

# Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation

György Wersényi

**Abstract**—Virtual audio simulation uses head-related transfer function (HRTF) synthesis and headphone playback to create a sound field similar to real-life environments. Localization performance is influenced by parameters such as the recording method and the spatial resolution of the HRTFs, equalization of the measurement chain as well as common headphone playback errors. The most important errors are in-the-head localization and front-back reversals. Among other cues, small movements of the head are considered to be important to avoid these phenomena. This study uses the BEACHTRON sound card and its HRTFs for emulating small head-movements by randomly moving the virtual sound source to emulate head-movements. This method does not need any additional equipment, sensors, or feedback. Fifty untrained subjects participated in the listening tests using different stimuli and presentation speed. A virtual target source was rendered in front of the listener by random movements of  $1^\circ$ – $7^\circ$ . Experiments showed that this kind of simulation can be helpful to resolve in-the-head localization, but there is no clear benefit for resolving front-back errors. Emulation of small head-movements of  $2^\circ$  could actually increase externalization rates in about 21% of the subjects while presentation speed is not significant.

**Index Terms**—Auditory display, head-tracking, head-related transfer function (HRTF) synthesis, virtual localization.

## I. INTRODUCTION

THE former Graphical User Interface for Blind Persons Project (GUIB) was focused on creating a virtual audio display (VAD) for the elderly and the visually disabled [1], [2]. These individuals do not have the possibility to use graphical user interfaces and they need special tools if they want to use personal computers. This project included a number of experiments, such as finding the proper mapping between icons or events on the screen and sound samples (called Earcons), possibilities of different input media (touch screens, Braille keyboards), and evaluation of different playback systems [3]–[5]. First, a multichannel loudspeaker array was tested and was found to be inappropriate. Subsequently, a headphone-based approach employing head-related transfer function (HRTF) filtering was used. Both methods used the BEACHTRON sound card to render the spatial sound. Although the GUIB project ended years ago, some psychoacoustic measurements have been made with this system. Those investigations focused, e.g.,

Manuscript received May 05, 2008; revised September 08, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Malcolm Slaney.

Gy. Wersényi is with the Department of Telecommunications, Széchenyi István University, H-9026, Győr, Egyetem tér 1., Hungary (e-mail: wersenyi@sze.hu).

Digital Object Identifier 10.1109/TASL.2008.2006720

on headphone playback errors, localization blur, and spatial resolution of the VAD.

## II. HEAD-TRACKING AND VIRTUAL LOCALIZATION

The purpose of our current investigation is to find tools to improve the localization performance with the system mentioned above. In a parallel investigation, we tested additional high-pass and low-pass filtering of sound sources to bias correct localization judgments in the median plane [6].

One of the main goals of this study is to decrease front-back reversals or in-the-head localization rates. During headphone playback, these errors influence the localization [7]–[14]. State-of-the-art multimedia virtual simulators use head-tracking devices, simulation of room reverberation, and different methods to create the best-fitting HRTF set [15]–[19]. Head-tracking has been shown to be important for reducing such errors [20]–[23]. Furthermore, *small* head-movements (often unwanted) of about  $1^\circ$ – $3^\circ$  could influence in-the-head localization in free-field listening through small changes in the interaural differences. We conjecture that such small head-movements could reduce in-the-head localization and the small changes in the interaural level and time differences may lead to better results. Dynamic changes introduced by small movements of the head, whether intentional or unintentional, can be relevant [14], [24], [25]. The idea is to add some kind of “jitter” to the interaural differences through the applied HRTFs to simulate the effects of micro head-movements.

State-of-the-art methods use headphones with a head-tracking device. Such a device has some sort of feedback and additional hardware (e.g., laser pointer and receivers, magnetic sensors, etc.), and typically requires considerable computational resources. In such systems, it is possible to change the HRTFs dynamically, to create a correct spatial event, and compute the appropriate HRTFs synchronized to the listener’s head-movements.

