Toronto, Canada
International Symposium on Room Acoustics
2013 June 9-11

Audio-visual interaction of size and distance perception in concert halls – a preliminary study

Hans-Joachim Maempel (maempel@sim.spk-berlin.de)
Department III for Acoustics and Music Technology / Studio Technology and IT
Staatliches Institut für Musikforschung Preußischer Kulturbesitz (SIMPK)
Tiergartenstr. 1
10785 Berlin, Germany

Matthias Jentsch (matthiasjentsch@web.de)
Audio communication group
Technische Universität Berlin
Einsteinufer 17c
10587 Berlin, Germany

ABSTRACT

The perception of rooms involves various unimodal and multimodal aspects on different perceptual levels. Rather abstract yet self-evident aspects are the source distance and the room size. We investigated to what extent the perceived room size and egocentric source distance as supramodal aspects are based on the auditory and the visual modality, i.e. experimentally influenced by the acoustic and optical stimulus. The statistical determination of the respective contributions demands the mutually independent variation of optical and acoustical room properties, usually referred to as conflicting stimulus paradigm. Simulation data of four rooms were collected acoustically by acquiring binaural room impulse responses for different head orientations and optically by acquiring stereoscopic images of the rooms including the electroacoustic sound source. In the laboratory, respective acoustic scenes were played back applying dynamic binaural synthesis, whereas the optical scenes were presented by the use of a stereoscopic display. Test participants were asked to assess the source distance and the room size. Results show main effects of the modalities rather than an interaction effect. It was found that distance perception in rooms is predominantly based on the acoustic stimulus characteristics whereas room size perception predominantly relies on optical information. There is no evidence in favour of the proximity effect hypothesis. Results do also not confirm an auditory or audio-visual underestimation of the source distance but show a general underestimation of room size and indicate an audio-visual asymmetry regarding the accuracy of room size perception. Maximum accuracy and cross-modal consistency of judgements were distinctively observed in low-absorbent rooms.

1 INTRODUCTION

1.1 Subject

When perceiving opto-acoustic scenes, e.g. musical renditions in concert halls, various auditory, visual and multimodal properties may emerge to or be focused on by the recipient on different perceptual or cognitive levels, i.e. on different degrees of abstraction. Loudness, brightness, audio-visual matching, and valence are examples of an auditory, a visual, an intermodal and a supramodal quality, respectively. A general question asks, to what extent are these properties

based on auditory and visual information, i.e. influenced by the acoustic and optical component of a stimulus? This paper reports on a preliminary study that was performed in order to tentatively apply the experimental method and the technical setup of a subproject within the framework of the SEACEN research unit and thereby focuses on the perceived source distance and the perceived room size as supramodal properties. The study aims (1) to quantify the proportionate contribution of hearing and seeing to size and distance perception, (2) to compare the inclusion of hearing and seeing with respect to these perceptual properties, and (3) to reveal and to quantify a potential audio-visual asymmetry in size and distance perception. As a by-product perceptual estimates may be related to physical values; however initially we had to explore how subjects use one- and three-dimensional scales for reporting the estimated size.

1.2 State of the art

There are several studies dealing with the purely auditory perception of room acoustics. They mostly sought to find physical or technical ('objective') characteristics that are apt to predict perceptual ('subjective') properties of rooms. The perceived room size was found to be predicted to some extent by several room acoustic parameters: the reverberation time RT_{60},[1] the clarity indices for music C_{80} and speech C_{50},[2,3] the definition index D_{50},[3] and the room gain G_{RG}.[4] The characteristics of early reflections are also assumed to play a role.[5] The perception of distance is determined by at least the sound pressure level, the direct-to-reverberant (D/R) energy ratio,[6-8] and the filtering by air absorption.[9] In virtual environments (VEs), auditorily perceived distances less than about 2 m appear to be systematically overestimated, whereas distances greater than 2 m are reported to be underestimated[9-15] and to be perceptually compressed[9,15] with reference to the physical distance.

Studies dealing with the visual perception of distances are primarily focused on the issue of assessment accuracy. The studies showed that accuracy depends on several factors such as the grade of virtuality (reality, augmented reality, virtual reality),[16] and – at least in outdoor environments – the level of cognitive abstraction (perceived, remembered, inferred).[17] Some factors did, however, not show a significant influence on the accuracy of distance perception in VEs such as the restriction of the field of view (FOV)[18] and the focal length[19] of the camera lens. Comparing different measurement protocols it was found that timed imagined walking and verbal estimation yielded similar results.[20] Nevertheless, visual distances are systematically underestimated within VEs no matter if head-mounted displays or large screen immersive displays are applied.[13,20-23]

In the case of audio-visual perception in VEs distances beyond 2.5 m are also reported to be underestimated and compressed.[15] Comparatively, underestimation was found to be greater under the acoustic than under the optical and the opto-acoustic condition.[13] Several studies give evidence to the fact that optical stimuli may influence auditory, and of course intramodal and supramodal properties, thus the perceived distance also depends on the modalities involved.[24,25] Experiments applying conflicting stimuli revealed the phenomenon that localisation being based on the pure acoustic stimulus component may be strongly biased by the position of an optical stimulus component, which is generally referred to as *visual-capture effect*. It occurs when varying the lateral or the distant position. In case of a distance shift it is often referred to as *proximity image effect*: In 1968, Gardner found that in an anechoic environment listeners tend to localise sounds at the position of the nearest visible loudspeaker even when the true sound source was a more distant and invisible loudspeaker.[26] This experiment has been criticised methodically[27] and replications showed the effect to be unstable when providing distinct acoustic distance cues (reflections).[28] Since Sandvad found that most subjects are able to correctly assign photographs of rooms to their respective acoustic stimuli,[29] it stands to reason that there is also an audio-visual interaction effect for aspects of spatiality as also assumed by Kohlrausch and van de Par.[30 (pp. 115–116)] Larsson and Väljamäe observed that the estimated room width is visually under-

estimated and auditorily overestimated whereas the opto-acoustic condition yielded a largely correct estimation.[31] Comparatively, in another experiment the estimated room size depended rather on the category of the optical virtuality (reality versus VE) than on the modalities actually involved.[32]

Obviously, bringing the studies together and extracting general findings is quite difficult due to different – and often highly specific – independent variables, empirical paradigms and methods, auralization concepts (modeled versus data-based), and formats of acoustic reproduction. Against the background of little and specific previous knowledge we consider a research strategy leading from the general to the specific. Thus we are not interested in quantifying specific room acoustic predictors yet, and in order to gain a better insight in the mechanisms of audio-visual perception, we consider neither a comparison between simulation systems nor a comparison between measurement protocols to be urgent at the time being. Rather we sought to focus on the modalities by using a specific virtual environment and a specific measurement protocol only as research tools and by raising at least one more general research question (1.1, objective 1).

