# Perception of Self-motion and Presence in Auditory Virtual Environments

Pontus Larsson[1], Daniel Västfjäll[1,2], Mendel Kleiner[1,3]
[1]Department of Applied Acoustics, Chalmers University of Technology, Göteborg, Sweden.
[2]Department of Psychology, Göteborg University, Göteborg, Sweden
[3] Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY, USA
{pontus.larsson@ta.chalmers.se, daniel.vastfjall@psy.gu.se, kleinm2@rpi.edu}

**Abstract**

*Apart from inducing a high sense of presence, creating a sensation of self-motion is often a key issue of many Virtual Reality applications. However, self-motion perception (vection) has been primarily investigated for visual stimuli. This study explored the possibility of inducing vection by realistic auditory stimuli. Furthermore, influence of various audio design parameters on auditory-induced vection was studied. The results suggest that sound source characteristic is a primary determinant of auditory-induced vection, especially for an environment with only a single sound source. However, it was also found that the type of sound source may play less of a role when the environment contains multiple sound sources. Auditory-induced vection is also depending on whether or not room acoustic cues are included in the simulation. It is however likely that it is the interaction between type of sound source and the environment that is important.*

*Keywords---* **Self-motion perception, Vection, Auditory cues, Ecological acoustics, Binaural reproduction, Auralization, Presence**.

## 1. Introduction

In many Virtual Reality (VR) applications creating a compelling sensation of self-motion is central. A large body of research reports that illusory self-motion can be elicited by visual stimuli. These illusory sensations of self-motion, often called vection, can be induced by large visual stimuli that move in a uniform manner. Observers then perceive as moving in the opposite direction to the visual stimulus [1].

However, in the real world, several other types of cues, primarily vestibular, auditory and somatosensory sensory signals, provide important self-motion information [2]. VR fun-rides, flight simulators and other types of motion simulators often utilize vestibular cues to enhance the sensation of self-motion.

Nonetheless, creating vestibular cues by means of e.g. motion platforms is most often technically complex, expensive and requires safety measures to be taken in order not to harm the user. It therefore seems more attractive to optimize other cues, such as auditory, to enhance the sensation of self-motion in VR-simulations.

However, research on illusory self-motion induced by sound-fields received little attention until recently [2-5]. One of the few studies available was performed by Lackner [2] where subjects were exposed to both real (loudspeaker array presentation) and virtual (dichotically presented) rotating sound fields. In this study it was found that subjects experienced self-rotation in both real and virtual conditions but that the real sound field was significantly more effective than the virtual one. Furthermore, when the contours of the experimental room were visible to the subject, auditory stimulation did not elicit illusory self-rotation.

The present study investigates some of the parameters of a rotating, virtual sound field, which may affect the illusion of self-rotation. The experiment resembles the study performed by Lackner [2] but our main goal here is to investigate how self-rotation is elicited by realistic stimuli such as those created for use in Virtual Environments and computer games.

## 2. Hypotheses

Based on the ideas of ecological acoustics [6,7], we believe that the character of the sound source is a relevant parameter that needs to be considered when studying auditory-induced illusory self-motion. Including easily recognizable sound sources in an acoustic motion simulation simply helps us resolving the conflict “I am moving” versus “The source is moving”.

For example, a rotating sound field containing sounds which immediately can be identified as being created by immovable objects, such as church bells or fountains, tells us “I am moving” while sound sources that are easily characterized as moving (such as sounds of driving cars, bicycles etc.) signal “The source is moving”. Synthetic sounds or, in general, sounds that do not provide any information on what type of source emitted them, are unable to provide such motion identification information. Instead, the auditory system has to rely totally on binaural cues or room acoustic cues in trying to resolve the motion identification conflict.

Moreover, we hypothesize that the intensity of auditory-induced vection is contingent also on the number of concurrent sound sources present in a rotating auditory stimulus, given that all sources move at the same angular speed. This is because we simply have more cues available when resolving the self vs. object motion conflict.

