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

Supporting online material

Tapio Lokki[1*], Jukka Pätynen[1], Sakari Tervo[1], Samuel Siltanen[1], Lauri Savioja[1]

[1]Department of Media Technology, Aalto University School of Science,
POBox 15400, FI-00076 Aalto, FINLAND
[*]To whom correspondence should be addressed; E-mail: Tapio.Lokki@aalto.fi

June 2011

This document contains supporting extra material for the JASA Express Letter (http://dx.doi.org/10.1121/1.3579145) and the supporting binaural demonstration video (http://link.aip.org/mm/JASMAN/1.3579145/507105jasv1.mp4)

S1. Demonstration video

The visualization of the reflections on the demonstration video is based on the hall simulation data which consists of the arrival times, amplitudes, and directions of the direct sound and reflections. When the exact reflection points are known the surface is drawn around them by utilizing the plane equation. The texture of the surface is determined by the property of the reflection; light surfaces visualize the temporal envelope preserving (TEP) reflections and the dark green surfaces illustrate the temporal envelope distorting (TED) reflections, as illustrated in Fig. 1.

The sound track of the video was rendered with virtual acoustics as explained in Section S3, except that a different reproduction method was used. The binaural sound track for headphone listening was carried out with a virtual loudspeaker array. In this technique the spatial sound is first computed for 24 loudspeakers surrounding a virtual listening position. Each channel is then convolved with a head-related transfer function defined by the direction of the corresponding loudspeaker. Here, there were 12 loudspeakers at ear level, at 30 degree intervals, and six loudspeakers at 60 degree intervals, at 45 degrees above and -30 degrees below the ear level, respectively.

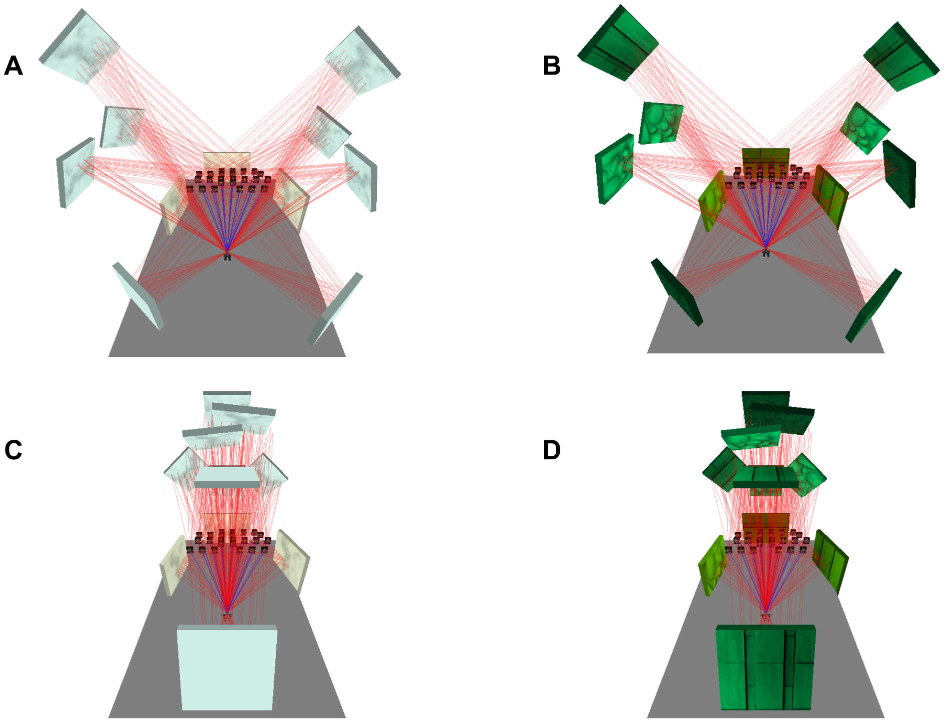

Figure 1: Four frames from the video illustrating temporal envelope preserving (**A**) and temporal envelope distorting (**B**) side reflections, and temporal envelope preserving (**C**) and temporal envelope distorting (**D**) median plane reflections.

S2. Harmonic complex tones of musical instruments

We recorded a single note c#4, the fundamental frequency of 277 Hz, of some symphony orchestra instruments in an anechoic chamber. The frequency responses of these instruments, shown in Fig. 2, contain a high number of harmonics, from 10 to over 25. As the fundamental frequency is equal for all instruments, human hearing can identify the differences between them from the relative amplitudes of harmonics.

