Moving Sounds Enhance the Visually-Induced Self-Motion Illusion (Circular Vection) in Virtual Reality

BERNHARD E. RIECKE
Max Planck Institute for Biological Cybernetics
ALEKSANDER VÄLJAMÄE*
Chalmers University of Technology
and
JÖRG SCHULTE-PELKUM
Max Planck Institute for Biological Cybernetics

While rotating visual and auditory stimuli have long been known to elicit self-motion illusions ("circular vection"), audio-visual interactions have hardly been investigated. Here, two experiments investigated whether visually induced circular vection can be enhanced by concurrently rotating auditory cues that match visual landmarks (e.g., a fountain sound). Participants sat behind a curved projection screen displaying rotating panoramic renderings of a market place. Apart from a no-sound condition, headphone-based auditory stimuli consisted of mono sound, ambient sound, or low/high spatial resolution auralizations using generic Head-Related Transfer Functions (HRTFs). While merely adding non-rotating (mono or ambient) sound showed no effects, moving sound stimuli facilitated both vection and presence in the virtual environment. This spatialization benefit was maximal for a medium (20°×15°) FOV, reduced for a larger (54°×45°) FOV and unexpectedly absent for the smallest (10°×7.5°) FOV. Increasing auralization spatial fidelity (from low, comparable to 5-channel home theatre systems, to high, 5° resolution) provided no further benefit, suggesting a ceiling effect. In conclusion, both self-motion perception and presence can benefit from adding moving auditory stimuli. This has important implications both for multi-modal cue integration theories and the applied challenge of building affordable yet effective motion simulators.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors, Human information processing*; H.5.1 [**Information Interfaces and Presentation, (e.g. HCI**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation, (e.g. HCI)**]: User Interfaces—*Input devices and strategies, Interaction styles*; J.4 [**Social and Behavioral Sciences**]: Psychology

General Terms: Experimentation, Human factors, Measurement

Additional Key Words and Phrases: audio-visual interactions, presence, psychophysics, self-motion simulation, spatial sound, vection, virtual reality,

1. INTRODUCTION

When presented with a moving visual stimulus that covers a large part of our visual field of view (FOV), one can experience a quite compelling illusion of self-motion ("vection") in the direction opposite of the visual stimulus motion. Most people know the phenomenon of vection from real-world experience: When one sits in a train waiting

---

Author's address: Bernhard E. Riecke, Max Planck Institute for Biological Cybernetics, Spemannstraße 38, 72076 Tübingen, Germany; e-mail: bernhard.riecke@tuebingen.mpg.de; Aleksander Väljamäe, CRAGmsa, Division of Applied Acoustics, Chalmers University of Technology, Sven Hultins gata 8a, 41296, Gothenburg, Sweden, e-mail: aleksander.valjamae@chalmers.se *(the first two authors contributed equally to this work); Jörg Schulte-Pelkum, Max Planck Institute for Biological Cybernetics; e-mail: joerg.sp@tuebingen.mpg.de;

to depart from the train station and watches a train on the neighboring track pulling out of the station, one can have the strong impression of moving oneself, even though it was in fact the train on the adjacent track that just started to move. A similar effect can be observed when one sits in a car waiting for the traffic light to turn green and a close-by large truck slowly starts to move.

While visually induced self-motion illusions have been described more than a hundred years ago and since been studied extensively [Dichgans and Brandt 1978; Fischer and Kornmüller 1930; Howard 1986; Mach 1875; Warren and Wertheim 1990], there has been surprisingly little research on auditory induced vection or cross-modal influences. The current study aims to close this gap by studying joint audio-visual contributions to vection.

Vection has typically been investigated by seating participants in the center of a rotating optokinetic drum that is painted with simple geometrical patterns like black and white vertical stripes. When stationary observers are exposed to such a moving visual stimulus, they will at first correctly perceive motion of the visual stimulus (object motion). After a few seconds, however, this perception typically shifts toward oneself being moved and the moving visual stimulus slowing down and finally becoming earth-stationary. This self-motion illusion is referred to as circular vection, and the illusion has been studied extensively for more than a century [Fischer and Kornmüller 1930; Mach 1875]. Excellent reviews on the phenomenon of vection are provided by [Dichgans and Brandt 1978; Howard 1986; Warren and Wertheim 1990]. More recently, the vection literature has also been revisited in the context of virtual reality (VR) and self-motion simulation applications [Hettinger 2002; Riecke et al. 2005b; Riecke et al. 2006; Schulte-Pelkum and Riecke 2008a]. The finding that vection can be reliably induced using VR setups has opened up new research possibilities, since VR allows to disambiguate sensory cues in a way that is not possible in the real world. Using VR, psychophysicists can dissociate the relative contributions of different sensory modalities for self-motion perception. For example, van der Steen and Brockhoff [2000] investigated visual-vestibular interactions for linear and circular vection and found that visually induced vection is quite “robust” against vestibular disturbances: Rather strong vestibular stimulation that was incongruent with visual stimulation was required to disrupt visually induced vection.

The current study addresses the question whether the self-motion illusion induced by a moving *visual* stimulus displaying a rotation around the earth-vertical axis (circular vection) might be enhanced by adding a correspondingly moving *auditory* cue. If this were the case, it would have important implications for both our understanding of multi-modal self-motion perception and optimizing virtual reality applications that include simulated movements of the observer.

1.1 Visual and auditory cues for self-motion perception

Being able to move about one’s environment and change one’s viewpoint is a fundamental behavior of humans and most animals. During self-motion, be it active locomotion or passive transportation, virtually all sensory modalities receive characteristic patterns of sensory stimulation. The visual and vestibular modalities are commonly regarded as the most important ones to perceive self-motion [Howard 1982]. Even though the auditory modality plays a rather important role in everyday life when moving about, there has been surprisingly little research on the relation between auditory cues and induced self-motion sensations. This is all the more striking as auditorily induced circular vection and audiokinethic nystagmus have been reported as early as 1923 [Dodge 1923] and later been replicated several times [Gekhman 1991; Hennebert 1960; Lackner 1977; Marmekarelse and Bles 1977]; Lackner demonstrated, for example, that an array of speakers simulating a rotating sound field can indeed induce vection in blindfolded participants [Lackner 1977]. Although rotating auditory cues can affect eye movement, it is less clear whether these eye movements qualify as an actual audiokinetic nystagmus [McFarland and Weber 1969; Watson 1968]. Only recently has auditory vection received more interest, and a small number of studies were able to induce auditory vection in at least some of the participants, both for rotational and translational motions [Kapralos et al. 2004; Larsson et al. 2004; Riecke et al. 2005; Sakamoto et al. 2004; Väljamäe et al. 2004; 2005; 2008a; 2008b]. However, compared to visually induced vection, which is quite compelling and can even be indistinguishable from real motion [Brandt et al. 1973], the auditory induced self-motion illusion is much weaker and less compelling and typically occurs only in about 25-

60% of the participants. Furthermore, participants need to be blindfolded – as soon as the stationary surroundings become visible, the illusion vanishes. Hence, even though auditory vection *can* occur, auditory cues alone are clearly insufficient to reliably induce a compelling self-motion sensation that could be employed in motion simulation applications. The question remains whether the low effectiveness of auditory cues for vection that has been found so far is due to the artificial sounds used in those studies, or whether these results reflect a low weight of auditory cues for self-motion perception in general. Some recent studies were able to show that auditory vection depends on the rendering accuracy and/or type of sounds used [Larsson et al. 2004; Väljamäe et al. 2004].While most researchers used artificial sounds (e.g., pink noise) [Kapralos et al. 2004; Lackner 1977; Sakamoto et al. 2004], Larsson et al. [2004] and Riecke et al. [2005] hypothesized that the nature or interpretation of the sound source might also be able to affect auditory vection. In line with their hypothesis, they were able to demonstrate that sound sources that are typically associated with stationary objects (so-called "auditory landmarks" like church bells) are more effective in triggering auditory circular vection than artificial sounds like pink noise or sounds that normally stem from moving objects (e.g., footsteps). These results strongly suggest the existence of higher cognitive or top-down contributions to vection, as the interpretation or meaning associated with a sound source affected the illusion. These results challenge the prevailing opinion that vection is mainly a bottom-up driven process. A more in-depth discussion of top-down and higher-level influences on auditory as well as visual vection can be found in Riecke et al. [2005] and Schulte-Pelkum and Riecke [2008a].

The finding that auditory cues *can* induce vection indicates that there might be more influence of auditory cues in self-motion perception than previously thought. The authors are aware of only one study that explicitly studied auditory vection in context of multi-sensory stimulation: Schinauer et al. [1993] investigated auditory-vestibular interactions during vection and were able to show that in ambiguous situations, e.g., near-threshold vestibular stimulation, auditory information can influence self-motion perception. In that study, a binaurally recorded moving sound field increased the perceived self-rotation magnitude when it turned opposite to the physical rotation direction (this was called the *compatible* condition), whereas ratings were reduced when the sound rotated in the same direction as the physical rotation (*incompatible* condition).

The current study takes a first step to investigate visual-auditory interactions in vection. The question we ask is whether auditory information presented in addition to visual information can enhance vection.

1.2 The problem of self-motion simulation in VR

In recent times, self-motion simulation has gained more and more importance both in the fields of professional training of pilots and vehicle drivers as well as in entertainment industry, and recent technical advances in VR technology has made motion simulation devices more affordable to research institutions. However, the question how to optimally simulate motions through space is still a matter of debate. If the aim of VR is to provide a sensory experience that is as close to the real world as possible, being able to simulate realistic self-motion is a key necessity for most VR applications. There are a number different approaches to simulating self-motion in VR, including motion platforms, free walking using head-mounted displays (HMDs), locomotion interfaces such as treadmills, or simply just presenting visual information about the self-motion. Each of these approaches has distinct disadvantages: The drawback of using motion platforms is that they require a considerable technical and financial effort, and even then performance in VR is not necessarily comparable to corresponding real-world tasks, which can be rather problematic for, e.g., driving or flight simulations [Boer et al. 2000; Burki-Cohen et al. 2003; Mulder et al. 2004]. An often used alternative is to allow users to freely walk around while wearing a position-tracked head-mounted display. For most tasks, however, this requires a rather large walking area in which the observer's position is precisely tracked – which is often infeasible or simply too costly. Using locomotion interfaces like treadmills or bicycles to allow for proprioceptive cues from physically walking or cycling etc. is often believed to be an optimal solution – there are, however, many open design and implementation issues that need to be carefully evaluated to come up with an optimal (and affordable) solution for a given task, especially if self-rotations are involved [Hollerbach 2002]. There has been only little research on the perception of self-motion (vection) using treadmills, and informal observations suggest that participants hardly ever report compelling sensations of self-motion that is comparable to vection as experienced in optokinetic drums, even in the most advanced linear treadports. Durgin et al. [2005] state, for example, that "during treadmill locomotion, there is rarely any illusion that one is actually moving forward" (p. 401). Finally, when only visual information about the self-motion is provided, users hardly ever have a convincing sensation of self-motion, especially for the relatively small FOVs that are common for off-the-shelf VR display devices.

Thus, despite tremendous progress in VR simulation technology, self-motion simulation in VR still poses a major challenge, and self-motion simulation is typically not as effective and convincing as corresponding real-world motions. This can lead to a number of problems including disorientation, reduced or mis-adapted task performance, general discomfort, and motion sickness (see, e.g., the discussion in [Chance et al. 1998; Riecke and Wiener 2007; Schulte-Pelkum and Riecke 2008a]). One of the causes of such negative side effects is the conflict between the different sensory cues indicating, for example, different directions and amounts of self-motion [Flanagan et al. 2004].

