Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Free-field Virtual Psychoacoustic and Hearing Impairment: Paper 373

Prior exposure to room acoustics and its effect on localization

Samuel W. Clapp, Bernhard U. Seeber

Audio Information Processing, Technical University of Munich, Germany
samuel.clapp@tum.de, seeber@tum.de

Abstract

In the majority of studies of psychoacoustics in rooms, test subjects do not interact with the stimuli presented to them. One form of interaction that has been increasingly deployed in recent studies is head movements, which have been shown to improve realism and reduce front-back confusions. Recent research and advances in computing power have paved the way for new forms of interaction, making it possible for subjects to modify their own position and that of sound sources in a virtual room, with fast updates to the room simulation.

In this study, two conditions are tested in order to examine how localization judgments of a short speech target in the presence of a noise distracter are affected by prior experience of a room's acoustics. In the first condition, no prior exposure to the simulated room's acoustics is given. In the second condition, listeners are able to first explore the room by manipulating the source position within the room geometry and hearing the resulting auralizations, before being tested again with the same target stimuli. These conditions test the hypothesis that active exploration and learning of a room's acoustics can be accomplished with practice, and can assist in source localization in non-ideal conditions. Results show large localization errors occur without prior exposure to the room's acoustics, always pulled in the direction of the noise source. Reduced localization errors are seen following the interactive exploration session. In both conditions, errors are larger the farther away in azimuth the target is located from the noise source.

Keywords: Room Acoustics, Sound Localization, Virtual Acoustics, Real-Time Auralization

Prior exposure to room acoustics and its effect on localization

1 Introduction

It has been shown in several previous studies that the human auditory system has the ability to adapt to the acoustics in reverberant environments over time. This ability is strongly related to the precedence effect, referring to a group of phenomena that describe how human listeners localize sound in the presence of additional sound reflections. Adaptation to reflections has been shown in studies involving a direct sound signal together with one or two later sound reflections.

Studies by Clifton [1] and Clifton et al. [2] looked at conditions under which a change occurs in the echo threshold, the shortest time delay between two sound stimuli at which at listener can discern two distinct sound events. In the first study [1] listeners were presented with clicks from two loudspeakers with a time delay of 5 milliseconds, and only one click was perceived by listeners, as predicted from the precedence effect. When the leading and lagging speakers were switched, listeners would suddenly hear clicks from both speakers, denoting a breakdown of the precedence effect. The effect would then build up again over a series of 8-10 clicks. A following study [2] examined a similar effect, playing listeners a “conditioning” train of clicks, and then changing the time interval, frequency content, or intensity. This study found that changes that would correspond to a highly improbable room acoustic event (such as a suddenly moving wall) were most likely to cause a breakdown of the precedence effect.

Brandewie and Zahorik [3] looked at how speech intelligibility in a noisy room improves with longer exposure to the acoustics of that room. A virtual room with a target talker and noise source was simulated to listeners over headphones, and listeners identified words spoken by the target talker (from a closed set) at varying signal-to-noise ratios. The words were presented to the listeners either in isolation, or as part of a sentence (with a preceding adapter), with the additional length of the target signal providing listeners a greater opportunity to adapt to the room acoustics. An average decrease of 2.68 dB in speech reception threshold was seen when the longer signal was used, showing evidence that adaptation to room acoustics is beneficial for speech intelligibility in rooms.

Previous work in this group by Menzer et al. has examined the capability of the human auditory system to learn the entire reflection patterns in a room [4]. This work employs a novel technique of comparing accurately simulated rooms to simulations with “scrambled” reflections, those with the same temporal reflection pattern, but with the spatial location of each reflection randomly changed, and how accurately listeners can discern source movement in each of these types of rooms. In a normal room with stable surfaces, reflections from a new source position will correspond to the same surfaces. However, in an auralization with spatially scrambled reflections, source movement would also correspond to wall movement, leading to unnatural sounding stimuli. Another study from this group by Seeber et al. (Paper 775, ICA 2016), examined the effects of passive prior exposure to room acoustics, delivered through adaptor stimuli consisting of trains of adapter noise bursts presented from random directions in a simulated room. Listeners localized a target noise burst following either a short adapter train of two noise bursts, a long train of 14, or no adaptor. It was found that localization improved following the adaptor stimuli, as compared with no adaptor, indicating a benefit in localization ability when some previous exposure to the room's acoustics is obtained.

