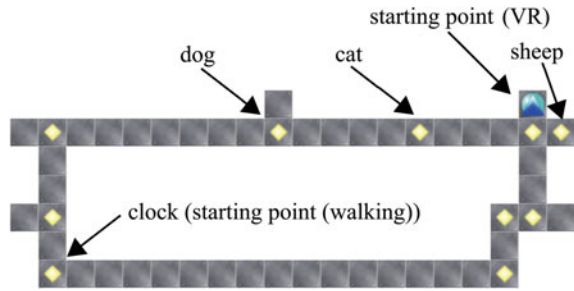
5.3 Experiment 2: Forming a Cognitive Map of an Actual Building

The results of Experiment 1 indicate that cognitive maps seem to be shaped correctly after navigation through auditory virtual environments. Experiment 2 examined whether training with auditory virtual maps is beneficial for navigating the real world.

Tasks

The maze used in this experiment was a replica of the corridor structure of an actual university building (Fig. 16). This experiment used the same four congenitally blind

Fig. 16 Visual representation of an auditory maze based on the floor structure of an actual corridor



adult subjects of Experiment 1. The sound assigned to landmarks was made audible only when the user was virtually present in the respective cell.

Participants first navigated freely through the auditory virtual maze up to 20 min to explore the maze. Then, immediately following navigation, they were asked to walk through the actual maze (i.e. a real corridor structure). To start, they were guided to the position of the landmark clock, which is different from the start position for the virtual navigation (Fig. 16). Then, the following four sequential tasks were assigned. During the tasks, for safety reasons, participants walked with a sighted guide and were asked to indicate vocally whenever changing their heading.

- **Task 2-1** Go to the landmark “dog.”
- **Task 2-2** Go to the landmark “sheep” without encountering the cat.
- **Task 2-3** Go to the landmark “dog,” taking the shortest route.
- **Task 2-4** Go back to the landmark “clock,” taking the shortest route.

Results and Discussion

This experiment aimed at examining whether participants were able to generate cognitive maps after navigating the virtual auditory maze. Task 2-2 requires participants to develop a good mental representation of the floor plan, including landscape locations. Therefore, if all these tasks are successfully completed, the cognitive maps can be regarded as well formed.

Table 3 shows the evaluation of Task 2-2 walking tasks by the experimenter for each participant. The rating scale is defined as follows:

- 4 Participants found the most direct route to the destination without wandering astray
- 3 Participants wandered but finally reached the destination
- 2 Participants reached the destination with some verbal assistance after wandering for some time or going off course
- 1 Participants could not find a way to the destination.

CSM, in Tasks 2-1 and 2-2, and CKR, in Task 2-1, wandered around before reaching the destinations. CTH became disoriented in Task 2-2. After asking for verbal assistance, CTH reached the destination and completed Tasks 2-3 and 2-4. CMR did not reach the correct destinations in Tasks 2-1 through 2-3. In Task 2-4, CMR reached

Table 3 Evaluation of Experiment 2

		Task			
		2-1	2-2	2-3	2-4
Participant	CSM	3	3	4	3
	CTH	4	2	4	4
	CKR	3	4	4	4
	CMR	1	1	1	2

the clock via an incorrect route. After the experiment, CSM and CKR were asked about their reasons for wandering back and forth in Task 2-1. CSM reported being confused because the starting point of the real environment (i.e. the clock) was not provided for the free navigation in the virtual environment. CKR reported that he walked from one end of the corridor to the other one to find out the map scale, which was not divulged to the participants.

Why could CMR not complete these experimental tasks after showing good performance in Experiment 1? A salient difference between these two experiments is the tactile maps. In Experiment 1, participants drew tactile maps after navigating through the mazes. This procedure might have reinforced the formation of cognitive mapping. If so, it suggests that the combination of navigation of virtual spaces and map-drawing is very effective in forming suitable and robust cognitive maps. Another reason could be differences in walking style. After the experiment, CMR reported difficulty in walking with a sighted guide, since she was used to walking with a guide dog in daily life. CMR was the only person having such a dog. This suggests a possible influence of the use of a guide dog on the acuity of spatial cognition in orientation and mobility, including the formation of cognitive maps. This observation raises interesting questions for future research.

