
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Lokki, Tapio; Pätynen, Jukka; Tervo, Sakari; Siltanen, Samuel; Savioja, Lauri
Engaging concert hall acoustics is made up of temporal envelope preserving reflections

Published in:
Journal of the Acoustical Society of America

DOI:
[10.1121/1.3579145](https://doi.org/10.1121/1.3579145)

Published: 01/01/2011

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Lokki, T., Pätynen, J., Tervo, S., Siltanen, S., & Savioja, L. (2011). Engaging concert hall acoustics is made up of temporal envelope preserving reflections. *Journal of the Acoustical Society of America*, 129(6), EL223-EL228. <https://doi.org/10.1121/1.3579145>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Engaging concert hall acoustics is made up of temporal envelope preserving reflections

Tapio Lokki,^{a)} Jukka Pätynen, Sakari Tervo,
Samuel Siltanen, and Lauri Savioja

Department of Media Technology, Aalto University School of Science, FI-00067 Aalto, Finland
Tapio.Lokki@aalto.fi, Jukka.Patynen@aalto.fi, Sakari.Tervo@aalto.fi,
Samuel.Siltanen@aalto.fi, Lauri.Savioja@aalto.fi

Abstract: Strong, exciting, and engaging sound is perceived in the best concert halls. Here, it is shown that wideband early reflections that preserve the temporal envelope of sound contribute to the clear and open acoustics with strong bass. Such reflections are fused with the direct sound due to the precedence effect. In contrast, reflections that distort the temporal envelope render the sound weak and muddy because they partially break down the precedence. The presented findings are based on the earlier psychoacoustics research, and confirmed by a perceptual evaluation with six simulated concert halls that have same monaural room acoustical parameter values according to ISO3382-1.

© 2011 Acoustical Society of America

PACS numbers: 43.55.Hy, 43.55.Fw, 43.66.Jh, 43.66.Nm [NX]

Date Received: March 05, 2011 Date Accepted: March 25, 2011

1. Introduction

A good concert hall amplifies sound so that in most of the seats music is clear, intimate, and engaging. Even in the back of the seating area the listener can distinguish different instruments and hear their individual contribution to the musical ensemble, even if the reflections are the dominant energy contribution of the sound field. The reflections contribute to the loudness of the sound and they envelope the listener, giving an impression of reverberation. The perception of loudness and reverberation can be predicted well with the standardized room acoustics parameters. However, there are remarkable perceptual differences between halls having the same objective measures.

This Letter proposes that a concert hall provides sound with strong bass and good clarity if the early reflections do not absorb sound and they preserve the temporal envelope of a sound signal. Moreover, the reflected sound has to reach the listener from another direction than the direct sound for open and enveloping sound. The perceptual consequences of the directions and types of early reflections that reach the listener within 100 ms after the direct sound—an approximate duration of 1/16 note in *Allegro* tempo, is explained. In addition, the effects of different early reflections are demonstrated with virtual acoustics in [Mm. 1](#).

[Mm. 1](#). Binaural demonstration video. This is a file of type “mpg” (7.1 MB).

2. Early reflections and precedence effect

Human hearing cannot distinguish reflections arriving shortly after the direct sound.¹ Instead, these reflections are fused with the direct sound and they do not change the perception of the sound source direction.² These properties of human hearing are explained by the precedence effect.³ With the sound of a click the effect lingers for less than 10 ms after the direct sound. However, for signals starting with a transient and followed by a steady state, such as speech or music, the precedence effect window is longer, 30 ms on average.¹ Psychoacoustics studies with clicks or noise bursts show that reflections after

^{a)} Author to whom correspondence should be addressed.

the precedence effect window are heard as echoes. However, in concert halls, several consecutive reflections follow the direct sound and no echoes are heard if a reflection is not considerably louder than the others.³ Thus, with music, the early reflections can be fused with the direct sound for longer than 30 ms after the direct sound. Such extended precedence effect has been reported when one click was added between the lead and lag clicks⁴ and when several reflections are present at fixed delay times.⁵

Audio engineers make the sound more powerful and rich by adding delayed copies of the sound within the precedence window.⁶ Delayed copies of the sound correspond to reflections from any direction. Copies of the sound increase the overall loudness⁷ and there is neurological evidence that the precedence effect works for a reflection from any direction.⁸ With musical signals, the wideband reflections after the precedence effect window, about 30 ms, are still fused with the direct sound, but they also give hints about the space. Here, it is shown that reflections coming from the same direction as the source lower the quality of sound and may give an impression of a more distant sound source. Such reflections interfere with the direct sound providing identical comb filters to both ears, resulting in a colored sound. In contrast, the reflections coming from other directions make the sound stronger because the interaural differences diminish the perceived effect of the comb filters, as shown with the measured data.⁹ Earlier studies proposed the need of lateral reflections for low interaural cross correlation to create spacious sound, but they did not discuss the detrimental effect of median plane reflections when the listener is facing the orchestra. Moreover, when the listener is not facing the orchestra the perception hardly changes, a fact that the previous studies could not explain.