2 METHOD

2.1 General methodological considerations

The attempt to experimentally quantify the contribution of the modalities to perceptual and cognitive properties raises several methodological issues.[33] As a prerequisite for a discussion, we distinguish between the physical and psychological realm by applying the terms *optical/acoustic* in order to denote characteristics of the stimuli, and *visual/auditory* in order to denote characteristics of the respective percepts and cognitions. We noticed that even studies explicitly aiming at investigating *audio-visual interaction* usually treat hearing and seeing as levels of factors (e.g. modality, congruency) rather than as factors themselves. At least formally, those designs are not apt to achieve this aim. Just the *cooperation* of seeing and hearing (in the present study with regard to opto-acoustic rooms containing a set of sound sources) may be experimentally investigated by the mutually independent variation of the *presence* of the optical and the equivalent acoustic component of the stimulus. This type of variation, here referred to as co-presence (CP) paradigm, leads to three conditions: acoustic, optical and opto-acoustic. Different rooms may or may not be included in a respective test design. The *interaction* of seeing and hearing is experimentally accessible by the mutually independent variation of the respective *characteristics* of the optical and the equivalent acoustic component of the stimulus. A respective design based on the so-called conflicting stimulus (CS) paradigm includes several opto-acoustic rooms representing different categorical or metrical levels of these characteristics.

Ideally, a general quantification of the proportionate contribution of the auditory and the visual system to perceptual or cognitive properties is desired, e.g. in the form of the respectively explained variance of a dependent variable. A reasonable comparison between the explained variances demands, however, an identic range of variation of both the acoustic and the optical experimental conditions, i.e. quantitatively commensurable independent variables. This again demands a qualitative commensurability of the independent variables, which is not directly given for acoustic and optical characteristics since they are described by different physical quantities, e.g. sound pressure level and illuminance. Commensurability may, however, just be indirectly achieved by orienting the values of the optical and acoustical variables toward the values of commensurable higher order variables, typically material variables (e.g. wall covering and upholstering) or structural variables (e.g. source distance and room dimensions), provided that the optical and the acoustic characteristics are equivalent, i.e. originate from the same room.

Due to complexity and envelopment a mutually independent variation of optical and acoustic room characteristics is only practicable by the use of simulated rooms. The required opto-acoustic equivalence is guaranteed just for real rooms and largely fulfilled just for their data-based simulations. In contrast, its degree is not quantifiable for simulations based on numeric models. Moreover, both the optical and the acoustic simulation are required to preserve as many physical cues as possible. This may be achieved by (a) three-dimensionality, (b) a high resolution, and (c) largely immersive stimuli. There are, however, limitations for the optical component in this preliminary study regarding the field of view (FOV).

2.2 Design

Table 1 shows a test design according to the considerations made above, i.e. dissociating the acoustic and the optical component and integrating both the CP paradigm (shaded cells) and the CS paradigm (white cells and shaded diagonal). Due to the high number of cells just four rooms were included, and dependent samples were drawn, resulting a full factorial repeated measures (RM) design. Row A and column O represent the unimodal acoustic and optical conditions, respectively. The characters within the cells indicate the modalities involved in the perceptual or cognitive measures collected. As supramodal properties, distance and room size perception access both the auditory and the visual modality.

There were 24 audio-visual conditions in total to be experimentally realized. Since the partial designs represent different paradigms and thereby share four conditions (shaded diagonal), in case of a separate analysis of the paradigms twelve conditions had to be analysed for the CP paradigm and 16 conditions for the CS paradigm. With the objective of increasing ecological validity both music and speech were used, thereby introducing a third factor: *content type*. Because the optical conditions were realized by showing an electroacoustic sound source (loudspeaker) instead of a natural source, the music and the speech condition for column O are identical, resulting a total of 44 stimuli to be presented, 32 opto-acoustic stimuli to be analysed according to the CS paradigm, and 20 acoustic, optical and opto-acoustic stimuli according to the CP paradigm.

Table 1: Test design

Audio-visual measures		**Factor I: Acoustic room**				
		O	1	2	3	4
Factor II: Optical room	A		av	av	av	av
	1	av	av	av	av	av
	2	av	av	av	av	av
	3	av	av	av	av	av
	4	av	av	av	av	av

Note that the variation of conditions is complex, thus size, source distance and other physical characteristics of the rooms are confounded. This is inevitable partly as a matter of principle re-

garding the impossible disentanglement of structural characteristics (e.g. room volume) and lower, purely optical or acoustic characteristics (e.g. reverberation time), and partly due to the methodologically required data-based simulations of real rooms. Thus, the confoundation issue has to be accepted as long as a systematic dissociation of the acoustic and the optical stimulus component as well as their commensurability are required.

2.3 Stimuli

Four accessible rooms varying in primary structure (shape and size), numerically described by the room volume V, and in secondary structure (surface structure, surface materials, upholstery), numerically described by the average absorption coefficient $\alpha_{ø}$, were selected. These criteria determine the reverberation time (RT) being an important acoustic predictor for the assessment of the room size.[5] Table 2 shows basic characteristics of the rooms. The rooms' labels are composed of a digit indicating their ascending order with respect to reverberation time, and two characters indicating their characteristics regarding size (small–large) and average absorption (low–high).

