Binaural bone-conducted sound in virtual environments: Evaluation of a portable, multimodal motion simulator prototype

Aleksander Väljamäe\(^1\), Ana Tajadura-Jiménez\(^1,\dagger\), Pontus Larsson\(^1,\ddagger\), Daniel Västfjäll\(^1,\ddagger,\S\) and Mendel Kleiner\(^1,\S\)

\(^1\)Chalmers Room Acoustics Group — Multisensory Applications, Division of Applied Acoustics, Chalmers University of Technology, Sven Hultins gata 8a 41296, Gothenburg, Sweden
\(^2\)Department of Psychology, Göteborg University, Gothenburg, Sweden

(Received 16 December 2006, Accepted for publication 24 August 2007)

Abstract: Virtual and augmented reality applications provide us with increasingly compelling multisensory worlds. Although spatial sound technologies are often used in such applications, headphone based sound reproduction may result in an undesired “mediation awareness” for an end-user. An alternative can be provided by bone-conducted sound technologies, traditionally used in hearing aids applications. Recent studies with bilaterally fitted bone-conduction transducers suggest that binaural sound cues can be rendered using this technology. In this paper we used binaural bone-conducted sound reproduction for enhancing a multi-modal self-motion simulator prototype. Similar to previous results from headphone based reproduction, the present study shows that the addition of moving sound images to visual stimuli significantly increase vection and spatial presence responses. These results provide empirical evidence that dynamic auditory scenes can be created using spatial bone-conducted sound with at least 45° horizontal resolution. The present research demonstrates the feasibility of using binaural bone-conducted sound in mediated environments.

Keywords: Spatial bone-conduction, Augmented reality, Multisensory interaction, Vection

PACS number: 43.66.+y [doi:10.1250/ast.29.149]

1. INTRODUCTION

Traditionally, bone-conducted (BC) sound actuators have been used in hearing aid applications. However, recently this technology has also gained interest from the perspective of novel navigation aids (e.g. [1]) and entertainment (e.g. an underwater sound player [2], or an electric toothbrush conducting sound via teeth and the jaw bone [3]). Bone-conducted sound technology may also have potential for auditory augmented reality applications. A common approach in these applications is to mix virtual sound objects with recorded surrounding acoustic environment using signal processing methods [4]. Alternatively, unaltered air conducted (AC) sounds can be naturally combined with sounds delivered via participants’ skull in BC based auditory mixed reality. In addition, BC sound rendering technology avoids “mediation awareness” which may be produced by ears occlusion in headphone reproduction. For example [5], showed that earplugs contribute to the sensation of unconnectedness to the surround (or, in authors terms, less “environmentally anchored presence”). Bone-conducted sound can be elicited by human head vibrations, which are transmitted to the cochlea through the skull bones [6] (for a recent review on BC sound see [7]). BC sound is experienced by most people in everyday life; approximately 50% of the sound energy when hearing one’s own voice is transmitted through bone conduction [8]. Recent evidence from patients with bilaterally fitted Bone-Anchored Hearing Aids (BAHA) suggests that binaural auditory processing can occur when using bone-conducted sound [9,10]. The BC transducer placement directly affects transcranial attenuation between the ears, which for the mastoid stimulation can vary in a range of 0–15 dB (see [7] and references therein). Although this is less than the 15–20 dB attenuation which is typically found for the AC acoustic path, it results in sound lateralization to the side with a higher level of stimulation [6].
tion can be also caused by time delay in between both cochleae. However, for low frequencies the human skull behaves like a rigid body, and the propagation velocity approaches infinity. For example, differences of 0.2 ms are obtained for sound below 0.8 kHz, which is significantly less than the 0.65 ms interaural time difference in the AC case [7]. Thus, feasibility for rendering spatial audio cues via BC remains largely an open research question.

Authors are aware of only one study which directly tested the feasibility of binaural BC sound rendering [11]. This study addressed the localization performance in the horizontal plane for virtual sound sources synthesized using individualized Head-Related Transfer Functions (HRTFs). No significant difference in localization errors between the BC and AC based rendering was found for targets separated by 45 degrees resolution (positions at 0°, ±45°, ±90°, ±135°, and 180°). This study provide the first direct evidence of a successful application of HRTFs for a two-channel BC headset, with no specific compensation for possible transcranial cross-channel interference.

