THE IMPORTANCE OF THE DIRECT TO REVERBERANT RATIO IN THE PERCEPTION OF DISTANCE, LOCALIZATION, CLARITY, AND ENVELOPMENT

David Griesinger
23 Bellevue Avenue
Cambridge, MA 02140 USA
www.davidgriesinger.com

Abstract

The Direct to Reverberant ratio (D/R) - the ratio of the energy in the first wave front to the reflected sound energy - is absent from most discussions of room acoustics. Yet only the direct sound (DS) provides information about the localization and distance of a sound source. This paper discusses the perception of DS in a reverberant field depends on the D/R and the time delay between the DS and the reverberant energy. Threshold data for DS perception will be presented, and the implications for listening rooms, hall design, and electronic enhancement will be discussed. We find that both clarity and envelopment depend on DS detection. In listening rooms the direct sound must be at least equal to the total reflected energy for accurate imaging. As the room becomes larger (and the time delay increases) the threshold goes down. Some conclusions: typical listening rooms benefit from directional loudspeakers, small concert halls should not have a shoe-box shape, early reflections need not be lateral, and electroacoustic enhancement of late reverberation may be vital in small halls.

1. Introduction

My own research into spatial acoustics – which I presume to be the heart of "immersive audio" – was greatly influenced both by my preoccupation with sound recording and reproduction, and by the work of pioneers such as Michael Barron, who studied the perceptual effects of adding (lateral) reflections to the direct sound from a loudspeaker. Much of interest has come from this type of study. My own contributions include research into how our physiology decodes reverberant sound, how we determine the loudness of the reverberant component of a sound field, and the effects of different combinations of early and late reflections on recorded sound.

For example, we have found that (lateral) reflections in the time range of 10 to 50ms contribute a sense of distance to a closely miked source. The source is pushed away from the listener, and behind the loudspeakers. A room impression is generated in front of the listener. Reflections later than 50ms create a "spaciousness" impression that surrounds the listener. The effectiveness of these reflections in producing the surround impression increases as the delay increases, up to about 160ms. If there are a collection of reflections of approximately equal level between 50 and 160ms a strong enveloping reverberance is perceived, with no sense of echo.

Michael Barron found much the same result using single reflections, as shown in the well known drawing in Figure 1. Note the vertical axis is the ratio of reflection to direct (R/D). This paper will concern itself with the inverse: the direct to reverberant ratio (D/R). The vertical axis of Barron's diagram covers the D/R range of +20dB to 0dB.

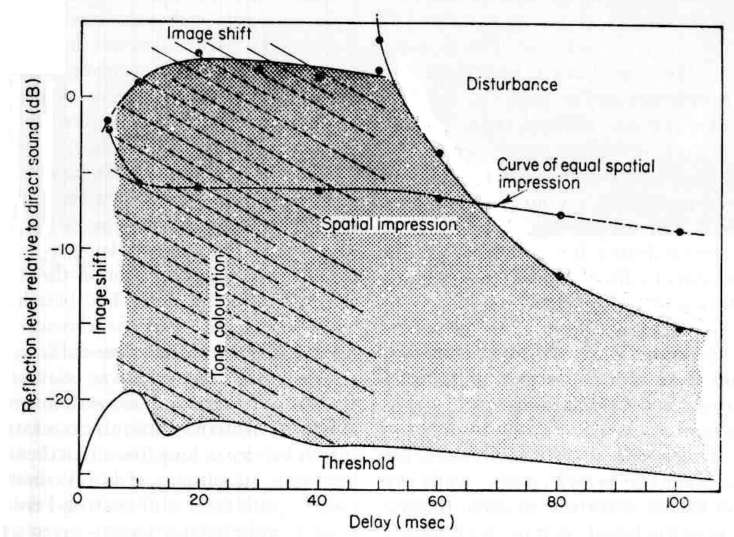

Figure 1: The subjective effects with music of a single side reflection as a function of reflection level and delay relative to the direct sound (azimuth angle 40 degrees). From Barron.

Barron's and our research has been both interesting and useful in understanding sound recording and reproduction. Recording engineers often start with a collection of tracks that are essentially free of reverberation. They mix these tracks together and then add (naturally or artificially generated) reflections and reverberation to them to create a natural sounding final product.

We have found that for nearly all recorded music – classical or popular – the D/R is never less than +4dB. Given a choice of how to set both the early reflection level and the reverberation level, both sound engineers – and even the acousticians I managed to test (Beranek for example) choose a direct to reflected energy ratio (D/R) between plus 4dB and plus 6dB. The "curve of equal spatial impression" in Barron's figure is drawn at this level.

But Barron was not principally interested in recorded sound. He was trying to understand concert halls. Very few seats in a hall have a D/R in this range. In a typical shoebox hall the critical distance (the distance where the direct sound and the reflected sound are equal) is 7 meters or less. All seats beyond this distance from a source will have a D/R less than one. In fact, in many halls half the seats have a D/R of minus 8dB or less.

This paper shows the results of experiments that probe the case where the direct sound is typical of real halls. A primary tool is a simple image source models of a hall which compute a binaural signal corresponding to the sound of a single instrument at a particular seat. The direct component of the signal is then varied experimentally, and a subject reports how the spaciousness and the localization of the source varies.

