



ASSESSMENT OF ROOM SIZE AND POSITION OF THE LISTENER BY NORMAL SIGHTED PERSONS BASED ON ACOUSTIC RESPONSE OF THE ROOM

Marko Horvat¹, Kristian Jambrošić¹, Juraj Francetić², Hrvoje Domitrović¹, Monika Rychtarikova³, Vojtech Chmelik⁴

¹ University of Zagreb, Faculty of Electrical Engineering and computing, Unska 3, 10 000 Zagreb, Croatia, Email: marko.horvat@fer.

² Brodarski Institute, Zagreb, Croatia

³ KU Leuven, Leuven, Belgium

⁴ Slovak University of Technology, Bratislava, Slovakia

ABSTRACT

As a part of a research conducted together with KU Leuven and Slovak Technical University in Bratislava, this paper investigates the possibility for a person with normal sight to estimate the size of an indoor space, i.e. a room, as well as to assess their position inside the room using only acoustical cues obtained from acoustical response of the room to an impulsive stimulus. Two experiments were designed, the first one dealing with room size estimation and the second one addressing the issue of egocentric localization. Both experiments utilize virtual environments defined in a room acoustics simulation software, in order to be able to change the acoustic treatment of the rooms in a well-defined manner, i.e. to use predefined values of the average absorption coefficient and to change the scattering coefficient of chosen room surfaces. The room size estimation experiment utilizes four rectangular rooms; a small one with a size of an average room in an apartment, two middle-sized rooms with the same volume, but radically different shape, and a large room, with the ratio of room volumes of 1:8:64. The egocentric localization is investigated on three positions in the middle-sized, ordinarily shaped room only. Two impulsive stimuli used in the experiments are hand claps as a stationary source produced by persons themselves, and footsteps as a moving source produced by another person, both chosen as good representations of real-life situations. The listeners were chosen from the student population and had various level of knowledge on acoustics, music production and architecture. They received no specific training before the tests, but were guided only with their experience. Similar experiments were conducted at KU Leuven, where binaural reproduction over headphones was applied. In these experiments spatial reproduction over loudspeakers was utilized using 2nd order Ambisonics as a platform for encoding the sound field in a virtual environment and later reproduction. The results of the experiments have undergone statistical analysis, in order to determine the extent to which the abilities of egocentric localization and room size assessment are developed for normal sighted persons.

INTRODUCTION

Auditory information represents a valuable supplement to the visual information obtained on the environment a person finds themselves in.

Moreover, in many cases the auditory information is crucial for decision-making process of purely primordial nature, for example, when a source of potential danger can be heard, but not seen. Although the quantity of received auditory

information is smaller than the one contained in visual information by several orders of magnitude, auditory information is often sufficient for transmitting all the necessary information. To quote a renowned expert in the field of acoustics: “Sound without picture is radio. Picture without sound is just video surveillance.”

Auditory information serves as a complement to visual information for a normal sighted person. For a blind or visually impaired person faced with the lack of visual information, however, auditory information becomes much more. The amount of information is the same for sighted and visually impaired, but the latter ones have the ability to process this information in a much more refined way, and to extract cues normally sighted person do not need and, therefore, ignore.

Many different research have been conducted on the difference between the ways normal and visually impaired perceive the environment around them, or even the difference between early- and late-onset blind people [Hull, 1990; Kujala, 1997; Lessard, 1998; Voss, 2008]. The phenomenon of echolocation was investigated as well [Gougoux, 2005, Herrensens, 2011].

Following the work of Chmelik [Chmelik, 2013], who has performed experiments on self-localization and room size assessment based on auditory cues obtained from virtual environments, the idea behind this paper was to recreate and extend his experiments by implementing them in a virtual acoustic environment recreated by a multi-channel loudspeaker system, being perhaps a situation that is more similar to real situations and environments.

EXPERIMENTAL SETUP

Two experiments were included in the investigation presented in this paper. In the first one, the listeners were given the task to estimate their position in a room based only on the response of the room to predefined sound stimuli. In the second one, the listeners were asked to try to differentiate between various room sizes, again using only the response of the room to an impulsive stimulus.

VIRTUAL ENVIRONMENTS AND STIMULI

The stimuli used in the experiments are taken from real-life situations and have an impulsive character.

A stationary virtual listener is represented with hand claps produced by the listener himself. A moving virtual listener or another virtual person moving through the room produces the sound of footsteps.

