HAL
archives-ouvertes.fr

Interaction between auditory and visual distance cues in virtual reality applications

Nicolas Côté, Vincent Koehl, Mathieu Paquier, Frédéric Devillers

▶ **To cite this version:**

Nicolas Côté, Vincent Koehl, Mathieu Paquier, Frédéric Devillers. Interaction between auditory and visual distance cues in virtual reality applications. Forum Acusticum 2011, Jun 2011, Aalborg, Denmark. pp.1275-1280, 2011. <hal-00606228>

HAL Id: hal-00606228
http://hal.univ-brest.fr/hal-00606228

Submitted on 5 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Forum Acusticum 2011 - Aalborg - Denmark

Interaction between Auditory and Visual Distance Cues in Virtual Reality Applications

Nicolas Côté, Vincent Koehl and Mathieu Paquier
European Centre for Virtual Reality (LISyC EA 3883), University of Brest (UEB), France

Frédéric Devillers
European Centre for Virtual Reality (LISyC EA 3883), École Nationale d'Ingénieurs de Brest (ENIB), France

Abstract

Virtual reality applications rely on the accurate localization of virtual objects. The position of these objects is defined by three coordinates: azimuth, elevation and distance. Even though several studies investigated the perception of auditory and visual cues in azimuth and elevation, little has been made on the distance dimension. It has been shown that listeners under-estimate egocentric distance in virtual environments. This study aims at investigating the way humans perceive visual and auditory distance cues of virtual objects. For egocentric distances from 2 to 20 m, subjects were asked to estimate the object distance in three contexts: auditory modality alone, visual one alone, combination of both modalities. Even though egocentric distance is under-estimated in the three contexts, the results suggest a higher influence of visual information than auditory information on the perceived distance.

PACS no. 43.66.Qp, 43.60.Jn

1. INTRODUCTION

Accurate distance perception is an essential task in everyday life. Both auditory and visual modalities have an influence on such a task. However, vision enables more accuracy in object localization or distance determination than audition. Auditory distance perception has received relatively less scientific attention than either directional auditory localization (i.e. azimuth and elevation) or visual distance perception. Auditory distance estimation is a far more complicated task than the azimuthal localization of a sound source. According to Zahorik et al. [1], sound source location are under-estimated for physical egocentric distance higher than 1.9 m and overestimated for lower physical distance. This tendency to a specific distance value of 1.9 m is considered as the "default" distance in absence of other distance cues.

By using either auditory or visual cues, egocentric distance are under-estimated by human subjects. It has been suggested in the literature that this under-estimation corresponds to an adaptation of humans over his evolution that provides a "margin of safety" for possible danger in natural environment, e.g. [1]. Da Silva [2] suggested that perceived distance is best estimated by Steven's power law according to the following function:

$$d' = kd^a \tag{1}$$

where d' is the perceived distance, d is the physical distance and k and a are two coefficients.

Several procedures can be used for the estimation of target distances. Examples of such procedures are: estimation based on distance scales (e.g. feet or meters), ratio scale between multiple conditions or directed action such as blind-walking task. In this procedure, observers first see/hear the object and then are asked to walk without vision to a specific point. Then, based on their sense of self-motion, subjects are able to update the target position. However, response procedures have an influence on the perceived values. For instance, directed action enables more accurate distance estimations than verbal reports [3].

Human distance perception in real environments has been exhaustively studied. Even though experiments conducted in such environments have more ecological validity, they enable less control on the experimental conditions than experiments that make use of virtual environments. Similar under-estimations of the target auditory distance have been observed with real and virtual sound sources, which suggest possible use of auditory displays for auditory distance perception studies. However, the literature reports a

(c) European Acoustics Association

larger compression of target egocentric distance in visual virtual environments than in real environments, e.g. Willemsen and Gooch [4].

This paper investigates the perceived egocentric distance in auditory and visual virtual environments. The same procedure is applied to the auditory modality, the visual one and a combination of them. For this purpose, Section 2 describes the perceived cues that influence distance estimation. Section 3 introduces acoustic and visual displays used in virtual reality applications. Then, in Section 4, the test procedure is described and, in Section 5 the results are analyzed.

2. DISTANCE PERCEPTION

Physical egocentric distance is under-estimated by humans, either for audition or vision. The following section describes the major indicators analyzed by humans to determine the distance of real or virtual objects. Table I provides a list of the major cues that have an influence on human distance perception. The last paragraph discusses the combination of multiple factors.

