TRAVELLING WITHOUT MOVING: AUDITORY SCENE CUES FOR TRANSLATIONAL SELF-MOTION

Aleksander Väljamäe

Department of Signals and Systems,
Chalmers University of Technology,
SE-41296, Göteborg, Sweden.
sasha@s2.chalmers.se

Pontus Larsson, Daniel Västfjäll, Mendel Kleiner

Department of Applied Acoustics,
Chalmers University of Technology,
SE-41296, Göteborg, Sweden.
{pontus.larsson,daniel,mk}@ta.chalmers.se

ABSTRACT

Creating a sense of illusory self-motion is crucial for many Virtual Reality applications and the auditory modality is an essential, but often neglected, component for such stimulations. In this paper, perceptual optimization of auditory-induced, translational self-motion (vection) simulation is studied using binaurally synthesized and reproduced sound fields. The results suggest that auditory scene consistency and ecologically validity makes a minimum set of acoustic cues sufficient for eliciting auditory-induced vection. Specifically, it was found that a focused attention task and sound objects' motion characteristics (approaching or receding) play an important role in self-motion perception. In addition, stronger sensations for auditory induced self-translation than for previously investigated self-rotation also suggest a strong ecological validity bias, as translation is the most common movement direction.

1. INTRODUCTION

Creating a sense of illusory self-motion for the end-user of a Virtual Environment (VE) is crucial for many Virtual Reality (VR) applications, e.g. various motion simulators. Illusory self-motion, also referred to as locomotion or vection, can be described as a sensation of actual movement relative to a stable surrounding environment. While a large body of research has been focused on vection elicited by visual stimuli (e.g. [1] and references therein), research on auditory induced vection (AIV) has received little attention until recently [2] cf. [3], [4] and [5] cf. [1], [6-11].

Auditory induced self-motion can be elicited using moving sound fields, either real (e.g. loudspeaker array presentation) or virtual ones (typically headphone reproduction). Binaural technology provides the most flexible way of creating virtual sound environments, where non-spatialized ("dry") sounds are convolved with pre-measured Head-Related Transfer Functions (HRTFs) of the corresponding spatial positions [12]. Binaural synthesis of moving soundfields is a computationally demanding task, where various factors have to be taken into account, e.g. spatial resolution of HRTFs catalogue, room acoustics model, and sound sources characteristics. Therefore, finding the auditory cues which are most instrumental in inducing vection is an important step towards perceptually optimized VR motion simulators.

In our previous experiments [10], [11] on rotational AIV, we decided to concentrate on the ideas of ecological acoustics which studies sound perception from the perspective of every-day listening experiences [13]. We hypothesized that the type of sound source is an important parameter when studying AIV. Unlike artificial sounds (e.g. pink noise), ecological sound sources can be classified by a listener into spatially "still" (e.g. a church bell) or "moving" (e.g. footsteps) categories. A major finding in our previous research is that the experience of self-motion is significantly higher for sound fields with sound sources from the "still" category representing clearly recognizable acoustic landmarks. The higher speed of the sounds' rotation and the larger number of sources also positively affect the AIV ratings [10], [11].

In the current study we address translational AIV and present our first experiment with artificial stimuli containing noise and tonal sound. Taking into account the findings in [8] and [9], we investigate how factors from ecological acoustics, motion metaphors (e.g. engine sound) and selected attention affect AIV. The results of this experiment serve a basis for the follow-up studies with ecological sound, which is reported in [14].

2. AUDITORY MOTION PERCEPTION

Knowledge on auditory motion perception is essential for determining salient auditory cues contributing to auditory-induced vection. The mechanism of auditory motion perception is a complex phenomenon with many parameters involved and it remains to be an active area of research. Recent evidence from brain studies show that a specific "movement-sensitive" area in auditory cortex is most likely to exist (e.g. [15] and references therein) thus indicating separate mechanisms for stationary and moving sounds localization.

Three main cues for discrimination of auditory motion are intensity, binaural cues and the Doppler effect. Intensity cues arise from the changes in sound pressure level emitted by a moving sound source. Binaural cues reflect the interaural time and level differences (ITD and ILD) at listener's ears. The Doppler effect results in perceived frequency shifts in the case of motion between a sound source and a listener. Lufti and Wang [16] thoroughly examined these three cues and showed that for sound object velocities below 10 m/s, intensity and binaural cues were the most instrumental in providing travelled distance information. The Doppler shifts were pre-dominant for sound object velocity and acceleration judgments. For a higher velocity (50 m/s), the Doppler shift tended to dominate in all discrimination tasks. It is important to note, however, that cue dominance depended not only on the task but also varied between tested individuals.

The intensity cue was found to be dominant for travelled distance perception in an earlier study by Rosenblum et al. [17], where this finding was explained from an ecological acoustics perspective. Recently it has been shown that continuous intensity changes can elicit illusion of pitch shift, which are roughly four times larger than the actual frequency shift caused by the Doppler shift [18]. It was also shown that continuous intensity changes in sound stimuli can solely lead to an illusory pitch shift. The authors concluded that Doppler shift perception in everyday listening is almost entirely driven by the intensity cue [18].