In all of the previously listed studies, the adaptation to the room acoustics was a largely involuntary process, caused by exposure to that room's spatial and temporal reflection pattern, and general previous experience with hearing in rooms. Advances in computing power and virtual reality technology have provided the opportunity for listeners to more interactively explore virtual spaces through manipulating sound source and listener positions and having the simulation update itself in real-time. This allows for testing scenarios that more closely resemble real-life scenarios, and thus might have greater validity in "everyday" conditions. For this reason, we are interested in exploring active forms of learning room acoustics, with two-way communication between listener and stimulus, and whether this form of learning room acoustics can produce similar localization improvement as the more passive type of learning that comes solely from exposure time to a reverberant environment.

2 Methods

2.1 Software and hardware

This test was conducted in the Simulated Open Field Environment (SOFE) at the Audio Information Processing Group at the Technical University of Munich. The SOFE consists of software for room acoustics simulation, and a horizontal ring of 96 closely-spaced loudspeakers for presenting room auralizations to listeners. The room simulation software employs the image source method to simulate the entire room impulse response. Image sources are simulated up to a very high order (200 for the simulated rooms in this study) to model high reflection density in late room reverberation, and a temporal jitter is applied to later image sources to simulate diffusion. Each individual room reflection is rendered by the nearest loudspeaker in the array (with elevated sources rendered to the loudspeaker located on the speaker nearest to the cone of confusion). [5]

A high-performance version of the room simulation software was recently developed, which is capable of simulating 80000 image sources in 2.3 milliseconds on an eight-core desktop PC, and was used to simulate all of the room impulse responses used in this study. This software makes possible real-time room acoustics simulations that adapt to input from the user and re-calculate room simulations with minimal latency. In addition, a real-time convolution system to generate multi-channel loudspeaker auralizations that smoothly transition between different room impulse responses as the simulation parameters change, and was used in this study for the interactive exploration phase.

Listeners' localization judgments were collected using a proprioception-decoupled pointer (the *ProDePo*-method) [6], a trackball-controlled laser pointer that projects onto a white paper ring sitting just above the loudspeakers.

2.2 Simulated room

A rectangular room with a width of 5 meters, a length of 9 meters, and a height of 2.3 meters was used to create the test stimuli. Frequency-dependent absorption coefficients were simulated for each surface, with heavy carpet on concrete used on the floor, and painted brick for the walls and ceiling, yielding a broadband reverberation time of approximately 1 second. For the speech target source position, three different angular positions (19 degrees right, 11 degrees left, and 41 degrees left) and two source-receiver distances (2 meters and 4 meters) with respect to the listener were used, for a total of 6 different target positions. The noise source was always located in the same position, 2 meters from the listener at 120 degrees right. The room layout is shown in Figure 1.

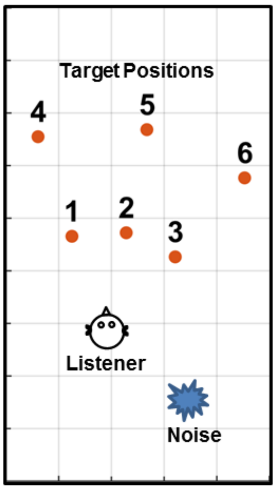


Figure 1. Diagram of test room showing positions of the listener, speech targets, and noise. Each box is 1 square meter.

Two other rooms were created for the creation "dummy stimuli" that were inserted in between each of the trials with stimuli from the test room, in order to minimize a learning effect occurring over the course of the trials. One was an anechoic room, leading to an auralization with target and noise each playing from just a single loudspeaker. The other was a room twice as large the test room in every dimension, leading to a volume eight times as large, and thus a longer reverberation time. The listener, target, and noise source positions were also scaled by a factor of two.

