Auditory Room Learning and Adaptation to Sound Reflections

Bernhard U. Seeber and Samuel Clapp

Abstract Sound reflections are abundant in everyday listening spaces, but they are rarely bothersome, and people are often not even aware of their presence. As shown in several studies, this is partially due to adaptation of the human auditory system to the spatiotemporal reflection pattern, namely, through an increase in the echo threshold that follows repeated exposure to the same reflection pattern. This raises the question of whether adaptation mechanisms to room reflections lead to improved localization accuracy as well—a measure more tangible for everyday listening. Moreover, this benefit would only be useful if it could be maintained through changes in the reflection pattern such as those produced by head turns or body movement within the room, or from sources at different locations. Therefore, a particular mechanism is hypothesized by the current authors based on learning a representation of the room geometry, rather than learning of or adapting to a specific reflection pattern. This chapter reviews and discusses the available literature on the build-up of the precedence effect and related effects in speech understanding in reverberation. In light of the hypothesis of room learning, it aims to trigger a discussion about the underlying mechanisms.

1 Introduction

One of the most remarkable abilities of the human auditory system is how it can function successfully in highly challenging acoustic environments. Nearly every built environment—where humans spend most of their time—contains surfaces that reflect acoustic energy. When a sound is emitted in such a space, listeners not only receive a signal that is traveling directly from the sound source to the ears but multiple

B. U. Seeber (✉) · S. Clapp
Audio Information Processing, Department of Electrical and Computer Engineering,
Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany
e-mail: seeber@tum.de

S. Clapp
e-mail: samuelclapp@fb.com

J. Blauert and J. Braasch (eds.), *The Technology of Binaural Understanding*,
Modern Acoustics and Signal Processing,
https://doi.org/10.1007/978-3-030-00386-9_8

delayed copies of the original signal superimposed upon it. The specific pattern of these delays (in time and space) is determined by the geometry of the surfaces in the environment and the positions and orientations of both the sound sources and the listeners.

It would be extremely difficult to localize sound effectively in reverberant environments without any sort of mechanism to deal with reflected sound energy. However, as most know from personal experience, normal-hearing human listeners are quite good at this, even in the absence of other sensory cues (Hartmann 1983; Blauert 1997). The mechanisms underlying this ability are understood in the context of the so-called "Precedence effect", a name given to a group of related phenomena that are briefly discussed in the following—compare Blauert (1997), Litovsky et al. (1999), Brown et al. (2015). The precedence effect has been studied extensively with stimuli played from both a single leading location and a single lagging location, separated by some time interval on the order of milliseconds. Results from one listener in such an experiment by Seeber and Hafter (2011) are depicted in Fig. 1. For short lead-lag delays up to 12 ms (here for a spoken word), a single sound location was reported both in the localization responses (top) and in the fusion responses (bottom). Hence, despite being played from loudspeakers separated by 60°, at short delays the lead and lag stimuli are perceptually fused into a single auditory event. This (fused) auditory event is located between the loudspeakers for delays up to 2 ms, an effect known as "summing localization", which is widely used in stereophony. For longer delays the sound is perceived as coming from the leading loudspeaker, hence the name "precedence effect" or "localization dominance" of the first incoming wavefront. Above the "echo threshold", here around 12–24 ms, lead and lag are segregated into two distinct auditory events, one perceived at the lead and one at the lag location. "Lag discrimination suppression" is the third phenomenon besides localization dominance and fusion subsumed under the term "precedence effect" (Yang and Grantham 1997; Litovsky et al. 1999; Brown et al. 2015). It indicates the listener's difficulty to determine binaural parameters of the lag stimulus at short lead-lag delays, with the difficulty decreasing as the delay increases.

In such precedence effect experiments, stimuli are usually presented in isolation, whereas in most listening situations, sources repeatedly emit sound, thereby giving the auditory system the opportunity to reassess and integrate information about the source and the room over time (Clifton and Freyman 1996; Hafter 1996). The focus in the current chapter is on the role of the context, that is, on the question of how signals that immediately precede the test stimulus affect the perception of that stimulus. This is an important question in terms of understanding spatial hearing in rooms because, in the course of a normal day, people spend minutes or hours at a time in one place, and repeatedly in the same places from day to day. This gives our auditory system the chance to collect information about the space via the acoustic signals reaching the ears. There is much evidence, both in studies of the precedence effect and in the articles discussed in this chapter, that the accumulation of this acoustic information results in later reflections being suppressed in favor of the direct sound, shown by significant increases in the echo threshold and lag discrimination suppres-

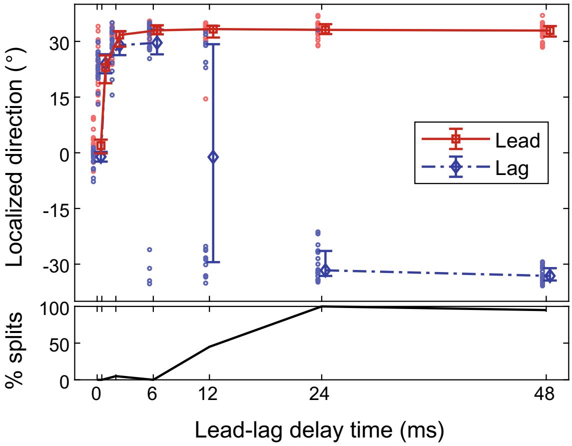

Fig. 1 **(Top)** Localized positions in the precedence-effect paradigm. The same sound, a recording of the spoken word *shape*, is played at equal level from both the lead loudspeaker (plotted at +30°) and the lag loudspeaker (at −30°), separated by the lead-lag delay given on the abscissa. Plotted are individual responses **(small circles)** as well as medians with quartiles. Medians are connected by lines for readability. Results were obtained with a light-pointer method. Lead and lag locations were randomized (from ±30°) and the listener was in one run instructed to point to the leftmost sound, if two locations were perceived, and in another run to point to the rightmost sound location. The listener thus pointed in separate trials to the lead and lag locations. By knowledge of the actual lead and lag location, data were analyzed into pointing to the lead **(red squares)** and the lag **(blue diamonds)**. **(Bottom)** Percentage of trials in which more than one sound event, that is, split images, was perceived. All data stem from one normal-hearing listener. Replotted from Seeber and Hafter (2011)

sion. This process is assumed to assist with localization and speech understanding in reverberation.