Many alterations of the reverberant signal are possible. For example, specific reflections can be amplified or attenuated, the directional properties can be eliminated or swapped over particular time ranges, etc. Interesting results emerge. Although these results are not directly applicable to recorded sound or reproduction, they have important implications for hall design.

2. Experiences

Before going into details, I want to present some personal experiences which were greatly influential if forming my interest in the properties and consequences of direct sound in halls. Perhaps the first of these was during the installation of the reverberation enhancement system in the Deutchestaatsoper in Berlin. Barenboim had decided to present Wagner's Ring cycle in the Staatsoper, for the first time since WWII. But he was not happy with the acoustic. The Staatsoper has a natural reverberation time at 1000Hz of 0.9 seconds, and Barenboim wanted something closer to Bayreuth, at 1.7seconds. Albrecht Krieger had heard my demonstration at the Schaubuhne Berlin, and suggested to Barenboim that we try the system in the Staatsoper. As a result, Krieger and I installed a shoestring system before the opening of Rhenigold, and I adjusted it till I liked it with my own singing. Krieger was delighted.

But Barenboim was NOT delighted. Horrible, he said. Good on the orchestra, horrible on the singers. He gave me 20 minutes to get it right – and I did. I installed a shelving filter in the microphone inputs, which reduce the reverberant level – not the reverberation time – by 6dB above 500Hz. Barenboim was delighted. Don't ever change it, he said, and jumped back into the pit.

The sound was indeed wonderful, and remains so to this day. The singers are almost unaffected by the reverberation enhancement. Most singers like the effect. One of them claimed that the hall remains as dry as ever – but I think he never heard the hall with the system off. This is distinctly not my impression as a singer. I find the enhancement is very audible, and quite helpful. Barenboim was right (as usual). The sound in the audience is glorious. The orchestra is full and rich (the RT below 500Hz is 1.7 seconds), while the singers are clear. There is enormous dramatic intensity. The hall is intimate to begin with, but there is a tremendous dramatic connection between audience and singer – they seem to be singing or talking directly to you. You are not at some comfortable distance observing the scene – you are in it. In the few times the system has needed adjustment (some of the old East German equipment was due for retirement long before) Barenboim has heard the problem immediately from the pit, and has vocally demanded repair. It took two years for anyone to realize there were electronics in the hall. All the improvement was attributed to Barenboim's skill as a music director – perhaps rightly so.

When we installed system in the Amsterdam Musiktheater, I was able to spend a lot of time with Peter Lockwood, the assistant conductor. He sat with me in a variety of seats while listening to rehearsals. We carried a remote control that allowed us to vary the D/R in half-dB steps. We lowered the D/R gradually, and the sound took on definite richness and depth. (The same reverberant level shelving filter used in Berlin was employed.) But a one point Peter said STOP – that's too much! I could not hear the difference. Listen, he said. With that one extra half-dB the singer moved back 10 feet! So I listened – and for the first time I appreciated the critical effect reverberation has on the apparent distance of an actor. When the reverberation is below some threshold the singers are perceived as close to the listener – dramatically they occupy the same space. Just a tiny bit more and they leave the space of the listener, and occupy a different space far away. Peter wanted the dramatic connection – and do did Haenchen, the music director.

Haenchen was responsible for us being in the Musiektheater. He had recently come from the position of music director at the Semperoper Dresden, which is a small opera with an RT occupied of 1.5 seconds. The Musiektheater is a larger space, with a RT below 1.3s at 1000Hz. Haenchen wanted me to reproduce the sound of the Semperoper in Amsterdam. As it happened, I had visited the Semperoper some years before and had binaural recordings of how it sounds. It is small and quite reverberant. The orchestra is loud, the singers are often overpowered, and often unintelligible. I could plausibly reproduce this acoustic in Amsterdam. Haenchen did not like it at all. I am convinced he was not responding to artifacts in the system, but to the lack of articulation in the singers. Turn it off! is all he could say. It is interesting – and the topic of another paper – how human perception can adapt to a particular acoustic and think it wonderful. But when there is the opportunity to change the acoustic rapidly in an A/B comparison, very different preferences are expressed.

But Hanenchen was very happy with the final adjustment we made with Peter. When the opera directors I have worked with are given the chance to choose immediately between orchestral richness and dramatic intensity they choose dramatic intensity. When this choice does not exist, they LOVE the sound of the orchestra, and the singers be damned.

After a similar experience with Michael Schonewand in Copenhagen, I was asked to install a system in a shoe-box shaped drama theater in a building next to the old Royal Theater. The object of the system was to improve the intelligibility of actors during a conventional (speech) drama. We had previously installed about 60 Genelec 1029 loudspeakers around the audience, and a pair of Gefeil line-array microphones at the balcony fronts. Alas, the microphones were well beyond the critical distance, and it was not possible to pick up the direct sound from the actors cleanly. I designed a fast acting gate that opened at the beginning of every syllable, and promptly closed just after the syllable ended. This signal was routed with complicated varying delays to the loudspeakers, and the system worked. Both loudness and intelligibility were improved.

