PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

3D localization of speech by mildly and moderately hearing-impaired persons in ecological environments

Laurent S. R. SIMON[1,2], Andrea KEGEL[1], Hannes WÜTHRICH[3], Norbert DILLIER[1,2]

[1]ENT department, University Hospital, Zürich, Switzerland
[2]University of Zürich, Switzerland
[3]Sonova AG, Stäfa, Switzerland

Abstract
Studies about 3D audio perception with hearing devices are sparse. There are a number of studies on localization on the horizontal plane, but none about localization in elevation. In this study, 20 hearing-impaired subjects took part in a 3D localization test. Participants had to point their head in the direction of various target stimuli. These stimuli were generated by convolving a speech signal with 4^{th} order HOA room impulse responses recorded with an Eigenmike. 4^{th} order HOA background audio recordings were played in addition to the target in order to reproduce realistically difficult environments. In total, 32 different directions were generated for both a noisy restaurant and a busy street background sounds. Three different hearing device beamformers were compared along with an unaided condition.
Keywords: Hearing devices, 3D audio, Localization

1 INTRODUCTION

Studies on 3D audio localization with hearing aids are scarce. [1] looked at the localization performance of hearing-impaired people and normal-hearing people for 40 randomly generated positions, and concluded that no matter the type of hearing aids used by the participants, none of the hearing-impaired participants had any perception of elevation, and the perception of azimuth was worse for the hearing-impaired participants than for the normal-hearing participants. In the Best study, participants were allowed to take their test hearing aids home for four to six weeks in order to get used to them. [2] also observed that hearing-impaired had almost no perception of elevation. However, these studies were conducted in anechoic chambers with only one single sound source active at a time.

The lack of studies on elevation can be explained by the world we live in: humans mostly interact with objects and people on their horizontal plane, or so people think. The distant car can be assimilated as being on the same plane as the person speaking next to us. Humans cannot easily move in elevation, and therefore everything is placed at a similar height as them. Localization studies therefore often limit themselves to the azimuthal plane. But the results of such tests are biased by the fact that the participants expect the sounds to come from the azimuthal plane: would their ability to localize sound be modified if they had no prior information about the elevation of the source, especially if they cannot estimate the elevation of a sound source?

Recently, an interest in 3D audio reproduction has been shown for hearing aid evaluation [3, 4, 5, 6, 7, 8]. Virtual Sound Environments (VSEs) offer the possibility to simulate plausible or realistic auditory environments similar to those perceived as difficult by the hearing-impaired people. Particularly, hearing-impaired people mention noisy restaurants, pubs and train stations as conditions where they struggle to understand speech [9]. Those environments typically feature numerous speakers, additional noise, and can be more reverberant than the environments perceived as easy such as living rooms.

In this context, the study presented here evaluates the localization performance of hearing-impaired people with and without hearing aids, using three different types of beamformer for the hearing aid cases. This paper presents the experimental setup. It is followed by the results of the perceptual experiment and a discussion on the results in the light of past studies on spatial audio.

2 EXPERIMENTAL SETUP

The experiment was conducted in a semi-anechoic room with an RT30 of 199ms. An array of 32 8020 DPM Genelec loudspeakers was used for the simulation of complex auditory environments:

- 8 loudspeakers every 45° azimuth at an elevation of -30°
- 12 loudspeakers every 30° azimuth at an elevation of 0°
- 8 loudspeakers every 45° azimuth at an elevation of +30°
- 4 loudspeakers every 90° azimuth at an elevation of +60°.

Loudspeakers were placed 1.5m away from the center of the reproduction system, except for the 4 top loudspeakers, which were 98cm away from the center. The position of the loudspeakers was inspired by [10]. The loudspeakers were hidden behind an acoustically transparent curtain.

The levels and delays of all the loudspeakers were corrected during a calibration process.

3D audio encoding and decoding was performed in Spat[1] with MAX[2]. Analysis of the measurements was performed in Matlab and R.

For this study, all Spat room effects were removed from the processing chain.

10 mildly hearing-impaired (people who have a Pure Tone Average hearing loss value between 26dB and 40dB) and 10 moderately hearing impaired (people who have a PTA hearing loss value between 41dB and 55dB) participants took part in the experiment. The test consisted of two sessions, taken on different days.