Still, it is known that moving visual stimuli *can* in certain situations be sufficient for triggering a compelling sensation of (illusory) self-motion, as is illustrated by the train illusion described above. This motivated us to investigate how far we can get without moving the observer at all, and how using VR technology might allow to optimize self-motion perception compared to the traditionally used optokinetic drums displaying abstract black and white patterns (instead of a natural scene as in the train illusion example). Given that auditory cues have been only scarcely investigated in multi-sensory VR simulations – even though affordable 3D audio technology is available – there might be potential to increase VR simulations at relatively low cost.

Other factors that have been shown to facilitate auditory vection include the realism of the acoustic simulation and the number of sound sources [Larsson et al. 2004; Riecke et al. 2005]. So far, though, there has been hardly any research on cross-modal contributions to auditory vection, and we are only aware of a study by Väljamäe et al. [2006] that showed that vibrations can enhance auditory vection, in line with experiments by Schulte-Pelkum et al. [2004] that showed a similar benefit of vibrations for visually-induced vection.

Therefore, the first experiment of the current study investigated whether *additional* spatial auditory cues can be utilized to *enhance* visually induced self-motion[1]. Even though there is a large body of literature on visual vection, audio-visual interactions for vection have hardly (if at all) been investigated before. Instead of using the classic black-and-white striped patterns as vection-inducing visual stimulus – which is not really suitable for VR applications – we opted here for using naturalistic visual and auditory stimuli that have previously been shown to be more powerful in inducing visual and auditory vection than more artificial and thus ecologically less valid stimuli [Riecke et al. 2006; Larsson et al. 2004]. Furthermore, using naturalistic stimuli allowed us to use the sense of presence in the presented environment as a corroborative measure of our simulation success. Presence is often referred to as the perceptual illusion of non-mediation [Lombard and Ditton 1999], and different aspects of presence have been shown to correlate differentially with different aspects of the self-motion illusion [Riecke et al. 2006; Schulte-Pelkum and Riecke 2008a]. While Experiment 1 compared a spatialized audio condition with a mono and a no-sound condition, the second experiment addressed the influence of the visual FOV and the rendering quality of the auditory cues.

2. EXPERIMENT 1 – INFLUENCE OF ADDITIONAL AUDITORY CUES ON VISUALLY INDUCED VECTION

Previous studies have only used visual and auditory vection-inducing stimuli in isolation. Here, we were interested in joint audio-visual contributions and interactions to vection. The first experiment was designed to test two different hypothesis of how additional auditory stimuli might potentially facilitate visually induced vection:

Hypothesis 1: Influence of adding non-spatialized (mono) auditory stimuli – unspecific *facilitation.* First, adding sound to the visual scene might increase the overall believability of the motion simulation and the resulting presence and involvement in the simulated scene and thus *unspecifically* facilitate vection, independent of the spatial content

---

[1] The first experiment is in part based on a conference presentation at the 8th Annual Workshop on Presence [Riecke et al. 2005a]

or location of the auditory stimulus. To investigate if such unspecific facilitation occurs, we compared a no-sound condition with a simple mono rendering of an auditory landmark in the scene (the sound of the fountain on the market place scene that was used as the visual stimulus).

Hypothesis 2: Influence of adding a spatialized, concurrently rotating auditory landmark – specific *facilitation.* Second, if an auditory simulation includes spatialized auditory landmarks, the spatial content of the auditory simulation could *specifically* enhance vection by providing acoustic flow and additional information about the spatial location of an auditory landmark and hence the current orientation of the observer. This hypothesis was tested by comparing the above-mentioned mono-condition with a spatialized acoustic rendering of the correct location of the landmark generated using binaural synthesis and a generic head-related transfer functions (HRTFs) catalog (see subsection 2.1.1 below). In addition, it is conceivable that the simulation might appear more realistic in the spatialized condition if the auditory landmark appears properly externalized and spatialized. This might also increase overall believability and presence in the simulated scene [Hendrix and Barfield 1996; Ozawa et al. 2004; Väljamäe et al. 2004].

Fig. 1. **Top:** 360° roundshot photograph of the Tübingen market place, which was wrapped onto a cylinder to provide an undistorted view of the scene for the simulated viewpoint centered in the cylinder. **Bottom:** Participants were seated at a distance of about 1.8m from a curved projection screen (left) displaying a view of the market place (right).

2.1 Methods

Twenty naive participants (eight male) took part in this experiment (mean age: 28.7 years, SD: 8 years) and were paid at standard rates. All participants had normal or corrected-to-normal vision and were able to locate the spatialized sound source without any problems. For both experiments reported in this manuscript, participants gave their informed consent prior to the experiment. The current study has been approved by the local ethics committee and has therefore been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

2.1.1 *Stimuli and Apparatus* .

Setup . Participants were comfortably seated at a distance of 1.8m from a curved projection screen (2m curvature radius) on which the rotating visual stimulus was displayed (see Fig. 1, bottom). The visual stimulus consisted of a photorealistic view of the Tübingen market place that was generated by wrapping a 360° roundshot (4096 × 1024 pixel) around a virtual cylinder (see Fig. 1, top). The simulated FOV was set to $54 \times 45$and matched the physical FOV under which the projection screen was seen by the participants. The computer-rendered images were projected non-stereoscopically using a JVC D-ILA DLA-SX21S video projector with 1400×1050 pixel resolution at an update rate of 60Hz (V-sync on). In order to increase immersion and block vision of the outside room, black curtains covered the side and top of the cabin surrounding the projection screen. A video clip of the visual stimulus is available at http://www.kyb.mpg.de/bu/people/bernie/Riecke_et_al_08_TAP/Riecke_et_al_08_TAP_movie.mpg. Vection responses were recorded using a force-feedback joystick (Microsoft force feedback 2) that was mounted in front of the participants. Visual circular vection was induced by rotating the visual stimulus around the earth-vertical axis with alternating turning direction (left/right) for consecutive trials.

Viewing instructions. Participants were instructed to watch the stimuli in a natural and relaxed manner, just as if looking out of the window of a moving vehicle. Furthermore, they were told to neither stare through the screen nor to fixate on any position on the screen in order not to suppress the optokinetic reflex. Instead, they were instructed to concentrate on the central part of the projection screen.

Auditory stimuli. In some conditions, auditory cues were displayed using active noise-canceling headphones (Sennheiser HMEC 300) that participants wore throughout the experiment. The auditory cues consisted of a mix of several recordings of river and fountain sounds that resembled the sound of the fountain that is located on the market place used as the visual stimulus. It turned out that a simple fountain recording did not sound very convincing, and did not have the characteristics and frequency spectrum required for optimal sound localization and masking of surrounding sounds of the lab. To achieve this, several qualitatively different fountain-like sounds were mixed and equalized to yield a relatively flat frequency spectrum with a frequency range of about 0.1 to 13 kHz. Active noise cancellation was applied throughout the experiment to eliminate auditory cues from the surrounding room that could have interfered with the experiment.

Spatialized sound rendering. In the spatialized auditory condition, we used a generic HRTFs and a Lake DSP system (Huron engine) with BinScape rendering which provided an auditory rendering that closely mimics 5-channel Dolby surround systems, and places the five virtual sound sources at positions of 0, ± 45, ± 135. Note that this produces a non-uniform distribution of sounds source with angular distances varying between 45° and 90°. Binaural (headphone-based) synthesis of rotating sound fields is based on the convolution of a given sound with the HRTFs corresponding to its virtual positions in space with respect to the listener's head [Kleiner and Dalenbäck 1993]. Prerecorded HRTFs catalogs have finite spatial resolution. Hence, interpolation is used to assure the smooth rendering of a moving sound source. A typical resolution of current HRTFs catalogs is 5° in the horizontal plane [Algazi et al. 2001; Gardner and Martin 1995], which allows for perceptually transparent synthesis of spatially dynamic auditory stimuli (see Experiment 2). However, lower resolutions can be used to assure real-time, interactive rendering of complex virtual acoustic scenes. As a consequence, binaural synthesis with low spatial resolution provides reduced spatial auditory information about the location and motion of sound sources and can lead to noticeable artifacts (e.g., reduced smoothness of the motion trajectory of the auditory object). Previous research on purely auditory-induced vection showed that such artifacts can significantly decrease spatial presence in blindfolded participants [Väljamäe et al. 2004]. When auditory cues were combined with matching naturalistic visual cues as in the current experiments, however, such artifacts seemed to play less of a role, potentially because auditory localization is far inferior to vision, which might have resulted in a visual capture or ventriloquism effect.

Perception and interpretation of the auditory stimuli. Previous studies demonstrated that the fountain sound used here represents an "auditory landmark" – that is, an ecologically plausible sound source that is normally expected to be stationary [Larsson et al. 2004]. Such auditory landmarks were shown to be more effective in inducing vection than sound sources that are not expected to be stationary (like the sound of footsteps or a driving car). This study also showed that modeling the proper room acoustics of the market place scene did not affect auditorily induced vection, we did not simulate room acoustic cues for the experiments presented in this study. Note that in the auditory conditions, the fountain sound was always audible (as we have omni-directional hearing), even when the visual counterpart was outside of the current FOV. Participants perceived the spatialized fountain sound properly externalized and associated it readily with the visual counterpart as intended. None of the participants commented on any mismatch between the spatialized auditory cues and visual counterpart. In the mono sound condition, the sound was perceived "inside the head" as is to be expected for mono sound.

2.1.2 *Procedure and experimental design.* Each participants performed 48 trials, consisting of a factorial combination of 3 auditory conditions (no sound, mono sound, HRTF-spatialized sound; these conditions were randomized within each session) × 2 turning directions (left/right; alternating) × 2 sessions × 4 repetitions of each condition. Participants were instructed to indicate the **onset of vection** by deflecting the joystick in the direction of perceived self-motion as soon as it was sensed. The amount of deflection indicated the **vection intensity**, and the time between vection onset and maximum vection (joystick deflection) reached indicated the **vection buildup time**. After each trial, participants indicated the **convincingness** of the perceived self-motion on a 0-100% rating scale (in steps of 10%) using a lever next to the joystick. Participants were instructed that the value of 0% corresponded to no perceived motion at all (i.e., perception of a rotating stimulus and a stationary self) and that of 100% to very convincing sense of vection (i.e., perception of a stationary stimulus and a rotating self).

Participants started each trial by pressing a dedicated button on the joystick, which caused the static image to start rotating clockwise or counterclockwise (alternating, in order to reduce motion after-effects) around the earth-vertical axis with constant acceleration for 3s, followed by a constant velocity phase at 30°/s. The maximum duration of constant velocity rotation was 39s, after which the stimulus decelerated at a constant rate for 3s. Stimulus motion stopped automatically once maximum joystick deflection (vection intensity) was sustained for 10s (otherwise it continued for 39s) to reduce the potential occurrence of motion sickness. Participants were asked to initiate each trial themselves to ensure that they could prepare for the next trial and paid attention to the stimulus[2].

Between trials, there was a pause of about 15 seconds to reduce potential motion aftereffects. In order to familiarize participants with the setup, a practice block containing 4 trials preceded the main experimental blocks. Furthermore, because none of the participants had experienced vection in the laboratory before, they were exposed, prior to beginning the practice block, to a visual vection stimulus for about 2 minutes or until they reported a strong sense of self-motion. In addition to the vection measures, **presence** was assessed after the experiment using the 14-item Igroup Presence Questionnaire (IPQ) by Schubert et al. [2001] for each of the three auditory conditions. The 14 items of the presence questionnaire are grouped into four subscales ("Involvement", "Realism", "Space", and "Being there"), and we aimed at testing if these different aspects of presence are affected differentially by the additional auditory cues.

---

[2] This procedure is not uncommon in psychophysical studies and implies that they might have been able to anticipate vection. We are, however, not aware of any study showing that this anticipation has any detrimental effect on the resulting data. If anything, we would rather expect that it might reduce the within-subject variability or random noise, as participants could start the next trial when they were ready for it and focusing on the stimulus to be presented.

Such a potential effect could be important both for our understanding of presence and for applications, where auditory cues could be employed to specifically enhance certain aspects of the VR simulations.