The human auditory system is capable of identifying coherent components of partially coherent signals and forms separate auditory events for each of these components.³ Therefore, different instruments can be distinguished when listening to a symphony orchestra. However, in a concert hall, the reflections and reverberation blend the sounds of different instruments, unless the early reflections help in the identification of instruments by supporting the direct sound of each instrument. To form a single auditory event with the direct sound the reflections have to be replicas of the direct sound. Such a replica, having the same spectrum and phase, i.e., the same temporal envelope as the direct sound, is produced only with a large flat surface. Such surfaces seldom exist in concert halls because surfaces are usually made diffusive to prevent clearly audible echoes. Diffusive surfaces spread the sound in space and time, thus randomizing the phase, and often also attenuating sound. As a consequence, in particular at high frequencies, the temporal envelope of the reflected and delayed sound differs from the temporal envelope of the direct sound, resulting in a partial breakdown of the precedence effect.¹⁰ In addition, resonant wall structures are even worse because they might attenuate sound near the fundamental frequencies of musical instruments, again deteriorating the precedence effect. Therefore, all of these temporal envelope distorting (TED) reflections do not fuse with the direct sound. During a concert, the instruments can no longer be heard separately and the articulation of notes is obscured. In contrast, when the precedence effect is not violated, the wideband temporal envelope preserving (TEP) reflections are fused with the direct sound, resulting in brighter and stronger sound with precise articulation as demonstrated in [Mm. 1](#).

The sound produced by musical instruments contains a high number of harmonics. The number and relative amplitude of the harmonics determine the timbre of different instruments. At high frequencies the harmonics are closely spaced in relation to the equivalent rectangular bandwidth (ERB) and they interfere on the basilar membrane. Thus, the resulting motion of the basilar membrane depends on the relative phase of the harmonics.¹¹ If a reflection changes the phases of high frequency harmonics, the reflected sound has a different temporal envelope than the direct sound. Two different envelopes on the basilar membrane, within circa 1.25 ERB band,¹¹ change the pattern of interspike intervals in the auditory nerve, which can affect both the clarity of the pitch¹¹ and timbre.¹² Therefore, it is most evident that TED reflections are a hindrance to hearing the harmonics of individual instruments clearly, due to the

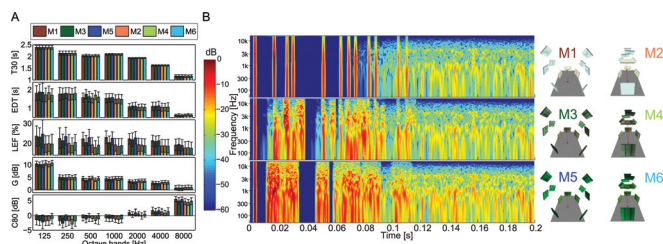


Fig. 1. (Color online) (A) The means and standard deviations of objective room acoustic parameters, computed with 24 omnidirectional sound sources as defined in ISO3382-1, for six studied concert halls (M1–M6). From top: reverberation time (T30), early decay time (EDT), lateral energy fraction (LEF), strength (G), and clarity (C80). (B) Spectrograms of the six studied simulated concert halls, showing one early part of only one of the 24 spatial impulse responses. The icons show the spatial distribution and type (white = TEP, dark = TED) of the reflecting surfaces in 3D.

partial breakdown of the precedence effect. This assumption is supported by recent neurological measurements,¹³ which prove the importance of the temporal envelope for the precedence effect.