Psychoacoustic studies have shown that the phases of sinusoidal components in a complex tone are not heard if they are separated sufficiently in frequency (*1*). In such a case the wave pattern of the basilar membrane produces distinct peaks for all separated sinusoidal components. However, when these components are closer than circa 1.25 ERB, two of them interfere at the basilar membrane (*1, 2*) and phases are therefore required for perception because together they can influence to the movement of the basilar membrane. Figure 2 shows that in a harmonic complex tone, at high frequencies over 4 kHz, at least three harmonics occupy 1.25 ERB. If the played note were one octave lower the harmonics would be twice as dense, resulting in at least three harmonics at 1.25 ERB even at 2 kHz.

The temporal envelope of a tone at one ERB is determined by the amplitude and phase of harmonics. With music in a concert hall, the played notes are long enough that they reach the ears of a listener as a sum of the direct sound and reflected sounds from the walls. If the reflections are of TED type, the resulting interference at the harmonics will alter the movement of the basilar membrane, as opposed to a case with only the direct sound or the direct sound with TEP reflections. When the reflections preserve the temporal envelope at one ERB band, the basilar membrane vibration, according to the direct sound, is not distorted, and the TEP reflections are fused with the direct sound due to the precedence effect.

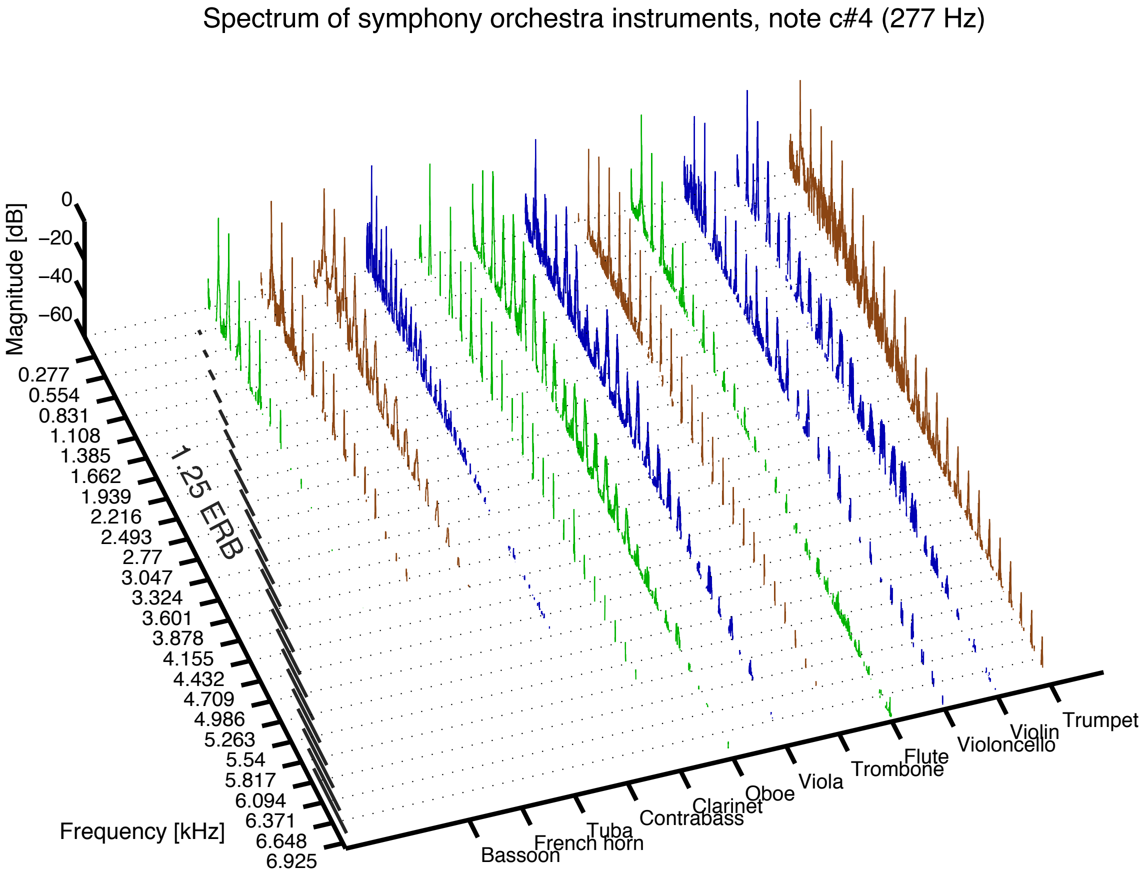

Figure 2: The spectrum of the common symphony orchestra instruments and widths of 1.25 equivalent rectangular bands (ERB). Instrument groups are marked with different colors; woodwinds (green), brass (brown) and strings (blue).

