THE EFFECTS OF EARLY REFLECTIONS ON PROXIMITY, LOCALIZATION AND LOUDNESS

D Griesinger David Griesinger Acoustics, Cambridge, Massachusetts, USA

1 INTRODUCTION

"Proximity" – the perception of being sonically close to a sound source - is not commonly found as a descriptor of concert hall sound. But Lokki et al. [8] have identified proximity as perhaps the most important acoustic perception affecting preference. To quote from Lokki, Pätynen, Kuusinen, and Tervo (2012, p. 3159*)* [8]:

"An interesting fact is that neither Definition and Reverberance nor EDT (early decay time) and C80 explain the preference at all. In contrast, the preference is best explained with subjective Proximity and with Bassiness, Envelopment, and Loudness to some extent. Further there is no objective measure that correlates to Proximity and overall average of preference."

In our view this is a powerful and damming statement. Why do commonly thought-of acoustic qualities like Definition and Reverberance, as well as the measures EDT and C80 have no consistent effect on preference, and why has one of the most important determinants of preference been both previously unknown and un-measureable?

It is the "direct sound", the component of a sound-field that travels directly from a source to a listener, that contains the information needed to localize the source and to perceive its closeness. It is possible to accurately localize a source, and perceive that it is in some way close to us, even when the direct to reverberant ratio (D/R) is less than -10dB. To do this our ears and brains have somehow separated the direct sound from the reflections that follow.

It is not obvious how this feat is performed. The mechanisms behind our abilities to separate signals from noise and other signals have been poorly understood. We believe the mechanism for both source separation and the perception of proximity relies on the phase alignment of the upper harmonics of music and speech. We believe that understanding this mechanism has the potential to revolutionize research into speech and musical acoustics. [3] [7]

In our view acoustic research has been stymied by several problems. It is well known in the audio field that it is nearly impossible to tease out differences in sound quality without instant A/B comparisons. But to do this for concert halls requires that the sound in different halls and seats be exactly reproduced in a laboratory, with the ability to instantly switch from one sound to another. In our view the perception of proximity relies on the ear's ability to detect the direct sound as different from other sounds, reflections, and noise through the regularity of the phases of the upper harmonics from vowels and musical notes. Reflections from all directions will randomize these phases. We need to be able to predict with accuracy where in a given venue the reflections become strong enough to make the detection of the direct sound impossible, and to do this we must be able to reproduce in the laboratory both the phase qualities of the direct sound, and the amplitude, direction, and time delays of the reflections that follow. We cannot do this if our measurement and reproduction methods do not reproduce these phases accurately, and capture the precise ratios between the direct sound and the reflections. Lokki is showing that the task is difficult, but it can be done. It requires a new method of working.

The perception of proximity has not been given the attention it deserves in part because currently standard methods of hall measurement, which use omnidirectional sources and receivers, are incapable of capturing it. To make matters worse, it is almost impossible to preserve phases when a sound source is panned between multiple loudspeakers.

A major additional problem is that hearing research has concentrated on oversimplified signals, such as noise and sinewaves. None of these signals carry much information, and do not resemble the signals human hearing has evolved to decode. Possibly because of this misstep acoustic research has ignored the importance of phase for signals above 1500Hz. But our research finds that the perception of proximity depends on the regular peaks in the sound pressure envelope created by the phase alignment of harmonics above 1500Hz. While it is true that the ear is insensitive to the phase of the carrier above 1500Hz, it is acutely aware of the envelope of signals. The harmonics of speech and music create peaks in the envelope of the sound pressure that are highly audible, and we believe the audibility of these peaks is of vital importance to the human ability to separate signals from each other and from noise.

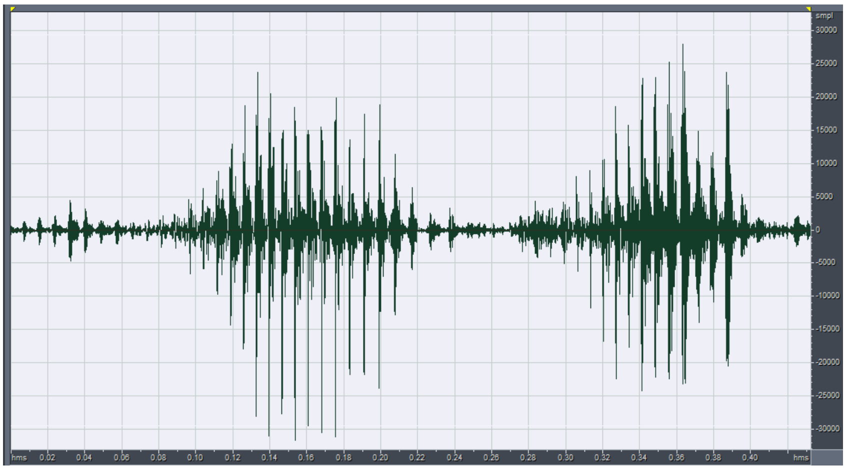


Figure 1: The syllable “one” first filtered through a 2kHz 1/3 octave 2nd order Butterworth filter, and then though a 4kHz filter. Note the prominent envelope modulation at the fundamental period, with peaks more than 10dB above the minima between peaks. Although the ear is not sensitive to the phase of the carrier at these frequencies, it is highly sensitive to these peaks. When they are present such a source can be sharply lateralized by interaural time differences (ITD) alone. If you listen to these filtered waveforms there is also a prominent perception of the fundamental tone. The horizontal scale is 0 to 0.44 seconds. (figures created by the author)

The idea that phase is both audible and important is not new. Blauert (1983, p. 153) remarks that for speech “Evaluation of the envelope can already be detected at a carrier frequency of 500 Hz, and it becomes more accurate as the frequency is increased. When the entire signal is shifted as a unit, lateralization is quite sharp throughout the entire frequency range.”

But reflections from all angles interfere with the phases of harmonics in the direct sound. They lose their alignment, and the sharp peaks at regular intervals become random noise. The ability to separate the direct sound from other signals, reflections, and noise is degraded, and the sense of proximity is lost.

