Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Challenges and Solutions in Acoustical Measurement and Design: Paper ICA2016-379

What is “proximity”, how do early reflections and reverberation affect it, and can it be studied with LOC and existing binaural data?

David Griesinger[a]

[a] David Griesinger Acoustics, United States, dgriesinger@verizon.net

Abstract

Lokki has recently shown that there is no correlation between most current ISO 3382 hall measurements and preference, and that the most prominent perception preferred by his listeners is currently unnamed and unmeasurable. He chose to name the perception “Proximity” because he found it was related to the auditory sense of being close to the performers. This paper proposes that proximity arises from the phase alignment of the upper harmonics of speech and most musical instruments. We present data from other fields that shows that the loss of phase alignment due to early reflections or masking can greatly decrease the ability to separate signals from noise and other signals. We will then show how convolving some existing binaural data from Boston Symphony Hall with Lokki’s anechoic recordings can create a realistic binaural rendition of an instrumental ensemble, which can be used to test the effects of early reflections on proximity, localization, and loudness in these seats. We find that in all our seat positions attenuating the side wall reflection in Boston either improves the sound, or is (in very good seats) inaudible. These effects are predicted by the author’s binaural measure LOC.

Keywords: Proximity, Binaural Technique, Localization, Concert Halls, LOC

What is "proximity", how do early reflections and reverberation affect it, and can it be studied with LOC and existing binaural data?

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. 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 proximity, the perception that the sound source is acoustically close to a listener, been both previously unknown and un-measureable?

An important clue can come from a simple experiment with live sound or an electronic string quartet using Lokki's anechoic recordings. [1] As the quartet plays tirelessly on stage we walk backward away from them into the hall. (If the ensemble is live it is very important not to look at them, as vision will almost always trump hearing.) At first it is easy to localize each instrument, and tell which played each note. The sound has an exciting, attention-grabbing quality. As you walk back the sound remains close and exciting. But suddenly in distance range of one or two rows all the instruments blend together into a fuzzy ball of sound. The attention-grabbing effect is gone. We call the distance at which this happens the Limit of Localization Distance, or LLD.

The difference is not subtle, and there is very little difference in the LLD for different individuals. Binaural recordings made in front of and behind the LLD sound very different, even on loudspeakers. The LLD appears to be a property of the sound field, and how the ear and brain system has evolved to decode it. We propose that the presence or absence of proximity can be predicted by the LLD.

We need to predict from measured data whether a seat position is inside or outside the LLD. To develop such a measure we need to accurately reproduce a particular hall sound in the laboratory, with the ability to modify the reflection patterns at will. Lokki's method is a step in the right direction. This preprint describes a binaural technique that can do the same thing with much less complexity. We will present results from a data set in Boston Symphony Hall that demonstrate the power of the method.

2 Proximity, Localization and Phase

We have been studying what we now call proximity for more than ten years. We propose that 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 a source and to perceive its closeness. We believe that our ability to localize must result from abilities our ears and brains have evolved to distinguish the direct component of a sound field from reflections that follow.

How is this done? We find that the direct component and the following reflections have very different amplitude envelopes. The direct sound from speech and most musical instruments is created by a repeating impulse, the opening of the vocal cords, a reed, rosin on a bow, etc. As a consequence once in each fundamental period the harmonics are forced to align in phase, creating peaks in the pressure amplitude at the fundamental period.