Although the impact of TED reflections on sound perception has not been understood earlier in concert hall acoustics, the effect of two incoherent sounds have been recognized in other fields. In sound reproduction, the decorrelation of two sound signals is used to widen the sound source, to create diffuse sound fields, to increase the perceived distance and to produce externalization in headphone reproduction.¹⁴ In psychoacoustics, timbre of the sound has been shown to vary when harmonics are in different phases and when the amplitude pattern differs.¹² The research on cochlear implants has shown that pitch and harmonicity serve as the strongest cues to combine auditory objects.¹⁵ Finally, the temporal envelope of a reflection plays a major role in the identification of vowels¹⁶ and the temporal envelope has been found more important for speech intelligibility than the temporal fine structure.¹⁷

3. Perceptual evaluation with simulated concert halls

The consequences of TEP and TED early reflections and their directions relative to a listener were studied with a listening test. Six concert halls were simulated in one listening position with 24 spatial impulse responses, each containing a direct sound, 11 early reflections, and a late reverberation of 2.0 s. A set of measurements from an existing concert hall, with 24 loudspeakers on the stage and a spatial microphone probe in the listening position, were used as a reference. Each loudspeaker represents a small number of musicians on stage and the layout of the loudspeakers is designed to correspond to the seating arrangement of a symphony orchestra. For artificial concert hall renderings, 24 individual spatial impulse responses, representing the number of sources on stage, were generated as follows.

Direct sound and early reflections at 5–120 ms after the direct sound were simulated with the image source method from 11 surfaces, illustrated in *Mm. 1*. The amplitude of the direct sound and early reflections followed the $1/r$ law and the air absorption was simulated with a linear phase finite impulse response (FIR) filter fitted to standardized (ISO9613-1) equations. The directivity of the musical instruments was taken into account by filtering the direct sound and early reflections with linear phase FIR filters, which follow the measured directivities.¹⁸ The late reverberation, added to the simulated early part of the response, was extracted from the reference measurements with 24 loudspeakers on the stage. The time of arrival for the measured direct sound was matched with the simulated direct sound. In each response the late reverberation starts 60 ms after the direct sound and a fade in time of 60 ms was used. Thus, the late reverberation was fully present at 120 ms after the direct sound in the simulations.

All simulated halls had 11 early reflections. As illustrated in *Fig. 1(B)*, three early reflections were from the orchestra shell and eight reflections from the side (M1,

M3, M5) or behind the orchestra and from the ceiling (M2, M4, M6). Concert halls M1 and M2 had 11 TEP reflections from hard flat surfaces. Such a reflection does not violate the temporal envelope of sound. Concert halls M3 and M4 had 11 TED reflections of six different types. The TED reflections were measured in a semianechoic space with six different diffusing structures on top of a hard surface. The measured structures were made of wooden beams with different heights and wooden boards with different spacing. As the measured structures introduced high frequency attenuation, the attenuated energy was compensated by adding 6 ms of spectrally shaped noise 3 ms after a reflection. Together, the measured reflection and the compensation noise had an average flat frequency response. The level and temporal envelope of resolved harmonics (up to eight first harmonics) were not modified, but the temporal envelopes of unresolved harmonics at high frequencies were more or less scrambled. Concert halls M5 and M6 had 11 TED reflections, which were obtained by spreading the energy of a TEP reflection to 10 ms time span, by producing a 10 ms long noise burst with an average flat frequency response. Such random reflections change considerably levels of some resolved harmonics, in addition to high frequency temporal envelope scrambling. Finally, it is important to notice that the total sound energy remained unchanged in all of the six simulated halls (M1–M6), resulting in the same standardized monaural room acoustical parameter values, see Fig. 1(A). Lateral energy fraction was the same in (M1, M3, M5) and in (M2, M4, M6), respectively.

The spatial impulse responses for each simulated source were processed for a 14-channel reproduction. Eight loudspeakers were at ear level at 45° intervals. Four loudspeakers were above the ear level at 45° elevation and at 90° intervals. The last two loudspeakers were 40° below ear level at azimuth angles –22° and 22°. The direct sound and 11 reflections were positioned using the vector base amplitude panning.¹⁹ The measured late reverberation was processed with spatial impulse response rendering^{20,21} in order to recreate the surrounding late sound field with all the reproduction loudspeakers. Finally, 14-channel spatial responses were convolved with anechoic instrument tracks,²² composed by A. Bruckner and G. Mahler.

4. Listening test method and results

A sensory evaluation method, called Flash Profile,²³ was tailored for finding the perceptual differences between six simulated concert halls. Nineteen screened assessors, with normal hearing and musical background, listened to concert hall renderings (M1–M6) in parallel, i.e., they could switch between the samples in real time. The task of the assessors was first to elicit discriminative attributes with which they could order the six concert halls on a continuous scale. After some practice they rated the samples, presented in fully randomized order, twice with two of their own attributes for two musical signals.