Table 2: Basic characteristics of the selected rooms

Label	**1SH**	**2SL**	**3LH**	**4LL**
Name	EN 324	EN 190	H 104	UdK CH
Function	Control room	Seminar room	Lecture hall	Concert hall
Volume V	230 m^3	190 m^3	3300 m^3	4200 m^3
Av. absorption coefficient $\alpha_{ø}$	0.36	0.17	0.28	0.17
Reverberation time RT_{60}	0.36 s	0.71 s	1.07 s	1.79 s
Source-receiver distance d	2.75 m	1.93 m	5.88 m	5.05 m

The audio content was taken from the CD *Music for Archimedes* (Bang & Olufsen 1989) and had been recorded monophonically in an anechoic room with a distance of 1 m. For the speech sample, a sentence (duration 16 s) spoken by a female voice was used, and the music sample was a 16 s clip from a classical piece played by violoncello solo. The different sound pressure levels of the natural sound sources were preserved over the signal chain.

The acoustics of the rooms were acquired by playing back bass-emphasized linear sine sweeps through a 3-way dodecahedron loudspeaker with omnidirectional characteristics positioned at the center of the stage or of the room's front area. At the receiver position, measurement signals were recorded by the in-ear-microphones of the automated, motion-controlled head-and-torso-simulator (HATS) *FABIAN*.[34] Binaural room impulse responses (BRIRs) were measured for different horizontal head orientations with a range of ±80° and an angular resolution of 1°. *FABIAN* was positioned at about two times the critical distance from the sound source. Furthermore, standard room acoustic measures were taken according to DIN EN ISO 3382-1,[35] e.g. RT_{30}, EDT, BR, TS, C_{80}, G.

Perspectively-correct optical information was acquired by taking stereoscopic photographs in rectilinear geometry. They show the room and the dedocahedron loudspeaker from the receiver position (figure 1). The pictures were taken by means of a Pentax K-x reflex camera (f / 22, f = 18 mm). In order to produce an interocular separation the camera was installed on a slide bar. The applied interocular distance was 7 cm.

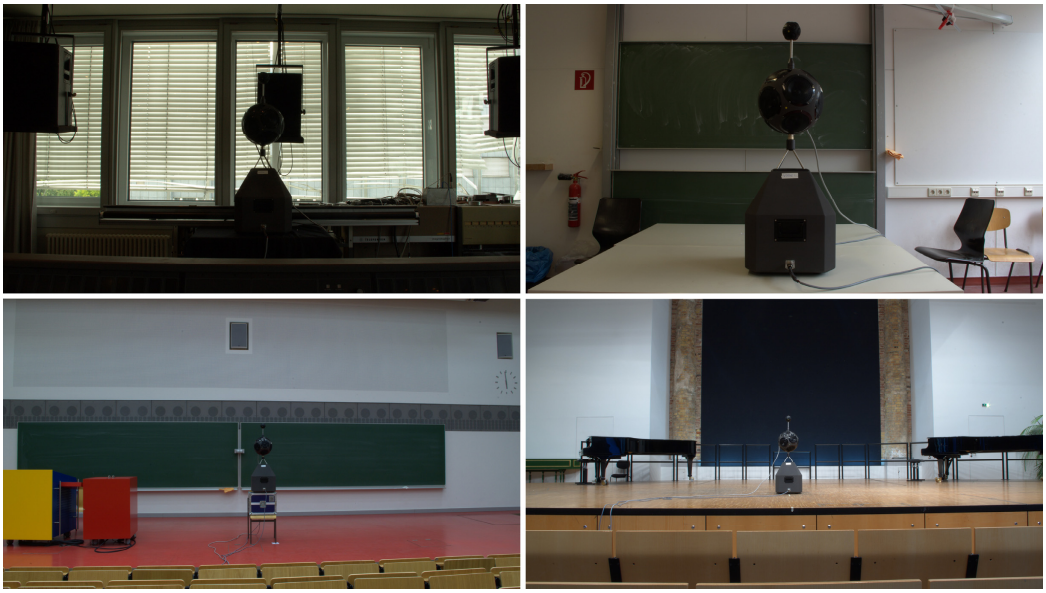

Figure 1: Optical stimuli: dodecahedron speaker in selected rooms (top left: 1SH; top right: 2SL; bottom left: 3LH; bottom right: 4LL)

2.4 Measures

In the rooms geometric measures were taken by means of a laser rangefinder. From the subjects, amongst others, the egocentric source distance and the room size were collected by means of absolute magnitude estimation, i.e. with reference to an internal reference (table 3). While the distance judgment was operationalized by a scale covering a range from 1 to 30 m, the size judgement was operationalized both one- and three-dimensionally. Participants were asked to assess the room volume, here referred to as perceived size 1D (see 3.3), by applying a symmetric seven-grade scale, the poles of which were named *small–large*, and also to separately assess the length, the width and the height by applying scales ranging from 1 to 30 m each.

Table 3: Physical and perceptual measures

Physical Measure	Symbol [unit]	Origin	Perceptual Measure	Symbol [unit]	Origin
Physical distance	d [m]	scenes	Perceived distance	d' [m]	subjects
			Perceived size 1D	s'_{1D}	subjects
Physical length	l [m]	rooms	Perceived length	l' [m]	subjects
Physical width	w [m]	rooms	Perceived width	w' [m]	subjects
Physical height	h [m]	rooms	Perceived height	h' [m]	subjects
Physical volume	V [m^3]	$V=l\cdot w\cdot h$	Perceived volume	V' [m^3]	$V'=l'\cdot w'\cdot h'$
Physical size	s [m]	$s=\sqrt[3]{V}$	Perceived size 3D	s'_{3D} [m]	$s'_{3D}=\sqrt[3]{V'}$

Based on the latter measures the perceived volume (V') and the perceived size 3D were calculated that are expressed in the units m^3 and m, respectively, and therefore may be easily compared with the equivalent physical measures. The issue of distortion of distance perception in terms of Stevens' Power Law is not addressed, since values of the exponent n are not far from 1.[36]

2.5 Sample

The sample consisted of 35 voluntary participants (27 male, 8 female) aged between 23 and 55 years and with normal hearing as per self-report, who were primarily students of the TU Berlin's audio communication master programme.