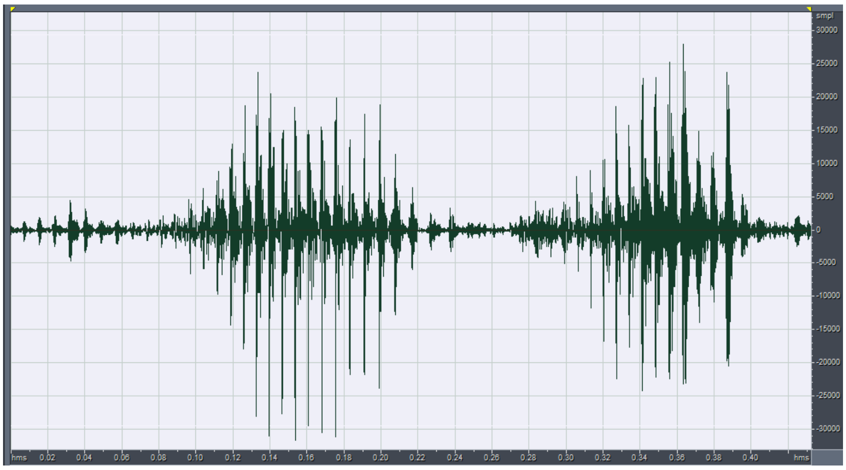

Figure 1: The syllable “one” first filtered through a 2kHz 1/3 octave 2nd order Butterworth filter, and then though a 4kHz filter. Note the prominent envelope modulation at the fundamental period, with peaks more than 10dB above the minima between peaks. Although the ear is not sensitive to the phase 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 has an important effect on the envelope of waveforms 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.” Licklider [7] proposed an autocorrelation mechanism located in the basilar ligament that explains our acute sense of pitch. His mechanism, in combination with the envelope peaks, not only can explain pitch acuity and the detection of ITD, it can enable the separation of several pitched signals into independent neural streams.

But reflections from all angles interfere with the phases of harmonics in the direct sound. The phase alignment is lost, 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.

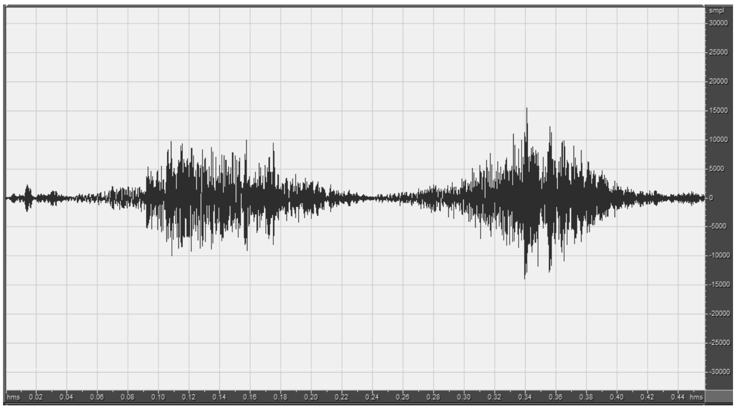

Figure 2: The same signals as figure 1, but altered in phase by a filter made from three series allpass filters of 371, 130, and 88 samples and allpass gains of 0.6. Notice that the peaks at the fundamental period have largely disappeared. When you listen to these signals no fundamental tone is heard.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

We believe a primary mechanism for both source separation and the perception of proximity resides in the spiral ganglia just below the hair cells in the basilar ligament, and that the mechanism relies on the phase alignment of the upper harmonics of music and speech. The properties of the spiral ganglia are only just beginning to be studied. [3] [5] [6] They are now known to be extremely fragile to continuous loud noise, and are thought to be vital to our perception of speech in noisy environments. Aarabi et al find that randomizing the phase of speech in the presence of noise can easily double the word error rate. [10]. Why has this mechanism and its effects on our perception of acoustics been unknown or forgotten for so long?

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. To detect the presence or absence of proximity we need a laboratory system that accurately re-creates the phase relationships between the harmonics of each individual instrument, along with the reflection field that follows it. The current ISO 3382 measurement techniques were developed primarily through listening tests that used two channel anechoic recordings of orchestras played back through two loudspeakers on a concert stage. The phase information of each instrument was already lost, and proximity is not there to be heard. The same thing happens if a single instrument is reproduced through multiple loudspeakers, as happens with Wave Field Synthesis (WFS), or most Ambisonic systems.

Lokki [9] uses an electronic orchestra with a single pair of speakers for each instrument, one pointing up and one pointing out. Each pair plays independent anechoic recordings of orchestral parts. Separate three-dimensional intensity impulse responses from each instrument are recorded at the seat position under test. The direct sound from each instrument is reproduced through a single loudspeaker as close as possible to the original azimuth of the instrument, and the reflections are reproduced in a similar way without panning. His system is the first I have heard that reproduces the sense of proximity. Lokki's system is a big step forward. But does it really reproduce the sound of a particular seat in a particular hall? How can we know?