2.3 Test Stimuli

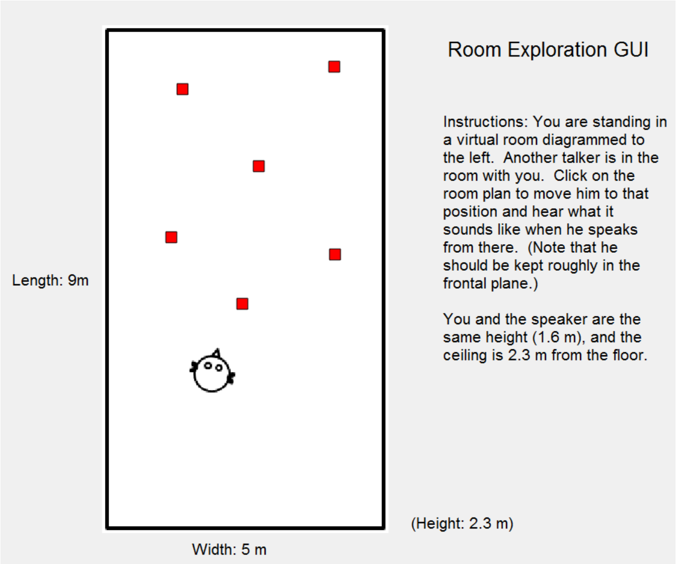
Each test stimulus consisted of pink noise played from the noise location, and male speech played from one of the target locations. The male speech stimulus was very short, consisting of the words "In the", and lasting approximately 0.3 seconds. The noise lasted for 1.3 seconds, with the speech target always occurring 0.5 seconds after the start of the noise. The noise would thus continue for another 0.5 seconds after the end of the speech target signal, with the reverberant tail of the noise included in the stimulus. The signal-to-noise ratio was -6 dB for Positions 1-3, and -12 dB for Positions 4-6.

2.4 Test Procedure

The test took place in two sections. In the “No Learning” (NL) condition, no prior training was given. The test was completed in darkness. After each stimulus played completely, the laser pointer was activated, and subjects moved the laser to the direction from which they perceived the target stimulus. As mentioned in Section 2.2, “dummy” stimuli (either anechoic or from a larger room) were inserted between each stimulus to minimize learning effects occurring during the test itself.

In the “Active Learning” (AL) condition, listeners were able to interactively explore the acoustics of the test room using a Matlab GUI, shown in Figure 2, and the real-time convolution software described earlier. A diagram of the room was shown, and users could adjust the source position by tapping on a touch screen and hearing what the target speech sounded like from different locations in the test room. (The target was presented together with the noise, to allow for a similarity to the final test condition.) Red squares on the screen marked the locations they had already heard. Subjects could take as long as they wanted with this section.

Four subjects completed 10 repetitions of each stimulus, divided into 2 runs for each part. For the AL condition, the room exploration was completed before each run. Half of the subjects completed the NL condition first, and the other half the AL condition first. For those who completed the AL condition first, a break of at least 20 minutes was taken before completing the NL condition, to allow them to “forget” the training they had received. Completing the test, including the interactive exploration component, lasted approximately 30-40 minutes.



Instructions: You are standing in a virtual room diagrammed to the left. Another talker is in the room with you. Click on the room plan to move him to that position and hear what it sounds like when he speaks from there. (Note that he should be kept roughly in the frontal plane.)

You and the speaker are the same height (1.6 m), and the ceiling is 2.3 m from the floor.

Figure 2. Matlab GUI used for interactive room exploration.

3 Results

Results are shown in Figure 3 for each speech target position. The data are plotted in degrees with respect to the “true” source location, with medians and quartiles shown. Across all conditions, median localization errors are generally greater the farther away the target is located from the noise in azimuth. For all speech target positions, the AL condition exhibits a more accurate median location judgment with respect to the “true” direction of the target position. Errors are nearly always to the right, i.e. towards the noise source, indicating that localization of the target is somehow pulled towards the noise. For Positions 1, 4, and 5, this improvement is quite substantial, around 20 degrees. For Positions 2 and 6, the improvement is around 10 degrees. For Position 3, the median localization judgment for the AL condition is actually negative (i.e. to the left), the only point where this occurs. In several target positions, but most noticeably Positions 3 and 6, there is also a decrease in the interquartile range of localization judgments in the AL condition as compared to the NL condition.

