AUDITORY-VISUAL INTERACTION IN REAL AND VIRTUAL ROOMS

PACS REFERENCE: 43.55.Hy, 43.55.Ka, 43.66.Ba

Larsson, Pontus ; Västfjäll, Daniel ; Kleiner, Mendel
Department of Applied Acoustics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
+46 (0) 31 772 2192
+46 (0) 31 772 2212
pontus.larsson@ta.chalmers.se

ABSTRACT

In this paper, we discuss some of the perceived auditory and visual qualities of rooms that in previous experiments have been shown to give rise to cross-modal effects. Moreover, the issue of simulation fidelity when using virtual reality systems to perform cross-modal interaction is discussed and an experiment where room acoustical qualities were assessed in different visual conditions is reviewed. The conclusions are that one can expect auditory-visual interaction in the perception of both real and virtual rooms, but that simulation fidelity might affect the results when performing experiments in virtual conditions.

INTRODUCTION

To be able to consistently and objectively measure room acoustic qualities and to find correlations between such objective measurements and subjective experience has for long been a goal of those involved with architectural acoustics and psychoacoustics (Beranek 1992). By utilizing such relations between perceptual factors and objective indices, the acoustician may be able to predict and objectively measure the quality of a certain hall or auditorium. However, establishing these relations appears rather complex due to the many factors that influence the subjective judgment. Current theoretical and empirical findings indicate that perceptual information from one sense (i.e. vision) influences evaluation and perception of information in other senses (i.e. hearing) (Stein & Meredith, 1997). Examples of such "cognitive associations" between perceptual modalities can easily be found in everyday situations such as when information from visual senses ("this room looks like a typical concert hall") and auditory senses ("this room sounds like a typical concert hall") combine to form a total evaluation of the situation. It is thus reasonable to assume that these cross-modal interactions affect subjective experience, evaluations, and judgments of quality of rooms. However, little research have addressed the interaction of information from different senses within this specific area, but rather studied them separately.

Computer auralization and virtual reality technology has brought means to efficiently study cross-modal interaction phenomena. However, it is questionable at what extent the results from such studies agree with real, unmediated situations (Larsson et al., 2000). Can one for example assume that cross-modal effects that arise in mediated systems, perhaps lacking certain characteristics in one or more modalities, are the same as those that would occur in reality?

The current article addresses cross-modal interaction in room acoustic perception and proposes an account of cross modal phenomena in room acoustic perception. In relation to this, the issue of simulation fidelity in virtual environments and presence as a quality measure is also discussed.

AUDITORY-VISUAL INTERACTION IN ROOMS

Although very few studies have systematically tried to establish auditory and visual room factors susceptible to cross-modal interaction effects, there seems to be a general consensus among acoustic scholars that such effects do exist. Egan (1998) means for example that visual factors affecting the acoustical success of concert halls are color (e.g. conductors seem to prefer white and gold interiors to blue interiors) and use of materials (e.g. musicians believe that wood is essential for a concert hall to sound well). Another acoustician means that one of the components of both perceived loudness and acoustical intimacy is visual in nature (Hyde, J. R., personal communication, January 25, 2002). Ando's (1996) investigations of correlation between lightning and preferred listening level and time delay between direct sound and one single reflection in a room to some degree confirms this hypothesis. The results from this experiment indicate that listeners prefer lower sound pressure levels at high illumination. On the contrary, the preferred time delay was highly independent of the lighting levels. Ando explained this with the fact that both listening level and lighting involve the right cerebral hemispheric dominance, while changes in time delay involves the left-hemispheric dominance and are thus less likely to give rise to interaction effects.

In a study performed by Hidaka and Beranek (2000) it was suggested that discrepancies between measured ($[1\text{-}IACC_{E3}]$) and subjectively rated spaciousness in a particular opera house could be explained by the beauty of the opera house. Obviously, aesthetics is a complex and hard-to-measure visual characteristic of any visual object or space. Nonetheless, this and similar visual factors' influence on spaciousness and other acoustical qualities could probably be investigated by applying the theories developed within architectural theory and visual perception (Hesselgren, 1954).

