The Matter of Attention and Motivation – Understanding Unexpected Results from Auditory Localization Training using Augmented Reality

Song Hui Chon*
University of Colorado Boulder

Sungyoung Kim[†]
Rochester Institute of Technology

ABSTRACT

We present the results from a seven-week auditory localization training using Microsoft HoloLens. Eight participants were divided into two groups. Both groups showed a generally declining pattern in localization performance over the eight tests, unlike the results from our previous study. The decreasing slope was smaller for the train group than for the control group, which might reflect some mild effect of training. There was a one-time performance improvement after two trainings, which was not observed from subsequent tests. The training program might have been too simple to maintain participants' attention for weeks. Possible extraneous factors such as the academic calendar are discussed that might have had an impact on this decreasing pattern against the hypotheses.

Keywords: Localization, Training effect, Augmented reality, HRTFs.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality; Applied Computing—Education—Interactive learning environments

1 INTRODUCTION

Auditory localization is essential in everyday life to determine the direction and position of a sound event [2]. This process is highly individual [11][15], which utilizes cues included in the head-related transfer functions (HRTFs) [14]. Many studies in the field therefore have investigated effective measurement and estimation of individualized HRTFs for precise localization [1][8][9][10][16]. And these individualized HRTFs are in general costly in terms of measurement efforts and equipment required.

While some particular tasks might require highly individualized HRTFs to achieve a specific level of localization performance, there are also reports of generalized HRTFs providing effective means for auditory localization training [6][7][13]. Particularly, Ohuchi and colleagues [12] developed an auditory localization training game for the visually impaired using generalized HRTFs on a virtual auditory display. They had ten participants in two groups (train and control), and the train group showed improvements after a 10-day training of 15 minutes each.

Based on existing literature, a question rises whether generalized HRTFs can be still and possibly more effective for auditory localization improvement, especially when the training is coupled with augmented reality (AR) technology. In particular, AR allows a trainee to map congruent visual and auditory cues and form isomorphic mapping between two modes. To answer this question, we developed an auditory localization training program using Microsoft HoloLens, which is an AR head-mounted display (HMD) device. An AR HMD is an attractive choice for training programs as it facilitates multimodal integration of visual, auditory, and kinaesthetic information, which will likely bring a more successful outcome. The HoloLens utilizes Microsoft's own generalized HRTFs, which we used for our training program.

We carried out an earlier study [3]. The training was on the horizontal plane only for the first two weeks, and the vertical displacements were incorporated in the last two weeks. The same test was conducted three times during the four weeks—once before any training began, once after the first two weeks of training, and once after the entire four weeks of training. The test trials included the horizontal only condition as well as the vertical displacement condition. All participants improved their localization accuracy over four weeks. In fact, the improvement was made in the first two weeks (especially in the horizontal only conditions) to what seems to be the ceiling level, which remained constant for the next two weeks.

We then conducted a forth test with the identical setup ten weeks after their last training [4]. All five participants showed that they retained the training effect even after a 10-week break. Based on these findings, we carried out a follow-up study with two groups. In the new experiment our goals were 1) to examine whether the train group could replicate the earlier results, and 2) to compare the train group's performance with the control group's.

2 LOCALIZATION TRAINING PROGRAM

We used the same training program that we used for earlier studies [3][4]. The only difference is that in earlier studies the vertical displacements could include a negative angle (which means a sound source placed at a lower height than the participant's ear level). This seems to lead to more confusion than training; therefore, we decided to restrict the vertical displacements to positive angles for the current study. Our localization training program utilizes Microsoft's built-in generalized HRTFs and spatializer on a HoloLens. An auditory target was placed at a randomly selected spot on the circumference of a six-meter radius centered at a participant. There were 16 predetermined positions uniformly distributed on the circumference on the horizontal plane (i.e., the participant's ear level). One of these predetermined spots would be randomly chosen for a trial. Once a trial was over, this particular point was removed from the selection pool so that it would not be tested again in the same module. If a trial included a distractor, its location was randomly determined in the exactly same way with the target's location. Therefore, it was possible to have both target and distractor placed on the same spot.