Other participants mentioned that they became disoriented by physically turning right and left repeatedly in the virtual environment. This disorientation might be attributable to mental rotation in incomplete maps under formation. Typically, the physical experience while walking is key to the formation of cognitive maps (Herman et al. 1982). In contrast, repeated rotation without any real physical walking might have induced such confusion. Exploring optimal procedures for the formation of cognitive maps of virtual auditory environments is certainly a further interesting area for future research.

Overall, the experimental results show that blind participants are able to form dependable cognitive maps via virtual navigation of unfamiliar environments. Furthermore, the results show that the experience to navigate virtual auditory mazes can transfer to the ability of navigating real environments with similar geometries. These results mean that dynamic 3D auditory displays, including binaural ones, are an effective assistive tool to improve orientation and mobility of visually-impaired people by adequate training.

In this context, Seki et al. (2011) made important contributions. Among other things, they provide a training system with more realistic virtual worlds, namely, with authentic traffic sounds. These are important cues that give blind people better mobility. This system, the *Wide-Range Auditory Orientation Training System*, (WR-AOTS), is described in Seki (2016).

6 Transfer Effects from Playing the Auditory Action Game

As mentioned in Sect. 4, BBBeat is an action-game edutainment software for training sound-localization skills. In this section, transfer effects⁵ as a result of playing this auditory game are described and discussed (Honda et al. 2007, 2009).

6.1 Transfer Effects and Spatial-Hearing Training

Transfer effects are commonly observed for various motor- or verbal-learning tasks. For instance, previous studies reported transfer effects of playing visual-action video games (Castel et al. 2005; Fery and Ponsérre 2001; Green and Bavelier 2003). However, few studies have examined transfer effects of playing auditory-action games. BBBeat is an auditory action-game-type edutainment application resembling the “whack-a-mole.” The players virtually hear the hum of honeybees instead of seeing annoying moles. They are then prompted to localize the honeybee position and to hit it with a hammer as quickly and accurately as possible (Fig. 17). It has been observed that players move their heads frequently to detect the position of the hum. When hitting a bee, vibration feedback is given, and another honeybee is spawned. Honda et al. examined the various transfer effects from playing BBBeat using pre- and post-test performance results of blindfolded individuals. In the experiments, participants were separated into two groups, maintaining the same proportion of males to females. Participants of the training group were asked to play the game for seven days (30 min per day) within a two-week period. In contrast, the control group did not play the game at all during this period.

Based on the results of this experiment, transfer effect with regard to sound-localization performance for real sound sources were examined (Honda et al. 2007), and a follow-up test was conducted to investigate the persistence of the transfer effects. The task was to identify a sound source among 36 loudspeakers distributed around the listener. Results revealed that the hit rate of the training group increased by approximately 20%, which is around twice that of the control group (statistically significant). Interestingly, a follow-up test, which was conducted one month later, showed that transfer effects persisted.

⁵**Transfer effect** This is defined as the ability to extend what has been learned in one context to new contexts. This is also called “transfer of learning.”

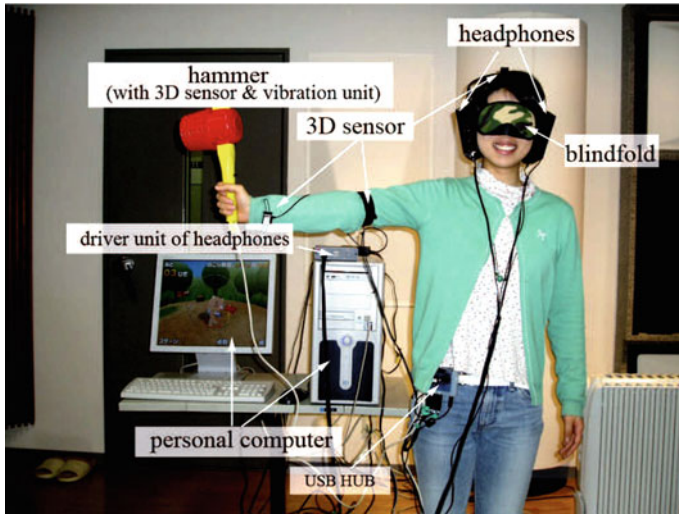


Fig. 17 Participant putting on attachments to play BBBeat

Readers may think that the above results are quite intuitive, since sound-localization performance was improved by the previous sound-localization training with BBBeat. However, the findings indicate a further interesting aspect, that the results show clear effects in the real world even though they were cultivated by training in a virtual environment. It must be admitted, however, that this transfer effect occurred between very similar tasks. Therefore, even better examples of transfer effects regarding skills that are useful in daily life are described in the following subsections (Honda et al. 2009).