2.6 Technical Setup

In the experiment the audio content was convolved with the respective BRIR according to the test subject's momentary head orientation being measured by a head tracker. The time variant convolution was performed by the department's *fwonder* application suppressing crossfade artifacts. Furthermore, the applied binaural synthesis system features a sufficient spatial resolution, involves electrostatic headphones (STAX SR-202), compensates for the headphone transfer functions, minimizes system latency to a level below perceptual threshold, and allows for an individual adaption to the subjects' interaural time differences (ITDs). Thus, it provides an high degree of plausibility.[37] The level of the BRIR set of each room was adapted to the respective strength factor *G* in order to preserve the loudness differences between the rooms. The images were presented via a 61" 3D DLP Monitor providing Full HD resolution. Subjects wore shutter glasses due to active stereoscopy. The experimental process was controlled by means of *Pure Data* (*Pd*, core by Miller Puckette) sending Open Sound Control (OSC) messages to the audio-video-player. Presenting the questionnaire and collecting the subject's response data was done by applying *LimeSurvey* (by Carsten Schmitz et al.).

2.7 Procedure

At first subjects were given the opportunity to familiarize themselves with the questionnaire and the experimental process through a testing phase. Then subjects were asked to run through the test sequence and to judge the stimuli using the items provided by the questionnaire. Thereby, stimuli were presented in randomized order.

2.8 Statistical analysis

Initially, in the course of data exploration and cleansing, one case (subject) had been completely deleted due to obviously untrue data. Furthermore, six data points going beyond the x_{95}-x_5 inter-percentile range had been regarded as outliers, thus removed and replaced with mean, and ten missing values had been replaced with mean. Variables for speech and music stimulus responses had been averaged in order to raise both ecological validity and reliability of the resulting aggregate measure. Subsequent analyses comprised plots of mean values for both the CP and the CS paradigm (see 2.2), regression analyses where appropriate, and RM analyses of variance (ANOVA). In case of a violated sphericity assumption the degrees of freedom were conservatively adjusted (Greenhouse-Geisser). In order to allow for different approaches to effect size comparisons, the sample statistics η^2, η_G^2,[38,39] and η_P^2 were mostly specified.[40 (pp. 222–224)]

3 RESULTS

3.1 Perceived source distance

Co-presence paradigm

The main effects of the opto-acoustic CP (levels O, A, OA) and of the room (4 levels), as well as of the respective interaction effect on the perceived distance were tested by means of a full-factorial RM ANOVA. Beforehand, a Kolmogorov-Smirnov (KS) test had indicated that data of two of the twelve distance variables violated the assumption of normally distributed error components. This should, however, not question the *F*-test result since both the main effects and the

interaction effect turned out to be highly significant (opto-acoustic CP: $df = 1.927$, $F = 26.817$, $p < .0005$, $\eta_P^2 = .448$; room: $df = 1.563$, $F = 151.342$, $p < .0005$, $\eta_P^2 = .821$; opto-acoustic CP × room: $df = 4.577$, $F = 30.953$, $p < .0005$, $\eta_P^2 = .259$).

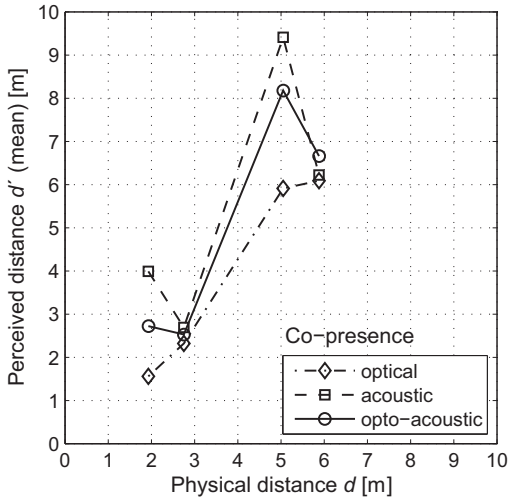

Figure 2: Comparison between perceived and physical distance for different levels of opto-acoustic co-presence

Plotting the mean values of the perceived distance against the physical distance for each opto-acoustic condition (figure 2) indicates that (a) auditorily perceived distances are greater than visually perceived distances, (b) subjects tend to average those distances when perceiving audio-visually, (c) auditorily, visually and audio-visually perceived distances appear to be consistent rather in high-absorbent than in low-absorbent rooms, and (d) subjects tend to auditorily overestimate distances in low-absorbent rooms.

Conflicting stimulus paradigm

In order to examine the main effects of the optical room characteristics (4 levels) and of the acoustic room characteristics (4 levels), as well as of the respective interaction effect on the perceived distance, a full-factorial RM ANOVA was performed. Beforehand, KS tests indicated that the data do not violate the assumption of normally distributed error components.

Table 4: Results of RM ANOVA for *d'*. Sources of Variation are based on the CS paradigm.

Source of Variation	***SS***	df_{adj}	***MS***	***F***	***p***	η^2	η_G^2	η_P^2
Optics	574.038	2.578	222.626	38.082	0.000	0.121	0.159	0.536
Error (Optics)	497.438	85.090	5.846					
Acoustics	1142.884	1.401	815.700	43.017	0.000	0.240	0.274	0.566
Error (Acoustics)	876.760	46.237	18.962					
Optics × Acoustics	16.568	6.203	2.671	1.071	0.382	0.003	0.005	0.031
Error (Optics × Acoustics)	510.386	204.705	2.493					

ANOVA results (table 4) show a highly significant influence of both the optical and the acoustic characteristics of the stimuli on the perceived distance. Besides η_P^2 allowing for comparability of

effect sizes across different factorial designs by partialling out the influence of the respective other independent variables, and η_G^2 facilitating comparisons across different studies, the classical η^2 is specified because the present study is geared to the proportionate comparison between the effect sizes of the two factors within one design (see 2.2). Thus, the optical and acoustic characteristics account for 12 % and, respectively, 24 % of the total variance of the perceived distance. According to Cohen the effects may be classified as medium and large, respectively.[41 (pp. 413–414)]

