Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments

Aleksander Väljamäe[1], Pontus Larsson[2], Daniel Västfjäll[2,3] and Mendel Kleiner[4]
[1] Department of Signals and Systems, Chalmers University of Technology, Sweden
[2] Department of Applied Acoustics, Chalmers University of Technology, Sweden
[3]Department of Psychology, Göteborg University, Göteborg, Sweden
[4]Program in Architectural Acoustics, Rensselaer Polytechnic Institute, Troy, NY, USA
{aleksander.valjamae@s2.chalmers.se, pontus.larsson@ta.chalmers.se,
daniel.vastfjall@psy.gu.se, kleinm2@rpi.edu}

Abstract

It is likely that experiences of presence and self-motion elicited by binaurally simulated and reproduced rotating sound fields can be degraded by the artifacts caused by the use of generic Head-Related Transfer Functions (HRTFs). In this paper, an HRTF measurement system which allows for fast data collection is discussed. Furthermore, effects of generic vs. individualized HRTFs were investigated in an experiment. Results show a significant increase in presence ratings of individualized binaural stimuli compared to responses to stimuli processed with generic HRTFs. Additionally, differences in intensity and convincingness of illusory self-rotation ratings were found for sub-groups of subjects, formed on the basis of subjects' localization performance with the given HRTFs catalogues.

Keywords--- **Self-motion perception, Auditory presence, Binaural reproduction, Individualized Head-Related Transfer Functions.**

1. Introduction

Creating a sense of presence in the end-user of a Virtual Environment (VE) is one of the main goals of Virtual Reality (VR) technology. The feeling of presence is often described as a sensation of "being there", whereas several other definitions exist [1]. Being "spatially present" in a specific context provides a stable reference frame, which allows for a good spatial orientation and spatial updating. Illusory ego- or self-motion (vection) can be described as a sensation of actual movement relative to a stable surrounding environment and has been shown to be closely related to spatial presence. For example, positive correlation between presence ratings and on-set times for induced illusory self-motion has been recently shown in experiments with visual stimuli [2].

A large body of research is concentrated on the illusory self-motion elicited by visual stimuli. On the contrary, research on auditory illusory self-motion received little attention until recently [3-6]. In our first experiments conducted within the European project POEMS (Perceptually Oriented Ego-Motion Simulation) [7], we focused on illusory self-rotation induced by sound-fields [8]. In this study, rotating sound sources were presented to subjects via headphones and the sensation of ego-motion in the opposite direction was expected.

These previous experiments were based on the ideas of ecological acoustics and it was hypothesized that the type of the sound source is an important parameter when studying auditory-induced illusory self-motion. Opposite to "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 was that both presence ratings and experience of self-motion were highest for the sound fields containing sound sources from the "still" category. Speed of sounds' rotation and number of sources also positively affected the ratings [8].

The stimuli in [8] were created using binaural technology, which is a two-channel spatial sound rendering technology where headphones are typically used for playback [9]. Pre-measured catalogues of Head-Related Transfer Functions (HRTFs) are used for binaural sound synthesis, where a non-spatialized ("dry") sound is convolved with transfer functions corresponding to the desired spatial position of the source. Larsson et. al. [8] used a catalogue of non-individualized HRTFs provided for the CATT Acoustics auralization software [10].

However, binaural sound synthesized with non-individualized HRTFs often can be perceived as distorted because of the mismatch between the listener's own and generic transfer functions. When generic HRTFs are used the most common problem is in-head localization (IHL), where sound sources are not externalized but are rather perceived as being inside the listener's head [11]. Another known artifact is a high rate of reversals in perception of spatial positions of the virtual sources, where binaural localization cues are ambiguous (cone of confusion), e.g. front-back confusion [9]. Errors in elevation judgments can be also observed for stimuli processed with non-individualized HRTFs [12]. Applying individualized HRTFs for auditory VEs can significantly reduce the artifacts described above [11].