2.1. Visual cues

Distance perception of a visual object depends on several visual and non-visual cues [5]. Visual cues can be divided in two groups, monocular cues and binocular ones. Perception of monocular cues requires one eye only whereas perception of binocular cues requires both eyes. Non-visual cues comprise the *a priori* information on the visual object and the environment. According to Cutting and Vishton [6], the visual environment of an observer can be divided in three subspaces: *personal space* (within arm's reach), *action space* (2–30 m), and *vista space* (beyond 30 m).

2.2. Auditory cues

Various acoustic and non-acoustic factors influence the perceived egocentric distance of sound sources. Acoustic cues can be divided in two groups, monaural cues and binaural ones. However, the binaural cues have reduced utility for sound source at distances beyond 1 m [7]. Non-acoustic cues correspond to *a priori* information such as the familiarity with the sound source or the listening environment. For instance, Kopčo et al. [8] showed that repeated presentations of the same sound source at different distances in the same listening environment enable subjects to detect relative changes for acoustic cues such as intensity or spectrum.

2.3. Cue combination

In natural environments, multiple distance cues are available to listeners. Perceived distance estimations are based on one or more cues and their congruence. Subjects are thus able to combine all cues to localize the perceived object. However, all cues do not equally contribute to perceived distance [9]. For instance, in case of multi-modal perception (e.g. visual and auditory), the visual information may enhance the auditory localization [10].

Table I. Visual and auditory cues.

Modality	Cues
Visual (monocular)	accommodation, size, occlusion, shading, texture gradient, perspective, atmospheric haze
Visual (binocular)	convergence, binocular disparity
Auditory (monaural)	intensity, spectrum, direct-to-reverberant ratio
Auditory (binaural)	interaural time differences (ITD), interaural level differences (ILD)
Common	familiarity, motion

3. VIRTUAL ENVIRONMENTS

Virtual reality applications make use of specific visual displays, such as Head-Mounted Displays (HMDs) or Cave Automatic Virtual Environments (CAVEs), in order to immerse the observer in the visual scene. These displays are usually combined to stereoscopic rendering technique, which improves depth perception in a realistic way. However, such stereoscopic visual displays have several limitations [11]. For instance, they yield to an accommodation-convergence mismatch. Indeed, in real environments accommodation and convergence are linked together whereas by using visual displays (including TV screens) observers accommodate on the image plane. In addition, they provide a restricted field of view and a low quality of graphics rendering compared to a real environment.

Typical auditory displays used in virtual reality applications are Vector Base Amplitude Panning (VBAP) technique, Wave Field Synthesis (WFS) or Binaural rendering. The latter technique makes use of direction-dependent filters called Head-Related Transfer Functions (HRTFs). These filters encode the amplitude and phase differences for each ear and each direction in both dimension azimuth and elevation. Binaural rendering is a convenient auditory display that enables sound scene reproduction in three dimensions (azimuth, elevation, distance) and accurate immersion of the listener uncoupled from the real listening room. In addition, realistic auditory displays require a room effect simulation. Bronkhorst [12] compared real binaural recordings to a room acoustics simulation which uses binaural reproduction techniques. Since both sound signals provided similar results it is thus assumed that virtual acoustics can be used for distance perception experiments.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

Table II. Test conditions.

Variable	Description
Modality	auditory, visual, bimodal
Distance	2, 3, 5, 10, 20 m
T_{60}	370, 860 ms
Visual cues	room
	room & pillars & neon lights & carpet
Offset	8 conditions

Experimental research on distance perception using virtual environments suggests different conclusions for auditory displays and for visual ones. Bronkhorst [12] suggested that differences for virtual and real acoustic environments are relatively small and, therefore, virtual acoustics can be used for distance perception experiments. However, studies on visual distance performances showed differences in distance perception between real and virtual visual environments [4]. The under-estimation of visual egocentric distance in virtual environments may have several causes: the reduced field of view, the rendering quality and the test procedure. Willemsen and Gooch [4] suggested that the display itself is the primary source of the distance compression in virtual environments. For instance, Plumert et al. [13] suggested that large-screen visual displays may provide better distance perception than HMD displays.

