ECOLOGICAL ACOUSTICS AND THE MULTI-MODAL PERCEPTION OF ROOMS: REAL AND UNREAL EXPERIENCES OF AUDITORY-VISUAL VIRTUAL ENVIRONMENTS

Pontus Larsson, Daniel Västfjäll, Mendel Kleiner

Chalmers Room Acoustics Group
Department of Applied Acoustics
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
pontus.larsson@ta.chalmers.se

ABSTRACT

An ecological approach to multimodal perception of virtual environments suggests that different perceptual mechanisms should cooperate in forming an impression of the complex surrounding. Traditionally, Virtual Environments (VE's) has primarily been developed for the visual modality. It is hypothesized that multi-modal stimulation in VE's raises the experience of presence perceived by the user. Furthermore, it is believed that auditory cues also can improve memory. In Experiment 1, 40 subjects were assigned either to a unimodal (vision only) or bimodal (vision and hearing) virtual environment. The subjects had two memory- and navigation tasks, one where auditory cues had no apparent connection to visual information and one where auditory and visual cues carried similar information. Completion time for both tasks was measured. Statistical analysis showed as expected that no improvement of memory occurred for the unrelated task, while the auditory information yielded a significant effect in the second memory task. Ratings showed that subjects in the bimodal condition experienced significantly higher presence, were more focused on the situation and enjoyed the VE more than subjects receiving unimodal information did. Experiment 2 tested the hypothesis that varying degrees of visual realism would affect judgments of aural room qualities in a between-subjects design using 80 undergraduates. The results suggested that auditory stimuli in virtual environments can serve both as an information-carrying channel as well as way to improve the experience of presence in a VE and that memory performance may serve as a measure of presence in VE's.

1. INTRODUCTION

A basic tenet of ecological approaches to perception is that man perceives complex perceptual information in everyday-life [1]. According to such a view, complex perception relies on complex environmental information rather than the integration of simple sensations. Consequently, common psychophysical and psychoacoustic dimensions such as big and loud are to narrow and inappropriate descriptions of perception of complex environment [2],[3]. Ecological acoustics is therefore trying to go beyond classical psychoacoustic definitions and study events (as opposed to the sounds and the visual images) that give rise to certain complex perceptions [2]. A specific application is multimodal Virtual Environments (VE) for rendering of visual, aural and tactile information [4]. However most of these VE's focus on visual rendering of information [5]. Nevertheless, the importance of including all modalities to achieve a high level of virtualization of the environment and a high degree of presence for the user is beginning to be recognized [5]. Systems including multimodal information may also simply be more efficient and better systems than unimodal ones, because they better represent real life and the complexity of real-life experiences. Research on simultaneous presentation of auditory and visual information have shown that both interaction and synergetic effects can be observed, but that visual domination is strong in perceptually ambiguous situations [6]. Stein and Meredith [7] and McDonald and Ward [8] proposed that all modalities share an "information bank" in the human cortex, and thus that cross-modal interchange or modality fusion have a neurological basis. Drawing on this tenet, Västfjäll et al. [9] attempted to outline a model for ecological perception of auditory-visual information in rooms. According to this model a number of perceptual and judgmental attributes such as perceived room size, distance to the sound source, and perceived reverberation time are contingent on information from both visual and auditory information. That is, an individual will base his or her decision on the size of the room on both the visual and aural impression and previous experience of how other rooms looked and sounded like. In most cases the visual impression matches the aural impression. However, in some cases (as may be possible in a virtual environment) the visual impression may be mismatched with the aural impression (i.e. "this room sounds much bigger than it looks like"). In such cases it is likely that the visual impression will dominate the perception. In support of this, Västfjäll et al. ([9], [10]) found that participants both seeing and hearing a concert hall gave a significantly more accurate estimate of the room size than did participants who only heard the same room.

The main bulk of research has however not considered potential subjective effects of combined auditory and visual information in Virtual Environments. An assumption is that a high degree of presence or simulation fidelity is reached when visual and auditory information merges. The current research thus focuses on the combined effect of visual and aural in real and virtual environments.

2. EXPERIMENT 1. PRESENCE IN AN AUDITORY-VISUAL ENVIRONMENT

Experiment 1 tested the hypothesis that participants experiencing a bimodal (auditory-visual) VE would experience significantly more presence than participants experiencing a unimodal (auditory) VE. Moreover, it was hypothesized that the relative importance of information from different modalities is dependent on the type of task and virtual environment. In some cases information from visual and auditory cues are congruent, or matched. In such cases multiple cue utilization can be facilitated. In other cases visual and auditory cues are incongruent or mismatched. Cue integration will then be more difficult. In short, it is believed that matched multimodal cues will enhance task performance (if the cues are relevant to the task) and increase presence whereas mismatched multimodal cues will lead worsened task performance and decreased presence. Further, if multimodal cues are mismatched, it is believed that individuals will rely on visual cues over other cues due to visual dominance [11].