By contrast, our system is built on different methods. Instead of moving the head, using feedback and additional equipment, we simply simulate these movements by moving the virtual sound source. This is achieved by small changes in the HRTFs that are not synchronized with the actual position of the head. The idea behind this is that motion is relative: turning the head one degree to the left in front of a steady source corresponds to a steady listener and moving the sound source one degree to the right. However, motion of the source does not necessarily substitute for voluntary motion of the listener if the motion is large. In this manner, small head-movements can be replaced by the movement of the sound source. This virtual movement of the source is achieved through HRTF synthesis only. The goal of

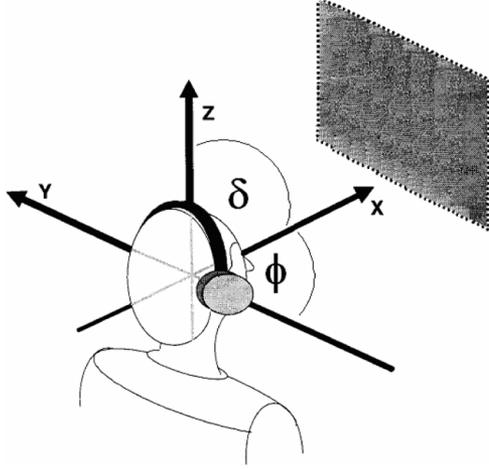


Fig. 1. Illustration of the 2-D VAD. The acoustic surface is parallel to the  $Z$ - $Y$  plane. The origin (the reference location) is in the front of the listener ( $\varphi = \delta = 0^\circ$ ).

the investigation is to explore whether source motion can replace additional hardware and head-tracking devices. To simulate only small movements of the head, and to investigate the influence on in-the-head localization and reversal rates, changes in the HRTFs about  $1^\circ$ – $4^\circ$  are applied.

### III. MEASUREMENT SETUP

The virtual audio display is simulated in front of the listener as a 2-D sound screen as seen in Fig. 1. The BEACHTRON system uses the HRTFs of a “good localizer” from measurements by Wightman and Kistler [26]–[30]. According to the user’s manual, the simulation method of the BEACHTRON card includes a distance model (for simulating atmospheric loss) but no room models or any other additional signal processing on the HRTF set [26]. Real-time filtering by the head-related impulse responses (HRIRs) and equalization for the Sennheiser HD540 headphone is made in the time-domain by FIR filters of length 75 taps (per ear). Furthermore, it is possible to set the head-diameter to obtain a better interaural time difference simulation.

The investigation was made in an anechoic chamber. Fifty untrained male university students, all with normal hearing (tested in a standardized audiometric scan) between 19 and 22 years participated.

The program sets the virtual sound source in the “front” direction ( $\varphi = \delta = 0^\circ$ ), which we considered as the target source location. Without rendering, this is a stationary source, a reference condition. During rendering, the sound source is moved randomly by changing the values of the following three parameters, denoted by  $A$ ,  $B$ , and  $C$ . Here  $A$  specifies the direction and extent of the movement both horizontal and vertical ( $0^\circ$  to  $10^\circ$ ; the case  $A = 0$  corresponds to no movement, and creates the reference condition of a stationary source). The parameter  $B$  specifies the number of times the source location is changed (1 to 100). The parameter  $C$  specifies the number of times the stimulus is presented in one location (1 to 1000).

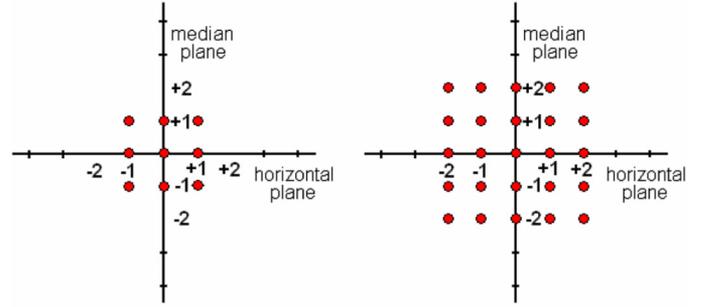


Fig. 2. Representation of parameters  $A = 1$  (left) and  $A = 2$  (right). The total number of rendered source positions is  $(2A + 1)^2$ .