The main aim of this paper was to test the applicability of binaural bone-conduction sound to self-motion simulator design where dynamic auditory information is required. Virtual reality and other media applications often aim at creating a compelling sensation of self-motion without physically moving the users. During such illusory self-motion sensation, also termed as vection, moving stimuli are perceived by a user as being stationary relative to her illusory locomotion. For example, vection sensation may arise in real life when seeing a departing train on a neighboring track while expecting your railroad car to move. Although a large body of research has studied visually induced vection (e.g. see reviews by [12,13]), it is only recently that the influence of auditory cues and their contribution to the self-motion experience has gained more attention [14–16].

Similarly to visual vection, auditory-induced vection (AVI) can be elicited using real or virtual sound fields moving relative to the listener’s point of audition (see Fig. 1). In a number of experiments we used binaural synthesis and headphone reproduction for inducing circular and linear AVI [16,17]. The sensation of vection in such acoustic simulations can be easily destroyed by other sensory modality providing information about a stable external environment. Therefore, one usually needs to apply special procedures in order to induce vection by auditory cues (e.g. blindfolding participants). This makes pure auditory-based self-motion systems useful only for very specific applications (e.g. audio only games [18] or radiodrama). However, auditory cues were shown to interact with vibrotactile [19] and visual information [20] where they significantly enhanced users' vection responses in multimodal motion simulators.

Enhancement of visually-induced vection by other sensory modalities is especially important when the strength of visual self-motion cues is reduced. For example, commercially available applications typically employ smaller visual displays (e.g. arcade-type or hand-held video games) than in high-end self-motion simulators. Recent audio-visual studies show that adding rotating acoustic fields to visual stimuli significantly enhance circular vection responses [20], especially, when the visual display size is reduced [21].

In the present experiment we studied whether audio-visual interactions can also occur in a portable motion simulator prototype [22], where a two-channel bone conduction headset was used to deliver spatial sound. In this way, vection-inducing auditory cues could be mixed with user’s external sound environment. In this study we tested several hypotheses:

**Hypothesis 1** — If binaural sound can be rendered via BC sound headset we expected to replicate previous findings reported in [20] and [21] showing auditory enhancement of rotating visual scene. However, such enhancement could be altered due to transcranial cross-channel interference, novelty of BC sound experience or presence of external sound environment.

**Hypothesis 2** — Our previous study showed that vection sensation can be significantly reduced when listeners’ attention is focused on a sound outside a moving auditory scene [17]. These results may be in line with other studies showing that attention can modulate the processing of visual or vestibular motion cues [23,24]. In a mixed reality system, based on bone-conducted sound, external sound cues can interfere with auditory motion cues. Therefore, in the current experiment we additionally studied whether directing participants’ attention to a sound outside the moving acoustic field may alter the expected auditory enhancement of vection and spatial presence.
Hypothesis 3 — Our surrounding environments contain not only a few distinct, clearly audible sound sources, but also some ambient sounds produced by various distant, unlocalizable sound objects enveloping listeners. The present study also investigated whether such ambient, “spaciousness”-rich sound might increase realism of the presented environment and thus influence vection and spatial presence ratings.

2. METHOD

2.1. Apparatus and Stimuli

The experiment took place in a special laboratory setup (see Fig. 2) surrounded by black curtains, which hid the external environment. Visual stimuli were displayed on a flat 17-inch LCD monitor at a distance of 0.5 m, which corresponded to a \(37° \times 30°\) field of view. Spatial sound was rendered by a Lake Huron DSP system and a generic HRTFs catalogue at two spatial quality levels. For the medium-fidelity, “BinScape” setting was used which employs 5 spatial locations corresponding to \(0°, 45°, 90°, 135°, 180°\) \([25]\). This medium-fidelity condition will be referred further in the text as having maximum 45° spatial resolution. For the high-fidelity, “HeadScape” setting was used at a uniform resolution of 5°. Virtual reality programming library VElib \([26]\) was used for stimuli playback and joystick data collection.

Auditory virtual space was reproduced using bilaterally placed bone-conduction transducers based on BEST (Balanced Electromagnetic Separation Transducer) technology \([27]\) manufactured by Ouido Equipment \([28]\). The headset is specially designed for commercial use in a variety of communication scenarios and it is comfort-optimized. The BC headset places transducers bilaterally on the temporomandibular condyle area (point joining temporal and jaw bones) as shown in Fig. 3. Several studies indicate little difference between head sensitivity at the mastoid and the condyle areas for BC sound \([29,30]\) with the condyle being more suitable for BC sound headset design. The transducers had the effective frequency range of 250–6,000 Hz as measured on the artificial mastoid using transducer vibration levels (personal communication, \([28]\); see also \([27]\) for measurement procedure details).