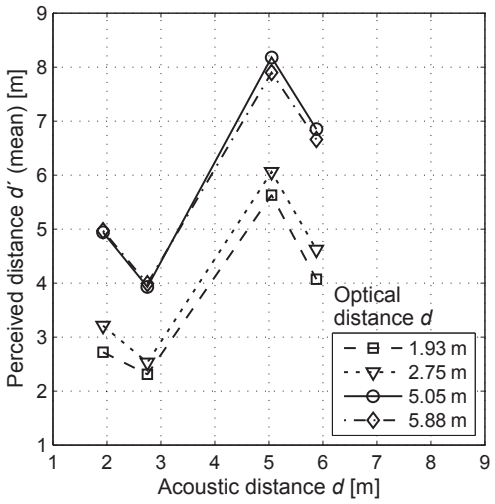

Figure 3: Comparison between the perceived and the acoustic distance for different levels of optical distance

Comparing the perceived with the physical distances (figure 3) indicates that, as expected, higher optical and acoustic distances lead to higher perceived distances in principle. The rank order of similar acoustic distances was, however, perceptually confused by the subjects. This effect occurs concordantly for both small and large distances (and room sizes, respectively). It may be plausibly explained by the different absorption of the respective rooms (cf. table 2). Specifically, in a low-absorbent room a smaller distance was perceived to be even greater than a larger distance in a high-absorbent room. A minor confusion of rank order could also be observed for similar, large optical distances (and room sizes, respectively).

3.2 Perceived room size

Interrelation between one- and three-dimensionally collected volume

In search of an adequate scale for the collection of room size assessments, the perceived volume was plotted against the perceived size 1D (mean values; the volume values were calculated from the respective mean values of length, width and height).

As shown in figure 4, the interrelation of the data is modeled best (R^2_{adj} = .983, p < .0005) by means of a cubic regression function (coefficients: a_1 = 163.546, a_2 = –94.492, a_3 = 20.582). Obviously, subjects did not map the one-dimensional, graphically equidistant scale explicitly asking for "volume"[42 (p. 83)] to equidistant volumes derived from their own separate length, width and height assessments. In this respect, the room volume appears to be mentally represented as an one-dimensional average edge length rather than a three-dimensional concept. Thus, the scale in question was henceforth denoted by *Perceived size 1D*; furthermore, for comparisons

with physical measures, the perceived volume and the perceived size 3D calculations are applied (see 2.4).

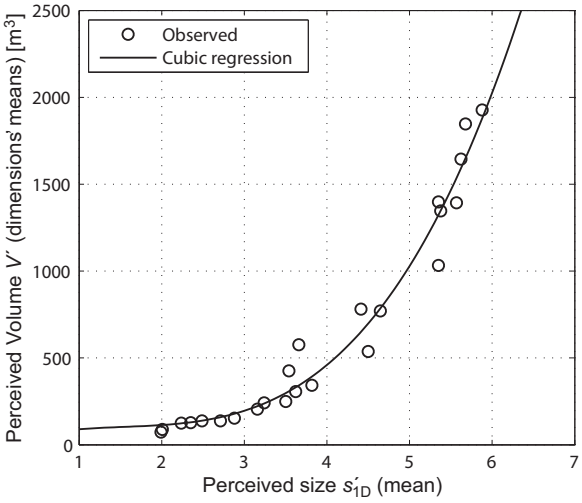

Figure 4: Interrelation between perceived volume and perceived size 1D

Co-presence paradigm

A RM ANOVA was performed in order to test the main effects of the opto-acoustic CP (levels O, A, OA) and of the room (4 levels), as well as of the respective interaction effect on the perceived size 1D. The error components of four out of twelve *s*' variables turned out not to be normally distributed. Only the main effect for room and the interaction effect are statistically significant (room: $df = 2.353$, $F = 387.790$, $p < .0005$, $\eta_P^2 = .876$; opto-acoustic CP × room: $df = 4.772$, $F = 7.536$, $p < .0005$, $\eta_P^2 = .186$). The power for the opto-acoustic CP effect is, however, poor ($1–\beta = .323$). An inspection of the cell means reveals that the interaction effect is of hybrid type. A comparison between the mean values of the perceived size 1D and the physical size for each CP condition (not depicted) shows similar tendencies like those of the distance variables (figure 1), namely a perceptual confusion of the rank order of the two small rooms. Inspecting the perceived volume means (not depicted) indicates that the physical volume is least underestimated under the optical condition and most under the acoustic condition. For the opto-acoustic condition the factor of underestimation was determined by means of a regression analysis ($R^2 = .933$, p = .034) and amounts to closely 40 %.

Conflicting stimulus paradigm

In order to examine the main effects of the optical room characteristics (4 levels) and of the acoustic room characteristics (4 levels), as well as of the respective interaction effect on the perceived size 1D, again a full-factorial RM ANOVA was performed. Beforehand, KS tests indicated that data do not violate the assumption of normally distributed error components.

The ANOVA results reveal highly significant main effects and a significant interaction effect. The optical and acoustic characteristics account for 32 % and 19 %, respectively, of the total variance of the perceived size 1D. The effect sizes may be regarded as large and medium, respectively.[41] An inspection of the cell means reveals that the type of interaction is hybrid. Its effect size is, however, very small.

Table 5: Results of RM ANOVA for s'_{1D}. Sources of Variation are based on the CS paradigm

Source of Variation	***SS***	df_{adj}	***MS***	***F***	***p***	η^2	η_G^2	η_P^2
Optics	459.347	1.669	275.255	66.972	0.000	0.319	0.395	0.670
Error (Optics)	226.340	55.071	4.110					
Acoustics	226.340	1.437	186.868	50.116	0.000	0.186	0.276	0.603
Error (Acoustics)	176.756	47.404	3.729					
Optics × Acoustics	11.042	7.040	1.568	2.381	0.023	0.008	0.015	0.067
Error (Optics × Acoustics)	153.020	232.333	0.659					

Comparing the perceived sizes 1D with the physical volumes (figure 5) indicates that, as expected, higher optical and acoustic volumes lead to higher perceived sizes 1D in the main. But in the case of the small rooms the rank order of similar acoustic sizes was perceptually confused by the subjects. Like the analogous finding regarding the distance variables (cf. figure 2) this may be due to the different absorption of the small rooms (cf. table 2). Incidentally, subjects also perceptually confused the rank order of the small optical volumes and partially of the large optical volumes.