Furthermore, the current experiment also investigates the influence of acoustic rendering quality on auditory-induced vection. A recent finding from research in the visual domain is that spatial presence is a possibly vection-mediating factor [1]. That is, if we have the feeling that we are in a particular spatial context, we can also be more easily convinced that we are actually moving. An intriguing question is thus whether adding room acoustic simulation (auralization) can facilitate auditory-induced vection since it has been shown that post-experimental subjective ratings of presence is contingent on the quality of auralization [8,9].

It may also be the case that if room acoustic cues are included in the simulation, these cues directly facilitate the resolution of the object/self-motion conflict. From a physical perspective, it is clear that for some types of environments, such as asymmetrically shaped rooms, there is indeed a measurable difference in the listener's position between rotating the listener and moving the sound source (see Figures 1 and 2).

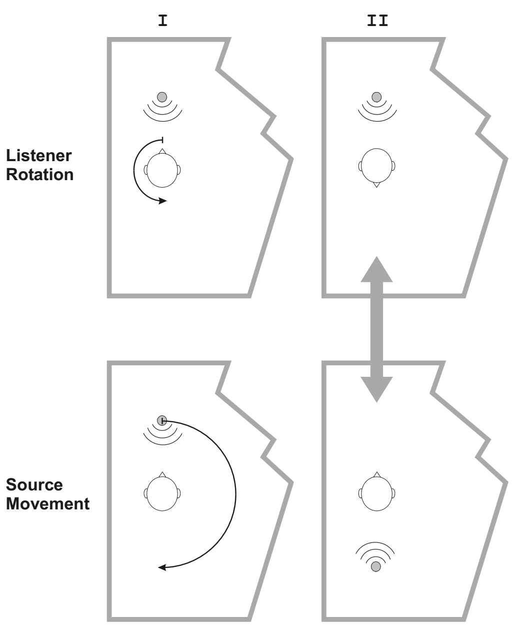


**Figure 1 Physical comparison of Listener Rotation (LR) and Source Movement (SM). Unless the surfaces are perfectly absorbing and source / listener are located as shown in the left panels, the impulse responses in the listening position will be different for LR and SM.**

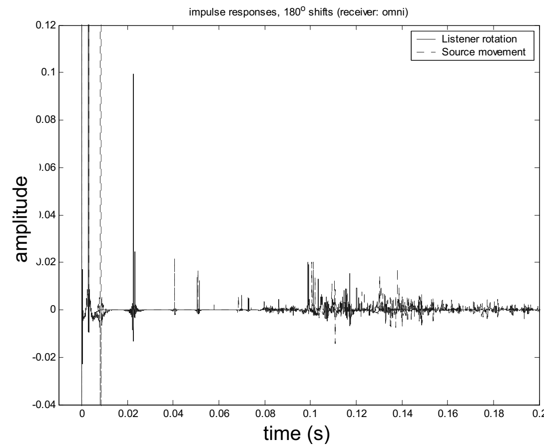


**Figure 2 Simulated impulse responses for LR (solid) and SM, for the source / listener locations shown in Figure 1, right panels.**

An informal listening session on simulations of these two situations (listener rotation and source movement) confirms that they are indeed perceptually different, however further investigations have to be undertaken in order to reveal whether this difference has any effect on vection responses.

In sum, the hypotheses for the current experiment can be stated as;

$H_1$: Rotating auditory stimuli consisting of one or more sound sources will elicit a stronger, more compelling sensation of self-rotation if the sound source(s) can be identified as being still compared to if the sound source(s) can be identified as being moving or if it is unidentifiable (artificial).

$H_2$: A rotating auditory stimulus consisting of several concurrent sound sources will elicit a stronger, more compelling sensation of self-rotation compared to if the stimulus includes only one sound source.

$H_3$: An acoustic simulation of listener rotation where realistic room acoustics cues are included will give rise to a stronger sense of presence and self-rotation compared to a simulation where only direct sound is rendered.