One of the goals of this study was to test the performance of the HRTFs measurement system designed by Chalmers Room Acoustic Group (CRAG). We decided to repeat some of the experiments on illusory self-rotation presented in [8], this time using individualized HRTFs. Responses from verbal probing in [8] showed that in some cases the participants experienced artifacts in presented sound scenes, such as non-circular trajectories and nonrealistic, very close distances to the moving sources. We believe that these artifacts may have been caused by the use of non-individualized HRTFs and that these artifacts in turn might have influenced ego-motion and presence ratings.

The main hypothesis for the current experiment stated that the higher presence ratings should be achieved with individualized HRTFs, since previous studies indicate that spatialization and localization may be linked to presence experiences [13-14]. A second hypothesis was that improved spatial quality of auditory scene might affect subjects' experience of ego-motion.

Auditory localization performance can depend on several factors in complex spatial sound scenes. Langendijk et. al. [15] studied the effect of localization of target sounds in the presence of one or two distracter sounds, which were interleaved but not overlapped with the target sound in time. They found that the localization performance was degraded as the number of distracters increased. In auditory VE with multiple ecological sounds target-distracter pairs can easily occur.

We decided to investigate how distracters added to the auditory VE can influence presence ratings. Our third hypothesis was that adding the auditory distracters, which are irrelevant to the sound scene, would decrease the localization accuracy (divided attention). This in turn might decrease the influence of non-individualized HRTFs artifacts on the sound scene perception in VE.

2. HRTF measurement system

The procedure of measuring catalogues of individualized HRTFs is a cumbersome and time-consuming procedure (for reviews see [9, 12, 16-17]). In line with POEMS project requirements, it was decided to develop a HRTFs measurement system which would allow for fast data collection in non-anechoic and somewhat noisy environments such as offices. In the CRAG HRTF measurement system the transfer functions are recorded for a grid of spatial positions on a virtual sphere, which center is aligned with a subject's head-related coordinate system.

2.1. Physical setup

In our laboratory setup we built an array of 32 loudspeakers as shown in Fig. 1, which can be seen as one sector or "vertical slice" of a virtual sphere with a radius of 1.25 meters (far-field acoustical mode measurements). Loudspeakers were non-uniformly placed at 16 elevation angles, which guarantee a resolution of less than 10 degrees in the vertical plane. A test person must be shifted 19 times (20 sectors) for the full HRTF catalogue measurement resulting in less than 8 degree resolution in the horizontal plane. If higher frontal resolution is required, one additional sector can be measured, which results in 63 measured azimuth positions in the horizontal planes with elevation angles of 16, 6, -4 and -14 degrees.