4. METHOD

4.1. Environment and stimuli

Since our research questions focus on the influence of acoustic and visual cues on the perceived distance, the same visual object and auditory source were processed through different conditions. There were four experimental variables: presentation modality (auditory-only, visual-only and bimodal), target distance, room effects and amount of visual information. The present study investigates the distance perception within the *action space*. In addition, the influence of auditory cues on visual localization accuracy were assessed by using 8 conditions with spatially incoherent auditory and visual cues. For these conditions, the auditory cues were moved back or forward compared to the visual cues. The test conditions are summarized in Table II.

The visual target consisted in a virtual blue loudspeaker of $40 \times 60\,\mathrm{cm}^2$. The visual environment was a virtual room corresponding to the extension of the real test room through the visual display. Figure 1 shows a picture of the test room including the virtual extension of the room. The visual display was a $2.4 \times 1.8\,\mathrm{m}^2$ stereoscopic passive screen with a 1280×1024 resolution. The subjects sat on a chair placed at 2 m in front of the middle of the screen resulting in a $\pm 31°$ field of view. The visual environments were rendered by the ARéVi library [14], which uses OpenGL. A fixed inter-pupil distance of 6.5 cm was used for the stereoscopic visual rendering.

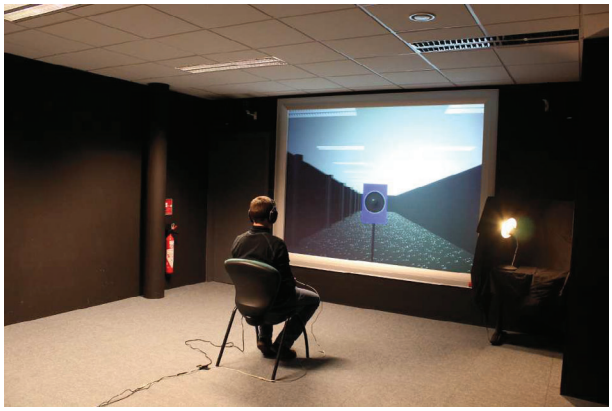

Figure 1. Test room and virtual environment.

The auditory source corresponds to a speech signal composed of two French sentences, spoken by a male and a female talker. The sound stimuli is processed by a binaural rendering system and reproduced through headphones. For this purpose, the auditory source has been convolved with Binaural Room Impulse Responses (BRIRs) at the different distances listed in Table II. Since the use of non-individualized Head-Related Transfer Functions (HRTFs) does not affect distance estimation accuracy [15], the BRIRs under use have been produced with artificial heads recordings. The BRIRs are composed of two parts:

1. The early reflections (up to the second order) have been simulated by the Matlab script "Roomsim" [16]. The early reflections were simulated using the test room size (i.e. combination of the real and virtual prolongation of the test room: $27.1 \times 5.5 \times 2.60\,\mathrm{m}^3$) with two different sets of absorption coefficients. In addition, an air absorption model has been used.
2. The diffuse field part came from a database of real BRIRs [17]. This database has been used to provide a realistic interaural cross-correlation.

The two parts were produced at a 44 100 Hz sampling frequency and combined to provide realistic reverberation conditions. The listening level was set to 63 dB at each of the two ears of the subject for a sound source at 2 m and stayed within $[63.8; 58.7]$ dB for all auditory-only conditions. The binaural stimuli were sent directly to the *Lexicon Alpha* soundcard and played back over a *Sennheiser HD 650* headphones.

4.2. Procedure

Even though a triangulated walking task provides more accurate distance estimations, this procedure uses angular auditory and visual cues whereas this study focuses on distance perception only. Therefore,

a static judgment task procedure has been used in this study. After presentation of the auditory and/or visual stimuli, subjects were asked to report their egocentric distance judgments by using a keypad. Subjects were limited in time (12 s) to enter their distance judgments.

All participants were first provided with a written description of the experimental task. After reading the instruction, an experimenter presented an equivalent verbal description of the task. The experiment consisted of two sessions:

1. During the first session the subjects judged the auditory- and visual-only conditions. The conditions were assessed in blocks, half of the subjects starting with the auditory block and half with the visual block.
2. During the second session the subjects judged the three auditory-visual blocks. The last block included the 8 spatially incoherent auditory and visual cues conditions.

The two sessions were separated in time by at least 36 hours. The participants had 8 and 4 training trials in the first and the second sessions, respectively. In each session, all combinations of target distance and reverberation and/or amount of visual information were presented in random order, with four repetitions per condition. Since the visual and auditory rendering were not modified according to the position of the subjects, they were asked not to move their head during the test.