2.1. Method

The predictions were tested in a between-group experiment where 40 undergraduates or graduates were assigned either to perform tasks in either a visual (unimodal) (n=20) or auditory-visual (bimodal) virtual environment (n=20).
A digital model of Örgryte Nya Kyrka in Gothenburg, Sweden was used (see figure 1). The model was originally created in the room acoustic prediction program CATT-Acoustic [12] and transferred to the real-time VR software EON Studio [13]. Textures and two avatars were added to the EON simulation, as well as simulation objects for handling of peripherals (tracker, joystick, TCP/IP communication etc.).
The auditory scene was rendered using Lake Technologies AniScape software and CP4 hardware [14]. The reverberation tail was created in CATT-Acoustic based on the same digital model as for the visual stimuli. The music piece "Swanee River" performed by a female singer (from Yamaha DSP test disc) was used as anechoic input source to the audio workstation. The sound source was visually represented as a female avatar moving along a pre-determined path in the virtual church. For both conditions a Sony Glasstron HMD operating in stereoscopic mode was used as visual display. The graphics was rendered by a PIII-600 NT Workstation equipped with an ELSA Gloria XXL graphics board. Audio was presented using Beyerdynamic DT990 headphones and the participants' head rotation was tracked using Polhemus FASTRACK.

Figure 1. *Screenshot from the digital model of Örgryte Nya Kyrka. Participants' starting position and first view of the VE is shown*

2.2. Tasks

Participants had two tasks. First, they were asked to count the number of windows in the virtual church. This was done to ensure that all participants navigated around in the environment. Further, this task was mainly based on visual search and not related to the musical excerpt replayed in the bimodal condition (non-congruent task). Second, participants were instructed to find four different balls positioned in the church. Each ball contained one of the phrases 1. "Swanee river", 2. "Sadly I roam", 3. "Old folks at home", 4. "My heart grows weary" that were taken from the text of the musical excerpt "Swanee River" replayed to the participants in the Auditory-visual condition. The second task was thus related to the musical excerpt replayed to the participants in the bimodal task (congruent task). Search time was recorded for both tasks.

2.3. Results

For the memory tasks two dependent measures were used: number of recalled windows and number of recalled phrases. In addition, search time was recorded for both the windows and phrases tasks. First mean number of windows recalled was submitted to independent t-test to check for differences between the two groups. As predicted no differences was found between the two experimental conditions; $t(38) = -1.17$, $p>.05$. The mean number of windows recalled however indicated that participants in the visual and auditory condition gave a more accurate estimate of number of windows ($M = 12.5$, $SD = 2.31$) than the visual only condition ($M = 11.3$, $SD = 1.14$) (actual number of windows was thirteen).
Participants' mean number of phrases recalled was then submitted to independent t-tests. As predicted, the analysis yielded a significant effect; $t(38) = 3.51$, $p<.001$ where participants in the bimodal condition recalled a higher number of phrases ($M = 3.45$, $SD = .76$) than the unimodal condition ($M = 2.50$, $SD = .95$) (See Figure 2, upper graph). As may be seen in Figure 2, participants in the bimodal condition rated the presence, enjoyment, and external awareness item (focused) item significantly higher than participants in the unimodal condition.