A test was arranged. Five of the principle drama directors in Copenhagen were invited to listen in various seats to a live performance of “Uncle Vanya” in Danish, with a full paying audience. The system was turned on and off every 10 minutes, with no audience complaint. At the end of the performance we all got together. The directors were unanimous. The system works, we don’t like it, turn it off. Why? The directors were not sure. Finally a reason emerged. The system made the actors louder, and more intelligible, but they sounded further away. The value of D/R had been pushed too low by the system, and the actors disappeared into a different space. “I would rather the audience did not understand the words than to lose the dramatic connection between the actor and the listener” said one of them. “If the audience can’t hear the words they have to concentrate harder – which is just what I want.” But all the directors agreed that they had to train the actors better. The older actors were always intelligible. Some aspect of training had been neglected. When I related this to Schonewand, who had just conducted “Tristan”, he was most amused, and not surprised. “These young actors don’t know how to speak Danish!” he said.

All these experiences have in common the critical relationship between D/R and audience involvement. At the IOA conference in Oslo, Krokstad gave a wonderful lecture, where he insisted that what we as acousticians (or sound engineers) were seeking was involvement, not envelopment. His final slide was of the Theatre de Colon in Buenos Aires. ‘Is this the opera house of the future’ he asked? Our experiences in the new Oslo opera indicate that he is probably right.

3. Main Points

1. The ability to hear the *Direct Sound* – the sound energy that travels to the listener without reflecting – is a vital component of the sound quality in a great hall.

The direct sound – some call it the first wavefront – provides the human brain with the only accurate information about the direction (elevation and azimuth) of the sound source. Where there is sufficient direct sound energy, and sufficient time between the direct sound and the reverberation that the brain can separately perceive the direct sound, both the intelligibility and the sense of connection to the sound source are optimized.

As we will see, the ability to separately perceive the direct sound depends on time. Providing the reverberation from a previous sound as sufficiently decayed, the onset of a new sound is always uncorrupted by reflections. But the brain may not be able to separate this onset from the reverberation that follows if the time delay is too short, or if the sound has too gradual a rise-time.

If we want to optimize audience *involvement* as Krokstad insists, we need to make the direct sound separately perceivable.

2. Current acoustic measures neglect the audibility of direct sound and audience involvement.

Measures such as C80 lump the direct sound together with a great many early reflections, particularly in small halls. These halls can have what look like good clarity values, but localization and sonic distance can be poor. The current IEC definition of EDT ignores the direct sound, unlike the original definition by Jordan. (And the value obtained depends on the sample rate.) This is why EDT and RT are almost always the same. Calculations of acoustic parameters based on a Schroeder integral obscure the actual D/R in practice, since they assume an infinitely long excitation. Musical notes are seldom infinitely long. Short notes excite the hall less, which increases the D/R and increases the effective critical distance.

We need much better ways of measuring the direct sound and its relationship to later reverberation. Several such measures will be demonstrated here.

3. Hall shape does not scale

Our ability to perceive the direct sound depends on its *level* compared to reflected sound, and on the *time-gap* between the two. Both the direct to reverberant ratio (d/r) and the time-gap change as the hall size scales – but human hearing (and the properties of music) do not change. A hall shape that provides great sound to a high percentage of 2000 seats may produce a much lower percentage of great seats if it is scaled to a capacity of 1000. Smaller venues produce a shorter delay between the direct sound and the later reverberation. If audience involvement is to be preserved we must:

a) specify a shorter reverberation time
b) bring the audience closer to the musicians

If we demand a longer reverberation time, we must design the hall so the longer reverberation time does not result in a higher reverberation level, as conventional theory says it must. This seemingly impossible trick can in fact be done, both electronically and with (expensive) architecture. (For example, the new Oslo opera house is very well liked as a symphony hall.).

4. Small halls have too many early reflections, not too few.

Although acousticians generally believe clarity can be improved in small halls by *adding* early reflections, the opposite is often the case. Small halls suffer from too high a reverberant level, and a too rapid a build-up of reverberation. Both problems can be alleviated by selectively absorbing or reflecting the earliest arrivals away from the audience. This delays the onset of reverberation long enough to allow the brain to detect the direct sound.

5. The way sound builds-up in a hall after the onset of a sound is at least as important as the way the sound decays – as we will see in the next section.

4. Reverberation build-up

The most important parameter that a sound engineer adjusts when adding reverberation to a music mix is the level – the loudness – of the reverberation. This control is equivalent to adjusting the D/R in a hall. Changing the D/R by one or two dB has a far greater effect on the sound than changing the reverberation time by 30 percent or more.
But sound engineers know that the pre-delay control, the one that changes the delay before the onset of reverberation is also very important to the sound. As the pre-delay increases the effective loudness of the reverberation increases, allowing the engineer to increase the D/R while preserving the hall spaciousness. The clarity of the recording increases. Too much pre-delay and the reverberation is perceived as an echo – and this is clearly undesirable. For this reason some devices include a method of stretching out a series of early reflections to fill the empty time between the direct sound and the bulk of the reverberation. (At Lexicon we called these controls "shape" and "spread".)