TABLE I  
VALUES FOR THE PARAMETERS  $B$  AND  $C$  (USED FOR BASIC SIGNAL PRESENTATION CONDITIONS) AS A FUNCTION OF PRESENTATION SPEED

	Presentation speed (approx.)	$B$	$C$
Slow	1 Hz	5	1000
Avg.	2 Hz	20	100
Fast	5 Hz	100	50

After setting these parameters, a white noise signal of 10 ms was rendered. The total duration  $T$  of the stimulus was

$$T = B * C * 10 \text{ (ms)}. \quad (1)$$

For example, by setting  $A = 2$ ,  $B = 50$ , and  $C = 5$ , the following presentation could be made. A random generator calculates an actual source location within  $\pm 2^\circ$  around the origin that includes  $-2^\circ$ ,  $-1^\circ$ ,  $0^\circ$ ,  $+1^\circ$ ,  $+2^\circ$  in all directions (see Fig. 2). These dots represent potential source locations. With  $B = 50$ , 50 actual source locations will be determined and in each location the sound file will be presented five times (50 ms). Because the potential number of different locations is only 25,  $B = 50$  means that all of them will be selected twice in a random order. For small  $A$ 's, each potential source location is used several times. As a consequence, when the number of potential source locations is higher than  $B$ , only part of all possible source locations could be used. The actual length of the stimulus is not crucial, because it is a consequence of parameter  $C$ . Using a 10-ms sound file and a number of 50 for parameter  $C$  is the same as using a 50-ms sound file and a value of 10 for parameter  $C$ . By reducing the number of  $C$  and increasing the number of  $B$ , we can simulate faster head-movements.

During rendering, subjects are asked to report 1) whether the perceived location is in the head, 2) the front-back reversals, and 3) whether they experience the percept of a stationary or a moving source (perception of movement). This latter question is a control, because our goal is to simulate a sound source that *appears to be steady*, and thus we would like the subjects not to detect any movement. In other words, we want to confine the location of the virtual source within the limits of localization blur. We assumed that about  $1^\circ$ – $3^\circ$  of random movement will be perceived as a stationary source.

At the start of the experiment, all subjects were exposed to the reference condition where  $A = 0$ , corresponding to a stationary

TABLE II  
SUMMARIZED RESULTS FOR 50 SUBJECTS USING WHITE NOISE EXCITATION. CLASSIFICATIONS FOR IN-THE-HEAD LOCALIZATION (IHL) AND FRONT-BACK REVERSALS (FBR) ARE SHOWN FOR VARIOUS SPEEDS OF PRESENTATION. DISTRIBUTION OF INCORRECT ANSWERS IS REPRESENTED BY THE TOTAL NUMBER OF “OTHER” AND “BACK” ANSWERS

Number of answers	IHL				FBR			FB errors (total)	
	NN	YY	YN	NY	FF	RF	RR	other	back
Slow	26	3	17	4	24	6	20	30	56
Avg.	26	6	17	1	17	9	24	47	67
Fast	28	6	14	2	23	2	25	38	70

source in front of them, followed by stimuli with different  $A$ ,  $B$ , and  $C$  parameters. During a measurement, a nonmoving sound source ( $A = 0$ ) is sometimes rendered.

For simplifying and reducing the parameters, three basic signal presentation setups were determined. Slow (S) means a few possible source locations, but in each location a longer playback time. On the other hand, fast (F) includes more sound source locations and shorter playback time per location. An average (A) was also selected in between (Table I). At slow presentation the HRTFs were updated about in every second, while at average this update is about two times a second, and at fast presentation it is about five times a second.