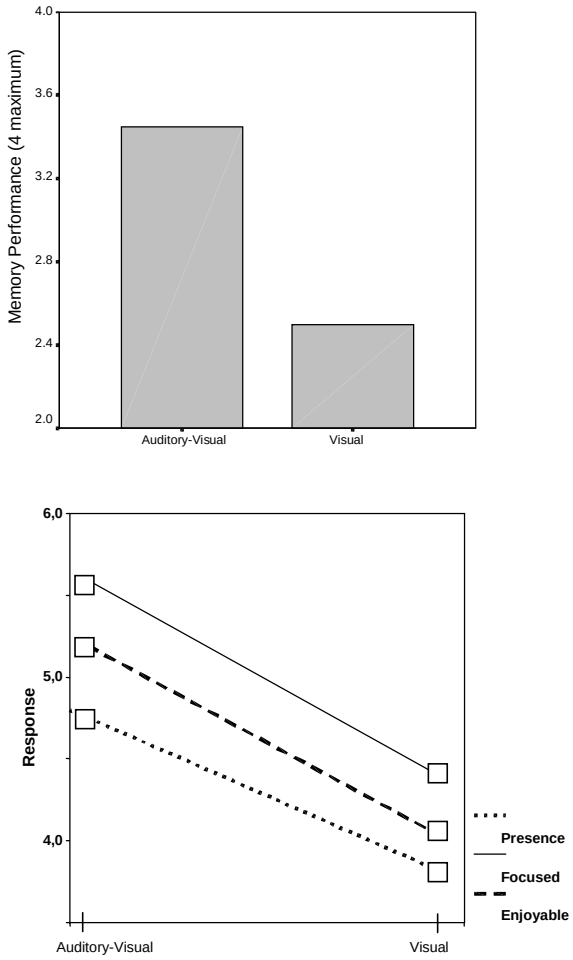

Figure 2. *The upper graph displays participants mean memory performance in the congruent task for auditory-visual and auditory condition. The lower graph shows mean ratings of presence, focused and enjoyment items by participants in the auditory-visual and auditory condition.*

In conclusion, it was found that participants receiving bimodal information experienced a higher degree of presence, were more focused on the situation and enjoyed the VR experience more than participants receiving only unimodal (visual) information did. Furthermore, bimodal processing of information significantly improved memory and task performance as compared to unimodal processing of the same information.

3. EXPERIMENT 2. ROOM ACOUSTICS PERCEPTION IN REAL AND VIRTUAL ENVIRONMENTS.

Experiment 2 aimed at studying the joint effect of visual and auditory information on ratings of room acoustic qualities. In order to do so, bimodal conditions (sound and visual input) were contrasted with unimodal condition (sound only). Moreover, simple pictorial reproduction is contrasted with Virtual Reality models (VRML) of rooms and actual experiences of the same rooms. It was hypothesized that an increasing level of visual realism and presence would significantly affect judgments of aural qualities.

3.1. Method

80 undergraduates were assigned to one of four conditions:

1) Participants rated the sounds only *(Sound condition)*

2) Participants rated the sounds as when viewing still pictures taken of the room *(Picture condition)*

3) Participants navigated in a virtual model of the rooms while rating the sounds *(VR condition)*

4) Participants rated the sounds replayed over headphones on location in the rooms *(Real condition)*

A between-subjects design was used. The visual stimuli were virtual (photographs or VRML-models) or real concert halls, theaters and practice rooms in Musikhögskolan in Gothenburg, Sweden (See Figure 3). The auditory stimuli were auralizations of these rooms made with CATT-Acoustic [12].

Figure 3. *Photograph and Screenshot from the concert hall used in experiment 2.*

3.2. Measures

Participants rated each sound with respect to a number of adjectives that previously have been found to be susceptible to cross-modal influences [9],[10],[15],[16],[17]. The adjectives were: auditory source width (ASW), aurally perceived room size, and aurally perceived distance to sound source.

3.3. Results

As may be seen in Figure 4, highly significant differences between the different conditions in ratings of Auditory source width (ASW) were obtained. Following the significant main effect obtained in the ANOVA, Tukey post hoc-test showed that both the VR and reality conditions were significantly different from the sound and sound and picture condition. The analysis of ratings of distance to sound source and perceived room size also showed highly significant differences between the different conditions. However, Tukey post-hoc tests showed that these differences are accounted for by the reality condition that deviates significantly from the three other conditions.

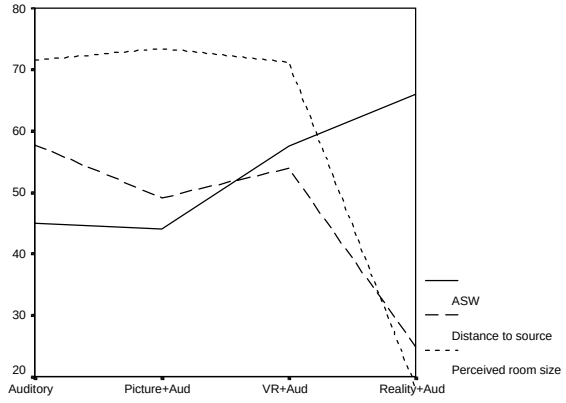

Figure 4. *Mean ratings of auditory source width, aurally estimated distance to sound source, and aurally perceived room size by participants in auditory, picture and auditory, VR and auditory, and Reality and auditory conditions.*

