Free-Field Virtual Psychoacoustics and Hearing Impairment:
Paper ICA2016-816

Are the precedence effect and spatial impression the
result of different auditory processes?

M. Torben Pastore\(^{(a)}\), Elizabeth Teret\(^{(b)}\), Brandon Cudequest\(^{(b)}\), Jonas Braasch\(^{(b)}\)

\(^{(a)}\)Rensselaer Polytechnic Institute, USA, m.torben.pastore@gmail.com
\(^{(c)}\)Rensselaer Polytechnic Institute, USA

Abstract
The sensation of auditory spaciousness is closely related to the pattern of reflections in a room, often described by a room impulse response (RIR). The early and late portions of RIRs are correspondingly the basis of most metrics of auditory spatial impression. Even though this pattern of reflections might reasonably be expected to indicate many sound source locations, listeners commonly localize sounds to their sources. This phenomenon, called the precedence effect (PE), is often thought to result from the suppression of reflections. This appears to present a paradox. How do we gain spatial information from reflections we are thought to suppress? Also, is this process different for early and late reverberation? In three parallel studies, we addressed this question by examining the specific roles of specular (early) and diffuse (usually late) reflections. The first experiment compared listener performance under conditions that elicited the precedence effect with diffuse or specular reflections. The second investigated listeners’ ability to match reverb times using different types of stimuli. The third compared apparent source width (ASW) resulting from physically wide sources to narrow sound sources with side reflections. Our results suggest that the PE, ASW, and listener envelopment are the result of closely related auditory processes. We also find that listeners do not appear to have a clear internal temporal representation of the decaying late reverb tail of a room impulse response. Possible implications for current metrics of spatial impression are considered.

Keywords: Precedence Effect; Auditory Source Width; Listener Envelopment; Spatial Impression)
Are the precedence effect and spatial impression the result of different auditory processes?

1 Introduction

In reverberant environments, multiple reflections off nearby surfaces present the auditory system with conflicting directional cues. If reflections arrive soon enough after the direct sound, listeners report hearing a single sound originating from the direction of the actual sound source. This phenomenon is commonly referred to as the precedence effect (PE) and shows that auditory temporal processing affects its spatial output. As the delays between the direct sound and its reflections increase, the perceived sound source widens until it breaks up into the direct sound and late arriving, distinct reflections. This delay, which varies between 4 ms and 80 ms depending on the stimulus, is called the echo threshold.

The echo threshold is implicitly (and sometimes explicitly, e.g., [1]) implicated in our current understanding of spatial impression. Most current understandings of spatial impression are based on the interaural decorrelation of early and late reflected sound energy. The sensation of auditory source width (ASW) is generally understood to be the result of decorrelated early reflections that arrive before the echo threshold, resulting in the perceived widening of a fused, localizable sound source. Listener envelopment (LEV), on the other hand, is usually attributed to diffuse late reverberation that arrives after the echo threshold, so that listeners feel surrounded by the sound even though they cannot localize it at its source. Listeners usually report that ASW and LEV can be experienced together or separately, but as separable percepts.

When considering spatial impression in light of what we know about the precedence effect, an apparent conundrum arises. While the precedence effect mechanism is often explained as some form of inhibition of reflections that arrive after the direct sound, spatial impression, especially ASW, is understood to be the result of listeners’ sensitivity to these same reflections. How do we extract spatial information from reflections we are supposedly suppressing? Also, the search for a definitive temporal division between early and late reflected energy that predicts ASW and LEV has been the focus of a great deal of research. For example, Dick and Vigeant, [2] recently showed that the temporal transition between RIR portions that elicit LEV and those that do not was extremely variable and that the transition existed along a logistic function, not precisely at 80 ms, as our standards would suggest.