## 3. Method

The stimuli used to test the hypotheses presented above were all binaural simulations of a virtual listener standing in one place and rotating a certain number of laps. Sound sources that were included in the simulation were *never actually moving*; only *the character* of the sound source was varied as discussed below.

Stimuli were rendered offline in CATT-Acoustic v 8 [10] by using the "Walkthrough Convolver". The parameters varied were:

1) Auralization quality (Marketplace or anechoic rendering).

2) Number of concurrent sound sources (1 or 3),
3) Type of input source sound (still, moving or artificial)
4) Turn velocity (20, 40 or 60 degrees per second)
5) Turn direction (left or right)

The acoustic model used to render the stimuli in this experiment is shown in Figure 3. The size of the model is approximately (W x L x H) 40x70x16 m. The main reason for choosing this model and not an indoor environment was that sound sources such as buses and fountains could be naturally associated with this type of outdoor context. Furthermore, a highly detailed visual model of this environment already exists which allows for performing similar experiments in auditory-visual conditions. Two different types of auralizations were created; one were realistic absorption and diffusion values were assigned to the model's surfaces (RT= 1.2s @ 1 kHz) and one where only the direct sound (source-receiver) was included, i.e. an anechoic auralization.

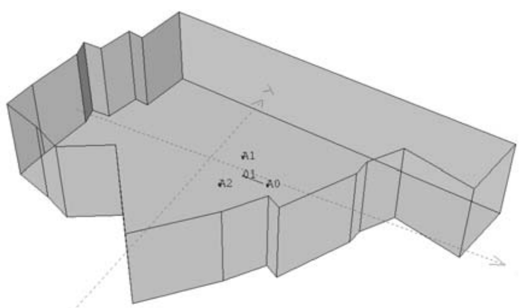


**Figure 3 The acoustic model used in the experiment: A model of the market place in Tübingen, Germany. A0, A1, A2 denotes source positions and 01 denotes the receiver position.**

Binaural Room Impulse Responses (BRIR) were calculated for head azimuth angles between 0-358.5 degrees at a resolution of 1.5 degrees yielding a total of 240 BRIR-pairs per source-receiver combination. Non-individualized HRTFs measured from a human subject were used (CATT1_Plain) in all stimuli. In addition, the stimuli were also equalized for the headphones used in the experiment (Beyerdynamic DT-990, see next section).

For stimuli with three concurrent sound sources, three different source positions located at 0, 120 and 240 degrees relative the receiver's median plane were used (see figure 3, positions A0, A1, and A2). For the single source stimuli, the source located at 0 degrees (position A0) was used. The distance between the sources and the receivers was 5 meters.

Concerning parameter 3, three different source sounds in each category (Still (S), Moving (M) or Artificial/Ambiguous (A)) were selected and assigned to the different virtual sources. These were: S1) Bus on idle, S2) Small fountain S3) Barking dog, M1) Footsteps, M2) Bicycle, M3) Driving bus, A1) Stationary pink noise, A2) Pink noise bursts, 250 ms + 250 ms of silence, and A3) Pink noise bursts of random length and temporal distribution.

By using the Walkthrough Convolver, various listener rotations could be simulated. The velocity of these listener rotations followed the following profile: 1) Stationary listener for 3s 2) Acceleration to maximum velocity (20, 40 or 60 degrees / second) for 3s 3) Constant rotation speed for 60s 4) Deceleration for 6s. Both left and right turns were simulated.

### 3.1 Apparatus

The experiment was conducted in a semi-anechoic room. Stimuli were presented with Beyerdynamic DT-990Pro circumaural headphones driven by a NAD Amplifier, model 3020.

Given the results in Lackner [2], some special measures were taken in order to achieve auditory-induced vection and to make the experience more convincing. First of all, a special seating arrangement, shown in Figure 4, was used. The arrangement consists of an ordinary office chair mounted on an electrically controllable turntable placed on a wooden base plate. The purpose of using this type of seat was to make the participant believe that rotational movements actually occurred during the experiment (although they in fact did not). Furthermore, the arrangement also prevented the participants from having any contact with footrests or the floor. In addition, four loudspeakers visible to the participant as he/she entered the test room, were placed around the experimental chair. Finally, the participant was also blindfolded during each trial.