4. GENERAL DISCUSSION

The current research argued that an ecological approach to perception of VE's would entail rendering of complex information (in our case visual and aural rendering of rooms with realistic aural and visual properties) as it is in real life. However, little is known about the complexities of the interplay between senses why vision and auditory perception often have been considered separately [1]. Previously, mainly efforts have been made to realistically render visual information in VE's, but lately multimodal VE's have gained popularity. The current research sought out to study the interplay between visual and auditory perception of VE's using sophisticated algorithms for realistic aural rendering of rooms. The results from Experiment 1 showed that combination of visual and auditory information in a VE improved sense of presence and enjoyment in contrast to only visual information and thus supports the use of multimodal systems. Moreover, the results showed that performance and memory was better in bimodal as compared to unimodal perception of information. Experiment 2 showed that that the perceived auditory quality of a room is affected by visual information and increasing visual realism. Analogously to the results from research on audio-visual home theatre systems and studies of ecological perception, these results thus suggest that the quality of a VE may be improved by adding information from other modalities than vision [18], [19], [20]. A question then is if a higher degree of realism or complexity of auditory virtualization in multimodal VE can further improve presence and task performance? It is reasonable to assume that a higher level of auditory virtualization will improve presence, especially in virtual soundscapes containing multiple and moving sound sources. Future research should therefore address the subjective impact of improved auditory virtualization in multimodal VE's by the use of for instance fast real-time auralization methods and realistic room acoustic representation.

5. REFERENCES

[1] J. Gibson, *The ecological approach to visual perception*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1979.

[2] W. W. Gaver, "How do we hear in the world? Explorations in ecological acoustics" *Ecological Psychology*, vol. 5, pp. 285-313, 1993.

[3] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory event perception." *Ecological Psychology*, vol. 5, pp. 1-29, 1993.

[4] M. Slater and S. Wilber, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments" *Presence: Teleoperators and Virtual Environments*, vol. 6, pp. 603-616, 1997.

[5] J. Blauert *et al*, "An interactive virtual-environment generator for psychoacoustic research. I: Architecture and implementation", *Acta Acustica*, vol. 86, pp. 94-102, 2000

[6] S. D. Lipscomb, "Cross-modal integration: Synchronization of auditory and visual components in simple and complex media", in *proceedings of Forum Acusticum*, Berlin, Germany 1999.

[7] B. E. Stein and M. A. Meredith, *The merging of the senses*, MIT press, Cambridge, MA, 1997.

[8] J. J. MacDonald and L. M. Ward (2000). "Involuntary listening aids seeing: Evidence from human electrophysiology", *Psychological Science*, vol. 11, pp 167-171, 2000.

[9] D. Västfjäll et al "Auditory-visual interaction in virtual room acoustics", *submitted for publication*, 2000.

[10] D. Västfjäll et al., "Auralization and visualization in merging: Cross-modal effects in room acoustic perception", *submitted for publication*, 2001

[11] M. I Possner, M. J. Nissen and R. M. Klein, "Visual dominance: An information-processing account of its origins and significance". *Psychological Review*, vol. 83, pp. 157-171, 1976.

[12] CATT, Mariagatan 16A, SE-41471 Gothenburg, Sweden, info@catt.se, http://www.catt.se

[13] EON Reality Inc., 31 W. Tech. Drive, Suite 150B, Irvine CA, sales@eonrality.com, http://www.eonreality.com

[14] Lake Technology Limited, G.P.O. Box 736, Broadway Post Office, NSW 2007, Australia, info@lake.com.au, http://www.lakedsp.com

[15] P. Larsson *et al*, *Subjective testing of the performance of reverberation enhancement using virtual environments.*

Report F 99-05, Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden 1999.

[16] P. Larsson, D. Västfjäll and M. Kleiner. ”The actor-observer effect in virtual reality presentations”, *CyberPsychology & Behavior*, vol. 4, no. 2, 239-247, 2001.

[17] P. Larsson, D. Västfjäll and M. Kleiner, “Auditory-visual interaction in Virtual reality: Auditory cues improve memory and presence in virtual environments” (submitted for publication).

[18] S. Bech, V. Hansen and W. Woszczyk, “Interactions between audio-visual factors in a home theater system: Experimental results”, *99th AES convention*, New York, NY, October 1995, Preprint No. 4096.

[19] C. Hendrix and W. Barfield, “The sense of presence within auditory virtual environments”, *Presence: Teleoperators and Virtual Environments*, vol. 5, 290-301, 1996.

[20] W. Woszczyk, S. Bech and V. Hansen, “Interactions between audio-visual factors in a home theater system: Definition of subjective attributes”, *99th AES convention*, New York, NY, 1995 October 6-9, Preprint No. 4133