The intensity cue dynamics give rise to several secondary cues contributing to the auditory motion perception. When a sound source passes a listener, a "point of closest passage" is clearly marked by the highest intensity peak [18]. Intensity can be also used for a time-to-arrival estimation by tracking the intensity change rate called an acoustic tau [19], [20]. However, the acoustic tau and the auditory motion parallax (auditory equivalent to visual parallax) have a minor impact on the auditory motion perception compared to stronger cues as intensity and reverberation [21].

Studies on sound intensity perception revealed another interesting effect related to the perception of approaching or looming sound sources. In his recent study Neuhoff [22] showed asymmetry in perception of rising and falling intensity where continuous intensity increase resulted in a stronger perceived loudness change compared to the same intensity fall. This sound "looming" effect also resulted in a different perception of distances travelled by approaching or receding illusory sound sources, which were simulated by rising and falling intensity. Several concurrent studies corroborated the fact that looming sounds have perceptual and behavioural priority and that sounds perceived as approaching have greater biological salience than receding ones [23].

An alternative way of auditory motion perception mechanism was suggested in "snapshot hypothesis" by Grantham in [24], where he proposed that, instead of direct perception of sound objects' velocity, listeners base their judgement on the total distance travelled by these objects. Recent findings by [25] suggest that both direct perception of motion cues and displacement detection can take place. In this light, the effects of attenuation of high frequencies due to air absorption on sound distance perception [26] can play a role in sound motion judgements. Distance perception also depends on the type of sound source and on a listener's familiarity to it [12]. More accurate results in judgements on travelled distance have been found for ecological sounds and sounds that are within listeners' reach [27].

The acoustic environment plays an important role in auditory distance perception, especially for indoor conditions where the ratio between direct and reflected sound is known to be one of the most salient cues [28]. Rosenblum et al. [29] uses the term "echolocation" for the human ability to track the echoic changes while moving in a reverberant environment. Knowledge on a sound source directivity pattern may also play a role in determine a source or self-motion [30].

To summarize, the presented information show that the perception of auditory motion can be influenced by the ecological context of the surrounding soundscape. It supports the suggestion by Popper and Fay [31] that the main function of the auditory localization mechanism may be to provide an input to the listener's perceptual model of the environment rather than exact estimates of sound sources' location and trajectories.

3. WORKING HYPOTHESES

In this experiment, a context-free scenario based on artificial sounds (noises and tones) was used. The hypotheses listed below were tested in the experiment and the results were used for a refinement of the experimental methodology in the follow-up experiments with ecological sounds in [14].

H1 – **Looming sounds**. According to the looming effect, sounds with falling and rising intensity are perceived differently [22]. We hypothesize that if approaching sounds are biologically more salient than receding ones, simulation of moving towards a sound will give a stronger AIV sensation than scenarios where the listener is leaving the sound.

H2 – **Acceleration effect**. The ability of discriminating between the sound sources moving with a constant velocity or acceleration ones is an interesting but rarely addressed question in auditory motion research. Such ability was originally shown in [32] and more evidence was indirectly given by the Doppler effect study in [18], where illusory pitch shifts were found to be dependent on the intensity changes mimicking either accelerating or constant velocity sound source. As the acceleration is a necessary component for the human vestibular system in the perception of self-motion, we believe that the accelerating sounds will have a stronger effect on AIV than the sounds with a constant velocity.

H3 – **Focused Attention**. We hypothesize that in a focused attention task, where participants have to concentrate on intensity changes in one specific sound, the AIV experience will be negatively affected. It is known that monitoring the changes in bodily orientation from vestibular or visual information requires a significant degree of attention or cognitive load [3], [33], [34]. Therefore, distracting the listener from auditory streams which are providing salient information for AIV can negatively affect the self-motion sensation.

H4 – **Auditory motion detection threshold**. We hypothesize that participants experiencing AIV will be slower in detecting a sound object motion. This argumentation was inspired by findings in [35], which showed elevated thresholds for object motion detection when experiencing visually induced vection. Similar experiments in auditory domain were suggested in [36] but, to the best of the authors' knowledge, where never reported before. Testing this hypothesis was done in conjunction with stimuli used for testing H3 and our aim was to get a first insight in developing a proper methodology for further studies on this effect in the auditory domain.

4. METHOD

Experiment was conducted using virtual auditory space and the stimuli were synthesized in Matlab^{TM} using a catalogue of generic HRTFs. Binaural synthesis was used to simplify the experimental setup, which was intended to resemble an optimized, cost-effective VR motion simulator. Moreover, the authors were interested in how the AIV will be affected by imperfections in spatial sound rendering due to generic HRTFs.

The generic HRTFs catalogue was measured from KEMAR mannequin using the procedure described in [11]. For stimuli synthesis only one horizontal plane (-4 degree elevation) with a 5 degree resolution was used. In this experiment no acoustic environment rendering was applied.