**Figure 4 Participant on experimental chair**

### 3.2 Measures

As the subject heard the sound, he/or she was asked to report when vection was perceived simply by saying the direction (left/right). The experiment leader then noted the time and direction. After each trial, subjects were also asked to fill in a single-page questionnaire containing the following items:

1) Intensity of vection (0-100)
2) Compellingness of vection (0-10)
3) Sensation of source vs. self motion (-5-5)
4) Localization of sounds (0-6)
5) Sensation of sounds coming from different directions (0-6)
6) Envelopment of sound (0-6)
7) Realism (0-6)
8) Magnitude estimation, presence (0-100)

### 3.3 Participants procedure and design

Twenty-six participants (13 female) with a mean age of 24 (SD 3.7) participated. They participated on two separate occasions approximately two weeks apart.

A within-subjects design was used where Type of sound source (3: moving, still, artificial), Velocities (3: 20, 40, 60 deg/sec), and Acoustic rendering: (2: marketplace-anechoic) was varied for both single and multiple sound sources. For each type of sound source three different sounds were used. To avoid exposing participants to all combinations, this variable was a between-group variable. Also, the single and multiple sound sources were separated and tested on different occasions. Thus, each participant was exposed to 18 combinations on each occasion. Different randomized orders of presentation were used for each participant. However, all participants were exposed to the single sound sources on the first occasion.

Upon arrival to the laboratory, the experiment leader thoroughly instructed participants on the use of equipment and scales. After a period of relaxation, the participant was seated in the chair and the task (binary vection) and questionnaire was then introduced. Following this, the participant was blindfolded and a series of two test sounds were replayed. After the test sounds, the participant was again reminded about the procedure and the main experimental task started. After completing all ratings, participants were debriefed, thanked and paid for their participation.

## 4. Results

The data was analyzed separately for single and multiple sound sources, respectively. Initial data screening included tests of possible effects of the between group variable sound (three different types of sounds for different participants) for each type of sound source, tests of effects of direction (left or right), as well as tests of demographic factors (age, gender). No systematic effects of any of these variables were obtained and therefore they were discarded in subsequent analyses. Also the three rating scales concerning sound characteristics were mainly included as controls and are not directly relevant for testing the hypotheses. The only systematic effects on these scales were, as expected, that the marketplace environment was more immersive and that sounds could be better localized than in the anechoic environment. For this reason, no further data are presented related to these three ratings scales. However, it should be noted that the fact that participants rated the marketplace sounds as being more easily localizable is somewhat surprising. A possible explanation to this effect is that the higher realism and increased externalization in the marketplace conditions made participants believe that they could localize sounds more easily, although they in fact could not.

### 4.1 Single sound sources

*Binary vection and vection onset time.* The number of participants experiencing vection as indicated by the binary vection measure is shown in Figure 5. Overall, vection was relatively low (the range was 6-13 [23-50%] of the participants), but it can be seen that the still sound sources as expected induced higher vection than both moving and artificial. This pattern was also more pronounced for the marketplace environment than the anechoic. No systematic effects of velocity were evident, why the data was collapsed over these conditions. A McNemar-test on these data showed that a significant higher number of vection responses (36) was obtained in the still, than the moving (21) or artificial (27) conditions for the marketplace environment. This effect was however not obtained for the anechoic environments where the number of responses in the still sound source condition was similar (29) as the moving (22) and artificial (24) conditions.