Before running the experiment, a pilot study using 3 blindfolded participants with blocked ears confirmed that the two-channel BC interface can render the rotating virtual sound fields. In the main experiment no ear blocking was used.

2.1.1. Within-subjects stimuli

In this experiment naturalistic, ecologically valid stimuli was used which reliably induced circular vection in the previous studies \([16,20]\). The stimuli rotation pattern was the following: 3 s-stationary phase, 3 s-acceleration, 46 s at constant rotation speed at \(30°/s\), and 3 s-deceleration. The rotating visual stimuli represented the Tübingen market square and was rendered using a round-shot photograph \((4,096 \times 1,024\) pixel) mapped on a virtual cylinder (see Fig. 2).

The naturalistic soundscape contained several components (see Fig. 1). First, the rotating acoustic fields contained two ecological sounds (“bus on idle” and “fountain”), which both corresponded to objects in the visual scene and could be classified by listeners as acoustic landmarks. Larsson et al. \([16]\) has recently shown that such sound sources are more instrumental in creating circular AIV compared to acoustic fields containing artificial sounds (noises and tones) or naturalistic sounds representing moving objects. Second, a non-rotating engine sound served as a sonic metaphor of the user’s virtual vehicle. Our previous study on auditory-induced linear...
vection suggests that sound representing one’s motion dynamics (e.g. footsteps, engine sound) may significantly enhance the self-motion sensation [31]. Third, in half of the trials, an additional ambient sound, recorded binaurally on a market square, was added to the Lake-rendered spatial sound (this ambient sound contained typical small town square sounds which are difficult to localize, e.g. distant speech, sparrows’ tweets, etc.). No acoustic environment rendering (early reflections, reverberation etc.) was applied.

In half of the trials an auditory focused attention task was introduced. A “church bell” sound was used for this focused attention task and it appeared either at 34 s or 43 s from stimuli playback start. In this task participants had to concentrate on the sounds and click the joystick button when hearing the church bell sound. Using two timings for the church sound appearance aimed to reduce the possible effect of expectation on the task performance.

The experimental within-subjects design thus contained the following factors: 2 turning directions (left/right) × 3 types of rotating spatial sound fidelity (monophonic (diotic), BinScape and HeadScape sound presentation) × 2 additional ambience (on/off) × 2 focused attention (task/no task).

2.1.2. Between-subjects stimuli

The experiment contained one spatial between-subjects factor which split the experimental sample into groups of 8 participants each. This factor represented the immersion level and was used for the simulator prototype evaluation study not covered in this paper. The first group of participants were exposed only to audio-visual rendering inside the simulator prototype. For the second group, additional low-level vibrations (below 3 cm/s²) under the seat and the footrest were applied (see [19] for research on audio-vibrotactile interactions in self-motion simulations). For the third group of participants, an additional cockpit occluding participants’ view from the external visual frame of reference (see Fig. 4) was introduced apart from the vibrations.

2.2. Measures

The vection illusion strength can be reliably accessed using a simple verbal measure where participants rate (on a 0–100 scale) the subjective sensation of experiencing self-motion [13]. After each trial participants were asked to rate self-motion intensity and also their sensation of presence (on a 0–100 scale) created by the presented scene. Spatial presence was defined as “a sensation of being actually present in the virtual world.” Accessing users’ presence sensation is believed to be crucial for human-centred evaluation of multi-modal virtual environments [32].

2.3. Procedure

The experiment was completed by 24 participants (mean age 24.4; SD = 3.5); 13 male took part. After receiving written and verbal instructions, participants were seated and the two-channel BC headset was positioned. A short sound-test was performed to check the good positioning of the BC headset-participants had to hear out two speakers in diotically presented speech mixture. A short training session was performed before the experiment started (2 stimuli presented).

Each trial was started by participants (by clicking the button of a joystick). Participants then had to proceed in two different ways depending on the attention condition. Participants either had to relax and wait until the end of the full stimulus playback or perform the focused attention task. Clicking the joystick button in these trials made the stimuli playback stop after 3 seconds of deceleration. After each trial, participants had to rate the intensity of self-motion and presence. Apart from the verbal responses to the questionnaire, verbal probing was done by the experiment leader. After completing the experiment, participants were debriefed, thanked and paid for their participation.

Due to the need to restart the Lake System each time a new sound-rendering engine was applied, stimuli were presented in 3 blocks containing different sound rendering qualities (other factors were randomly presented within each block). The order of these 3 blocks was randomly changed between subjects.