6.2 Transfer Effects on Face Contacts

Normally sighted people devote attention to nonverbal information in interpersonal communication. For example, eye contact in face-to-face situations plays a regulatory function in everyday conversation (Kendon 1967). Eye contact is a relevant critical component of rewarding social exchange for sighted people (Ellsworth and Ludwig 1972). In contrast, visually impaired people use more non-visual cues for social interaction (Fichten et al. 1991). The difference in communication cues affects the impressions of visually impaired people.

Several researchers attempted to find effective training methods for the communication skills of visually impaired people (Erin et al. 1991; Sanders and Goldberg 1977; Raver 1987). For example, Sanders and Goldberg (1977) proposed a training program using auditory feedback for correct eye/face contacts to increase the rate of eye contact. These findings suggest that the communication skills could be enhanced

by training sound-localization skills. They reported that the eye contacts of clients, consisting of almost totally blind men, increased to over 80%, and the effect remained at the 74% level after 10 months.

These findings suggest that communication skills can be improved by training sound-localization skills. Therefore, Honda et al. (2009) examined whether sound-localization training is transferred to the rate of face contacts.

Experimental Conditions

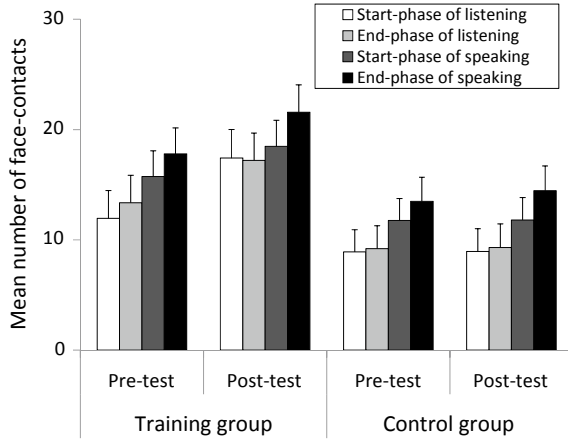
In this experiment, all blindfolded participants conducted the task in a soundproof room. Participants sat on a chair allowing free head and body movement. They were asked to discuss some topics with two interviewers (pre-test). Two trained female experimenters assisted as interviewers to ensure uniformity of tasks. Two video cameras recorded the scenes from the interviewers' positions. The interviewers provided scripted questions for the participants. Each interviewer was asked to confirm whether the participants showed face-contact behaviors on each topic with four phases, namely, (i) the start-phase of listening, (ii) the end-phase of listening, (iii) the start-phase of speaking, and (iv) the end-phase of speaking. Another interviewer confirmed whether the participants showed face-contact behavior to the querying interviewer during question-and-answer communication. Reliability obtained using the corresponding rate between the interviewers was 83%. All participants were asked to perform the same task again two weeks later (post-test). In the post-test, several topics were altered and the position of the interviewers was exchanged.

Results and Discussion

Figure 18 shows the results of the experiment. A three-way ANOVA was performed on the number of face-contacts in the communication task, considering the group (training or control), the test phase (pre-test or post-test), and the interview phase (start phase of listening, end phase of listening, start phase of speaking, or end phase of speaking) as factors. Results indicate that interaction between the group and the test phase is significant ($F(1, 37) = 5.71, p < 0.05$). The interaction can be observed in the results shown in Fig. 18. Post-hoc analysis (Ryan method, $p < 0.05$) reveals that the face-contact of the training condition ($m = 14.71$) increased significantly after playing the bee-hitting virtual auditory game ($m = 18.67, p < 0.01$). Additionally, the training group in the post-test showed more face-contact than the control group in the post-test ($m = 11.13, p < 0.05$). However, interaction between the group and the interview phase and interaction between the test and the interview phase are not significant. Furthermore, no significant three-way interaction was found.

These results indicate that, by playing the bee-hitting virtual auditory game rendered by a dynamic binaural display, face-contacts in social interaction increased significantly. This indicates that skills acquired while playing the auditory virtual game transferred to participants' communication skills during social interaction.