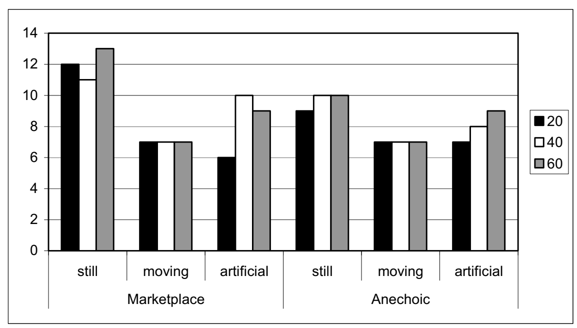


**Figure 5 Frequency of binary vection.**

Visual inspection of the vection onset-time suggests that overall onset time is shorter for the anechoic than the marketplace environment (Figure 6). Second, the artificial sound sources resulted in longer vection onset time as compared to still and moving sound sources for both types of renderings. It should however be noted that because these means are based on six to thirteen observations, no inferential parametric statistical tests can be performed and these results should be interpreted with caution.

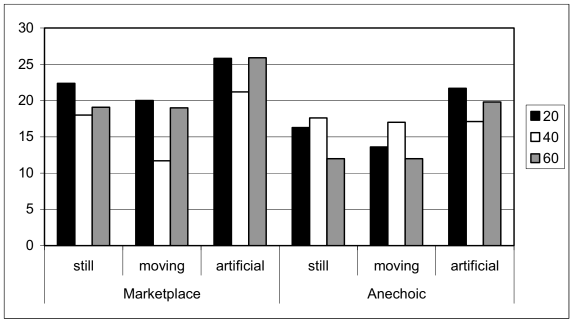


**Figure 6 Vection onset time in seconds.**

*Rating scales.* Next, the rating scales that participants filled out after each sound presentation were analyzed. 3 (sound source) x 3 (velocity) x 2 (rendering) within-subjects ANOVAs were performed on each of the rating scales. Grenhousse-Geisser correction was used to correct for unequal variances. A significance level of $p<.05$ was adopted as a criterion for the inferential statistics.

*Vection intensity and Convincingness of vection*. The means for the intensity ratings are shown in Figure 7. The ANOVA for intensity yielded a significant main effect of sound source ($F(2,25) = 5.66$, $p<.001$) where the still sound sources were significantly higher ($M = 36.3$) than the moving ($M = 20.3$) and artificial ($M = 20.1$). Neither the main effects of velocity and rendering, nor any interactions reached significance ($F>1$).

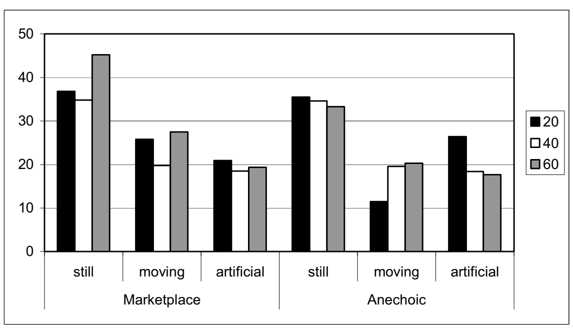


**Figure 7 Vection intensity**

Similar results were obtained for the convincingness ratings. A main effect of sound source ($F(2,25) = 7.12$, $p<.001$) where the still sound sources were significantly higher ($M = 5.1$) than the moving ($M = 3.4$) and artificial ($M = 3.4$). A main effect of rendering was also found ($F(2,25) = 4.76$, $p>.01$, where the rendered marketplace was perceived as more convincing ($M = 4.80$), than the anechoic ($M = 3.11$). As may be seen in Figure 8 there was also a significant interaction between sound source and rendering ($F(2, 38) = 3.90$, $p<.05$), where the still sound source condition were more convincing in the marketplace than the anechoic environment. No other effects reached significance.

