Audiovisual perception of real and virtual rooms

Hans-Joachim Maempel*, Michael Horn*

* Department for Acoustics and Music Technology / Studio Facilities and IT
Staatliches Institut für Musikforschung Preußischer Kulturbesitz
Tiergartenstraße 1, Berlin, Germany
email: [last name]@sim.spk-berlin.de
www: www.sim.spk-berlin.de

Abstract

Virtual environments utilized in experimental perception research are normally required to provide rich physical cues if they are to yield externally valid perceptual results. We investigated the perceptual difference between a real environment and a virtual environment under optical, acoustic, and optoacoustic conditions by conducting a 2×3 mixed design, with *environment* as a between-subjects factor and *domain* as a within-subjects factor. The dependent variables comprised auditory, visual, and audiovisual features including geometric estimates, aesthetic judgments, and sense of spatial presence. The real environment consisted of four visible loudspeakers in a small concert hall, playing back an anechoic multichannel recording of a string quartet. In the virtual environment, deemed the Virtual Concert Hall, the scene was reproduced three-dimensionally by applying dynamic binaural synthesis and stereoscopic projection on a 160° cylindrical screen. Most unimodal features were rated almost equally across the environments under both the optical/acoustic and the optoacoustic conditions. Estimates of geometric dimensions were lower (though not necessarily less accurate) in the virtual than in the real environment. Aesthetic features were rated almost equally across the environments under the acoustic condition, but not the optical, and similarly under the optoacoustic condition. Further results indicate that unimodal features of room perception might be subject to cognitive reconstruction due to both information acquired from another stimulus domain and abstract experiential knowledge of rooms. In conclusion, the validity of the Virtual Concert Hall for certain experimental applications is discussed.

First presented at the 2^{nd} International Conference Stereo and Immersive Media 2016, extended and revised for JVRB

Keywords: Virtual reality, audiovisual perception, concert hall, room acoustics, binaural synthesis, stereoscopy, simulation

1 Introduction

1.1 Subject

Today, virtual environments (VEs) play an important role in education, science, and the arts. In engineering, they may serve as development tools or as research objects themselves. Due to their capacity to reproduce scenes that are dangerous, unlikely to occur, or non-existent in real environments (REs), they furthermore serve as valuable tools for experimental research into perception and behaviour [dKIKS03]. In the latter case, VEs are often required to provide all perceptually relevant cues in a physically correct manner if they are to yield externally valid experimental results [Rum16, p. 33-34]. This requirement is particularly crucial when optical and acoustic stimuli must be physically and experientially congruent, as is generally the case in real environments. On a technical level, the criterion of physical correctness is currently still best met by applying data-based rather than numerically-modelled simulations. That is to say, simulation data acquired in situ (in the RE which is to be simulated) are preferred to data generated through computer-aided modelling. However, even data-based simulations can yield poor validity – either for physical reasons such as biased, incomplete, or masked cues, or for psychological reasons such as subjects' sense of artificiality and feeling that something is missing (e.g. sensory input to other modalities, social interaction). For example, under differing conditions, egocentric distances greater than about 2 m have been reported to be either underestimated (e.g. Armbrüster et al. [AWK$^+$08], Bruder et al. [BAOL16]) or overestimated [MJ13] in both optical and optoacoustic VEs. Thus, when substituting a VE for a RE during experimentation, an empirical comparison of their perceptual features is prerequisite [dKIKS03].

Unfortunately, this kind of validation is not always done. We investigated the difference between a real and a virtual performance room regarding auditory, visual, geometric and aesthetic features. The system under test was the Virtual Concert Hall (VCH), located at the Technical University of Berlin (TUB). As a research tool, it was designed strictly according to previously elaborated methodological criteria [Mae17] which are formulated by means of a consistent terminology (e.g. *domain* vs. *modality*, *optical* vs. *visual*, *acoustic* vs. *auditory*, *property* vs. *feature*, *unimodal* vs. *intermodal* vs. *supramodal*) used also in this article. The Virtual Concert Hall is capable of reproducing 3D sound and vision by applying dynamic binaural synthesis and stereoscopic projection onto a 160° cylindrical screen (Figures 1, 2). The simulation data were acquired in situ in the form of orientational binaural room impulse responses and stereoscopic panoramic images (for a detailed description see Maempel & Horn [MH17]).

1.2 State of the Art

The majority of VEs examined in comparison with respective REs have been designed to provide optical stimuli only, and data collection and analysis have predominantly focused on egocentric distance estimates and their accuracy (for an overview cf. Loomis & Knapp [LK03]). Generally, humans in real environments under optical full-cue conditions can estimate egocentric distances between 4 m and 12 m quite accurately. Under optical reduced-cue conditions and in optical VEs, however, distances were increasingly underestimated above a physical distance of about 3 m [LdSPF96, PKCR05, ZPCK09]. The accuracy of the estimates was shown not to be decreased by the field of view [KL04] and tended to be increased by stereoscopy, shadows, and reflections [GGA$^+$09]. Turning to the acoustic stimulus domain, egocentric distance perception was reported to be compressed in REs featuring full-cue conditions [ZBB05]. Specifically, distances were increasingly underestimated above physical distances of 2 to 7 m [LKG99, MNG13]. Kearney et al. [KGBR10] contrasted an optoacoustic RE with the combination of an optical RE and an acoustic VE. Egocentric distances were estimated mostly correctly at below 4 m and increasingly underestimated above 4 m in both the RE and the partial VE. Rébillat et al. [RBECK11] report a similar effect above 2.5 m in an optoacoustic VE displaying artificial content. Chan et al. [CLE$^+$09] report, briefly and in an abstract matter, on a systematic investigation of the differences between RE and VE in the acoustic, optical, and optoacoustic domains using a head-mounted display (HMD) and binaural recordings, and found "qualitative differences in the type of errors between the two environments" and relatively larger underestimations under "both unimodal" conditions (p. 30).

To date, no studies comparing VEs and REs were found that systematically investigated purely auditory features such as loudness and reverberance, and few comparative studies have investigated aesthetic features based on both optical and optoacoustic stimuli. Regarding the optical stimulus domain, Billger et al. [BHSR04] report a less accurate perceptual assignment of colours in the VE condition. In contrast, de Kort et al. [dKIKS03] observed no significant differences in the correctness of colour estimations between RE and VE. However, they did find that virtualisation resulted in underestimation of heights and less complete/correct retrospective room sketches, as well as significantly reduced mean factor scores of the perceptual components evaluation, ambience, privacy, and security. Moreover, principal component analysis (PCA) solutions were structurally different between the RE and the VE. Kuliga et al. [KTDH15] report significant differences in the perception of aspects of "Atmospherics" (p. 5). Providing optoacoustic stimuli, Bishop & Rohrmann [BR03] asked participants to complete an extensive questionnaire covering cognitive and affective aspects. They found that ratings of pleasure, naturalness, appreciation, and general real-

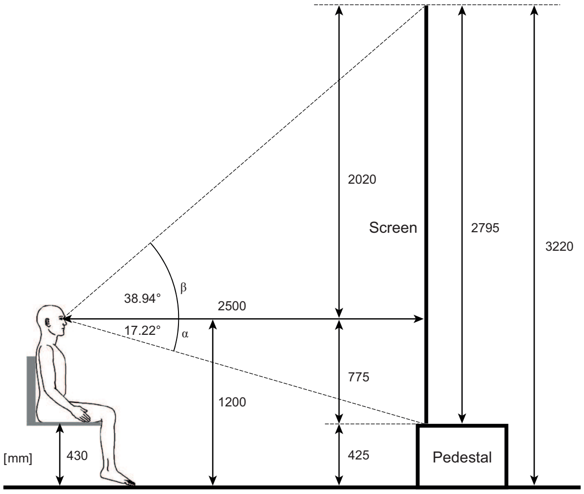

Figure 1: Optical projection setup of the VCH at the TUB (side view)