The conditions in virtual environments were modified by altering the values of absorption coefficient of all room surfaces except the floor, which was highly reflective in all cases. The values of absorption coefficient were set to 0.1, 0.2 and 0.4. In each of the three cases, the scattering properties of the surfaces were changed in three steps by altering the value of the scattering coefficient in the following way: initially, scattering of 0.05 was set on all surfaces; then it was set to 0.9 for the ceiling and left at 0.05 on the remaining surfaces; finally, it was set to 0.9 for one of the side walls and left at 0.05 on other surfaces.

Virtual environments were defined and auralized using ODEON room acoustics modelling software. The resulting impulse responses were convolved with source sound files that contained the sounds of sound claps and footsteps and then encoded in 2nd order Ambisonics. Decoding and reproduction were done using Reaper audio software package. To remove the influence of loudness changes due to different room volumes and acoustic treatment, all sound files intended for reproduction were RMS-normalized to the same level, and the reproduction level was kept constant during both experiments.

TEST SUBJECTS

The test subjects that took part in the listening experiments were recruited from student population, thirty six listeners altogether. The age range of the listeners spanned from 21 to 28, and no hearing impairment or damage was reported. According to their own statements, approximately two thirds of the listeners have come in contact with acoustical principles at various knowledge levels, through interest in music and/or courses on acoustics taken at the faculty. Approximately ten percent of the listeners claimed to have extensive knowledge on acoustics and/or musical education.

TESTING PROTOCOL

The listening tests took place at the University of Zagreb, specifically, in the Auralization Laboratory, in which a multichannel loudspeaker system in a quasi-spherical 4-8-4 configuration has been

installed. Such a configuration allows for reproduction of spatial audio recordings and can be used for reproduction of 2nd order 3D Ambisonics recordings. Limited to horizontal plane only, the system offers the possibility to reproduce up to 3rd order 2D Ambisonics recordings.

Due to the fairly large quantity of test material that needed to be evaluated, the tests have been carried out in two stages, with 18 evaluations in each stage. Instructions were given to the listeners on their sitting position and orientation, and the importance of maintaining both was emphasized, so that the listeners would remain within the sweet spot of the system and that their orientation would match the one defined for the virtual listener. To direct the attention of the listeners to the task at hand, rather than keeping it on the environment the experiment takes place in, the interior of the Laboratory is decorated with neutral colours, and the lighting is kept discrete, but still sufficient for the listeners to see the test form and the images of the rooms provided for them as the supplement to oral and written explanations.

In the self-localization experiment the task for the listeners was to listen to recordings on all three positions (A, B and C) for a given acoustic treatment of the room (nine cases in total) and to assign the positions in the room to the recordings they heard, thus yielding one of six possible permutations (e.g. BAC, CBA, etc.). The task put before the listeners in the room size assessment experiment was to listen to recordings in four different rooms for a given acoustic treatment, and to assign room numbers to these recordings, obtaining one of twenty four possible permutations (e.g. 3124, 4132, etc.).

To familiarize the listeners with their task, an example sequence of sounds was reproduced to the listeners, and the correct solution was given as well, so that the listeners could grasp what was asked of them. Upon completing the example sequence, the listening experiment continued with evaluations for each of the nine different acoustic treatments and with two different sound stimuli.

SELF-LOCALIZATION

The virtual room defined for the self-localization experiment was a shoe-box type room 12 m long, 7 m wide and 3 m high, with a volume of 252 m³.

In the first part of the experiment the virtual listener was stationary and occupied one of the three predefined positions in the room, as shown in Figure 1, and virtually clapped his hands to generate the sound stimulus consisting of six consecutive claps. As an average, the height of the virtual listener's ears was set to 1.7 m, and his hands as the source of sound were set to a height of 1.5 m and 0.7 m in front of the listener's position. As mentioned before, scattering coefficient was changed on one of the walls, in this case the left side wall.

The reverberation times in the room are shown in Figure 2 as a function of frequency, with the absorption coefficient applied to all surfaces except the floor as parameter.