After completing the presence questionnaire, participants were asked to rate the **spatial aspects of the auditory simulation** on a 7-point Likert-like scale in terms of three questions: 1.) "I was able to localize sounds" (henceforth referred to as "**localization**"); 2.) "I sensed that the sound was surrounding me" ("**surrounding**"); and 3.) "I sensed that the sound was moving" ("**moving**"). The values of 0 and 7 indicated full disagreement and agreement, respectively. After completing the experiment, participants were debriefed, thanked, and paid for their participation.

Normalization procedure. The data showed considerable overall between-subject differences in all dependent measures. To remove those systematic between-subject differences without changing the mean values of the individual measures, the following normalization procedure was applied: Each data point per participant was divided by the ratio between the mean performance of that participant across all conditions and the mean of all participants across all conditions.

2.2 Results and discussion

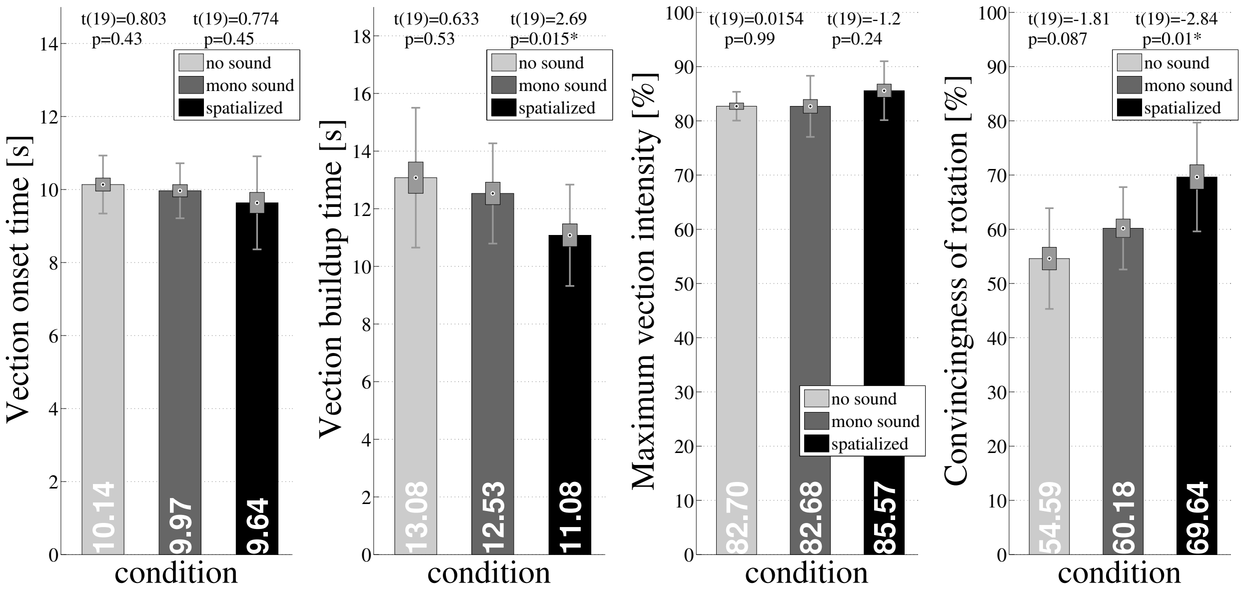

Fig. 2. Normalized mean values for the four vection measures, averaged over the 20 participants. Boxes indicate one standard error of the mean, whiskers depict one standard deviation. The results of pairwise post-hoc comparisons between the no-sound and mono condition (testing hypothesis 1) and between the mono and spatialized sound condition (testing hypothesis 2) using paired t-tests are indicated in the top inset of each plot. Due to the 2 post-hoc t-tests per dependent measure, the α-levels were adjusted from 5% to 5/2 = 2.5%. An asterisk '*' indicates significant differences. Note the small but consistent vection-facilitating effect of the spatialized auditory rendering of the fountain sound (right, black bars) as compared to simple mono display (middle, dark gray bars) or the no-sound condition (left, light gray bars). Differences between the mono sound and no sound condition did not reach significance.

Data from the different dependent measures of vection, presence, and auditory ratings were submitted to separate repeated-measures 1-way ANOVAs for the three auditory conditions (no sound, mono sound, and HRTF-spatialized

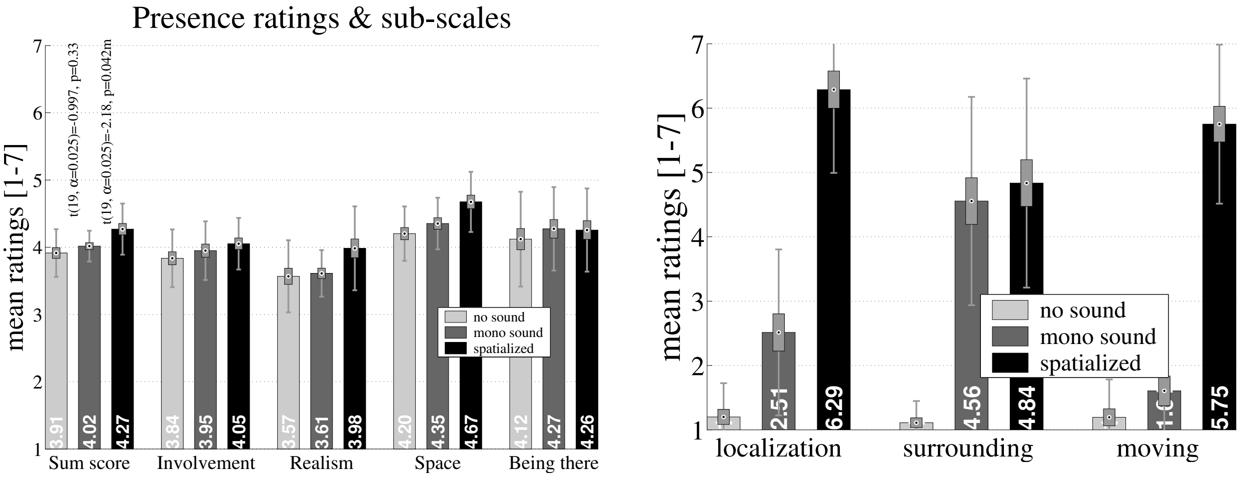

Fig. 3. **Left:** Presence ratings for the three auditory conditions, plotted as in Figure 2. The sum score over all 14 items of the Igroup Presence Questionnaire (left three bars) were split up according to the four original sub-scales described by Schubert et al. (2001): "Involvement", "Realism", "Space", and "Being there". Even though the effect size was quite small ($\leq 6\%$), there was a consistent trend towards higher ratings for the spatialized sound condition, and the effect strengthsη^2 were as high as 19%. **Right:** Mean ratings of the spatial aspects of the auditory stimulation. Note that participants unexpectedly rated the mono and spatialized sound similarly in terms of feeling surrounded by it. Note that we included audio ratings in the no-sound condition to test for reliability of participants' answers: They were clearly expected to answer with "0", which almost all of them did, thus confirming the overall reliability of their responses.

sound). Vection data were averaged over the two turning directions (left/right) prior to that. Greenhouse-Geisser correction was applied whenever unequal variances occurred. The results of the statistical analysis are summarized in Tables I, II, and III. The data are plotted in Figures 2 and 3. To test for hypothesis 1 and 2 explicitly, additional post-hoc paired t-tests with adjusted α-levels were performed for the vection data and presence sum scores, as indicated in the top inset of Figures 2 and 3.

2.2.1 *Vection measures.* The ANOVA results in Table I showed that both the buildup time and convincingness of vection was significantly affected by the auditory stimulus. The effect strength η^2 of .158 and .332, respectively, indicates that 15.8% and 33.2% of the variance observed in the data is accounted for by the auditory contribution. Note that neither the onset time nor the intensity of vection was significantly affected by the experimental manipulation, though.

	Vection onset time	Convincingness of vection	Vection intensity	Vection buildup time
auditory condition	F(1.37, 26.1) = .936	**F(2, 38) = 9.46**	F(1.36, 25.9) = 1.62	**F(2, 38) = 3.55**
	p = .37	**p < .0005*****	p = .22	**p = .038***
	η^2= .047	**η^2= .332**	η^2= .078	**η^2= .158**

Table I. Table of ANOVA results and effect strengths η^2 (which indicates the percentage of variance explained by a given factor) for the four vection measures. Significant effects are typeset in boldface.

Influence of adding non-spatialized (mono) auditory stimuli (Hypothesis 1). Even though there was a trend towards increased convincingness ratings when the mono sound was added, the post-hoc statistical analysis in Figure 2 showed no significant difference between the no sound and mono sound condition for any of the vection measures, thus providing no support for hypothesis 1.

Influence of adding a spatialized, concurrently rotating auditory landmark (Hypothesis 2). Comparing the mono sound condition with the spatialized sound condition demonstrates a small but consistent vection-facilitating effect of the sound spatialization. The strongest effect was observed for the convincingness ratings (16% increase, $\eta^2 = .299$) and the vection buildup time (12% decrease, $\eta^2 = .275$), cf. Table V. As expected from the ANOVA analysis, the other vection measures show only small and insignificant trends, albeit in the expected direction.

2.2.2 *Presence ratings*

	Presence sum score	Involvement subscale	Realism subscale	Space subscale	"Being there" subscale
auditory condition	**F(2, 38) = 4.14** **p = .024*** η^2**= .179**	F(2, 38) = .904 p = .414 η^2= .045	F(1.48, 28.1) = 2.63 p = .103 η^2= .121	**F(2, 38) = 4.55** **p = .017*** η^2**= .193**	F(2, 38) = .223 p = .802 η^2= .012

Table II. Table of ANOVA results and effect strengths η^2 (which indicates the percentage of variance explained by a given factor) for the presence sum score and the four subscales. Significant effects are typeset in boldface.

The left subplot of Figure 3 suggests a gradual – albeit small – increase in presence when adding mono and spatialized sound. The ANOVA analysis showed that this trend reached significance for the presence sum score and the "space" subscale, and explained 17.9% and 19.3% of the observed variance, respectively (see Table II). In addition, the "realism" subscale showed a marginally significant trend with η^2= .121. Note that neither the "being there" nor the "involvement" subscales were affected by the auditory cues ($\eta^2 < 5\%$). Figure 3 suggests that the presence-facilitating effect of the added sound cues is mostly do to the spatialization of the sound (hypothesis 2). Post-hoc t-tests on the presence sum score showed indeed no significant differences between the no-sound and mono sound condition, while comparing the mono and spatialized sound condition showed at least a marginally significant trend (cf. Fig. 3), which explained 20% of the observed variance (cf. Table V). Even though the current data is not fully conclusive, they suggests that adding spatialized auditory cues did not result in an *unspecific* increase in the overall feeling of presence and "being there" (in contrast to predictions from hypothesis 1), but rather *specifically* enhanced the perceived realism and feeling of the simulated space surrounding the observer (consistent with hypothesis 2, thus paralleling the results for the vection measures), despite the naturalistic and compelling visual cues.

2.2.3 *Auditory ratings*

	localization	surrounding	moving
auditory condition	F(1.25, 23.8) = 77.2 p < .0005*** η^2= .802	F(1.07, 20.3) = 32.1 p < .0005*** η^2= .628	F(1.34, 25.5) = 86.5 p < .0005*** η^2= .820

Table III. Table of ANOVA results and effect strength (partial η^2) for the ratings of the spatial aspects of the auditory stimulation.