Fig. 18 Number of face contacts in the communication task for the training group in comparison to those of the control group



6.3 Transfer Effects on Collision-Avoidance Behavior

In order to avoid looming objects, the generation of an appropriate response includes five tasks, that is, (i), detection of a looming stimulus, (ii), localization of the stimulus position, (iii), computation of the direction of the stimulus movement, (iv), determination of an escape direction and, (v), selection of proper motor actions (Liaw and Arbib 1993). Previous studies revealed that visual information tends to be used more efficiently than auditory information regarding accuracy of estimating time-to-arrival (Schiff and Oldak 1990). However, visually impaired people are obviously restricted in their use of visual cues. Consequently, it is very important for them to formulate and execute avoidance behaviors using acoustical information for the correct location of approaching objects (i.e. perceived sound sources). Furthermore, when an object is on a collision course toward persons, they must move aside with minimal distance from their own position to that of the object, because avoidance with greater distances might cause another collision with surrounding obstacles. Therefore, when visually impaired people try to conduct appropriate avoidance behaviors, it is crucial that they perceive sound-source positions accurately. Appropriate avoidance behavior thus relies on good sound-localization skills. Consequently, Honda et al. (2009) was interested in whether sound-localization training with virtual auditory games are transferred to avoidance behaviors in response to approaching auditory objects.

Experimental Conditions

Figure 19 illustrates the collision-avoidance task. In this scene, all blindfolded participants were asked to avoid an approaching object when they felt that it was moving on a collision course (relevant path). Furthermore, they were asked to perform avoidance maneuver with minimal displacement from their position. They were further instructed not to avoid an approaching object when they felt it was moving on an irrelevant path. The distance between the relevant path and the two irrelevant paths was 80 cm. The colliding object was a toy car (width: 30 cm, weight: 2.5 kg).

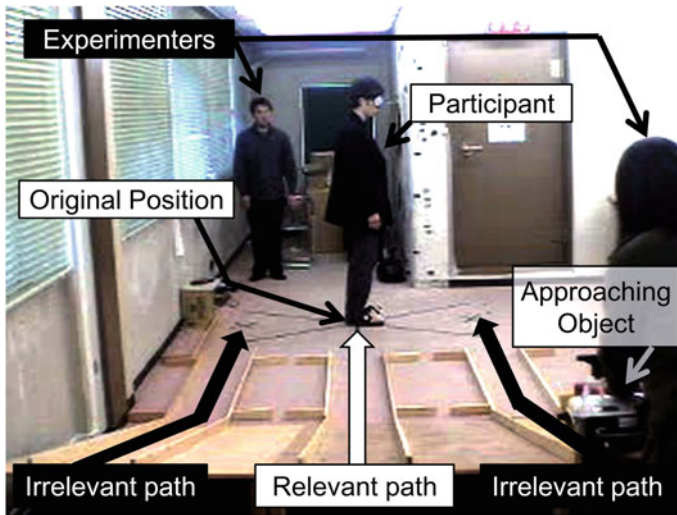


Fig. 19 Illustration of a collision-avoidance task

The approaching stimulus was presented randomly and repeatedly to participants from either relevant or irrelevant paths. The A-weighted sound pressure level of the approaching sound was 75 dB, and the background noise level in the room was 35 dB. The toy car was placed at 50 cm height on lanes (initial velocity: 0 m/s) and slid along the lane slopes at 2 m/s. Three lanes were used. The center lane was used for the relevant (i.e. collision) path, and lanes of both sides were for irrelevant paths. The distance between the participants and the lanes was 4.0 m. The task of the participants was to localize the approaching object solely based on auditory cues and to decide their behaviors within 2 s. The trials numbered 36 in all. The approaching object was sent on one of the three lanes, selected randomly, 12 times for each course. The body direction of the participants was changed for each trial. Consequently, the toy car approached from either front, back, left, or right. The experimenter then checked whether participants had completed the avoidance behaviors for each trial. Additionally, the distances from the participants' start position to the end point of their actions was measured. All participants were asked to perform the same task again two weeks later (post-test).