Auralization and virtual environments

Results from research on simulation of room acoustics (auralization) provide further examples of cross-modal effects in rooms. Nataniel et. al. (1997) used visualization and auralization to investigate the visual influence on auditory distance perception in a concert hall. The results in this study showed that auditory distance judgements are lower when the concert hall stage was visually close and higher when it appeared visually distant. In general, an often unwanted characteristic of such virtual environments (VEs), is that visual and aural impressions are unmatched (for example, "this room sounds much bigger than what it looks like"). Larsson et. al. (2002) presented a hypothesis of how such unmatched impressions are perceived. 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, for example, the size of the room on both the visual and aural impression and previous experience of how other rooms looked and sounded like. Larsson et. al. hypothesize that in the case when unmatched auditory and visual stimuli are received, it is likely that the visual impression will dominate the perception. In support of this it was found that subjects both seeing and hearing a concert hall gave a significantly more accurate estimate of the room size than did subjects who only heard the same room. Related to these findings are studies of reverberation enhancement systems. In rooms equipped with such systems, auditory and visual stimuli may also may be unmatched if, for example, the reverberation unit is set to produce unrealistically long reverberation time. It has been found that the degree of mismatch between aurally and visually perceived room size is often mentioned as being one of the factors influencing the acceptance of such a system (Svensson, 1994)

Architectural theories

Auditory-visual perception of rooms has also been addressed by architectural scholars to some extent. Hesselgren (1954), for example, states that "If the visual perception does not confirm the meaning of the auditory perception (for example; long reverberation means big room), one often feels this as a annoying disharmony in the total experience of the room" (p. 203, translated from Swedish). Another architectural scholar, Dyrssen (1998) means that the total experience (of music and architecture) can be reached both by "[…] being inside the music and by being inside a physical space". Dyrssen continues with saying that "To be inside a whole where details, material, color, shape, dimensions, sound and lines agree and support each other is an experience of harmony – which does not necessarily mean lack of tension, conflict or dynamics […]" (p. 20, translated from Swedish). According to Dyrssen's analysis, the musical content's suitability for a certain hall is thus strongly related to the overall experience. It might therefore be necessary to include or at least be aware of such relationships between musical and architectural form in further studies of cross-modal interaction effects in rooms.

Connections to basic research on auditory-visual interaction

In addition to the above-mentioned experiences of acoustic and architectural scholars, one could also find support for cross-modal interaction phenomena in rooms within the field of experimental psychology. For an overview of basic research on auditory-visual interaction effects, see e.g. Kohlrausch & van de Par (1999). The research performed within this area focus mainly on the perception of direct sound and in large ignores the influence of the room context. However, it is likely that these findings and theories to some extent also can be transferred to either the virtual or the real room context. It is for example likely that the often referred to "ventriloquist effect" can be applied to the perception of sound sources in rooms.

INFLUENCES OF VISUAL SIMULATION FIDELITY ON ROOM ACOUSTIC PERCEPTION

Auralization and visualization technologies clearly provide efficient means to perform studies of cross-modal interaction phenomena in rooms, since they allow visual and auditory parameters to be changed easily. Nonetheless, the reliability of subjective experiments conducted by the use of such mediation systems is most likely contingent on the quality of the medium (Larsson et. al., 2000). Västfjäll et. al. (2002) performed an experiment with the aim of 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 was 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. In the study, 80 undergraduates were assigned to one of four conditions in a between-groups design: 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). The visual stimuli were virtual (photographs or VRML-models viewed on a standard PC-monitor) or real concert halls, theaters and practice rooms in Musikhögskolan in Gothenburg, Sweden (See Figure 1). The auditory stimuli were auralizations of these rooms created with CATT-Acoustic.

Figure 1. Photograph and Screenshot from one of the rooms used in the experiment. (Västfjäll et. al. 2002)

Participants rated each sound with respect to a number of adjectives that previously have been found to be susceptible to cross-modal influences (Nataniel et. al., 1997, Larsson et. al 2002). The adjectives were: Auditory Source Width (ASW), aurally perceived room size, and aurally perceived distance to sound source.