The data for the three auditory ratings are summarized in Figure 3 (right) and Table III. As expected, the auditory ratings were quite high for the spatialized audio condition and close to the bottom of the scale for the no-sound condition (in fact, all but one participants reported consistently with "1"). Interestingly, 12 of the 20 participants indicated that that they were to some extent able to localize sounds in the mono condition (indicated by ratings above 1 in the "localization" question). Moreover, five participants reported that they perceived the sound in the mono condition as moving. This result was rather surprising, and might suggest some kind of visual capture of the auditory cues, much like a ventriloquist effect: It seems as if a compelling moving visual stimulus can to some extent suggest a spatialization of a sound cue, even when the sound cue was clearly not spatialized (i.e., a mono rendering). Unexpectedly, the sense that the sound was surrounding the participant was rated equally high for the mono sound and the spatialized sound condition. Again, one might argue that this shows a cross-modal benefit of providing a compelling, naturalistic visual stimulus. These results might be interesting in terms of applications, as they seem to suggest that the requirements in terms of auditory spatialization fidelity could be relaxed somewhat if compelling visual cues are provided.

2.2.4 *Factor analysis for presence questionnaire and correlation to vection measures.* A factor analysis on the 14 individual questions of the presence questionnaire showed a two-dimensional structure of presence: Factor 1 contained items about spatial presence and the realism of the scene (ten items), while Factor 2 contained items addressing involvement and attentional aspects (4 items, which coincide with the original "involvement" subscale of the questionnaire [Schubert et al. 2001], see Figure 3). Factor 1 and 2 explained 59.4% and 16.3% of the variance in the presence data, respectively. Interestingly, the same two-dimensional structure of presence with the same two factors were found in an earlier experiment on visually induced circular vection that did not contain any spatialized auditory cues [Riecke et al. 2006]. In that study, Factor 1 (realism/spatial presence) showed a positive correlation with the convincingness ratings of vection, while Factor 2 (attention/involvement) was negatively related to the vection onset latency. A similar correlation analysis between the presence and vection measures of the current experiment replicated the positive correlation between Factor 1 (realism/spatial presence) and the convincingness of the self-motion illusion (Pearson correlation coefficient $r = .707$, $p < .0005^{***}$). None of the other correlations between Factor 1 or 2 and the four vection measures showed any significant correlation, though (all $r < .222$, $p > .35$). In sum, this suggests that the convincingness of the self-motion illusion is closely related to the realism and spatial presence aspects of the simulation, which has interesting implications not only for our basic understanding of self-motion perception, but also in terms of designing convincing self-motion simulations.

The three additional post-experimental questions about spatial aspects of the auditory stimulation ("localization", "surrounding", and "moving", see subsection above and Fig. 3) did not show any significant correlations with any of the presence or vection measures. Thus, the spatial aspects of the auditory cues might either not be closely related to presence and/or vection, or the effect size was too small and/or masked by the noise in the data.

2.2.5 *Importance of sound spatialization and not just sound per se.* Even though adding mono sound marginally increased the convincingness of the motion simulation by about 10%, neither the presence ratings nor any of the other vection measures were affected at all. That is, merely adding an audio cue that is associated with the fountain on the market place but not spatially aligned with it did not increase vection or presence significantly. This argues against an *unspecific* benefit of just adding audio cues (hypothesis 1). Only when the sound source was actually perceived to originate from the same location as it's visual counterpart did we observe a facilitation of vection and presence. This argues for a *specific* facilitation due to the spatialization of the sound source, thus supporting hypothesis 2. These results suggest that cross-modal consistency is indeed an important – albeit often overlooked – factor in improving VR simulations from a perceptual point of view. This issue is all the more relevant as most existing VR simulations have rather poor audio quality, especially in terms of localizability of the sound sources (and externalization if headphone-based auralization is used).

2.2.6 *Why do auditory cues only help so little?.* In the debriefing after the experiment, participants rated the motion simulation as being much more convincing when spatialized sound was included. Nevertheless, the effect size of adding spatialized sound was rather small, both in terms of vection and rated presence. One potential reason for the relatively small auditory contribution might be the different vection-inducing potential of the visual and auditory stimuli: On the one hand, the visual stimuli covered a relatively large FOV and depicted a globally consistent natural scene, which has previously been shown to be a rather powerful vection-inducing stimulus [Riecke et al. 2006]. This was confirmed by the overall high intensity and convincingness of vection and the relatively low vection onset latencies even without the auditory cues, which might have resulted in a ceiling effect. On the other hand, auditory cues are known to be in general far less potent in inducing vection than visual cues [Larsson et al. 2004; Väljamäe et al. 2004; 2006]. Furthermore, Experiment 1 used only a relatively crude auditory spatialization (5 spatial locations, no headphone equalization, only one auditory landmark, and no room acoustics simulation). Together, these factors might have contributed to a visual dominance and/or ceiling effect.

The second experiment was designed to address these issues and investigate audio-visual contributions to vection in a setting where

(1) the vection-inducing power of the visual cues is *reduced* (by decreasing the visual FOV from 54° × 45° (Exp. 1) to 20°× 10° or 10°× 7.5°) and

(2) the vection-inducing power of the auditory cues is *increased* by using a high-quality auralization with 5° resolution, headphone equalization, adding a second auditory landmarks, and additional ambient sound in half of the trials.

3. EXPERIMENT 2 – INFLUENCE OF VISUAL FOV AND AUDITORY SPATIAL RENDERING QUALITY

The second experiment further investigated the facilitation of visually-induced circular vection by concurrently presenting rotating sound fields. As merely adding non-spatialized (mono) sound affected neither vection nor presence in Experiment 1 (thus providing no support for hypothesis 1), the no-sound condition was omitted for the second experiment. Instead, the mono sound condition served as a baseline condition for two spatialized auditory conditions with different rendering quality. In addition to hypothesis 2 (adding spatialized sound facilitates vection and presence), three more hypotheses were tested in Experiment 2:

Hypothesis 3: Influence of auditory spatialization fidelity. To investigate the effect of auditory spatialization fidelity on vection and presence, a mono sound condition was compared with two auditory spatialization conditions of different rendering quality: a low-resolution condition (5 spatial locations BinScape setting as in the previous experiment, which the quasi-standard for consumer surround sound systems) and a high-resolution auditory rendering with 72 spatial locations resulting in a 5° spatial resolution in the horizontal plane (which is the typical resolution of most HRTF catalogs). If localization accuracy was the determining factor for the auditory effects found in experiment 1, one might expect that rotating sound fields synthesized using high spatial resolution rendering might have a stronger impact on self-motion responses than low fidelity auralization.

Hypothesis 4: Influence of visual FOV and interaction with auditory facilitation of vection. As an attempt to magnify the auditory contributions to vection and avoid possible ceiling effects and visual dominance that might have occurred in Experiment 1, the visual FOV was reduced to a medium (20°×15°) or small (10°×7.5°) size for the second experiment. If the relatively small effect size for the auditory facilitation of vection in experiment 1 was (at least in part) caused by a visual dominance or ceiling effect, one might expect an increased auditory contribution to vection when decreasing the visual FOV. This would predict that the vection-enhancing effect be strongest for the smallest visual field size.

Hypothesis 5: Influence of ambient sound. Most natural environments contain not only a few distinct, clearly audible sound sources (like the fountain sound used in Experiment 1), but also some ambient, hardly localizable sounds produced by for example wind, distant traffic, or humans talking, thus creating a uniform auditory space surrounding the listener. To investigate whether such ambient sound might contribute to audio-visual vection and presence in a simulated environment – a question that has apparently not been addressed before – the normal sound rendering with the two individual auditory landmarks was compared to a condition where ambient sound was added. The ambient condition was intended to clarify whether the vection facilitation effect observed in Experiment 1 was caused by an auditory scene "spaciousness" rather then by the congruency of audio-visual rotational stimuli. If this was the case, the ambient sound-scape would be expected to enhance the self-motion sensation.

3.1 Methods

Twelve naive participants (10 male) with a mean age of 26 years (SD: 5 years) took part in Experiment 2 and were payed at standard rates. All participants had normal or corrected-to-normal vision and were able to localize the spatialized sound without any problems.

3.1.1 *Stimuli and apparatus.* The stimuli and procedures were identical to the first experiment apart from a few changes described below. While the first experiment was conducted in a lab in Tübingen, Germany, the second experiment was performed in Göteborg, Sweden using a replica of the setup in Tübingen that was developed in the context of an European project on Perceptually Oriented Ego-Motion Simulation ("POEMS", see www.poems-project.info). The visual stimulus was the same as in Experiment 1 (a photorealistic view of the Tübingen market place), but with an added simulated black mask that reduced the FOV to 20°×15° (medium FOV condition) or 10°×7.5° (small FOV condition) as shown in Figure 4. As vection was expected to be reduced for the smaller FOV, the constant velocity phase was extended from 39s (Experiment 1) to 49s. Virtual acoustic scenes were synthesized using a Lake Huron DSP system and presented using Beyerdynamic DT-990Pro circumaural headphones (that did not use active noise cancellation). Headphone equalization was applied to in order to prevent coloration artifacts and to increase sound externalization.

Fig. 4. The VR setup and stimuli used for the second experiment were identical to the one used in Experiment 1 apart from a black mask that limited the FOV to either 20°×15° (medium FOV condition) or 10°×7.5° (small FOV condition) as illustrated by the white rectangles.

As an attempt to enhance the auditory contribution to vection and thus make the vection-inducing potential of the visual and auditory cues at least somewhat more comparable, the following additional changes were made to the auditory simulation, as compared to Experiment 1:

Sound rendering resolution and spatialization. In addition to the monophonic sound condition which served as a baseline, two different spatialized auditory rendering conditions were used: A high fidelity binaural synthesis with a 5° resolution ("HeadScape" setting of the Lake Huron system, mimicking a 72 channel/spatial locations reproduction, henceforth referred to as "high resolution" condition), and a medium-fidelity binaural synthesis where the generic HRTF catalog was reduced to a new set with just 5 spatial locations (Dolby surround-like "BinScape" setting as in Experiment 1, henceforth referred to as "low resolution" condition).

Number of sound sources. Väljamäe et al. [2004] and Larsson et al. [2004] demonstrated that auditorily induced vection in blindfolded participants can be enhanced if more than one spatialized sound source is used. This motivated us to include a second auditory landmark for Experiment 2: the sound of a van on idle, which was added to the auditory scene at the location where the visual stimulus depicts a white van (see Fig. 1). The two auditory landmarks (van and fountain) were spaced 120° apart. As before, the frequency range of the spatialized sound ranged from 0.1 to 13 kHz.

Ambient sound. To investigate the influence of ambient sound on vection and presence in VR, a non-rotating[3], ambient soundscape was added to the visual scene in half of the trials. The ambient sound was binaurally recorded on the actual market place that served as the visual stimulus, and contained typical small town square background sounds like distant speech, wind, traffic noise and bird's tweets. Due to geometry and acoustic properties of the market place, the recorded ambient sound was difficult to localize despite being clearly audible.

3.1.2 *Procedure and experimental design.* Each participant was presented with 24 trials as a result of a of $3 \times 2 \times 2 \times 2$ design containing 3 auditory spatialization conditions (mono sound, low resolution and high resolution spatial sound; in separate blocks) $\times$ 2 turning directions (left/right; alternating) $\times$ 2 different FOVs (medium, 20°×15°, or small, 10°×7.5°; randomized) $\times$ 2 ambiance conditions (with or without ambient sound; randomized). The three auditory conditions were presented in three separate blocks due to the need to restart the Lake auralization system each time a new sound rendering engine was used. The order of the three blocks was balanced across participants. Before the actual experiment, participants received verbal instructions and performed a short training session (2 trials) to acquaint them with the experimental procedures. At the end of each trial, participants provided verbal ratings of the convincingness of the perceived self motion and the overall presence defined as "the sensation of being actually present in the virtual world". Due to time constraints, we refrained from using the full presence questionnaire for each of the twelve different conditions and asked just this one question instead.