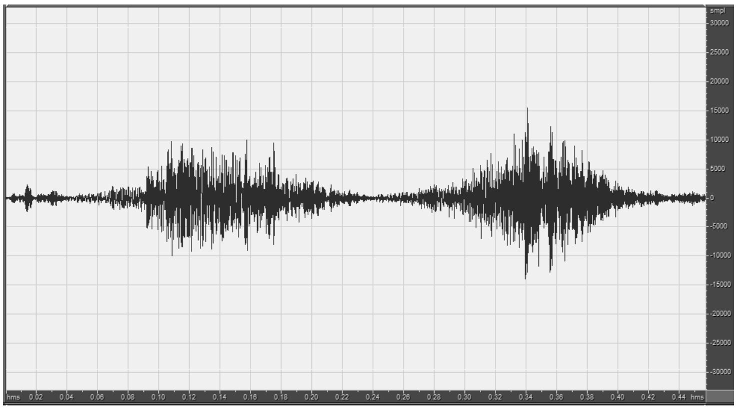


Figure 2: The same signals as figure 1, but altered in phase by a filter made from three series allpass filters of 371, 130, and 88 samples, and with allpass gains of 0.6. Notice that the peaks at the fundamental period have largely disappeared. When you listen to these signals no fundamental tone is heard. There is garbled low frequency noise instead.

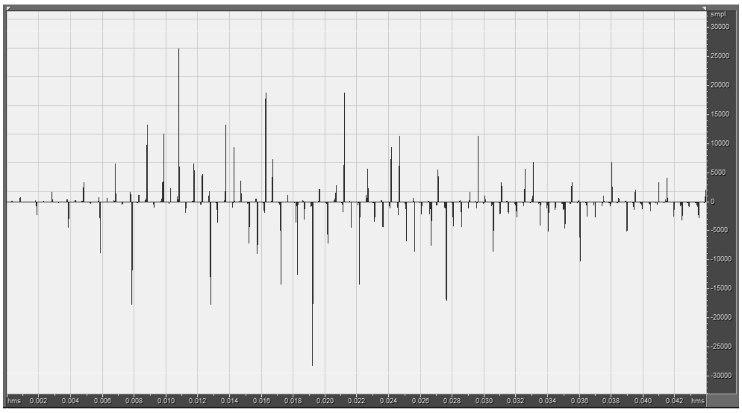


Figure 3: The impulse response of the all-pass filter used to create figure 2 from figure 1. The horizontal scale is 0-44ms. This filter is inaudible to pink noise, but highly audible to speech. When tested with current acoustic measures the filter should not affect the sound, as C80 = C50 = infinity, and STI is 0.98. But speech and music through this filter sound distant, muddy and unpleasant.

This preprint describes a series of experiments that use a version of Lokki's virtual orchestra [9] to study the effects of early reflections through binaural technology. With this method it is possible to take existing binaural impulse response data from the stage to a particular seat, and use it and Lokki's anechoic recordings to synthesize the sound of a musical ensemble. We change the ITD and ILD of the direct sound component of the measured impulse response to make a new IR for each instrument that puts each instrument at the correct azimuth, and then convolve the ensemble of impulse responses with Lokki's anechoic recordings. The resulting sounds are mixed together and played back through headphones calibrated individually for each listener. (The method of calibration is discussed in another preprint for this conference.) The playbacks are startlingly realistic, and are closely similar to the author's personal binaural recordings of live performances in the same set of seats.

When the BSH measurements were made we also recorded the impulse responses at each seat with a four-channel soundfield microphone. This data shows the direction of the strongest early reflections. Using this information we break the measured impulse response into small segments, each segment containing just the direct sound or the sound of a particular reflection. Separately convolving the segments allows us to instantly evaluate the effect of each reflection on the sound of the players. The differences in the sound examples included in this preprint, while best heard through individually calibrated headphones, can in many cases be heard with closely spaced loudspeakers, as are often found on desktop computer systems.

The work in this preprint is confined to the author's data set for BSH. Almost all the seats studied are considered by the author (and the ticket price) to be very good. The difference in sound with and without the first lateral reflection can be subtle, but it can be reliably heard. The author has similar data sets for other venues waiting to be analyzed in this way. But for BSH the results are conclusive. Early lateral reflections in most seats in the rear half of the hall are nearly always detrimental to sound quality. Deleting them from the impulse responses improves both proximity and envelopment with no effect on loudness. One of the seats – in row DD on the right hand side of the hall – is in the author's collection of live performance recordings, and has poor proximity. The binaural synthesis of this seat is quite similar, but deleting the right side wall reflection brings the sound back to life. In the front half of the hall when a seat is not in the center section the reflection off the near side wall shifts the localization of instruments and muddies the sound. The difference is subtle, but the author has noticed it in many of the performances he has heard in the hall, and a few years ago both the author and Mr. Lokki noticed it during a live performance in the hall. It is gratifying to finally be able to reproduce it in the laboratory.

Much of the material in this preprint can be found in more detail in reference [1].

2 PROXIMITY

What is "proximity", and how can we learn to recognize it? Perhaps the best way to introduce the concept is to describe a simple experiment. Human perception of source distance and direction is predominantly visual. At a music performance with eyes open we are usually sure we are hearing the precise location of each musician. But this impression can change dramatically when the eyes are closed. When working on hall acoustics we find it very useful to walk slowly away from a small group of musicians or actors while looking at the floor. In practice we often use a version of Tapio Lokki's virtual orchestra, which plays away tirelessly on stage with no visual reference for who is playing which line. Close to the group each virtual performer and musical line can be localized precisely by ear, and the lines they are playing or speaking can be easily distinguished separately from the others. As we walk back very little changes at first. The spread of azimuth diminishes, but the ability to localize and follow the lines is unchanged. The performers are still "present".