Table 1 summarizes the detailed composition of the 18 training modules. These modules were designed using cross-combination of three factors: three vertical displacement levels (horizontal only; 0 to 22.5 degrees upward; 0 to 45 degrees upward), two loudness level (loud and soft), and three distractor conditions

*e-mail: songhui.chon@colorado.edu
[†]e-mail: sxkiee@rit.edu

(none, one, or two). The six modules in Type 1 involved no vertical displacements, which means all sound objects were randomly placed on a horizontal plane. These modules were used for training for the first two weeks. The modules in Type 2 had both random horizontal positions as well as vertical displacements in the range of 0 to 22.5 degrees (i.e., upward) from the participant's ear level. This group of modules was used for training for the next two weeks. The last two weeks of training was performed with the six modules in Type 3. These modules were very similar to those in Type 2; one difference was that the vertical displacements were randomly chosen out of the range between 0 to 45 degrees for Type 3. Each module has eight trials.

Half of the modules were presented at the full loudness level of the HoloLens (the "Loud" conditions), whereas the other half used the low volume setting (the "Soft" condition; 6 dB lower than the "Loud" condition).

TABLE 1. THE COMPOSITION OF THE 18 TRAINING MODULES

Type	Module	Displacement	Loudness	Sounds
1	1-1A	Horizontal only	Loud	Target only
	1-2A			Target + 1 Distractor
	1-3A			Target + 2 Distractor
	1-1B		Soft	Target only
	1-2B			Target + 1 Distractor
	1-3B			Target + 2 Distractor
2	2-1A	Horizontal plus 0 – 22.5 degrees vertical displacement	Loud	Target only
	2-2A			Target + 1 Distractor
	2-3A			Target + 2 Distractor
	2-1B		Soft	Target only
	2-2B			Target + 1 Distractor
	2-3B			Target + 2 Distractor
3	3-1A	Horizontal plus 0 – 45 degrees vertical displacement	Loud	Target only
	3-2A			Target + 1 Distractor
	3-3A			Target + 2 Distractor
	3-1B		Soft	Target only
	3-2B			Target + 1 Distractor
	3-3B			Target + 2 Distractor

There was a blue cross in the middle of the participant's visual field as a reference to the center of his/her gaze. This is the only virtual visual cue during a training trial. Only after a participant indicated the estimated position of the target with the selection gesture while keeping the blue cross fixed on the estimated spot, the program would provide visual feedback—the target's actual position and the user's estimated position, as well as the localization score based on the distance of these two points. These visual information facilitated the participant in (1) developing an isomorphic mapping between auditory and visual positions, and (2) improving their localization performance based on the distance between the actual and the estimated points. Participants can adaptively learn through the isomorphic feedback, which is essential for long-term memory development of the trained skill [5].

2.1 Localization Score

A localization score was calculated and recorded after each trial based on the distance of the two points d according to the following equation, $score = 1 - d/8$, where 8 meters is slightly smaller than the distance between two points forming a 90-degree angle (which is about 8.5 m). Therefore, the score of zero was assigned if the distance between the two points exceeded 8 meters. This score was displayed along with the locations of the target (and distractor(s)) to the participant during training. The localization score was not shown to the participant during test, but was still recorded in data files.

2.2 Sound Objects

The target and distractor sounds are always the same mono signals: the target is a female singing improvisation (without any linguistic information), and the distractor is a piano accompaniment of the singing. We decided to use musical stimuli that have more harmonic energy in lower frequencies and less energy in higher frequencies. High-frequency energy is known to be important in localization, so we purposefully chose our stimuli to make the task more difficult. All sounds are looped so that participants could take as long as they wanted. This helped participants get used to the task at a gradual and individually adequate pace.

3 Seven-Week Study

An individualized and self-paced study was conducted for seven weeks in a quiet laboratory on Rochester Institute of Technology campus. Ten undergraduate students were recruited from an advanced audio engineering course. All reported normal hearing with no known hearing problems. After the initial test (Test 0), the participants were divided into two groups of the similar mean and standard deviation. Two students did not finish the study; therefore, the analyses were carried out on the results from the remaining eight participants.