Results and Discussion

Figure 20 shows the results of the experiment. A three-way ANOVA was performed on the mean avoidance distances from the original position for the object approaching from irrelevant paths considering the group (training or control), the test phase (pre-test or post-test), and the direction to the approaching object (front, back, left, or right) as factors. Results show that a two-way (group \times test) interaction is significant ($F(1, 25) = 6.93, p < 0.05$). Post hoc analysis (Ryan method, $p < 0.05$) reveals

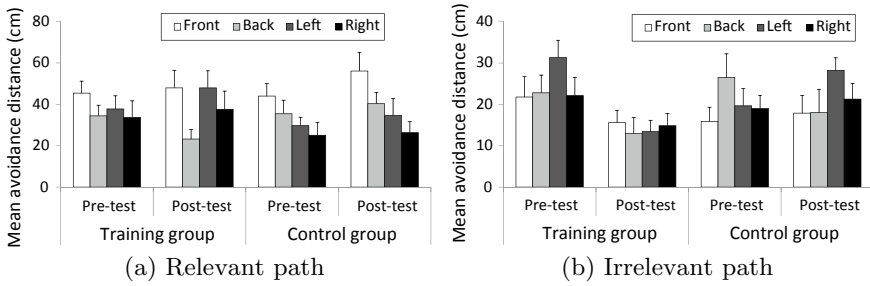


Fig. 20 Mean avoidance distances in the avoidance task

that avoidance distances of the training group ($m = 24.51$) decreased significantly by playing the bee-hitting virtual auditory game, BBBeat ($m = 14.19, p < 0.01$).

The results indicate that the avoidance distance after sound-localization training in the virtual auditory game decreased for objects approaching from the irrelevant path.

These findings indicate that auditory training using the bee-hitting virtual auditory game with binaural display modified the detailed manners of executed avoidance behaviors, which relate to sound-localization skills.

7 Concluding Remarks

In this chapter some interesting aspects of human active listening have been described. The question raised is, how can active listening be defined when considering these aspects? Listener movements induce dynamic ear inputs. As an operational definition, the following is proposed:

Active listening is a mode of multisensory spatial hearing that takes advantage of dynamic information induced by listeners' movements, irrespective of being intentional, conscious, or unconscious.

In fact, as shown in Sect. 2, even unconscious dynamic change of ear inputs may significantly change listener experience during spatial hearing.

While evidence has accumulated confirming that active listening facilitates sound-localization performances, a few recent studies have revealed that it may suppress sound-localization accuracies (see Sects. 1 and 3). However, these examples should not be regarded as an inconsistency, but rather an indication of the diversity of the roles of active listening.

Facilitation of sound-localization performance seems to occur when overall sound localization is requested after presentation of sound stimuli, and the listeners have terminated their direction-finding movements, whereas deterioration is observed when instantaneous sound-image positions are reported while listeners are moving during the sound presentation. In other words, the former would resolve ambiguities in the ill-posed problem caused by the scarcity of acoustic input to only two channels (i.e. the two ears), while the latter helps stabilize the perceptual auditory space during variation of the ear-inputs signals. This is in a way similar to *saccadic suppression* in vision. Knowledge of active listening and the psychoacoustic effects going with it are key to advancing binaural technologies. For instance, knowledge of how sound localization in dynamic scenarios can be enhanced can be directly applied to performance enhancement of, for example, dynamic 3D auditory displays, including binaural ones. In this context, knowledge of suppression and masking effects supports the economic use of computational resources during listener movements.

In Sects. 5 and 6, two edutainment applications of dynamic binaural displays were introduced. These are applied to support active listening for the training of auditory spatial-perception acuity, particularly of visually impaired people. One of the applications is a maze game and the other is an action game. The maze game facilitates users' ability to draw cognitive maps, as well as the evaluation of this capability. Moreover, a transfer effect was found to the navigation of real environments having a similar geometry as the virtually experienced maze. Playing the action game improved players' sound localization performances. Again, the experience at virtual sound localization in playing the action game transferred to improve players' sound-localization performances of real sound sources. Moreover, clear transfer effects to skills useful in daily life were observed, including increased eye-contact frequency during conversation and improved ability to avoid an approaching object. These results indicate good potential for application of dynamic binaural displays to improve spatial-hearing abilities and, hopefully, other skills that have the potential to enhance quality of life.

In summary, active listening plays an important role in making human spatial hearing more reliable and richer. Binaural technologies that support active listening are key to high-definition communication. Dynamic binaural displays that support active listening are universally applicable to enhance quality of experience in virtual and real auditory dynamic scenes.