3.2 Results and discussion

Data from the joystick deflection (vection onset time, buildup time, and intensity) and participants' verbal responses (convincingness of vection and presence) were submitted to separate 3 (sound spatial fidelity) $\times$ 2 (FOV) $\times$ 2 (ambiance) within-subject ANOVAs. Left/right turning direction had been averaged prior to that. Greenhouse-Geisser correction was used whenever unequal variances occurred. The results of the statistical analysis are summarized in Tables IV and V. The factor ambient sound did not show any significant effects, and the data was correspondingly pooled. The pooled data for the vection and presence measures are plotted in Figures 5 and 6 in the identical format as Experiment 1 for easier comparison. As for Experiment 1, additional post-hoc pairwise t-tests with adjusted α-levels were performed to explicitly test for hypothesis 2 and 3. Results are displayed in the top insets of Figure 5.

---

[3] Due to technical difficulties, we were unfortunately not able to simulate ambient sound that rotated in synchronization with the visual stimulus. This might, of course, potentially have created a conflict between the rotating visual stimulus and auditory landmarks and the non-rotating nature of the ambient recording, despite the ambient recording being hard to localize.

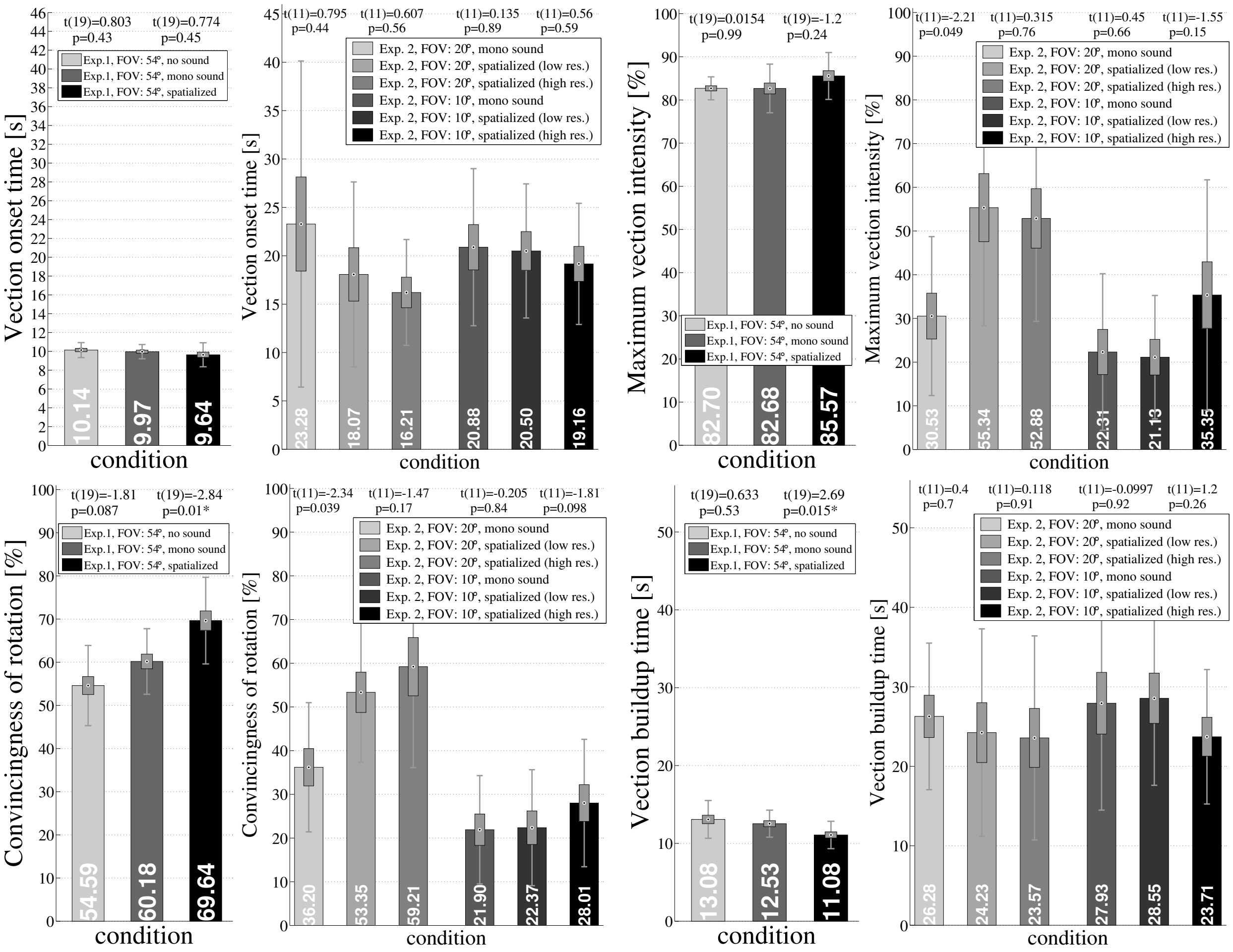

Fig. 5. Normalized means of the vection measures of Experiment 2 for the three auditory conditions and two visual FOVs, plotted as in Figure 2. The results of pairwise post-hoc comparisons between the mono and low resolution auditory condition (testing hypothesis 2) and between the low and high auditory resolution conditions (testing hypothesis 3) using paired t-tests are indicated in the top inset of each plot. Due to the 4 post-hoc t-tests per measurand, the α-levels were adjusted from 5% to 5/4 = 1.25%. For comparison, the data from Experiment 1 was re-plotted with the same scale. Note that the smaller FOV in Experiment 2 decreased vection in all dependent measures, compared to Experiment 1. The vection-enhancing effect of the spatializing the auditory cues was most pronounced for the medium FOV condition (20°×15°).

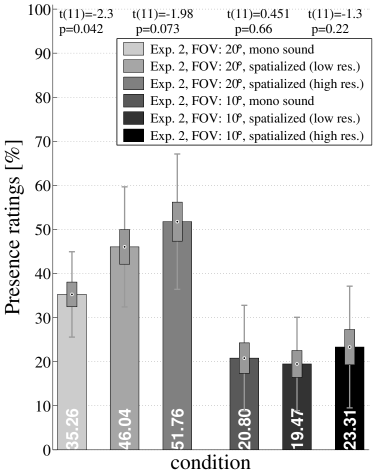

Fig. 6. Normalized presence ratings for the three auditory conditions and two FOVs of Experiment 2, plotted as in Figure 5. Note that spatializing the auditory cues enhanced presence only for the medium FOV (20°×15°) condition.

	Vection onset time	Convincingness of vection	Vection intensity	Vection buildup time	Spatial presence
Field of view (FOV)	F(1,11) = 0.11	**F(1,11) = 13.77**	**F(1,11) = 7.23**	F(1,11) = 0.65	**F(1,11) = 15.74**
	p = 0.75	**p = 0.003****	**p = 0.02***	p = 0.44	**p = 0.002****
	η^2= .010	η^2**= .556**	η^2**= .397**	η^2= .056	η^2**= .589**
Auditory rendering	F(2,17) = 0.94	**F(1,16) = 5.89**	*F(2,21) = 3.34*	F(2,18) = 0.38	**F(2,19) = 4.1**
	p = 0.38	**p = 0.022***	*p = 0.056m*	p = 0.64	**p = 0.04***
	η^2= .079	η^2**= .342**	η^2*= .233*	η^2= .033	η^2**= .271**
Interaction rendering–FOV	F(1,15) = 0.74	**F(2,21) = 4.31**	*F(2,21) = 3.11*	F(2,21) = 0.21	**F(2,18) = 8.29**
	p = 0.44	**p = 0.038***	*p = 0.074m*	p = 0.77	**p = 0.004****
	η^2= .063	η^2**= .282**	η^2*= .220*	η^2= .019	η^2**= .430**

Table IV. Table of ANOVA results and effects strengths (partial η^2) for the five different dependent variables. Significant effects are typeset in boldface, marginally significant effects in italics.

3.2.1 *Decreasing the FOV reduces vection and presence.* As expected, the reduction of the size of the visual FOV significantly decreased presence as well as the intensity and convincingness of vection (see Table IV and Figures 5 and 6). Vection onset time and buildup time showed a similar same tendency toward reduced vection for the smaller FOV, but this trend did not reach significance. Compared to Experiment 1, where the FOV was more than twice as large, vection was overall much weaker for Experiment 2 as predicted (see Fig. 5).

3.2.2 *Influence of adding spatialized, concurrently rotating auditory landmarks (Hypotheses 2) and increasing the auditory spatialization fidelity (Hypothesis 3).* Increasing the auditory rendering quality significantly enhanced the convincingness of the self-motion illusion (see Table IV). In addition, the vection intensity was marginally increased ($p = .056$). This increase in the convincingness and intensity of vection accounted for $\eta^2 = 34.2\%$ and $\eta^2 = 23.3\%$ of the observed variance, respectively. None of the post-hoc t-tests showed any significant differences between the mono and low resolution condition (hypothesis 2) or the low resolution and high resolution condition (hypothesis 3), though (see Fig. 5). Thus, we are unable to clearly disambiguate between hypothesis 2 and 3 from the current data,

	Exp. 1 (54° FOV)		Exp. 2 (20° FOV)		Exp. 2 (10° FOV)	
	η^2	% change	η^2	% change	η^2	% change
Vection onset time	.031	3.3% decrease (n.s.)	.054	22.4% decrease (n.s.)	.002	1.8% decrease (n.s.)
Vection build-up time	**.275**	**11.6% decrease (*)**	.014	7.8% decrease (n.s.)	.001	2.2% increase (n.s.)
Vection intensity	.071	3.5% increase (n.s.)	.307	81.3% increase (n.s.)	.018	5.3% decrease (n.s.)
Convincingness of v.	**.299**	**15.7% increase (*)**	.333	47.4% increase (n.s.)	.004	2.1% increase (n.s.)
Presence	*.20*	*6.2% increase (m.s.)*	.324	30.6% increase (n.s.)	.018	6.4% decrease (n.s.)

Table V. Table of the relative effect strength η^2 and relative effect sizes (relative change in %) between the mono sound condition and the low resolution (5 locations BinScape) rendering for Experiment 1 and 2. η^2 serves as a measure of the variance explained by that factor. An Asterisk "*" indicates significant effects, whereas "*m.s.*" and "n.s." denotes marginal or no significant effects, respectively. Note that the small FOV condition did not show any significant effects, and sometimes even showed trends in the opposite direction (typeset as underlined).

even though the ANOVA showed significant overall effects of increasing the auralization fidelity. This suggests a *gradual* increase in vection and presence with increasing auditory rendering quality. Furthermore, it seems that the resolution of spatial sound plays only a minor role in audio-visually induced vection, in agreement with other studies on cross-modal motion perception where spatial aspects of moving stimuli seem to be dominated by vision and not audition [Soto-Faraco et al. 2003] (see also the subsection below and the general discussion.)

3.2.3 *Influence of visual FOV and interaction with auditory facilitation of vection (Hypothesis 4).* The data revealed an unexpected differential effect of FOV on the sound-induced facilitation of vection: In line with our predictions (hypothesis 4), the medium FOV condition (20°×15°) induced stronger sound-induced vection facilitation effects than the large FOV (54°×45°) used in Experiment 1 (see Figure 5 and Table V). Interestingly, though, the smallest FOV condition (10°×7.5°) did *not* show a further increase in auditory facilitation of vection as would be predicted by hypothesis 4, but instead showed no significant facilitation of vection at all when adding spatialized sound (see Figure 5 and Table IV and V). This unexpected interaction between the FOV and auditory rendering quality reached significance for the convincingness of vection and showed a marginal trend for the vection intensity (see Table IV). A similar interaction was observed for the presence ratings: The medium FOV condition shows an enhancement of presence due to the added spatialized sound, whereas the small FOV condition showed no such effects at all (see Figure 6 and Table IV and V).

Clearly, auditory motion cues were insufficient to reliably enhance vection or presence when the visual field was reduced to 10°×7.5°. The analysis of the relative effect strength in Table V further corroborates this lack of any systematic influence of audition for the smallest FOV. The underlying reasons are not fully understood, but participants informal observations suggest a potential higher cognitive influence [Riecke et al. 2005]: During verbal probing during and after the experiment many participants mentioned a mismatch between the realistic auditory environment and the very small visual field. Participants often interpreted the condition as being inside a vehicle while looking out of a narrow window ("periscope" feeling), and observation that might be of interest with respect to the reduced presence and immersion elicited by such small FOVs (see Fig. 6).