At the heart of this inquiry lies an interesting challenge. Our current methods of studying and predicting spatial impression rely on the analysis of room impulse responses (RIR). This is a powerful method because rooms can be approximated as linear time-invariant (LTI) systems. Given the impulse response of a room, we can accurately predict the acoustic attributes of any signal played into that space by convolving the signal with the RIR. However, the mechanisms and performance of the auditory system are decidedly non-linear and vary greatly over time (e.g., neural adaptation, non-linear compression at the auditory nerve, auditory attention, the wide range of the echo threshold based on stimulus inputs, etc.). In using RIRs to predict per-
ceived attributes of room acoustic environments, we implicitly assume that listeners’ perception for any signal can be predicted based on convolution of that signal with the RIR. In so doing, we study and try to predict the performance of a non-LTI system using an LTI model.

Three recent experiments in our lab have prompted us to consider these issues based on results that do not easily fit with our current understanding of room acoustics. While these investigations were not directly aimed at answering the larger question of this paper, the results bear upon the underlying relation between the precedence effect and spatial impression and our current focus on the RIR to predict perceived spatial impression.

2 Experiment 1

The first experiment investigated the precedence effect for specular and diffuse reflections that arrive before the echo threshold. We hypothesized that if the PE is caused by active suppression, the temporal integrity of the reflections should not affect the dominance of the leading stimulus. To test this hypothesis, a simplified lead-lag paradigm was used to present stimuli over headphones for diffuse versus specular early reflections. The left panel of Figure 1 illustrates this testing paradigm. The direct sound coming from the speaker has a shorter path than the reflection of the side wall, so it arrives earlier. This stimulus condition is modeled as shown in the right side of the figure. The first arriving stimulus (lead) has an ITD of \(+300\) µs which is equal but opposite to the \(-300\) µs ITD of the later arriving stimulus (lag). The delay between the lead and lag was varied from 1–5 ms. Given that the direct-to-reverberant ratio in enclosed spaces can exceed 1, the precedence mechanism was also tested by increasing the level of the lag. Results for conditions where the ratio of intensities between the lead and lag were 0 dB, 4 dB, and 8 dB are shown here.

The stimulus was a 200-ms duration Gaussian noise with a center frequency of 500 Hz and a bandwidth of 800 Hz. For the ‘specular’ stimulus condition, the lag stimulus was a delayed copy of the lead stimulus, as in [5] and [6]. For the ‘diffuse’ condition, the lag was convolved with a Hanning-windowed 2-ms Gaussian noise. For the ‘diffuse correlated’ lag stimuli were identical in both left and right channels. For the ‘diffuse decorrelated’ condition, different Gaussian
noises were used for the Hanning-windowed diffusion filter, so that the lag was decorrelated between the left and right channels. The resulting lag stimulus can be thought of as many copies of the original lag stimulus superposed at many different time delays and amplitudes, similar to late reverberation, but presented before the echo threshold instead of after it.

Figure 2: The averaged normalized performance of eight subjects lateralizing 800-Hz bandwidth, 200-ms duration noise bursts in the presence of one modeled specular (black) or diffuse (red) reflection. Each panel shows listener performance grouped by lag level, shown at the top. The delay between lead and lag, in ms, is shown along the x-axis of each panel. o’s indicate the average result across all listeners and presentations, error bars show the standard error of the mean. Dashed lines indicate the results of the reference condition tests, where only a single noise was presented with an ITD corresponding to the lead, the midline, and lag. Auditory objects lateralized fully to the perceived position of the lead would be shown on the dotted line labelled ‘lead.’ Lateralization towards the midline and lag positions are indicated similarly. After [4].

Figure 2 shows the average response of eight listeners and error bars show the standard error of the mean. For lead-lag delays of 1–5 ms and for all three stimulus conditions, listeners demonstrated localization dominance. For the specular condition, listeners’ perceived lateralization shows an oscillatory pattern across lead-lag delays, which was shown in [5] to be likely the result of ILDs created by differences in comb filtering between lead and lag stimuli in the left and right channels. For the diffuse lag conditions, however, show no such pattern across lead-lag delays emerged. What is clear from these results is that the degree to which the lead dominated listeners’ perceived lateralization was dependent on the temporal features of the lag. This further suggests that an understanding of the PE based on active suppression of the lag will not describe these results. Rather, it appears that processing of the lead-lag stimulus pairs includes both lead and lag.