4.1. Stimuli

In the stimuli synthesis two different initial distances to the sound sources were used - “distant”, simulating approaching sound objects and “close”, simulating the receding sound objects. Figure 1 illustrates 3 types of excerpts for eliciting translational AIV in the forward direction. Further in the text these 3 stimuli types will be referred as “distant” (approaching), “close” (receding) and “mixed” (one approaching one receding). These stimuli types contained two sound sources with additional “anchor” sound introduced in some conditions (see Table 1 for stimuli design).

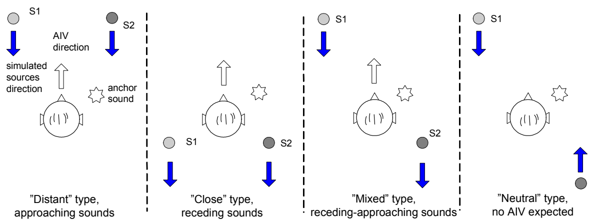

Figure 1. *Stimuli types used in experiment. The arrow in front of the listener indicates expected (AIV) direction, filled arrow indicates the motion direction of the virtual sound objects.*

In addition to the 3 stimuli types in Table 1, four excerpts of “neutral” stimuli type were used in order to provide a baseline for the detection threshold estimation task described in section 4.3. In the “neutral” stimulus the same two sounds objects were moving in the opposite directions (see Fig. 1), and therefore no AIV was expected. Neutral excerpts were always accompanied with the anchor sound.

Table 1. *Experimental design for distant, close and mixed type stimuli (expected AIV direction: forward or backward)*

		No anchor		Anchor	
distance	*velocity*	*AIV direction*		*AIV direction*	
		forw.	backw.	forw.	backw.
Distant (approach.)	const. v accel	x x	x	x x	x
Close (receding)	const. v accel.	x x	x	x x	x
Mixed	const. v	x	x	x	x

In our definition, the “anchor” sound is a sound following moving listener and often caused by him (e.g. sounds of ones own breath, coins in the pocket, engine sound, etc.). The anchor sound was inspired by participants’ verbal responses from [10], where footstep sounds elicited an illusion of moving with a crowd. In the case of AIV the anchor sound should be perceived as accompanying listener’s illusory self-motion. The anchor sound can be seen as an auditory correspondence to the visual self-avatar, which has proven to be an important component for compelling VE experiences [37]. The effects of such sonic self-avatars on AIV were further studied in [14].

Apart from distance and direction, other parameters varied in stimuli design (Table 1) were: 1) sound sources velocity - constant (v = 1 m/s - pedestrian speed in the city) or increasing ($a = 0.012$ m/s^2 – smallest value used in [9]); 2) direction of sound sources movement (forward and backward), which is opposite to expected AIV direction (see Fig. 1).

All stimuli were approximately 1 minute long, with a difference of several seconds between excerpts with accelerating and constant speed sounds. For the accelerating type, sources’ “stationary” phase lasted the first 4 seconds and then the sounds started to accelerate. For the constant velocity type, after the same 4 second stationary phase and 3 second acceleration, sounds speed were constant. A Hann half-window of 0.5 second duration was applied to smooth stimuli on- and off-sets.

Bandlimited pink noise was used for the synthesis of two moving sound sources. Two regions from an idealized critical band filter bank [38] were used - 510-920 Hz (6^{th} -8^{th} band) and 1270-2000 (11^{th}-13^{th} band) – in order to maximize auditory stream separation. This frequency range was chosen to restrict distortions caused by the use of a generic HRTFs catalogue, i.e. a limited spatial resolution and HRTFs’ mismatch to participants’ ears. As ITD dominates spatial sound localization below 1600 Hz, this binaural cue was a main source for such distortions in this experiment and the effects of ILD and spectral cues mismatch were minimized [12].

The anchor sound was represented by a frequency modulated tone with 300 Hz carrier and 20 Hz modulation (modulation index = 0.5). In order to help participants to focus their attention on the anchor sound, it was appearing 3 seconds before the two moving sound sources playback. After 33 seconds of “still” period, the anchor sound intensity started to decrease mimicking a receding sound source with an acceleration of 0.1 m/s. This acceleration was specifically chosen for the detection threshold estimation, where a slow fading would spread detection times over a longer period and thus help timing data post-processing. The instructions for detection of intensity fall are described in the procedure section. The anchor sound was placed on the listener’s right side (see Fig.1) with the intensity subjectively corresponding to a proximal position (1-2 meter range).

Intensities of moving sound sources were changing according to the inverse square law (6 dB level change per distance doubling). However, for the anechoic conditions the distance perception depends on various factors (type of sound source, listener expectations and knowledge, etc.) and sometimes higher intensity changing rates like 9 or 12 dB per distance doubling are used for a more subjectively adequate simulation [26]. In the experiment, the following assumptions about initial distances to the moving sound sources were used 1) distant type – 50 meters from a listener for constant velocity sounds and 20 meters for accelerating sounds (main motivation for this difference was to have the “point of closest passage” at a similar time in all stimuli) 2) close type – 1 meter 3) mixed type - 5 meters for nearby source and 50 meters to distant (see Fig. 1).