During the first session, the hearing thresholds of the participants were measured, and the tested hearing device was fitted for them. The participants were then instructed of their task and proceeded to a training phase, during which they rehearsed their task in the presence of the experimenter, who ensured they had well understood the instructions. Following this was the main part of the experiment, during which the participants had to face towards a reference point, indicating the 0° position both in azimuth and in elevation, and then localize as quickly as possible a continuous speech target signal that was played alongside a background sound recording. When they had localized the speech signal, they had to point with their head a laser pointer towards the direction of the sound and validate the position by pressing a button on a Microsoft gaming controller.

For the mildly hearing-impaired group, the tested hearing device was a pair of Phonak Audeo Receiver In the Canal (RIC) hearing aids with an open dome, meaning that they could still hear partially the sounds directly. For the moderately hearing-impaired group, 7 participants were equiped with Phonak Bolero BTE hearing aids with occluded ear-pieces. One moderately-impaired participant used a pair of Audeo hearing aids with an occluded ear-piece. The two remaining moderately-impaired participants used Audeo hearing aids with closed domes.

Fig.1 shows the audiogram boxplots of the two groups.

The participants wore a NGIMU headtracker upon which the laser pointer was fixed, ensuring that the direction of the headtracker and that of the laser pointer were the same.

Two auditory environments were simulated: a loud restaurant environment (RT60 = 572ms, EDT = 123ms) and a noisy street environment (RT60 = 837ms, EDT = 37ms). Both of them were recorded with an Eigenmike microphone and processed to obtain 4^{th} order background recordings. They were recorded at 67dBA. Both recordings were edited to obtain 1 minute-long recordings.

For each of the two enviroments, three sweep signals were recorded with a 8020 DPM Genelec loudspeaker at a distance of 1.5m from the Eigenmike microphone in the 0° direction. The recorded sweeps were then averaged and inverted in order to obtain the desired impulse responses.

[1] https://forumnet.ircam.fr/product/spat-en/
[2] https://cycling74.com/

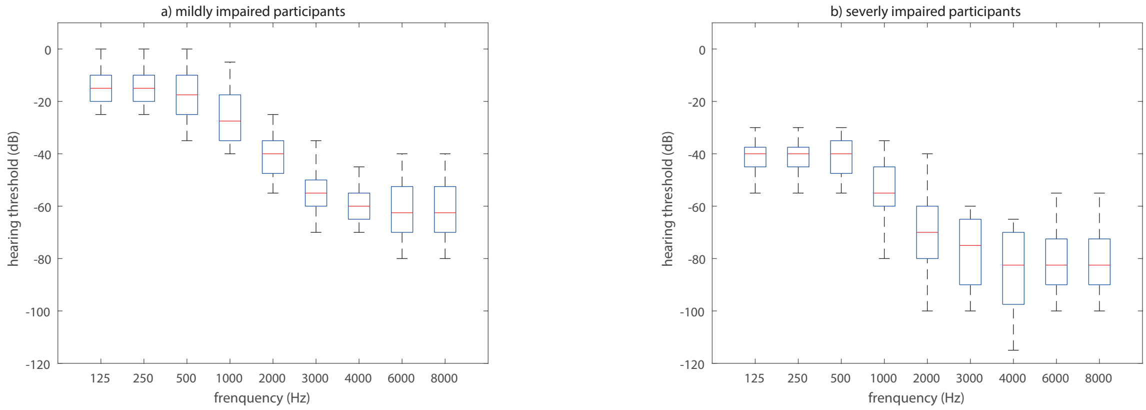

Figure 1. Boxplots of the mildly-impaired group (a) and of the moderately impaired group (b) hearing thresholds.

32 positions were used in this experiment. They were initially randomly-generated for a previous experiment conducted with normal hearing participants. Each octant of space contained 4 possible directions, in order to guarantee a relatively homogeneous distribution of the source positions. Elevation of sources was constrained to $\pm 40°$.

Four hearing aid conditions were tested: unaided, aided with beamformer A, B, or C. Beamformers combine the signals from the two microphones of each ear in different ways. It was therefore assumed that the beamformers might affect auditory perception in different ways. During the first session, the unaided conditions were tested once all the aided conditions were tested. During the second session, the unaided conditions were tested first.