3.2.4 *Influence of ambient sound (Hypothesis 5).* We hypothesized that adding a binaurally recorded ambient sound field might increase vection and/or presence, as ambient sound is present in almost all natural environments and might thus increase the overall believability and realism of the VR simulation (hypothesis 5). The data showed, however, no influence of adding ambient sound whatsoever, thus providing no support for hypothesis 5. One might interpret this as corroborating the finding from Experiment 1 that vection and presence were only facilitated by *moving* spatialized sound (hypothesis 2), but not by adding *non-moving* (mono) sound (hypothesis 1). In fact, the binaurally recorded ambient sound in Experiment 2 was clearly perceived as externalized (i.e., it was not perceived to originate inside the head as is the case for normal mono or stereo recordings displayed via headphones), but did not rotate with the visual stimulus. That is, similar to the mono condition in Experiment 1, merely adding spatialized but stationary ambient sound to a moving visual scene does not seem to have any effect on illusory self-motion perception.

Despite not directly affecting vection or presence, the ambiance condition did, in fact, clearly enhance participants' perception of auditory spaciousness – even though many of them commented on the mismatch between the lively auditory environment and static empty visual scene rendered using still image, which corroborates the importance of cross-modal consistency for VR simulations in general and self-motion perception in particular [Schulte-Pelkum and Riecke 2008a].

4. GENERAL DISCUSSION AND CONCLUSIONS

Visually induced self-motion illusions ("vection") are among the most compelling illusions and have been subject to extensive research for more than a century [Fischer and Kornmüller 1930; von Helmholtz 1896; Mach 1875]. Under certain conditions, auditory cues can induce similar self-motion illusions when being blindfolded [Dodge 1923; Lackner 1977; Larsson et al. 2004]. Such auditory vection is, however, far less compelling and is experienced by only about 20-60% of participants, depending on the presented stimuli [Lackner 1977; Larsson et al. 2004]. Thus, auditory cues *alone* seem to be insufficient for reliably inducing self-motion illusions, which limits their potential for self-motion applications.

The current study was designed to investigate whether auditory cues can nevertheless *enhance* the visually-induced self-motion illusion ("circular vection", Experiment 1), and how such a facilitatory effect might depend on the visual and auditory vection-inducing potential, which was manipulated by varying the visual FOV and the auditory rendering quality, respectively (Experiment 2).

While most visual vection studies up to now have used moving abstract geometric patterns (like black-and-white striped or dotted patterns) as vection-inducing stimuli, we used ecologically more plausible stimuli, namely the view of a town square projected using a high-end virtual reality system and the sound of a fountain that is also visible in the visual stimulus. Previous studies have demonstrated that using such naturalistic visual and auditory vection-inducing stimuli significantly enhances vection as well as presence [Larsson et al. 2004; Riecke et al. 2005; Riecke et al. 2006; Schulte-Pelkum and Riecke 2008a]. Note that using ecologically valid stimuli is also critical when considering the current results from an applied perspective of self-motion simulation and effective virtual environments. Furthermore, using naturalistic stimuli allowed us to complement the vection measures with measures of spatial presence and involvement in the simulated scene. Previous studies suggest that spatial presence in the virtual scene might correlate with vection and, in some situations, even be a mediator of vection [Riecke et al. 2006; Schulte-Pelkum and Riecke 2008a].

4.1 Neither mono sound nor stationary ambient sound enhance visually induced circular vection

Experiment 1 revealed that simply adding mono (i.e., non-spatial) sound to a visual motion simulation does not enhance circular vection or presence in the simulated scene. This was corroborated by Experiment 2, where the addition of ambient, not rotating sound fields improved neither vection nor presence, even though the ambient sound enhanced as expected the perceived spaciousness and overall believability and naturalism of the auditory simulation. Note, however, that the ambient sound was a stationary binaural recording and thus did not rotate concurrently with the visual stimulus, which might, of course, have countervailed a potential facilitatory effect of the enhanced spaciousness. This might not seem likely, as the ambient recording was not well localizable, but it is nevertheless a possibility we cannot exclude given the current data.

4.2 Adding spatialized sound that moves in sync with the visual cues enhances both vection and presence

While neither non-spatialized (mono) nor stationary ambient sound affected vection, both Experiment 1 and 2 showed a small but significant enhancement of the visually induced rotational self-motion illusion when the mono sound was replaced by a concurrently rotating spatialized sound. This seems to suggests a *specific* facilitation of vection due to the spatialized aspect of the auditory stimuli. Apart from a specific vection-inducing influence of the rotation sound sources, there are three more factors that might have unspecifically contributed to the vection-facilitating effect of concurrently rotating auditory landmarks:

First, the rotating sound field might have increased the perceived auditory realism, as the HRTF-rendered sound had a frequency spectrum that resembled more closely what one would hear in the real world situation. In particular, the spatialized sounds appeared properly externalized and spatialized, which has been shown to increase the naturalness of the sound as well as the overall believability of the simulation and the perceived "auditory presence" therein [Hendrix and Barfield 1996]. Recently, subjective evaluation methods of spatial audio have employed a new auditory spaciousness attribute termed "auditory presence", similar to attributes used in virtual reality presence research. Auditory presence has been defined as "the sense of being inside an (enclosed) space or scene" (see Rumsey [2002] and references therein). This auditory presence attribute of spatial sound quality is used for describing the recreated outdoor environments and generally refers to the background ambient sound energy arriving from various directions. In our case, rotating acoustic fields definitely magnified the auditory sensation of the market place.

Second, and closely related to the first issue, the concurrently rotating sound field improved the cross-modal consistency, which in turn might have increased the believability and/or perceived reliability of the audio-visual simulation and thus indirectly enhanced circular vection. In a sensor fusion model, this could be modeled by giving a stronger weighting to the audio-visual cues (which indicate self-motion) while leaving the weights of the other sensory cues (e.g., kinaesthetic and vestibular cues, which indicate the absence of any self-motion) unaltered.

Third, the meaning associated to rotating sound cues has also been shown to affect auditorily induced vection: Sounds that originate from objects that are expected to be stationary ("auditory landmarks" like church bells or a fountain sound) elicit stronger vection than objects that normally move (e.g., the sound of footsteps or a driving car) or are ambiguous (e.g., pink noise) [Larsson et al. 2004; Riecke et al. 2005]. These studies suggest that apart from providing physical motion cues, sound can also have a strong cognitive or higher-level influence on experienced self-motion and presence in virtual environments. Similar cognitive and higher-level influences have also been observed for visually induced vection (see discussion in Riecke et al. [2005] and Schulte-Pelkum and Riecke [2008a]).

4.3 Increasing auditory spatialization fidelity beyond 5 spatial locations did not enhance vection further

Even though adding spatialized sound enhanced both vection and presence, further increasing the auditory simulation fidelity (by increasing the resolution of the auditory rendering from 5 to 72 spatial locations (i.e., 5° resolution)) did not further affect any of the vection or presence measures. This finding was somewhat surprising, as sound spatializations with resolutions of more than 20° have been shown to produce audible artifacts and reduce localization accuracy [Langendijk and Bronkhorst 2000]. This might suggest a ceiling effect for the spatial fidelity of the auditory rendering: Once a certain quality level is achieved, spending more time and effort on further increasing the audio rendering quality might not be necessary for a given goal and does, in fact, not necessarily enhance the effectiveness of the overall simulation – even though there might be audible differences. This is in agreement with recent research showing that auditory perception and localization is surprisingly robust against degradation of both signal quality and spatialization [Best et al. 2005; Brungart et al. 2005; Moeck et al. 2007]. These findings pose a new interesting research question – what is the optimal auditory rendering quality for a given goal? In other words, what is the minimum spatial resolution and auditory fidelity that is still sufficient to achieve the desired effect (here: an auditory enhancement of visually induced self-motion illusions)?

There are a number of other factors that might have contributed to the observed ceiling effect (i.e., the lack of a clear benefit) when increasing auditory spatialization fidelity beyond five spatial locations: We used only one or two sound sources which were approximately at ear-level and rotated in the horizontal plane. Hence, accurate elevation simulation/perception (which is one of the challenges in HRTF rendering) was probably not essential for the current task. One might even argue that perfectly accurate sound localization might not have been that critical for the given task, as the sound field always moved in a simple and predictable manner around the observer. Furthermore, the visual cues moving in sync with the auditory cues might have resulted in a ventriloquist effect which further reduced the importance of perfectly accurate sound localization. That is, even when the fountain was outside of the current FOV, the visual scene provides a natural context which allows one to estimate where the fountain/van sound sources should be, at least when the visual FOV was sufficiently large.

Increasing the number of consistently moving sound sources has been shown to enhance auditorily induced vection [Larsson et al. 2004; Väljamäe et al. 2004]. Using individualized instead of generic HRTF catalogues for the binaural sound synthesis of the rotating sound fields, however, did not improve auditory vection – even though it successfully improved externalization, reduced the occurrence of perceptual artifacts (e.g. distorted trajectories of the rotating sound objects or in-head localization) and even increased auditory spatial presence [Väljamäe et al. 2004]. Taken together, these findings suggest that localization cues and spatial sound fidelity do play a role in enhancing the vection sensation. Note that a rather coarse spatialization fidelity of 5-6 spatial locations seems to be sufficient for many situations, though. This might turn out to be advantageous for many applications, given that such multi-channel sound systems are readily available and affordable.

4.4 The auditory contribution to vection is differentially effected by the visual FOV

Interestingly, the effect size of the auditory facilitation of vection was most pronounced for the medium FOV condition (20° × 15° FOV) and somewhat reduced for the large FOV condition (54° × 45°), whereas the small FOV condition (10° × 7.5°) showed virtually no auditory contribution. This implies an unexpected differential influence of the visual FOV on the auditory contribution to visually induced vection, which has to the best of our knowledge not been reported before.

Although a direct comparison of the absolute values for the different dependent measures between Experiment 1 and 2 is somewhat problematic due to potential differences in the participant population and the minor differences in the experimental procedures, the current results – and in particular the analysis of the *relative* effect sizes – suggests that the contribution of the auditory cues might be maximal not for the smallest visual FOV as initially hypothesized, but for the medium FOV condition. In terms of applications, this suggests that self-motion simulations with medium-sized FOVs might benefit the most from audio-visual interactions. Interestingly, these happen to be roughly in the range of the more affordable VR systems: Apart from a few rather costly head-mounted displays (HMDs), the most commonly used HMDs currently have FOVs between about 21° × 16° and 48°× 36°. The physical FOV of desktop-VR solutions strongly depends, of course, on the viewing distance, but typically has a similar range. Further studies that directly manipulate the visual FOV within one experiment and span a wider range of FOVs would be needed to provide more conclusive answers, though. It is, however, noteworthy that the visual condition which provided the weakest vection due to the smallest FOV did not show the strongest auditory contribution. This result was rather surprising, as one might argue (as we have, in fact, done above) that the vection-inducing potential of the visual and auditory stimuli were most similar in the small FOV condition, and most cue integration theories would predict larger cross-modal benefits the better the strength or weighting of the two cues are matched [Ernst and Bulthoff 2004].

The lack of any significant auditory contribution for the smallest FOV and the origin of the differential influence of auditory cues on vection depending on the visual FOV are not well understood, and we can only speculate about the underling reasons. When participants were debriefed after the experiments, some of them mentioned that the small FOV was just not big enough to really get the impression of one consistent natural scene. Furthermore, Figure 4 suggests that the fountain was no longer easily recognizable for the smallest FOV, even if one knows that it should be there. One might argue that these factors might have disrupted the audio-visual consistency, which in turn might have interfered with an auditory facilitation of vection.