S3. Simulation of the concert halls

The acoustic responses of the six artificial concert halls were composed of the direct sound, 11 early reflections and the late reverberation. A set of measurements from an existing concert hall, with 24 loudspeakers on the stage and a spatial microphone probe in a constant 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, twenty four individual spatial impulse responses, representing the number of sources on stage, were generated as follows.

The direct sound and early reflections were simulated with the image source method from 11 surfaces, illustrated in Fig. 1, thus reflections were at 5-120 ms after the direct sound. The amplitude of the direct sound and early reflections followed $1/r$ law and high frequency attenuation due to air absorption was simulated with a linear phase finite impulse response (FIR) filter fitted to standardized equations (*3*). The direct sound and early reflections were also filtered with a linear phase FIR filter that was fitted to measured directivities of musical instruments (*4*). The late reverberation added to the simulated early part of the response was measured in a real concert hall 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.

The simulated concert halls had three types early reflections. It is important to notice that the total sound energy remains unchanged in all of the six artificial halls (M1-M6), resulting in the same standardized monaural room acoustical parameter values (*5*). Lateral energy fraction was the same in (M1, M3, M5) and in (M2, M4, M6), respectively.

- Concert halls M1 and M2 had 11 TEP reflections from the hard flat surfaces. Such a reflection is illustrated in Fig. 3B and the ERB band analysis shows that a TEP reflection does not violate the temporal envelope the outputs of any auditory filter of the basilar membrane.

- Concert halls M3 and M4 had six different type of TED reflections. The TED reflections were measured in a semi anechoic space with six different diffusing structures on top of a hard surface. One of the measured structures is illustrated in Fig. 3C. 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 ERB band analysis shows that level and temporal envelope of resolved harmonics (up to 8 first harmonics) are not modified, but the temporal envelopes of unresolved harmonics at high frequencies are 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. This was performed by producing a 10 ms long noise burst with an average flat frequency response. Again, Figure 3D shows one example of the used random reflections. The ERB band analysis reveals that in addition to high frequency temporal envelope scrambling levels of some resolved harmonics are considerable changed.

In all six concert halls the 24 sound source positions were associated with anechoic symphony music. Six sources were used for violins, three for violas, three for violoncellos, two for contrabasses, four for woodwinds, two for French horns, 3 for trumpets, trombones, and tuba, and finally one sound source for timpani.

A spatial sound reproduction with 14 Genelec 6020 loudspeakers was used in the perceptual evaluation. Eight loudspeakers were at ear level at 45 degree intervals. Four loudspeakers were above the ear level at 45 degrees elevation and at 90 degree intervals. The last two loudspeakers 40 degrees below ear level were at azimuth angles -22 and 22 degrees. The direct sound and 11 reflections were positioned using the vector base amplitude panning (*6*). The measured late reverberation was processed with spatial impulse response rendering (*7, 8*) in order to recreate the surrounding late sound field with all the reproduction loudspeakers.

References and Notes

1. Brian C J Moore. Interference effects and phase sensitivity in hearing. *Phil. Trans. R Soc. Lond. A*, 360(1794):833–858, 2002.

2. B. C. J. Moore and K. Ohgushi. Audibility of partials in inharmonic complex tones. *J. Audio Eng. Soc.*, 93(1):452–461, 1993.

3. ISO 9613-1. Acoustics — attenuation of sound during propagation outdoors — part 1: Calculation of the absorption of sound by the atmosphere, 1993.

4. J. Pätynen and T. Lokki. Directivities of symphony orchestra instruments. *Act. Acust. Acust.*, 96(1):138–167, January/February 2010.

5. ISO 3382-1:2009. Acoustics – measurement of room acoustic parameters – part 1: Performance spaces. International Standards Organization, 2009.

6. V. Pulkki. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.*, 45(6):456–466, June 1997.

7. J. Merimaa and V. Pulkki. Spatial impulse response rendering I: Analysis and synthesis. *J. Audio Eng. Soc.*, 53(12):1115–1127, 2005.

8. V. Pulkki and J. Merimaa. Spatial impulse response rendering II: Reproduction of diffuse sound and listening tests. *J. Audio Eng. Soc.*, 54(1):3–20, 2006.