3 RESULTS

Fig. 2.a) shows an overview of the perceived azimuth vs the target azimuth. An offset can be observed around 180°: sounds are perceived more counterclockwise than they should. The reason for this offset is still unclear. The problem already occurred to a lesser extent in a previous experiment conducted with normal hearing participants, bursts of pink noise, no reverberation nor background noise. Possible reasons for this bias considered so far indicate that:

- it is unlikely to result from the calibration of the system, as it would result in an overestimation of the angle to one side of the loudspeaker and an underestimation of the angle on the other side of the loudspeaker
- it does not come from the beamformers, otherwise there would be a constant shift.
- it is unlikely to be related to the rotation of the HOA impulse response, as the problem was already present in the B1 experiment.

The best explanation is therefore the movement of the body and the fact that when people are turning their head to the side, they tend to bend their head on one side, which could bias the headtracker results. It is however surprising that the bias happens to one side and not to the other.

One participant was not included in these results as his perceived angles seem to be random: there is no link between the target angle and the perceived one. Two other participants had to be removed because of a hardware issue during their test: only one session was valid for those two participants. This resulted in 17 participants whose results could be shown and analysed.

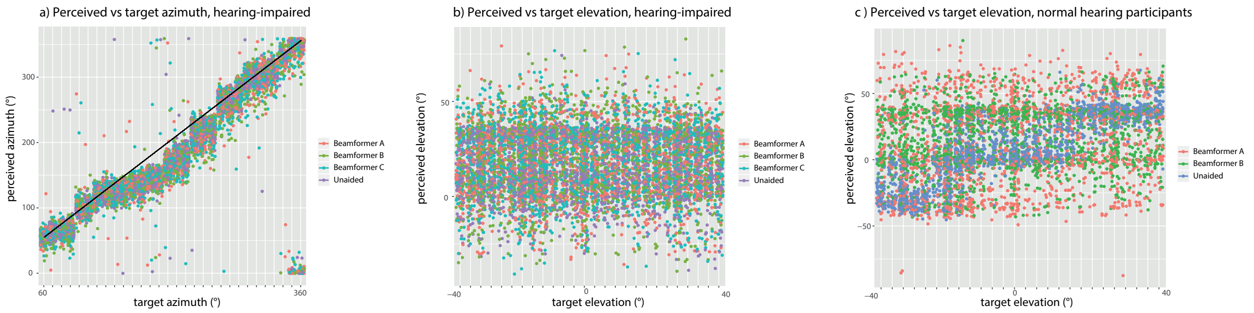

Figure 2. Perceived angles vs target angles. a) azimuth angles. The continuous line indicates the identity. b) elevation angles. c) elevation angles for a previous experiment conducted with normal hearing participants and bursts of pink noise, in the absence of background noise.

The perception in elevation shows that aided or not, the hearing impaired participants had no idea of the sound elevation. They almost all mentioned having problems for perceiving the elevation of sounds, both during the experiment and in real life.

In a previous experiment conducted using non-reverberant noise bursts, normal-hearing participants had a poor elevation judgment when aided but a good one when unaided, as shown on Fig. 2.c). It was therefore thought it would still be the case at least with the mildly impaired participants. The addition of room effect could not be an explanation for these differences, but the high frequencies audibility could: unaided, the participants do not have enough high frequency cues to localize sounds in elevation; aided, they have less frequency cues.

Participants perceived especially rarely sounds below the horizontal plane. Informal tests with normal hearing participants confirmed this. Despite the fact that when trying to simulate a sound coming from below, the lowest loudspeakers play louder sounds than the remaining loudspeakers, the sound still does not appear to come from below. This could be a combined effect of the asymmetrical reproduction system (four less loudspeakers below than above the horizontal plane, and carpet on the floor but not on the ceiling) and of the reverberation (diffuse sound and early reflections may come from all directions).

Statistical analysis of the reaction time, of the absolute error of localization in azimuth, and of the absolute error of localization in elevation was conducted. As the playback of the stimuli was continuous, it was expected that the error of localization should be little dependent on the position of the sound: no matter where the sound was coming from, the participants were supposed to end up facing the direction of the speech stimulus. In order to guarantee a potential average across positions, a correction offset was subtracted from each participant's reaction time results. For each participant and each direction, the minimum reaction time was subtracted from all the reaction time results of this participant in this direction. In addition, the data for the three dependent variables was log-transformed in order to ensure a normal distribution of the results.