Parameter  $A$  was increased throughout the experiment and was terminated when subjects reported the percept of a moving sound source (by answering “yes” to the third question). First, they were exposed to the reference situation ( $A = 0$ ). Second, parameter  $A$  was set to 1, and the listening test was repeated, etc. After reporting the perception of movement, the simulation was stopped. This means that subjects exceeded the limit of the individual localization blur that would influence the measurement and the evaluation [31].

Seven different sound files were used for the investigation: white noise, 1500 Hz low-pass and 7000 Hz high-pass filtered noise, a female voice saying “welcome,” and three commonly used MS Windows sounds. These were “empty recycle bin,” “critical stop,” and the “exclamation”—all familiar with MS Windows users (they can be found in the Windows/Media folder). The selection criteria for these included spectral content, familiarity, and possible use for Earcons in the future [6], [33].

During evaluation, subjects answered the following questions: “Is the sound source externalized or in-the-head?”, “Where is the simulated sound source in the virtual space?”, and “Do you have the percept of a moving source?”. The first and last questions are a 2-category-forced-choice (yes/no) but the second is a 3-category-forced-choice (front/back/other direction).

The goal of Experiment 1 was to find the optimal value for parameter  $A$ , and used only white noise excitation. Experiment 2 then used all seven sound files, but only two values for  $A$ ,  $0^\circ$  (as a reference) and  $2^\circ$ .

## IV. RESULTS

### A. Experiment 1

1) *Results for In-the-Head Localization:* The effectiveness of source motion in reducing in-the-head localization was evaluated by classifying the subjects into one of four categories

- Subjects for whom the emulation of head-movement was not necessary for resolving in-the-head localization (they do not have it even without emulation). This group is labeled NN.
- Subjects for whom the emulation of head-movement did not help by resolving in-the-head localization (they have it from the beginning and also with emulation). This group is labeled YY.
- Subjects for whom the emulation of head-movements did help to resolve in-the-head localization (first they have it, later they do not). This group is labeled YN.
- Subjects for whom the emulation indicated in-the-head localization (they did not have it, but with the emulation they do). This group is labeled NY.

At an average rendering speed, from the 50 subjects, 17 found the emulation helpful. Most of them (26) did not need it because they externalized the sound source from the beginning. For six subjects, the emulation did not help at all. It is interesting that one subject first reported externalized source, then in-the-head localization during the emulation. Table II shows the results for in-the-head localization (IHL), front-back reversals (FBR) and front-back errors (FB errors) for slow (S), average (A), and fast (F) source motion.

2) *Results for Front-Back Reversals:* Similar evaluation of measured data can be made for front-back reversals using the same representation method in the tables and measurement parameters. With HRTF rendering, virtual sources that are supposed to be directly in front of the listener are often perceived to be in back, which is a significant localization error. Even using HRTFs from a good localizer can lead to a high rate of reversals. In this evaluation, we only have three sets for classification:

- Subjects for whom the emulation of head-movement was not necessary and they had a correct, frontal image even without emulation. This group is labeled FF.
- Subjects for whom the emulation of head-movements did help to increase the number of frontal judgments. This group is labeled RF.
- Subjects for whom the emulation of head-movement did not help at all. This group is labeled RR.

Table II shows these results at the three different presentation speeds. It is clear that this emulation is not very effective in increasing correct judgments of subjects. Only 2 to 9 subjects out of 50 gained something from the emulation, most of them gained nothing. The error rates are about 50%, and the best results are for the average speed (18%). From the total answers of 150 (for all speeds), only 17 indicated any kind of improvement (11%).