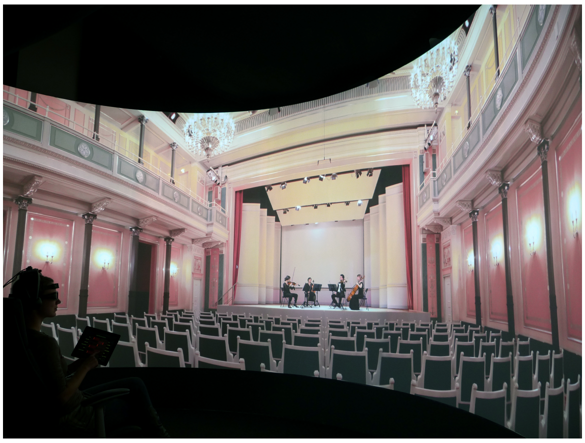

Figure 2: VCH at the TUB in use

ism (among others) were significantly decreased in the VE; familiarity, however, was rated higher. With the exceptions of de Kort et al. [dKIKS03] and Chan et al. [CLE+09], the above-cited studies all realised the experimental comparison of REs and VEs as a within-subjects factor, thereby raising the issue of order effects. Ziemer et al. [ZPCK09] demonstrated that prior exposure to a RE may, but does not necessarily, enhance the accuracy of egocentric distance estimation in a VE that models it. However, repeated measurement designs also entail uncontrollable validity problems for those collected features which are not verifiable: for example, aesthetic impression. Because participants know the experimental factor, they might – consciously or unconsciously – incorporate biases into their assessment (e.g. their assumptions of the study's purpose or their attitudes towards VEs).

The study at hand differs significantly from the cited comparative studies. This is due not only to differing stimulus domains, dependent variables, and designs, but also to the use of significantly richer stimulus content and more robust simulation techniques. In all cited studies providing optical stimuli, 3D numeric models were used as content, and in about half of them the optical 3D content was presented in 2D (these factors are unknown for Chan et al. [CLE+09]). The cited studies providing acoustic stimuli either used artificial audio content (e.g. noise bursts [RBECK11]); employed two-channel stereophony [BR03] or binaural recordings [CLE+09] instead of virtual acoustics; or applied sound field synthesis [KGBR10, MNG13, RBECK11]. In brief, no study was found comparing an optoacoustic RE to an optoacoustic VE which featured rich-cue conditions. By contrast, the VE under examination provides optical and acoustic stimuli that are data-based (thus detailed and occurring in nature) and reproduced in 3D, the independent variable *environment* was established as a between-subjects factor in order to avoid order effects and their associated validity issues, and participants were required to stay seated in both the RE and the VE to minimize interference by changes of position.

1.3 Research questions and hypotheses

In view of the validation purpose of the experiment, we posed the following core research questions: (1) Does the simulation of a spatial environment influence certain perceptual features? (2) Does this influence depend on the stimulus domain (acoustic, optical, optoacoustic)? Regarding distance perception, question (1) obtains the level of empirically-founded unspecific research hypothesis: Namely, that the simulation of an object and its spatial environment may bias its perceived egocentric distance (H_0: $\mu_{\mathrm{R}} = \mu_{\mathrm{V}}$; H_1: $\mu_{\mathrm{R}} \neq \mu_{\mathrm{V}}$). Regarding other aspects of perception, question (1) remains empirically unfounded, and must be investigated through an exploratory approach. This is also generally the case for question (2). Nevertheless, this approach can be expected to clarify which perceptual features are susceptible to alteration during the virtualisation of a RE in certain stimulus domains. Since a factor *domain* must be introduced to investigate question (2), the following question is also investigated: (3) Does the stimulus domain also directly influence the perceptual features under investigation? The respective results will be reported; however, as they are ancillary to the article's core aims, they will not be conclusively discussed.

2 Method

Since the VE under test was completely developed, we focused on non-expert features and stimulus content suited to the practical application of the VE as a research tool.

2.1 Perceptual features

With regard to the research questions on audiovisual room perception to be investigated by means of the VE, we were interested primarily in those auditory, visual, and audiovisual features that are apt to perceptually describe performance rooms in a multifaceted way (Table 1).

The audiovisual features of interest can be subclassified as geometric, aesthetic, and presence features. The geometric features comprised *source distance*, *source width*, and *room size*. Since test participants cannot be expected to maintain linearity when assessing three-dimensional room volume by means of a single one-dimensional scale [MJ13], we asked for separate length, width, and height estimates. The cube root of the product of these measures comprises the derived one-dimensional feature *room size*, which we took as a discrete object of analysis. Finally, we followed Osgood & Suci in regarding aesthetic judgment as an orthogonal structure also known as *semantic space*, consisting of three dimensions: *evaluation*, *potency*, and *activity* [OS55]. To help keep the test trials short, we abandoned the application of three measurement models (i.e. three scales) consisting of several indicator variables each. Instead, we applied one typical indicator variable per dimension: *pleasantness*, *powerfulness*, and *excitement*. Since 'sense of presence' is defined inconsistently throughout the literature, and diverse questionnaires had been proposed (cf. Maempel & Weinzierl [MW12] for an overview and Schuemie et al. [SvdSKvdM01] for a detailed review), we considered only the clear and comprehensible feature *spatial presence* as an important aspect regarding the perception of virtual performance rooms. As an intermodal feature, we also asked for the *audiovisual matching* of the acoustic and the optical rooms.

Loudness, *transparency*, *reverberance*, and *envelopment* are standard features in room acoustics. The *apparent source width* is only apparently missing in this list; we consider it a geometric feature [Mae17], so it merges into *source width* in the list of audiovisual features. Since the common auditory feature *bass* denotes both a pure tone colour and a musical instrument, the more distinct and neutral feature *lows* was chosen instead; consequently, the proper complement *highs* was used instead of the common feature *treble*.

The visual features do not (and are not required to) conform to physical or physiological component models because colour perception and naming is subject to categorization in the course of cognitive processing, resulting in unified percepts (for an overview see Cohen & Matthen [CM10]). While the feature *colour intensity* is related to chromaticity, *hue* is related to the dominant physical wavelengths. The scale is designed to cover only those prominent colour categories that are expected to be affected by the experimental factors. Since the concert halls to be simulated are neither blue nor green nor pink, we are particularly interested in the differentiation between yellow and red. The risk of an inconsistent understanding of the attributes was minimized by providing each test participant with a definition in advance.

2.2 Design

The two core research questions required two independent variables (factors). To prevent order effects and test participant response biases related to research question (1) due to participants' noticing the levels of the factor *environment*, this factor was realized as a between-subjects factor: the RE and VE levels were assigned to the separate groups *real* (R) and *virtual* (V), respectively. With a view toward achieving a reasonable total sample size, the second factor was realized as a within-subjects factor, i.e. through repeated measurements of participants' responses to *acoustic* (A), *optical* (O), and finally *optoacoustic* (OA) stimuli. In this manner, the research questions led to a two-factor mixed design (Table 2).

