Experiments on audio-visual room perception: A methodological discussion

Hans-Joachim Maempel[1]

[1] *TU Berlin, Audio communication group, 10587 Berlin, E-Mail: hans-joachim.maempel@tu-berlin.de*

Introduction

Against the background of the growing audio-visual integration and convergence of media, information about the interaction of hearing and seeing gain importance. Cross-modal interaction denotes the phenomenon that one sense influences the modality-specific percept of another and/or the corporate modality-unspecific percept. Studies show that audio-visual interaction effects apply all stages of the perception process.

Interaction effects regarding localization have been studied relatively well, and it stands to reason, that there is also an audio-visual interaction for aspects of spatiality [1]. However, there are few studies dealing particularly with the audio-visual perception of room acoustics, e.g. [2] [3] [4] [5] [6] [7] [8]. Omitting a detailed review in the paper on hand, a synopsis indicates that the previous studies may be characterized by the consideration of few – often highly specific – independent and dependent variables, by the application of numerical room models, by the generation of artificial sound fields using static binaural synthesis or loudspeakers in an anechoic room, and by the application of different empirical paradigms and methods. Apparently this contributes to that results lack consistency and connectivity.

Methodological considerations

The following considerations might contribute to the disentanglement of experimental strategies towards the experimental investigation of audio-visual interaction effects.

1. It must be acknowledged, that there is little previous knowledge about the topic, suggesting a funnel-shaped research strategy leading from the general to the specific. Thus, a largely integrated exploratory approach emphasizing ecological validity appears to be reasonable.

2. Rooms are auditorily perceived as a transmission system rather than as an object. Thus, the auditory perception of a room always depends on sonic events that bear a meaning themselves. Consequently, stimulus content should be diversified, as also suggested in [8].

3. Occasionally, a general or preponderant visual dominance is assumed a priori [8]. However, the contribution of a sense to the characteristics of a percept is not only a function of the specific modality itself, but also of the current perceptual task on a given condition and at a specific stimulus configuration. Hence, opto-acoustically symmetric test designs are required, and unisensory or respectively unimodal characteristics should be clearly distinguished from intersensory or respectively intermodal characteristics.

4. A very clear distinction between physical and psychological measures adds to avoid the premature application of well-known psychoacoustic associations to the special condition of audio-visual perception. This distinction may be done terminologically by the strict use of the terms optical/ acoustical for the physical realm (stimuli) and auditory/visual for the psychological realm (percepts), and empirically by collecting separate data in the respective realm.

5. The contribution (i.e. the explained variance) of the optical and the acoustical stimulus component to a perceptual effect cannot be quantified in general (i.e. independently of their specific levels), as long as the commensurability of the factors is not assured, i.e. each component is not measured in the same physical unit. Optical and acoustical measures usually are not quantifiable by the use of identical units, whereas structural measures (e.g. room size) or modality-unspecific measures (e.g. the overall aesthetic judgment) are.

6. An absolute quantification of audio-visual interaction effects (e.g. the determination of a point of subjective equality) demands a perfect mutual match of the optical and acoustical characteristics of the investigated rooms. The degree of this match cannot be numeralized for modeled rooms. This is a substantial argument in favor of using data-based (existing) rooms.

7. Frequently, the presence or absence of the optical and acoustical stimulus component is systematically varied in order to determine the contribution of the domains to the quality of the percept. However, results based on this co-presence paradigm are only valid for the particular relation of the optical and acoustic *characteristics* under test, usually matching each other. In order to give evidence of the *inter*-modality of room perception, an opto-acoustic variance *between* the respective *characteristics* needs to be generated, known in the neuro- and cognitive sciences as conflicting stimulus paradigm. Depending on their grade of divergence and acuity, conflicting stimuli can lead to fused, ambiguous or segregated percepts or perceptual streams, respectively.

Proposed Method

The subproject 9 "Audio-visual perception of acoustical environments" within the framework of the SEACEN project [9] attempted to come up to the above-mentioned methodological requirements.

Table 1 shows a taxonomy of the involved variables and thereby illustrates the transmission of room characteristics from the physical realm to the psychological realm, temporarily split into two domains or modalities, respectively. According to the integrated approach, all variables are collected in one extensive experiment containing different rooms and test participants.

Table 2 shows the test design integrating both the co-presence paradigm (colored cells) and the conflicting stimulus paradigm (uncolored and green cells). Due to the high number of cells, data collection for independent samples would not be practicable. For this reason, dependent samples are drawn, and the number of rooms under test is limited to six. The characters within the cells indicate the modalities of the perceptual measures: Within the co-presence paradigm, collecting visual or auditory measures would not make sense under the acoustical and optical condition, respectively.