Very surprisingly, both D/R and reverberant level have been routinely ignored in hall acoustics. Two years ago I was asked to write a review article for the IEEE on hall acoustics. The article was supposed to present a mathematical theory of halls that would explain why they sounded the way they do. Conventional theory – Morse and Beranek – could not explain the obvious differences in sound between even the best halls, such as Boston Symphony Hall (BSH), and the Amsterdam Concertgebouw (CB). A simple binaural model of these halls showed clear differences in sound even when the RT was adjusted to be identical. I decided to look carefully at the way sound builds up.

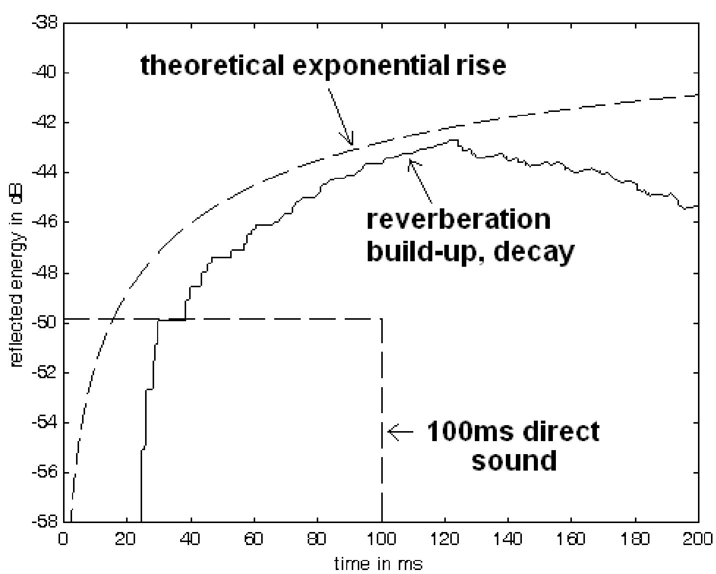

Figure 2: Build up and decay of reverberation in Boston Symphony hall at a position 70 feet from the stage. The excitation is a note 100ms long. The dashed curve is the theoretical build up of energy from an infinitely long excitation of the same amplitude. Note the ~23ms time delay before the amplitude of the reverberation exceeds the amplitude of the direct sound.

BSH is a wonderful hall. There is clarity and localization over a wide range of seats. From about row 20 to row 40 the sound is nearly identical and very good. It degrades slowly further back. The sound is identical because once the threshold for the detection of direct sound is exceeded, very little changes as the D/R increases further. The loudness is almost entirely provided by the reflected energy.

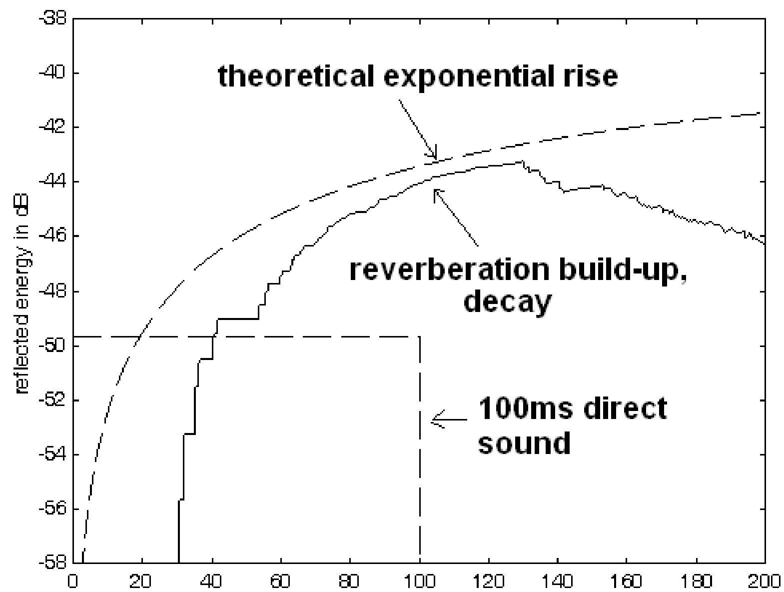

Figure 3: Build up and decay of reverberation in the Concertgebouw at 60 feet from the stage. The excitation is a note 100ms long. The dashed curve is the theoretical build up of energy from an infinitely long excitation of the same amplitude. Note the ~35ms time delay before the amplitude of the reverberation exceeds the amplitude of the direct sound.

Figures two and three show a very clear difference in pre-delay between the two halls. Boston has the shorter pre-delay, and is perceived as less reverberant, and less spacious, at least to the author. The Concertgebouw is more reverberant and spacious, with a marvelous clarity about the direct sound. But with some instruments, such as solo piano, there can occasionally be disturbing echoes.

Note that the direct sound plays a vital role in these two halls. Without the direct sound there is no pre-delay, and the sound is identical. If we measure the threshold for the detection of direct sound in the two halls we find a lower threshold in the Concertgebouw, implying that there are more seats with good localization and clarity.