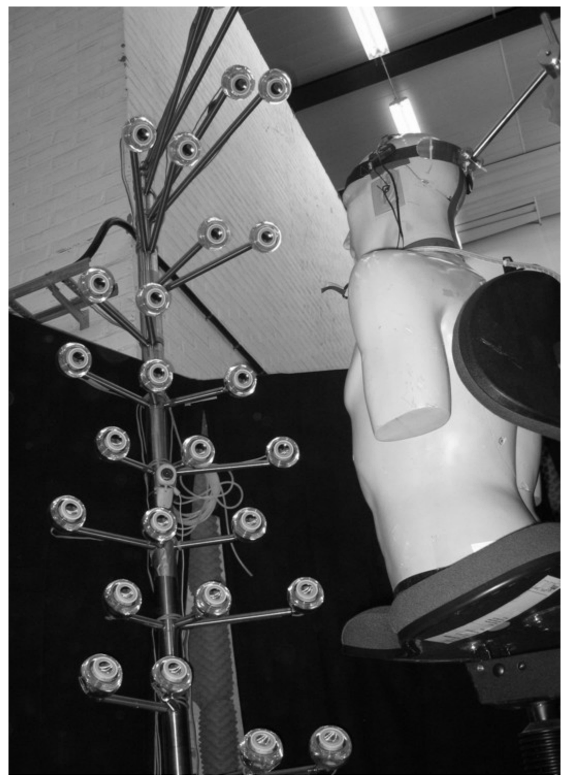

Figure 1 HRTFs catalogue measurement system

In our current setup, HRTFs catalogue measurements are conducted in a big room with a reverberation time of approximately 0.7s and noise floor of 30 dBA. The test subject is seated on a swivel chair and the head is fixed with a special headrest and an elastic band. The position of the measurement system in the room ensures that no early reflections arrive within the first few milliseconds after the direct sound. Additionally, the floor is covered with sound absorbing material.

Apple Pro™ loudspeakers were chosen for the loudspeaker array because of their small size and spherical shape so as to reduce scattering of sound by neighboring loudspeakers. DPA 4060 miniature condenser microphones are placed at the entrance of the blocked-ear canal of the subject. This type of HRTF measurements gives results comparable to those measured with a microphone probe immediately inside the ear canal opening [18], but the procedure is faster and more convenient for the subject under measurement. An M-Audio Audiophile 2496 PCI-bus card is used for playback and recording of measurement signals because of its large dynamic range and the availability of driver routines for Linux. Specially written software for raw HRTFs data collection is used.

2.2. Post-processing of measured data

In our system we use frequency sweeps (quadratic chirp) as the deterministic stimulus for measurements of Head-Related Impulse Responses (HRIRs). This method was chosen due its immunity to harmonic distortions introduced by the measurement chain and the low crest factor of the signal (ratio between peak and root-mean-square in the voltage values) [19]. The latter property helps to ensure desired 20 dB signal-to-noise ratio for a desired frequency range. The chirp is band-limited from 0.1 to 15 kHz and its duration is 2048 samples (≈43 ms, f_s = 48 kHz).

After the measurement procedure, the raw HRIRs are processed using Matlab™ software. The raw HRIRs are first deconvolved with the chirp signal and a half-Hanning window is applied to keep first 256 samples of the response. At this stage, the Interaural Time Delay (ITD) is estimated using the difference between onset times of the left and right HRIRs. This allows for optimizing in further processing steps by working only with the HRTFs magnitude responses. The ITD is added only to the final HRIR dataset. This decision was motivated by the assumption that it is possible to use linear-phase HRTFs for binaural sound synthesis without adding perceptually significant artifacts [20].

Furthermore, the raw HRIRs collected by the system have to be compensated for the artifacts coming from transducers limitations. For this purpose "baseline" or free-field responses are measured with the microphones placed at the center of the virtual sphere. Free-field corrected HRTF magnitude responses can then be obtained by division of the raw HRTF by the corresponding baseline TF data in frequency domain.

Interpolation is needed to obtain a final HRTFs catalogue with uniformly distributed spherical grid. For the current experiment, a 5-degree resolution in the horizontal plane was required for smooth rendering of moving sound sources. This is comparable with an average localization blur, which is 3 degrees for frontal positions and up to 20 degrees for peripheral and rear positions [11]. Spherical spline interpolation in the frequency domain is known to give best results compared to other interpolation types [21]. Before interpolation, magnitude responses are smoothed using the procedure described in [22]. Instead of using a perceptually motivated reduction of magnitude response, smoothing is applied to all data points. After interpolation, HRTF magnitude responses are used for creating linear-phase FIR filters. A circular shift equal to the earlier estimated ITD is introduced within filter pairs, which represent HRTFs for certain spatial locations.

The processing steps presented above were used for creating the stimuli for the current experiment. One of the goals of this study was to test the performance of the CRAG HRTF measurement system and apply all necessary corrections at the post-processing stage if needed. At this stage we found good results for the HRTFs for the frequency range from 0.1 to 13 kHz. Deficiencies observed for frequencies above 13 kHz had no affect to the experiment as ecological sounds with frequencies from 0.1 up to 10 – 12 kHz were used.