At a particular distance everything changes. The sound of the ensemble collapses into a fuzzy ball, and the performers lose their sense of closeness. We will refer to the distance from the sound sources found in this way as the *limit of localization distance*, or LLD. The difference in the distance between a position in the hall where the sound is sharp and clear to a position where the sound is fuzzy can be less than a meter or two. One can learn to hear the difference in sound by walking back and forth around the LLD – a new experience for most people. A/B comparison of binaural recordings of live music in front of and behind the LLD can also be startlingly different, even when played through loudspeakers. There is a high degree of agreement between individuals for the distance from the stage of the LLD. The property that creates it appears to be a property of the sound field, not of the individual.

We hypothesize that the peaks in the pressure envelope created by the alignment of harmonic phases in the direct sound are essential for the sense of closeness, and also enable the ability to separate sounds. We infer from the experience above that the ability of the ear to separate direct sound from reflections works well down to a certain value of the direct to reverberant ratio, D/R, in a particular venue. Below this value localization and separation of individual sources become difficult, and the sense of closeness disappears. In recent work it has been shown that loss of the phase relationships of the upper harmonics can also severely impact the intelligibility of speech in the presence of noise. See references [1] and [10].

Reflections in the first fifty milliseconds are believed to add beneficially to intelligibility and loudness. But the experiments that justify this claim were based on single reflections and are not applicable when multiple early reflections exceed the direct sound energy. A combination of reflections can easily mask the direct sound from speech and music to such an extent that even the most sophisticated machine – the human ear - cannot detect it. We need to know when reflections are beneficial, and when they are not.

3 LOKKI'S HALL SIMULATOR

Fortunately it is possible with a hall simulator like that of Lokki et al. [9] or an accurate individual binaural system to study the effects of early reflections in detail. To do so requires that we have available to us separate impulse responses for each instrument from a source loudspeaker that emulates the directivity of that instrument. This is the basis of Lokki's recording method. He sets up a large loudspeaker array using two drivers for each instrument, one pointing forward and one pointing up. In theory the balance between the two can adjust the directivity to some degree, but I am not sure he does this. The loudspeakers he uses are about the size of a human head, but the high frequency waveguide is probably more directive. In any case the directivity is close enough to real instruments that the sound is convincing.

It is not necessary to have the entire orchestra setup before gathering the impulse response data. Only one speaker array is needed. It can be moved to the various locations and the measurement can be repeated. Lokki gathers the impulse responses both binaurally with a dummy head and with a three-dimensional first-order microphone array. From this microphone the azimuth of each virtual instrument can be detected. The direct sound captured by the microphone is convolved with the anechoic recording of the instrument and sent to the closest frontal loudspeaker in his playback system, which for the front image consists of 9 loudspeakers 22.5 degrees apart. The principle direction of the reflections from each instrument one at a time is determined on a sample to sample basis, and convolved sound is directed to the closest speaker to that direction in the surround array. The same is done for all the instruments, and all the signals are combined. The result is the most realistic reproduction of a particular hall that the author has heard.

The system is complex and expensive. Might there be an easier way?

4 BINAURAL TECHNIQUE

It is reasonable to assume that if a person could make a recording of the sound pressure at each eardrum, and then play it back in such a way that the sound pressure was precisely duplicated, that the original sound impression would be exactly reproduced. Manfred Schröder demonstrated this technique using a crude dummy head microphone and a very sophisticated playback system that employed loudspeakers, mathematically derived crosstalk canceling filters, and steel probe tubes at the listener's eardrums. The system was deemed successful at the time.

These experiments were continued at IRCAM in Paris. In this case the recording was from steel probe tubes at the subject's eardrums, and playback was with Schröder's method. Subjects needed to have their heads in a clamp both for the recording and the playback, or risk damage to the eardrums. But the IRCAM experiments are reported to have been very successful. The original sound field was reproduced exactly.

There is a better way to the same result. The author has been making binaural recordings of live concerts as part of this acoustical work for many years. Initially this was done with small microphones taped to his glasses bows just above the pinna. After many tests he found a type of earphone that worked well for reproducing the recordings – a small on-ear headphone by Koss called the Portapro. But since 2007 he has used probe microphones with a soft, very flexible tip to record the pressure at the eardrums. He also developed a dummy head with exact copies of his pinna, ear canals, and eardrum impedance. This head – and another like it – allows him to record sound accurately in three places at once. Both the probes and the heads are equalized such that the frequency response matches the frequency response of a loudspeaker in front of the head up to at least a frequency of 6kHz, at which frequency there begin to be response notches that you should not attempt to equalize.

5 RECORDING BINAURAL SOUND

The effort that went into making the dummy heads was necessary for testing headphones (if only for the author's use) but it is not essential for binaural recording of halls. We have found that at least several commercially available heads with realistic pinna can give useful results if they are equalized to match the response of a frontal loudspeaker. In some cases an existing IR can be equalized simply by looking at the frequency response as a function of time. The response of an IR in the time range of 100-200ms as measured by Adobe Audition, for example, should be quite flat, with a gentle roll-off above 3kHz. In general, it seldom is. But with some practice you can correct it with parametric filters using an impulse response recorded in a hall. In some ways this is better than measuring the head alone with a calibrated loudspeaker, because a correction from hall data compensates both for the response of the head and the power response of whatever loudspeaker was used for the source in the measurement.

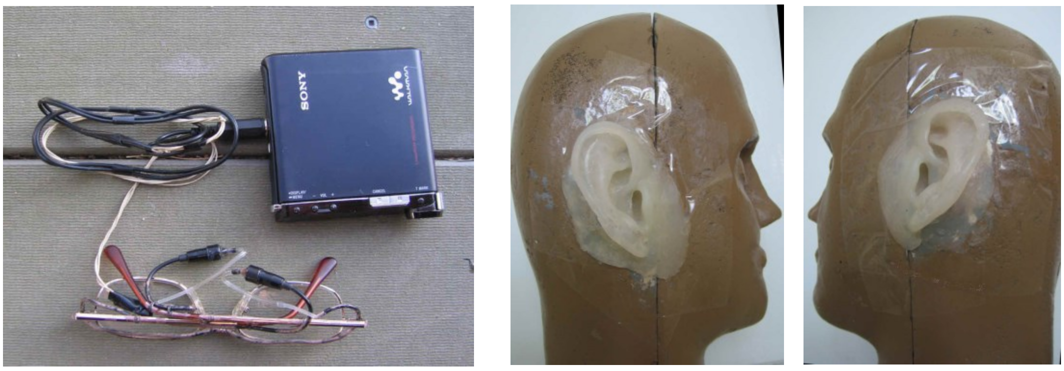


Figure 4: the author's recording equipment

5 PREPARING HEADPHONES FOR BINAURAL LISTENING

The Koss Portapro headphones were a step forward, and were very useful in learning to recognize proximity. But the image was not always frontal, and it never had a sense of realism. The author has been using headphones for more than forty years as an on-location recording engineer. He learned very early that headphones do not reproduce the same frequency balance as loudspeakers. They are untrustworthy as recording tools.