The results from this experiment (see Figure 2) showed that different visual conditions influenced the ratings of ASW. Statistical contrasts 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, that these differences were accounted for by the reality condition that deviates significantly from the three other conditions. The results from this experiment thus support the hypothesis that cross-modal interaction in rooms is contingent on the quality of the stimuli.

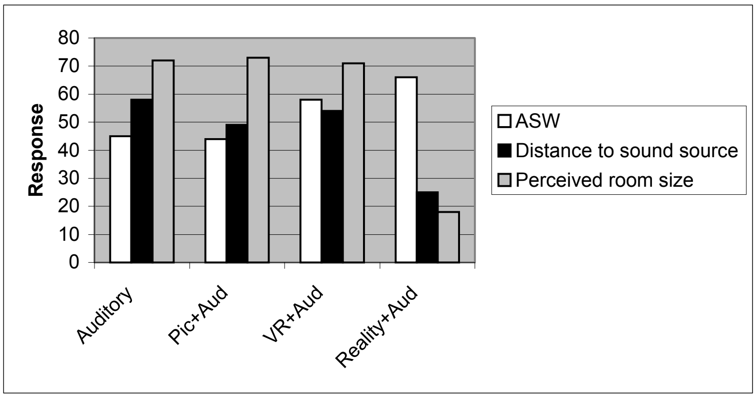

Figure 2. Mean ratings of auditory source width, aurally estimated distance to sound source, and aurally perceived room size by participants in auditory, picture and auditory, VR and auditory, and Reality and auditory conditions. (Västfjäll et. al. 2002)

SIMULATION FIDELITY AND PRESENCE IN VIRTUAL ENVIRONMENTS

Within the area of Virtual Reality, a lot of effort is put on trying to establish factors that define simulation quality or fidelity and to develop measures of simulation fidelity. Such a measure is the sense of presence (generally described as “the sense of being there” (Lombard & Ditton, 1998) experienced within the environment. Several methods of how to carry out presence measurements have been suggested (Larsson et. al., 2001, Freeman et. al. 2000).

Nonetheless, presence in VEs has mainly been related to spatialized sounds, i.e. rendering of direct sound only (Barfield & Hendrix, 1996). Barfield and Hendrix found that adding spatialized sound to a visual VE increased presence but did not increase the rated realism of the VE. This could probably be explained by the semantic load on the word "realism" as it might be associated with "visual realism" and not realism in general. Furthermore, it is unclear to the authors of the current article whether spatialization is a characteristic fundamental in contributing to the sense of realism. In an anechoic of semi-anechoic this might be the case whereas in a concert hall or a church, proper rendering of reverberation is probably more important.

Larsson et. al. (2002) performed an experiment aimed at investigating the aural simulation quality on the sense of presence where subjects experienced an auditory-visual virtual model of a church. In the experiment, subjects were assigned to either a low-quality auralization condition (the auditory model was a generic, shoe-box shaped hall with shorter reverberation time than the actual church) or high-quality auralization condition (the auditory model was based on auralizations of the actual church and thus matched with the visual impression). The study showed that subjects in the high auralization condition experienced a higher degree of presence as compared to subjects in the low-auralization condition. Moreover, subjects in the high-auralization condition experienced a significantly higher degree of sound localization and that the sound contributed more to the total experience. This indicates that presence indeed is a good overall measure of simulation fidelity.

DISCUSSION

A conclusion of the findings reviewed above is that one can expect auditory-visual interaction in the perception of both real and virtual rooms. As of yet, no one has tried to systematically organize and characterize these effects in order to be able to get a more global understanding of the perception of rooms. Rather, the approach taken in previous research has mainly been exploratory in nature. It is suggested that more detailed experiments on single perceptual dimension such as auditory room size or perceived reverberation time is performed. Nonetheless, it is at the same time important to identify relevant perceptual parameters and their structure and cause. Future research could take either of these approaches to study cross-modal interaction in room acoustics. It is important to acknowledge the relevance of both approaches in giving new insights in human perception. It may be argued that the semantic tests performed in this kind of research taps cognitive evaluations of incoming information, why it is important to also study affective reactions to multi-modal information, by using behavioral or physiological indicators.