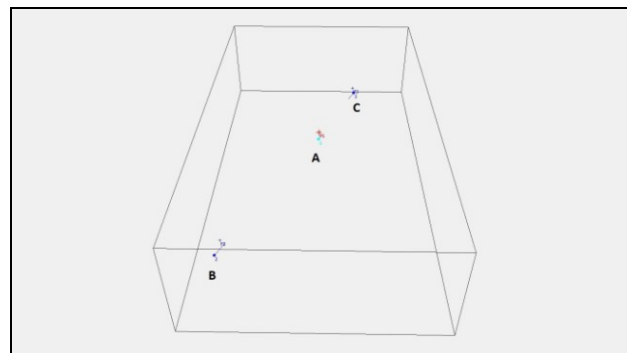


Figure 1. Room 1: medium-size, 12 m x 7 m x 3 m, volume 252 m³; stationary virtual listener at A = central position, B = corner position, C = frontal position; his own hands as the source of sound

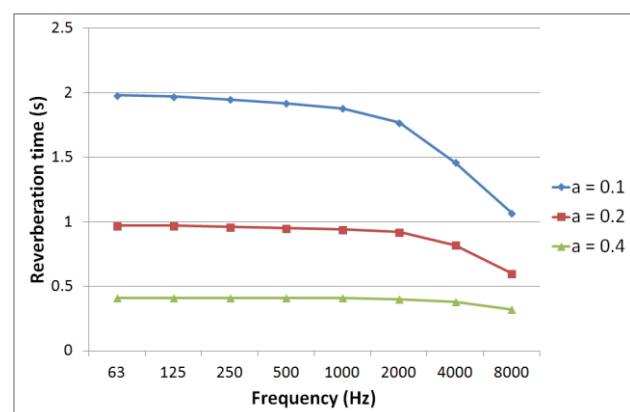


Figure 2. Reverberation time in room 1

In the second part of the self-localization test the stimulus was changed to the footsteps of a virtual walker moving along the length of the room, as displayed in Figure 3. The length of a step was 0.6 m and the height of the sound source was set to 0.1 m. The listener remained stationary and his orientation

was kept towards the middle of the room, not following the movement of the virtual walker. The listening positions were changed to avoid the learning effect. The wall with adjustable scattering is again the left side wall (the farthest wall in Figure 3).

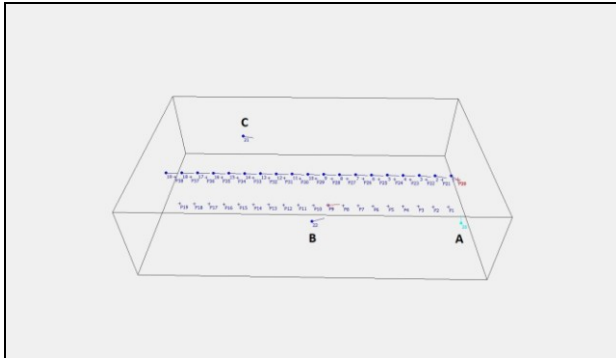


Figure 3. Room 1: stationary virtual listener at A = corner position, B = central right lateral position, C = back left lateral position; footsteps of a virtual walker as the source of sound

ROOM SIZE ASSESSMENT

In the first part of the room size assessment experiment the listener was stationary and occupied the central position in the room, with his orientation aligned along the largest dimension of the room. The source of sound is again a series of sound claps produced by the virtual listener himself. In the second part the listener is moving along the central axis of the room stretched along its length, producing 11 steps altogether. The sound of these footsteps is used as the stimuli required for the experiment.

Four virtual rooms were defined for the experiment on room size assessment. Room 1 was taken as is from the experiment on self-localization. Room 2 has the same volume as room 1, i.e. 252 m^3 , but its shape was changed to a long hallway with dimensions $35 \text{ m} \times 2.4 \text{ m} \times 3 \text{ m}$. The appearance of room 2 is shown in Figure 4, and the reverberation times in Figure 5.

Room 3 is the smallest room with dimensions $4 \text{ m} \times 3 \text{ m} \times 2.5 \text{ m}$, yielding the volume of 30 m^3 , as shown in Figure 6. The reverberation times in room 3 are shown in Figure 7.