In the early 1980s he began to make his own probe microphones and measure the sound pressure at his eardrums from a frontal loudspeaker. The spectrum at the eardrums while wearing headphones was then adjusted with a graphic equalizer to match the spectrum from the loudspeaker. The technique is invasive and at the time it could be painful, but it works. Headphones equalized this way were far more accurate for on-location recording, and the author has used this method ever since. Currently we use the small soft probe microphones developed for binaural recording for equalizing headphones for binaural playback. See figure 4.

Measuring headphones with probe microphones is not the only method that works. DIN Standard 45619 from June 1975 describes a method of equalizing headphones that uses loudness matching to indirectly measure the pressure at the eardrums. The standard dictates that to equalize a headphone you play a third octave noise band through a loudspeaker in front of a listener, and have the listener quickly switch back and forth from listening to the noise from the speaker, and listening to the same noise through the headphones under test. The signal to the headphones is adjusted until the two noises have the same loudness. The result as a function of frequency is the desired frequency response correction for the headphones. Adjusting an equalizer to this curve results in a very good sounding headphone for that particular listener. With many – but not all – headphones music and speech are perceived in front of the listener, and with accurate timbre. Recordings from a similarly equalized dummy head can be quite convincing.

The problem with equalizing headphones with probe microphones and with DIN 45619 is that a response that is accurate for one listener is usually far from accurate for another listener. Variations of +-5dB or more between individuals in the frequency range from 500Hz to 6kHz are the rule, not the exception, even for so-called “free-field” open phones. Headphone design went a different direction from the 1975 DIN standard in an attempt to find more universal response curves – unsuccessfully in our opinion. To achieve frontal localization head tracking has been assumed to be necessary. It is not. Individual equalization is necessary, not head tracking. To us there is no universal headphone equalization, although some particular headphone designs are more independent of the listener than others. (This does not include most insert phones, which can be highly variable.) See figure 5.

DIN 45619 is effective but tedious to use. We have developed a computer app that achieves the same result without needing to take the headphones on and off. It uses an equal loudness method similar to the ISO method for finding equal loudness curves. The method is described in another preprint for this conference. We will have the app set up at the conference, and attendees are encouraged to try it. (We will also have a number of headphones that have good independence of response for different individuals available for hearing the results of this work and some of our binaural hall data.)

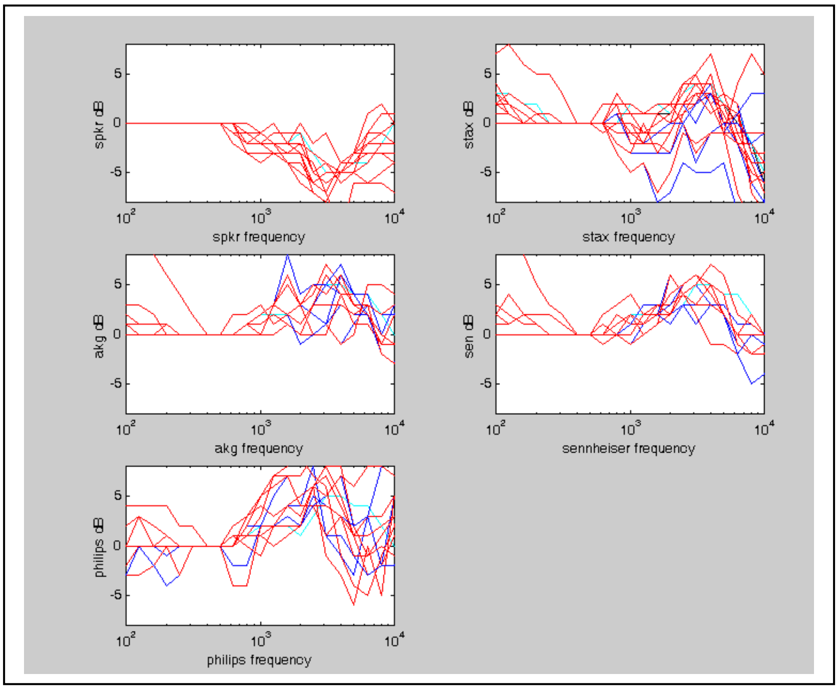


Figure 5: Headphone frequency response as measured by equal loudness for four different headphones by a group of students in Finland using the headphone measuring app. The top left graph shows the equal loudness contours of each student in the test. The other graphs show the difference between their individual equal loudness curves and the equal loudness curves from the headphones. The headphones were AKG 701, Stax 303 Classic, Sennheiser 250, and a Phillips insert phone.

Although 1/3 octave equalization of headphones always improves the realism of the sound, some headphones are better than others for binaural reproduction. Circumaural “free field” open phones that are typically used for binaural listening are usually not the best choice, and circumaural closed phones may well be worse. By design they create notches in the frequency response that may resemble the individual’s 90 degree azimuth HRTFs, but this is undesirable for binaural reproduction. The HRTFs are supposed to be in the recording, not the headphone. The notches these headphones create also vary each time the headphones are put on the head, which makes them impossible to invert, even with a mathematical inverse filter.

On ear headphones such as the Koss Portapro mentioned above, and my current favorites the Sennheiser 100, 200, 250-II, and 350 designs provide a more startling realistic playback of my binaural recordings after they are equalized. (Caution – the frequency responses of the different models are different, but the uniformity of different examples of the units I have tested has been good.)