Even though the FOV is receiving a lot of attention both in vection research and in terms of VR systems, there are, of course, many other factors that should not be neglected. Tan and colleagues demonstrated, for example, that physically larger displays can enhance performance in various spatial task, even if the physical FOV subtended by the different displays tested was identical [Tan et al. 2004; 2006]. Even if the physical dimensions and subtended FOV are similar, turn perception can differ systematically if a curved projection screen is used instead of a flat projection screen [Schulte-Pelkum et al. 2004]. Similarly, turn perception depends systematically on the FOV and display device and geometry used, with larger, curved projection screen leading in general to more accurate turn perception than head-mounted displays (HMDs) or projection screen that are flat and/or offer a smaller FOV [Riecke et al. 2005; Schulte-Pelkum et al. 2004].

In terms of visually induced self-motion illusion, the visual FOV has unarguably a strong vection-facilitating effect. If a smaller FOV is provided, the resulting self-motion illusion seems to depend not only on the absolute overall area of stimulation, but also on the spatial layout of the moving stimulus: If the vertical FOV is limited to 30°, circular vection in an optokinetic drum is still quite strong. When the horizontal FOV is limited to 30° instead, however, circular vection is considerably reduced [Brandt et al. 1973]. In the context of self-motion simulations, this argues for the usage of panoramic displays with a horizontal FOV that goes well beyond the vertical FOV. Using the current methodology to investigating these issues in more detail would allow us to further optimize self-motion simulations from a perceptual perspective.

4.5 Potential relations between vection and presence

The data from Experiment 1 and 2 showed a small but significant vection- and presence-facilitating effect of adding concurrently rotating spatialized sound to a vection-inducing visual stimulus. Furthermore, the convincingness of the self-motion was highly correlated to the perceived spatial presence in Experiment 1. Despite this correlation, it remains, though, an open question whether there might also be a *causal* relationship or mediation between presence and vection.

On the one hand, it is conceivable that an increase in presence can mediate an increase in vection, as suggested by Riecke et al. [2006]. On the other hand, it is also feasible that there might be a reciprocal relation between self-motion perception and presence, as suggested by Riecke et al. [2004] and discussed in more detail in Schulte-Pelkum and Riecke [2008b]. If compelling vection is perceived in VR, then this might in turn also increase presence. In almost any instance when we move in the natural world, we also perceive self-motion. Hence, if self-motions simulated in VR are unable to evoke a natural percept of self-motion, the overall believability of the VR simulation and presence in the virtual environment in particular might also be reduced. In the long run, gaining a deeper understanding of any potential causal relations between presence and the effectiveness of a simulation for a given task or goal (here: self-motion perception) would be quite helpful for optimizing VR simulations from a perceptual point of view.

4.6 Mutual cross-modal benefits

Even though adding non-rotating mono or ambient sound did not directly enhance vection or presence, Experiment 1 demonstrated unexpected secondary benefits of adding mono sound: While the mono sound was clearly not spatialized and would have been perceived as internalized and not moving if no additional visual cues were provided, participants in our experiments – where naturalistic visual stimuli were provided throughout the experiment – reported that the mono sound was surrounding them to the same extend as HRTF-based spatialized sound (cf. Fig. 3). Furthermore, more than half of the participants reported being able to localize the mono sound cue to some extent, and a quarter of them even reported some auditory motion. This could be interpreted as some sort of visual capture or "ventriloquism effect", and suggest a mutual cross-modal benefit: On the one hand, providing a compelling visual stimulus can enhance the auditory percept of a simple mono sound. On the other hand, adding spatialized auditory landmarks that move concurrently with the visual stimulus can enhance both vection and presence consistently. There does, however, seem to exist a window of optimal cue integration: While reducing the visual FOV from 54° × 40° to 20° × 15° increased the auditory contribution to vection and presence as expected, this auditory contribution almost vanished when the visual FOV was further reduced to 10° × 7.5°.

4.7 Why should sound be added to a visual simulation?

Apart from a specific vection-enhancing effect, adding spatialized auditory cues to VR simulations can have a number of further advantages, as is discussed in more detail in [Larsson et al. 2008; Väljamäe et al. 2008; Väljamäe et al. 2008]: Adding auditory cues is known to increase presence in the simulated world, especially if spatialized auditory cues are used that are perceived as properly externalized and can be well localized, for example by using generic HRTFs [Hendrix and Barfield 1996; Ozawa et al. 2004] or individualized HRTFs [Väljamäe et al. 2004]. This is in agreement with the observed presence-facilitating effect of spatialized auditory cues in the current study.

Furthermore, auditory cues provide the advantage of extending the perceivable virtual space beyond the limits of the visual FOV of the setup. This makes auditory cues perfectly suited for warning signals or for guiding attention. The omni-directional characteristics of human hearing enables us to get also a decent impression of the size and layout of a (real or simulated) scene without the need to turn our head and face the direction of objects of interest [Pope and Chalmers 1999].

Apart from providing spatial information (a domain where visual cues seem to dominate), auditory cues are also excellent indicators of the temporal characteristics of sounding events, and the temporal resolution in humans is, in fact, much higher for auditory than for visual cues. By taking into account temporal and spectral properties, auditory cues can, for example, inform us about the type, material, and mechanical properties of sounding objects themselves (when listening to the sound of footsteps, for example, the auditory cues can be quite informative about the material and elastic properties of both the shoes and the floor) and the dynamic properties of the interaction between objects (e.g., pace of the stride, velocity and momentum of a moving object when hitting a target) [Luciani 2004]. Auditory cues can thus be used to convey the spatio-temporal structure of motions and may compensate for visual imperfections via cross-modal interaction mechanisms [Väljamäe et al. 2006]

In general, whenever the corresponding situation in the real world would be accompanied with specific sounds, one would probably expect to hear those sounds in VR, too. This is of particular importance for achieving high perceptual and behavioral realism in specific applications like driving and flight simulations, where adding appropriate engine sounds or environmental sounds is of crucial importance. One of the most frequent usages of audition is probably due to its clear potential to elicit emotional responses, a fact that is well-known and frequently employed by, for example, the movie industry. Last but not least, including auditory cues can also be particularly important for people who's preferred modality or cognitive style is auditory (as opposed to visual or kinaesthetic).

4.8 Conclusions and outlook

Adding spatialized auditory cues to (predominately visual) VR simulations and ego-motion simulations in particular can have a number of advantages including an increase in the perceived self-motion. Relatively little research has been performed in this area, and additional studies are required to investigate these issues further. It is conceivable, however, that the requirements for visual rendering quality could be relaxed when appropriate simulation of the auditory modality (and potential other modalities) is provided [Durlach and Mavor 1995]. As high quality auditory rendering can be achieved at relatively low cost, adding spatialized auditory cues might allow us in the future to increase simulation effectiveness while reducing the overall simulation effort, especially when the attention guiding potential of auditory cues is employed. Using a selective rendering approach, guiding attention has, for example, been shown to reduce computational costs of the visual rendering considerably [Cater et al. 2003; Sundstedt et al. 2004]. This is promising for the usage of auditory cues for optimizing VR simulations both on a computational and perceptual level. If this were the case, it would have important implications for both our understanding of multi-modal self-motion perception and optimizing virtual reality applications that include simulated movements of the observer. As the current study demonstrated, adding HRTF-based auralization using headphones and spatial sound of relatively low resolution (comparable to 5-channel surround sound systems) can reliably be used to improve self-motion perception as well as presence in VR, even when the visual rendering is already of high quality and realism. This has many practical advantages, especially for applications where speaker arrays are unsuitable or where external noise must be excluded.

In conclusion, performing basic psychophysical experiment does not only foster our understanding of the underlying mechanisms and contributions, but can also be used as a means to guide further research and development on optimizing self-motion simulations from a perceptual perspective: The better we understand how the different sensory cues are perceived and integrated in the human brain to result in a desired outcome (here, the compelling illusion of self-motion and/or presence), the more we can use this knowledge to plan ahead and allocate our resources intelligently in order to maximize the overall effectiveness of a simulation while minimizing the simulation and research effort. The resulting simulation software and hardware will thus be more effective with respect to the desired goal and can then serve as an improved research paradigm and setup that enables us to extend the range of questions that can be experimentally addressed beyond the previous limitations.

ACKNOWLEDGMENTS

This research was funded by the EC grant POEMS-IST-2001-39223 (see www.poems-project.info), the Swedish Science Council (VR), and the Max Planck Society. We would like to thank Jan Oliver Hirn for his assistance in running Experiment 1, Ana Tajadura-Jimenez for programming the stimuli and assistance in running Experiment 2, Franck Caniard for programming the stimuli, for helpful discussions and technical assistance, and Pontus Larsson and Daniel Västfjäll for useful comments on the design of Experiment 2.

REFERENCES

ALGAZI, V. R., DUDA, R. O., THOMPSON, D. M., AND AVENDANO, C. 2001. The CIPIC HRTF database. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 99–102.

BEST, V., VAN SCHAIK, A., JIN, C., AND CARLILE, S. 2005. Auditory spatial perception with sources overlapping in frequency and time. *Acta Acustica 91*, 421–428.

BOER, E. R., GIRSHIK, A. R., YAMAMURA, T., AND KUGE, N. 2000. Experiencing the same road twice: A driver-centred comparison between simulation and reality. In *Proceedings of the Driving Simulation Conference 2000*. Paris, France.

BRANDT, T., DICHGANS, J., AND HELD, R. 1973. Optokinesis affects body posture and subjective visual vertical. *Pflugers Archiv-European Journal of Physiology 339*, 97–97.

BRANDT, T., DICHGANS, J., AND KOENIG, E. 1973. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research 16*, 476–491.

BRUNGART, D., SIMPSON, B., AND KORDIK, A. 2005. Localization in the presence of multiple simultaneous sounds. *Acta Acustica united with Acustica 91,* 3, 471–479.

BURKI-COHEN, J., GO, T. H., CHUNG, W. Y., SCHROEDER, J., JACOBS, S., AND LONGRIDGE, T. 2003. Simulator fidelity requirements for airline pilot training and evaluation continued: An update on motion requirements research. In *Proceedings of the 12th International Symposium on Aviation Psychology, April 14-17.* Dayton (OH), USA, 182–189.

CATER, K., CHALMERS, A., AND WARD, G. 2003. Detail to attention: Exploiting visual tasks for selective rendering. In *EGRW '03: Proceedings of the 14th Eurographics Workshop on Rendering*. 270–280.

CHANCE, S. S., GAUNET, F., BEALL, A. C., AND LOOMIS, J. M. 1998. Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence - Teleoperators and Virtual Environments 7,* 2 (Apr.), 168–178.

DICHGANS, J. AND BRANDT, T. 1978. Visual-vestibular interaction: Effects on self-motion perception and postural control. In *Perception*, R. Held, H. W. Leibowitz, and H.-L. Teuber, Eds. Handbook of Sensory Physiology, vol. VIII. Springer, 756–804.

DODGE, R. 1923. Thresholds of rotation. *J. Exp. Psychol. 6*, 107–137.

DURGIN, F. H., PELAH, A., FOX, L. F., LEWIS, J. Y., KANE, R., AND WALLEY, K. A. 2005. Self-motion perception during locomotor recalibration: More than meets the eye. *Journal of Experimental Psychology: Human Perception and Performance 31,* 3, 398–419.

DURLACH, N. I. AND MAVOR, A. S., Eds. 1995. *Virtual reality: Scientific and technological challenges.* National Academy Press.

ERNST, M. O. AND BULTHOFF, H. H. 2004. Merging the senses into a robust percept. *Trends in Cognitive Sciences 8,* 4 (Apr.), 162–169.

FISCHER, M. H. AND KORNMÜLLER, A. E. 1930. Optokinetisch ausgelöste Bewegungswahrnehmung und optokinetischer Nystagmus [Optokinetically induced motion perception and optokinetic nystagmus]. *Journal für Psychologie und Neurologie*, 273–308.