One way to understand this phenomenon is to consider it as inhibition and adaptation process. The auditory system, after having obtained sufficient information about the acoustic environment, suppresses information from the directions of strong reflections. This view is grounded in discrimination suppression experiments that show reduced access to binaural cues in the lagging sound. Starting with early views of the precedence effect as inhibition (McFadden 1973), corresponding models use inhibition of monaural and binaural information after the sound onset as a general suppression process or suppress specific lead-lag delays or interaural time differences (ITDs), that is, time differences in the arrival of a signal at the closer ear and the arrival at the farther ear—compare Hartung and Trahiotis (2001), Lindemann (1986a). For example, in such models, inhibition equipped with a forgetting time increases upon repeated presentation from the same direction, thus demonstrating adaptation in terms of an increasing echo threshold. A model proposed by Djelani and Blauert (2002) exhibits a direction-specific buildup of the precedence effect that could be viewed as an adaptation of binaural neurons that signal particular directions, namely, the echo threshold is increased for directions from which reflections were

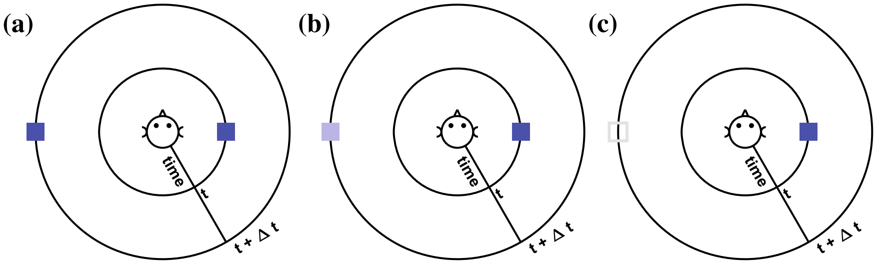

Fig. 2 A schematic diagram of the buildup of the precedence effect. In these plots, the radius represents time, and the azimuth the direction of the sound stimulus. The sub-figures represent the stages of the build-up of the precedence effect for the case of a leading click coming from the right of the listener, followed by a lagging click from the left, separated in time by some interval Δt. **(a)** Initial exposure to the lead-lag pair, where both clicks are perceived. **(b)** Intermediate phase after a few presentations of the lead-lag click pairs, where the lagging click is still perceived but is beginning to be suppressed. **(c)** Complete build-up of the precedence effect after repeated presentation of lead-lag click pairs, where the lag is still acoustically present but no longer perceived as a discrete auditory event

repeatedly presented. Here the question comes up of whether such direction-specific adaptation is ecologically useful, or even ecologically valid, given that real-world acoustic environments are much more complex. Figure 2 shows spatiotemporal diagrams that illustrate direction-specific adaptation.

The (direction-specific) adaptation observed in precedence effect experiments and addressed in the present chapter should be considered separate from "binaural adaptation", a term coined by Hafter for phenomena related to the localization of binaural click trains that do not contain reflections (Hafter 1996). Hafter and colleagues studied the relative weight given to individual clicks in a click train when localizing the complete click train. For high-rate click trains, localization is determined almost exclusively by the first click, indicating onset dominance (Hafter and Dye 1983; Stecker and Hafter 2002). A restart of the adapted binaural system, seen by an increased weight of a click, occurs, for example, after a gap in the train (Hafter and Buell 1990).

The common way of acoustically experiencing a room is not a static process since listeners and sound sources almost never remain completely fixed in place. In addition, listeners constantly make small adjustments to their head orientation and posture. This strongly affects the interaural cues. Thus, listeners usually perceive a specific space by experiencing its reflection pattern that varies in time. In a simple "adaptation" process, any movement of sources or listeners would thus require an ongoing re-adaptation to the current configuration. Of course, as long as the listener remains within the same room, these changes in the reflection pattern will abide by an underlying logic as dictated by the geometry of the room—a natural scenario that listeners encounter daily. The current authors thus postulate another potential way to understand this process, namely, through "room learning" or "abstraction" as proposed in Seeber et al. (2016), Menzer and Seeber (2014). Rather than simply suppressing information impinging on the listeners from specific directions as in an adaptation process, in the “abstraction” process, the direct sound and early reflections are used by the auditory system to develop an abstract representation of the room geometry. The positions of reflecting surfaces (and thus the room boundaries) are inferred via the timings and locations of reflections. Based on such an abstract room representation, signals arising from early reflections can be anticipated and, subsequently, their interaural binaural cues can be suppressed. While this appears indistinguishable to an adaptation process for static sources, a room abstraction process can allow for suppression even after head turns or position changes in the room. Thus, in such a process, the perception of a given room would not require a new period of adaptation following source or listener movements, once the auditory system has acquired the necessary information to develop a model of the geometry of the space. A schematic diagram of the two proposed processes following head rotation is depicted in Fig. 3. In this chapter, several studies will be examined in light of these two proposed processes—adaptation and abstraction—with the aim of exploring whether the results provide evidence of the existence of one or the other, or even of both.

2 Context Effects with Simple Lead-Lag Stimuli

In seeking to understand how the auditory system deals with reflections, many studies have made use of the simplest case, with a leading stimulus from one horizontal location and a lagging stimulus from another horizontal location, separated by a time interval. In the *real world*, this would correspond to a room with a single reflective surface, and all other surfaces being completely absorptive. The direct sound would come from the lead location and the reflection from the lag location. In the introduction, it was discussed how these lead-lag stimuli are perceived based on the duration of the interval between the lead and the lag. Here we consider how stimuli that immediately precede a test stimulus affect the perception of the latter. This is particularly important with respect to everyday listening in rooms, where longer periods of time are spent in specific spaces and thus, the acoustic context plays an important role.

One of the first and best-known examples for the impact and build-up of context is the “Clifton effect” (Clifton 1987). A lead-lag click pattern with an interstimulus interval of 5 ms was played in free field from loudspeakers located at ±90°, that is, from perpendicularly to the left and the right of the listener. First, a click train of several seconds was played, with the right loudspeaker leading. Then, the positions of the lead and lag were suddenly flipped, and listeners were asked whether they heard clicks from the left loudspeaker, right loudspeaker, or from both. Immediately after the locations were switched, all listeners heard both the leading and the lagging clicks distinctly. However, after a certain number of clicks in the new spatial orientation, listeners returned to hearing clicks solely in the direction of the lead loudspeaker.