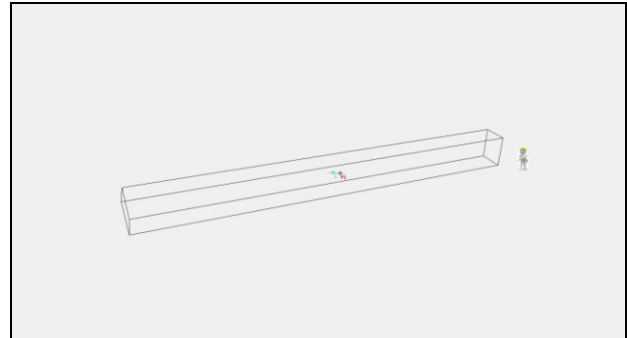


Figure 4. Room 2: hallway, dimensions $35 \text{ m} \times 2.4 \text{ m} \times 3 \text{ m}$, volume 252 m^3

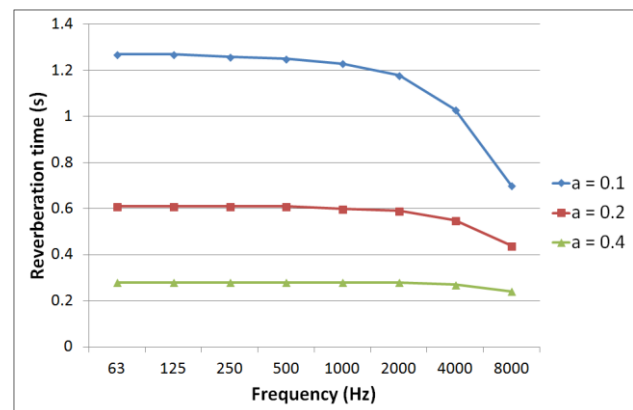


Figure 5. Reverberation time in room 2

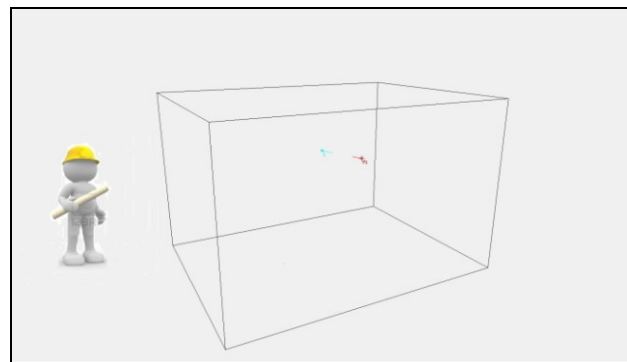


Figure 6. Room 3: small room, dimensions $4 \text{ m} \times 3 \text{ m} \times 2.5 \text{ m}$, volume 30 m^3

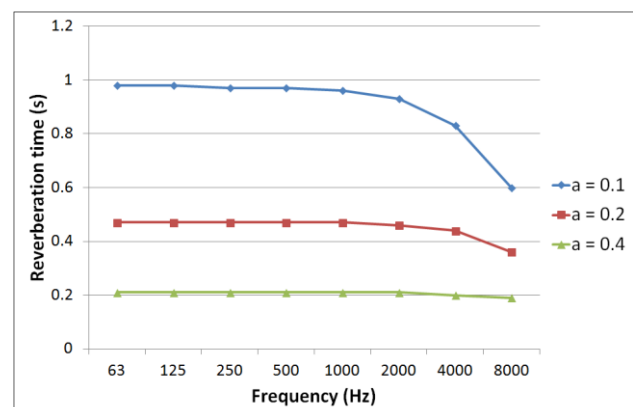


Figure 7. Reverberation time in room 3

Finally, room 4 is the largest room obtained by doubling all dimensions of room 1, which gives room dimensions of 24 m x 14 m x 6 m and the volume of 2016 m³. The room is shown in Figure 8, and the corresponding reverberation times in Figure 9.