FLANAGAN, M. B., MAY, J. G., AND DOBIE, T. G. 2004. The role of vection, eye movements and postural instability in the etiology of motion sickness. *Journal of Vestibular Research-Equilibrium & Orientation 14,* 4, 335–346.

GARDNER, W. AND MARTIN, K. 1995. Hrtf measurements of a KEMAR. *J. Acoust. Soc. Am. 97*, 3907–3908.

GEKHMAN, B. 1991. Audiokinetic nystagmus. *Sensornye Sistemy 5,* 2, 71–78. (in Russion).

HENDRIX, C. AND BARFIELD, W. 1996. The sense of presence within auditory virtual environments. *Presence - Teleoperators and Virtual Environments 5,* 3, 290–301.

HENNEBERT, P. E. 1960. Audiokinetic nystagmus. *Journal of Auditory Research 1,* 1, 84–87.

HETTINGER, L. J. 2002. Illusory self-motion in virtual environments. In *Handbook of Virtual Environments*, K. M. Stanney, Ed. Lawrence Erlbaum, Chapter 23, 471–492.

HOLLERBACH, J. M. 2002. Locomotion interfaces. In *Handbook of Virtual Environments*, K. M. Stanney, Ed. Lawrence Erlbaum, Chapter 11, 239–254.

HOWARD, I. P. 1982. *Human visual orientation*. J. Wiley, Chichester, New York, Engl.

HOWARD, I. P. 1986. The perception of posture, self motion, and the visual vertical. In *Sensory processes and perception*, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds. Handbook of human perception and performance, vol. 1. Wiley, New York, 18.1–18.62.

KAPRALOS, B., ZIKOVITZ, D., JENKIN, M., AND HARRIS, L. 2004. Auditory cues in the perception of self-motion. In *Proceedings of the 116th AES convention, Berlin 2005*. Berlin, Germany.

KLEINER, M. AND DALENBÄCK, B-I. ANDSVENSSON, P. 1993. Auralization – an overview. *J. Audio Engineering Soc. 41,* 11, 861–875.

LACKNER, J. R. 1977. Induction of illusory self-rotation and nystagmus by a rotating sound-field. *Aviation Space and Environmental Medicine 48,* 2, 129–131.

LANGENDIJK, E. H. A. AND BRONKHORST, A. W. 2000. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *Acoustical Society of America Journal 107*, 528–537.

LARSSON, P., VÄLJAMÄE, A., VÄSTFJÄLL, D., AND KLEINER, M. 2008. Auditory-induced presence in mediated environments and related technology. In *Immersed in Media Experiences: Presence Psychology and Design (Handbook of Presence)*. Lawrence Erlbaum. in print.

LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2004. Perception of self-motion and presence in auditory virtual environments. In *Proceedings of 7th Annual Workshop of Presence*. 252–258. Available: www.kyb.mpg.de/publication.html?publ=2953.

LOMBARD, M. AND DITTON, T. 1999. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication 3,* 2.

LUCIANI, A. 2004. Dynamics as a common criterion to enhance the sense of presence in virtual environments. In *Proceedings of 7th Annual Workshop Presence 2004*. 96–103.

MACH, E. 1875. *Grundlinien der Lehre von der Bewegungsempfindung*. Engelmann, Leipzig, Germany.

MARMEKARELSE, A. M. AND BLES, W. 1977. Circular vection and human posture ii: Does the auditory-system play a role. *Agressologie 18,* 6, 329–333.

MCFARLAND, W. AND WEBER, B. 1969. An investigation of ocular response to various forms of sound field auditory stimulation. *Journal of Auditory Research 9,* 3, 236–239.

MOECK, T., BONNEEL, N., TSINGOS, N., DRETTAKIS, G., VIAUD-DELMON, I., AND ALLOZA, D. 2007. Progressive perceptual audio rendering of complex scenes. In *I3D '07: Proceedings of the 2007 symposium on Interactive 3D graphics and games*. ACM, New York, NY, USA, 189–196.

MULDER, M., VAN PAASSEN, M. M., AND BOER, E. R. 2004. Exploring the roles of information in the control of vehicular locomotion - from kinematics and dynamics to cybernetics. *Presence - Teleoperators and Virtual Environments 13*, 535–548.

OZAWA, K., CHUJO, Y., SUZUKI, Y., AND SONE, T. 2004. Psychological factors involved in auditory presence. *Acoustical Science and Technology 24*, 42–44.

POPE, J. AND CHALMERS, A. 1999. Multi-sensory rendering: Combining graphics and acoustics. In *Proceedings of the 7th International Conference in Central Europe on Computer Graphics*. Czech Republic, 233–242.

RIECKE, B. E., SCHULTE-PELKUM, J., AVRAAMIDES, M. N., AND BÜLTHOFF, H. H. 2004. Enhancing the visually induced self-motion illusion (vection) under natural viewing conditions in virtual reality. In *Proceedings of 7th Annual Workshop Presence 2004*. 125–132. Available: www.kyb.mpg.de/publication.html?publ=2864.

RIECKE, B. E., SCHULTE-PELKUM, J., AVRAAMIDES, M. N., VON DER HEYDE, M., AND BÜLTHOFF, H. H. 2006. Cognitive factors can influence self-motion perception (vection) in virtual reality. *ACM Transactions on Applied Perception 3,* 3, 194–216.

RIECKE, B. E., SCHULTE-PELKUM, J., AND BÜLTHOFF, H. H. 2005. Perceiving simulated ego-motions in virtual reality - comparing large screen displays with HMDs. In *Proceedings of the SPIE*. Vol. 5666. San Jose, CA, USA, 344–355.

RIECKE, B. E., SCHULTE-PELKUM, J., CANIARD, F., AND BÜLTHOFF, H. H. 2005a. Influence of Auditory Cues on the visually-induced Self-Motion Illusion (Circular Vection) in Virtual Reality. In *Proceedings of 8th Annual Workshop Presence 2005*. 49–57.

RIECKE, B. E., SCHULTE-PELKUM, J., CANIARD, F., AND BÜLTHOFF, H. H. 2005b. Towards lean and elegant self-motion simulation in virtual reality. In *Proceedings of IEEE Virtual Reality 2005*. Bonn, Germany, 131–138.

RIECKE, B. E., VÄSTFJÄLL, D., LARSSON, P., AND SCHULTE-PELKUM, J. 2005. Top-down and multi-modal influences on self-motion perception in virtual reality. In *Proceedings of HCI international 2005*. Las Vegas, NV, USA, 1–10.

RIECKE, B. E. AND WIENER, J. M. 2007. Can people not tell left from right in vr? point-to-origin studies revealed qualitative errors in visual path integration. In *Proceedings of IEEE Virtual Reality 2007*. 3–10.

RUMSEY, F. 2002. Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm. *J. Audio Eng. Soc 50,* 9, 651.

SAKAMOTO, S., OSADA, Y., SUZUKI, Y., AND GYOBA, J. 2004. The effects of linearly moving sound images on selfmotion perception. *Acoustical Science and Technology 25*, 100–102.

SCHINAUER, T., HELLMANN, A., AND HÖGER, R. 1993. Dynamic acoustical stimulation affects self-motion perception. In *Contributions to Psychological Acoustics. Results of the 6th Oldenburg Symposium on Psychological Acoustics*, A. Schick, Ed. Contributions to Psychological Acoustics. Oldenburg, 373–385.

SCHUBERT, T., FRIEDMANN, F., AND REGENBRECHT, H. 2001. The experience of presence: Factor analytic insights. *Presence - Teleoperators and Virtual Environments 10,* 3, 266–281.

SCHULTE-PELKUM, J. AND RIECKE, B. E. 2008a. An integrative approach to presence and self-motion perception research. In *Immersed in Media Experiences: Presence Psychology and Design (Handbook of Presence).* Lawrence Erlbaum. in print.

SCHULTE-PELKUM, J. AND RIECKE, B. E. 2008b. An integrative approach to presence and self-motion perception research. In *Immersed in Media Experiences: Presence Psychology and Design (Handbook of Presence).* Lawrence Erlbaum. in print.

SCHULTE-PELKUM, J., RIECKE, B. E., AND BÜLTHOFF, H. H. 2004. Vibrational cues enhance believability of ego-motion simulation. In *International Multisensory Research Forum (IMRF).* Available: www.kyb.mpg.de/publication.html?publ=2766.

SCHULTE-PELKUM, J., RIECKE, B. E., VON DER HEYDE, M., AND BÜLTHOFF, H. H. 2004. Influence of display device and screen curvature on perceiving and controlling simulated ego-rotations from optic flow. Tech. Rep. 122, MPI for Biological Cybernetics. Avaliable: www.kyb.mpg.de/publication.html?publ=2574.

SOTO-FARACO, S., KINGSTONE, A., AND SPENCE, C. 2003. Multisensory contributions to the perception of motion. *Neuropsychologia 41,* 13, 1847–1862.

SUNDSTEDT, V., DEBATTISTA, K., AND CHALMERS, A. 2004. Selective rendering using task-importance maps. In *ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV).* 175.

TAN, D. S., GERGLE, D., SCUPELLI, P., AND PAUSCH, R. 2006. Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction 13,* 1 (March), 71–99.

TAN, D. S., GERGLE, D., SCUPELLI, P. G., AND PAUSCH, R. 2004. Physically large displays improve path integration in 3D virtual navigation tasks. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 439–446.

VÄLJAMÄE, A., KOHLRAUSCH, A., VAN DE PAR, S., VÄSTFJÄLL, D., LARSSON, P., AND KLEINER, M. 2008. Audio-visual interaction and synergy effects: implications for cross-modal optimization of virtual and mixed reality applications. In *Immersed in Media Experiences: Presence Psychology and Design (Handbook of Presence).* Lawrence Erlbaum. in print.

VÄLJAMÄE, A., LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2004. Auditory presence, individualized head- related transfer functions, and illusory ego-motion in virtual environments. In *Proceedings of 7th Annual Workshop of Presence*. Valencia, Spain, 141–147. Available: www.kyb.mpg.de/publication.html?publ=2954.

VÄLJAMÄE, A., LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2005. Travelling without moving: Auditory scene cues for translational self-motion. In *Proceedings of ICAD 2005*. Limerick, Ireland.

VÄLJAMÄE, A., LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2006. Vibrotactile enhancement of auditory induced self-motion and spatial presence. *Journal of the Acoustic Engineering Society 54,* 10, 954–963.

VÄLJAMÄE, A., LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2008a. Auditory landmarks enhance multimodal circular vection simulations in virtual reality. *Journal of the Acoustic Engineering Society*. submitted.

VÄLJAMÄE, A., LARSSON, P., VÄSTFJÄLL, D., AND KLEINER, M. 2008b. Sound representing self-motion in virtual environments enhances linear vection. *Presence: Teleoperators and Virtual Environments 17,* 3. in press.

VÄLJAMÄE, A., VÄSTFJÄLL, D., LARSSON, P., AND KLEINER, M. 2008. Perceived sound in mediated environments. In *Immersed in Media Experiences: Presence Psychology and Design (Handbook of Presence).* Lawrence Erlbaum. in print.

VAN DER STEEN, F. A. M. AND BROCKHOFF, P. T. M. 2000. Induction and impairment of saturated yaw and surge vection. *Perception & Psychophysics 62,* 1 (Jan.), 89–99.

VON HELMHOLTZ, H. 1896. *Handbuch der physiologischen Optik.* L. Voss, Hamburg-Leipzig.

WARREN, R. AND WERTHEIM, A. H., Eds. 1990. *Perception & Control of Self-Motion*. Erlbaum, New Jersey, London.

WATSON, J. E. 1968. Evaluation of audiokinetic nystagmus as as test of auditory sensitivity. *The Journal of Auditory Research 8*, 161–165.

Received Month Year; revised Month Year; accepted Month Year