This idea has implications for hall design. What happens if we build a shoebox hall with half the dimensions of BSH? (This is a very popular idea). Instead of 2700 people, the new hall will hold about 700. What happens to the build-up and decay? Let's assume we do not change the design of the hall – so the reverberation time will be half. Let's also keep the source and receiver position exactly the same, except scaled. In this case the only variable that should change is time. RT should go down a factor of two, as should the critical distance. We would expect the D/R at our chosen seat to remain the same – as should the average D/R through the hall. Is this what happens?

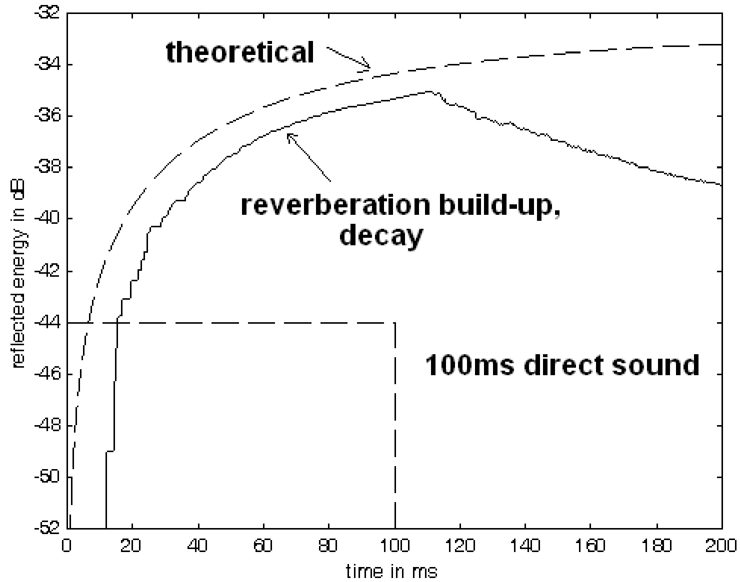

Figure 4: Build up and decay in a hall half the size of BSH. The RT decreases to about 1.1s, from a value of 2.2s in the full size BSH model. Note as expected the onset time delay has decreased a factor of two. But notice the D/R has also decreased, from about minus 6dB to minus 8.5dB. In spite of the far lower RT, these changes cause a significant reduction in clarity and an increase in perceived distance to the source. An audience member can compensate by buying a more expensive seat – but the number of good seats decreases considerably.

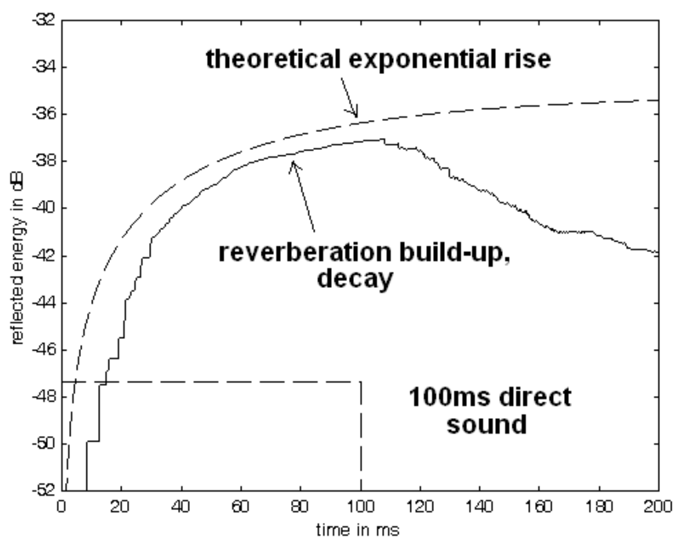

Figure 5: A similar diagram for an existing small shoebox hall, of about 350 seats. The measured RT is 1.0s. The D/R has is -10dB, and the onset delay – although larger than in figure 4, is still quite small. This hall is reasonably articulate in the middle seats on the floor. Closer to the stage the sound is too loud and harsh. Further back the sound is blurred, with poor separation between notes and instruments.

Figures 4, and 5 show the folly of assuming that because a shoebox shape works well (sometimes) for a large hall, the shape is ideal for a smaller hall. Unlike our expectation, the D/R has decreased, even without the increase in RT clients prescribe for small halls. The D/R decreases because sound builds up quicker in a small hall, so a note of a certain length creates a stronger reverberation.

Alas, the folly is entrenched. A great many small shoebox halls are being built. The good news is that some of them sound better than one would expect from these models. The best drastically increase the cubic volume per seat – becoming essentially a large hall with a small number of seats. When the extra volume is above the listeners, the sound can be pretty good. While I was in Helsinki I visited a ~400 seat small hall. The hall was large and high for the number of seats it contained. The RT was ~ 2.0 seconds, as specified by the client – but in spite of this many of the seats sounded at least OK. The best sound was in row 4. In row 8 it was OK. Row 11 it was muddy and blurred – and in row 13 (there were 14 rows) it was good again. Why it gets better further back I do not know. The large stage area was unused in the violin-piano concert I heard, adding volume and RT to the hall. The two musicians played as close as possible to the audience.