A repeated-measures ANOVA was conducted on the data. It showed that session number, the processing, the type of background environment, and the direction of the sound had a significant effect on the reaction time (resp. $p < 0.01$, $F = 9$, $p < 0.01$, $F = 9$, $p < 0.01$, $F = 88$, and $p < 0.01$, $F = 5$). It also showed that the type of background environment and the position of the sound had a significant effect on the absolute error in azimuth (resp. $p < 0.01$, $F = 12$, and $p < 0.01$, $F = 12$), and that only the position had a significant effect on the absolute error in elevation ($p < 0.01$, $F = 14$). In each case, the street environment led to faster reaction time and less errors than the restaurant environment.

More particularly, there were large differences of observed effects between the participants. For 8 participants, there was no visible effect of the processing on the reaction time nor on the absolute error in azimuth. The remaining 9 participants could be grouped depending on the effect the processing had on their reaction time:

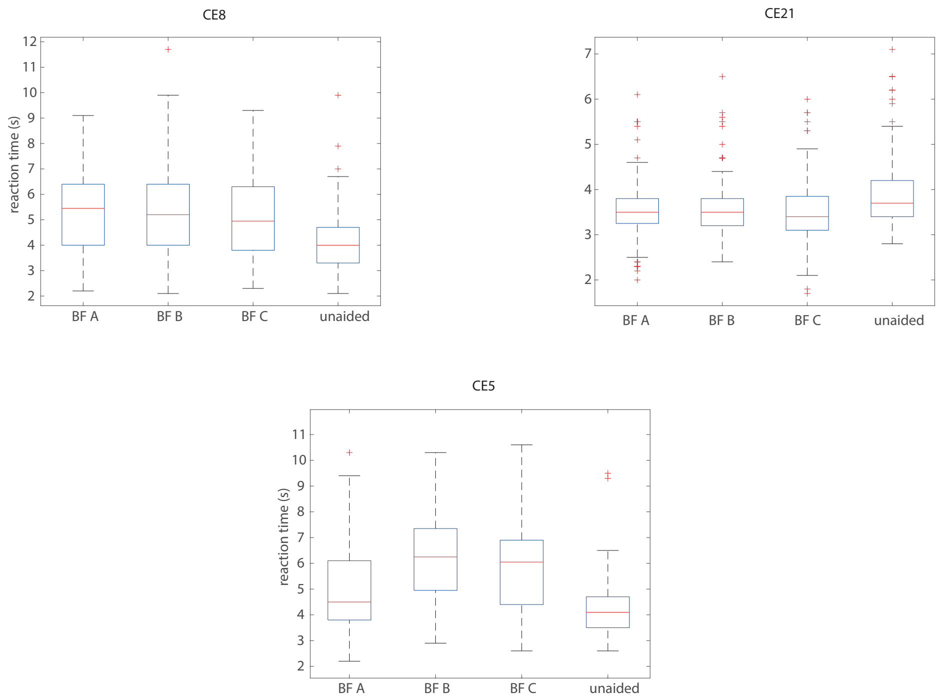

Figure 3. Three examples of the effect of the processing on the reaction time. Participants CE8 and CE21 were mildly impaired and had a different reaction time for the unaided conditions compared to the aided conditions. Participant CE5, moderately impaired, was faster unaided than aided, and faster with beamformer A than with beamformers B and C.

- Five participants, all moderately impaired, were faster both unaided and with the beamformer A
- Three participants (two mildly impaired, one moderately impaired) could localize sounds faster with the unaided condition
- One mildly impaired participants took longer to localize sounds when unaided than when aided.

This last category is unexpected. If the hearing impairment caused the localization to take longer, it should take the moderately impaired participants longer to localize sounds unaided, but that is not the case. Only one participant behaved like this. Examples of these three categories can be seen in Fig. 3.

4 DISCUSSION

The main finding of this experiment is similar to those of [1]: participants had no perception of elevation no matter which setting was used. Several possible causes can be hypothesized. In the unaided case, the hearing impairment of the participants made them oblivious to the spectral cues that are used for localization in elevation, as those can mostly be found above 5kHz [11]. In the aided case, the participants could not use the spectral cues that the hearing aids made available to them. The spectral cues produced by the BTE hearing aids