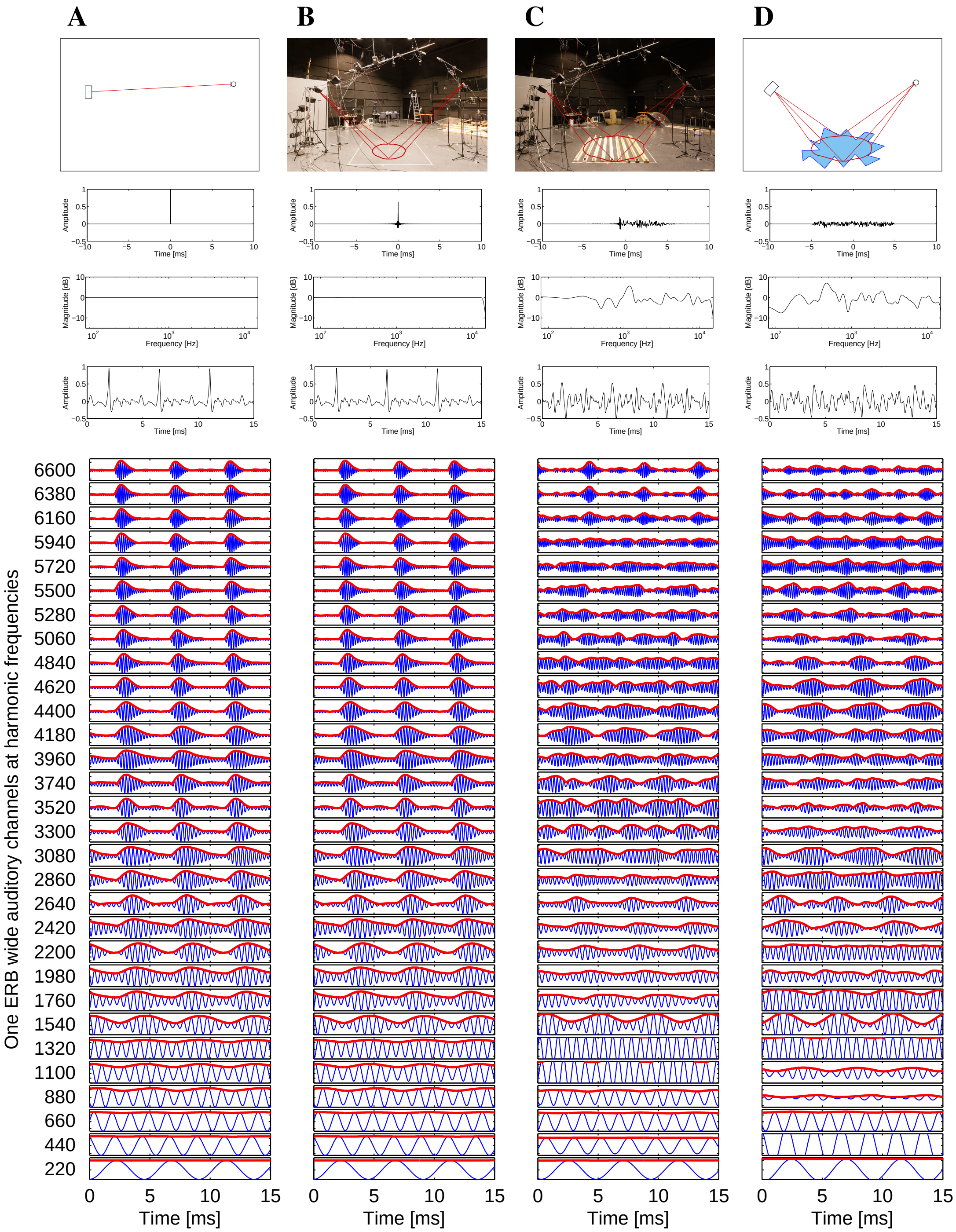

Figure 3: From top to bottom: Illustration of a direct sound or a reflection, response in the time domain, frequency response, direct sound or reflection convolved with a part of the trumpet sound of A3 (220 Hz), amplitude and envelope of convolved trumpet sound at ERB bands, i.e., the waveforms (blue) and envelopes (red) at the outputs of auditory filters on the basilar membrane. (**A**) Direct sound (**B**) TEP reflection (**C**) TED reflection at high frequencies (**D**) TED reflection at all frequencies.