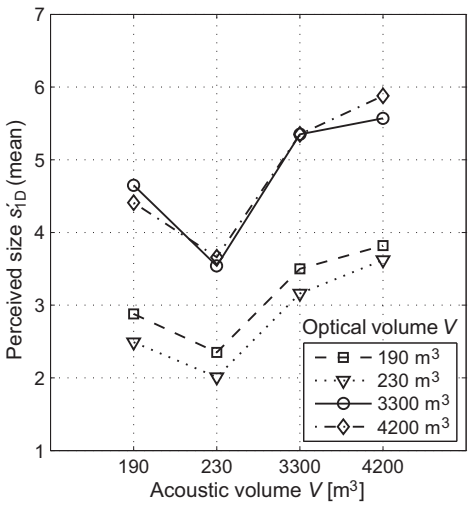

Figure 5: Comparison between the perceived size 1D and the acoustic volume for different levels of the optical volume

The CS paradigm may also be applied in order to examine quantitative asymmetries between two modalities. For example, it is well-known that there is an asymmetry in the perception of audio-visual synchrony: Optical and acoustic stimuli are perceived as the most synchronous if the sound is delayed by about 45 ms with reference to a corresponding moving picture.[e.g. 43-45] As an analogy in the spatial domain, we raise the question as to whether an asymmetry with respect to room size perception exists. Well-established measures quantifying an inter-modal (a)symmetry are the point of subjective equality (PSE), the upper and the lower detection threshold, and the preferred offset. The measures are based on certain response paradigms, e.g. discrimination tasks, rank order judgements or evaluative judgements. Albeit the present study is not methodologically geared to the determination of these measures, an accuracy measure derived from the available data might indicate the existence and the scale of a possible audio-visual asymmetry.

To this end, for each experimental condition, the mean perceived size 3D was divided by the expected perceived size 3D and multiplied by 100. Under CS conditions, the expected perceived size 3D is assumed to be the average of the optical size and the potentially divergent acoustic size. In line with Grechkin et al. in principle,[16] we define the accuracy measure ACC = $100 \cdot 2 \cdot s'_{3D} / (s_{acoustic} + s_{optical})$ [%]. Furthermore, we use $\Delta s = s_{acoustic} - s_{optical}$ [m] as a measure of opto-acoustic room size deviation.

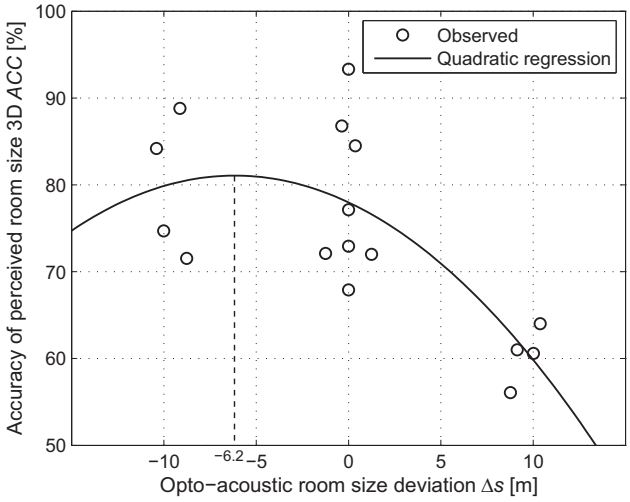

Figure 6: Accuracy of perceived room size 3D related to opto-acoustic room size deviation

The relation of the accuracy measure and the deviation measure is modeled best (R^2_{adj} = .540, p = .006) by an inverse quadratic function (coefficients: a_0 = 77.981, a_1 = –1.001, a_2 = –.081) that may be interpreted as an optimum curve (figure 6). Its maximum (ACC = 81 %) is located at Δs = –6.179 m. This is an indication for an audio-visual asymmetry in room size perception in the sense that the underestimation of perceived room size may be easier counteracted by a larger optical size than by a larger acoustical size.

4 DISCUSSION

We investigated the cooperation and the interaction of the auditory and the visual modality with respect to room size and egocentric distance perception by applying two design paradigms, the co-presence and the conflicting stimulus paradigm. All in all 35 % (perceived distance) and 51 % (perceived size) of the judgements could be statistically explained by the physical characteristics of the stimuli. In contrast to the prevalent postulate of an 'audio-visual interaction' at least no interaction effect of the acoustic and the optical stimulus characteristics with respect to size and distance perception could be found that was statistically significant or not negligible with regard to effect size. In fact main effects of medium and large size were observed.

The egocentric source distance was shown to be visually well-estimated in general and auditorily strongly overestimated leading also to a considerable audio-visual overestimation for both congruent and incongruent (conflicting) stimuli and to rank order confusions of similar distances (i.e. differing not more than 1 m). This effect was, however, observed only in reverberant / low-absorbent rooms – thus being in line with Cabrera et al. who found RT to strongly influence room size perception.[5] Conflicting stimuli containing lower optical distances may counteract the auditory overestimation of distance in reverberant rooms. Since the optical and acoustic stimulus distances contribute directly (i.e. as main effects) to distance perception and therefore may trade off mutually, there is no evidence for the existence of a proximity effect in terms of a visual dominance. If at all, we might speak, in contrast, of an *auditory capture effect* because acoustic