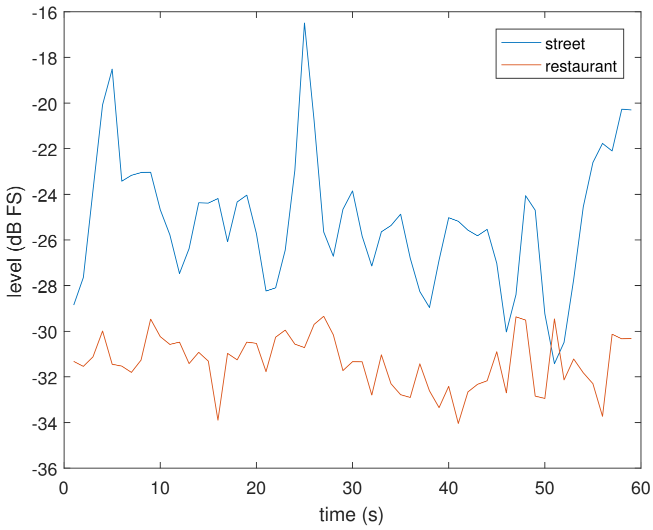

Figure 4. Evolution across time of the mean RMS level of both background sound recordings. Estimation of the rms level was done on the omnidirectional component of the HOA recordings.

are very different from those obtained through natural hearing. According to [12], the spectral cues produced by BTE hearing aids varied little across frequencies and show smaller differences between two different elevations than the spectral cues obtained through natural hearing or In-The-Ear (ITE) hearing aids. This makes the use of these cues more difficult. In this paper, no ITE nor Completely-In-the-Canal (CIC) hearing aids were used. They are expected to provide closer spectral and interaural cues than those obtained with BTE hearing aids. [1] compared BTE and CIC, but with hearing-impaired participants that were not used to CIC: the participants were either not hearing aid users at all or they were BTE hearing aid users. Studies on brain plasticity [13, 14, 15] showed that training helps learn new Head-Related Transfer Functions (HRTF). Training the hearing aid users to localizing with hearing aids may therefore help them use auditory cues that are different from those they used or that they have not used for a long time.

The type of environment had a significant effect on the reaction time and on the absolute error of localization in azimuth, despite the similar average level of both background sound recordings. However, both background sound recordings varied in terms of level homogeneity, as can be seen in Fig.4. The RMS level was estimated on the omnidirectional component of the recording, without any A-weighting. This explains the difference of global level between the two recordings, the street recording having some engine noises that include low frequency content. Average level was estimated on 1s-long windows. The figure shows that the level of the restaurant recording varied by up to 4.5 dB across time, whereas the level of the street recording varied by up to 15 dB. The faster reaction time and lower errors were then obtained with the street environment. It could be caused by the discrepancies of level, but also by the type of sounds (in the restaurant environment, noise was mostly present in the form of other speakers, the noise's spectro-temporal characteristics were therefore more similar to those of the target than the street noise's spectro-temporal characteristics) or by the difference of early reflections and reverberation on the target sound. The street environment had a larger RT60 value but a significantly shorter EDT value (37ms instead of 123ms for the restaurant environment, as described in the experimental setup). This could cause the reverberation to be less disruptive for the localization of sounds and therefore make the localization faster with less errors. This result suggests that one should test several environments when evaluating localization in noise.

For eight participants, there was no significant effect of the processing on the reaction time, whereas there was a significant effect for the remaining nine participants. The most likely explanation for this is that these participants may have tried to localize the sounds as quickly as possible, as instructed. This should have caused an increase of the absolute errors of localization for the most difficult conditions. However, the error of localization in elevation was seemingly random. An increase of this error would not have been visible. The standard deviation of the error of localization in the azimuthal plane was larger for the 8 eight participants who showed no significant effect of the processing on the reaction time (7.9) than for the 9 participants who did (6.9), but the difference is small and is therefore unlikely to explain this lack of sensitivity. Participants sometimes lacked of focus (two participants forgot the instructions half way through their first session), but if it were a real issue, the difference of std between the two groups should be larger. This difference of behaviour therefore requires further testing.

For the nine participants for whom a significant effect of the reaction time was observed, the moderately impaired participants showed an improvement when using beamformer A compared to beamformers B and C. This implies that the beamformer could be varied depending on the situation for moderately impaired participants: using a classifier, it would be desirable to switch to beamformer A when the localization of sounds is important, and possibly switch to beamformer B or C when the localization is less important if one of these provides an improvement of intelligibility. For the mildly impaired participants, results show better results only in the unaided case, thus suggesting that the gain of audibility of spatial cues is not sufficient to compensate for the distortion of cues caused by the hearing devices.