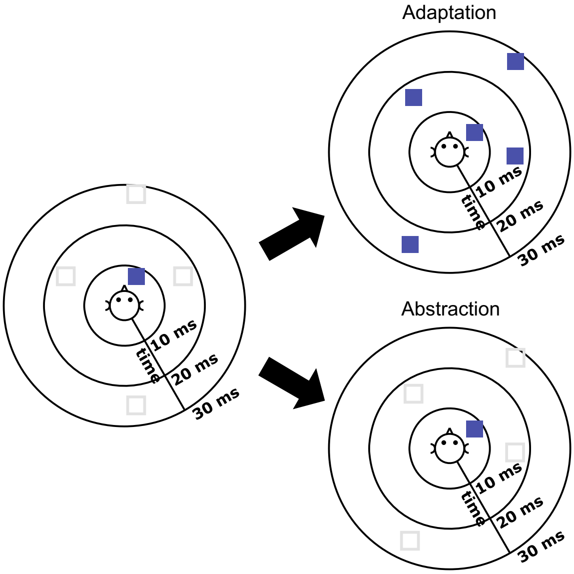

Fig. 3 A schematic diagram illustrating predicted perception in a room (simplified to the direct sound and four first-order reflections) following head-turning under the "adaptation" and "abstraction" processes. The left scheme represents perception after repeated exposure to the direct sound with reflections, where only the direct sound is perceived distinctly and the reflections are suppressed. Both schemes on the right show the new orientation of the direct sound and reflections with respect to the listener following the listener turning the head 30° to the left. The expectation in an "adaptation" process **(top right)** is that the precedence effect breaks down and needs to build back up again because the head-turning results in a new spatial orientation of reflections with respect to the listener. In an "abstraction" process **(bottom right)**, however, the listener continues to suppress the reflections in favor of the direct sound because the room is still the same

This experiment demonstrates the involvement of dynamic processes in the precedence effect, with both fusion and the echo threshold increasing over time in response to repeated exposure to the same stimulus. The Clifton effect was modeled by Djelani and Blauert (2002) based on an approach by Lindemann (1986a, b) by using an interaural cross-correlation function with a dynamic inhibition algorithm. Peaks in the cross-correlation function (corresponding to the directions of sound events) inhibited the function at other delay values (i.e. other directions). In addition, the strength of the inhibition was increased when it was triggered regularly and repeatedly. The model successfully reproduces the results of Clifton (1987). At the first presentation of a lead-lag stimulus, two peaks in the binaural activity map appear, corresponding to both the lead and lag locations and indicating a situation above the echo threshold, where fusion has not yet taken place. However, after 3–4 presentations of the lead-lag stimulus at regular intervals, the peak corresponding to the lag has disappeared, indicating that fusion has now taken place. The Lindemann model assumes an adaptation process for replicating the Clifton effect and does not need to estimate the room geometry, just the spatial locations of lead and lag sound sources as inferred from the binaural signals. The binaural features of the lag become suppressed over time. Consequently, the lag auditory events cease to be spatially separate.

The psychoacoustic data by Rachel Keen (then Clifton) indicate that suppression of the lag does not happen immediately. She examined different click pair rates and found that at a rate of 1/s, the time required for lag suppression to build up completely was 8–10 s, while for faster click pair rates of (2–4)/s, the time was only 3–5 s. Thus, it appears to require 8–12 click pairs for the lag to be suppressed rather than a set length of exposure time to a given reflection pattern (Clifton and Freyman 1989, 1996). Freyman et al. (1991) confirmed the total number of click pairs in the conditioning train to be the most salient quantity in predicting an increase in the echo threshold, rather than the click pair rate or the total duration. An increase in the echo threshold was observed when increasing the total number of click pairs from 3 to 5 to 9, with only a very small change when increasing from 9 to 17 clicks. This suggests that a plateau is reached with nine click pairs in the conditioning train. Note that while click trains are an interesting stimulus since they provide a quantified amount of information per click, build-up appears to occur faster for continuous speech (Djelani and Blauert 2001).

It is interesting that it requires a number of click pairs (greater than one) to increase the echo threshold, as a new click pair in a train does not actually contribute any new information, as it is identical to the preceding ones. This is congruent with an adaptation process, where an inhibitory process requires repeated stimulation to build up over time. Likewise, it could also be explained by the room learning process, which requires repeated presentation in order to extract reliable information, namely, a sufficient number of observations.

In a controlled laboratory setting, the click pairs in the conditioning train and in the test stimulus can be made exactly identical. However, it would be extremely rare for this type of scenario to happen in a natural environment, so for these effects to have any sort of validity outside of the laboratory, it is of interest to know whether they can be observed for stimuli that have the same spatiotemporal arrangement, but differ in other aspects.

In another experiment by Freyman et al. (1991), clicks were replaced with short white noise bursts, using either the same or different noise tokens for every burst. The echo threshold increased in both cases, indicating that the exact waveform of the bursts is not critical. Clifton et al. (1994) went on to vary the frequency content and intensity of the test click pair with respect to the conditioning train, but now they measured the lag discrimination ability rather than the echo thresholds. There was little to no difference in discrimination performance when changes in frequency content or intensity were introduced between the conditioning click pairs and the test click pair, versus when these parameters were held identical between the two pairs. These results indicate that the spatiotemporal arrangement of the click pairs is the salient feature that goes with increased echo thresholds and lag discrimination suppression.

The results derived from different noise tokens and intensity changes between conditioning train and test stimulus can be understood via either the adaptation or the abstraction process. However, it is unclear whether a tonotopically operating model would predict the psychoacoustic results of cases where the frequency content differs between conditioning train and test stimulus, as might happen in a real environment with a time-varying signal. The model of Djelani and Blauert (2002) passes the ear signals through a band-pass filter bank before activating the inhibition algorithm in its binaural processor. Thus, the build-up of inhibition in one critical band should not necessarily cause build-up in another one, unless that band receives the same reflection pattern. To fix this problem, binaural filters could be made much wider than peripheral auditory filters. In contrast, a model employing abstraction does not have a problem explaining this result. Room geometry is not frequency-dependent, and so a given spatiotemporal reflection pattern holds for any stimulus, regardless of its frequency content. Therefore, if the auditory system can build an internal geometric model of a room, it does not matter if there is a difference in the frequency content of the conditioning and test stimuli.

Another interesting question is whether conditioning trains that consist of clicks at only the lead or only the lag location can affect the echo threshold. In an abstraction process, one would not expect a single click to have an effect on echo thresholds, as a single click implies an anechoic environment with no room geometry information. In an adaptation process, it is, however, possible that a single click might reduce the sensitivity to other locations. This holds when inhibition would be expected to build up at locations other than the conditioning click location, albeit the Clifton effect shows direction specificity in the build-up.

Freyman et al. (1991) found that when only the lead or only the lag click pairs were presented in the conditioning train, it resulted in a reduced echo threshold for the test click pair, as compared to the case where no conditioning train was used. Thus, listeners were more attuned to reflections after hearing clicks at just the lead or just the lag location in the conditioning train. Freyman and Keen (2006) confirmed these findings and showed that the echo threshold even reduces to that of the single click baseline. In their experiment 3, the build-up click train was followed by a lead-only click train. Echo thresholds were reduced, but not to that without the build-up click train, indicating a partial break-down. When only a single lead- or lag-only click was inserted into the end of the build-up click train, echo thresholds were unaffected—a certain number of lead-only or lag-only clicks seems needed to disturb the build-up. A recent study by Bishop et al. (2014) also looked at the effects of lead- and lag-only clicks in the conditioning train. After the conditioning train, a 4-s test stimulus of click pairs (with time delays varying across trials) was played, and listeners were asked how many clicks at the lag location they heard. When a lead-alone conditioning train was used, approximately 9% more lag clicks were heard as compared to the case of a silent conditioning stimulus, whereas a lag-alone conditioning train resulted in approximately 7% fewer lag clicks being heard (averaged across all listeners and lead-lag time delay values). Thus, in this study, conditioning clicks at the lead location slightly increased the sensitivity to the lag location, in agreement with Freyman et al. (1991). However, the conditioning clicks at the lag location slightly decreased sensitivity to the lag location, in contrast to the results of Freyman et al. (1991). None of these results directly support either the adaptation or the abstraction hypothesis, but the absence of strong effects without a lag click pair being present agrees with the abstraction hypothesis.