The results include the clustering of elicited discriminating attributes, the ordering of them in a common factorial space and the ordering of simulated concert halls in the same multidimensional space. The integrated picture of the observations and of the relationships between the descriptive attributes is demonstrated with multiple factor analysis (MFA) on the centered and scaled data in Fig. 2(C). The first two principal components explain 74.1% of the variance of the data. The attributes formed two groups by hierarchical agglomerative clustering based on Euclidean distances, in conjunction with Ward's minimum variance method, as shown in Fig. 2(A). In total, 38 attributes were used in rankings, but the analysis is performed with the 18 reliable attributes. These 18 were reliably repeated ($p < 0.05$), when results of two ratings were associated to the test of the significance of the RV coefficient with the Pearson-type III approximation.²⁴ The number of reliably repeated attributes was rather small, probably due to short assessor training. This assumption is supported by the fact that clustering of second ratings with all 38 attributes found only eight attributes that were not clustered to two main clusters. Finally, the profiling of six simulated concert halls is illustrated in Fig. 2(B).

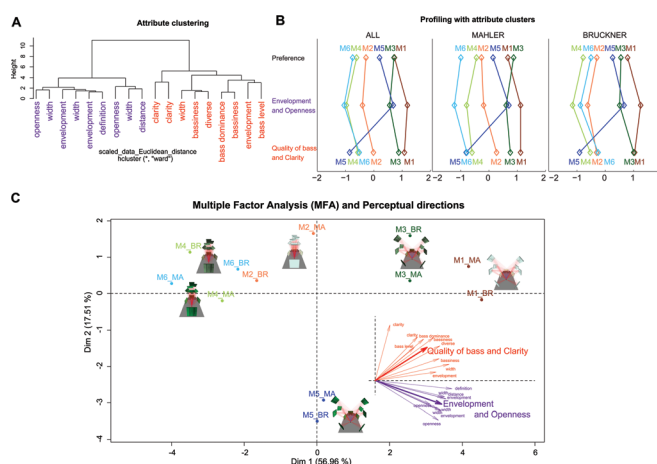


Fig. 2. (Color online) The results of the perceptual evaluation. (A) Clustering of elicited discriminating attributes. (B) Profiles of simulated concert halls with attribute clusters and preference. (C) Ordering of simulated concert halls in a common factorial space and perceptual directions of the largest variances.

The elicited attributes define the perceptual differences between six concert halls. The first cluster is dominated by envelopment and width, but also openness and distance attributes are used. The second cluster consists mainly of quality of bass attributes. Many assessors defined the quality of bass as a vigorous bass. The MFA analysis orders the simulated concert halls according to the largest variances of used attributes, namely envelopment and quality of bass attribute groups. When all reflections are surrounding the listener (M1, M3, M5), the music is always perceived more enveloping and open than in the case of the median plane early reflections (M2, M4, M6). The highest envelopment and openness are always perceived when the surrounding reflections are TEP reflections (M1). The surrounding TEP reflections (M1) and high frequency TED reflections (M3) were rated the highest according to quality of bass and clarity. As expected, the TED reflections at all frequencies (M5) render the sound more muddy with weaker bass. The attributes were originally elicited in Finnish language, which has no such words for sound as engagement or excitement, thus they are not seen as individual attributes in the results.

After sensory evaluation, all 19 assessors rated the samples according to their preference. The means of preference ratings are depicted in Fig. 2(B). A two-way analysis of variance showed a main effect on concert hall, $F(5, 216) = 27.45$, $p = 0.000$, but no significant interaction with music and concert halls, $F(5, 216) = 1.59$, $p = 0.163$. Post hoc analysis using Tukey's HSD criterion indicated that M5 ($M = 0.202$, 95% CI $[-0.055, 0.460]$) has significantly lower preference ratings than M1 ($M = 0.728$, 95% CI $[0.482, 0.974]$), $p = 0.034$, and M3 ($M = 0.711$, 95% CI $[0.480, 0.942]$), $p = 0.045$.

To conclude, in the majority of existing concert halls most early reflections are not fully preserving the temporal envelope of the sound. Such early reflections do not fuse with the direct sound because they partially break down the precedence effect, which results in less enveloping and slightly muddied sound, in particular at low frequencies. In contrast, early TEP reflections from the side are perceived as the most open, clear, and enveloping. In addition to envelope distortion, small and scattering surfaces cause frequency-dependent absorption, which also weakens the contribution of early reflections to total sound. It is most obvious that the differences would be even bigger between TEP and TED reflections when TED reflections also attenuate sound. However, research on perceptual consequences of this attenuation, as well as the low frequency attenuation of the direct sound due to the seat dip effect, remains for future work.

