How level, delay, and spatial separation influence the echo threshold

Korbinian Wegler[1], Florian Wendt[1,2], Robert Höldrich[1,2]
[1] *University of Music and Performing Arts Graz, Austria*
[2] *Institute of Electronic Music and Acoustics*
Email: wegler@student.tugraz.at, wendt@iem.at, hoeldrich@iem.at

Introduction

The echo threshold is a common measure to quantify the strength of the precedence effect. Depending on the experimental design, it is defined as the delay or level of the reflected sound compared to the direct sound at which the reflection is perceived as a separate auditory event. Respective studies are derived from a setup consisting of a sound source in the free field and a specularly reflecting vertical wall, cf. Figure 1. The typical paradigm is to determine the delay of the lagging reflection with a fixed level or - vice versa - to determine its level with a fixed delay. The spatial separation of lead and lag is known as a third setup parameter influencing the echo threshold, e.g. [1]. Nevertheless, most studies disregard this parameter and playback directions of lead and lag are kept constant during the experiment (see [2] for a thorough review).

This contribution studies the influence of delay, level, and spatial separation of the reflection on the echo threshold. Obviously, three parameters increase the number of possible conditions, which makes a systematical investigation difficult. Therefore, two listening experiments are conducted. Results of the first experiment define significant ranges of delay, level, and spatial separation for the tested setup. These ranges are then examined by the second experiment in a multi-stage paradigm. In each stage another parameter is varied and the results are input to the subsequent stage.

Measuring the echo threshold

The echo threshold is subject to many studies on the precedence effect. And almost as numerous as the studies itself is the variety of setup parameters, stimuli, playback methods, and measurement procedures used to quantify echo thresholds with listening experiments.

Most studies use two brief transient stimuli, as the precedence effect is known to be most active for transient onsets, e.g. [3].
In some early studies, planar wall panels were used for creating reflections, e.g. [4], but in common stimulus paradigms both sound stimuli are played back over loudspeakers or headphones. For the latter we distinguish between different simulations of the free-field stimuli. Interaural difference stimuli regard either only temporal (ITD) or intensity (ILD) aspects to produce lateralization of the lead and lag. These paradigms are rather artificial, as stimuli of the real word are always a combination of both ITD and ILD. Therefore, in a more natural headphone paradigm the stimuli are filtered with head-related transfer functions (HRTFs) for the desired lead and lag virtual locations. In this way ITD, ILD, and even spectral cues are regarded by the rendering.
Regarding measurement procedure, echo thresholds are attained directly or indirectly. For the comparatively fast direct procedure, listeners are given direct control over a setup parameter, e.g. the level of the reflection, and adjust it to match their criterion, e.g. [5]. For indirect paradigms listeners are typically instructed to report either the number of sounds perceived, or to localize, lateralize, or discriminate the location or directionality of the reflection, e.g. [6].

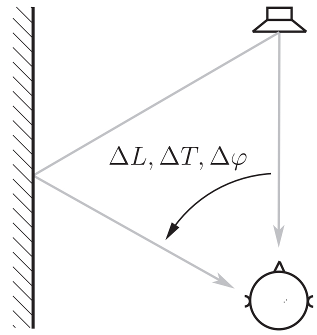


Figure 1: Illustration of a typical setup used to study the echo threshold. Compared to the direct sound, the specular reflection arriving at the listener's ears is attenuated by ΔL, delayed by ΔT, and spatially separated by $\Delta\varphi$.

Experiment 1

The first listening experiment studies significant ranges of delay, level, and spatial separation for the echo threshold, defined as the reflection level ΔL as a function of reflections' delay and direction. An ongoing sequence of 10 ms-long white noise bursts (instant on- and offset) with a period of 300 ms was used as stimulus. Playback employed headphones and direct sound and reflection were simulated by filtering the stimulus with corresponding HRTFs of the Neumann KU100 dummy head [7]. The direct sound was simulated always directly in front of the listener ($\varphi = 0°$) and reflection angles varied in 15°-steps from $\varphi = 0° \ldots \pm 90°$ (left/right). The listeners were given control over the level of the lagging copy. Starting from -50 dB relative to the direct sound, their task was to adjust the level in 1 dB-increments until the reflection was just barely audible as an image distinct from the direct sound.