3. RESULTS AND DISCUSSION

None of the 24 participants reported any discomfort of wearing the BC headset after the 50-minute experiment. All experimental results were averaged between left and right rotation directions. The results were submitted to
3 (spatial sound) × 2 (ambience) × 2 (attention task)
ANOVA.s for self-motion intensity and spatial presence
ratings. Greenhouse-Geisser correction was applied when-
never unequal variances occurred.

Before the analyses of the within-subjects effects,
separate ANOVAs were conducted with the spatial three
factor (immersion level) as a between subject variable. No
significant interaction between within-subjects variables
and between-subjects immersion level factor was found for
any of the used measures. Therefore, the experimental data
were further treated as having only the within-subjects
factorial design.

As predicted in Hypothesis 1, the main effect of spatial
sound fidelity reached significance both for vection
intensity ($F(2, 46) = 8.66, p < 0.001$) and spatial presence
($F(2, 44) = 6.75, p < 0.005$) ratings. The plotted means in
Fig. 5 (left panel) and Bonferroni pairwise comparisons
showed that these effects were due to the significant
difference between BinScape (max 45° resolution) and
mono conditions. It can be seen that this effect falls off
when the higher, 5° spatial sound resolution, was used
(HeadScape). This might be due to the fact that the spatial
resolution of spatial bone-conduction can be lower than the
one found for air-conducted sound perception. However,
the effects presented in Fig. 5 (left panel) are consistent
with our previous findings on audio-visually induced
circular vection where the rotating acoustic fields were
rendered using headphones [21].

In the study reported in [21] the same methodology and
verbal measures as the ones described above were used. In
the experimental design, the same three levels of spatial
sound fidelity were used in combination with two different
visual field sizes ($20° \times 15°$ and $10° \times 7.5°$) shown on a
projection screen. For comparison purposes, the right panel
in Fig. 5 shows vection intensity and spatial presence
ratings obtained for the $20° \times 15°$ visual field of view.
Similar to the present study, a significant enhancement
effect of BinScape spatial sound condition was found. One
can also note that the vection intensity ratings are smaller
than in the current study, which is most likely due to the
different size of visual field of view used.

The focused attention task (Hypothesis 2) did not
significantly alter neither illusory self-motion ($F(1, 23) =
0.3, p = 0.6$) nor spatial presence ($F(1, 23) = 0.02, p =
0.8$) ratings. It also did not interact with ambience and
spatial sound factors. Dedicated studies should address the
question to which extend auditory enhancement of visual
vection is dependent on attentional resources. However, the
current results provide an initial support for using auditory
augmented reality in applications with self-motion induc-
ing scenarios (e.g. hand-held games).

Opposite to hypothesis 3, the addition of ambient sound
to rotating acoustic field did not affect presence and vection
ratings. It seems that adding spaciousness to the auditory
scene do not play a significant role in the auditory
enhancement of vection. Similarly, in [20] spatially sta-
tionary sound did not differ significantly from the visual
only condition suggesting the importance of auditory
motion cues rather that mere sound realism (cf. engine
metaphor in [31]).

How important is the congruence of auditory and visual
motion cues? Unfortunately, the current study and the
headphone-based studies described earlier ([20] and [21])
did not examine a conflicting situation where the rotating
visual scene would be coupled with auditory motion in the
opposite direction. In another study on auditory enhance-
ment of passive self-rotation in darkness [33], such
conflicting situations were created. Only “compatible”
acoustic fields rotation significantly enhanced self-motion
reports compared to stationary and “incompatible” sound

![Fig. 5](image-url) Effects of spatial sound fidelity on vection intensity and presence ratings rendered via BC headset, visual display of $37° \times 30°$ (left panel) and circumaural headphones, visual display of $20° \times 15°$ (right panel). Whiskers show the standard
effects of the means (* marks significance at $p < 0.05$ significance level, ** at $p < 0.01$, *** at $p < 0.005$).
conditions. Taking into account the higher impact of visual cues when perceiving dynamic multisensory information ([34] and references therein), one should expect similar result for conflicting audio-visual rotational cues. Further, dedicated studies should investigate the effects of auditory motion cues congruence and their minimal spatial resolution on such sound enhancement of visually induced circularvection.