4.2. Measures

To assess the AIV, two direct verbal measures were used in this experiment: vection intensity and convincingness of vection. Vection intensity corresponded to the level of subjective sensation when experiencing self-motion. On the convincingness scale participants had to report how convinced they were of having been moving in the direction of the experienced self-motion. It should be noted that the convincingness and intensity ratings are often highly correlated. Ratings of both measures were given on a 0-100 scale.

Apart from the direct measures listed above, an indirect binary measure, reflecting the number of ego-motion experiences, was used (participants were asked to verbally indicate the direction of self-translation). While an onset time for vection experience is often used in experiments with visual stimuli, in the present study the onset time was not measured since previous experiments [10] on auditory-induced vection indicated that this measure showed large inter-individual variance.

According to the hypothesis of change in detection threshold when experiencing AIV (see H4), the reaction times for intensity change in the anchor sound were monitored during the experiment (see procedure for further details)

4.3. Procedure

In the first experiment 24 naive participants (11 male) with a mean age of 24 (SD 3.8) took part. Before coming to the experiment all participants filled in two web-based questionnaires on mental imagery [39] and need for cognition [40].

The experiment was conducted in a special laboratory setup with black curtains surrounding the participant (see Fig. 2). Stimuli were played back with Beyerdynamic DT-990Pro circumaural headphones. Taking into account the experimental procedure in [6] and our previous experiments in [10] and [11], special measures have to be taken in order to amplify AIV sensation. During current experiment, participants were blindfolded and seated on a chair mounted on a wheeled platform coupled with a footrest as shown in Figure 2. The fact that participants knew that the platform could potentially move was intended to increase the convincingness of the simulation setup [10].

Participants were instructed verbally and a short training session was performed before the experiment start (2 stimuli presented). For the anchor sound stimulus, participants were asked to concentrate on the tonal sound and to stop the playback verbally when they heard anchor sound fading or moving away. The reaction times for this detection task were monitored on the basis of listener’s verbal response. After the stop in the stimulus playback, participants had to give ratings on intensity and convincingness of self-motion if such sensation was perceived. They also were asked about the direction of the perceived AIV. Sound excerpts without anchor sound were presented in full length and were followed by the same questionnaire as for the other excerpts. Stimuli were presented in randomized order with small breaks after each 6 excerpts. 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.

Figure 2. *Laboratory setup: a participant sitting on a chair mounted on a wheeled platform coupled with a footrest.*

5. RESULTS

The experiment results from 2 (anchor) x 3 (distance) ANOVA showed no significant effects for the mixed type stimuli and in next subsections only 2 distance types (distant and close) were used for analysis. Results presented in the next subsections did not correlate neither with the need for cognition and mental imagery scores or gender of participants.

It is noteworthy to mention that 9 from 24 participants were sometimes reporting auditory-induced sensation of translation along vertical axis (e.g. forward-down).

5.1. Binary vection measure

One of the measures for the experienced AIV perception is a binary vection measure, which shows how many participants experienced vection for a particular stimulus type (Fig. 1). Results from this experiment showed that inducing self-translation sensation by purely auditory means is more successful compared to self-rotation experiments in [10] – current range of 33-79% (8-18 from 24) is higher compared to rotational binary vections range 23-50% (6-13 from 26).

Figure 3 presents the percentage of experienced binary vections for 8 stimuli types where results for the constant velocity stimuli are averaged for backward and forward directions (see Table 1 and further discussion in section 5.3). It can be seen that higher amount of binary vections are experienced in the no anchor sound condition. Additionally, an asymmetric pattern emerges for the velocity parameter in close and distant conditions. While the stimuli with the constant velocity give almost the same ratings for binary vection - 73% (close) vs. 71% (distant), accelerating sounds show a large shift in favour of the distant, approaching type excerpts (58% vs. 79%, see also Table 3). This asymmetry might be accounted for the difference in the sound sources velocity for the period when sounds are in listeners’ proximity, as previous findings in [10] showed that higher sound velocity positively affects AIV. However, this is only true for the close type condition, where receding sounds are much slower for the excerpts with accelerating stimulus, as in the distant type the accelerating sounds achieve speed of only $\approx$ 0.7 m/s when passing by the listener. Taken together, a first evidence for the difference in AIV sensation induced by accelerating or constant velocity sounds was found and this effect was further investigated in the follow-up experiment reported in [14].