5. ANALYSIS OF RESULTS

A total of 24 subjects participated in the experiment. Participants were naive with respect to the purpose of the experiment. They had normal or corrected to normal vision and reported no auditory impairments. Except for 15 trials (over 4 608), the subjects were always able to provide a score within the 12 s time-window. This section presents only the spatially coherent conditions.

An analysis of variance (ANOVA) was performed on the perceived distance from the first session to measure whether subjects provided equal distance estimations in case they had the auditory block first or the visual block first. The ANOVA shows neither an effect of *order* ($F(1, 1890) = 1.09$, $p = 0.30$) nor an interaction between the order and target distance ($F(4, 1890) = 0.33$, $p = 0.86$). In other words, a prior visualization of the virtual object and virtual test room had no influence on the auditory distance perception.

Figures 2 to 4 show the relationship between the target distance and the averaged perceived distance for the auditory-only, the visual-only and the bimodal

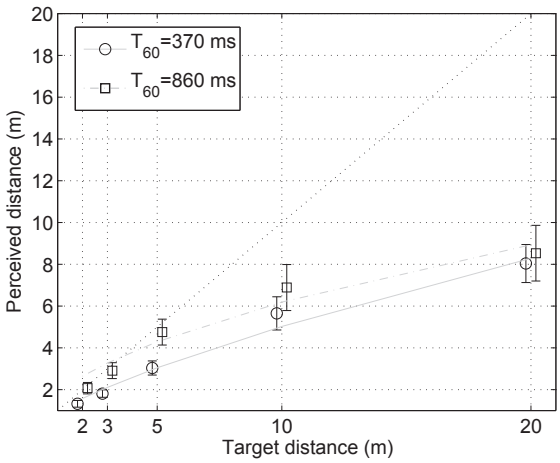

Figure 2. Relationship between the perceived *auditory* distances and target distances.

conditions. Error bars correspond to the 95% confidence intervals. The gray curves correspond to the estimated power law functions, see eq. (1). The dotted line represents ideal performance.

5.1. Auditory modality

The auditory distance values were analyzed with a repeated measures ANOVA with the within-subject factors *reverberation* and *target distance*. The ANOVA indicates a highly significant influence of the target distance ($F(4, 720) = 41.84$, $p < 10^{-4}$), an effect of the reverberation time ($F(1, 720) = 8.3$, $p = 0.0084$) but the interaction between distance and reverberation was not significant ($F(4, 720) = 2.16$, $p = 0.08$). This result follows assumption made by Bronkhorst and Houtgast [18] that perceived distance increases as the direct-to-reverberant energy ratio decreases. For the long reverberation time, subjects provided accurate distance estimations for distances below 10 m whereas target distances are under-estimated for distances above 5 m, see fig. 2. By using the perceived and the target egocentric distance values, the k and a coefficients in eq. (1) are estimated in a least-square sense. The estimated coefficients, $0.96\,d^{0.72}$ ($RMSE = 0.39$) for the short reverberation time and $1.83\,d^{0.53}$ ($RMSE = 0.57$) for the long reverberation time, show a compression of the target distance ($a < 1$).

In addition, the variability in individual auditory distance estimations is high and it increases with increasing target source distance. Similar results have been observed by Zahorik et al. [1].

5.2. Visual modality

A repeated measures ANOVA with the within-subjects factors *target distance* and *environment* was performed on the perceived visual distance values.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

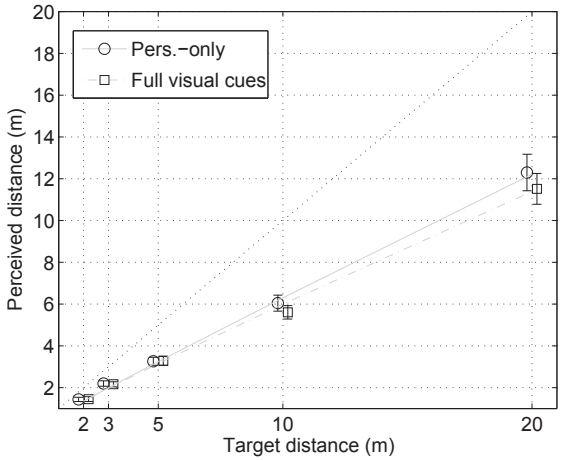

Figure 3. Relationship between the perceived *visual* distances and target distances.