The level of the direct sound was fixed with $L = 54\,\mathrm{dB(A)}_{\mathrm{eq}}$ to ensure good audibility while minimizing fatigue of the listeners. To support the buildup of the precedence effect [8], the level of the stimuli sequence was faded in by 2 s at the beginning of each condition. Measured delays were 20 ms, 40 ms, and 60 ms. In this way listeners had to adjust the reflection level of $3_{\,\mathrm{delays}} \times 7_{\,\mathrm{directions}} \times 2_{\,\mathrm{sides}} = 42$ conditions, presented as individual random permutation.

Playback employed a Beyerdynamic DT770 pro headphone with a Focusrite Scarlett 2i2 audio interface. Thirteen experienced listeners (one female, twelve male) participated in the experiment with a mean duration of 25 minutes.

Experimental results

The evaluation of the raw data of Experiment 1 revealed the criterion for hearing an echo to be highly individual as individual echo thresholds for the same condition spread over 30 dB. To compensate for this individuality, the results were normalized. Similar to the normalization process in [9], each level of a condition ΔL_i in dB adjusted by the ith listener is corrected by subtracting the individual mean level $\Delta \bar{L}_i$ over all conditions and adding the mean level $\Delta \bar{L}_{\mathrm{all}}$ calculated over all listeners and conditions:

$$\Delta L_{i,\mathrm{norm}} = \Delta L_i - \Delta \bar{L}_i + \Delta \bar{L}_{\mathrm{all}}. \tag{1}$$

A statistical analysis of the data reveals the reflection side not to be a significant parameter (t-test: $p = 0.43$) and the obtained results from left and right are pooled. Dashed lines in Figure 2 show normalized and pooled reflection levels as mean values and corresponding 95%-confidence intervals. It can be seen that the echo thresholds ΔL decrease progressively with increasing delay ΔT and resemble those obtained by similar studies, e.g. [5]. An analysis of variance (ANOVA) reveals the delay to be a significant parameter ($p \ll 0.01$) for all 7 reflection directions. The reflection direction $\Delta\varphi$ on the other hand does not yield monotone trends. Although it is a significant paramter for $\Delta T = 40\,\mathrm{ms}$ and $\Delta T = 60\,\mathrm{ms}$ ($p < 0.01$), Tukey's range test reveals that there are only 2 significant groups: $\Delta\varphi = 0°$ and $\Delta\varphi \geq 30°$. For $\Delta T = 20\,\mathrm{ms}$ no significance is obtained.

Experiment 2

Results from Experiment 1 suggest that measuring the echo threshold by adjusting the reflection direction $\Delta\varphi$ might be difficult. An informal listening test by the authors supported this hypothesis and by varying solely the reflection angle, while keeping constant its delay and level, it was not possible to measure the echo threshold. Therefore the Experiment 2 is conducted in two stages: Similar to Experiment 1, the first stage of Experiment 2 measured the echo threshold defined as the reflection level for 4 delays (20 ms, 40 ms, 60 ms, 80 ms) and three reflection directions (0°, 30°, 60°). Subsequently in the second stage, previously obtained individual reflection levels were used to measure the echo threshold defined as the reflection delay.