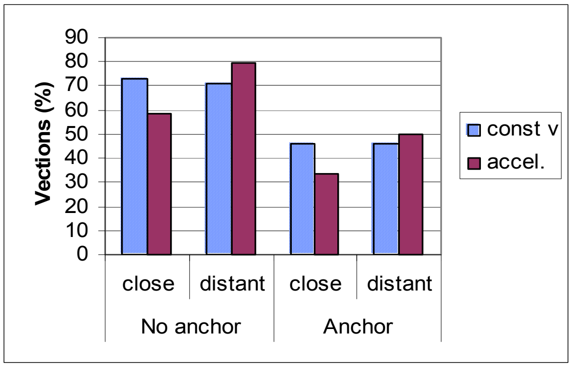

Figure 3. *Reported binary vections (100% correspond to stimuli where all 24 participants experienced vection). "Close" type corresponds to receding sounds and "distant" to approaching sounds.*

Furthermore, simple chi-square analysis shows that binary vection measures for both close and distant stimuli types in the no anchor condition differ significantly from what may be expected from chance, $\chi^2(1) = 4.54$, p.<.05 (close) and $\chi^2(1) = 8.91$, p.<.01. No such difference was apparent in the anchor condition. To further corroborate these results a Friedman rank-test was performed on the vection data (vection coded as 1, no vection coded as 0). The mean ranks were 2.16 (anchor, close), 2.25(anchor, distant), 2.70 (no anchor, close) and 2.89 (no anchor, distant). The test statistic for this analysis was highly significant, $\chi^2(3) = 18.51$, p.<.001.

5.2. Intensity and convincingness

Two separate ANOVAs with factorial design of 2 (anchor) x 2 (velocity) x 2 (distance) were conducted for vection intensity and convincingness. The same trends as for binary vection responses were found for anchor and distance parameters. For vection intensity the main effect of anchor was significant F(1, 23) = 29.16, p.<.001, means 18.6 (anchor) vs. 35.6 (no anchor). No other effect reached significance.

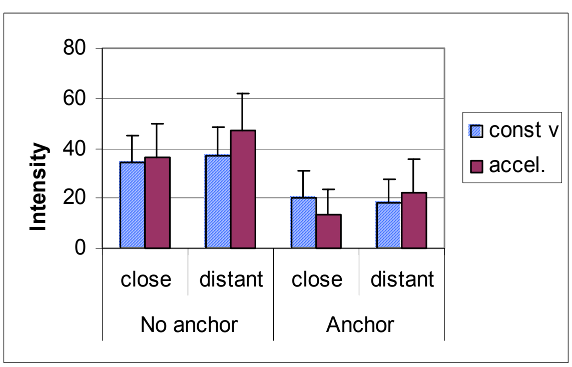

Figure 4. *Intensity ratings with 95% confidence interval (upper bound). "Close" type corresponds to receding sounds and "distant" to approaching sounds.*

Similar results were found for the convincingness data: the main effect of anchor was significant F(1, 23) = 28.86, p.<.001, mean 21.7 (anchor) vs. 40.6 (no anchor). In addition, the main effect of distance was significant, F(1, 23) = 3.99, p.<.005, with a higher mean (35.1) for the distant than the mean (27.3) for the close condition. No other effects or interactions reached significance.

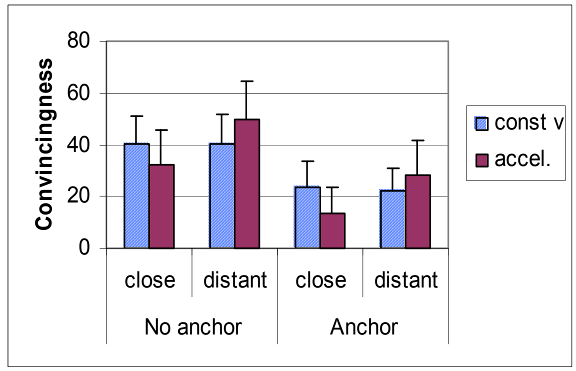

Figure 5. *Convincingness ratings with 95% confidence interval (upper bound). "Close" type corresponds to receding sounds and "distant" to approaching sounds.*

5.3. Front-back reversals

The experiment results showed that the binaural stimuli designed specifically for AIV in forward/backward often did not give the expected self-motion direction sensation. The participants often misinterpreted two moving sounds direction, i.e. front-back reversals occurred [26] where created virtual sources were localized at the opposite to expected side. Due to the high rate of front-back confusions – approximately 30% of the total number of reported AIV, we decided to combine forward-backward stimuli pairs (see Table 1) into one and to discard the direction parameter in the results analysis.

Three listeners perceived the motion of the virtual sources only in the frontal hemisphere thus reversing some parts of perceived motion trajectories (see Fig 6a). This effect was observed from the sudden change in perceived AIV direction in the participants' reports. As no information about perceived movement of the sound objects was asked, one could assume that front-back confusions occurred also in the cases when no AIV was reported.

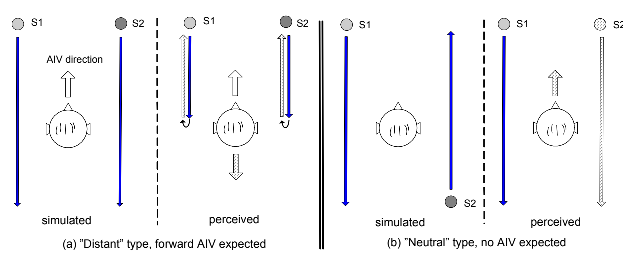

Figure 6. *Examples of back-front reversals (dashed lines represent perceived "reverse" trajectories and resulted AIV direction): a) sound objects motion is perceived only in front of the listener leading to switching of simulated forward AIV to backward direction; b) sound objects motion direction is coupled leading to undesired AIV responses.*

It is interesting to note, that usually there is a significant asymmetry between front-back (FB) and back-front (BF) reversals with stronger tendency for FB confusion (roughly 3 times higher than BF) [26]. On the contrary, in our experiment this difference was not observed and almost equal numbers of reversals in both directions were found for most of the participants. Moreover, on the average BF reversals were even more frequent – 25% vs. 15%. Unfortunately, the neutral type stimuli were also often - 50% of the times on average - inducing AIV as shown on Fig 6b. Analysis of these "undesired" self-motions reports also showed equal occurrence of both reversal types.