3. Method

Twelve subjects (five male) with a mean age of 24 (SD 2.2) from the previous study described in [8] participated in the experiments. All subjects had normal hearing verified by a standard audiometric procedure [23]. After completing the experiment, subjects were debriefed, thanked and paid for their participation.

3.1. Measures

To assess auditory induced vection and subjective presence sensation, three direct measures were used in this experiment: presence, vection intensity and convincingness of vection.

Presence was defined in the questionnaire as "a sensation of being actually present in the virtual world", which corresponded to the single perceptual dimension without any interaction with the VE. Vection intensity corresponded to the level of the subjective sensation when experiencing self-motion. On the Convincingness scale subjects had to report how convincing the sensation of self-motion was. It should be noted that the convincingness and intensity ratings often are highly correlated. Ratings of all three 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-rotation). While on-set time for vection experience is often used in experiments with visual stimuli, the previous experiment [8] on auditory-induced vection indicated that this measure showed large inter-individual variance. The present study measured onset time, but since no systematic effects were found, results from this measure will not be presented.

3.2. Stimuli

In the current experiment, rendering of acoustic environment was not considered as being important; therefore all stimuli were simulations of an anechoic environment rendered off-line. Three parameters varied in the presentation of the rotating sound field:

- Rotation velocity (20 or 60 degrees/second)
- Number of concurrent sound sources (1 or 3)
- Type of HRTFs catalogue used for stimuli synthesis (individualized or generic catalogue)

Since results from the previous experiment showed that the "still" type sound sources are the most instrumental in inducing ego-motion, only sounds from this category were used: “bus on idle”, “small fountain”, and “barking dog”. The stimuli duration was approximately 1 minute and consisted of the following parts: 3 seconds in stationary listening position, 4 seconds acceleration to maximum velocity, 60 seconds constant rotation speed and 4 seconds deceleration.

The stimuli were synthesized using one horizontal slice of HRTFs at -4 degree elevation. The stimuli synthesized with individualized HRTFs was contrasted with one processed with generic HRTFs catalogue, which resulted in 4 pairs of stimuli kept together for the verbal probing purposes. HTRFs measured from the KEMAR (Knowles Electronic Manikin for Acoustic Research) mannequin were used as the non-individual catalogue. Headphone equalization was applied to the final sound excerpts in order to prevent coloration artifacts.

For testing the effects of irrelevant auditory distracters, 20 clicks (6 kHz carrier, 4 ms duration) were added at random time moments to the two stimuli from the main set described above (1 and 3 sound sources, synthesis with generic HRTF, velocity of 60 degrees/second). Clicks were also convolved with KEMAR HRTFs and appeared at random positions in space. During the experiment stimulus with the clicks always followed the same stimulus without distracters hence creating two pairs for verbal probing.

Apart from the pair restrictions described above, all 10 stimuli were presented in the randomized order for proper statistical analysis.

3.3. Procedure

The experiment was conducted in a semi-anechoic room, where stimuli were played back with Beyerdynamic DT-990Pro circumaural headphones. Participants were asked to report verbally the direction of their rotation – i.e. left or right, if they felt self-motion during the particular stimulus playback. Stimulus was stopped after the positive ego-motion response and subjects were asked to rate presence, intensity and convincingness.

In the current experiment presence is studied from the ego-motion perspective and rapid interruption, which could certainly influence the presence ratings, is acceptable. If the ego-motion sensation was not reported during the stimulus playback, only the presence rating was asked. Apart from the verbal responses to the questionnaire, verbal probing was done by the experiment leader.