Participants came in once a week. In each session, the train group performed training before completing test (except for Tests 0 and 7), whereas the control groups completed only the test modules. Seven weekly trainings and eight tests were performed.

3.1 Test Program

A test program was developed with six training modules, using cross-combination of two factors: two loudness level (loud and soft) and three distractor conditions (none, one, or two). Table 2 summarizes the composition of the six test modules. The modules were presented in a randomized order for each session. Each module had eight trials. For each trial, the target (and distractor(s) for some modules) would be randomly placed on the horizontal plane along the circumference of a six-meter-radius circle centered at the participant.

TABLE 2. THE COMPOSITION OF THE SIX TEST MODULES

Module	Loudness	Sounds
1	Loud	Target only
2		Target + 1 Distractor
3		Target + 2 Distractor
4	Soft	Target only
5		Target + 1 Distractor
6		Target + 2 Distractor

This test program was different from our earlier studies [3][4]. The earlier version had two modules with vertical displacements, which we decided not to include. Instead, we added two modules with two distractors. Each test module had eight trials.

For each test trial, no visual feedback was given to the participant regarding their localization performance. This was an intentional design to prevent test modules from giving inadvertent training to participants. It also means that the control group never received feedback during the entire seven weeks of study on how well or poorly they performed.

4 Results

A three-way mixed analysis of variance was performed on each participant's average localization score as dependent variable. Group was the between-subject independent variable (IV) which did not show any significant difference, $F(1, 7) = 0.865$, $p = .388$. Test and Module were repeated IVs. The only significant difference was found with the main effect of Test, $F(7, 42) = 2.778$, $p < .05$. However, this significance simply reflects three clusters emerging from the average localization score in Tests {0, 2, 4}, {1, 3}, and {5, 6, 7}, with significant intercluster differences.

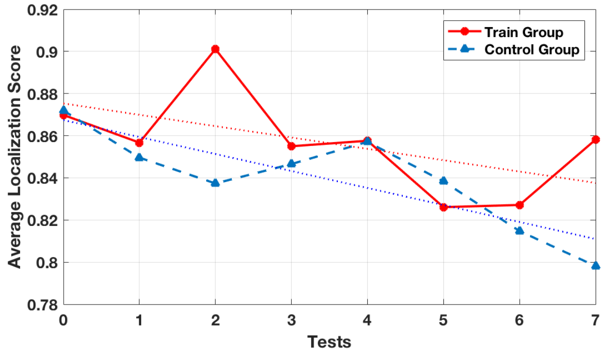

Figure 1: Average localization score for each test for two groups.

Since there was no significant effect of module, the average localization score was calculated for each test across all modules. Figure 1 presents the average localization score for each test for each group. The dotted lines show decreasing linear regression slopes for the two groups, which are different from the improvement pattern in our previous study. A paired-sample t-test indicated that the train group's localization performance between Test 0 (*M*= 86.98, *SD* = 9.37) and Test 7 (*M* = 85.81, *SD* = 9.72) did not change, $t(191) = 1.28$, $p < .20$. In contrast, the control group's Test 7 (*M* = 79.81, *SD* = 10.09) was significantly worse than Test 0 (*M* = 87.19, *SD* = 6.62), $t(191) = 7.79$, $p < .001$. This unexpected decrease in performance could be explained at least in part considering the academic calendar. We started the study already halfway into the fall semester and finished the week before the final exam week. Students gradually became busier and more tired during the entire seven weeks of study, which probably affected their baseline attention level. This decreased attention likely in turn influenced their performance. Note that Test 7 score for the test group is significantly higher than for the control group, $t(191) = 5.78$, $p < .001$. This could suggest a positive training effect, making the decreasing slope smaller for the train group.

The best performance in the train group was seen in Test 2 after two training sessions (*M* = 90.11, *SD* = 9.78), which was a statistically significant improvement over Test 0, t(191) = 3.10, *p* < .005. In our previous study, the average performance improved after two weeks (or four sessions) of training. Perhaps the maximal improvement could be achieved after two trainings, although it failed to retain in the current study, in contrast to our earlier findings [4] where all participants kept the improved training effect.