5.4. Detection threshold estimation

Vection experiments in the visual domain showed that the self-motion sensation affects object motion perception thresholds [35]. The aim of the detection threshold task in this experiment was to take the first steps in the validation of this effect for the auditory domain. Unfortunately, the problem of front-back reversals also occurred with the neutral type stimuli (see Fig. 1), were two sound sources were perceived as moving in one common rather than in opposite directions, which in turn lead to unwanted AIV sensations. Some participants' responses (for roughly half of the participants) allowed for detection times comparison between cases with or without AIV experience. An indication of the threshold increase for reported self-motion cases was observed, however, this trend did not reach significance.

This part of the experiment gave us a first insight into the methodology for a new subjective corroborative measure of AIV, which might be based on auditory motion detection threshold estimation. In the refined experimental design several parameters have to be more carefully selected: 1) precise measurement of detection times 2) truly "neutral" stimuli not eliciting self-motion but resembling the spectral content of AIV oriented stimuli; 3) intensity change rate if it is used a motion cue for the detection task. Apart from the tasks involving distance and velocity judgements for auditory motion, sound localization tasks might be also a proper substitute to this methodology.

6. DISCUSSION AND CONCLUSIONS

This study aimed at identifying the auditory cues important for translational self-motion sensation. One major finding was that a focused attention task significantly reduced AIV. Moreover, asymmetry in AIV was found between stimuli containing approaching or receding sound sources, where approaching sounds had marginally stronger impact on AIV. The current results also show that auditory induced self-motion is a more reliable phenomenon for simulated translational movements than for rotational movements [10], and that only a set of minimal acoustic cues is sufficient for successful translational AIV simulation.

The finding that a selected attention task reduces AIV ratings might be biased by the difference in the procedure for stimuli conditions with or without the anchor sound. In the anchor condition stimuli playback was interrupted which could have affected participants' AIV ratings. However, continuous stimuli playback was used in our follow-up study [14] with the similar experimental methodology, and we found similar trends. We therefore suggest that the self-motion sensation can be negatively affected if the listener is distracted from the auditory stream which provides salient information for AIV. This reasoning is in line with findings from vestibular [33] and visual [34] self-motion research.

The result of asymmetry in translational AIV for the distant (approaching) and close (receding) type stimuli supports our hypothesis that looming sounds might increase translational AIV experience. The perception of looming objects is more biologically salient than for receding ones and evidence for this effect has been found both in auditory only [23] and audio-visual domain [41].

On the other hand, we have not found any systematic gender differences in the perception of the looming effect as was recently presented by [42]. Taking into account that the looming effect is significantly stronger for tonal than for noise sounds [22], the noise stimuli used in the present experiment could have prevented gender differences. As was suggested by [22], tonal sounds are more likely to represent separate sound objects or acoustic landmarks; on the contrary broad-band noise usually represents a surrounding environment with multiple sound sources. The looming effect perspective can bring to a new understanding of the results presented in [8], where approaching noise stimuli were found to be more instrumental for eliciting self-motion than the receding ones.

An alternative explanation for the asymmetry found in AIV responses for approaching and receding conditions is the influence of the "point of closest passage". In the distant condition the sound sources were passing by the listener, which was not the case for the receding sounds (see Fig. 1). The "point of closest passage" might be an important component for the self-motion sensation and follow-up translational experiments reported in [14] give further evidence to salience of this cue for AIV. When the experimental procedure was changed and participants had to stop the sound playback when experiencing AIV, most of the times the self-motion sensation was built-up at the time of the "point of closest passage".

In the current study we do not find a significant difference for the AIV between the sound objects moving with acceleration or constant velocity. The previous study in [10] showed that the higher sound objects' velocities and the higher number of sound sources were more instrumental for rotational AIV. Similar results can be predicted for translational AIV – the recent study in [9] showed that higher values of acceleration were more instrumental for auditory-induced self-motion. It is interesting to note that in the current experiment, the "mixed" type stimulus, where one receding source was separated in time from another approaching source (see Fig. 1), can be seen as a single source stimulus compared to other conditions with two sounds moving together. Therefore, the lower AIV ratings trend for the mixed stimuli type can be accounted to the lower number of sound objects moving in one direction.

Difference in stimuli with accelerating or constant velocity sound could also cause asymmetry for distant and close types perception, however further study in [14] on this parameter suggests it might be closely related to the looming effect.