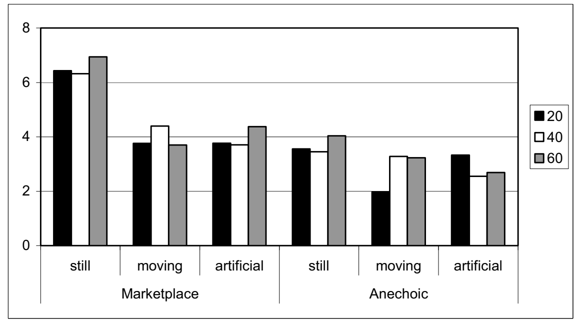


**Figure 8 Convincingness.**

Ratings of object- vs ego-movement only showed a significant interaction ($F(2,19) =4.20$, $p<.02$) between rendering and environment, in that more object motion was perceived for the marketplace environment under 20 $^{o}$/s and 40 $^{o}$/s velocities, while the reverse was true for the 60 $^{o}$/s velocity.

*Presence and realism.* For the magnitude estimation presence scale the only significant effect was a main effect of sound ($F(2, 38) = 3.69$, $p<.05$) where the still sound sources were higher ($M = 64.0$) than both moving ($M = 55.4$) and artificial ($M = 55.5$).

For the realism scale none of the main effects reached significance, but the interaction between sound source and rendering was significant ($F(2, 38 = 4.77$, $p<.01$). Analyses of the means showed that that both the still and moving sound sources received higher realism ratings in the marketplace condition, while no such difference was evident for the moving sound sources.

### 4.2 Multiple sound sources

*Binary vection and vection onset time*. The number of participants experiencing vection as indicated by the binary ego-motion measure is shown in Figure 9. Overall, vection was higher than for single sound sources, but still far from all participants experienced vection (the range was 7-17 [28-66%] of the participants). Figure 9 indicates that the stationary sounds sources as expected induced higher vection than both moving and artificial. This pattern was also more pronounced for the marketplace environment than the anechoic. No systematic effects of velocity were evident, why the data was collapsed over these conditions. McNemar-tests on these data showed that a significant higher number of vection responses (50) was obtained in the still, than the moving (39) or artificial (36) conditions for the marketplace environment. This pattern was however not as pronounced for the anechoic environments where the number of responses in the still sound source condition was similar (42) as the moving (35) and artificial (33) conditions.

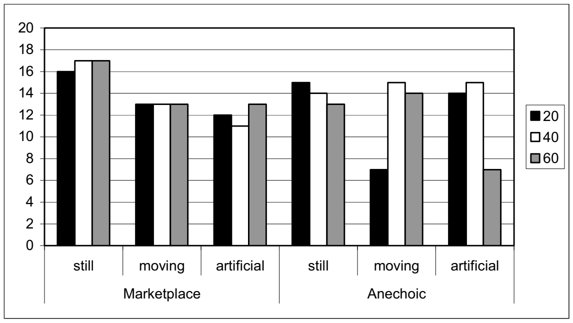


**Figure 9 Binary vection, multiple sound sources**

Visual inspection of the vection onset-time suggests that overall onset time in the anechoic and the marketplace environment are similar (Figure 10). There is however a trend across conditions that vection onset time is shorter for the 60 $^{o}$/s velocity (as compared to 20 $^{o}$/s and 40 $^{o}$/s).

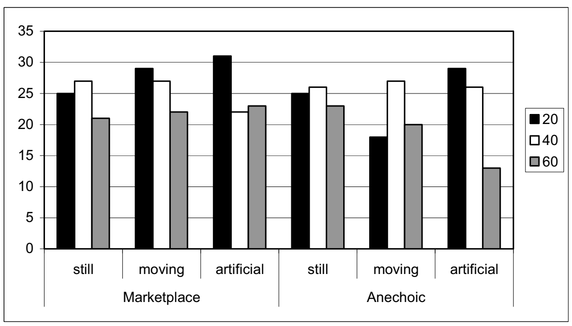


**Figure 10 Vection onset time in seconds for multiple sound sources.**

*Vection intensity and Convincingness of vection.* The ANOVA for intensity yielded a significant main effect of sound source ($F(2,25) = 12.15$, $p<.001$) where the still sound sources were significantly higher ($M = 39.2$) than the moving ($M = 30.6$) and artificial ($M = 29.7$). Also a significant main effect of velocity ($F(2,25) = 7.74$, $p<.001$) was retrieved where the 60 $^{o}$/s velocity was higher ($M = 38.5$) than both 20 $^{o}$/s ($M = 30.5$) and 40 $^{o}$/s ($M = 32.0$). No other effects reached significance.