Interestingly, there was no appreciable difference in average listener performance between the interaurally correlated and decorrelated diffuse lag stimulus conditions. However, variability was greater for the decorrelated diffuse condition, suggesting the decorrelation of the lag did have some effect on listeners’ perceived auditory events.

What is of particular interest to this discussion, however, was revealed in exit interviews. For
the specular condition, listeners reported percepts similar to what we might expect for ASW: a localizable, horizontally broadened percept. The broadening of the percepts could also be observed in the variability of listeners’ responses. However, for the diffuse stimulus conditions, many listeners reported a fairly punctate image that was easily localizable surrounded by a ‘cloud’ of sound that was essentially ‘everywhere.’

Our headphone experiment did not use head-related transfer functions, so the exact relation between the percept listeners reported and ASW and LEV should be approached with caution. Nevertheless, the results of this experiment prompted us to consider an unanticipated possibility. What if the transition from ASW to LEV is not directly based on the relative temporal location in the room impulse response of the ‘early’ and ‘late’ reverberation but rather the degree to which reflections are diffuse. An look for a time range within the IR to identify the transition between ASW and LEV, we might instead consider the transition from specular reflections to the point where reverberation is diffuse to the point of being beyond the spatial acuity of the auditory system. In this case, the time it takes for reverberation to become approximately diffuse, the mixing time, of a room might be worth considering, especially as the size and shape of performance venues differ more and more from traditional concert halls.

3 Experiment 2

As mentioned above in the Introduction, many of our attempts to predict listeners’ perception of acoustic environments are based upon measured RIRs. This implicitly suggests that we would expect listeners to perceive different signals that were convolved with the same RIR to have equivalent perceived sound attributes. For example, we might expect listeners’ perceived reverb time to be based on the RIR, and therefore the same across different signals. One way to consider this proposition is to ask if listeners can in fact extract the impulse response of a room regardless of the signal that is played into it. If the diffuse reverberation is truly beyond the ability of the auditory system to fully process it, we might expect that the IR would be impossible to deconvolve from different signals, and so listeners would not be able to match the duration of the IR when different signals were convolved with the same IR.

For a first approximation to this answer, we turn now to results from another experiment in our lab. Our initial intent with this experiment was to investigate the extent to which impulse responses can be used to predict subjective experience across different sound stimuli is examined with a series of perceptual tests.

Listeners were presented with a signal that was convolved with a simulated room impulse response (IR). The IR consisted of an impulse followed by a 20 ms gap, and then Gaussian noise to model late reverberation. The IR was then windowed to have an exponential decay corresponding to one of a range of reverberation times (RT) from 1 s to 2 s in steps of 0.25 s. Listeners were asked to adjust their perceived reverberation time of a long-duration orchestral sample with no clear onset of offset to match their perceived RT of the reference. The listener-adjustable signal was convolved with the same set of IRs as in the reference condition, but with a range that extended 0.25 s shorter and longer than the presented range, was available for the adjustable stimulus.
Figure 5 shows mean results for Experiment 2, with the standard deviation indicated by vertical lines. The listener-adjustable stimulus is held constant and indicated above each figure panel. The x-axis shows the signal that was used for the reference stimulus. The data show that the source signal had an influence on perceived reverberation. Listeners demonstrated less accuracy and increased variability when matching reverberation times between different signals as compared to when they matched RTs across identical signals. This suggests that the auditory system does not have a well-developed temporal representation of the diffuse reverb tail. Further, any spatial information contained in the dichotic stimuli did not lead to any real improvement in listeners’ ability to match RTs across signals, suggesting that any spatial information included in binaural cues was not helpful to the task.