3 Binaural technique

This preprint describes a series of experiments that use a version of Lokki's virtual orchestra 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 along with Lokki's anechoic recordings to synthesize the sound of a musical ensemble. It is also possible to compare the sound from an electronic ensemble to live musicians. Binaural technology, if it can be made accurate and convincing, can be a reference by which other methods of orchestral reproduction can be compared. But to work properly, it is absolutely necessary to have accurate equalization, all the way from the sound outside the microphone *to the eardrum* of the listener.

Experiments at IRCAM [11], as well as the author's work, has shown that if the equalization is correct at the eardrum external frontal localization can be achieved without head tracking. But to do this it is essential to measure the eardrum pressure directly or indirectly for each listener. For many years the author has been recording data and live music with probe microphones at his eardrums. Equalizing headphones with the same probes at the eardrums allows the author to play back these recordings with startling accuracy. The experience can be stunningly beautiful – recreating the experience of the performance better than any commercial recording technique.

Two of our three dummy head microphones are fitted with silicone castings of my own pinna and ear canals, and one is a standard Neumann KU-81. All of them are equalized for flat response from a frontal plane wave up to a frequency of about 6kHz. The author finds the difference between them to be minimal. The author's personal eardrum recordings are equalized the same way. This type of equalization makes the dummy head microphones and my own eardrum recordings similar in timbre to a high-quality studio microphone, but with a far different directivity. An additional advantage of this equalization is that our binaural recordings can sound natural when played on loudspeakers.

But do recordings from my ears played through headphones equalized for my ears work for other listeners? In general, the answer is no. But we have found that most of the variation between individuals comes from the transfer function from the headphone to the eardrum, which varies enormously between individuals. If we individually calibrate a pair of headphones to a particular listener the result with my personal recordings can be surprisingly robust. Almost all listeners report frontal localization and a realistic timbre. The presence or absence of proximity can be easily determined.

The most obvious method of matching these recordings a listener is to use probe microphones at the eardrums, and match the headphone response to the response at the eardrum to a frontal loudspeaker. The procedure is fast and accurate, but invasive.

We have developed a computer application that uses equal loudness methods similar to ISO 226 to match a pair of headphones to a listener using their own eardrums as microphones. This method is described in another preprint for this conference. The method results in the perception of accurate timbre, and almost always permits frontal localization without head tracking. Once fitted with individually equalized headphones most listeners find our binaural recordings to be realistic.

4 Measurements in Boston Symphony Hall

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, but this error can be corrected. Impulse responses were recorded with the author's dummy head avatar, and a small soundfield microphone.

In all the measurements the low frequencies of the direct sound are steeply rolled-off by the seat-back effect, and there is a high frequency boost from about 2kHz to 5kHz. See the left hand picture in figure 3. But to sound natural the later reverberation should be roughly flat from

60Hz to about 3kHz, rolling off as time goes on. This spectrum depends on the frequency response of both the dummy head and the source. The second picture in figure 3 shows the spectrum after about 160ms in our tests after careful equalization.

Once equalized it is possible to convolve the IR with one of the Lokki's anechoic voices – I prefer the soprano – and listen with calibrated headphones. It should sound completely natural. The equalization needs to be adjusted if sounds tinny or shrill. Like Manfred Schroeder, we also compensate for the reverberation time of the unoccupied hall by changing the RT of our impulse responses with a Matlab script to about 1.8 seconds at 2000Hz.