In general, the results from our studies show that translational AIV is more easily induced than rotational AIV, even with fewer special measures (blindfolding, wheeled platform but not special instructions, see [10] for more details) applied to achieve self-motion sensations. This is not surprising as translational movement is more common experience from an ecological perspective. Interestingly, even with artificial noises representing sound sources, participants tended to create a specific scene context, e.g. being in the train, metro or driving in a tunnel. Moreover, the almost equal percentage for front-back and back-front reversals suggests that an ecological context can influence the auditory scene perception. Sound localization in binaural synthesis systems has previously been found to be rather asymmetric in favour of sound appearance behind a listener [26]. The reason why such asymmetry was not found in the current results can be explained by the fact that people are simply more used to move in forward direction.

This experiment and the previous findings [10] suggest that the auditory scene consistency and ecological validity plays a crucial role in AIV. In the current experiment we deliberately used only the most salient acoustic cues (sound intensity and ITD) for moving sound fields simulation, and the quality of the rendering was determined by a generic HRTFs catalogue with a relatively low spatial resolution. The results suggest that this reduced level of details in spatial sound rendering can be sufficient for creating a self-motion sensation. This, in turn, would allow allocating sound processing resources for other tasks including, for example, low latency rendering of real-time interaction.

7. FUTURE WORK

Two follow-up experiments on translational AIV have been conducted and the results will be reported in [14]. These findings suggest that reverberation and auditory scene "spaciousness" might play an important role in AIV, which will be investigated in the future experiments. Participants' verbal responses show that auditory-induced sensation of translation along vertical axis (e.g. elevator sensation) can take place, which can be an interesting topic for future AIV studies. The refinement of detection threshold estimation procedure as a corroborative measure of AIV is included in the currently conducted self-motion experiments.

8. ACKNOWLEDGEMENTS

The work presented in this paper was supported by the EU FET Presence Research Initiative project POEMS (Perceptually Oriented Ego-Motion Simulation), IST-2001-39223 and the Swedish Science Council (VR). The first author of this paper would like to thank for the support from Alfred Ots' scholarship foundation.

9. REFERENCES

[1] G.J. Andersen, "Perception of self-motion: Psychophysical and computational approaches", *Psychol. Bull.*, vol. 99(1), pp. 52-65, 1986.

[2] V. Urbantschitsch, "Über Störungen des Gleichgewichtes und Scheinbewegungen [On disturbances of the equilibrium and illusory motions]", *Zeitschrift für Ohrenheilkunde*, vol. 31, pp. 234-294, 1897.

[3] T. Mergner and W. Becker, "Perception of horizontal self-rotation: multisensory and cognitive aspects", in *Perception and Control of Self-motion: Resources for Ecological Psychology*, R. Warren and A.H. Wertheim, Eds. Hillsdale, NJ: Erlbaum, 1990, pp. 219-263.

[4] S. yon. Stein, *Schwindel (Autokinesis externa et interna)* [*Vertigo (External and internal self-motion)*]. Leipzig: Lessier, 1910.

[5] R. Dodge, "Thresholds of rotation", *J. Exp. Psych.*, vol. 6, pp. 107-137, 1923.

[6] J.R. Lackner, "Induction of illusory self-rotation and nystagmus by a rotating sound-field", *Aviat. Space Environ. Med.*, vol. 48(2), pp. 129-131, 1977.

[7] Y. Suzuki, S. Sakamoto, and J. Gyoba, "Effect of auditory information on self-motion perception", in *Proc. ICA*, Tokyo, 2004.

[8] S. Sakamoto, Y. Osada, Y. Suzuki, and J. Gyoba, "The effects of linearly moving sound images on self-motion perception", *Acoust. Sci. & Tech*, vol. 25(1), pp. 100-102, 2004.

[9] B. Kapralos, D. Zikovitz, M. Jenkin, and L.R. Harris, "Auditory cues in the perception of self-motion", in *116th AES convention*, Berlin, 2004.

[10] P. Larsson, D. Västfjäll, and M. Kleiner, "Perception of self-motion and presence in auditory virtual environments", in *Proc. of Seventh Annual Workshop Presence 2004*, Valencia, Spain, 2004, pp. 252-258

[11] A. Väljamäe, P. Larsson, D. Västfjäll, and M. Kleiner, "Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments" in in *Proc. of Seventh Annual Workshop Presence 2004*, Valencia, Spain, 2004, pp. 141-147

[12] J. Blauert, *Spatial Hearing*. MIT Press, Cambridge, rev. edition, 1997.

[13] W.W. Gaver "How do we hear in the world? Exploration of ecological acoustics", *Ecol. Psychol.*, vol. 5(4), pp. 285-313, 1993.

[14] A. Väljamäe, P. Larsson, D. Västfjäll and M. Kleiner, "Sonic self-avatars and self-motion in virtual environments", manuscript in preparation

[15] J.D. Warren, B.A. Zielinski, G.R. Green, J. P. Rauschecker, and T. D. Griffiths, "Perception of sound source motion by the human brain", *Neuron*, vol. 34, pp. 139–148, March 2002.