distance variation contributed twice as much (24 %) to distance judgement as a commensurable optical distance variation (12 %). Insofar, the results confirm the findings of Zahorik[28] who replicated Gardner's[26] experiment in a semi-reverberant room and exceed them by just providing rich acoustic distance cues of reverberant rooms for different head orientations. In contrast, results do not confirm an auditory and audio-visual underestimation of the source distance (1.2). This disagreement might also be explained by the reproduction of largely original acoustic cues by means of a state-of-the art auralization technique in the present case, particularly in relation to an optical simulation with restricted FOV. Though authors reporting auditory and audio-visual distance underestimation in VEs frequently used highly immersive displays, they did not always apply audio reproduction techniques preserving all important room acoustic cues. E.g., Rébillat et al. played back ambient sound in order to mask the video projectors' noise,[15] and thereby possibly masked some acoustic reflections, too. Chan et al. played back microphone recordings via headphones;[13] this does not allow for a perceptual externalization of the sound sources. A second concern is the use of artificial sound content. Subjects are normally not familiar with artificial sounds such as low-pass filtered and modulated white noise[15] even though it provides good localisation cues. Moreover, an artificial sound normally is not equivalent to the optical stimulus component, and the process of cognitively matching two components is instructed rather than experienced. So both displaying loudspeakers and playing back white noise appears not to be the best choice in order to establish an ecologically valid audio-visual interrelation. An opto-acoustic equivalence is not only required for the content or the objects but also for the environment (2.1). Experimental setups involving numeric models, however, do not guarantee that a room 'sounds as it looks'.

Perceptual volume assessments that had been collected by means of a one-dimensional scale turned out to express rather an average room edge length (or a room size), than a room volume. We thus recommend to collect the perceived room volume by means of separate one-dimensional scales for length, width and height in order to take advantage of both the physical units and the additional information on room shape. In contrast to distance perception, the perception of room size appears to rely rather on visually based than on auditorily based information since optical room size variation contributed more (32 %) to room size judgement than a commensurable acoustic room size variation (19 %). Also in contrast to distance perception, we observed the physical room size to be underestimated (except the small room 2SL) under all CP factor levels (most auditorily, least visually), as well as under all CS factor level combinations – so large optical sizes could not sufficiently counteract small acoustic sizes and vice versa. This result is contrary to the findings of Larsson & Väljamäe.[31] Furthermore, underestimation increased with room size (compression effect). The differences between the modalities as well as the differences between the perceptual means and the physical values were again minimal for the low-absorbent rooms.

Only for the perceived room size we found an indication for an audio-visual asymmetry (figure 6): The empirically modeled accuracy was maximum at –6.2 m room size deviation indicating that optical room size has a greater perceptual weight than the acoustic room size. In principle, this is consistent with the results provided by the application of the CP paradigm (see above) that the physical volume is most underestimated under the acoustic condition, and with the results provided by the application of the CS paradigm regarding both the η^2 values and the cell mean values (figure 5) whereupon the combination of a large acoustic and a small optical room size is perceived to be smaller than the combination of a large optical and a small acoustic room size.

Further investigation is needed in order to exclude the found asymmetry to be an artifact due to few data points and in order to determine a PSE by means of discrimination tasks. The improvement of internal validity by properly dissociating the factors source distance and room size is surely a crucial, however costly issue. And the variation of rooms in accordance to more criteria

than absorption and volume might contribute to the disentanglement of the perceptual effects of different acoustic parameters. The current *SEACEN* subproject will raise the external validity by using more and different rooms and scenes, respectively, enhance the potential of immersion by applying a large 180° curved display, and broaden the ecological validity and the opto-acoustic equivalence by presenting natural optical stimuli in motion.

ACKNOWLEDGMENTS

This work was carried out as a part of the subproject 9 "Audio-visual perception of acoustical environments" within the framework of the *SEACEN* project and funded by the German Research Foundation (DFG MA 4343/1-1). We would like to thank Alexander Lindau, Frank Schultz, Fabian Brinkmann, Michael Horn and Dr. Steffen Lepa for supporting the acoustic measurements, the experimental setup and data collection.

REFERENCES

1 S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki, "Psychoacoustic Cues in Room Size Perception", in *116th AES Convention 2004, Berlin, Germany*, Convention Paper 6084.

2 D. Cabrera, C. Pop, and D. Jeong, "Auditory Room Size Perception: A Comparison of Real versus Binaural Sound-fields", in *Acoustics 2006, 20–22 Nov. 2006, Christchurch, New Zealand*, pp. 417–422.

3 D. Cabrera, "Acoustic clarity and auditory room size perception", in *14th International Congress on Sound & Vibration, Cairns, Australia, 9-12 July 2007*.

4 M. Yadav, D. Cabrera, and W. L. Martens, "Auditory Room Size Perceived From A Room Acoustic Simulation With Autophonic Stimuli", in *Acoustics Australia* 39 (3) (2011), 101-105.

5 D. Cabrera, D. Jeong, H. J. Kwak, and J.-Y. Kim, "Auditory room size perception for modelled and measured rooms", in *Internoise, the 2005 Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil, 7-10 August 2005*.

6 S. H. Nielsen, "Auditory distance perception in different rooms", in *Journal of the Audio Engineering Society* 41 (1993), pp. 755–770.

7 A. W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms", in *Nature* 397 (1999), pp. 517–520.

8 A. W. Bronkhorst and P. Zahorik, "The direct-to-reverberant ratio as cue for distance perception in rooms", in *Journal of the Acoustical Society of America* 111 (5) (2002), pp. 2440–2441.

9 J. M. Loomis, R. L. Klatzky, and R. G. Golledge, "Auditory distance perception in real, virtual, and mixed environments", in Y. Ohta and H. Tamura (Eds.) *Mixed reality: Merging real and virtual worlds*, Berlin et al.: Springer, Tokyo: Ohmsha 1999.

10 H. Y. Kim, Y. Suzuki, S. Takane, and T. Sone, "Control of auditory distance perception based on the auditory parallax model", in *Applied Acoustics* 62 (3) (2001), pp. 245–270.

11 P. Zahorik, "Assessing auditory distance perception using virtual acoustics" in *Journal of the Acoustical Society of America* 111 (4) (2002), pp. 1832–1846.

12 H. Wittek, *Perceptual differences between wavefield synthesis and stereophony*. PhD thesis, Dep. of Music and Sound Recording School of Arts, Univ. of Surrey 2007.