Table 1: Taxonomy of variables

Physical characteristics		**Psychological characteristics**	
opto-acoustic	**acoustical**	**auditory**	**audio-visual**
structural (room geometry room dimensions, source dimensions, source position) material (surface)	room acoustic parameters (RT, EDT, G, C80, LF, BR, IACC, …)	loudness timbre spatiality transparency balance intelligibility …	structural representation (room geometry, room dimensions, source dimensions, source localization) material representation (surface) matching presence aesthetics emotion
	optical	**visual**	
	color space (H, S, V)	brightness contrast hue colorfulnesss …	

The six rooms under test were selected in view of the variation of both the volume as a numeric measure for their primary structure (three levels) and the average absorption coefficient as a numeric measure for their secondary structure (two levels), resulting a large range of reverberation times.

Table 2: Test design. O=only optical, A=only acoustical, v=visual, av=audio-visual, a=auditory

Perceptual measures		Factor I: Acoustical room						
		O	1	2	3	4	5	6
Factor II: Optical room	A		a av	a av	a av	a av	a av	a av
	1	av v	a av v	a av v	a av v	a av v	a av v	a av v
	2	av v	a av v	a av v	a av v	a av v	a av v	a av v
	3	av v	a av v	a av v	a av v	a av v	a av v	a av v
	4	av v	a av v	a av v	a av v	a av v	a av v	a av v
	5	av v	a av v	a av v	a av v	a av v	a av v	a av v
	6	av v	a av v	a av v	a av v	a av v	a av v	a av v

Since the conflicting stimulus paradigm cannot be realized in natural rooms, the rooms have to be simulated, thereby providing as many physical cues as possible. Thus, the optical simulation combines a 180° field of view with stereoscopy, a high resolution and a cylindrical projection screen; the acoustical simulation is realized by the use of the SEACEN reference reproducing system applying dynamic binaural synthesis, extra-aural headphones and individual ITD correction. The respective stereoscopic panoramas and directional BRIRs are acquired in situ. The design demands a dissociation of the varied room characteristics and the performance. This becomes practicable by an anechoic audio recording to be convolved with the BRIRs and a greenbox video recording to be composited with the panoramas. With respect to the required ecological validity and the variation of content, a speech and a string quartet performance of high artistic quality are produced. The resulting 3D audio and video streams serve as stimuli presented to each test participant individually.

Acknowledgements

This work was carried out within the framework of the SEACEN project and funded by the German Research Foundation (DFG MA 4343/1-1).

References

[1] Kohlrausch, A.; van de Par, S.: „Audio-Visual Interaction in the Context of Multi-Media Applications“. In: J. Blauert (Hg.) *Communication Acoustics.* Springer, Berlin et al., 2005, S. 109-138.

[2] Cabrera, D.; Nguyen, A.; Choi, Y. J.: “Auditory versus visual spatial impression: a study of two auditoria”. In: *Proceedings of ICAD 2004 - Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004.*

[3] McCreery, A.; Calamia, P. T.: “Cross modal perception of room acoustics (A)”. In: *J. Acoust. Soc. Am.* 120 (2006), p. 3150, Abstract.

[4] Larsson, P.; Väljamäe, A.: “Auditory-visual perception of room size in virtual environments”. In: *Proceedings of the 19th International Congress on Acoustics, Madrid, 2-7 September 2007*, PPA-03-001.

[5] Sandvad, J.: “Auditory perception of reverberant surroundings. In: *J. of the Acoustical Society of America* 105/2 Pt. 2 (1999), p. 1193.

[6] Valente, D. L.; Braasch, J.: “Subjective expectations adjustments of early-to-late reverberant energy ratio and reverberation time to match environmental cues of a musical performance.” In: *J. Acoustica Special Issue: Virtual Acoustics* 94 (2008), p. 840-855.

[7] Vorländer, M.; Abbou-Elleal, E.: “Simultaneous visual and acoustical perception in room acoustics.” In: *Techni Acustica, Bilbao, 2003.*

[8] Larsson, P.; Västfjäll, D.; Kleiner M.: “Auditory-visual Interaction in Real and Virtual Rooms. In*: 3rd Convention of the European Acoustics Association, 2002 Sept. 16-21, Sevilla, Spain.*

[9] Weinzierl, S.; Lindau, A.; Brandenburg, K.; de Vries, D.; Maempel, H.-J.; van de Par, S.; Rafaely, B; Spors, S.; Vorländer, M.: „The SEACEN project“. In: *DAGA 2012, Darmstadt.*