5 Discussion

In this paper we presented a follow-up study of auditory localization training using AR. Unfortunately we did not find the positive effect of training, which was observed from our previous studies. This discrepancy could reflect the circumstantial differences: the first study was carried out during summer when students had more free time and much less stress than in the fall semester. Furthermore, the current study was conducted during the second half of a semester, when students would steadily become busier and more stressed. This different circumstance might help explain the downward trend found in both groups' weekly test results.

Another big difference was monetary compensation for the participants. The five participants in our earlier studies received remuneration for the time they spent on the study, whereas the participants for the current study were not; their participation was mandatory as a part of the audio engineering course. This could have had a big impact on their motivation for improvement. Additionally, the train group's localization ability might have improved, but the test might have been unable to measure it because the test might have been too simple to be repeated for seven weeks, especially without any modules with vertical displacements. The fact that the train group showed a significant performance improvement after two weeks could suggest that the peak performance improvement could be attained after two weeks of training, even though this temporary positive training effect did not last. Although we did not find the hypothesized training effects, the current study still illuminates some important points in experimental design that were previously not considered, such as the inadvertent effect of the academic calendar especially in the absence of remuneration. Without a sufficient motivation, students would not want to spend their attention on a task where the outcome does not have a return value to them. Based on these findings, we will conduct another study in the future with improved experimental design that can stimulate and maintain participants' attention and motivation over the entire duration of training.

Acknowledgements

This study was supported by Yamaha Corporation.

References

[1] D. R. Begault, E. M. Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, 49(10): 904–916, 2001.

[2] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[3] S. H. Chon and S. Kim. Auditory localization training using generalized head related transfer functions in augmented reality, *Journal of Acoustical Science and Technology*, 39: 312–315, 2018.

[4] S. H. Chon and S. Kim. Long term effects of auditory localization improvement – a pilot study, In *Proceedings of the 15th International Conference of Music Perception and Cognition (ICMPC), Graz, Austria, July 23–28, 2018*.

[5] J. Corey. *Audio Production and Critical Listening*, 2nd ed. Focal Press, 2016.

[6] P. M. Hofman, J. G. A. Van Riswick, and A. J. Van Opstal, Relearning sound localization with new ears. *Nature Neuroscience*, 1: 417–421, 1998.

[7] A. Honda, H. Shibata, J. Gyoba, K, Saitou, Y, Iwaya, and Y. Suzuki, Transfer Effects on Sound Localization Performances from Playing a Virtual Three-Dimensional Auditory Game. *Applied Acoustics*, 68: 885–896, 2007.

[8] Y. Iwaya. Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears. *Acoustical science and technology*, 27(6): 340–343, 2006.

[9] C. Jin, P. Leong, J. Leung, A. Corderoy, and S. Carlile. Enabling individualized virtual auditory space using morphological

measurements. In *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing)*, pp. 235–238, 2000.

[10] B. F. G. Katz and G. Parseihian. Perceptually based head-related transfer function database optimization. *Journal of the Acoustical Society of America*, 131(2): EL99–EL105, 2012.

[11] B. C. J. Moore. *An introduction to the psychology of hearing*. Academic Press, 2012.

[12] M. Ohuchi, Y. Iwaya, and T. Munekata. Training effect of a virtual auditory game on sound localization ability of the visually impaired. In *Proceedings of the 11th Meeting of the International Conference on Auditory Display*, 2005.

[13] E. M. Wendzel, M. Arruda, D. J. Kistler, and F. L. Wightman, Localization using non-individualized head-related transfer functions. *Journal of Acoustical Society of America*, 94 (1): 111–123, 1993.

[14] B. Xie, *Head-related transfer function and virtual auditory display*. J. Ross Publishing, 2013.

[15] X.-L. Zhong and B.-S. Xie. Head-Related Transfer Functions and Virtual Auditory Display. *Soundscape Semiotics - Localization and Categorization*, Dr. Hervé Glotin (Ed.), ISBN: 978-953-51-1226-6, InTech, 2014. DOI: 10.5772/56907.

[16] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis. HRTF personalization using anthropometric measurements. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop*, pp. 157–160. IEEE, 2003.