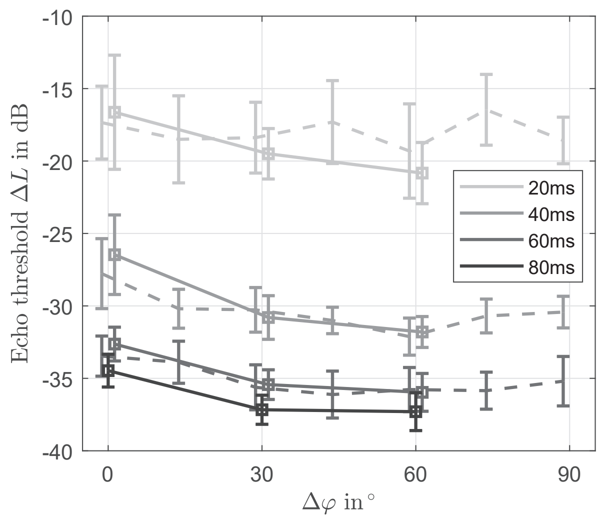


Figure 2: Means and corresponding 95%-confidence intervals of the echo threshold defined as reflection level ΔL for different delays ΔT and reflection directions $\Delta\varphi$ obtained by Experiment 1 (dashed lines) and Experiment 2 (solid lines).

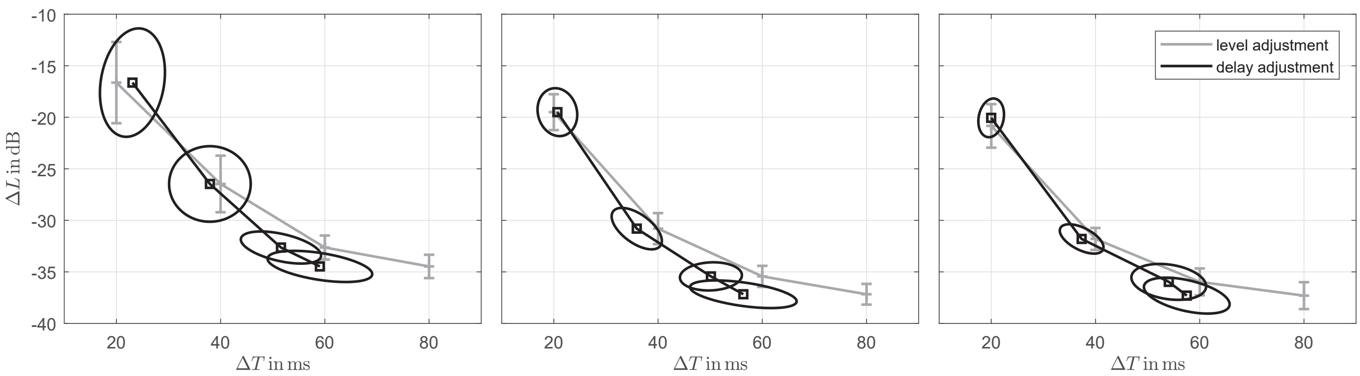
The level adjustment of Experiment 2 (first stage) was conducted similar than Experiment 1. Conditions of the delay adjustment (second stage) started with an initial value of $\Delta T = 5\,\mathrm{ms}$ relative to the direct sound, which could be incremented in 10%-steps.

Fifteen listeners (one female, fourteen male; including all listeners from Experiment 1) participated in the experiment and $4_{\,\mathrm{delays}} \times 3_{\,\mathrm{directions}} \times 2_{\,\mathrm{sides}} \times 2_{\,\mathrm{stages}} = 48$ conditions were evaluated.

Experimental results

Reflection levels of the first stage of Experiment 2 are normalized according to Eq. (1) and are given in Figure 2 as means and 95%-confidence intervals by solid lines. Unsurprisingly, the results strongly resemble corresponding levels of Experiment 1. Similarly, the delay is found to be significant ($p \ll 0.01$) for all directions, whereas the direction is significant only for $\Delta T \geq 40\,\mathrm{ms}$ (2 groups: $\Delta\varphi = 0°$ and $\Delta\varphi \geq 30°$).

