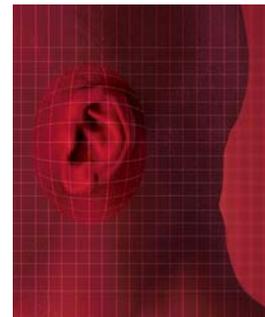


Human Processing of Reflected Sound

by Steve Barbar



*A response to "Reverberation Time - What Could Be Simpler?"
Syn-Aud-Con Newsletter Volume 37 May 09*

I must preface the following with the acknowledgement that most on the SAC Email Group have interests that focus on applications of sound reinforcement. This text will widen the discussion to include acoustical sources – however, it is germane to a common dialog about how we as humans interpret sound, what we garner from current “standard” measures, and how we might derive something that is a better match to what we perceive.

In your reverberation article, you said:

A strong direct field will produce an EDT that is less than (steeper slope) than the T30. This is a desirable attribute for increased speech intelligibility or music clarity.

Yes and no. "Increased speech intelligibility or music clarity" from reflected energy is dependent on the absolute level of the direct sound, the time gap between the direct and reflected energy, and the ratio of reflected energy that is merged with the direct sound regardless of frequency and the location of the reflection(s) compared to the energy that is not merged into the foreground sound stream.¹ So it should come as a bit of a shock that current acoustic measures ignore both the ratio of direct to reflected and or reverberant energy, as well as the initial time gap! For example, RT60, T30, EDT, and C80 all ignore the absolute level of the direct sound, the absolute level of the reverberant sound, as well as the effects of varied language or the effects of musical style

on the audibility of the D/R. Yet it is the absolute level of the direct sound, particularly in the range of vocal formants that is crucial to phonemic recognition. This is also what our neurology interrogates to impart the sensation of acoustic distance, as well as the sensation of intimacy and involvement. Another important element is the duration of the direct sound (as noted by Dale Shirk in several posts on the SAC Email Group). The duration of sound events for speech and music can be substantially different. In spoken English, phonemes occur at a rate of roughly 150ms when the person is talking quickly. Notes can have much greater duration. For sounds that have steady state amplitude, the level of the reverberation is dependent on the duration of the sound generated by the source compared to the reverberation time. Short sounds will not fully excite a large internal volume, and the resulting D/R may be much higher than what occurs with sources that have sustained output.

The concept of providing ample direct sound is a bit like preaching to the choir for this group. When the direct sound is comprised mainly of the output from an electro-acoustic system, the dispersion of the EA system becomes the dominant factor in the D/R. The ability to increase the direct sound without increasing the reflected or reverberant sound in this manner can constitute “*a desirable attribute for increased speech intelligibility or music clarity*” - as stated in the reverberation article referenced above.

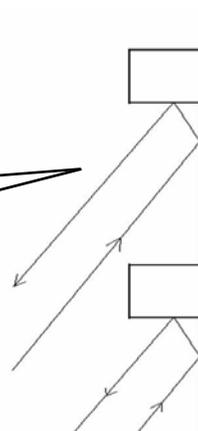


Figure 1 - Boston Symphony Hall

Steve Barbar, president of E-coustic Systems of Belmont, MA designs and develops acoustic enhancement systems.

But the bigger picture is not just quantifying SPL, or intelligibility, or RT60, RT30, EDT etc. It is determining what these measures mean with respect to what we experience. In an ideal world, I would prefer to have one meter, with BAD printed on the left face and GOOD printed on the right with a zero center scale in dB. Pushbuttons would enable me to interrogate MY listening experience. For instance, I could press a button labeled ‘Loudness’ and then press a button for a source like ‘Orchestra’, hold up the meter and see how this rated against my preferences. Since I was the one holding the meter, I could then argue with myself about the results if I thought they were off. But this is entirely the point – our hearing is hardwired, and although our neurology is not, the manner in which we interpret sound as a species is remarkably similar.