In general, both presence andvection ratings showed that moving auditory scenes can be rendered via the BC sound interface, at least at a spatial resolution comparable to 5-channel surround sound systems. In this study we did not adapt binaural cues to the transcranial attenuations of the bone-conduction path. More systematic studies should investigate the properties of interaural transcranial cues and their differences from binaural cues in traditional AC-based technologies. In addition, more dedicated research should address also the minimum audible angle of spatial sound rendered via bone-conduction. However, it should be noted that such measurement results would largely depend on the location of the BC transducers placement.

4. CONCLUSIONS

The present study investigated the feasibility of binaural sound rendering via bone-conduction by applying it for enhancement of a multimodal motion simulator prototype. In agreement with previous findings, the addition of rotating acoustic fields to the visual scene significantly increased illusory self-motion and spatial presence ratings. Thus, our empirical data demonstrates that binaural bone-conducted sound can be used for creating dynamic virtual acoustic spaces. In addition, the study confirmed our previous research showing that sound can reliably enhancevection experiences thus playing an important role in the design of cost-effective motion simulators. Finally, more systematic psychoacoustic studies on spatial bone-conducted sound are essential for the development of future applications.

ACKNOWLEDGEMENTS

The work presented in this paper was funded by EU grant POEMS-IST-2001-39223 (www.poems-project.info) and the Swedish Science Council (VR). The authors would like to thank Max-Planck Institute for Biological Cybernetics for providing the visual stimuli and the VELib programming environment, Tobias Good and Oiiido Equipment AB for providing the two-channel bone-conduction headset and Stefan Stenfelt for insightful references on bone-conducted sound perception.

REFERENCES

[21] A. Väljamäe, Paper D in “Self-motion and presence in the perceptual optimization of a multisensory virtual reality...

[22] IST-FET POEMS (Perceptually Oriented Ego-Motion Simulation) project. http://www.poems-project.info/


A. VA¨LJAMA¨E graduated from Tallinn University of Technology in 2000, degree in radio and communications engineering. He received his M.Sc. and Ph.D. degrees from Chalmers University of Technology, Gothenburg, Sweden, in 2003 and 2007 respectively. His research interests are mainly in multisensory cognition and perception, spatial sound, auditory presence and its applications. Currently he is a postdoctoral fellow at Laboratory for Synthetic Perceptive, Emotive and Cognitive Systems (SPECS), Universitat Pompeu Fabra, Barcelona, Spain.

Ana Tajadura-Jiménez graduated from Universidad Politécnica de Madrid in 2002. In 2003 she received M.Sc.’s degree and in 2007 licentiate’s degrees in Electrical Engineering from Chalmers University of Technology, Gothenburg, Sweden. Currently she is a doctoral student at the Division of Applied Acoustics at Chalmers University of Technology studying human responses to affective multisensory media.

Pontus Larsson studied electrical engineering at the Chalmers University of Technology in Göteborg, Sweden, and obtained his M. Sc. degree in 1999. He then worked as a research assistant at the Division of Applied Acoustics where he received his Ph.D. degree in 2005. Since then Dr. Larsson has been working as a researcher and teacher at Applied Acoustics. His main research areas are multimodal interaction in virtual environments, auralization and sound design. He currently works as an acoustic consultant at WSP Sverige AB, Gothenburg, Sweden.

Daniel Västfjäll received his undergraduate degrees from Göteborg University. In 2002 he received a doctoral degree in Psychology from Göteborg University and in 2003 a second doctoral degree (Acoustics) from Chalmers University of Technology. Currently he is an assistant professor the Division of Applied Acoustics at Chalmers University of Technology and associate professor at Department of Psychology, Göteborg University, Sweden. His research focuses on the role of affect in judgment and decision-making, perception and psychophysics. A recent focus is on affective reactions to sound and vibrations, multimodal perception and the role of sound in the experience of real and virtual worlds.

Mendel Kleiner received his M.Sc.E.E. degree in 1969 and his Ph.D. in Architectural Acoustics in 1978, both from Chalmers University of Technology. He is now professor at the Division of Applied Acoustics at Chalmers University of Technology, Gothenburg, Sweden, teaching audio, electroacoustics, ultrasounds, and architectural acoustics. He is also working as a consultant in the fields of audio, acoustics and noise control. Prof. Dr. Kleiner heads the Chalmers Room Acoustics Group — Multisensory Applications. Currently his main acoustic research interests are in auralization and other virtual reality technologies, musical acoustics, and active sound field control. He is a fellow of the Acoustical Society of America, as well as member of the Institute of Electrical and Electronics Engineers, the Audio Engineering Society, and the Institute of Acoustics.