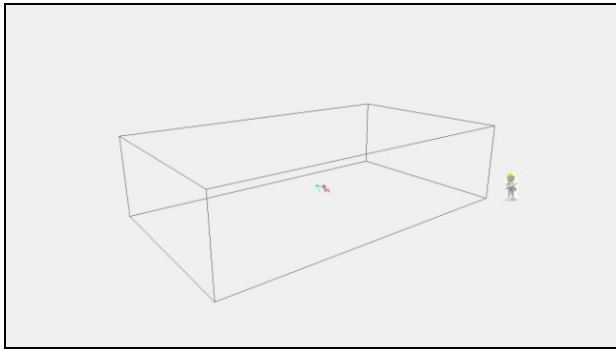


Figure 8. Room 4: large room, dimensions 24 m x 14 m x 6 m, volume 2016 m³

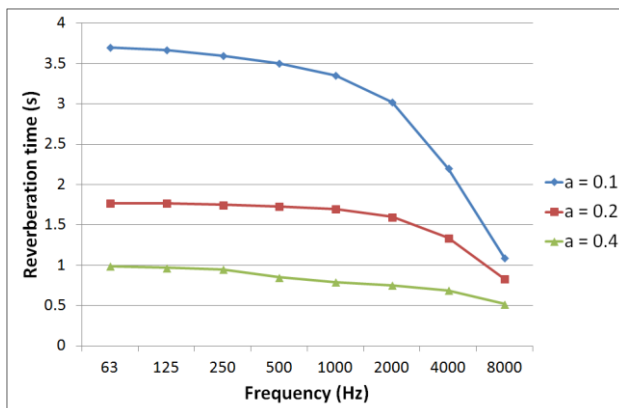


Figure 9. Reverberation time in room 4

RESULTS AND DISCUSSION

The results of the listening experiments were analyzed using the χ^2 -test. The null hypothesis was established that there is no way to differentiate between the positions in the room or between various room sizes based only on acoustical cues. Translated to the self-localization experiment, all six permutations obtainable from position markers A, B and C (e.g. ABC, ACB, BAC,...) are equally probable. Analogously for the room size assessment experiment, all twenty four permutations obtainable from room markers 1, 2, 3 and 4 (e.g. 1234, 2134, 4321, 3412, ...) are equally probable as well. The summarized results of this analysis are shown in Tables 1 and 2.

Table 1. The results of the statistical analysis for the self-localization test

Hand claps		Scattering coefficient ()		
$df = 5$		all 0.05	ceiling 0.9	wall 0.9
Absorption coefficient ()	0.1	$\chi^2 = 23.18$ $p < 0.001$	$\chi^2 = 6.94$ $p = 0.225$	$\chi^2 = 4.12$ $p = 0.533$
	0.2	$\chi^2 = 0.94$ $p = 0.967$	$\chi^2 = 3.41$ $p = 0.637$	$\chi^2 = 4.12$ $p = 0.533$
	0.4	$\chi^2 = 7.65$ $p = 0.177$	$\chi^2 = 2.35$ $p = 0.798$	$\chi^2 = 0.24$ $p = 0.999$
Footsteps		Scattering coefficient ()		
$df = 5$		all 0.05	ceiling 0.9	wall 0.9
Absorption coefficient ()	0.1	$\chi^2 = 10.18$ $p = 0.070$	$\chi^2 = 4.88$ $p = 0.430$	$\chi^2 = 0.65$ $p = 0.986$
	0.2	$\chi^2 = 20.41$ $p = 0.001$	$\chi^2 = 8.06$ $p = 0.153$	$\chi^2 = 10.53$ $p = 0.062$
	0.4	$\chi^2 = 2.41$ $p = 0.790$	$\chi^2 = 3.82$ $p = 0.575$	$\chi^2 = 1.00$ $p = 0.963$

Table 2. The results of the statistical analysis for the room size assessment test

Hand claps		Scattering coefficient ()		
$df = 23$		all 0.05	ceiling 0.9	wall 0.9
Absorption coefficient ()	0.1	$\chi^2 = 197.33$ $p < 0.001$	$\chi^2 = 228.00$ $p < 0.001$	$\chi^2 = 162.67$ $p < 0.001$
	0.2	$\chi^2 = 32.00$ $p = 0.100$	$\chi^2 = 98.67$ $p < 0.001$	$\chi^2 = 209.33$ $p < 0.001$
	0.4	$\chi^2 = 120.00$ $p < 0.001$	$\chi^2 = 102.67$ $p < 0.001$	$\chi^2 = 73.33$ $p < 0.001$
Footsteps		Scattering coefficient ()		
$df = 23$		all 0.05	ceiling 0.9	wall 0.9
Absorption coefficient ()	0.1	$\chi^2 = 119.29$ $p < 0.001$	$\chi^2 = 163.17$ $p < 0.001$	$\chi^2 = 179.63$ $p < 0.001$
	0.2	$\chi^2 = 142.60$ $p < 0.001$	$\chi^2 = 141.23$ $p < 0.001$	$\chi^2 = 130.26$ $p < 0.001$
	0.4	$\chi^2 = 116.54$ $p < 0.001$	$\chi^2 = 89.11$ $p < 0.001$	$\chi^2 = 138.49$ $p < 0.001$

The χ^2 -test is designed to show the overall differences between an expected and observed data, without revealing more precise information. The results of the tests applied to the data obtained from these experiments reveal only that the listeners are capable (or not capable) of detecting certain

differences between positions in a room, or some differences between rooms regarding their size. However, the results of the statistical analysis alone cannot give an answer to whether these differences were perceived correctly or not. Furthermore, the results of the room size assessment experiment do not meet all the requirements set for using the χ^2 -test in the sense that the ratio of the number of listeners to the number of categories is too low. For both reasons, the results of statistical tests are complemented with additional data given in graphic form in Figures 10 to 13.