5 PREPARING BINAURAL IMPULSE RESPONSES FOR AURALIZATION

The impulse response data used for the experiments in this preprint came from two different measurement sessions. Most came from a 2008 session in BSH with Leo Beranek and a group from Rensselaer Polytechnic Institute under Ning Xiang. While they were setting up for conventional measurements the author quickly measured his favorite seats with a sine sweep from a Genelec 1029 loudspeaker near the conductor’s position. The stage was fully covered with stage furniture, so there was little or no back-wall reflection. The loudspeaker used is very similar to the modern version of the same design used by Lokki. We did not utilize a second speaker of this type pointing up, so the directivity of the source is higher by 2-3dB than the arrays used by Lokki. A second group of data came from a similar session in 1012. In the second session I did not bring my dummy head, but obtained the data from Ning’s Head Acoustics head microphone.

In the first session two microphones were used in each seat position, a small soundfield microphone constructed by the author about 25 years ago, and the dummy head built with models of my own pinna, ear canals, and eardrum impedance. As usual it was equalized for flat response from the front up to about 6kHz. The soundfield data verified that the reflection we were studying was indeed from the right-hand side wall. It is worth noting that in figure 6 the data is from a 2000Hz octave band. In most displays of data of this type the apparent direction of reflections is based on the full bandwidth of an impulse response, which means that half the energy displayed is typically above 10,000Hz. We believe these graphs are misleading, as there is very little energy in that frequency range in concert halls. Plotting data from the 2000Hz and 4000Hz bands is more likely to be meaningful. At these frequencies a first-order microphone (and the human ear) is unable to resolve much of interest beyond about 75ms.

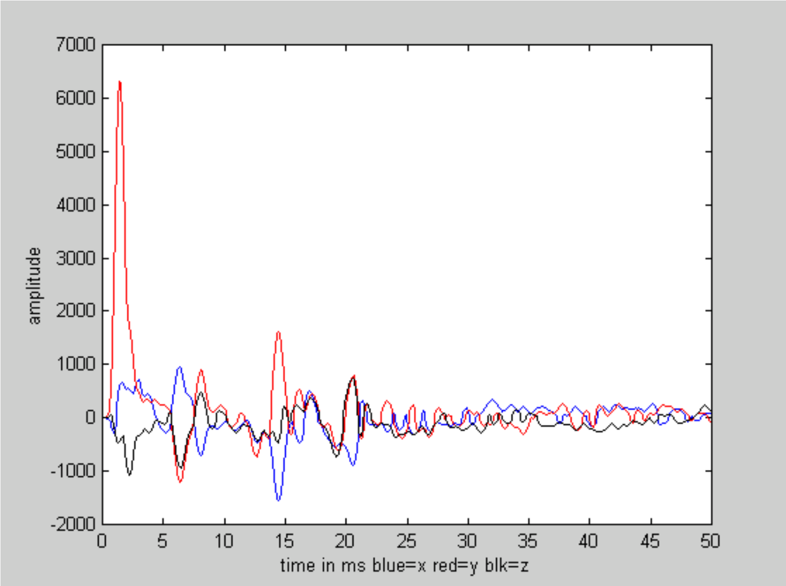
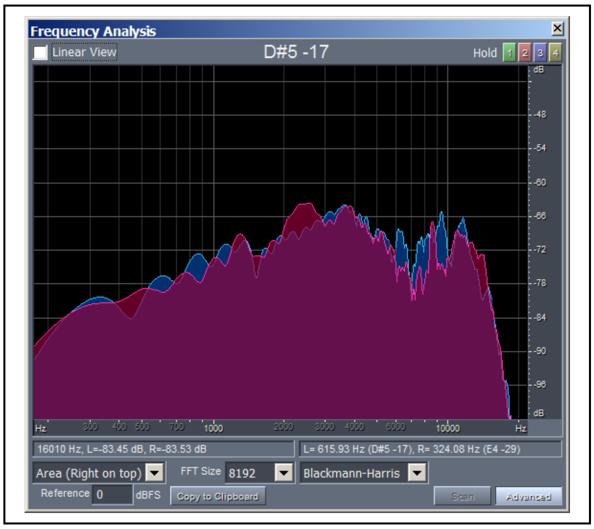


Figure 6: Direct sound and first reflection as seen by the soundfield microphone for BSH seat DD 11. The data is from a one octave band centered at 2000Hz. There is a strong side wall reflection at about 14ms after the direct sound, followed by a weaker ceiling reflection at about 20ms. We will test the effects on the sound when the side wall reflection is eliminated.

The loudspeaker has a flat response on axis, so in theory the direct sound impulse should have a flat response also. But in all the measurements it does not. Low frequencies are steeply rolled off by the seat-back effect, and there is a high frequency boost from about 2kHz to 5kHz. But most of the energy in all the impulse responses is in the reflections and the reverberation. To sound natural it is essential that the reverberation should have a smooth frequency response: Flat from 60Hz to about 3kHz, and then rolling off more and more as time goes on. With some trial and error and the parametric equalizer in Adobe Audition, this can be achieved. Doing so also cleans up the direct sound to some degree – but you should not attempt to make the direct sound have a flat frequency response. Check that in the reverberation the two channels are more less the same. They are likely to be different for any number of reasons. You need to make them as similar as possible.



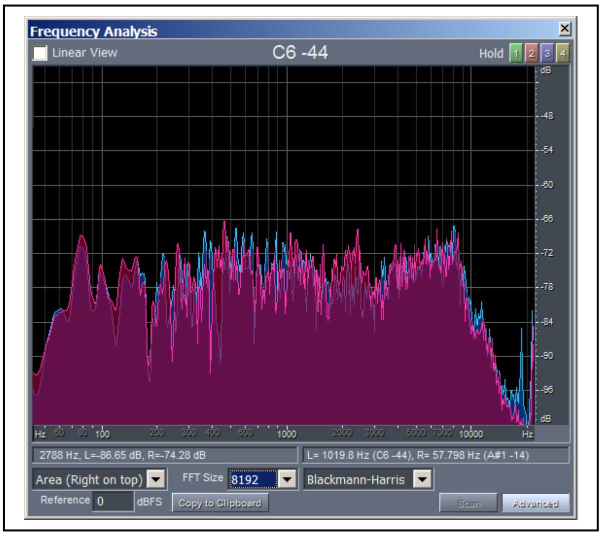


Figure 7: - left: frequency spectrum of the direct sound at seat DD11 as seen by the dummy head microphone. Right: the frequency spectrum at the same seat 160ms after the direct sound.