Regarding the factor *domain*, we deliberately did not counterbalance potential order effects by means of order variations: Given normal humans' strong visual dominance regarding geometric assessments, information initially perceived by eye was likely to superimpose upon and possibly bias information subsequently perceived by ear. To mitigate this possibility, we ap-

Modality	Nominalized feature	Ref. object	Scale pole labels, English translation (Scale pole labels, German original)
Audiovisual	*Pleasantness*		unpleasant - pleasant (unangenehm - angenehm)
	Powerfulness		powerless - powerful (kraftlos - kraftvoll)
	Excitement		calming - exciting (beruhigend - aufregend)
	Spatial presence		I feel present in the lab - I feel present in the hall (fühle mich im Labor - fühle mich im Saal)
	Audiovisual matching		sound/vision not matching - sound/vision matching (Ton/Bild unpassend - Ton/Bild passend)
	Source distance	Quartet	Distance [m] 0 - 20 (Entfernung [m] 0 - 20)
	Source width	Quartet	Width [m] 0 - 5 (Breite [m] 0 - 5)
	Room height	Room	Height [m] 0 - 25 (Höhe [m] 0 - 25)
	Room width	Room	Width [m] 0 - 50 (Breite [m] 0 - 50)
	Room length	Room	Length [m] 0 - 100 (Länge [m] 0 - 100)
Auditory	*Loudness*		gentle - loud (leise - laut)
	Lows		little lows - much lows (wenig Tiefen - viel Tiefen)
	Highs		little highs - much highs (wenig Höhen - viel Höhen)
	Transparency		hardly transparent - very transparent (kaum durchhörbar - sehr durchhörbar)
	Reverberance		dry - reverberant (trocken - hallig)
	Envelopment		slightly enveloping - strongly enveloping (schwach umhüllend - stark umhüllend)
Visual	*Source brightness*	Quartet	dark - bright (dunkel - hell)
	Room brightness	Room	dark - bright (dunkel - hell)
	Contrast	Room	weak contrast - strong contrast (schwacher Kontrast - starker Kontrast)
	Colour intensity	Room	pale colours - intense colours (blasse Farben - intensive Farben)
	Hue	Room	yellow-dominated - red-dominated (gelblich dominiert - rötlich dominiert)

Table 1: Collected perceptual features. The questionnaire was presented to participants electronically by means of a tablet computer. It contained bipolar rating scales and reference objects, where necessary; the geometric feature scales specified units, while the others did not. Interval scaling was assumed. The scale resolution was 127 steps. The test language was German; we use the English ‘transparent’ for the German ‘durchhörbar’, lit. ‘hear-through-able’.

Partial sample, time		*Environment* (between-subjects)	
		Real	Virtual
Domain (within-subjects)	Acoustic	n_R, t_1	n_V, t_1
	Optical	n_R, t_2	n_V, t_2
	Optoacoustic	n_R, t_3	n_V, t_3

Table 2: Test Design

plied the fixed stimulus order A → O → OA (Table 2), included stimulus-free pauses between the trials (see 2.5), and accepted a residual potential acoustic-to-optical order effect, which may be assumed to be weak and/or to affect the levels of the factor *environment* in a comparable way.

2.3 Sample

We strove to statistically reveal a medium effect size at a type I error level of $\alpha = .05$ and a test power of $1 - \beta = .80$. We needed to assume a large estimated correlation of around $r = .85$ among the levels of the factor *domain*, particularly levels O and OA. The required sample size had to be geared to the between-subjects effect, and was calculated *a priori* with the aid of the software package G*POWER 3 [FELB07, RFHN10]. Considering the disadvantageous condition of only two measurements in the case of auditory and visual features (see 2.6), the required total sample size was determined to be $N = 120$. This requirement was fulfilled by partial sample sizes totalling $n_\mathrm{R} = n_\mathrm{V} = 60$. Non-expert subjects unfamiliar with either the real or the simulated rooms participated voluntarily and received an incentive of 10 Euros each. The distribution of male and female participants was approximately equal across the partial samples ($f_\mathrm{R} = \{31, 29\}$, $f_\mathrm{V} = \{30, 30\}$), while the distribution of age classes (20-29, 30-39, 40-49, 50-69) was ($f_\mathrm{R} = \{30, 13, 13, 4\}$, $f_\mathrm{V} = \{34, 12, 12, 2\}$).

2.4 Stimuli

The Curt Sachs Hall at Berlin's Musical Instruments Museum, a small concert hall ($V \approx 1,030$ m^3, $RT30_\mathrm{mid} \approx 0.7$ s), served as the RE and provided the content for the VE. The optical and acoustic stimuli generated by the VE were data-based and reproduced three-dimensionally by means of stereoscopy and dynamic binaural synthesis.

Normally, the VCH is applied as a research tool, in which recordings of a string quartet are presented acoustically and optically in three dimensions (Figure 2). Because live musical performances cannot be exactly replicated again and again, in the case of this validation experiment the musicians had to be replaced by loudspeakers (Figures 3, 4). The optical absence of the moving musicians is an inevitable constraint of the validation experiment, as opposed to the experiments which the VCH is normally used for. The arrangement of the loudspeakers corresponds exactly, however, to the arrangement in a later application (see Maempel & Horn [MH17, 3.3]). Accordingly, the same music was used: an anechoic four-channel recording of the first part of the 2$^\mathrm{nd}$ movement of Claude Debussy's String Quartet in g minor, op. 10.

2.5 Procedure and data collection

All tests were performed individually. Participants of the R group ran through the levels of the factor *domain* by undergoing the following procedure: The participant put on a blindfold and ear-muffs, was guided to a defined seat in the lighted concert hall, and took a tablet computer. After the room light was turned off, the participant took off both the blindfold and the ear-muffs. The acoustic stimulus was presented and rated by the participant in the dark. After a five-second pause, the optical stimulus was presented by turning on the light, and was rated by the participant. The room light was then turned off for another five-second pause. Lastly, the room light was turned on while the acoustic stimulus was presented. This optoacoustic stimulus was rated by the participant. Participants of the V group were not required to wear a blindfold or ear-muffs. The levels A, O, and OA were realized by playing back sound and vision, respectively, within the VE. Like the concert hall, the VE was dark during both the A stimulus and the pauses. The lengths of the stimuli and the pauses were the same in the VE and the RE. Participants were required to stay seated in both the RE and the VE.

2.6 Data analysis

During the data analysis stage, descriptive and inferential statistics were applied. Arithmetic means, standard deviations, and 95% confidence intervals (CIs) for independent samples were calculated for each combination of factor levels. Means and CI bars

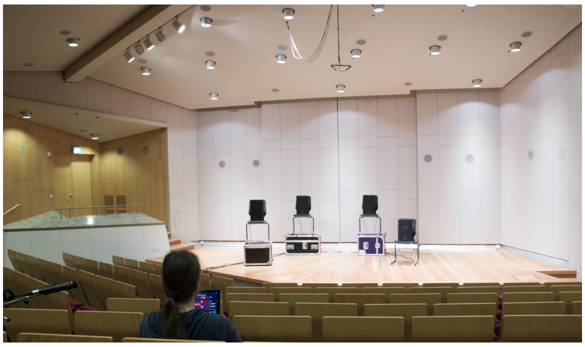

Figure 3: Optical RE

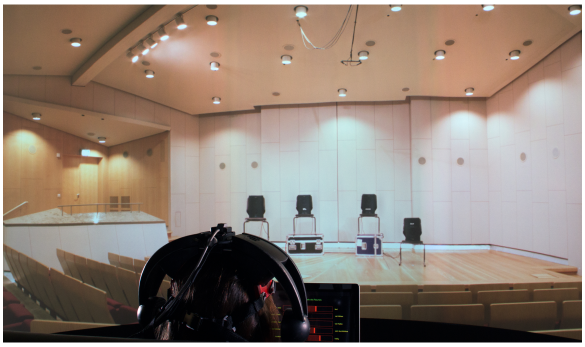

Figure 4: Optical VE (stereoscopy switched off for the photograph). In front: person with extraaural headset and tracking sensor

were plotted against the combinations for each perceptual feature; thereby the 127 step scales used in the electronic questionnaire were linearly transformed to scales ranging from -2 to +2 for the auditory, visual and aesthetic features, and to scales ranging from 0 to the values specified in Table 1 for the geometric features.