4 Experiment 3

The third experiment investigated the effect of the physical width of a sound source (PSW) on its perceived auditory source width (ASW). This was prompted by the traditional practice of placing a mono sound source on stage and predicting ASW based on the amount of lateral energy from sidewall reflections, based on experiments by Barron and Marshall [8]. We noted that, while the traditional ASW paradigm essentially sets up a precedence effect scenario, actual performances in concert halls most often involve musical ensembles that extend across
the stage, presenting a widely decorrelated sound field before side reflections are even considered. Given that this decorrelated sound source would arrive before sidewall reflections, we would further expect the effect of these later arriving reflections on ASW to be less than that of the direct sound. We therefore hypothesized that the traditional paradigm would require very low direct-to-reverberant ratios, i.e., high reverberant energies, to approximate the same ASW listeners would perceive when presented with a physically broad sound source.

![Schematic of the two stimulus conditions. See text for description. After [9].](image)

Figure 4: Schematic of the two stimulus conditions. See text for description. After [9].

The left side of Figure 4 shows the presented reference stimulus. A stereo clip of an orchestral motif was played over two speakers spaced at one of three azimuthal PSW angles: ±11.25°, ±22.5°, ±33.75°. The right side of Figure 4 shows the listener-adjustable stimulus. The same orchestral sample as in the reference condition was played from the center speaker and two speakers at ±45° played modeled reflections delayed by 28 ms and 32 ms, respectively. Listeners adjusted the level of the two modeled reflections together to match their perceived ASW for the reference condition.

The results show that as the width of the PSW condition increased, listeners required more reflected energy to perceived equivalent width. A plausible explanation is that in the listener-adjusted condition, the precedence effect minimizes to some extents the contribution of side reflections on the spatial range of a perceived event, rather requires more energy from the sides to broad the range. By contrast, the stereo presentation in the reference condition delivered balanced stimuli and thus evoke minimum precedence effect.
Exit interviews revealed that, for the ±33° reference stimulus, listeners found that in order to match the perceived width of the reference condition, they had to increase the level of reflected energy to the point where they perceived the sound coming from the sides, indicating that the PE had broken down because of the high lag level. This suggests that side-wall reflections would not give the same ASW as a physically broad sound source, and that the width onstage of the anticipated ensembles is important to consider in predicting ASW.

An additional experiment added modeled sidewall reflections to the reference condition. Results (not shown) indicated a very similar trend to the first experiment; listeners needed to increase the level of reflections for the mono source to levels well above that of the reference condition. Taken together, the results of these two experiments suggest a reconsideration of the current focus on sidewall reflections as the primary source of ASW in concert halls.

5 Conclusion

The three experiments discussed in this paper suggest a strong connection to the precedence effect, as already identified in the literature. However, the results of experiments 1 and 3 suggest that the dominance, in terms of localization, of the direct sound over reflections is not complete. That is, the characteristics of the lag stimuli, such as temporal coherence, have an effect on listeners' perception of the sound source that is intrinsic to the precedence effect. Consideration of this phenomenon could be pursued further with binaural precedence modeling in the future.

The results of experiment 2 suggest that listeners do not have the ability to deconvolve a room
impulse response from any signal, and so additional modeling of auditory perception is likely to be beneficial to attempts to predict listener outcomes for the built environment, including concert halls.

We propose that while RIRs describe room acoustics very successfully, their predictive power is greatly reduced once the signal gets past the ear canal. Since spatial impression is the product of both room acoustics and human auditory perception, the addition of auditory modeling to our current approach may offer deeper insights and greater predictive power. Auditory models exist (e.g. [10] and [11]) that predict listener performance under precedence conditions and respond to different stimuli differently. Such models could be combined with impulse response measurements created with different degrees of spatial extent onstage to give an enhanced understanding of the phenomena of spatial impression and improved predictions.

Acknowledgements

This research was funded in part by a Rensselaer HASS Fellowship. Brandon Cudequest is now at McKay, Conant and Hoover. Elizabeth Teret is now an independent acoustical consultant in Gardiner, Maine.

References