Taking into account results by Lackner [3], special measures were taken in order to achieve auditory ego-motion. Participants were seated on an ordinary office chair, which was mounted on an electrically controllable turntable with a wooden base plate. Subjects were instructed that turntable would be used during some of the experimental trials. However, the turntable was still throughout the experimental trials. This manipulation was foremost used to make participants believe that they could actually move during the experiment. In addition, the turntable height prevented the subjects from having their legs any contact with the floor. In order to make the experimental setup look more convincing, four loudspeakers, placed around the experimental chair, were visible to participant as he/she entered the test room. The loudspeakers were never in use during the experimental trials. Finally, during the experiment participants were blindfolded.

3.4. HRTFs quality-test

In order to evaluate the quality of the individualized HRTFs a short test was performed before or after the main experiment. The purpose of this quality-test was to justify the improved localization performance compared with localization when using the generic (KEMAR) HRTFs catalogue. Instead of common strategies of defining absolute localization accuracy (e.g. [24]), we decided to use a simplified procedure, where a level of the most usual binaural rendering artifacts acts as indirect quality measure (it was important to keep the quality-test short in duration, as it had to be conducted together with the main experiment). The major parameters then are the front-back confusions rate and the externalization of perceived virtual sources.

It was decided to evaluate 6 positions on the horizontal slice at –4 degrees elevation: three in front (315, 0 and 45 degrees) and their reverse positions on rear (225, 180 and 135 degrees). A fountain sound of frequency range from 2 to 12 kHz was used as a sound source, since it is well known that, apart from small head movements, spectral differences in HRTFs around 5 and 9 kHz help to resolve front-back confusion [9]. Figure 2 shows how spectral differences can vary between the subjects.

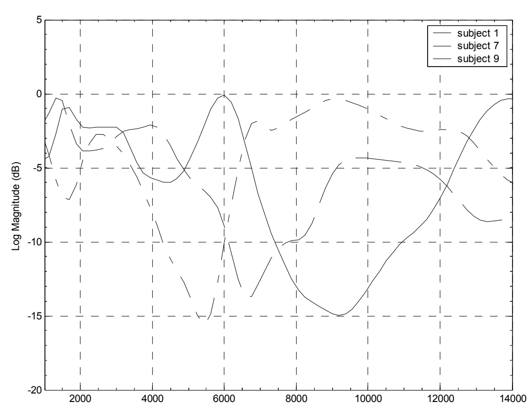

Figure 2 Difference in spectra between two front-back source locations on a cone of confusion (45 and 135 degrees azimuth, -4 degrees elevation) for three representative subjects

Four different measures were used for evaluating the HRTF quality: 1) the front-back confusion ratio, 2) the relative distance to the sound source, 3) errors in the elevation perception, and 4) responses from the verbal probing.

The front-back confusion ratio was based on the subjects’ estimates of the spatial position of the sound source. In order to simplify this task, participants were asked to use a clock metaphor when verbally indicating one of the six positions listed above (e.g. 0 degrees azimuth corresponds to 12 o'clock).

For the distance evaluation, participants were asked to rate the distance to the source in meters. Later, answers were converted to the relative scale from 0 to 100, where 100 corresponded to the maximum perceived distance to the source from the stimuli pair processed with individualized or generic HRTFs.

For the elevation perception measure, subjects were asked to indicate the position of the source regarding the horizontal plane in the head-related coordinates system. Subjects were asked to indicate the height of the sound source relative to this plane.

As a last binary measure of the HRTF catalogues' quality, results from the verbal probing in the main experiment were used. Based on the subjects' comments on perceived trajectories of virtual sources, their distance and overall spatial scene consistency, decision was made regarding the preference between stimuli processed by individualized or generic HRTF catalogues.

The quality-test stimuli consisted of three pairs of subsets synthesized by individualized and generic HRTFs catalogues. Each subset contained a fountain sound sequentially presented for 6 seconds from each of 6 spatial positions in random order. One pair was used for the elevation perception measure and two other pairs for collecting data on the front-back confusion ratio and the perceived source distance.