After checking the respective assumptions, a two-factor mixed-design analysis of variance (mixed ANOVA) was performed for each perceptual feature. Because ratings of auditory features were not collected under the optical condition and ratings of visual features were not collected under the acoustic condition, the 3×2 design was reduced to a 2×2 model for these features. In the case of audiovisual features (except *audiovisual matching*), a 3×2 model was applied. Contrasts between the *domain* main and interaction levels A-OA and O-OA were also tested. The significance level was Bonferroni adjusted ($\alpha' = .0253$). Because the intermodal feature *audiovisual matching* required an optoacoustic stimulus, only the effect of *environment* was tested by applying a two-tailed t-test.

One-sample Kolmogorov-Smirnov tests indicated that the assumption of normally distributed error components was met. Only the variable *colour intensity* showed a minor violation of the assumption (KS-$Z = 1.507$, $p = .021$) in the VE group under the optical condition, but this was deemed tolerable in view of the robustness of the mixed ANOVA. For the audiovisual variables, however, Mauchly's sphericity test indicated a significant violation of the sphericity assumption in the 3×2 analyses, which was compensated for by correcting the degrees of freedom using Greenhouse-Geisser estimates of sphericity. The effect size was reported as both partial eta squared ($\eta_p{}^2$) and classical eta squared (η^2). To classify effect sizes as small (s), medium (m) or large (l), Cohen's f was calculated [Coh88, p. 281] from classical eta squared [Lak13, p. 7]. In the case of the t-test Cohen's d was reported and classified [Coh88, pp. 20, 25-26].

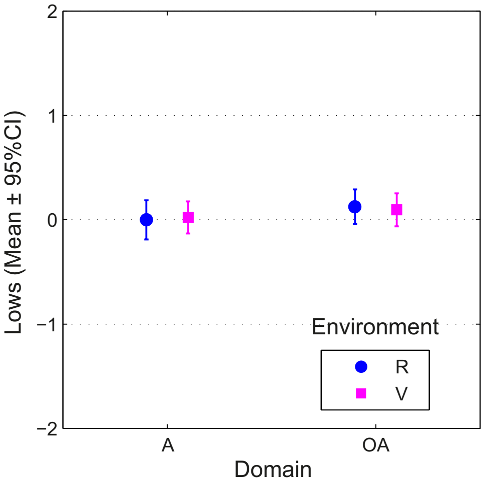

Figure 5: Means and CIs of *lows*

3 Results

3.1 Auditory features

The auditory features *lows*, *highs*, and *transparency* showed no significant effects of the factors *environment* or *domain* (see exemplary Figure 5 for *lows*). At level OA, *loudness* was rated slightly but significantly higher than at level A (Figure 6), $F(1, 118) = 5.443$, $p = .021$, $\eta_p{}^2 = .044$, $\eta^2 = .014$ (s). There was, however, no significant effect of *environment*.

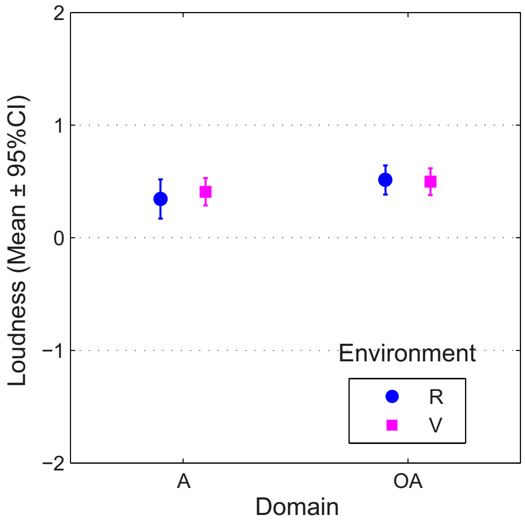

Figure 6: Means and CIs of *loudness*

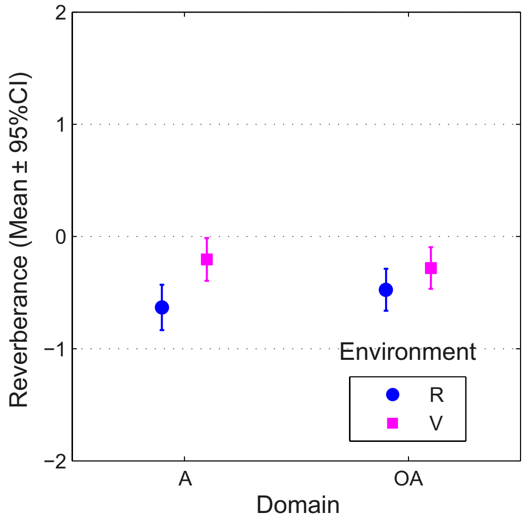

Figure 7: Means and CIs of *reverberance*

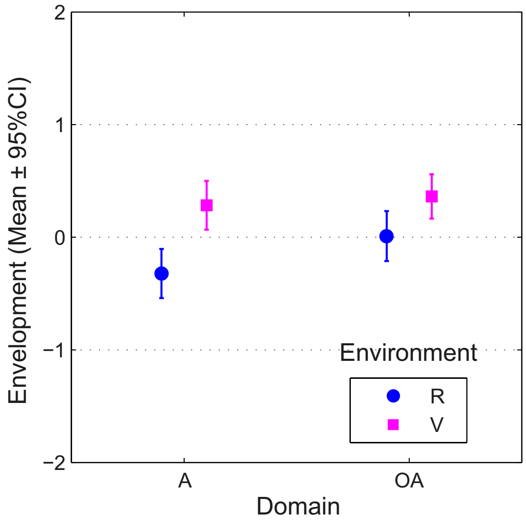

Figure 8: Means and CIs of *envelopment*

In contrast, *reverberance* was rated higher in the VE (Figure 7), $F(1,118) = 6.485$, $p = .012$, $\eta_p{}^2 = .052$, $\eta^2 = .041$ (m). This held true for *envelopment* as well, $F(1,118) = 12.764$, $p = .001$, $\eta_p{}^2 = .098$, $\eta^2 = .074$ (m); additionally, *envelopment* was slightly influenced by the factor *domain* (Figure 8), $F(1,118) = 7.213$, $p = .008$, $\eta_p{}^2 = .058$, $\eta^2 = .014$ (s).

3.2 Visual features

Three of the five visual features (*source brightness*, *contrast*, *colour intensity*) were not influenced by any of the experimental factors (see exemplary Figure 9 for *contrast*).

The factor *domain* did show an effect on *hue*, $F(1,118) = 7.403$, $p = .007$, $\eta_p{}^2 = .059$, $\eta^2 = .011$ (s): the feature was rated as slightly less yellowish under the optoacoustic condition ($M = -.79$, $SD = .71$) than under the optical ($M = -.94$, $SD = .68$). However, *environment* showed no effect on *hue*.

In contrast, *environment* did show a small effect on *room brightness* (Figure 10), $F(1,118) = 4.436$, $p = .037$, $\eta_p{}^2 = .036$, $\eta^2 = .028$ (s), as did *domain*, $F(1,118) = 4.179$, $p = .043$, $\eta_p{}^2 = .034$, $\eta^2 = .006$ (s). Specifically, at level O the virtual room appeared slightly brighter than the real room (interaction effect), $F(1,118) = 17.298$, $p = .000$, $\eta_p{}^2 = .128$, $\eta^2 = .027$ (s).