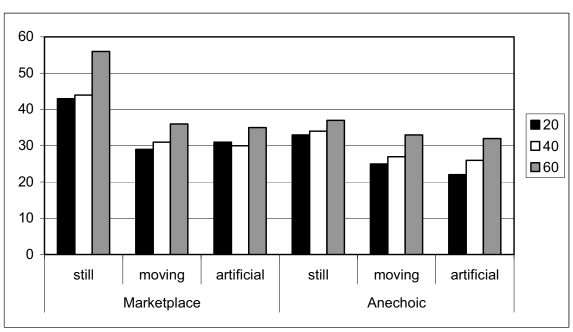


**Figure 11 Intensity of vection for multiple sound sources**

The same pattern was found for convincingness ratings. A main effect of sound source ($F(2,25) = 6.98$, $p<.001$) where the still sound sources were significantly higher ($M = 4.7$) than the moving ($M = 3.9$) and artificial ($M = 3.6$). A main effect of velocity was also found ($F(2,25) = 5.37$, $p>.01$, where the 60 $^{o}$/s velocity was higher ($M = 4.7$) than both 20 $^{o}$/s ($M = 3.1$) and 40 $^{o}$/s ($M = 4.1$). No other effects reached significance.

*Presence and realism.* For the magnitude estimation presence scale the three main effects all reached significance while no interaction effect was significant. A main effect on sound sources ($F(2, 38) = 5.07$, $p<.01$) showed that the still and moving sounds induced higher presence ($M = 63.1$ and $62.7$, respectively) than the artificial ($M = 52.3$). For velocity, ($F(2, 38) = 3.55$, $p<.05$) the 60 $^{o}$/s velocity induced higher presence ($M = 62.2$) than 20 $^{o}$/s and 40 $^{o}$/s velocities ($M = 58.8$ and $56.2$, respectively). Finally, the effect on rendering $F(1, 38) = 4.01$, $p<.01$) showed that the marketplace environment gave rise to higher presence ($M = 62.4$) than the anechoic ($M = 59.1$).

For the realism scale a main effect of sound source was found $F(2, 38) = 16.37$, $p<.001$) where the artificial sound received lower ratings ($M = 3.17$) than did still and moving (M:s 4.0 and 4.2, respectively). A main effect of rendering was also found $F(1, 38) = 6.44$, $p<.01$), where the marketplace environment ($M = 4.4$) was rated as being more realistic than the anechoic ($M = 3.5$). No other effects were significant.

### 4.3 Comparison of Single and Multiple sound sources

*Binary vection*. Overall, the frequency of "yes" responses on the binary vection measure increased with app. 20% when comparing single to multiple sound sources. Collapsed across rendering conditions the McNemar tests showed that the increase was significant ($p<.05$) for all cases.

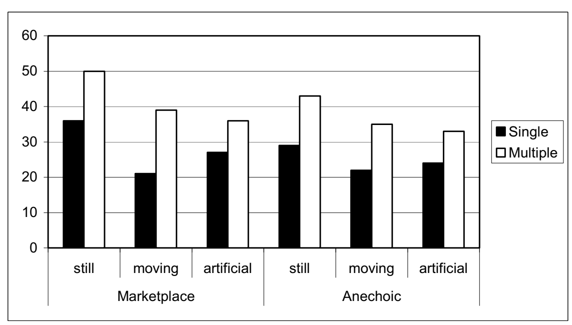


**Figure 12 Comparison of single and multiple sound sources on number of vection responses.**

## 5. Discussion

Overall, support was found for the hypothesis that still sound sources are more instrumental in inducing vection than both moving and artificial sound sources. Measures of vection and intensity/convincingness showed exactly this for both single and multiple sound sources. The effect was however more pronounced for single sound sources.