The direct sound component in most of these seats has a total energy at least 10dB less than the reflections and reverberation, so the direct sound spectrum is not audible. But the frequencies in the direct sound between 1000Hz and 6000Hz contain the harmonics that the ear uses to separate one instrument from another, and to detect proximity. Low frequencies are not required for this process. The brain automatically connects the later arriving low frequencies to the harmonics detected in the direct sound.

Once equalized it is possible to convolve the IR with one of the voices – I prefer the soprano – and listen with the calibrated headphones. It should sound completely natural. Sometimes it can seem a bit tinny, or shrill, or lacking in some other way. Figure out why, and fix it.

6 UNOCCUPIED VS OCCUPIED DATA

Our hall data was measured with an unoccupied hall and stage, but we would like to compare it to live recordings in the occupied hall. The reverberation time we measure in our data is about 2.3 seconds at 2000Hz. This is way too long. Fortunately a bit of mathematics can alter the IRs. We wrote a Matlab script that reduces the RT of the impulses to 1.8 seconds at 2000Hz.

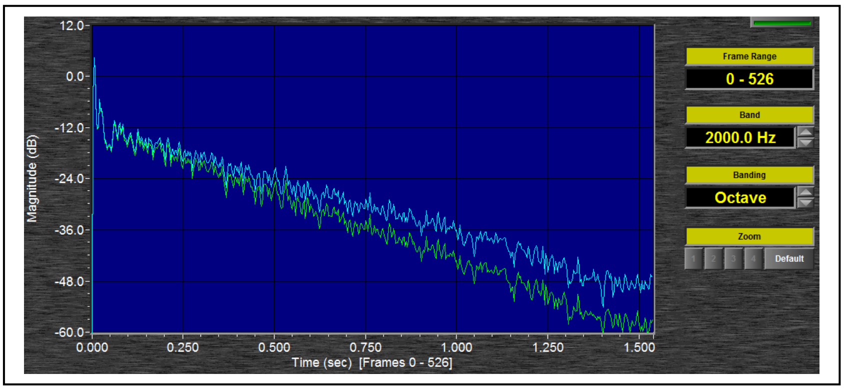


Figure 8: Blue – the decay curve of our data in the 2000Hz octave band, showing a 2.3s RT. Green: the same data as modified to 1.8s RT.

The sound synthesized with the longer reverberation time of the unoccupied hall is quite different from the sound synthesized from the reverberation time of the occupied hall. In seat DD11 there is even more reverberant masking, and the sound is even less attractive. This is not the case in a seat like R11. With the first reflection included the sound is still clear, and the extra reverberation is not problematic. However, if we delete the first reflection in this seat with the longer reverberation time something magic happens. There is better proximity, and the hall becomes much more audible and alive. The sound is similar to the front third of the Vienna Musicverreinsaal, or, even better, to many seats in the new Schermerhorn hall in Nashville.

7 CONVOLUTION

We are now ready to modify the direct sound component of the measured IR to create the different azimuths of each instrument. First copy the direct sound – assumed to everything from zero to 5ms from the beginning of the direct sound – into a new file, and set the region from which it came in the original IR to zero. Be sure the timing does not change. Check that the direct sound component has equal energy in both channels and a nearly identical spectrum. Make sure the peak amplitude occupies the same sample. All of this can be easily done in Adobe audition. If this is not the case, fix it.

For my experiments I used 7 front azimuths, left 22.5 degrees, 15 degrees, 7.5 degrees, 0 degrees, right 7.5 degrees, 15 degrees, and right 22.5 degrees. The most revealing of Lokki's ensembles for me is the Mozart. I have used just six of the performers, and the whole ensemble, but find the most revealing arrangement is the smaller one. The instruments in the ensemble were chosen to be:

Violin 1 left 15 degrees, Violin 2 left 7.5 degrees, Soprano 0 degrees, Cello right 7.5 degrees, Viola right 15 degrees, and Bass Viol right 22 degrees. (Lokki's recordings of the violins, cello, and viola were intended to be reproduced through multiple loudspeaker positions. I am playing them solo, so I raised their level by 4dB.)

To create these azimuths for my ears the recipe is easy. I attenuate the direct sound in the contralateral ear by 1.2dB and increase the time delay of that channel by one sample at 44.1kHz for every 7.5 degrees of right or left azimuth.

For the work in this paper I then make three copies the IR I wish to study. In one I have only the direct sound – from zero to 5ms. These are really multiple files that have been adjusted to the seven azimuths listed above. In another I have only the first reflection, assumed to be the same for each instrument. In the third I have everything else, assumed to be reverberation, and also the same for each instrument. The direct sound is then convolved separately for each instrument, and the results are summed. The same could be done for the first reflection and the reverberation, but it is quicker to sum the instruments, and then convolve the first reflection and the reverberation with the sum of the instruments.

It does not matter that the IRs for the reverberation and the first reflection are the same for all the instruments. There is no correlation between the instruments, and the difference between what the first reflection or the reverberation would be for each of them would not be audible.

All these convolutions can be done with Cool Edit or Adobe Audition. To save time I made a Matlab script that modifies the direct sound for azimuth and does all the convolutions. You simply tell it what seat IR to use. It outputs three sound files: The direct sound for all the instruments, the first reflection for all the instruments, and the reverberation for all the instruments. Summing the three files re-creates the sound of this ensemble in the hall – and it does. The script makes it easy easily generate the above three files for all the seats in the data collection. The same equalizations for the IRs works for all the seats in my data set, but I did take the time to be sure the direct sound of the measured data was balanced in the left and right channels. The head was not always properly centered on the source loudspeaker.