The table also shows the total number of individual false answers representing the difference between “rear” and “other”

TABLE III

SUMMARIZED RESULTS FOR 50 SUBJECTS USING WHITE NOISE STIMULUS. COLUMNS SHOW THE NUMBER OF SUBJECTS WHO PERCEIVED THE MOTION FROM  $A = 1$  TO  $A = 7$ , RESPECTIVELY. THE AVERAGE IN DEGREES AS WELL AS THE NUMBER OF PERSONS WITH A GAIN IS ALSO SHOWN

Value of $A$	1	2	3	4	5	6	7	Avg. (deg)	Gain (nr)
Slow	0	14	32	1	1	1	1	2.92	15/50
Avg.	0	2	33	14	1	0	0	3.98	16/50
Fast	0	0	37	6	6	1	0	3.42	14/50

TABLE IV

RESULTS FOR IN-THE-HEAD LOCALIZATION OF 50 SUBJECTS FOR ALL STIMULI AND PRESENTATION SPEED ( $A = 2$ ). THE LAST COLUMN SHOWS THE RELATIVE VALUE OF SUBJECTS HIGHLIGHTED FOR WHOM THE EMULATION WAS EFFICIENT (RELATIVE NUMBER OF YN ANSWERS FOR ALL STIMULI ON AVERAGE)

IHL		White Noise	1500 Hz LPF	7000 Hz HPF	Speech	Exclamation	Critical stop	Recycle bin	%
Fast	NN	28	3	11	3	8	10	9	20.57
	YN	14	10	8	11	8	16	10	<b>22.00</b>
	YY	6	23	24	23	21	15	23	38.57
	NY	2	14	7	13	13	9	8	18.86
Avg.	NN	26	9	4	14	5	8	10	21.71
	YN	17	14	9	7	13	8	9	<b>22.00</b>
	YY	6	20	24	19	19	20	20	36.57
	NY	1	7	13	10	13	14	11	19.72
Slow	NN	26	9	6	13	13	13	10	25.72
	YN	17	9	11	9	8	11	11	<b>21.71</b>
	YY	3	24	22	23	16	16	19	35.14
	NY	4	8	11	5	13	10	10	17.43

sound directions. The answer “rear” is about two times more frequent than “other” directions. During evaluation, every answer different from “front” was regarded as an error report. Results for slow and fast presentation tend to be worse, so in connection with front-back errors, a medium speed of presentation seems to be appropriate. Individual results of the data displayed in Table II can be found for white noise only in [32].

3) *Perception of Movement*: During reporting sensations about externalization and front-back errors, we wanted to know if the subjects perceive the motion or not. The experiment stops if they do. We analyzed this limit in connection with presentation speed and error values as well.

At average speed, two subjects reported the motion sensation at  $2^\circ$ , 33 subjects at  $3^\circ$ , 14 subjects at  $4^\circ$ , and one subject at  $5^\circ$  (see Table III); the average is  $3.98^\circ$ . An interesting result is that out of the 17 subjects for whom the emulation helped resolving in-the-head localization, only one person reported externalization and the perception of movement at the same distance; on the other hand, 16 subjects externalized the sound source  $1^\circ$  or  $2^\circ$  before they perceived movement. Taking this one subject out, we can conclude that for 16 out of 50 subjects (one-third), the motion contributed to externalization.

At fast speed, this number is 14 out of 50. At  $2^\circ$ , nobody perceived the movement. At slow speed, we have to exclude two of the 17 subjects for the same reason, that is externalization and perception of movement appear simultaneously. So the motion was helpful for 15 out of 50.

Summarizing these results, we can conclude that about one-third of the subjects gain something from the emulation for externalization, almost independent of presentation speed. The value  $A = 2$  is roughly optimal, because this seems to be the most appropriate for most of the subjects. We have to take into account that  $A = 2$  can be perceived as movement at slow speed of presentation. This is rather interesting, as one might expect

that perception of movement is more likely at a faster speed of presentation.