Second, some support was found for the notion that a realistically rendered environment may increase perception of self-motion. For single sound sources the marketplace resulted in slightly more vection responses and higher ratings of convincingness. While the effect of rendering on the binary vection response was replicated for multiple sound sources, no effects were obtained on intensity or convincingness. However, as expected, the realistically rendered environment received higher presence ratings for both single and multiple sources.

Third, in line with our hypothesis, multiple sound sources induced significantly more vection responses than the single sound source condition.

Finally, and somewhat unexpected, we found that velocity influenced vection for multiple sound sources. More specifically, the faster velocity simulations (60 $^{o}$/s) seemed to induce more vection as measured on the binary measure and on the rating scales. Even though this effect was not predicted, it mimics results on visual vection [1]. The reason why the velocity effect was not found for single sound sources is unclear. However, this might again indicate that single sound source environments provide unstable reference frames, which are unsuitable for inducing vection.

In summary, the present results suggest that the type of sound source is a primary determinant of auditory-induced vection, especially for an environment where there is only a single sound source. The present findings however suggest that the type of sound source may play less of a role when the environment contains multiple sound sources. Auditory-induced vection is also depending on the rendering of the environment. It is however likely that it is the interaction between type of sound source and the environment that is important. Our results suggests that the rendering mainly affected ratings of vection for the still sound sources while it had little effect for ratings of the moving and artificial sound sources.

The finding that auditory-induced vection is higher for multiple sound sources and that rendering in those cases become less important is a result directly applicable to the development of a perceptually driven ego-motion simulator; in scenes with multiple sound sources, simple room acoustic simulation could be used with the computational effort instead being allocated to realistic rendering of sound source movements.

The present experiments have concerned rotational movement. However, we believe that the ideas of ecological acoustics can be employed in the case of linear vection as well, which is something that will be investigated in future experiments. Furthermore, the possibility of using other measures of auditory vection, such as nystagmus [2] and postural responses [11], will be explored in future work. Measuring motion after-effects and motion sickness may also provide further insights in the area of auditory-induced vection.

## 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.

## References

[1] J. Schulte-Pelkum, B.E. Riecke, M. von der Heyde, H.H. Bülthoff. Circular vection is facilitated by a consistent photorealistic scene. In *Proceedings of Presence 2003*, Aalborg, Denmark, October 2003.

[2] J. R. Lackner. Induction of illusory self-rotation and nystagmus by a rotating sound-field. *Aviation, Space and Environmental Medicine*, 48(2), 129-131. 1977.

[3] Y. Suzuki, S. Sakamoto, J. Gyoba. Effect of auditory information on self-motion perception. In *Proceedings of ICA*, Tokyo, Japan. 2004.

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

[5] B. Kapralos, D. Zikovitz, M. Jenkin and L. R. Harris. Auditory Cues in the Perception of Self-Motion. In *116th AES convention*, Berlin, Germany. May 2004.

[6] W. W. Gaver. How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5, 285-313. 1993.

[7] W. W. Gaver. What in the world do we hear? An ecological approach to auditory event perception. Explorations in ecological acoustics. *Ecological Psychology*, 5, 1-29. 1993.

[8] C. Hendrix, W. Barfield. The sense of presence within auditory virtual environments. *Presence – Teleoperators and Virtual Environments*, 5(3), 290-301. 1996.

[9] P. Larsson, D. Västfjäll, M. Kleiner. Spatial auditory cues and presence in virtual environments. *Manuscript in preparation.* 2004.

[10] CATT-Acoustic 8.0 (Computer software). Gothenburg, Sweden. URL: http://www.catt.se

[11] W. Ijsselsteijn, J. Freeman, H. de Ridder, S. E. Avons, D. Pearson. Towards an objective corroborative measure of presence: Postural responses to moving video. Presented at *PRESENCE 2000 – 3rd International Workshop on Presence*, Techniek Museum, Delft, The Netherlands, March 27-28, 2000.