One other question that might be asked is the following. Once an echo threshold has been increased for a specific lead-lag orientation, how long does it persist in periods of silence? In an abstraction process, it could persist indefinitely, as a period of silence would not tell the auditory system that the listener has moved to a different space unless one assumes a forgetting time constant for the surrounding room. Similarly, in an adaptation process, it would depend on the forgetting time constant of the inhibition algorithm. Keen and Freyman (2009) looked at the effect on echo thresholds of a test click pair when the conditioning click pair train was followed by a variable amount of silence. They found virtually no difference in echo threshold after up to 3 s of silence following the conditioning train compared to when the test click was presented immediately after the conditioning train. This finding could support either process and in the future longer periods of silence could be investigated to determine a value for the forgetting-time constant.

The underlying mechanisms of the precedence effect and its build-up with continuous stimuli are more difficult to ascertain. Various studies have shown that onsets and offsets are weighted more heavily, as these periods in time are thought to give the most unimpaired information about the locations of the direct sound and reflections (e.g., Houtgast and Aoki 1994; Stecker and Hafter 2002). This onset dominance can be used, for example, to improve spatial coding with cochlear implants (Monaghan and Seeber 2016). However, it is also known that localization dominance is caused by the temporally overlapping part of continuous noises (Dizon and Colburn 2006; Seeber 2011), suggesting that, generally, information from temporal modulations is used. How room abstraction and adaptation processes would function for ongoing stimuli is difficult to judge without further study. Generally, identifying the locations of individual reflections from ongoing stimuli either for spatially specific suppression of binaural cues of reflections in an adaptation process or for an abstraction of room dimensions from individual reflections remains an issue which is not yet well understood.

3 Break-Down

So far, the building up of room adaptation or abstraction through repeated exposure to a given reflection pattern has been discussed, and it has been shown that, once it has been built up, it can persist for several seconds. Several studies have investigated the reverse process, namely, whether the build-up state breaks down when introducing new stimuli after an initial conditioning train. During a typical day, a listener will move from space to space, and each new space will require an adjustment in terms of which directions are suppressed. This raises the question of how long echo suppression persists for the *previous* space. Does it disappear immediately, or does it persist for some time following exposure to a new pattern?

Djelani and Blauert (2001) investigated this idea by varying the spatial orientation of the 30 lead-lag conditioning click pairs, followed by a test-click pair. When the 30 conditioning click pairs and the test click pair had the same orientation, the echo thresholds were large, indicating a strong build-up. When the test click pair pointed in the mirrored orientation, echo thresholds were shorter. This is in agreement with the idea that *adaptation is direction specific*. If, however, the 30th conditioning click pair pointed to the mirrored location of the previous 29 conditioning clicks, thus indicating a new room configuration, the echo threshold for the test-click pair remained mostly unchanged with respect to the case where all conditioning clicks have the same orientation. In short, one click pair in a different orientation does not completely destroy the adaptation built up by the previous 29, indicating that break-down is not immediate (see also Freyman and Keen 2006).

It makes sense that, if the build-up of echo suppression is not immediate, the break-down should not be immediate either. In the Djelani and Blauert (2002) model, the inhibition algorithm builds up over time through regular and repeated triggering. Therefore, it exhibits the behavior of an integrator or moving average that is affected by past information. It will take a certain number of clicks (i.e. triggers) in the new spatial orientation to *flush out* all of the information from the previous room. In an abstraction process, it is also possible that information about a specific space is integrated over time. If the auditory system has accumulated a lot of information from one space, a new orientation of clicks could take time for the auditory system to resolve and to indicate that the listener has moved to a new environment, particularly in the absence of information from other senses such as vision or proprioception.

Flipping the orientation of the triangular test room in Djelani and Blauert (2001) not only mirrored the direction of the reflections, it also changed the level of each ear signal since individual head-related transfer functions were used which contain interaural level and time differences (ILDs and ITDs). Krumbholz and Nobbe (2002) showed that the ILD in a click pair is more potent for causing a break-down than the ITD, a result confirmed by Brown and Stecker (2013).

Keen and Freyman (2009) investigated the break-down with combinations of a "Room A", that is, a lead click on the left side followed by a lag click on the right side, and a special "Room B", which was just the lead click from the left side without the lag. A sequence of Room A click trains increased the echo thresholds for a test click in Room A as expected. If these Room A click trains were followed by an increasing number of clicks from Room B, the echo thresholds gradually decreased with the increasing number of clicks from Room B, eventually reaching the same echo threshold as seen with presentations of Room B clicks only. One interesting point is that it took eleven clicks from Room B to completely break down the build-up caused by five clicks from Room A, indicating the possibility of an asymmetry between the break-down and build-up processes.

While clicks are very useful stimuli in such studies, as they provide quantified amounts of information, many of the stimuli that are encountered in everyday life are more continuous in nature. Therefore, in order to claim that these effects can occur in real-world scenarios, it would be beneficial to see evidence that they also arise with non-transient stimuli. Adaptation and break-down effects were demonstrated for the Clifton effect paradigm with continuous stimuli, including speech and noise. For example, Djelani and Blauert (2001) presented listeners with a lead from 45° left and a lag from 45° right. After an initial period of at least 3 s in this spatial orientation, listeners pressed a button that caused lead and lag to switch sides immediately. Listeners then reported whether they perceived a temporarily enhanced echo, defined as an echo that either clearly diminishes or fuses with the direct sound again over time. For a train of 2-ms noise bursts at a rate of 4/s, the maximum percentage of temporarily enhanced echoes was reported at delays of 10–20 ms for all listeners. Results were very similar when a speech stimulus was used. However, for a continuous-noise stimulus, only three out of six listeners showed similar results, while two reported hardly any temporarily enhanced echoes and one listener was uncertain. Although noise proved to be a more difficult stimulus for some listeners to detect echoes in a break-down scenario, it is nonetheless clear that break-down does not only occur with clicks.