We have to also mention here that this kind of localization blur of about  $3^\circ$ – $4^\circ$  is an average on the whole 2-D surface around the origin, including movements in the horizontal plane and in the median plane, while results of localization blur measurements usually represent horizontal and vertical directions separately [2], [33].

## B. Experiment 2

Based on the results from Experiment 1, we used only  $A = 0$  for reference and  $A = 2$  for emulation in Experiment 2. The goal of this second part was to determine any influence on the results using different stimuli.

1) *Results for In-the-Head Localization and Reversal Rates*: Tables IV and V show results for in-the-head localization and front-back errors for all stimuli and presentation speeds, respectively. The last column shows the relative values. For in-the-head localization, motion was helpful (YN) for 22%, 22%, and 21.71% of the subjects, i.e., 21.90% on average. Averaged values are 22.67% for the group NN, 36.76% for YY, and 18.67% for NY.

Similarly, Table V presents relative values in the last column. On average, there is an increase in correct judgments by only 14.38% of the subjects, while 39.43% have frontal perception with and without emulation. For 46.19%, emulation does not help at all.

2) *Individual Results for Different Stimuli*: Tables IV and V also provide data about the different types of stimuli. As expected, white noise is often externalized, followed by high-frequency sounds and female speech. For Windows samples, reducing the speed increased externalization rates. The file “exclamation.wav” is systematically worse than “critical stop.wav”

TABLE V

RESULTS FOR FRONT-BACK REVERSALS OF 50 SUBJECTS FOR ALL STIMULI AND PRESENTATION SPEED ( $A = 2$ ). THE LAST COLUMN SHOWS THE RELATIVE VALUE OF SUBJECTS HIGHLIGHTED FOR WHOM THE EMULATION WAS EFFICIENT (RELATIVE NUMBER OF RF ANSWERS FOR ALL STIMULI ON AVERAGE)

FBR		White Noise	1500 Hz LPF	7000 Hz HPF	Speech	Exclamation	Critical stop	Recycle bin	%
Fast	FF	23	15	21	25	21	23	19	42.00
	RF	2	7	7	4	8	7	7	<b>12.00</b>
	RR	25	28	22	21	21	20	24	46.00
Avg.	FF	17	13	20	24	20	19	16	36.86
	RF	9	9	5	1	10	9	9	<b>14.86</b>
	RR	24	28	25	25	20	22	25	48.28
Slow	FF	24	16	21	19	17	22	19	39.43
	RF	6	11	9	11	7	6	7	<b>16.29</b>
	RR	20	23	20	20	26	22	24	44.28

and “recycle.wav.” The results for the last two are almost identical.

For speech, average or slow presentation is recommended. While speech is externalized more frequently without emulation of movement at these speeds than Windows sounds, the emulation of head-movement has a bigger impact on the externalization of Windows-sounds than of speech. At fast presentation, speech is not the optimal signal for externalization.

There is no clear evidence that familiar sounds from the operating system would be better externalized than speech or noise. Familiar Windows-sounds may be more informative in meaning, but localization performance depends foremost, on spectral content, and second, on presentation speed.

Basically, the same conclusions can be drawn for front-back rates: increase due to the emulation can be observed for the Windows-sounds at medium speed and for speech at fast speed. It is common that motion is more helpful at slow presentation speed.

## V. CONCLUSION

Fifty untrained subjects participated in listening tests using HRTF synthesis and headphone playback. A virtual target source was simulated in front of the listener. An initial stationary presentation was followed by random movements of  $1^\circ$ – $7^\circ$  around the reference location in all directions. The goal was to simulate small head-movements and to evaluate front-back reversal and in-the-head localization rates. The most important idea was to replace unintentional, small head-movements during real-life listening with simulated object motion. In virtual simulation this corresponds to dynamic changes in the HRIRs. However, this random, small amplitude object motion must not be recognized by the listeners. The goal was to find whether this undetectable motion improves localization or not.