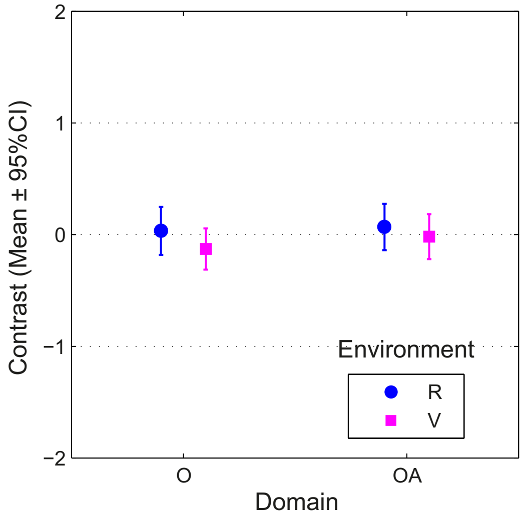

Figure 9: Means and CIs of *contrast*

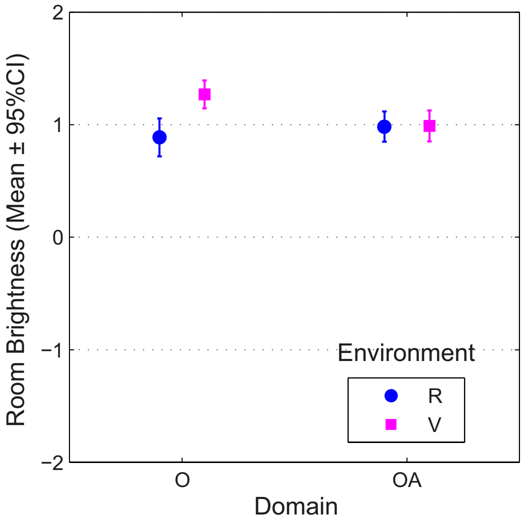

Figure 10: Means and CIs of *room brightness*

3.3 Geometric features

For all geometric features, significant main and interaction effects of both *environment* and *domain* were observed – with the exception of the feature *source distance*, which was not significantly influenced by *domain*.

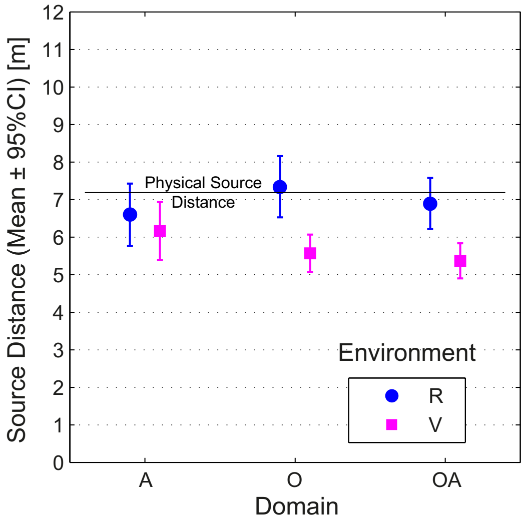

Figure 11: Means and CIs of *source distance*

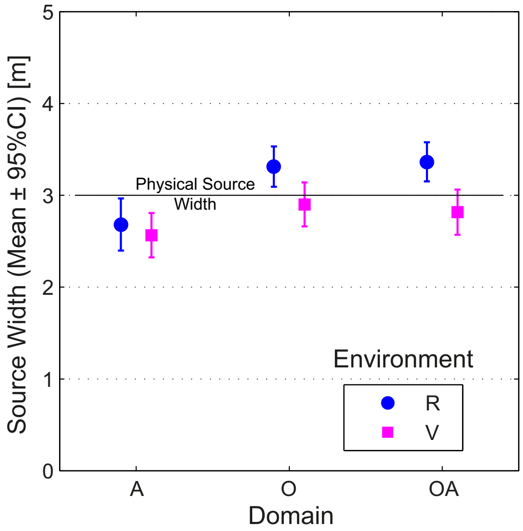

Figure 12: Means and CIs of *source width*

Estimates of *source distance* were significantly lower at V than at R level in general, $F(1,118) = 10.153$, $p = .002$, $\eta_p{}^2 = .079$, $\eta^2 = .028$ (s). Additionally, they depended on the combination of factor levels (interaction effect), $F(1.547, 182.501) = 3.424$,

$p = .047$, $\eta_p{}^2 = .028$, $\eta^2 = .018$ (s). In the RE, estimates were remarkably domain-independent and accurate, ranging from -8.2 % (A) to +2.2 % (O) of the physical source distance (Figure 11).

In the case of *source width* (Figure 12), estimates were a bit lower (and less accurate) in the VE than in the RE (main effect), $F(1,118) = 6.030$, $p = .016$, $\eta_p{}^2 = .049$, $\eta^2 = .020$ (s). Regarding the factor *domain*, the presence of the congruent optical stimulus led to higher (and more accurate) estimates (main contrast A-OA), $F(1,118) = 30.376$, $p = .000$, $\eta_p{}^2 = .205$, $\eta^2 = .053$ (m), as well as to relatively lower (but more accurate) estimates in the VE than in the RE (interaction contrast A-OA), $F(1,118) = 6.457$, $p = .012$, $\eta_p{}^2 = .052$, $\eta^2 = .011$ (s).

As with *source width*, estimates of *room size* were significantly lower in the VE than in the RE, $F(1,118) = 4.545$, $p = .035$, $\eta_p{}^2 = .037$, $\eta^2 = .020$ (s). Again, the presence of the congruent optical stimulus component caused both absolutely higher (however considerably less accurate) estimates (significant main contrast A-OA), $F(1,118) = 43.455$, $p = .000$, $\eta_p{}^2 = .269$, $\eta^2 = .059$ (m), and VE-related estimates that are relatively lower (however more accurate) than in the RE (significant interaction contrast A-OA), $F(1,118) = 7.669$, $p = .007$, $\eta_p{}^2 = .061$, $\eta^2 = .010$ (s). Large confidence intervals indicate test participants' general uncertainty regarding their room size estimates (Figure 13), which is presumably due to the invisible retral hemisphere.

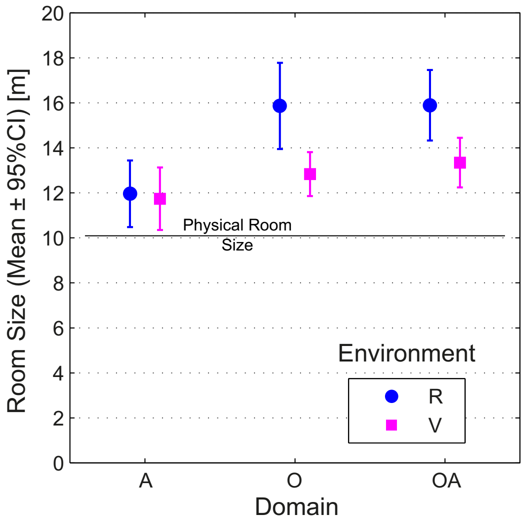

Figure 13: Means and CIs of *room size*

3.4 Aesthetic and presence features

The evaluative feature *pleasantness* (Figure 14) was clearly influenced by environment, $F(1,118) = 9.981$, $p = .002$, $\eta_p{}^2 = .078$, $\eta^2 = .023$ (s). It was rated significantly lower when the congruent acoustic component was absent (main contrast O-OA), $F(1,118) = 9.566$, $p = .002$, $\eta_p{}^2 = .075$, $\eta^2 = .018$ (s). More specifically, the VE was liked less than the RE when the congruent optical component was present (interaction contrast A-OA), $F(1,118) = 5.912$, $p = .017$, $\eta_p{}^2 = .048$, $\eta^2 = .013$ (s).