Acknowledgements Results introduced in this chapter have been partly supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grant-in-Aid for Specially Promoted Research (19001004), the Society for the Promotion of Science (JSPS) *Kakenhi* Grant-in-Aid for Scientific Research (A) (24240016, 16H01736), and the Consortium R & D Projects for Regional Revitalization of the Ministry of Economy Trade and Industry from 2003 to 2004. The authors would like to thank M. Cohen for his diligent proofreading of earlier versions of the manuscript, which greatly improved the readability of this chapter. Thanks are further due to two anonymous reviewers for their very constructive comments and advice.

References

- Afonso, A., B.F. Katz, A. Blum, C. Jacquemin, and M. Denis. 2005. A study of spatial cognition in an immersive virtual audio environment: Comparing blind and blindfolded individuals. In *Proceedings of International Conference on Auditory Display*.
- Begault, D.R., E.M. Wenzel, M. Godfroy, J.D. Miller, and M.R. Anderson. 2010. Applying spatial audio to human interfaces: 25 years of NASA experience. In *40th International Audio Engineering Society Conference*.
- Berkhout, A.J., D. de Vries, and P. Vogel. 1993. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America* 93: 2764–2778.
- Blauert, J. 1997. *Spatial Hearing*. Cambridge: The MIT Press. ISBN 0-262-02413-6.
- Blauert, J., H. Lehnert, J. Sahrhage, and H. Strauss. 2000. An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation. *Acta Acustica United with Acustica* 86 (1): 94–102.
- Brimijoin, W.O., A.W. Boyd, and M.A. Akeroyd. 2013. The contribution of head movement to the externalization and internalization of sounds. *PLoS ONE* 8: e83068.
- Castel, A.D., J. Pratt, and E. Drummond. 2005. The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychologica* 119 (2): 217–230.
- Cooper, J., S. Carlile, and D. Alasis. 2008. Distortions of auditory space during rapid head turns. *Experimental Brain Research* 191: 209–219.
- Ellsworth, P.C., and L.M. Ludwig. 1972. Visual behavior in social interaction. *Journal of Communication* 22 (4): 375–403.
- Enomoto, S., and S. Ise. 2005. A proposal of the directional speaker system based on the boundary surface control principle. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 88: 1–9.
- Erin, J.N., K. Dignan, and P.A. Brown. 1991. Are social skills teachable? A review of the literature. *Journal of Visual Impairment and Blindness* 85: 58–61.
- Fery, Y.-A., and S. Ponsere. 2001. Enhancing the control of force in putting by video game training. *Acta Psychologica* 44 (12): 1025–1037.
- Fichten, C.S., D. Judd, V. Tagalakakis, and R. Amsel. 1991. Communication cues used by people with and without visual impairments in daily conversations and dating. *Journal of Visual Impairment and Blindness* 85: 371–378.
- Gougoux, F., F. Lepore, M. Lassonde, P. Voss, R.J. Zatorre, and P. Belin. 2004. Neuropsychology: Pitch discrimination in the early blind. *Nature* 430: 309.
- Green, S.C., and D. Bavelier. 2003. Action video game modifies visual selective attention. *Nature* 423: 534–537.
- Herman, J.F., R.G. Kolker, and M.L. Shaw. 1982. Effects of motor activity on children's intentional and incidental memory for spatial locations. *Child Development* 53 (1): 239–244.
- Hirahara, T., Y. Sawada, D. Morikawa. 2011. Impact of dynamic binaural signals on three-dimensional sound reproduction. In *Proceedings of Inter-Noise 2011*.
- Hirahara, T., D. Yoshisaki, and D. Morikawa. 2013. Impact of dynamic binaural signal associated with listener's voluntary movement in auditory spatial perception. In *Proceedings of Meetings on Acoustics*, Vol. 19.
- Honda, A., H. Shibata, J. Gyoba, K. Saito, Y. Iwaya, and Y. Suzuki. 2007. Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Applied Acoustics* 68: 885–896.
- Honda, A., H. Shibata, J. Gyoba, Y. Iwaya, and Y. Suzuki. 2009. Transfer effects on communication and collision avoidance behavior from playing a three-dimensional auditory game based on a virtual auditory display. *Applied Acoustics* 70: 868–874.
- Honda, A., H. Shibata, S. Hidaka, J. Gyoba, Y. Iwaya, and Y. Suzuki. 2013. Effects of head movement and proprioceptive feedback in training of sound localization. *i-Perception* 4: 253–264.