4. RESULTS

4.1. Sub-groups based on the HRTFs quality-test

Results from the quality-test for all four measures highly varied among the subjects. However, when analyzing all 4 measures for each subject, these results could be combined into a final measure of preference between generic and individualized HRTFs used for the stimuli synthesis. Based on this binary measure, subjects were subdivided into two subgroups for further analysis: a G-group (better localization with individualized HRTFs) and a B-group (no clear preference between generic and own HRTFs). It has to be noted that localization performance could be degraded either by the HRTFs catalogues accuracy or due to individual localization abilities (see section 5 for a discussion).

For several subjects, the difference between the performance with individualized and generic HRTFs catalogues was not prominent but for a proper statistical analysis an equal number of participants was allocated to each group (i.e. median split). In this case the binary measure based on the verbal probing (see section 3.4) was used for the final decision.

Tables 1 and 2 show the average results for the three quality-test measures for all subjects and for the two subgroups formed. Table 1 shows average rates of front-back confusion, where responses from individualized and KEMAR HRTFs are compared. Means for all subjects showed an 11 % increase in the front-back confusion ratio for generic HRTFs. The sub-group analysis showed no effect for the B-group, but for the G-group 20 % improvement was found when individualized HRTFs were used.

HRTFs	All	G-group	B-group
Ind.	33	32	35
KEMAR	44	53	35

Table 1 Average rates (%) of front-back confusion for all subjects and the subgroups

Table 2 shows the responses for distance perception, where sub-group differences can be clearly seen. For the G-group, individualized HRTFs resulted in more distant percepts of the virtual source - 14 % improvement. For the B-group, the effect was reversed, resulting in small 8 % degradation for the stimuli processed with individualized HRTFs catalogues.

HRTFs	All	G-group	B-group
Ind.	61	72	50
KEMAR	58	58	58

Table 2 Average distance responses in relative scale (100 corresponds to the most distant percept)

The elevation perception measure was strongly biased by the high-frequency contents of the sound used for the quality-test. Sound with such characteristics is usually perceived as being located higher than its actual position [9]. This was clearly seen from the subject responses; more than 70 % of source positions were judged as being located above the horizontal plane. In general, both groups showed smaller deviations in the sound height judgments when individualized HRTFs were used. At the same time, the B-group showed roughly 3-times larger deviations in the answers to this measure compared to the G-group.

4.2. Main effects: Ego-motion and Presence

All dependent variables were submitted to separate 2(HRTF) x 2(Number of Sources) x 2(Velocity) ANOVAs. First, the analysis using the binary ego-motion measure yielded no statistically significant differences. The analyses of intensity and convincingness showed no significant main effects for HRTF's or velocity, but in both instances a significant main effect of number of sources was found. The means for the intensity measure were 20.5 (1 source) vs. 31.0 (3 sources), $F(1, 11) = 4{,}29$, $p<.05$. Similarly the means for convincingness were 22.2 (1 source) vs. 31.3 (3 sources), marginally significant for a $F(1,11) = 4.13$, $p =.06$. No other effects were significant for these measures.

For the presence ratings a significant main effect of HRTF was found where the individualized HRTFs yielded higher presence ($M = 61.8$) than did the generic KEMAR HRTFs ($M = 57.8$), $F(1,11) = 5.43$, $p<.05$. No other main effects or interaction effects reached significance.

Stimuli with auditory distracters affected only the presence ratings. While almost no effect was found for the stimuli containing multiple sources: M = 58.3 (with distracters) vs. M = 59.2 (without distracters), presence ratings for the single rotating sound source were higher for the stimulus with distracters (M = 58.8) compared to non-distracter condition (M = 52.9). In addition verbal probing showed a clear difference in overall judgment of the presented sound scene and the trajectories of the virtual sound sources. In the presence of distracters less distorted trajectories were perceived.