One of the best small concert halls is in Boston: Jordan Hall of the New England Conservatory, 1200 seats, built in about 1910. This hall is half of an octagon in shape, with a very high ceiling. The audience surrounds the stage, on the floor and in a single balcony. The RT is about 1.3s occupied. Because the audience is closer to the stage than they would be in a shoebox, and because there is a lot of extra volume on top, nearly every seat combines excellent clarity and localization with a beautiful sense of surrounding reverberation. While ideal for chamber music, small orchestras also sound wonderful. The hall fulfills Krokstad's ideal – audience involvement is maximum.

5. Thresholds

Thresholds for the onset of localization were measured using a binaural hall model and a source signal that emulated a cello, playing notes with a rise-time of 20ms, a fall time of 100ms, a duration of 200ms, and a gap between notes of 200ms.

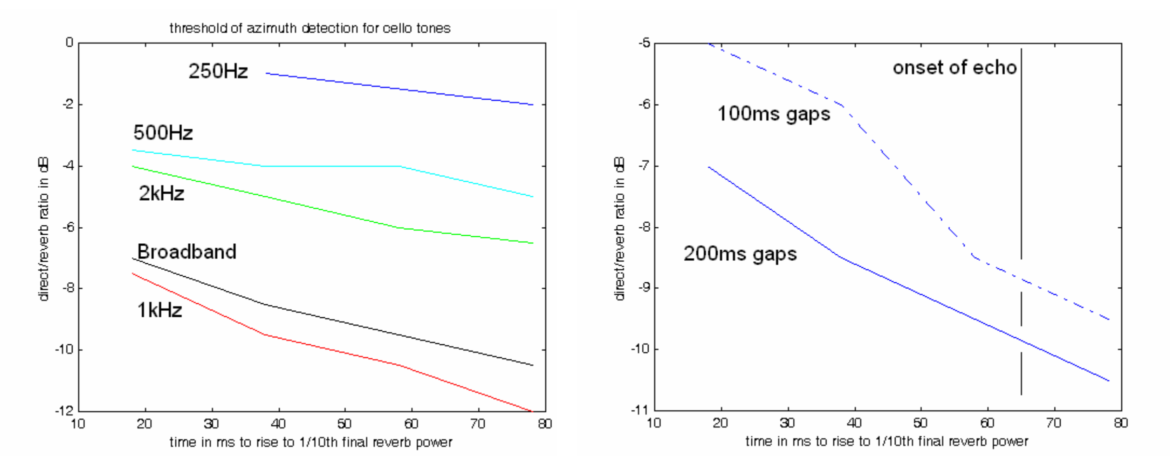

Figure 6: left – thresholds for the detection of localization as a function of the time delay between the direct sound and the build up of reverberation to an energy equal to the direct sound, and as a function of the D/R. Note the threshold decreases as the delay increases. Right – when the gap between notes decreases, the threshold increases.

Figure 6 shows some thresholds for localization of a cello tone. The threshold is the lowest for the 1000Hz band, and is very high for the 250Hz band. This shows the importance of maintaining a high D/R above 1000Hz in a good hall. How can this be done, particularly if a client has specified a constant reverb time of 2 seconds? The trick is to reduce the reverberant level at these frequencies by clever use of frequency dependent reflections and diffusion. If frequencies above 1000Hz can be preferentially reflected back to the front of the hall where the direct sound is strong and absorption is high, the D/R for the further seats will be raised. This is the magic provided by coffered ceiling and wall surfaces in Boston, Vienna, and Amsterdam.

6. Binaural measures

Ideally a measure for distance and clarity should use the same mechanisms that the human brain uses to determine these perceptions. This is too lofty a goal for me. I am content to design a non-linear measurement system that is based on human physiology, but works somewhat more mathematically. I have worked on two such measures, one for localization, and one for distance. Both measures take as inputs binaural recordings of a small number of instruments – ideally a solo instrument, actor, or singer. The recording is then analyzed with a hearing model and azimuth and distance information is extracted. The percentage of time where such an extraction is possible becomes a measure of the space.

It is almost always possible to localize some sounds in an acoustic space. The brain is good at separating the direct component of a sharp click from the reverberation. Remember it takes time for reverberation to develop. The click is short enough that very little reverberant energy is generated, and the reverberant level is low.

But musical notes are usually not really short, and they tend to mask one another. The perception of localization relies of very brief inputs in real information. Typically only the onsets of sounds can be localized, and that only if they have not been masked by previous sounds. So a measure based on the percentage of localized onsets makes sense.
My system for measuring localization has been previously described. We find the IACC in overlapping 10ms segments of a musical signal, and plot 1/(1-IACC) in 1/3 octave bands. The result is a map of the D/R of the onsets of sounds as a function of time and frequency. The percentage of time these are over some threshold can be counted. We can also plot the Azimuth determined by such a measure with azimuth and frequency as axis. This shows how consistent the azimuth is over frequency – also a measure of hall quality.