8 LISTENING

Listening to the reconstructed data ideally requires a pair of headphones individually equalized for the listener. A method for doing this is has been outlined above, and is described more fully in another preprint for this conference. However equalized headphones are not absolutely required. The difference in proximity can be heard even when the orchestra sounds inside the top of your head, as often happens with headphones. It can also often be heard with computer monitor loudspeakers. The recordings should sound natural however they are played, but to hear the difference clearly between when you delete the first reflection really requires a way of switching rapidly between the two conditions.

Our technique is to use Adobe Audition in its multitrack mode. The three files convolved by the Matlab script can be loaded in and played together. By pushing the mute button on the first reflection it can be switched on and off, although there is a delay of a second or two before the sound actually changes. The change is not always immediately audible. Remember that except for seat DD 11 the other seats in the data set are considered some of the best in the world, so if you can hear a difference it is a bit like gilding a lily. But the difference is there.

More interesting at first is to simply compare the different seats, with or without the first reflection.

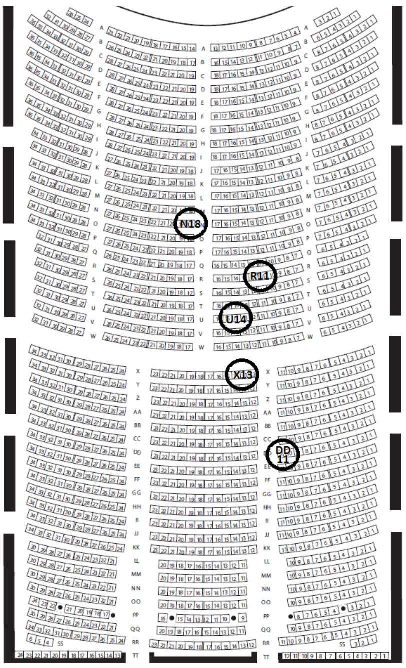


Figure 9: A seating chart for BSH showing the tested positions. The four closest to the stage are considered very good seats. Seat DD11 is not particularly good. To the author's ears and in his binaural recordings the sound is distant and muddy.

The author has a binaural recording of an orchestral performance in seat DD24, on the left side of the hall opposite to DD11. In the performance recording, (which you can hear at the conference) the sound was loud, fuzzy and distant. The author heard, but did not record, a performance four seats to the right of X13, in seat Y9. The sound was much better there than in DD24. It had proximity, but the beginnings of notes, where proximity can be detected, were often masked by reverberation. The piece was Strauss' Salome, which is thickly and continuously orchestrated.

In the binaural reconstruction from the measured data from seat DD11 the sound is similarly fuzzy. But when the first reflection – from the right side wall – is deleted the sound is at least as good as I remember from y9. But as in that performance experience, note onsets are often lost in the decay of previous notes.

The reconstruction of X13 sounds pretty good, with or without the first reflection. The direct sound is stronger there than in DD11, and there is less masking from the reverberation. Deleting the first reflection improves the clarity and the ability to localize quite a bit.

The sound in N13 as reconstructed from the binaural data is very fine. Sit there if you can! When we delete the right side wall reflection the difference is barely detectable. We find the sound to be a tiny bit better, but it certainly is not worse. R11 is different. Deleting the right side wall reflection noticeably improves clarity and localization, but it also enhances the audibility of the hall reverberation. I often sit during performances in seat R8, just three seats further to the right, and have begun to notice that the side wall reflection is audible, and does not an improve to the sound.

The data from seat U18 was from Ning Xiang's Head Acoustics dummy head, and the source loudspeaker was a large dodecahedral loudspeaker on stage. The frequency response of the reverberant sound was peculiar, but after several tries we got it more or less ironed out. However the first reflection from the right side wall was bizarre, with the frequencies around 2000Hz 10dB stronger than the others. Auralizing this seat produced a gastly result. There was clearly an interference lobe from the speaker that happened to bounce off that wall. We used Audition to equalize that reflection to flat, and the sound was much better. But it was still way to reverberant, more reverberant than the sound in X13. There was no sense of proximity, and a great deal of reverberant masking of notes. This was due to the omnidirectional source speaker. In our opinion one reason for the absence of the perception of proximity in previous concert hall research is the standardization of omnidirectional loudspeakers. We should use sources that have realistic directivity, or we will not hear what we are looking for.

But the beauty of this method is that we can compensate for the difference in loudspeakers. The Genelec is a bit too directive, the large dodecahedron too omnidirectional. Reducing the level of the reverberation relative to the first reflection and the direct sound brought the sound much closer to our other data. And deleting the first reflection made it clearly better.

8 VALUES OF LOC IN THE SYNTHESIZED SEATS

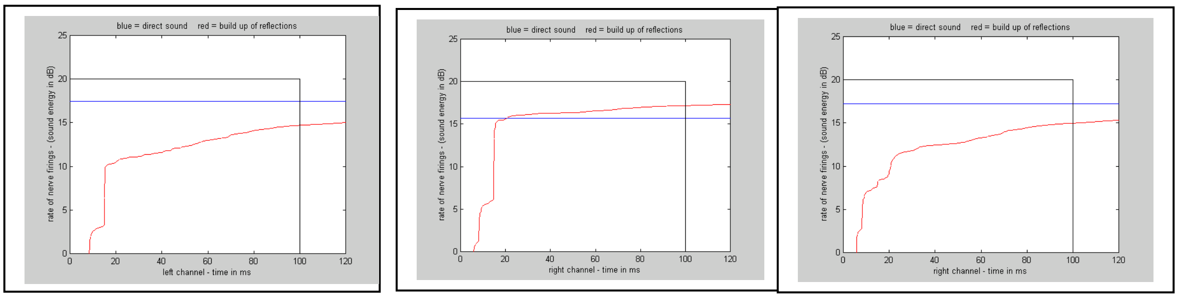


Figure 10: Graphs showing the level of the direct sound and the build-up of the reflected energy in seat DD11. Left: the blue line is the level of the direct sound, the red is the build-up of reflected energy in the left ear of the dummy head. Middle: The same data for the right ear. The direct sound is drawn to be lower than for the left graph because the drawing is normalized to the total energy. Right: the same drawing for the right ear after the first lateral reflection is deleted. LOC values are in order of the graphs, 6.7dB, 1.2dB, and 5.6dB.