[16] R. A. Lutfi and W. Wang, "Correlated analysis of acoustic cues for the discrimination of auditory motion", *J. Acoust. Soc. Am.*, vol. 106(2), pp. 919–928, 1999.

[17] L.D. Rosenblum, C. Carello, and R.E. Pastore, "Relative effectiveness of three stimulus variables for locating a moving sound source", *Perception*, vol. 16, pp. 175–186, 1987.

[18] M.K. McBeath and J.G. Neuhoff, "The Doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change", *Psychon. Bull. Rev.*, vol. 9(2), pp. 306-313, 2002.

[19] D.N. Lee, "Getting around with light or sound", in *Perception and Control of Self-motion: Resources for Ecological Psychology*, R. Warren and A.H. Wertheim, Eds. Hillsdale, NJ: Erlbaum, 1990, pp. 487-505.

[20] B.K. Shaw, R.S. McGowan, and M.T. Turvey, "An acoustic variable specifying time to contact", *Ecol. Psychol.*, vol. 3, pp. 253-261, 1991

[21] J.M. Speigle and J.M. Loomis, "Auditory distance perception by translating observers," in Proc. of the IEEE Symposium on Research Frontiers in Virtual Reality, 1993, pp.92-99.

[22] J.G. Neuhoff, "An adaptive bias in the perception of looming auditory motion", *Ecol. Psychol.*, vol. 13(2), pp. 87-110, 2001.
[23] D.A. Hall and D.R. Moore, "Auditory neuroscience: the salience of looming sounds", *Curr Biol.*, vol. 13(3), R91-3, 2003.
[24] D.W. Grantham, ''Detection and discrimination of simulated motion of auditory targets in the horizontal plane,'' *J. Acoust. Soc. Am.*, vol. 79, pp. 1939–1949, 1986
[25] S. Carlile, and C. Best, C. ''Discrimination of sound velocity in human listeners,'' *J. Acoust. Soc. Am.*, vol. 111, pp. 1026–1035, 2002
[26] D.R. Begault, *3D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.
[27] L.D. Rosenblum, A.P. Wuestefeld, and K.L. Anderson, "Auditory reachability: An affordance approach to the perception of sound source distance", *Ecol. Psychol.*, vol. 8(1), pp. 1-24, 1996.
[28] W. Bronkhorst and T. Houtgast, "Auditory distance perception in rooms," *Nature*, vol. 397, pp. 517-520, 1999.
[29] L.D. Rosenblum, M.S. Gordon, and L. Jarquin, "Echolocating distance by moving and stationary listeners", *Ecol. Psychol.*, vol. 12(3), pp. 181-206., 2000
[30] J.G. Neuhoff, M.A. Rodstrom, and T. Vaidya, "The audible facing angle", *J. Acoust. Soc. Am.*, vol. 109(5), pp. 2376-2377, 2001.
[31] A.N. Popper and R.R. Fay, "Evolution of the ear and hearing: issues and questions", *Brain Behav. Evol.*, vol. 50, pp. 213-220, 1997
[32] D.R. Perrott, B. Constantino, and J. Ball ''Discrimination of moving events which accelerate or decelerate over the listening interval'', *J. Acoust. Soc. Am.*, vol. 93, pp. 1053–1057, 1993.
[33] L. Yardley, M. Gardner, N. Lavie, and M. Gresty, "Attentional demands of perception of passive self-motion in darkness", Neuropsychologia, vol. 37, pp. 1293–1301, 1999
[34] G. Rees, C.D. Frith, and N. Lavie, "Modulating irrelevant motion perception by varying attentional load in an unrelated task", *Science*, vol. 278, pp. 1616–19, 1997
[35] T. Probst, S. Krafczyk, T. Brandt, and E. Wist. 1984. "Interaction between perceived self-motion and object-motion impairs vehicle guidance", *Science*, vol. 225, pp. 536-538, 1984.
[36] A.H. Wertheim, "Motion perception during self-motion: The direct versus inferential controversy revisited", *Behav. Brain Sci.*, vol. 17(2), pp. 293-355, 1994.
[37] M. Usoh et al., "Walking>virtual walking>flying, in virtual environments", in *Proc. of SIGGRAPH'99*, 1999, pp. 359-364
[38] B. Scharf, "Critical bands," in *Foundations of Modern Auditory Theory*, J. V. Tobias, Ed., Academic, San Diego, 1970 pp. 159–202.
[39] P.W. Sheehan, "A shortened form of Betts' Questionnaire Upon Mental Imagery", *J. Clinical Psych.*, vol. 23, pp. 386-389, 1967
[40] J.T. Cacioppo, R.E. Petty, and C.F. Kao, "The efficient assessment of need for cognition," *J. Pers. Assess.*, vol. 48, pp. 306-307, 1984
[41] J.X Maier, J.G. Neuhoff, N.K. Logothetis and A.A. Ghazanfar, "Multisensory integration of looming signals by rhesus monkeys", *Neuron*, vol. 43, pp. 177-181, 2004
[42] J.G. Neuhoff and T. Heckel, "Sex differences in perceiving loudness change and auditory looming", in *Proc. of ICAD'04*, 2004