5 CONCLUSIONS

This study showed the lack of localization in elevation for hearing-impaired people, whether equipped with BTE hearing aids or not. The lack of hearing aids reduces the audibility of the cues used for the localization in elevation, and the presence of BTE hearing aids provides spectral cues that are different and vary less depending on the elevation than the natural spectral cues.

The type of complex environment had an effect on the reaction time and localization error despite a similar average level for all background sounds. The environments varied in terms of homogeneity of the level, type of sound sources, and reverberation. This suggests that further studies on spatial perception with hearing aids should include several different environments.

Finally, the type of hearing aid processing had a significant effect on the reaction time for 9 out of 17 participants. The effect varied depending on the participant's hearing loss: participants with moderate hearing loss behaved significantly better unaided and with one of the beamformers than with the other two beamformers, whereas the mildly impaired participants performed better only in the unaided case.

ACKNOWLEDGEMENTS

This work was funded by CTI grant Nr. 25276.2 PFLS-LS.

REFERENCES

[1] Best, V.; Kalluri, S.; McLachlan, S.; Valentine, S.; Edwards, B.; Carlile, S. A comparison of CIC and BTE hearing aids for three-dimensional localization of speech, International Journal of Audiology, Vol 49 (10), 2010, pp 723-732.

[2] Byrne, D.; Noble, W. Optimizing Sound Localization with Hearing Aids, Trends in Amplification, Vol 3 (2), 1998, pp 51-73.

[3] Oreinos, C.; Buchholz, J. M. Objective analysis of higher-order Ambisonics sound-field reproduction for hearing aid applications. Proceedings of Meetings on Acoustics, 2013.

[4] Minnaar, P.; Albeck, S. F.; Simonsen, C. S.; Søndersted, B.; Oakley, S. A. G.;Bennedbæk, J. Reproducing Real-Life Listening Situations in the Laboratory for Testing Hearing Aids, Audio Engineering Society 135th Convention, 2013.

[5] Grimm, G.; Ewert, S.; Hohmann, V. Evaluation of Spatial Audio Reproduction Schemes for Application in Hearing Aid Research, Acta Acustica united with Acustica, Vol 101 (4), 2015, pp 842-854.

[6] Cubick, J.; Dau, T.; Validation of a Virtual Sound Environment System for Testing Hearing Aids, Acustica United with Acta Acustica, Vol 102 (3), 2016, pp 547-557.

[7] Oreinos, C.; Buchholz, J. M. Evaluation of Loudspeaker-Based Virtual Sound Environments for Testing Directional Hearing Aids. Journal of the American Academy of Audiology, Vol 27 (7), 2016, pp 541-556.

[8] Simon, S. R. L.; Wuethrich, H.; Dillier, N. Comparison of Higher-Order Ambisonics, Vector- and Distance-Based Amplitude Panning using a hearing device beamformer, 4th International Conferenceon Spatial Audio, 2017.

[9] Newman, C. W.; Hug G. A.; Jacobson G. P.; Sandridge S. A. Perceived Hearing Handicap of Patients with Unilateral or Mild Hearing Loss, Annals of Otology, Rhinology & Laryngology, Vol 106 (3), 1997, pp 210-214.

[10] Zotter, F.; Pomberger, H.; Noisternig, M. Ambisonic Decoding with and without Mode-Matching: A Case Study Using the Hemisphere, 2nd Int. Symposium on Ambisonics and Spherical Acoustics, 2010.

[11] Langendijk, E. H. A.; Bronkhorst, A. W. Contribution of spectral cues to human sound localization, J. of the Acoust. Soc. of Am. , Vol 112 (4), 2002, 1583 – 1596.

[12] Denk, F., Ewert, S. D.; Kollmeier, B., Spectral directional cues captured by hearing device microphones in individual human ears, J. of the Acoust. Soc. of Am., Vol 144 (4), 2018, 2072-2087.

[13] Stitt, P.; Picinali, L.; Katz, B. Auditory Accommodation to Poorly Matched Non-Individual Spectral Localization Cues Through Active Learning, Nature Scientific Reports, Vol 9, Article number: 1063, 2019

[14] Parseihian, G.; Katz, B. F. G. Rapid head-related transfer function adaptation using a virtual auditory environment, The Journal of the Acoustical Society of America, Vol 131 (4), 2012, pp 2948-2957.

[15] Hofman, P. M.; Van Riswick, J. G.; Van Opstal, A. J. Relearning sound localization with new ears, Nature Neuroscience, Vol 1 (5), 1998, pp 417-421.