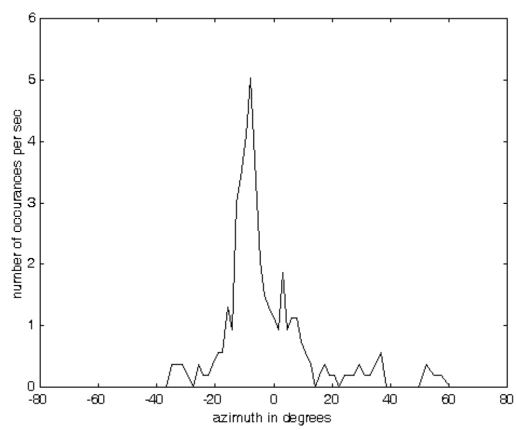

Figure 7: the number of times per second a note from a solo violin is localized in row 4 of the small hall near Helsinki. The indicated azimuth is correct – about 10 degrees to the left of center. The binaural recording was made with probe microphones at the ear drums of the author, and plays with startling realism with headphones calibrated the same way. A ~ 5 second segment was analyzed to make the graph.

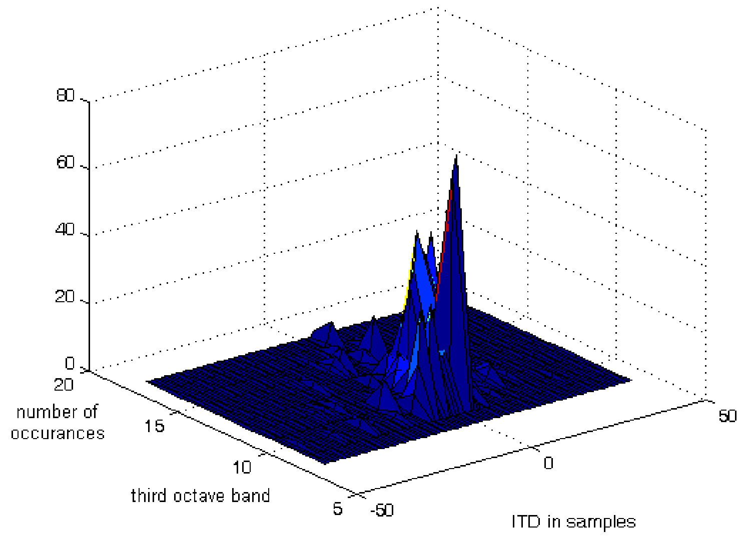

Figure 8: A surface showing the frequencies of maximum localization for the violin in figure 6. The major peak is for the 1000Hz 1/3 octave band.

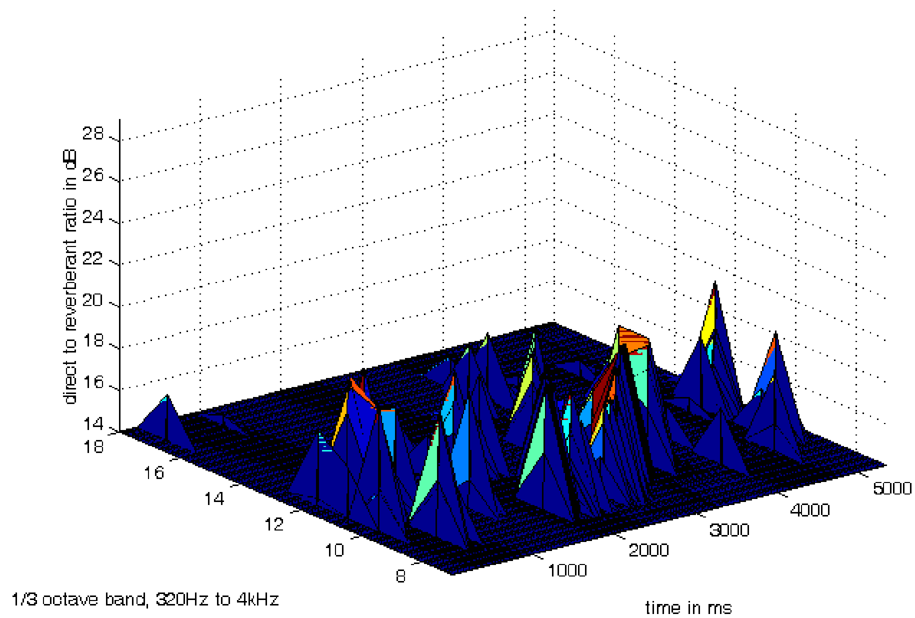

Figure 9: A time/frequency surface showing the peak in running IACC that occurs at the onset of each note of the violin in Figure 6. For a very short time the localization is quite clear.

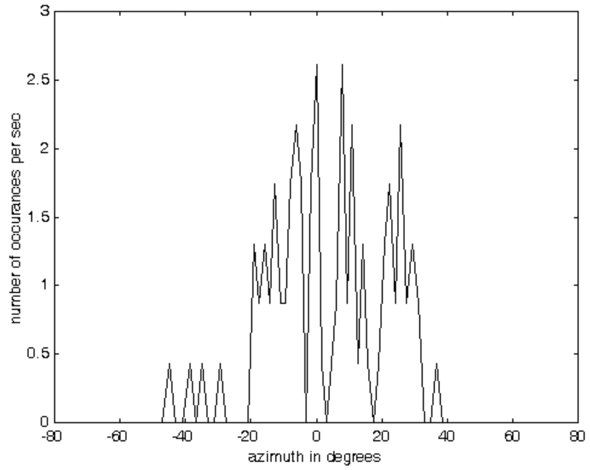

Figure 10: Number of localizations per second for the same violin in row 11 of the small hall in Helsinki. Note the small number of localizations, and the poor consistency of azimuth. Localization is very difficult, but sometimes possible, in this seat. The sound is muddy.