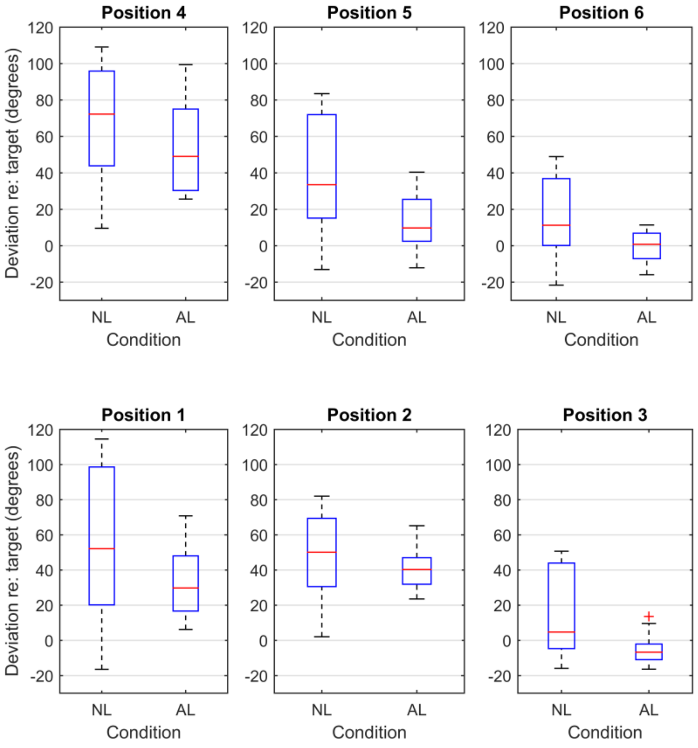


Figure 3. Localization results, showing medians and quartiles for each speech target position and learning condition. The vertical axis is in degrees relative to the "true" position of the speech target.

4 Conclusions and future work

These results, while preliminary in nature, suggest that virtual exploration of a reverberant room can offer some improvement for sound source localization under non-ideal circumstances, with a short duration target and poor signal-to-noise ratio. It also appears that the location of the noise distracter with respect to the target may have an effect on localization judgments in this scenario, biasing perception of the target towards the noise source. Future experiments can examine the role of the position of the noise source, as well as multiple noise sources or diffuse noise. Finally, there are also a wide variety of opportunities for adapting and enhancing virtual room exploration, including real-time adjustment of the levels of different stimuli in the room, incorporating source (i.e. listener) movement, and visual rendering. These techniques could potentially help both normal-hearing and hearing-impaired listeners the opportunity to "practice" in difficult acoustic situations and learn how to better navigate them.

Acknowledgements

Supported by the Bernstein Center for Computational Neuroscience Munich, BMBF 01 GQ 1004B.

References

[1] R. K. Clifton, "Breakdown of echo suppression in the precedence effect," *J. Acoust. Soc. Am.*, vol. 82, no. 5, pp. 1834–1835, 1987.

[2] R. K. Clifton, R. L. Freyman, R. Y. Litovsky, and D. McCall, "Listeners' expectations about echoes can raise or lower echo threshold.," *J. Acoust. Soc. Am.*, vol. 95, no. 3, pp. 1525–33, 1994.

[3] E. Brandewie and P. Zahorik, "Prior listening in rooms improves speech intelligibility.," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 291–299, 2010.

[4] F. Menzer, M. Kleinsteuber, and B. U. Seeber, "Influence of geometrically implausible reflections on decision variance in sound movement detection," in *Proc. Bernstein Conference*, 2015.

[5] B. U. Seeber, S. Kerber, and E. R. Hafter, "A System to Simulate and Reproduce Audio-Visual Environments for Spatial Hearing Research," *Hear. Res.*, vol. 260, no. 1–2, pp. 1–10, 2010.

[6] B. Seeber, "A New Method for Localization Studies," *Acta Acust. united with Acust.*, vol. 88, pp. 446–450, 2002.