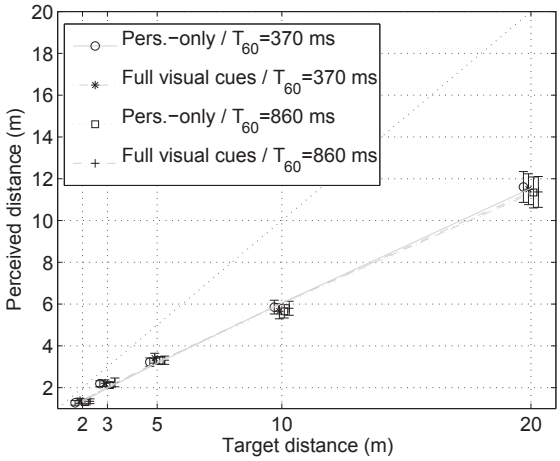

Figure 4. Relationship between the perceived distances and target distances for the *bimodal* conditions.

The analysis of this second block shows a high influence of the target distance ($F(4, 720) = 156.5$, $p < 10^{-4}$) and an effect of the environment ($F(1, 720) = 8.96$, $p = 0.0065$). This analysis revealed also the existence of a significant interaction between distance and environment ($F(4, 720) = 6.74$, $p = 0.0001$). Figure 3 shows the relationship between the perceived visual distance and the target distance. The environment based on perspective-only visual cues enables less distance errors than the full visual cues environment, especially for target distances above 5 m. The latter environment included more "anchors" (e.g. pillars and neon lights) to help subjects make distance estimates. According to Andre and Rogers [3], the amount of visual information has an effect on verbal report accuracy. The authors found that an increase of the amount of visual anchors available for the subjects led to an increase of the distance estimations. However, our results suggest that including more visual information in the virtual environment decreases perceived visual distance.

In addition, the relationship between the target and perceived distances is different from the auditory case. The interpolation of eq. (1) in a least-square sense shows estimated coefficients of $0.70\,d^{0.96}$ ($RMSE = 0.17$) and $0.71\,d^{0.93}$ ($RMSE = 0.24$) for the restricted and full visual cues, respectively. These values and fig. 3 show that target distance is not compressed ($a > 0.93$) but still under-estimated ($k < 1$).

An ANOVA performed on the perceived distance values with the within-subjects factors *target distance* and *modality* revealed that perceived distance is not influenced by the modality ($F(1, 1440) = 1.66$, $p = 0.21$) but a significant interaction exist between *target distance* and *modality* ($F(4, 1440) = 18.6$, $p < 10^{-4}$). A comparison of figs. 2 and 3 shows a lower perceived distances for target distance at 20 m for the auditory-only conditions than for the visual-only conditions. Below 20 m, both modalities provide equivalent perceived distance values.

5.3. Bimodal conditions

The bimodal distance estimations, depicted in fig. 4, show similar results to the visual estimations: the subjects did not compress target distance ($a > 0.90$). A repeated measures ANOVA with the within-subjects factors *target distance*, *reverberation* and *environment* was performed on the perceived distance values. The results reveal a significant effect of the target distance ($F(4, 1440) = 172.57$, $p < 10^{-4}$) but neither an effect of reverberation ($F(1, 1440) = 0.15$, $p = 0.70$) nor an effect of the environment ($F(1, 1440) = 2.72$, $p = 0.1125$). Both reverberation times provide the same bimodal distance estimations whereas reverberation has an influence on the auditory-only conditions. The restricted influence of the direct-to-reverberant energy ratio on the bimodal conditions can be considered as a visual capture effect of the perceived auditory distance.

Two ANOVAs were performed to compare the unimodal and bimodal conditions. The first one had the within-subjects factors *target distance*, *modality* and *reverberation* and the second one had the within-subjects factors *target distance*, *modality* and *environment*. In both cases, the perceived distance was not influenced by the modality. These results suggest that the bimodal conditions does not change visual distance localization accuracy.

The results show also an increase of the variability in individual distance estimations with increasing target source distance for the visual-only and the bimodal conditions. However, this variability is reduced in visual-only and bimodal conditions compared to the auditory-only conditions.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767

6. CONCLUSIONS

This study makes use of virtual rendering systems in order to investigate auditory and visual distance localization performances under various unimodal and bimodal conditions. Overall, the results of our study show that acoustic and visual cues can modified perceived auditory and visual distance in virtual environments, respectively.