A study by Clifton et al. (1994) shed further light on the underlying processes. This study looked at discrimination suppression as a function of different lead-lag delays. Listeners were presented with a conditioning train of lead-lag click pairs at a given interstimulus interval, and then a test click pair whose interstimulus interval was varied. For all listeners, discrimination performance for the test click pair with the same lead-lag time delay as the conditioning train was worse, compared to the condition where the conditioning train was not played first, thus confirming the build-up. However, discrimination performance followed roughly a V-shaped pattern, that is, with the performance being worst when the lead-lag delay of the test click pair was identical to that used for the clicks in the conditioning train. Yet, performance improved again for both shorter and longer delays in the test click pair, indicating that discrimination suppression is delay specific. Note that the lag location was mostly unchanged in both situations. This is an interesting result in favor of the abstraction process. When the time delay between lead and lag changes, this could only be caused by a reflecting surface moving either towards or away from the listener, which would be a large change in room geometry. The temporal change would indicate to a stored room geometry model that a large room change has taken place, hence discrimination suppression is shortly reduced. However, in an adaptation process, the location (as implied by the interaural cues) of the lag is suppressed. Consequently, one would assume that only a change in the interstimulus interval does not have such a large, measurable effect on lag-discrimination suppression.

4 From Single Reflections to Room Reverberation

Most of the “rooms” discussed thus far consisted only of a direct sound and a single reflection, whereas typical spaces encountered every day produce a much greater number of reflections. While comparatively unexplored, recent work has begun to investigate build-up and break-down in more complex and realistic room models.

Djelani and Blauert (2001) mimicked the Clifton effect paradigm with a room with a triangular floorplan, which was simulated using the image-source method up to second order, yielding a room-impulse response with a direct sound and 12 reflections. The listeners scaled the size of the triangular room adaptively in order to determine a minimum room size at which the reflections were just barely audible. When the conditioning train and the test click both came from the triangular room in the same orientation, the room needed to be 2–3 times larger in order for reflections in the test click to be perceived, compared to situations where the conditioning room had no reflections or was a left-right flipped room. In all conditions, the source was always placed directly in front of the receiver, so that the leading stimulus was always in the same direction and diotic, whereas the reflection pattern switched ears when the room was flipped.

These results are readily explained by both an adaptation or an abstraction mechanism. While the flipped room will have the same timbral qualities and share the same room acoustic parameters, the spatial locations of the reflections change from the left to the right side and vice versa. This will swap reflections to previously unadapted locations, thereby increasing their audibility in terms of the adaptation model. If the auditory system is storing an abstract geometrical representation of the room, it should likewise be able to identify that the room geometry has changed and, consequently, trigger relearning. One other observation about this particular room is that, in the test orientation, most of the reflections came from the left hemifield, and in the flipped orientation from the right hemifield. Lag suppression could thus build up to one side primarily.

The question remains as to whether the directions of all reflections are suppressed, only some, or if perhaps suppression is weighted by relative amplitude or time after the direct sound. Stecker and Hafter (2002) found that in a click-train stimulus with short inter-click intervals (i.e. $<5\,\text{ms}$), the first click was the most strongly weighted perceptually. For longer click intervals, clicks were weighted equally, and when there was a gap in the click train, the first click following the gap was weighted more strongly than those around it. As reflections in a room impulse response are not equally spaced in time, the suppression of any one single reflection could be influenced by the gap in time between itself and the one immediately before it, or by deviation from periodicity.

Rather than looking just at the echo threshold, some recent studies have examined potential benefits of the room-acoustic context on the localization of the direct sound. In the dissertation by Sudirga (2014), localization accuracy was measured in "fixed" rooms, where all trials within a block were simulated from different source locations within the same virtual room, and in "mixed" rooms, where the simulated virtual room varied from trial to trial. There was no significant difference in absolute localization error between the mixed-room and fixed-room paradigms, but the variability in responses was somewhat lower in the fixed-room paradigm—namely, by 1.2–2.4°. In other words, when listeners remained in the same (virtual) room from trial to trial, their localization judgments did not necessarily match the true (i.e. physical) sound source location any better, but localization was more consistent from trial to trial.

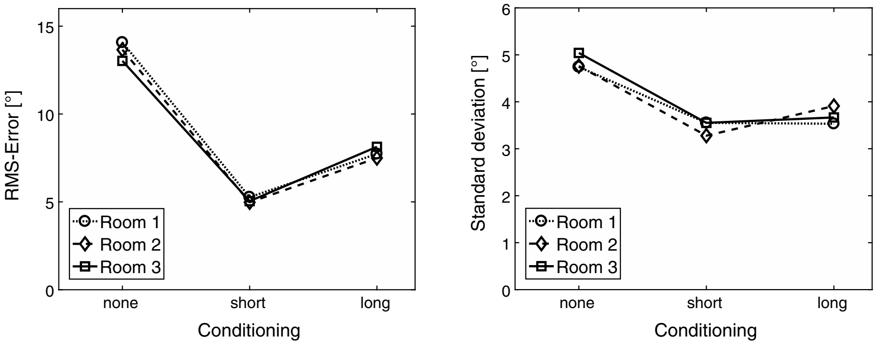

Fig. 4 RMS localization errors, **(left)**, and standard deviations **(right)**, averaged across eight listeners for localization of a click in virtual rooms (1, 2 or 3) when preceded either by silence (conditioning: none), or by a short or a long conditioning sequence of 2 or 14 clicks, respectively. The conditioning clicks were played from random locations left and right of the listeners, while the test clicks were restricted to the non-overlapping frontal region. The stimuli were played in loudspeaker-auralized virtual rooms composed of reflections up to a mirror-image order of 100. The three rooms varied in every trial to reduce across trial build-up. The results show a significant reduction of RMS error and standard deviation when a conditioning sequence preceded the test click. Since the conditioning clicks and the test clicks did not overlap in space, and since test conditions were otherwise identical in all conditions, the observed benefit can be attributed to the context built up by the conditioning clicks. Data replotted from Seeber et al. (2016)

Seeber et al. (2016) tested for direct evidence of room abstraction processes, again through the lens of localization in reverberant environments. Their study examined whether clicks from different source locations within a room improved localization from other source directions in that same room while the listener's location was kept constant. Such an explicit test for a transfer from one location to another one in the same room, if rendering positive results, could not be explained by an adaptation process because source and reflection positions change with each conditioning click and from the conditioning clicks to the test click. Instead, an improvement due to the room context would support the idea that the auditory system uses interaural cues to infer aspects of the geometry of the room.

In this experiment, a conditioning train of clicks was first played from random locations left and right of the listener in a virtual room. Then, a test click came from a location in the frontal region, and the listener had to indicate its perceived direction. Thus, the conditioning train never contained a click that occurred from the same location or even the region being tested. In this way, the listener has not been exposed to it before completing the localization task—a transfer due to context is required. Three different virtual rooms were used randomly from trial to trial to avoid adaptation from simply hearing an identical room over a block of trials. The test was performed in darkness, in an acoustically treated room.