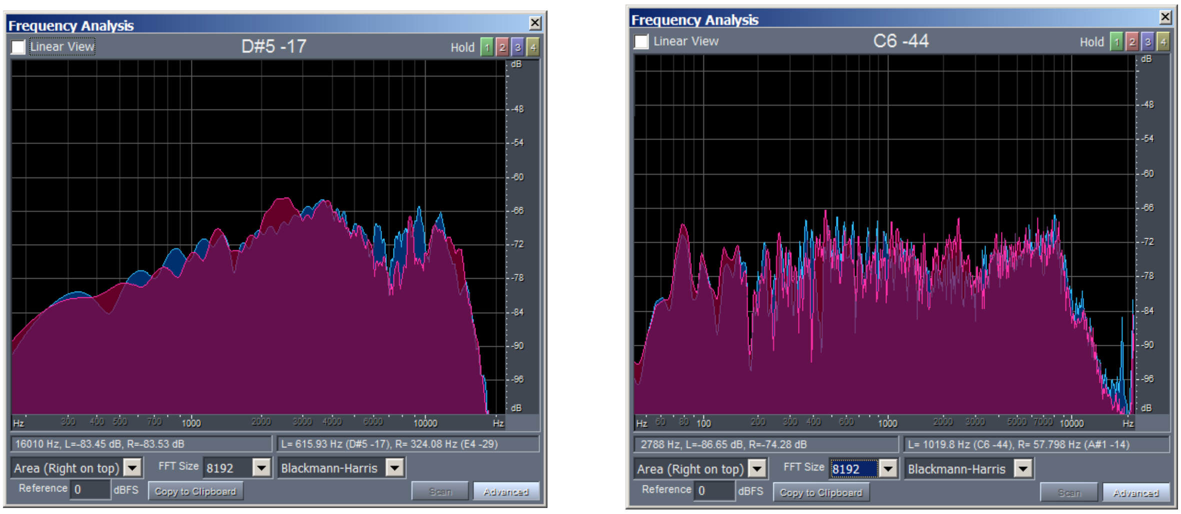

Figure 3: - left: frequency spectrum of the direct sound at seat DD11 as seen by the dummy head microphone. Right: the frequency spectrum at the same seat 160ms after the direct sound.

6 Convolution

The next step is to modify the direct sound component of the measured IR to create the different azimuths of each instrument. We find the most revealing of Lokki's ensembles is a six voice ensemble from an aria by Mozart. The azimuths chosen were: 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 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.

A Matlab script modifies the direct sound for azimuth and does all the convolutions with Lokki's anechoic instrument tracks. You simply tell the script what seat data to use. The script 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 by selectively muting one or more the effects of a particular reflection can be easily auditioned.

7 Results

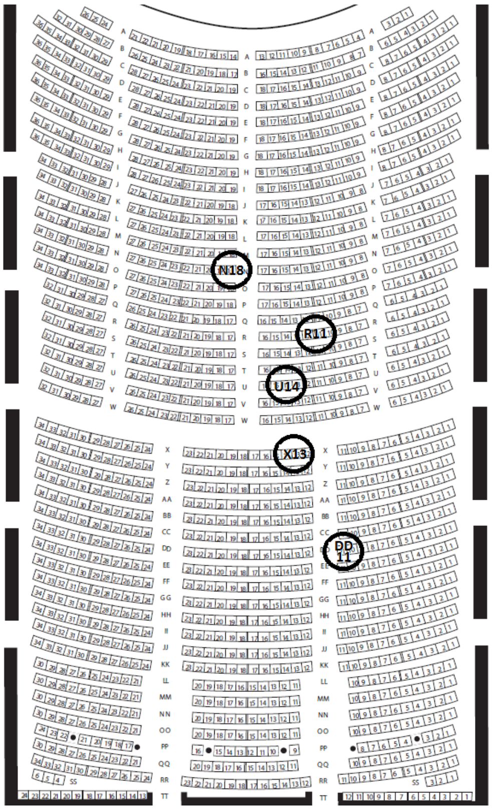

Figure 4 A seating chart for BSH showing the tested positions. The four closest to the stage are considered to be very good seats.

The author has binaural recording of an orchestral performance for most of these seats, although my recording for DD11 was from seat DD24, on the left side of the hall. In the performance recording the sound was loud, fuzzy and distant. The sound of the binaural reconstruction of seat DD11 is very similar. But when the first reflection from the right side wall is deleted the sound becomes clear and present, at least for lightly scored music. When loud notes are held reverberation can mask the direct sound.