When performing subjective tests on room acoustic qualities or on auditory-visual interaction, computer auralization and visualization techniques might be advantageous compared to performing the test in a real situation. In some cases it might even be necessary to perform the test by using some kind of mediation system. However, the research reviewed above indicates that the quality or simulation fidelity of the VE affects the ratings of acoustic attributes. To be able to draw any conclusions from tests performed with virtualization techniques it may therefore be necessary to develop and use reliable simulation fidelity measures in parallel. To date, several simulation fidelity and presence questionnaire exist. The use of such questionnaires in room acoustic quality experiments may lead to an increased understanding of the interrelation between simulation fidelity and auditory/visual quality and interaction. This would also give new insights on the influence of virtualization parameters (such as rendering quality and scene content) thus providing a basis for novel Virtual Reality technologies to be developed.

ACKNOWLEDGEMENTS

This article was written under the auspices of the EC ORSEV project - The Organ as a Symbol of the European Vision (Raphael Programme).

REFERENCES

Ando, Y. (1998). *Architectural acoustics*. New York, NY: Springer-Verlag.

Beranek, L. L. (1992). Concert hall acoustics - 1992. *Journal of Acoustical Society of America*, 92(1), 1-39.

Dyrssen, C. (1995). *Musical space - metaphors, rituals, institutions. A cultural study of architecture within and around music (in Swedish)*. Unpublished ph. d.-thesis, Chalmers University of Technology, Göteborg.

Egan, M. D. (1998). *Architectural acoustics*. New York, NY: McGraw-Hill, Inc.

Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & Ijsselsteijn, W. I. (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. *Presence-Teleoperators and Virtual Environments*, 9(2), 149-164.

Hendrix, C., & Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence-Teleoperators and Virtual Environments*, 5(3), 290-301.

Hesselgren, S. (1954). *Arkitekturens uttrycksmedel (in Swedish)*. Stockholm: Almqvist & Wiksell / Gebers Förlag AB.

Hidaka, T., & Beranek, L. L. (2000). Objective and subjective evaluations of twenty-three opera houses in Europe, Japan, and the Americas. *Journal of Acoustical Society of America*, 107(1), 368-383.

Kohlrausch, A., & van de Par, S. (1999). Auditory-Visual Interaction: From Fundamental Research in Cognitive Psychology to (possible) Applications. In *proceedings of SPIE, Human Vision and Electronic Imaging IV*, San Jose, California, January 25-28.

Larsson, P., Västfjäll, D., & Kleiner, M. (2002). Auditory-Visual Interaction in Virtual Room Acoustics. *In preparation.*

Larsson, P., Västfjäll, D., & Kleiner, M. (2002). Better presence and performance in virtual environment by enhanced binaural sound rendering. In *proceedings of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, Espoo, Finland, June 15-17, 2002.

Larsson, P., Västfjäll, D., & Kleiner, M. (2001). The actor-observer effect in virtual reality presentations. *CyberPsychology & Behavior*, 4(2), 239-246.

Larsson, P., Olsson, C., Västfjäll, D., & Kleiner, M. (2000). *Subjective testing of reverberation enhancement using Virtual Environments*. Göteborg, Sweden: Report F99-05, Department of Applied Acoustics, Chalmers University of Technology.

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer Mediated Communication*, 3(2).

Nathanail, C., Lavandier, C., Polack, J. D., & Warusfel, O. (1997). Influence of the visual information on auditory perception. Consequences on the subjective and objective characterisation of room acoustic quality. In *proceedings of the International symposium on simulation, visualization and auralization for acoustic research and education*, Tokyo, Japan, April 2-4.

Stein, B. E., & Meredith, M. A. (1997). *The merging of the senses*. Cambridge, MA: MIT press.

Svensson, P. (1994). *On reverberation enhancement in auditoria*. Unpublished ph. d. -thesis, Chalmers University of Technology, Göteborg.

Västfjäll, D., Carlsson, M., Larsson, P., Kleiner, M., & Dalenbäck, B-I. (2002): Visualization and auralization in merging: Cross-modal effects in room acoustic perception. *Submitted to Journal of Acoustical Society of America.*