We can build a similar measure for distance, in this case using the concept of harmonic coherence. Harmonic coherence is a monaural measure, which tests the degree to which harmonics of a pitched sound source are phase-locked to each other. This phase locking is highly audible, resulting in a sound described by Zwicker as "roughness". The phase locking is destroyed by reverberation, and can be used as a measure. When the phases are locked the fundamental frequency is easily determined by rectifying critical bands. For these examples I plot the strength and frequency of the extracted fundamental as a function of time. All critical bands with significant fundamental are combined. I will use sounds from singers binaurally recorded in Oslo.

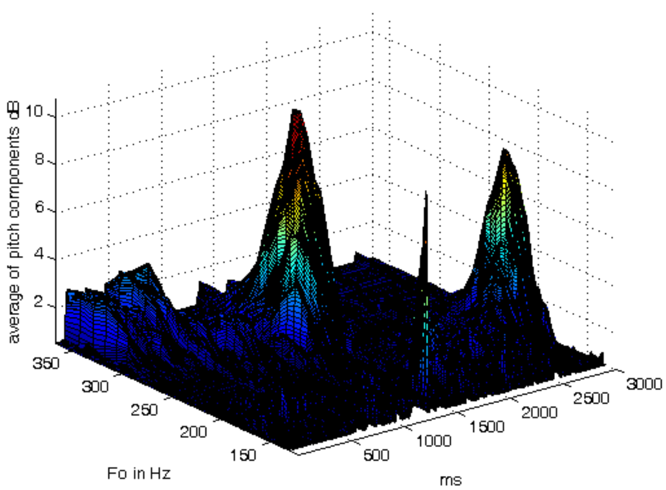

Figure 11: A strong baritone forcefully singing directly to the third balcony. The sound there is often muddy, but the fundamental pitch of this singer came through strongly at the beginning of two notes. He seemed to be speaking directly to me, and I liked it.

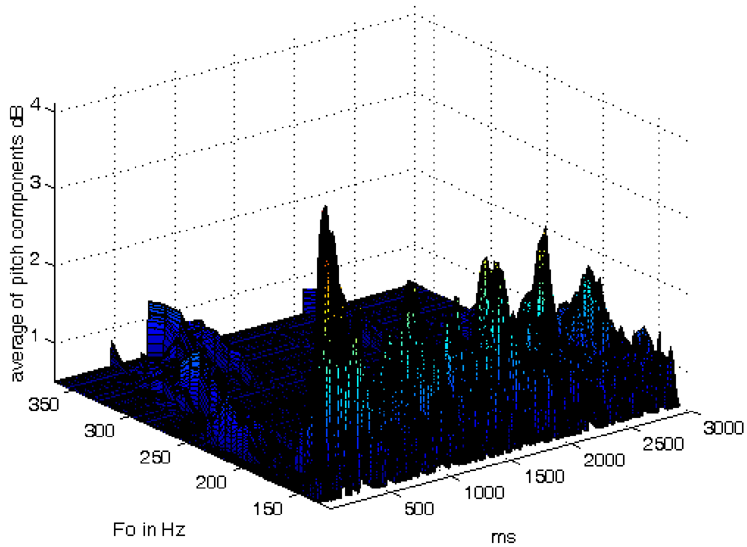

Figure 12: The king (in Verdi's Don Carlos) on the other hand, in his wonderful solo aria, was not able to reach the third balcony with the same strength. The fundamental pitches are not well defined. He seemed muddy and far away. Pathetic, but the pathos was muted.

7. Conclusions

Direct sound is vital to maximizing an audience's involvement with music and drama. If it is to be effective in this task the direct sound must be separately audible – at least some of the time. Because most seats in halls are well beyond the critical distance, the direct sound can only be heard during the onsets of notes, and then only if the onset is not masked by other notes.

But without the occasional detection of direct sound the ability to precisely localize actors or musicians disappears, and dramatic intensity is reduced.

The ability to detect direct sound depends on the time delay between the direct sound and the onset of reverberation, the effective direct to reverberant ratio, the frequency dependence of the direct to reverberant ratio, and the and on the type of music.

The detection can be maximized by carefully controlling the D/R at 1000Hz and above as the sound moves back through a large hall. In a small hall – which includes any hall less than ~ 2000 seats – a shoebox shape is probably not optimum. Effort should be made to increase the volume above the listeners, while reducing the average seating distance. For smaller halls – under 1000 seats – it can be advantageous to absorb or scatter the earliest reflections away from the audience. In these halls the direct sound is already strong. Early reflections only act to reduce localization. These reflections reduce the time delay needed to detect direct sound, and provide no other benefit.

The degree to which a person can localize a particular instrument, or detect its sonic distance can be measured through the application of a hearing model to a binaurally recorded signal of a solo instrument or voice. Matlab code for these models and measurements can be obtained from the author. They will all sooner or later end up on his web-page, along with how-to instructions for making and reproducing the binaural recordings.