The results of both stages are given in Figure 3 for reflection directions $\Delta\varphi$. Obtained levels ΔL of the first stage are plotted as means and 95%-confidence intervals for fixed delays 20 ms, 40 ms, 60 ms, and 80 ms. Obtained delays of the second stage on the other hand were measured with individual levels. For a compact representation of individual data points, 2-dimensional mean values with the corresponding 95%-confidence ellipses are used, cf. [10]. Half-axes of the 95%-confidence ellipses are mostly parallel to the figures axes, indicating the reflections delay and level to be independent from each other. Agreeing with the listeners comments, the higher answer spreads in Figure 3 (a) suggest that it is more difficult to determine if the reflection is considered as a separate auditory event, if direct sound and reflection arise from the same direction. Ideally, the curves of both stages would coincide. This is not the case and a systematic shift towards shorter delays is seen for the second stage for all directions $\Delta\varphi$.



(a) $\Delta\varphi = 0°$ (b) $\Delta\varphi = 30°$ (c) $\Delta\varphi = 60°$

Figure 3: Results of the Experiment 2 for directions $\Delta\varphi$. Means and corresponding 95%-confidence intervals of the adjusted level ΔL (first stage) are shown in gray; means and corresponding 95%-confidence ellipses of the adjusted delay ΔT (second stage) are shown in black.

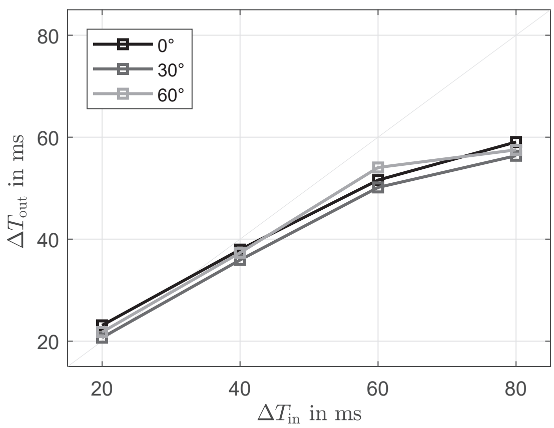


Figure 4: Mean delays ΔT_{out} adjusted in the second stage using individual levels over the delay ΔT_{in} used in the first stage to measure corresponding levels.

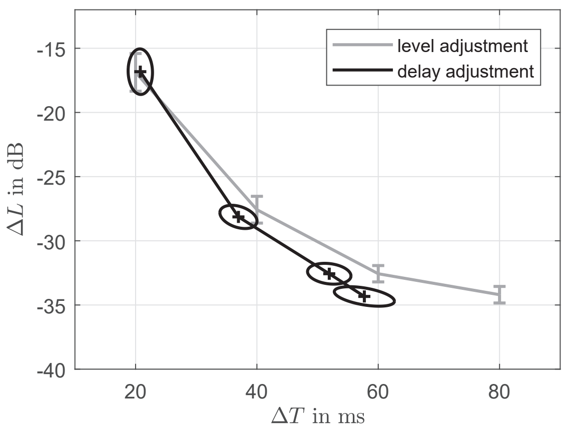


Figure 5: Combined representation of the echo thresholds of Experiment 2 shown as means and corresponding 95%-confidence intervals (gray, first stage) or 95%-confidence ellipses (black, second stage). The influence of the reflection direction is considered by $-3.6\,\text{dB} \cdot \sin(\Delta\varphi)$.

Moreover corresponding ellipses spread more along the ΔT-axis, suggesting that it is more difficult to adjust the delay for small levels.

An alternative representation of deviating curves is given in Figure 4, showing the direct relation between the delay ΔT_{in} used in the first stage and corresponding mean delays ΔT_{out} adjusted by the listeners in the second stage. Deviations are similar for all reflection angles, which allows modeling as a joint representation of the data of Experiment 2. The monotonic influence of reflection angle $\Delta\varphi$ is considered by a sine function, yielding an extension of Eq. (1):

$$\Delta L_{i,\text{norm}} = \Delta L_i - \Delta\bar{L}_i + \Delta\bar{L}_{\text{all}} + a \cdot \sin(\Delta\varphi). \qquad (2)$$

The parameter $a = -3.6\,\text{dB}$ is obtained by fitting the data for different directions using the least squares method and yields a combined representation of the data, cf. Figure 5. Obviously, by combining the data, sizes of confidence intervals and confidence ellipses shrink. Output delays of the second stage ΔT_{out} are found to be significantly ($p \leq 0.02$) lower than corresponding input delays of the first stage ΔT_{in} for $\Delta T \geq 40\,\text{ms}$.