The results are consistent with performances observed in experiments carried out in real environments: target distance is under-estimated in almost all of the conditions. However, the relationship between perceived and target distances is not consistent when compared between the unimodal conditions: the relationship follows a linear function in the visual-only conditions whereas it follows a power-law function in the auditory-only conditions. It results in a higher under-estimation of the target distance above 10 m for the auditory-only conditions than for the visual-only conditions. In addition, perceived distance values were similar in the visual-only and the bimodal conditions. This effect suggests that combined auditory and visual modalities does not change visual distance localization accuracy.

Topic for future investigations include the extension of this study with triangulated walking task procedure. However, such a procedure requires a dynamic and interactive visual and auditory rendering.

Acknowledgement

The authors would like to thank all test subjects. This research has been funded by the Finistère General Council (29), France.

References

[1] P. Zahorik, D.S. Brungart, and A.W. Bronkhorst. Auditory Distance Perception in Humans: A Summary of Past and Present Research. *Acta Acust. united with Ac.*, 91(3):409–420, 2005.

[2] J.A. Da Silva. Scales for Perceived Egocentric Distance in a Large Open Field: Comparison of Three Psychophysical Methods. *The Am. J. of Psychology*, 98(1):119–144, 1985.

[3] J. Andre and S. Rogers. Using Verbal and Blind-Walking Distance Estimates to Investigate the Two Visual Systems Hypothesis. *Attention, Perception, & Psychophysics*, 68(3):353–361, 2006.

[4] P. Willemsen and A.A. Gooch. Perceived Egocentric Distances in Real, Image-Based, and Traditional Virtual Environments. In *Proc of the IEEE Virtual Reality Conf.*, pages 275–276, 2002.

[5] M.W. Matlin and H.J. Foley. *Sensation and Perception.* Allyn and Bacon Boston, MA, Needhman Heights, Mass., 4th edition, 1997.

[6] J.E. Cutting and P.M. Vishton. *Perception of Space and Motion*, chapter Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth, pages 69–117. Academic Press, New-York, USA, 1995.

[7] BG Shinn-Cunningham. Distance Cues for Virtual Auditory Space. In *Proc. of the IEEE Pacific-Rim Conf. on Multimedia*, pages 227–230, Sydney, Australia, 2000. Citeseer.

[8] N. Kopčo, M. Schoolmaster, and B. Shinn-Cunningham. Learning to Judge Distance of Nearby Sounds in Reverberant and Anechoic Environments. In *Proceedings of the Joint Congress CFA/DAGA '04*, Strasbourg, France, 22-25 March 2004.

[9] M.S. Landy, L.T. Maloney, E.B. Johnston, and M. Young. Measurement and Modeling of Depth Cue Combination: In Defense of Weak Fusion. *Vision Research*, 35(3):389–412, 1995.

[10] P. Zahorik. Estimating Sound Source Distance With and Without Vision. *Optometry & Vision Science*, 78(5):270–275, 2001.

[11] J.P. Wann, S. Rushton, and M. Mon-Williams. Natural Problems for Stereoscopic Depth Perception in Virtual Environments. *Vision Research*, 35(19):2731–2736, 1995.

[12] A.W. Bronkhorst. Localization of Real and Virtual Sound Sources. *The J. of the Acous. Soc. of Am.*, 98(5):2542–2553, Nov. 1995.

[13] J.M. Plumert, J.K. Kearney, J.F. Cremer, and K. Recker. Distance Perception in Real and Virtual Environments. *ACM Transactions on Applied Perception (TAP)*, 2(3):216–233, 2005.

[14] ARéVi. *Atelier de Réalité Virtuelle (Virtual Reality Toolkit)*, Retrieved 31 March 2011. URL `http://svn.cerv.fr/trac/AReVi`.

[15] P. Zahorik. Auditory Display of Sound Source Distance. In *Proc. Int. Conf. on Auditory Display*, pages 326–332. Citeseer, 2002.

[16] D.R. Campbell, K.J. Palomaki, and G.J. Brown. Roomsim, a Matlab Simulation of Shoebox Room Acoustics for use in Teaching and Research. *Computing and Information Systems*, 9(3):48–51, 2005.

[17] M. Jeub, M. Schafer, and P. Vary. A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms. In *16th Int. Conf. on Digital Signal Proc.*, pages 1–5, 2009.

[18] A.W. Bronkhorst and T. Houtgast. Auditory Distance Perception in Rooms. *Nature*, 397 (6719):517–520, February 1999.

(c) European Acoustics Association, ISBN: 978-84-694-1520-7, ISSN: 221-3767