Experiment 1 led us to conclude that motion can help resolve in-the-head localization for about one-third of the subjects if we randomly move the simulated sound source by about  $1^\circ$ – $2^\circ$  using white noise stimuli. On the other hand, the emulation did not really influence front-back reversal rates. Simulated head-movements more than  $4^\circ$  will be perceived as a moving source.

Experiment 2 used seven different types of stimuli. Emulation of small head-movements of  $2^\circ$  could actually increase externalization rates in about 21% of the subjects. Broadband signals and stimuli with more high-frequency content are recommended for practical applications. Presentation speed is not critical, an average speed seems to be optimal for all kinds of stimuli. The same is true for front-back reversals, although only about 14%

found the emulation helpful. This rate seems to be too low to be useful, so this kind of emulation has much more impact on in-the-head localization. Practical applications such as creating Earcons, Auditory Icons, or virtual environments for blind persons may benefit from an implementation of an emulation like this during playback without the need of any additional hardware for head-tracking.

## REFERENCES

- [1] K. Crispian and H. Petrie, “Providing access to GUIs for blind people using a multimedia system based on spatial audio representation,” *Audio Eng. Soc. 95th Convention*, 1993, New York, Preprint 3738.
- [2] G. Wersényi, “Localization in a HRTF-based minimum audible angle listening test on a 2-D sound screen for GUI applications,” *Audio Eng. Soc. 115th Convention Preprint*, Oct. 2003, New York.
- [3] D. Burger, C. Mazurier, S. Cesarano, and J. Sagot, “The design of interactive auditory learning tools,” *Non-Visual Human-Comput. Interaction*, vol. 228, pp. 97–114, 1993.
- [4] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, “Earcons and Icons: Their structure and common design principles,” *Human-Comput. Interaction*, vol. 4, no. 1, pp. 11–44, 1989.
- [5] G. Awad, “Ein Beitrag zur Mensch-Maschine-Kommunikation für Blinde und Hochgradig Sehbehinderte,” Ph.D. dissertation, TU-Berlin, Berlin, Germany, 1986.
- [6] G. Wersényi, “Localization in a HRTF-based virtual audio synthesis using additional high-pass and low-pass filtering of sound sources,” *J. Acoust. Sci. Technol. Jpn.*, vol. 28, no. 4, pp. 244–250, Jul. 2007.
- [7] J. Blauert, *Spatial Hearing*. Cambridge, MA: MIT Press, 1983.
- [8] F. E. Toole, “In-head localization of acoustic images,” *J. Acoust. Soc. Amer.*, vol. 48, pp. 943–949, 1969.
- [9] P. Laws, “Zum Problem des Entfernungshören und der Im-Kopf-Lokalisierung von Hörereignissen,” Ph.D. dissertation, TU-Aachen, Aachen, Germany, 1972.
- [10] G. Plenge, “Über das problem der im-kopf-lokalisierung,” *Acoustica*, vol. 26, pp. 241–252, 1972.
- [11] N. Sakamoto, T. Gotoh, and Y. Kimura, “On “out-of-head localization in headphone listening,”” *J. Audio Eng. Soc.*, vol. 24, pp. 710–716, 1976.
- [12] J. Kawaura, Y. Suzuki, F. Asano, and T. Sone, “Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear,” *J. Acoust. Soc. Jpn.*, vol. 12, pp. 203–215, 1991.
- [13] P. A. Hill, P. A. Nelson, and O. Kirkeby, “Resolution of front-back confusion in virtual acoustic imaging systems,” *J. Acoust. Soc. Amer.*, vol. 108, no. 6, pp. 2901–2910, 2000.
- [14] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, “Augmented reality audio for mobile and wearable appliances,” *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 618–639, 2004.
- [15] J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss, “An interactive virtual-environment generator for psychoacoustic research I: Architecture and implementation,” *Acoustica*, vol. 86, pp. 94–102, 2000.
- [16] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. London, U.K.: Academic, 1983.
- [17] R. L. McKinley and M. A. Ericson, “Flight demonstration of a 3-D auditory display,” in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Mahwah, NJ: Lawrence Erlbaum, 1997, pp. 683–699.