4.3. Subgroup differences in ego-motion perception

HRTFs quality-test results presented in Tables 1 and 2 motivated a sub-group analysis of the data from the main experiment. Table 3 shows the results from the statistical analysis of 3 direct measures used in the main experiment, where 4 pairs of stimuli were used for non- and individualized HRTFs catalogues comparison. However, since the sample size was too small to allow for parametric statistical analyses, only trends are reported here.

As was shown in [8], multiple sources positively affect presence ratings and this can be seen for both subgroups in the table. The same trend continued for intensity and convincingness of ego-motion, but in the B-group the difference is negligible. On the contrary, the G-group showed clear discrimination in ratings when the number of sources presented in stimuli were increased.

Measures	Source type	G-group	B-group
Intensity	Single	10.4	30.6
	Multiple	30.0	32.1
Convincingness	Single	12.9	31.2
	Multiple	30.4	32.3

Table 3 Effects of concurrent sound sources number on intensity and convincingness average ratings in the sub-group analysis

5. Discussion

The major finding in the present study was that stimuli processed with individualized HRTFs catalogues resulted in a significant increase of presence ratings as compared to stimuli processed with generic HRTFs. Several other lines of research have independently showed that individualized HRTF increase spatial perception and spatial abilities [16, 24] or that more spatialized sound increase the sense of presence [8, 13-14]. However, to the best of our knowledge, this is the first study to show a direct link between individualized HRTFs and spatial auditory presence.

In addition, the results for ego-motion (intensity and convincingness ratings) showed consistency with the previous findings reported in [8] - a higher number of concurrently rotating sources and a higher rotational speed increased these ratings.

The sub-group differences shown for the HRTF quality-test (Tables 1 and 2) and the results of the main experiment (Table 3) can be explained either by the errors that occurred during the individualized HRTFs measurement procedure or by the subjects' auditory localization abilities. Localization performance varies between the individuals and terms “poor” or “bad” localizers are used in the literature, e.g. [9, 25].

In general, the results from the HRTF quality-test were influenced by several factors. First, participants were not trained to perform localization tasks in previous experiments and the quality-test procedure did not include a training period. Second, utilization of an ecological sound as a stimulus might bias the judgments of the participants. More work has to be done for designing fast and reliable procedures for evaluation of HRTFs catalogues' quality.

Results presented in Table 1 showed that the front-back confusion ratio was significantly increased for G-group subjects when using non-individualized HRTFs. However, no such difference was found for B-group. It is known that the performance of good localizers degrades when using bad localizers' HRTFs [9]. The opposite effect, when bad localizers improve their abilities using other person HRTFs catalogues, has not been fully evaluated [9]. Larger deviations in rated source heights found in the B-group for both individualized and generic HRTFs catalogues suggest the influence of individual performance. Table 3 presents the last evidence for difference in the sub-groups' performance: for intensity and convincingness ratings no discrimination between stimuli containing single or multiple sound sources was done by the B-group subjects.

When re-examining the data from the previous experiment [8] for the same participant sub-groups, the trends presented in Table 3 were not found. Therefore it is more likely that sub-group difference is due to the errors occurred during the HRTFs measurements procedure. However since different stimuli synthesis procedure was used in [8] a direct comparison of the subject responses is not possible and further studies of this finding is needed.

Preliminary tests with 2 stimuli with added clicks supported the hypothesis that adding distracters to the auditory VE might influence overall perception of the sound scene and decrease influence of artifacts caused by non-individualized HRTFs utilization. Specially designed experiment should shed further light on the effects of divided attention on quality judgments of VEs.

6. Conclusions and future work

There are several conclusions from this initial investigation. First, it was found that individualized HRTFs increase presence ratings. Second, the results were consistent with the previous results reported in [8], where the number of sound sources influenced both presence ratings and ego-motion experiences. Third, inter-group differences were found within the subjects, which were more likely caused by the errors occurred during the fast measurement of individualized HRTFs catalogues. Participants from the group with poor localization performance showed no discrimination in intensity and convincingness ratings for the number of presented sound sources. Finally, it is important to note that stimuli processed with generic (KEMAR) HRTFs also induced ego-motion regardless to the lowered rendering quality of spatial scene.