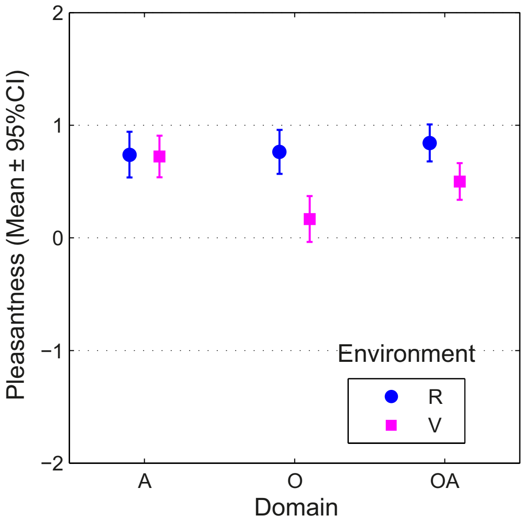

Figure 14: Means and CIs of *pleasantness*

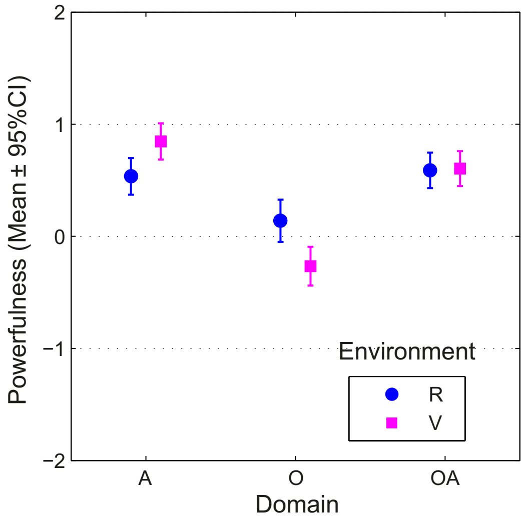

Figure 15: Means and CIs of *powerfulness*

In contrast, no significant main effect of *environment* on *powerfulness* was observed (Figure 15). Regarding the factor *domain*, the O condition yielded a considerably lower rating than the OA condition (significant main contrast O-OA), $F(1,118) = 88.750$, $p = .000$, $\eta_p{}^2 = .429$, $\eta^2 = .191$ (l). At the A level, ratings in the VE were higher than in the RE (interaction contrast A-OA), $F(1,118) = 5.414$, $p = .022$, $\eta_p{}^2 = .044$, $\eta^2 = .013$ (s), whereas at the O level they were lower than in the RE (interaction contrast O-OA), $F(1,118) = 9.006$, $p = .003$, $\eta_p{}^2 = .071$, $\eta^2 = .019$ (s). The feature *excitement* (Figure 16) was rated generally higher in the VE than in the RE, $F(1,118) = 17.867$, $p = .000$, $\eta_p{}^2 = .132$, $\eta^2 = .036$ (m). Without the congruent acoustic component, ratings generally dropped (main contrast O-OA), $F(1,118) = 11.061$, $p = .001$, $\eta_p{}^2 = .086$, $\eta^2 = .032$ (m). Participants felt less spatially present in the VE than in the RE in general (Figure 17), $F(1,118) = 23.550$, $p = .000$, $\eta_p{}^2 = .166$, $\eta^2 = .046$ (m), though negligibly so under the A condition and significantly more so under the O condition (interaction contrast O-OA), $F(1,118) = 19.945$, $p = .000$, $\eta_p{}^2 = .145$, $\eta^2 = .037$ (m). No significant main effect of *domain* was observed.

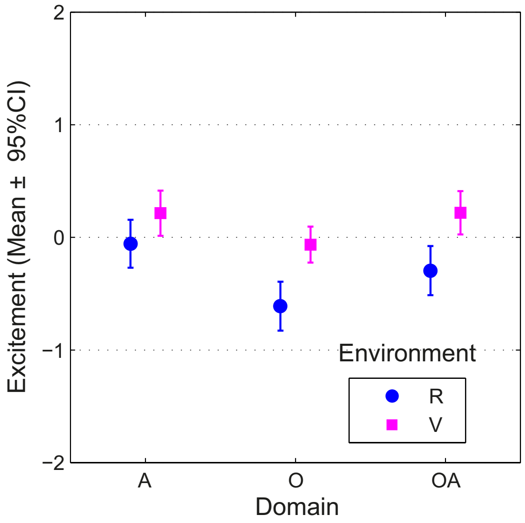

Figure 16: Means and CIs of *excitement*

The main effect of *environment* on *audiovisual matching* was significant, $t(118) = 2.924$, $p = .004$, $d = .518$ (m). Audiovisual matching was rated lower under the V condition ($M = -.107$, $SD = 1.024$) than under the R condition ($M = .421$, $SD = .954$).

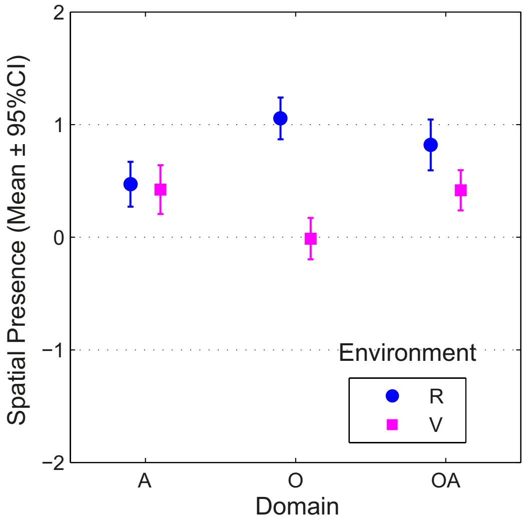

Figure 17: Means and CIs of *spatial presence*

4 Discussion

4.1 Auditory and visual features

All unimodal features that are determined by the amount or spatial distribution of stimulus energy (*lows*, *highs*, *loudness*, *source brightness*, *colour intensity*, *contrast*, *hue*) were rated almost equally across the environments in all domains. For *room brightness*, however, this only held true at the level OA, though the size of the interaction effect is small. Since the laboratory was a dark background against which the illuminated projection screen was silhouetted, in this case it might have acted as a reference that played a minor role when the perceived room was modally completed (OA condition). Auditory features which are also and/or predominantly determined by the temporal distribution of energy (*reverberance* and *envelopment*) were rated higher in the VE, regardless of *domain*. Again, the properties of the laboratory might have offered a reference in view of expectations regarding both the small, absorbent room and the likely sound of two-channel stereophony via headphones.

4.2 Geometric features

The domain-independent and largely accurate estimates of given physical distance (7.19 m) in the RE were generally in line with findings regarding optical REs reported by Loomis et al. [LdSFF92, LdSPF96]. Conversely, they did not fully accord with the underestimations in optical or acoustic REs reported by Plumert et al. [PKCR05] and Kearney et al. [KGBR10], with half of the experiments reported in Loomis et al. [LKG99], or with the average parameters of the power function reported by Zahorik et al. [ZBB05, p. 410]. The observed slight underestimation in the acoustic RE was also nowhere near as large as the (extrapolated) findings of Moulin et al. [MNG13] would suggest. Moreover, apparently due to the advanced binaural simulation technique utilized in the VE, neither a noteworthy deviation from estimates in the RE nor severely decreased accuracy were observed. The latter aspect is contrary to Rébillat et al. [RBECK11], who used wave field synthesis for their acoustic simulation. The underestimations in both the optical and the optoacoustic VE relative to the RE accord, however, with respective findings of Loomis et al. [LdSPF96] regarding diminished cues, as well as with Ziemer et al. [ZPCK09] and Kearney et al. [KGBR10]. The difference of source width estimates between the RE and the VE might be compared to the findings of de Kort et al. [dKIKS03] regarding height estimates, because both width and height are orthogonal to egocentric distance. In the present study, however, estimates in the optical VE were more accurate.

4.3 Aesthetic features

The observed negative impact of virtualisation on ratings of the aesthetic, evaluative feature *pleasantness* is largely in line with de Kort et al. [dKIKS03] who found decreased mean factor scores of the component *evaluation* in their optical VE, and with Bishop & Rohrmann [BR03], who reported significantly lower ratings of *pleasure* in their optoacoustic VE compared to an optoacoustic RE. Regarding the feature *audiovisual matching*, the VE did rate relatively lower than the RE; however, this is perhaps less remarkable than the low absolute rating of the RE itself. This raises the question of to what extent humans have the intermodal ability to relate complex optical and acoustic room properties to each other.