Figure 4 gives results from eight listeners. Both RMS localization errors and standard deviations of the localization responses were significantly reduced when a conditioning train of clicks (either a short train of 2 clicks or a long train of 15 clicks) preceded the test click. These results suggest that the conditioning clicks and their reflections create a context in which localization accuracy and precision of subsequent stimuli is improved. Unique to this study is the fact that the conditioning clicks and the test clicks did not come from the same location, and thus a binaural adaptation process cannot explain the improvement. The improvement, therefore, requires a transfer which is in line with some kind of a room abstraction process, particularly as it extends to new locations. Alternately, the context could also be built up by the clicks irrespective of room reverberation, similar to audio-visual contextual effects due to a “visual frame of reference”—compare Radeau and Bertelson (1976). While more research remains to be done in this area, it has become clear that experiments involving more spatiotemporally complex stimuli will be useful to probe the complex processes involved in room adaptation and possibly abstraction in the auditory system.

5 Room Adaptation and Speech Understanding

Understanding speech in reverberant environments is one of the most important tasks performed by the auditory system and the focus of much research activity, particularly for hearing-impaired listeners. Therefore, it is certainly of interest how prolonged exposure to a specific space may improve speech understanding. Research into speech understanding in reverberant spaces has often been studied with respect to the amount of reverberant energy and the shape of the reverberation decay, rather than the specific orientation of direct sound and reflections. Several studies discussed here employ a paradigm with two spoken test words, [sir] and [stir], which differ by the phoneme /t/. In recognition experiments, the amount of reverberation affects the phoneme boundary. In fact, reverberation can make it difficult to differentiate between two similar spoken words because it fills in the gaps that are perceivable in non-reverberant listening conditions.

Watkins (2005) created a continuum of stimuli by interpolating the temporal envelopes of [sir] and [stir] and embedded the respective test words into the context of a sentence. To test the effect of reverberation, the amount of reverberation in the context sentence and for the test words was varied independently. Listeners were played the entire context sentence and then asked whether they heard either [sir] or [stir]. Increasing the reverberation of the test word relative to the surrounding sentence resulted in more [sir] identifications, as there was no opportunity to hear the short glottal stop before the phoneme /t/. When the reverberation of the context sentence was increased relative to the test word, [stir], the number of correct identifications increased. This was interpreted as the reverberation context help uncover the silence before the /t/.

In a subsequent study, Watkins and Makin (2007) used a similar experimental paradigm, but with a noise context rather than a speech context. The same pattern of [sir] versus [stir] identifications as in the previous study could be seen, provided that the noise was broadband and contained temporal modulation in the envelope (i.e.

pauses) that allowed the listener to judge the level of reverberation. Narrow-band noise contexts outside of the relevant frequency bands for [sir]/[stir] identifications did not induce reverberation compensation, even when they had the same temporal envelopes as the broadband noise contexts that did induce compensation. In other words, reverberation compensation appears to be specific with respect to the frequency band.

These results thus help clarify which components of the context are used by the listener to estimate reverberation. A study by Nielsen and Dau (2010) complicates the interpretation since a non-reverberant speech context resulted in more [sir] identifications while other non-speech contexts (including white noise, speech-shaped noise, and amplitude-modulated noise, in addition to silence) with no reverberation to it resulted in more [stir] identifications. This suggests that the non-reverberant speech context itself impedes the detection of the consonant /t/, and that changes in the modulation spectrum of the context are critical for inducing reverberation compensation.

Beeston (2014) has developed a peripheral auditory model to perform the [sir]/[stir]-identification task. The model adjusts efferent suppression in a closed feedback loop based on estimating the amount of reverberation in the pauses of the signal. The model is able to emulate human performance in a [sir]/[stir]-identification task, thereby giving evidence that simple adaptation of peripheral suppression is sufficient for reverberation compensation of speech.

While these studies have investigated the effect of reverberation on a specific phoneme boundary, Zahorik and collaborators have examined more generally how speech understanding in reverberant environments is affected by repeated exposure to the same room acoustics. Brandewie and Zahorik (2010) measured speech reception thresholds (SRTs) in three versions of the same simulated rectangular room, but with different surface materials, to yield strongly different reverberation times. Preceding the target phrase with a sentence carrier and presenting it in the same room showed a significant improvement in SRT by 2.7 dB compared to a no-carrier condition. This is an important result, as it shows that understanding of regular words in noise can be improved by prior exposure to the room. Follow-up experiments with anechoic or monaural stimuli showed a much smaller, non-significant improvement, equivalent to a signal-to-noise improvement of only 0.8 dB. This suggests that the underlying mechanism is binaural and involves a spatial compensation of room reverberation.

Srinivasan and Zahorik (2013) further investigated the time course of this improvement with an open speech corpus and without the sentence carrier. Five different simulated rooms, including an anechoic room, were used to generate the test stimuli, which were presented in either a “blocked” (a block of stimuli all from the same room) or “unblocked” condition (room varied from trial to trial). There was no significant difference in performance between the blocked and unblocked conditions for the anechoic room, while there was a significant improvement in the reverberant rooms when presented in the blocked format. However, no significant improvement was seen over the time course of a group of blocked trials. This suggests that the processes resulting in improved performance operate on fairly short time scales, namely, hundreds of milliseconds to seconds (i.e. within the length of one trial).

It is difficult to tell from these results whether they suggest an adaptation or an abstraction process. Despite their differences in wall absorption, all rooms shared the same geometry, thus leading to the same binaural reflection pattern. An adaptation process should thus suppress the binaural features of reflections similarly in all rooms and show only small recognition differences across conditions, which is in disagreement with present results. On the one hand, an adaptation process could explain the results if one was to postulate that the reflection energy would affect the amount of suppression, which does not easily agree with the above-discussed results on echo thresholds and discrimination suppression, as these are somewhat unaffected by reflection energy. On the other hand, an abstraction process could readily estimate the room geometry in all conditions but would, likewise, have to reset its room-configuration concept upon changes in the absorption characteristics.

6 Self-Motion

In natural settings, people are rarely completely motionless, as, for instance, with their head on a chin-rest. At almost any time, when entering a new space, people will move through a range of different positions within that space. In addition, they are also making small head movements, which have been shown to be important for spatial hearing, particularly with regard to resolving front-back confusions—see, for example, Wightman and Kistler (1999), Thurlow and Runge (1967). Furthermore, listeners often create their own sounds while exploring space, such as with speech or footsteps. When listeners have some sort of input into the room exploration process, this can have an effect on adapting to a new acoustic space versus what can be gleaned from being passively presented with a given stimulus.