- [18] M. Cohen and E. Wenzel, "The design of multidimensional sound interfaces," in *Virtual Environments and Advanced Interface Design*, W. Barfield and T. A. Furness III, Eds. New York: Oxford Univ. Press, 1995, pp. 291–346.
- [19] F. Chen, "Localization of 3-D sound presented through headphone—Duration of sound presentation and localization accuracy," *J. Audio Eng. Soc.*, vol. 51, no. 12, pp. 1163–1171, 2003.
- [20] D. R. Begault, E. Wenzel, and 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.*, vol. 49, no. 10, pp. 904–917, 2001.
- [21] M. Kleiner, B. I. Dalenbäck, and P. Svensson, "Auralization—An overview," *J. Audio Eng. Soc.*, vol. 41, pp. 861–875, 1993.
- [22] R. L. Martin, K. I. McAnally, and M. A. Senova, "Free-field equivalent localization of virtual audio," *J. Audio Eng. Soc.*, vol. 49, no. 1/2, pp. 14–22, 2001.
- [23] W. Noble, "Auditory localization in the vertical plane: Accuracy and constraint on bodily movement," *J. Acoust. Soc. Amer.*, vol. 82, pp. 1631–1636, 1987.
- [24] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "The importance of head movements for binaural room synthesis," in *Proc. Int. Conf. Auditory Display*, Espoo, Finland, Jul. 2001, pp. 21–25.
- [25] D. R. Perrott, H. Ambarsoom, and J. Tucker, "Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions," *J. Acoust. Soc. Amer.*, vol. 82, pp. 1637–1645, 1987.
- [26] "BEACHTRON-Technical Manual. Rev. C," Crystal River Engineering, Inc., Palo Alto, CA, 1993.
- [27] S. H. Foster and E. M. Wenzel, "Virtual acoustic environments: The convolotron. Demo system presentation at SIGGRAPH'91," in *Proc. 18th ACM Conf. Comput. Graphics Interactive Tech.*, Las Vegas, NV, 1991, ACM Press, New York.
- [28] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, 1993.
- [29] D. J. Kistler and F. L. Wightman, "Principal component analysis of head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 88, p. 98, 1990.
- [30] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening I-II," *J. Acoust. Soc. Amer.*, vol. 85, pp. 858–878, 1989.
- [31] P. Minnaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 10, pp. 919–929, 2005.
- [32] G. Wersényi, "Simulation of small head-movements on a virtual audio display using headphone playback and HRTF synthesis," in *Proc. 13th Int. Conf. Auditory Display (ICAD)*, Montréal, QC, Canada, Jun. 2007, pp. 12–14, 6 pages.
- [33] G. Wersényi, "Localization in a HRTF-based minimum-audible-angle listening test for GUIB applications," *Electron. J. Tech. Acoust. (EJTA)*, vol. 1, 2007 [Online]. Available: <http://www.ejta.org>, 16 pp.



**György Wersényi** was born in Győr, Hungary, in 1975. He received the M.Sc. degree in electrical engineering from the Technical University of Budapest, Budapest, Hungary, in 1998 and the Ph.D. degree from the Brandenburgische Technische Universität (BTU) Cottbus, Germany, in 2002.

He has been a member of the Department of Telecommunications at the Széchenyi István University, Győr, Hungary, since 2002. He is an Associate Professor at the field of telecommunications, audio and video broadcast, and electroacoustics. His current research includes spatial hearing, listening tests, virtual audio displays, HRTF measurement, and dummy-head techniques.

Dr. Wersényi is a member of the Audio Engineering Society, the Hungarian Telecommunication Society, and the International Community for Auditory Display (ICAD).