4.4 Perceptual validity of the VE

It was observed regarding geometric, aesthetic, and presence features alike that noteworthy differences between the RE and the VE occurred only in the presence of the optical stimulus component. Obviously, the optical simulation causes these differences – not the acoustic. However, an assessment of the perceptual validity of the VCH must consider the fact that it was specifically designed for the reproduction of optoacoustically-conflicting stimuli (e.g., presenting a certain acoustic room while at the same time displaying a different optical room), and is therefore used primarily for this purpose. At this level of primary interest (OA), virtualisation had no significant influence on the ratings of either auditory or visual features (two-tailed t-tests at the level OA between the RE and the VE), except a significant influence on *envelopment*; that is to say, those effects of *environment* which occur are likely to be small, and the mean values of the unimodal features induced by the RE may also be reliably induced by the VE. In contrast, decreased estimates and ratings must be anticipated for the geometric features and for the aesthetic and presence features *pleasantness*, *spatial presence*, and *audiovisual matching*, as should increased ratings of *excitement*. Only for *powerfulness* the mean ratings were almost equal across the environments. In view of the planned application of the VCH, however, such simulation-induced differences of mean ratings do not play a major role; this is because different rooms are to be compared within the VE, and the perceptual deviation imputable to virtualisation may be assumed to be equal across the rooms. On the contrary, the VCH does not allow for valid questioning of the accuracy of distance estimates. At least in the acoustic domain, absolute geometric ratings in both environments match up closely.

4.5 Additional results

Based on the second factor *domain*, the experiment further yielded results beyond the issue of the perceptual validity of the VE under test. Due to the fixed order of the factor levels, potential order effects must be discussed first. Indications of the absence of an order effect are: Half of the dependent variables do not show a significant effect of *domain* at all, whereas in the case of significant effects, changes from one domain level to another often diverge in the RE and the VE (e.g. *room brightness*, *source width*, *source distance*, *spatial presence*). Assuming order effects are negligible, the presence of significant differences in the ratings of unimodal features between the levels A or O and level OA suggest potential crossmodal effects: the optoacoustic stimuli were perceived as slightly louder and more enveloping than the acoustic, and as slightly darker and more reddish than the optical. Two additional results further substantiate the assumption that basal features are subject to cognitive influence. (1)

Sound pressure levels were the same in the RE and the VE, and thus ratings of loudness were predictably similar. Surprisingly, however, the mean rating of room brightness at level O was higher in the VE than in the RE, even though the luminance of the VE was about two exposure values lower than the luminance of the RE (due to restrictions of the projection system and shutter glasses). (2) Where significant differences of ratings between the environments were observed, they were generally not equally distributed across the levels of *domain* (interaction trend or significant interaction effect). Specifically, rating differences between the RE and the VE were generally largest at the O level, smallest at the A level, and medium at the OA level (except *source width* and *powerfulness*). In the case of audiovisual features, this frequent trade-off is plausibly due to the merging of information from either domain within each environment, which would entail compensation for deteriorations of the optical simulation. Remarkably, this pattern may also be observed in the case of unimodal features, e.g. within: 1. Trends regarding the auditory feature *envelopment*; 2. A significant difference between the RE and the VE only at level A regarding the auditory feature *reverberance* (additional t-tests); 3. A significant interaction effect regarding the visual feature *room brightness*. Obviously, an additional and congruent optical rich-cue stimulus influenced the mean ratings of *reverberance* and *envelopment* (Figures 7, 8), while an additional and congruent acoustic rich-cue stimulus influenced the mean rating of *room brightness* (Figure 10). Indeed, crossmodal effects, albeit usually of low size, are well-known [Fas04]; they are normally explainable by the accumulation of stimulus energies and/or arousal states. In the present case, however, the mean ratings differing between the RE and the VE at levels A or O *converged* at level OA, thereby compensating for deteriorations caused by virtualisation – though the ratings were generated by different groups of participants without previous knowledge of the real or virtual environments. This observation points to the influence of a general experiential knowledge of rooms, similar across the groups. Thus, it is tentatively hypothesised that: (1) Even basal, unimodal, energetically-determined perceptual features of rooms are in part subject to cognitive reconstruction; (2) Both information acquired from another stimulus domain and abstract experiential knowledge of rooms may contribute to this reconstruction. Evidence that these effects stabilise the perceptually-based conception of rooms would be a strong argument in favour of VEs featuring rich-cue conditions, as far as the ecological validity of test stimuli is considered in their experimental applications.

5 Conclusion

Under the optoacoustic condition, the VE may be regarded as perceptually valid for the auditory and visual features, i.e. it does not bias the respective ratings induced by the RE. In contrast, the VE may not be regarded as perceptually valid for the geometric and aesthetic features (except *powerfulness*), i.e. it does bias the respective estimates and ratings induced by the RE. So whenever the RE serves as a crucial external criterion, e.g. when questioning the accuracy of distance estimates, it must not be replaced by the VE. Under the acoustic condition, however, the VE turned out to be perceptually valid for all features except *reverberance*, *envelopment*, and *powerfulness*. As a by-product, significant mean differences suggest that *loudness*, *envelopment*, *room brightness*, and *hue* might be subject to crossmodal effects. Since the results indicate that unimodal features of room perception might be subject to cognitive reconstruction due to both information acquired from another stimulus domain and abstract experiential knowledge of rooms, virtual environments serving as research tools for the investigation of room perception are recommended to provide multi-domain (e.g. optoacoustic) stimuli and to feature rich-cue conditions.

6 Acknowledgments

This work was carried out as a part of 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 all test participants.

References

[AWK$^+$08] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Will Spijkers, and Bruno Fimm, *Depth perception in virtual reality: Distance estimations in peri- and extrapersonal space*, CyberPsychology & Behavior **11**

(2008), no. 1, 9–15, ISSN 1094-9313, DOI 10.1089/cpb.2007.9935.

[BAOL16] Gerd Bruder, Ferran Argelaguet, Anne-Héléne Olivier, and Anatole Lécuyer, *CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups*, Presence: Teleoperators and Virtual Environments **25** (2016), no. 1, 1–16, ISSN 1054-7460, DOI 10.1162/pres$_{a0}$0241.

[BHSR04] Monica Billger, Ilona Heldal, Beata Stahre, and Kristian Renstrom, *Perception of Color and Space in Virtual Reality: A Comparison Between a Real Room and Virtual Reality Models*, Proceedings of SPIE, Human Vision and Electronic Imaging IX, vol. 5292, 2004, DOI 10.1117/12.526986, pp. 90–98.

[BR03] Ian D. Bishop and Bernd Rohrmann, *Subjective responses to simulated and real environments: a comparison*, Landscape and Urban Planning **65** (2003), no. 4, 261–277, ISSN 0169-2046, DOI 10.1016/S0169-2046(03)00070-7.

[CLE$^+$09] Jason S. Chan, Danuta Lisiecka, Cathy Ennis, Carol O'Sullivan, and Fiona N. Newell, *Comparing audio-visual distance perception in various 'real' and 'virtual' environments*, Perception (ECVP Abstract Supplement) **38** (2009), 30, DOI 10.1177/03010066090380S101.

[CM10] Jonathan Cohen and Mohan Matthen (eds.), *Color Ontology and Color Science*, MIT Press, 2010, ISBN 978-0-262-01385-7, DOI 10.7551/mitpress/9780262013857.001.0001.

[Coh88] Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Lawrence Erlbaum Associates, 1988, ISBN 0805802835.