- Honda, A., K. Ohba, Y. Iwaya, and Y. Suzuki. 2016. Detection of sound image movement during horizontal head rotation. *iPerception* 7: 2041669516669614.
- Ise, S. 1997. A principle of active control of sound based on the Kirchhoff-Helmholtz integral equation and the inverse system theory. *Journal of the Acoustical Society of Japan* 53: 706–713.
- Iwaya, Y., and Y. Suzuki. 2007. Rendering moving sound with the doppler effect in sound space. *Applied Acoustics* 68 (8): 916–922.
- Iwaya, Y., Y. Suzuki, and D. Kimura. 2002. The effects of head movement on sound localization with real and virtual sound sources. In *Proceedings of China-Japan Joint Conference on Acoustics (JCA2002)*.
- Iwaya, Y., Y. Suzuki, and D. Kimura. 2003. Effects of head movement on front-back error in sound localization. *Acoustical Science and Technology* 24: 322–324.
- Iwaya, Y., M. Toyoda, and Y. Suzuki. 2005. A new rendering method of moving sound with the doppler effect.
- Iwaya, Y., M. Otani, and Y. Suzuki. 2009. Development of virtual auditory display software responsive to head movement and a consideration on deration of spatialized ambient sound to improve realism of perceived sound space. In *Proceedings of International Workshop on Principles and Applications of Spatial Hearing (IWPASH)*.
- Iwaya, Y., M. Otani, and Y. Suzuki. 2011. Development of virtual auditory display software responsive to head movement and a consideration on deration of spatialized ambient sound to improve realism of perceived sound space. In *Principles and Applications of Spatial Hearing*, ed. Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato. Singapore: World Scientific.
- Kawaura, J., Y. Suzuki, F. Asano, and T. Sone. 1989. Sound localization in headphone reproduction by simulating transfer function from the sound source to the external ear (in Japanese). *Journal of the Acoustical Society of Japan* 45: 755–766.
- Kawaura, J., Y. Suzuki, F. Asano, and T. Sone. 1991. Sound localization in headphone reproduction by simulating transfer function from the sound source to the external ear (English translation). *Journal of the Acoustical Society of Japan (E)* 12: 203–216.
- Kendon, A. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26: 22–63.
- Leung, J., D. Alais, and S. Carlile. 2008. Compression of auditory space during rapid head turns. *Proceedings of the National Academy of Sciences of the United States of America* 105: 6492–6497.
- Liauw, J.-S., and M.A. Arbib. 1993. Neural mechanisms underlying direction-selective avoidance behavior. *Adaptive Behavior* 1 (3): 227–261.
- Lokki, T., and H. Järveläinen. 2001. Subjective evaluation of auralization of physics-based room acoustics modeling.
- Miller, J.D. 2001. Slab: A software-based real-time virtual acoustic environment rendering system. In *Proceedings of International Conference on Auditory Display*.
- Mills, W.A. 1958. On the minimum audible angle. *Journal of the Acoustical Society of America* 30: 237–246.
- Morimoto, M., and Y. Ando. 1980. On the simulation of sound localization. *Journal of the Acoustical Society of Japan (e)* 1 (3): 167–174.
- Ohuchi, M., Y. Iwaya, Y. Suzuki, and T. MuneKata. 2005. Training effect of a virtual auditory game on sound localization ability of the visually impaired.
- Ohuchi, M., Y. Iwaya, Y. Suzuki, and T. MuneKata. 2006. Cognitive-map forming of the blind in virtual sound environment In *Proceedings of International Conference on Auditory Display*.
- Paré, N.L.M., F. Lepore, and M. Lassonde. 1998. Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395: 278–280.
- Perrett, S., and W. Noble. 1997. The contribution of head motion cues to localization of low-pass noise. *Perception & Psychophysics* 59: 1018–1026.
- Picinali, L., A. Afonso, M. Denis, and B.F. Katz. 2014. Exploration of architectural spaces by the blind using virtual auditory reality for the construction of spatial knowledge. *International Journal of Human-Computer Studies* 72: 393–407.