The reconstruction of seat X13 sounds good with or without the first reflection. The direct sound is stronger there than in DD11, and there is less masking from the reverberation. But deleting the first reflection slightly improves the clarity and the ability to localize.

The sound in N13 as reconstructed from the binaural data is very fine. When we delete the right side wall reflection the difference is barely detectable. R11 is different. Deleting the right side wall reflection noticeably improves clarity and localization, and also enhances the audibility of the hall reverberation. I have sometimes sat for performances in seat R8, just three seats further to the right. The side wall reflection is audible, and does not improve clarity.

8 Values of LOC in seat DD11

LOC is a measure developed by the author for predicting from a binaural impulse response if a sound will be localizable. It is based on the idea that if the integrated log of the direct sound above 1000Hz is stronger than the integrated log of the reflections in a 100ms window, then localization is possible. [3] [12] [13]. Matlab code for calculating LOC is on the author’s website.

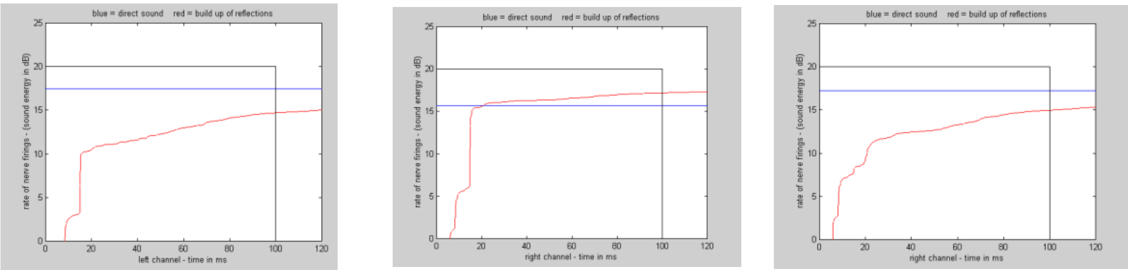

Figure 10: Graphs showing the level of the direct sound and the build-up of the reflected energy in seat DD11. Left: the blue line is the level of the direct sound, the red is the build-up of reflected energy in the left ear of the dummy head. Middle: The same data for the right ear. It can be seen that the strong reflection at 17ms in the right ear greatly increases the integrated reverberant level. 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.

The values of LOC for seat DD11 in the left ear of the dummy is high. The value in the right ear when the side-wall reflection is present is significantly below +3dB, which is generally defines the limit of localization. When the right side wall reflection is deleted, LOC is high, and proximity should, and does, result – if reverberation does not build up too much. Without the sidewall reflection DD11 would be a good seat with lightly scored music, but possibly muddy when long, loud notes are held.

9 Conclusions

We have proposed that binaural technique, when combined with careful equalization and individual headphone equalization, can be a valuable reference for concert hall reproduction. When combined with an electronic orchestra similar to Lokki’s, [9] measured binaural impulse responses from individual instruments can be manipulated to test the audible effects of reflections from different directions. The ability to study these factors is vitally important to understanding proximity, which may be one of the most important predictors of preference in halls. We have also shown that considerable benefit can be obtained from previous binaural data if care is taken to correct the data for equalization and other factors.

We have not performed elaborate listening tests with many subjects. We hope these experiments will inspire such work. But the differences we observe in the experiments above are not subtle. Audio examples from this work are on the author's web page. We urge readers to listen to them. You don't need elaborate equalization to hear the results. The sound may not be entirely natural, but the differences in proximity are very clear, even over loudspeakers. The sounds speak for themselves.

References

[1] Griesinger, D. Acoustic quality, proximity and localization in concert halls: the role of harmonic phase alignment. *Psychomusicology:Music, Mind & Brain, 2015, vol 25*, #3 339-344.

[2] Griesinger, D., (1997). The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces. Acta Acustica 83, #4, 721-

[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., 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

[11] Personal communication from Eckhard Kahle

[12] Griesinger, D. What is clarity and how can it be measured J. Acoust Soc. Am 133, 3224-3232 (2013)

[13] Beranek L. Concert hall acoustics: Recent findings J. Acoust. Soc. Am. 139(4) April 2016