13 J. S. Chan, D. Lisiecka, C. Ennis, C. OSullivan, and F. N. Newell, "Comparing audiovisual distance perception in various 'real' and 'virtual' environments", in *Perception* 38 (2009), ECVP Abstract Supplement, p. 30.

[14] G. Kearney, M. Gorzel, F. Boland, and H. Rice, "Depth perception in interactive virtual acoustic environments using higher order ambisonic soundfields", in *Proccedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics 2010*.

[15] M. Rébillat, X. Boutillon, É. Corteel, and B. F. G. Katz, "Audio, visual, and audio-visual egocentric distance perception invirtual environments", in *EAA Forum Acusticum 2011, Aalborg, Denmark*.

[16] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney, "How does presentation method and measurement protocol affect distance estimation in real and virtual environments?" In: *ACM Transactions on Applied Perception* 7 (4) (2010), pp. 26:1–26:18.

[17] W. M. Wiest and B. Bell, "Stevens's Exponent for Psychophysical Scaling of perceived, Remembered, and Inferred Distance", in *Psychological Bulletin* 98 (3) (1985), pp. 457–470.

[18] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson, "The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments", in *Perception* 34 (2) (2005), pp. 191–204.

[19] L. F. Kruszielski, T. Kamekawa, and A. Marui, "The Influence of Camera Focal Length in the Direct-To-Reverb Ratio Suitability and Its Effect in the Perception of Distance for a Motion Picture", in *AES 131st Convention, New York, NY, USA, 2011 October 20–23,* Convention paper 8580.

[20] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt, "Measurement protocols for medium-field distance perception in large-screen immersive displays", in *IEEE Virtual Reality 2009*, pp. 107–113.

[21] C. Armbruster, M. Wolter, T. Kuhlen, W. Spijkers, and B. Fimm, "Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space", in *Cyberpsychology & Behavior* 11 (1) (2008), pp. 9–15.

[22] A. Naceri, R. Chellali, F. Dionnet, and S. Toma, "Depth perception within virtual environments: a comparative study between wide screen stereoscopic displays and head mounted devices", in *Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns. Conference 2009,* pp. 460–466.

[23] I. V. Alexandrova, P. T. Teneva, S. de la Rosa, U. Kloos, H. H. Bülthoff, and B. J. Mohler, "Egocentric distance judgments in a large screen display immersive virtual environment", in *7th Symposium on Applied Perception in Graphics and Visualization, Los Angeles, CA, USA, July 24 – 25, 2010,* pp. 57–60.

[24] C. Nathanail and C. Lavandier, "Influence of the visual information on auditory perception. Consequences on the subjective 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, 1997*, pp. 285–290.

[25] P. Larsson, D. Västfjäll, and M. Kleiner, "Ecological Acoustics and the Multi-modal Perception of Rooms – Real and Unreal Experiences of Auditory-visual Virtual Environments", in *2001 International Conference on auditory Display, Espoo, Finland*.

[26] M. B. Gardner, "Proximity Image effect in sound localization", *Journal of the Acoustical Society of America* 43 (1) (1968). p. 163.

[27] D. H. Mershon, D. H. Desaulniers, T. L. J. Amerson, and S. A. Kiefer, "Visual capture in audition distance perception. Proximity image effect reconsidered", in *Journal of Auditory Research* 20 (1980), pp. 29–136.

[28] P. Zahorik, "Estimating sound source distance with and without vision", in *Optometry and vision science* 78 (5) (2001), pp. 270–275.

[29] J. Sandvad, "Auditory perception of reverberant surroundings", in *Journal of the Acoustical Society of America* 105 (2) Pt. 2 (1999), p. 1193.

[30] A. Kohlrausch and S. van de Par, "Audio-Visual Interaction in the Context of Multi-Media Applications", in *Communication Acoustics*, Berlin et al: Springer 2005, pp. 109–138

[31] P. Larsson and A. Väljamäe, "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.

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

[33] H.-J. Maempel, "Experiments on audio-visual room perception: a methodological discussion", in *DAGA 2012, Darmstadt,* pp. 783–784.

[34] A. Lindau and S. Weinzierl, "FABIAN - An instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom", in *24. Tonmeistertagung, Leipzig, 2006*.

[35] Deutsches Institut für Normung e.V., *DIN EN ISO 3382-1 Akustik - Messung von Parametern der Raumakustik - Teil 1: Aufführungsräume*, Berlin: Beuth 2009.

[36] E. H. Matsushima, F. F. G. da Silva, A. P. A. de Gouveia, R. R. M. Pinheiro, N. P. Ribeiro-Filho, and J. A. Da Silva, "The Psychophysics of Visually Directed Walking", in *23rd Annual Meeting of the International Society for Psychophysics, 2007, Tóquio*. p. 379–384.

[37] A. Lindau and S. Weinzierl, "Assessing the Plausibility of Virtual Acoustic Environments", in *Acta Acustica united with Acustica* 98 (5) (2012), pp. 804–810.

[38] S. Olejnik and J. Algina, "Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs", in *Psychological Methods* 8 (4) (2003), pp. 434–447.

[39] R. Bakeman, "Recommended effect size statistics for repeated measures designs", in *Behavior Research Methods* 37 (3) (2005), pp. 379–384.

[40] G. Keppel, *Design and analysis: A researcher's handbook*, Englewood Cliffs, NJ: Prentice Hall 1991.

[41] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Erlbaum 1988.

[42] M. Jentsch, *Audiovisuelle Raumwahrnehmung*, Master's thesis, Berlin: Techn. Univ., Audio comm. group 2012.

[43] B. Conrey and D. B. Pisoni, "Auditory-visual speech perception and synchrony detection for speech and nonspeech signals", in *Journal of the Acoustical Society of America* 119 (6) (2006), pp. 4065–4073.

[44] A. Vatakis and C. Spence, "Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task", in *Neuroscience Letters* 393 (2006), pp. 40–44.

[45] B. L. Heide and H.-J. Maempel, "Die Wahrnehmung audiovisueller Synchronität in elektronischen Medien", in *26th VDT International Convention, 25. –28.11.2010 Congress Center Leipzig, 2011, pp.* 525–537.