- Poletti, M.A. 2005. Three-dimensional surround sound systems based on spherical harmonics. *Journal of the Audio Engineering Society* 53: 1004–1025.
- Raver, S.A. 1987. Training blind children to employ appropriate gaze direction and sitting behavior during conversation. *Education and Treatment of Children* 10 (3): 237–246.
- Sanders, R.M., and S.G. Goldberg. 1977. Eye contacts: Increasing their rate in social interactions. *Journal of Visual Impairment and Blindness* 71: 265–267.
- Savioja, L., J. Huopaniemi, T. Lokki, and R. Väänänen. 1999. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society* 47: 675–705.
- Schiff, W., and R. Oldak. 1990. Accuracy of judging time to arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance* 16 (2): 303–316.
- Seki, Y. 2016. Wide-range auditory orientation training system. <https://staff.aist.go.jp/yoshikazu-seki/AOTS/WR-AOTS/index.html> (last accessed December 21, 2019).
- Seki, Y., Y. Iwaya, T. Chiba, S. Yairi, M. Otani, M. Ohuchi, T. Munekata, K. Mitobe, and A. Honda. 2011. Auditory orientation training system developed for blind people using PC-based wide-range 3-d sound technology. In *Principles and Applications of Spatial Hearing*, ed. by Y. Suzuki, D. Brungart, Y. Iwaya, K. Iida, D. Cabrera, and H. Kato. Singapore: World Scientific. ISBN: 978-981-4465-41-0.
- Suzuki, Y., S. Takane, S. Takahashi, and T. Miyajima. 2002. A preliminary development of high definition virtual acoustic display based on advise.
- Suzuki, Y., T. Okamoto, J. Treviño, Z. Cui, Y. Iwaya, S. Sakamoto, and M. Otani. 2012. 3d spatial sound systems compatible with human's active listening to realize rich high-level Kansei information. *Interdisciplinary Information Sciences* 18: 71–82.
- Takane, S., T. Miyajima, Y. Yamada, D. Arai, Y. Suzuki, and T. Sone. 1997. An auditory display based on virtual sphere model. In *Proceedings of International Symposium on Simulation Visualization and Auralization for Acoustic Research and Education*.
- Thurlow, W.R., and P.S. Runge. 1967. Effect of induced head movements on localization of direction of sound. *Journal of the Acoustical Society of America* 42: 480–488.
- Tobler, W. 1977. Bidimensional regression: A computer program. https://www.geog.ucsb.edu/~tobler/publications/pdf_docs/Bidimensional-Regression.pdf (last accessed December, 2019).
- Toshima, I., and S. Aoki. 2009. Sound localization during head movement using an acoustical telepresence robot: Telehead. *Advanced Robotics* 23: 289–304.
- Toshima, I., H. Uematsu, and T. Hirahara. 2003. A steerable dummy head that tracks three-dimensional head movement: Telehead. *Acoustical Science and Technology* 24: 327–329.
- Wallach, H. 1939. On sound localization. *Journal of the Acoustical Society of America* 10: 270–274.
- Wenzel, E.M., J.D. Miller, and J.S. Abel. 2000. Sound lab: A real-time, software-based system for the study of spatial hearing. In *Proceedings of AES 108th Convention*, p. Preprint:5140.
- Yairi, S., Y. Iwaya, and Y. Suzuki. 2006. Investigation of system latency detection threshold of virtual auditory display. In *Proc. of ICAD 2006—12th Meeting of the International Conference on Auditory Display*, 217–222, London.
- Yairi, S., Y. Iwaya, and Y. Suzuki. 2007. Estimation of detection threshold of system latency of virtual auditory display. *Applied Acoustics* 68: 851–863.
- Yairi, S., Y. Iwaya, and Y. Suzuki. 2008a. Individualization feature of head-related transfer functions based on subjective evaluation. In *Proc. of International Conference on Auditory Display (ICAD2008)*, Paris. 2008.
- Yairi, S., Y. Iwaya, and Y. Suzuki. 2008b. Influence of large system latency of virtual auditory display on behavior of head movement in sound localization task. *Acta Acustica united with Acustica* 94: 1016–1023.
- Zhang, C., and B. Xie. 2013. Platform for dynamic virtual auditory environment real-time rendering system. *Chinese Science Bulletin* 58 (3): 316–327.