Acknowledgments

We thank H. Vertanen for the help in listening tests and Dr. E. Kahle, Professor J. Blauert and Dr. D. Griesinger for discussions. The research leading to these results has received funding from the Academy of Finland (Project Nos. 218238 and 140786) and the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 203636.

References and links

- ¹H. Haas, "On the influence of a single echo on the audibility of speech," *Acustica* **1**, 48–58 (1951).
- ²H. Wallach, E. Newman, and M. Rosenzweig, "The precedence effect in sound localization," *Am. J. Psychol.* **62**, 315–336 (1949).
- ³J. Blauert, *Spatial Hearing. The Psychophysics of Human Sound Localization*, 2nd ed. (MIT Press, Cambridge, MA, 1997), pp. 201–287.
- ⁴M. Ebata, T. Sone, and T. Nimura, "On the perception of direction of echo," *J. Acoust. Soc. Am.* **44**, 542–547 (1968).
- ⁵H. Seraphim, "On the perceptibility of multiple reflections of speech sounds," *Acustica* **11**, 80–91 (1961).
- ⁶A. Noxon, "Sound fusion and the acoustics presence effect," in the 89th AES Convention, 1990, paper no. 2998.
- ⁷R. Freyman, D. McCall, and R. Clifton, "Intensity discrimination for precedence effect stimuli," *J. Acoust. Soc. Am.* **103**, 2031–2041 (1998).
- ⁸R. Y. Litovsky, B. Rakerd, T. C. T. Yin, and W. M. Hartmann, "Psychophysical and physiological evidence for a precedence effect in the median sagittal plane," *J Neurophysiol.* **77**, 2223–2226 (1997).
- ⁹Y. Huang, Q. Huang, X. Chen, T. Qu, and L. Li, "Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults," *Hear. Res.* **244**, 51–65 (2008).
- ¹⁰K. Terada, M. Tohyama, and T. Houtgast, "The effect of envelope or carrier delays on the precedence effect," *Acta Acust. Acust.* **91**, 1016–1020 (2005).
- ¹¹B. C. J. Moore, "Interference effects and phase sensitivity in hearing," *Philos. Trans. R. Soc. London, Ser. A* **360**, 833–858 (2002).
- ¹²R. Plomp and J. Steeneken, "Effect of phase on the timbre of complex tones," *J. Acoust. Soc. Am.* **46**, 409–421 (1969).
- ¹³B. Nelson and T. Takahashi, "Spatial hearing in echoic environments: The role of the envelope in owls," *Neuron* **67**, 643–655 (2010).
- ¹⁴G. Kendall, "The decorrelation of audio signals and its impact on spatial imagery," *Comp. Music J.* **19**, 71–87 (1995).
- ¹⁵J. Culling and C. Darwin, "Perceptual separation of simultaneous vowels: Within and across formant grouping by f_0 ," *J. Acoust. Soc. Am.* **93**, 3454–3467 (1993).
- ¹⁶A. Watkins and N. Holt, "Effect of a complex reflection on vowel identification," *Acustica* **86**, 532–542 (2000).
- ¹⁷R. Drullman, "Temporal envelope and fine structure cues for speech intelligibility," *J. Acoust. Soc. Am.* **97**, 585–592 (1995).
- ¹⁸J. Pätynen and T. Lokki, "Directivities of symphony orchestra instruments," *Acta Acust. Acust.* **96**, 138–167 (2010).
- ¹⁹V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *J. Audio Eng. Soc.* **45**, 456–466 (1997).
- ²⁰J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," *J. Audio Eng. Soc.* **53**, 1115–1127 (2005).
- ²¹V. Pulkki and J. Merimaa, "Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests," *J. Audio Eng. Soc.* **54**, 3–20 (2006).
- ²²J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acust. Acust.* **94**, 856–865 (2008).
- ²³V. Dairou and J.-M. Sieffermann, "A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile," *J. Food Science* **67**, 826–834 (2002).
- ²⁴J. Josse, J. Pagès, and F. Husson, "Testing the significance of the RV coefficient," *Comput. Statist. Data Anal.* **53**, 82–91 (2008).