The authors are planning to test their findings in sub-group differences using a higher number of participants for reliable statistical analysis. Modification of existing methodology of both measuring and evaluating of individualized HRTFs catalogues is also planned for upcoming work. Influence of distracters on presence ratings and illusory ego-motion sensation is another topic for the follow up experiments.

Acknowledgements

This work was supported by the EU POEMS project (IST-2001-39223). The first author of this paper would like to thank for the support from Alfred Ots' scholarship foundation.

References

[1] M. Lombard, T. Ditton. At the heart of it all: the concept of presence. *Journal of Computer Mediated Communication*, 3(2), 1997.

[2] 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.

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

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

[5] 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.

[6] 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.

[7] EU-Project POEMS (Perceptually Oriented Ego-Motion Simulation) URL: http://www.poems-project.info/

[8] P. Larsson, D. Västfjäll, M. Kleiner. Perception of self-motion and presence in auditory virtual environments. Submitted to *Proceedings of Presence 2004*, Valencia, Spain. October 2004.

[9] D.R. Begault. 3D sound for virtual reality and multimedia. Academic Press. 1994.

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

[11] J. Blauert. Spatial hearing. MIT Press. Cambridge, rev. edition. 1997.

[12] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, vol 94, 111–123. 1993.

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

[14] J. Freeman, J. Lessiter. Hear there & everywhere: the effects of multi-channel audio on presence. In *Proceedings ICAD 2001*. July 2001

[15] E.H.A. Langendijk, D.J. Kistler, F.L. Wightman. Sound localization if the presence of one or two distracters. *J. Acoust. Soc. Am.*, vol. 109(5), 2123-2134. 2001.

[16] H. Møller. Fundamentals of binaural technology. *Applied Acoustics,* vol. 36, 121-218. 1992.

[17] D. Hammershøi, H. Møller. Methods for binaural recording and reproduction. *Acta Acustica united with Acustica,* 88 (3), 303-311. 2002.

[18] F.L. Wightman, D.J. Kistler, S.H. Foster, J. Abel. A comparison of head-related transfer functions measured deep in the ear canal and at the ear canal entrance. In *Abstracts of the 17th Midwinter Meeting of the Association for Research in Otolaryngology*, 71. 1995.

[19] S. Müller, P. Massarani. Transfer function measurement with sweeps. *J. Audio Eng. Soc.* vol. 49 (6), 443-471. 2001.

[20] A. Kulkarni, S.K. Isabelle, H.S. Colburn. Sensitivity of human subjects to head-related-transfer-function phase spectra. *J. Acoust. Soc. Am.*, vol. 105(5), 2821-2840. 1999.

[21] K. Hartung, J. Braasch, S. Sterbing. Comparison of different methods for the interpolation of head-related transfer functions. In *AES 16th Int. Conf. on Spatial Sound Reproduction*, 319-329. 1999.

[22] J. Breebart. Modeling binaural signal detection. PhD thesis. Technische Universiteit Eindhoven. 2001.

[23] Home Audiometer (Hearing test software) URL: http://www.esser.u-net.com/homeaudiometer.html

[24] T. Djelani, C. Pörschmann, J. Sahrhage, J. Blauert. An interactive virtual-environment generator for psychoacoustic research II: Collection of head-related impulse responses and evaluation of auditory localization. *ACUSTICA acta acustica*, vol. 86, 1046–1053. 2000.

[25] E.M. Wenzel, F.L. Wightman, D.J. Kistler, S.H. Foster. Acoustic origins of individual differences in sound localization behavior. *J. Acoust. Soc. Am*, vol. 84, S79. 1988.