The figures show the distribution of the percentage of answers given by the listeners in listening experiments for each acoustical treatment. The correct sequences of positions in the self-localization experiment, and the correct sequences of rooms in the room size assessment experiment are marked with red. Incorrect sequences are marked in black.

The results of the self-localization test and the room size assessment test show a fundamental difference, which is reflected in the results of the χ^2 -tests as well. As described before, in the self-localization test three positions were defined in the room. Consequently, the listeners could have given $3! = 6$ different sequences of positions as their answer for a given situation, after they had listened to a sequence of three recordings. In the opinion of the listeners, all six sequences were viable candidates for the correct solution, as can be seen on the charts in Figures 10 and 11. In most cases the distribution of the percentages of answers does not deviate much from the expected null-hypothesis value of 16.67 %. This result implies that the listeners have given their answers randomly, being unable to hear the difference between the positions in the room. In most cases the correct sequence has not even been chosen more frequently than other sequences. The results of the room size assessment experiment offer a different picture. The experiment was more complex, as there were four rooms and $4! = 24$ possible ways to put them in order after listening to four recordings. As shown in Figures 12 and 13, not all 24 possibilities have been used. On average, their number was reduced to about one half, suggesting that there were audible differences between rooms. Hand claps as the sound source and low absorption in the room yield the rate of correct assessment of

room size of about 40-50 percent, which is then reduced as the amount of absorption in the room increases. When footsteps are used as the source of sound stimulus, the correct room size assessment stands fairly stable at a rate of about 25-30 percent, independent on the amount of absorption. In many cases the listeners singled out two or sometimes three sequences as correct ones, with one of them indeed being the correct sequence. In such cases the other sequence (or sequences) is different from the correct one only in the sense that two rooms have been swapped. As expected, room 1 as the middle-sized room was confused most often with other rooms. It has the same volume as room 2 (the hallway), and its volume puts it between rooms 3 and 4 (small and large).

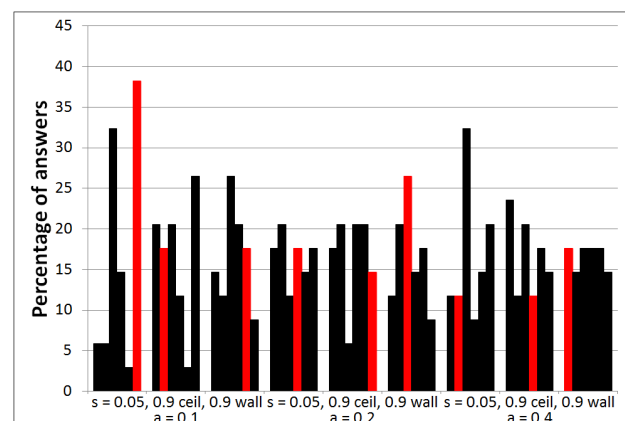


Figure 10. The self-localization experiment – hand claps as the source – expected value 16.67 %

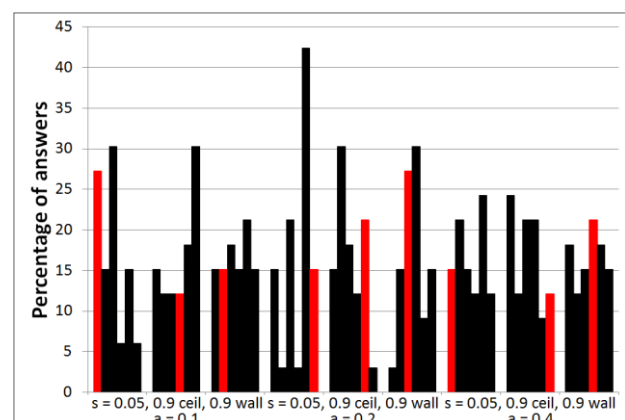


Figure 11. The self-localization experiment – footsteps as the source – expected value 16.67 %