[dKIKS03] Yvonne A. W. de Kort, Wijnand A. IJsselsteijn, Jolien Kooijman, and Yvon Schuurmans, *Virtual Laboratories: Comparability of Real and Virtual Environments for Environmental Psychology*, Presence: Teleoperators and Virtual Environments **12** (2003), no. 4, 360–373, ISSN 1054-7460, DOI 10.1162/105474603322391604.

[Fas04] Hugo Fastl, *Audio-visual interactions in loudness evaluation*, Proceedings of the 18th International Congress on Acoustics, Kyoto Japan, vol. 2, 2004, pp. 1161–1166.

[FELB07] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner, *G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences*, Behavior Research Methods **39** (2007), no. 2, 175–191, ISSN 1554-351X, DOI 10.3758/BF03193146.

[GGA$^+$09] Davide Gadia, Alessandra Galmonte, Tiziano Agostini, Alberto Viale, and Daniele Marini, *Depth and distance perception in a curved large screen virtual reality installation*, Proceedings of SPIE, Stereoscopic Displays and Applications XX (Andrew J. Woods and John O. Merritt Nicholas S. Holliman, eds.), vol. 7237, 2009, DOI 10.1117/12.805809.

[KGBR10] Gavin Kearney, Marcin Gorzel, Frank Boland, and Henry Rice, *Depth Perception in Interactive Virtual Acoustic Environments using Higher Order Ambisonic Soundfields*, 2nd International Symposium on Ambisonics and Spherical Acoustics, 2010, Available from ambisonics10.ircam.fr/drupal/index5259.html?q=proceedings/o6, last visited March, 1st 2018.

[KL04] Joshua M. Knapp and Jack M. Loomis, *Limited Field of View of Head-Mounted Displays Is Not the Cause of Distance Underestimation in Virtual Environments*, Presence: Teleoperators and Virtual Environments **13** (2004), no. 5, 572–577, ISSN 1054-7460, DOI 10.1162/1054746042545238.

[KTDH15] Saskia Felizitas Kuliga, T. Thrash, Ruth C. Dalton, and Christoph Hölscher, *Virtual reality as an empirical research tool - Exploring user experience in a real building and a corresponding virtual model*, Computers, Environment and Urban Systems **54** (2015), 363–375, ISSN 0198-9715, DOI 10.1016/j.compenvurbsys.2015.09.006.

[Lak13] Daniel Lakens, *Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs*, Frontiers in Psychology **4** (2013), 1–12, ISSN 1664-1078, DOI 10.3389/fpsyg.2013.00863.

[LdSFF92] Jack M. Loomis, José A. da Silva, Naofumi Fujita, and Sergio S. Fukusima, *Visual Space Perception and Visually Directed Action*, Journal of Experimental Psychology: Human Perception and Performance **18** (1992), no. 4, 906–921, ISSN 1939-1277, DOI 10.1037/0096-1523.18.4.906.

[LdSPF96] Jack M. Loomis, José A. da Silva, John W. Philbeck, and Sergio S. Fukusima, *Visual Perception of Location and Distance*, Current Directions in Psychological Science **5** (1996), no. 3, 72–77, ISSN 1467-8721, DOI 10.1111/1467-8721.ep10772783.

[LK03] Jack M. Loomis and Joshua M. Knapp, *Visual perception of egocentric distance in real and virtual environments*, Virtual and adaptive environments: Applications, implications, and human performance issues (Michael W. Haas and Lawrence J. Hettinger, eds.), Lawrence Erlbaum Associates, 2003, DOI 10.1201/9781410608888, pp. 21–46, ISBN 9781410608888.

[LKG99] Jack M. Loomis, Roberta L. Klatzky, and Reginald G. Golledge, *Auditory distance perception in real, virtual, and mixed environments*, Mixed Reality: Merging Real and Virtual Worlds (Yuichi Ohta and Hideyuki Tamura, eds.), Springer, 1999, pp. 201–214, ISBN 978-3-540-65623-4.

[Mae17] Hans-Joachim Maempel, *Apples and oranges: A methodological framework for basic research into audiovisual perception*, Preprint from Jahrbuch des Staatlichen Instituts für Musikforschung 2016, 2017, DOI 10.14279/depositonce-6424.

[MH17] Hans-Joachim Maempel and Michael Horn, *The Virtual Concert Hall - a research tool for the experimental investigation of audiovisual room perception*, International Journal on Stereo & Immersive Media **1** (2017), no. 1, 78–98, ISSN 2184-1241.

[MJ13] Hans-Joachim Maempel and Matthias Jentsch, *Auditory and visual contribution to egocentric distance and room size perception*, Building Acoustics **20** (2013), no. 4, 383–401, ISSN 1351-010X, DOI 10.1260/1351-010X.20.4.383.

[MNG13] Samuel Moulin, Rozenn Nicol, and Laetitia Gros, *Auditory Distance Perception in Real and Virtual Environments*, Proceedings of the ACM Symposium on Applied Perception SAP '13 (New York, USA), ACM, 2013, DOI 10.1145/2492494.2501876, p. 117, ISBN 978-1-4503-2262-1.

[MW12] Hans-Joachim Maempel and Stefan Weinzierl, *Demands on measurement models for the perceptual qualities of virtual acoustic environments*, 59th Open Seminar on Acoustics, 2012, pp. 149–154.

[OS55] Charles E. Osgood and George J. Suci, *Factor analysis of meaning*, Journal of Experimental Psychology **50** (1955), no. 5, 325–338, ISSN 0022-1015, DOI 10.1037/h0043965.

[PKCR05] Jodie M. Plumert, Joseph K. Kearney, James F. Cremer, and Kara Recker, *Distance Perception in Real and Virtual Environments*, ACM Transactions on Applied Perception **2** (2005), no. 3, 216–233, ISSN 1544-3558, DOI 10.1145/1077399.1077402.

[RBECK11] Marc Rébillat, Xavier Boutillon, Étienne Corteel, and Brian F. G. Katz, *Audio, visual, and audio-visual egocentric distance perception in virtual environments*, EAA Forum Acusticum, Aalborg, Denmark, 2011, pp. 482–487.

[RFHN10] Björn Rasch, Malte Friese, Wilhelm Hofmann, and Ewald Naumann, *Quantitative Methoden*, 3rd ed., vol. 2, Springer, 2010, ISBN 978-3-540-33309-8, DOI 10.1007/978-3-540-33310-4.

[Rum16] Olli Rummukainen, *Reproducing reality: Perception and quality in immersive audiovisual environments*, Ph.D. thesis, Helsinki: School of Electrical Engineering, 2016.

[SvdSKvdM01] Martijn J. Schuemie, Peter van der Straaten, Merel Krijn, and Charles A. van der Mast, *Research on Presence in Virtual Reality: A Survey*, CyberPsychology & Behavior **4** (2001), no. 2, 183–201, ISSN 1094-9313, DOI 10.1089/109493101300117884.

[ZBB05] Pavel Zahorik, Douglas S. Brungart, and Adelbert W. Bronkhorst, *Auditory Distance Perception in Humans: A Summary of Past and Present Research*, Acta Acustica united with Acustica **91** (2005), no. 3, 409–420, ISSN 1610-1928.

[ZPCK09] Christine J. Ziemer, Jodie M. Plumert, James F. Cremer, and Joseph K. Kearney, *Estimating distance in real and virtual environments: Does order make a difference?*, Attention, Perception, & Psychophysics **71** (2009), no. 5, 1095–1106, ISSN 1943-3921, DOI 10.3758/APP.71.5.1096.

Citation

Hans-Joachim Maempel and Michael Horn, *Audiovisual perception of real and virtual rooms*, Journal of Virtual Reality and Broadcasting, 14(2017), no. 5, May 2018, urn:nbn:de:0009-6-46683, DOI 10.20385/1860-2037/14.2017.5, ISSN 1860-2037.