Discussion and conclusion

The influence of level, delay and, spatial separation on the echo threshold was considered in this study. Two listening experiments were conducted. With Experiment 1 we could show that all three parameters can significantly influence the echo threshold. However, the spatial separation does not yield a monotone echo threshold and significance was obtained only between $\Delta\varphi = 0°$ and $\Delta\varphi \geq 30°$ for delays $\Delta T \geq 40\,\text{ms}$. Consequently, we were not able to directly measure the echo threshold by means of the reflection direction, while keeping constant the reflections delay and level.

Experiment 2 examined the influence of the reflections level and delay in a multi-stage paradigm. In the first stage of the experiment, listeners were instructed to adjust the level for fixed delays. The results were then input to the second stage and listeners had to adjust the reflections delay for individual levels obtained from the first stage. Ideally, both input and output delays would be identical. However, although level and delay were found to be independent parameters, the adjusted delays of the second stage were significantly lower than the input delays of the first stage for $\Delta T \geq 40\,\text{ms}$. This suggests that the criterion defining the echo threshold varied over time similarly for all listeners and was looser in the second stage: compared to the first stage less delay was needed to perceive the reflection as a separate auditory image. The explanation therefore could be an increasing sensitivity of the listeners to perceive the echo. A straightforward method to measure the echo threshold and simultaneously bypass the listeners criterion is not possible, neither directly nor indirectly.

Acknowledgments

The authors thank all listeners taking part in the experiments.

References

[1] B. Shinn-Cunningham and P. Zurek, and N. Durlach, "Adjustment and discrimination measurements of the precedence effect," in *The Journal of the Acoustical Society of America*, 93, 2923-32, 1993.

[2] A. Brown, C. Stecker, and D. Tollin, "The Precedence Effect in Sound Localization," in *Journal of the Association for Research in Otolaryngology*, 16, 10.1007/s10162-014-0496-2, 2014.

[3] Rakerd, Brad and Hartmann, William, "Localization of sound in rooms, III: Onset and duration effects," in *The Journal of the Acoustical Society of America*, 80. 1695-706. 10.11211.394282, 1987.

[4] B. Rakerd, and W. Hartmann, "Localization of sound in rooms, II: The effects of a single reflecting surface," in *The Journal of the Acoustical Society of America*, 78, 524-33, 10.11211.392474, 1085.

[5] B. Rakerd, W. Hartmann, and J. Hsu, "Echo suppression in the horizontal and median sagittal planes," in *The Journal of the Acoustical Society of America*, 107. 1061-4. 10.11211.428287, 2000.

[6] R. Litovsky and B. Shinn-Cunningham, "Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression," in *The Journal of the Acoustical Society of America*, 109, 1, 10.1121/1.1328792, 2001.

[7] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Fortschritte der Akustik*, Merano, Italy, 2013.

[8] T. Djelani and J. Blauert, "Investigations into the Build-up and Breakdown of the Precedence Effect," in *Acta Acustica united with Acustica*, 87, 253-261, 2001.

[9] P. W. Robinson, A. Walther, C. Faller, and J. Braasch, "Echo thresholds for reflections from acoustically diffusive architectural surfaces," in *The Journal of the Acoustical Society of America*, 134, 2755-2764 10.1121 1.4820890, 2013.

[10] F. Wendt and M. Frank, "On the localization of auditory objects created by directional sound sources in a virtual room," in *Proceedings of the 30th Tonmeistertagung*, Cologne, 2018.