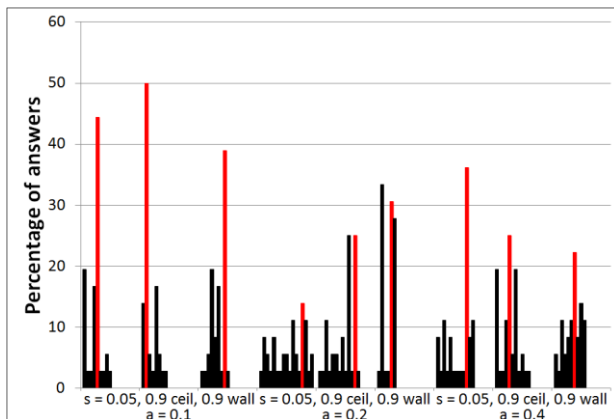


Figure 12. The room size assessment experiment – hand claps as the source – expected value 4.16 %

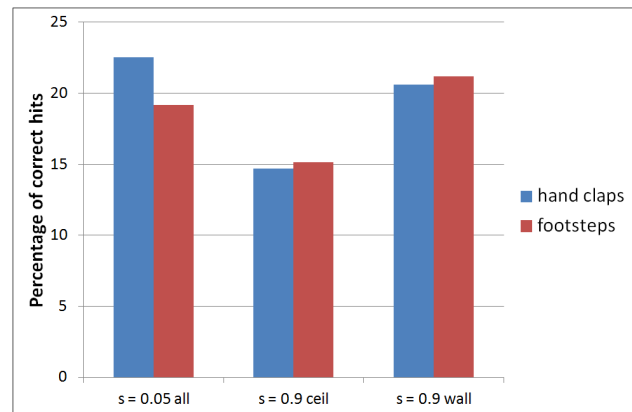


Figure 15. Self-localization: percentage of correct hits for different diffusion conditions

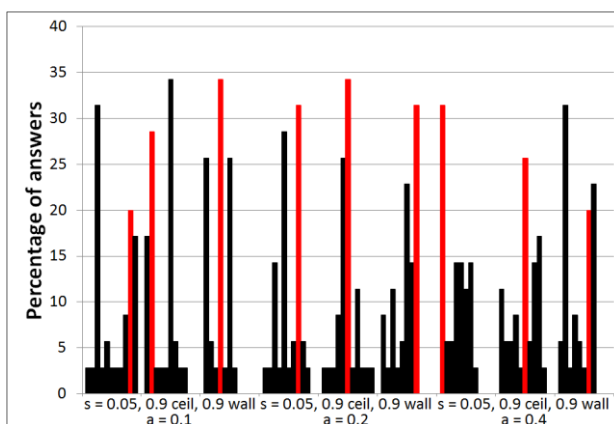


Figure 13. The room size assessment experiment – footsteps as the source – expected value 4.16 %

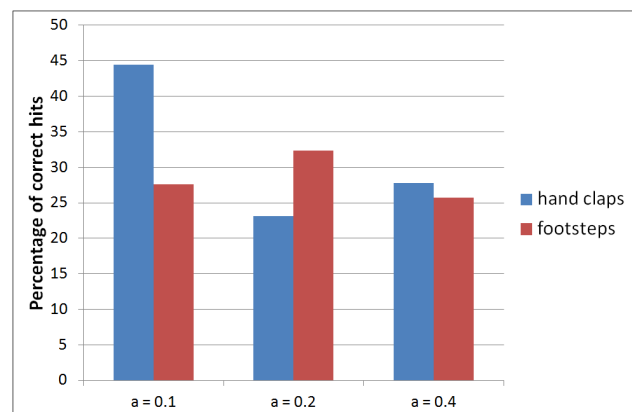


Figure 16. Room size assessment: percentage of correct hits for different absorption conditions

To examine the influence of acoustic treatment implemented in virtual rooms on the results of the listening experiments, the results are presented cumulatively on charts in Figures 14 to 17 as total percentage of correct hits for the given absorption and diffusion conditions.