Self-motion also has interesting implications for the differentiation of adaptation from abstraction processes. In particular, can the auditory system integrate proprioceptive and vestibular information to update its echo suppression? For instance, if a listener's auditory system has built up suppression to a lead at −45° left and a lag at 45° right, and they turn their head 15° to the left, will they then be adapted to a lead at −30° left and a lag at 60° right (in head-centered, i.e. binaural coordinates)? And if so, is this adaptation or abstraction, assuming that abstraction should be able to deal with all source positions and orientations within a given space? Answers to these specific questions have not yet been given in the literature, but related work on self-motion provides some clues.

Wightman and Kistler (1999) investigated the effect of head movements on resolving front-back confusions and confirmed that they do so. Interestingly, even with a fixed head, front-back confusions were resolved when listeners had control over the movement of the virtual source position (controlled via arrow keys on a computer keyboard), but not when the experimenter controlled source movements—despite both situations being acoustically identical. This suggests that having control over the position of the sound source must work in tandem with the acoustical signals reaching the two ears to show the improvement in spatial hearing. In a similar vein, Perrett and Noble (1997) showed that head movements also assist with vertical sound-source localization—compare Pastore et al. (2020), this volume.

Echolocation is most famously used by bats to navigate their physical surroundings, but blind and sighted humans can also utilize echolocation after some training. Echolocation requires the listener to produce their own sound stimuli, whereby the direct sound travels directly from the mouth to the ears, and the reflections return from surfaces in the environment. Therefore, it follows that listeners need to have low echo thresholds to echolocate, as the conscious perception of the echoes is what allows echolocation to work. Adapting to a reflection location and having its position suppressed would make it more difficult to echolocate. Thus it might not be surprising that training can affect the sensitivity to ITDs in the lagging sound (Saberi and Perrott 1990).

Wallmeier and Wiegrebe (2014) examined the role of self-motion and head movements in listeners trying to navigate a virtual corridor using only echolocation, with no visual cues. The listeners generated sound stimuli with their own mouths and heard the auralized response from a virtual corridor. They were then asked to orient themselves along the axis of the corridor. Listeners could adjust the orientation of the corridor virtually (like in a video game), adjust the physical orientation of the motorized chair in which they were sitting with head fixed, or adjust the motorized chair with head movements. The ability to move both the chair and one's head resulted in significantly better performance, indicating the importance of vestibular cues from the motion of the chair as well as from head movements.

This is a scenario where listeners are actually trying to avoid adaptation taking place since they need to be able to hear the echoes distinctly in order to complete the task. An "abstraction" of the room is what the listeners need in order to determine with some certainty where the walls are and in which direction the corridor leads. So, echolocation may be a special case, wherein higher-level cognitive processes attempt an abstraction process to learn the geometry of the space explicitly.

7 Conclusion

This chapter examined the literature for evidence of the processes involved in the build-up and break-down of the precedence effect, starting with pairs of leading and lagging clicks from different spatial locations. This literature overview was then expanded to stimuli with higher numbers of reflections, speech understanding in reverberant environments, and how self-motion can be incorporated into spatial hearing.

The current authors proposed and discussed two possible mechanisms for understanding how the auditory system builds up information to suppress reflections. The first and simpler mechanism has been termed "adaptation" and is affecting binaural and possibly also relevant monaural processes. When determining the location of leading and lagging signals in the binaural auditory system, binaural information from the lagging direction is increasingly suppressed, so that after a matter of several seconds, the listener only hears one sound event, rather than two, indicating a rise in the echo threshold. This paradigm has been modeled by Djelani and Blauert (2002) and been shown to predict the "Clifton Effect" (Clifton 1987). This binaural adaptation process might be supported by monaural adaptation processes occurring at the level of hair cells (Hartung and Trahiotis 2001), the cochlear nucleus (Buerck and van Hemmen 2007; Hafter 1996) or through the MOC-feedback loop (Beeston 2014).

The second mechanism, termed "abstraction", is more complex, but has the potential of being applied to dynamic complex listening scenarios. It is postulated that the auditory system can build up an abstract geometric model of an entire room and use this model to control reflection suppression. This model is predicted to survive, for instance, source and listener movements in the same room, wherein the reflection pattern changes at the listener's position, but in a way that is consistent with the room geometry. Controlled by proprioceptive information, the use of an abstract room representation avoids having the auditory system re-calibrate every time the reflection pattern shifts, as would have to be assumed for a pure adaptation process. Evidence for this mechanism comes from experiments showing improved localization ability for test positions not part of the exposure sequence but located in the same room (Seeber et al. 2016). A localization improvement of this kind would require a generalization or abstraction of information from the context given by the room rather than by individual reflections.

The work discussed in this chapter provides evidence for both processes in different scenarios. Some results could potentially be explained by either one. Generally speaking, echo suppression and build-up effects can be well explained by an adaptation process while context effects in localization, echolocation and speech understanding in reverberation may suggest an abstraction process. As future work will likely probe more complex and dynamic scenarios, it will help to disentangle the contribution of these different processes or, hopefully, to understand how they might work in tandem to aid listeners when exploring new acoustic environments.

Acknowledgements The present work was supported by the Bernstein Center for Computational Neuroscience Munich, BMBF 01 GQ 1004A and 01 GQ 1004B. SC received funding from the Elite Master Programme in Neuroengineering at the Technical University of Munich. The authors thank Dr. Piotr Majdak and two anonymous reviewers for valuable comments and suggestions.

References

Beeston, A.V. 2014. Perceptual compensation for reverberation in human listeners and machines. Ph.D. thesis, University of Sheffield, UK.

Bishop, C., D. Yadav, S. London, and L. Miller. 2014. The effects of preceding lead-alone and lag-alone click trains on the buildup of echo suppression. *The Journal of the Acoustical Society of America* 136 (2): 803–817.

Blauert, J. 1997. *Spatial Hearing*, 494. Cambridge, USA: MIT Press.

Brandewie, E., and P. Zahorik. 2010. Prior listening in rooms improves speech intelligibility. *The Journal of the Acoustical Society of America* 128 (1): 291–299.

Brown, A., and G. Stecker. 2013. The precedence effect: Fusion and lateralization measures for headphone stimuli lateralized by interaural time and level. *The Journal of the Acoustical Society of America* 133 (5): 2883–2898.

Brown, A., G. Stecker, and D. Tollin. 2015. The precedence effect in sound localization. *Journal of the Association for Research in Otolaryngology* 16: 1–28.

Buerck, M., and J. van Hemmen. 2007. Modeling the cochlear nucleus: A site for monaural echo suppression? *The Journal of the Acoustical Society of America* 122 (4): 2226–2235.

Clifton, R. 1987. Breakdown of echo suppression in the precedence effect. *The Journal of the Acoustical Society of America* 82 (5): 1834–1835.