LOC is a mathematical method of measuring the ability to localize sounds. It uses as input a binaural impulse response. A value of zero dB is considered to be the threshold at which male speech can be localized in the absence of noise, and in a uniformly decaying three-dimensional reverberant field. See [1] and [3]. We have been looking for a way to calibrate the measure LOC for real concert hall data, and also looking for a way to know the meaning of LOC if the measures are different in the left and the right ear. This seat give us an ideal opportunity to find out. As can be seen, the first lateral reflection at about 14ms is much stronger in the right ear than the left. The values of LOC reflect this difference. The LOC value in the right ear indicates poor proximity. However when that reflection is deleted that LOC value improves dramatically. The low value in the right ear would imply that localization and proximity would be poor, and this is definitely the case.

The implication of this is that when we measure LOC we should be concerned with the value in whichever ear LOC is the lowest. Having a high value of LOC in one ear is not sufficient for either the ability to localize or to perceive proximity.

9 CONCLUSIONS

We have shown that it is possible to manipulate existing binaural data in such a way that it can be auralized using Tapio Lokki’s anechoic recordings. Many experiments on the effects of reflections on sound become possible with this technique. In this preprint we focus on the effects of early lateral reflections on the sound in a famous shoebox hall. The results conclusively show that in every seat tested these reflections are either inaudible or undesirable. Deleting them from the measured data increases proximity and the sense of envelopment with no effect on bassiness or loudness.

We are just starting this work. The same data set includes a measurement in the front of the first balcony much further from the stage than DD11 which we have yet to auralize. The author has binaural recordings of performances from similar seats in the front of the first balcony, one with lightly scored music, and one with much more thickly scored music. The lightly scored music is absolutely beautiful in that seat, near the center. The average note length is short, and the late reverberation does not have time to build up excessively. The thicker music still has some sense of proximity, but there is a lot of masking of note onsets. One of these seats was far to the left. The left sidewall reflection was audible and disturbed proximity. We hope to verify these observations once we get more impulse response data. We will also try the more heavily scored pieces in Lokki's recordings.

We have also hinted at the beneficial effect of a somewhat longer reverberation time at seats where the direct sound is strong enough that it is not masked, and where the early reflections are weak enough that there is good proximity. This aspect deserves further study, as the result can be beautiful. In the author's opinion a good example of this can be heard in Nashville. Interested readers can hear examples of this hall during the conference from the author's binaural recordings.

We point out that although BSH is deservedly one of the world's finest halls, there are a great many seats closer to side walls than the ones studied here. We can simulate the sound in these seats with the current data set by simply increasing the level of the first reflection. The result is not pretty. In addition, there are a great many seats more distant than DD11. We can probably manipulate the data to auralize them too. It is not likely the sound will improve. Although fine shoebox halls can have seats with outstanding sound, we believe with careful design the number of these seats could be greatly improved. In the author's opinion an excellent example of how this can be done can also be heard in Nashville.

The author remains convinced that the current fetish about the necessity of early lateral reflections is a shibboleth. In his opinion the best hall sound for both symphonic music and chamber music has good proximity in all seats, adequate but not excessive late reverberation, particularly in the time range of 160ms, and sufficient early reflections that the late reverberation is not heard as an echo. The hall shape need not be a shoebox. Too much of the audience is too far away, and if the late reverberation is to be sufficient for good envelopment in the front, it may well be too strong for clarity in the rear.

Designs that bring the audience closer to the stage can work well if they have sufficient volume to generate strong late reverberation. They might look more like theaters, but with high ceilings that are not surrounded by absorbing surfaces. If in fact they are theaters, or high ceilings are not in the budget, modern electronic architecture can seamlessly add completely natural late reverberation. If you don't tell, no one will know.

The impulse responses and the convolved sound files used in these experiments will be on the author's web-site. A research version of our headphone app is available by request, and others interested in acoustic research of all types are encouraged to use it. We hope to have a simplified, platform-independent version for music lovers by the time of the conference.

10 REFERENCES

1. Griesinger, D. Acoustic quality, proximity and localization in concert halls: the role of harmonic phase alignment. *Psychomusicology:Music, Mind & Brain, 25*, (xx-xx). (in press)
2. Griesinger, D., (1997). The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces. Acta Acustica 83, #4, 721-731
3. Griesinger, D. (2011). Clarity, Cocktails, and Concerts: Listening in Concert Halls. *Acoustics Today*, 1, 15-23.
4. Kujawa, S., & Liberman, M. (2009). Adding insult to injury: Cochlear Nerve Degeneration after "Temporary" Noise-Induced Hearing Loss. *Journal of Neuroscience,* 29(45):14077–14085

5. Kujawa, S., & Liberman, M. (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hearing Research* http://www.sciencedirect.com/science/article/pii/S037859551500057X

6. Laitinen, M, Disch, S., Pulkki, V., (2013) Sensitivity of Human Hearing to Changes in Phase Spectrum. *Journal of the Audio Engineering Society, 61,* 860-877
7. Licklider, J. (1951) A duplex theory of pitch perception. *Experientia,* **Vol VII/4 128-134**.
8. Lokki, T., Pätynen, J., Kuusinen, A., & Tervo, S. (2012). Disentangling preference ratings of concert hall acoustics using subjective sensory profiles. *The Journal of the Acoustical Society of America,* **132,** 3148-3161.
9. Pätynen, J., & Lokki, T. (2011). Evaluation of Concert Hall Auralization with Virtual Symphony Orchestra. *Building Acoustics, 18***, 349-366**.
10. Shi, G., Sanechi, M., M., Aarabi, P., On the Importance of Phase in Human Speech Recognition. IEEE transactions on audio, speech, and language processing, vol. 14, no. 5, September 2006