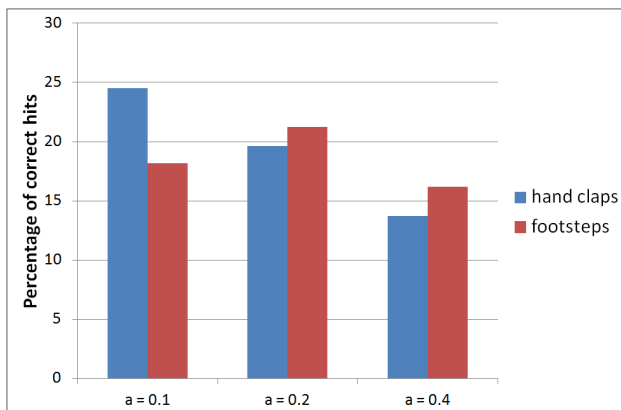


Figure 14. Self-localization: percentage of correct hits for different absorption conditions

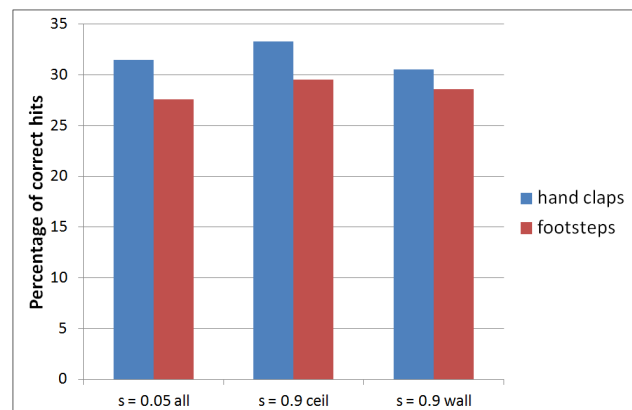


Figure 17. Room size assessment: percentage of correct hits for different diffusion conditions

In general, higher values of absorption decrease performance in both self-localization and room size assessment. Low values of absorption, in this case 0.1, contribute to the success of room size estimates, due to larger and more easily perceivable differences in reverberation times of individual rooms,

especially if combined with hand claps as the source of sound.

Diffusion properties of room surfaces and the way diffusion is implemented in the room do not seem to be important for the perception of room size, whereas diffusive ceilings seem to reduce the success rate in self-localization.

CONCLUSIONS

The results of the experiments described above show that the ability of self-localization in a room based on acoustical cues is not very well developed, while the performance in room size assessment is considerably better. In our opinion, the reason for this is the fact that over time people generally gain some experience in relating the acoustical response of room with its size merely by using different spaces, public or other, in everyday life. On the other hand, self-localization in a room based on sound is not crucial for normally sighted persons, such as the ones that took part in these experiments, because they use their sight to do that. On the other hand, it is expected that visually impaired people

will have a better developed ability of sound-based self-localization, simply by being forced to rely on sound rather than eyesight, which remains a topic for verification in further research.

The questions that still remain are the influence of the laboratory space itself on the listeners, i.e. whether they are able to imagine themselves in a virtual environment created by sound or not. Additionally, in a real situation, hand claps are produced by people themselves and thus know exactly when this sound will be heard, whereas in these experiments hand claps and the response of the room to them were reproduced by the loudspeaker system, which is not a real situation. Perhaps a way to improve the conditions of the experiment would be to trigger the sound reproduction system with a real hand clap by using gloves with motion or pressure sensors.

ACKNOWLEDGEMENT

This work has been supported by the European Community Seventh Framework Programme under grant No. 285939 (ACROSS).

REFERENCES

- [1] J. M. Hull: On Sight and Insight. A Journey into the World of Blindness. Oxford SPCK, 1990.
- [2] T. Kujala, K. Alho, M. Huottilainen: Electrophysiological evidence for cross-modal plasticity in humans with early- and late-onset blindness. *Psychophysiology* 34 (1997) 213-216.
- [3] N. Lessard, M. Pare, F. Lepore: Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395 (1998) 278-280.
- [4] P. Voss, F. Gougoux, R. J. Zatorre: Differential occipital responses in early- and late-blind individuals during a sound-source discrimination task. *Neuroimage* 40 (2008) 746-758.
- [5] F. Gougoux, P. Belin, P. Voss: Voice perception in blind persons: A functional magnetic resonance imaging study. *Neuropsychologia* 47 (2009) 2967-2974.
- [6] J. Herssens, L. Roelants, M. Rychtarikova, M. Heylighen: Listening in the absence of sight: The sound of inclusive design. Include 2011, London, 2011.
- [7] V. Chmelik: Principles of inclusive design in architecture and room acoustics. doctoral thesis, Slovak university of technology, Bratislava, 2013.