Clifton, R., and R. Freyman. 1989. Effect of click rate and delay on breakdown of the precedence effect. *Perception & Psychophysics* 46 (2): 139–145.

Clifton, R., and R. Freyman. 1996. The precedence effect: Beyond echo suppression. In *Binaural and Spatial Hearing in Real and Virtual Environments*, ed. R. Gilkey, and T. Anderson, 233–255. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Clifton, R., R. Freyman, R. Litovsky, and D. McCall. 1994. Listeners' expectations about echoes can raise or lower echo threshold. *The Journal of the Acoustical Society of America* 95 (3): 1525–1533.

Dizon, R., and H. Colburn. 2006. The influence of spectral, temporal, and interaural stimulus variations on the precedence effect. *The Journal of the Acoustical Society of America* 119 (5): 2947–2964.

Djelani, T., and J. Blauert. 2001. Investigations into the build-up and breakdown of the precedence effect. *Acta Acustica united with Acustica* 87: 253–261.

Djelani, T., and J. Blauert. 2002. Modelling the direction-specific build-up of the precedence effect. In *3rd European Congress on Acoustics—Forum Acusticum*, Sevillia, Spain.

Freyman, R., and R. Keen. 2006. Constructing and disrupting listeners' models of auditory space. *The Journal of the Acoustical Society of America* 120 (6): 3957–3965.

Freyman, R., R. Clifton, and R. Litovsky. 1991. Dynamic processes in the precedence effect. *The Journal of the Acoustical Society of America* 90 (2): 874–884.

Hafter, E. 1996. Binaural adaptation and the effectiveness of a stimulus beyond its onset. In *Binaural and Spatial Hearing in Real and Virtual Environments*, ed. R. Gilkey, and T. Anderson, 211–232. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Hafter, E., and T. Buell. 1990. Restarting the adapted binaural system. *The Journal of the Acoustical Society of America* 88 (2): 806–812.

Hafter, E., and R. Dye. 1983. Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number. *The Journal of the Acoustical Society of America* 73 (2): 644–651.

Hartmann, W. 1983. Localization of sound in rooms. *The Journal of the Acoustical Society of America* 74 (5): 1380–1391.

Hartung, K., and C. Trahiotis. 2001. Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli. *The Journal of the Acoustical Society of America* 110 (3): 1505–1513.

Houtgast, T., and S. Aoki. 1994. Stimulus-onset dominance in the perception of binaural information. *Hearing Research* 72 (1–2): 29–36.

Keen, R., and R. Freyman. 2009. Release and re-buildup of listeners' models of auditory space. *The Journal of the Acoustical Society of America* 125 (5): 3243–3252.

Krumbholz, K., and A. Nobbe. 2002. Buildup and breakdown of echo suppression for stimuli presented over headphones—the effects of interaural time and level differences. *The Journal of the Acoustical Society of America* 112 (2): 654–663.

Lindemann, W. 1986a. Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front. *The Journal of the Acoustical Society of America* 80 (6): 1623–1630.

Lindemann, W. 1986b. Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *The Journal of the Acoustical Society of America* 80 (6): 1608–1622.

Litovsky, R., H. Colburn, W. Yost, and S. Gunzman. 1999. The precedence effect. *The Journal of the Acoustical Society of America* 106 (4): 1633–1654.

McFadden, D. 1973. Precedence effects and auditory cells with long characteristic delays. *The Journal of the Acoustical Society of America* 54: 528–530.

Menzer, F., and B. Seeber. 2014. Does reverberation perception differ in virtual spaces with unrealistic sound reflections? *Proceedings of Forum Acusticum*, 1–4, Krakow, European Acoustics Association.

Monaghan, J., and B. Seeber. 2016. A method to enhance the use of interaural time differences for cochlear implants in reverberant environments. *The Journal of the Acoustical Society of America* 140 (2): 1116–1129.

Nielsen, J., and T. Dau. 2010. Revisiting perceptual compensation for effects of reverberation in speech indentification. *The Journal of the Acoustical Society of America* 128 (5): 3088–3094.

Pastore, M., Y. Zhou, and W.A. Yost. 2020. Cross-modal and cognitive processes in sound localization. In *The Technology of Binaural Understanding*, eds. Blauert, J. and J. Braasch, 315–350. Cham, Switzerland: Springer and ASA Press.

Perrett, S., and W. Noble. 1997. The effect of head rotations on vertical plane sound localization. *The Journal of the Acoustical Society of America* 102 (4): 2325–2332.

Radeau, M., and P. Bertelson. 1976. The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception & Psychophysics* 20 (4): 227–235.

Saberi, K., and D. Perrott. 1990. Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. *The Journal of the Acoustical Society of America* 87 (4): 1732–1737.

Seeber, B. 2011. The contribution of intrinsic amplitude modulation to the precedence effect at high frequencies. In *Fortschritte der Akustik – DAGA'11*, ed. J. Becker-Schweitzer and G. Notbohm, Dt. Ges. f. Akustik e.V. (DEGA), 833–834, Berlin.

Seeber, B., and E. Hafter. 2011. Failure of the precedence effect with a noise-band vocoder. *The Journal of the Acoustical Society of America* 129 (3): 1509–1521.

Seeber, B., M. Mueller, and F. Menzer. 2016. Does learning a room's reflections aid spatial hearing? In *Proceedings of the 22nd International Congress on Acoustics*, ed. F. Miyara, E. Accolti, V. Pasch, and N. Vechiatti, Buenos Aires, Argentina, ICA2016–775, 1–8.

Srinivasan, N., and P. Zahorik. 2013. Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *The Journal of the Acoustical Society of America* 133 (1): EL33–39.

Stecker, G., and E. Hafter. 2002. Temporal weighting in sound localization. *The Journal of the Acoustical Society of America* 112 (3): 1046–1057.

Sudirga, R.E. 2014. Effect of reverberation context on spatial hearing performance of normal hearing listeners. Ph.D. thesis, University of Western Ontario, Canada.

Thurlow, W., and P. Runge. 1967. Effect of induced head movements on localization of direction of sounds. *The Journal of the Acoustical Society of America* 42 (2): 480–488.

Wallmeier, L., and L. Wiegrebe. 2014. Self-motion facilitates echo-acoustic orientation in humans. *Royal Society Open Science* 1: 140185.

Watkins, A. 2005. Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acustica united with Acustica* 91: 892–901.

Watkins, A., and S.J. Makin. 2007. Perceptual compensation for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts. *Acta Acustica united with Acustica* 93 (3): 403–410.

Wightman, F., and D. Kistler. 1999. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America* 105 (5): 2841–2853.

Yang, X., and D. Grantham. 1997. Cross-spectral and temporal factors in the precedence effect: Discrimination suppression of the lag sound in free-field. *The Journal of the Acoustical Society of America* 102 (5): 2973–2983.