We have just reached the 30th anniversary of the development of the first digital reverberation system. Prior to this development executing a change in reverberation required some sort of physical manipulation. Unfortunately, what was available to manipulate physically often had no relationship to the physics of enclosed volumes (unless one manipulated a real volume like a reverb chamber). The initial deployment of this technology might be best described as giving a Model T to a caveman - you could sense that sooner or later he was going to get it, but it was going to be a bumpy ride. It explains a lot about music released in the early 80's. None the less, the enormous volume of recordings released since that time serves as a data record for preferences of D/R. We have performed experiments that use a "dry" sound source and require a listener to adjust the D/R until they think it is "optimum".² The results are almost unanimous, and fall between +4dB to +6dB D/R. Engineers and acousticians (including Beranek) choose the same values. These values of D/R are optimal based on the properties of music, and human neurology - otherwise, the data from recordings would be different, and all of us including producers, engineers, conductors, and musicians would desire and mandate something else. These ratios, however, are NOT the values found in most spaces. For example, the critical distance of Boston Symphony Hall is roughly 21 feet. This is the point where the D/R equals 0. Thus in contrast to the optimum ratios of D/R noted above as +4dB to +6dB, almost every seat in this venue has a negative D/R! Furthermore, this imparts that most of the perceived loudness is a function of the reflected energy - otherwise the music would not be loud enough throughout the venue. So why is Boston Symphony Hall rated so highly? It is the ability to hear the direct sound before the onset of the reflected energy that is vital to the perceived sonic quality in such spaces. Our work indicates that there may be an optimum value for ITG, and the ~25ms ITG found in Boston Symphony Hall may be close. Recent work in large volumes using acoustic enhancement suggests that the optimum time gap can be slightly less when the levels of direct, reflected and reverberant sound can be manipulated, and that this is very important for both intelligibility of choral music, as well as the sense of perceived intimacy with the sound source.

There is another phenomenon that takes place in the “Big 3” concert halls (Amsterdam, Vienna, and Boston). All of them have rectangular architectural elements that serve to redirect sound toward the stage at frequencies above 1 kHz (Figure 1).

Since the level of the direct sound is higher near the stage, it is difficult to perceive this energy in the front of the volume. Toward the rear of the volume, however, this energy serves to increase the D/R at frequencies that our neurology interrogates to determine acoustic intimacy, providing a greater sense of involvement. You can think of this as old school onset enhancement (or retro frequency dependent delay fills).. Since this energy is directed toward the audience and stage, it is largely absorbed – hence it does not contribute to late reverberation.

All of the Big 3 halls have reverberation decay that is very similar – yet each of the halls sounds different. My colleague David Griesinger decided to investigate the manner in which the reflected end reverberant energy builds up instead of the way that it decays. He constructed simple binaural image source models of Boston and Amsterdam using HRTF's measured from his eardrums.

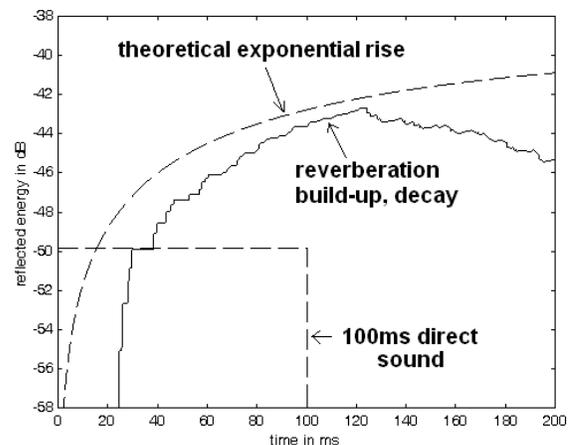


Figure 2 - Boston Model T10R

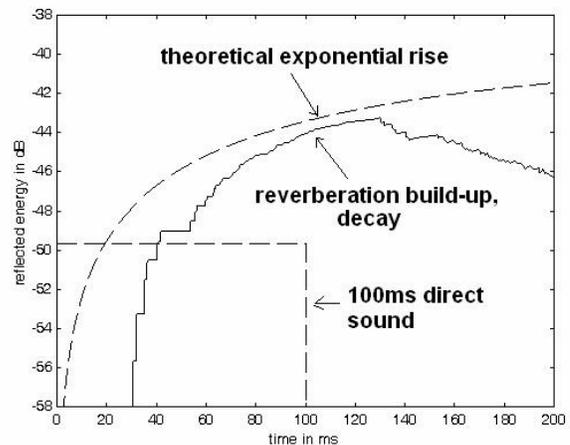


Figure 3 - Amsterdam Model T10R

From David:

The upward dashed curve shows the theoretical exponential rise of reverberant energy from a continuous source. The seat position in the model has been chosen so that the D/R is -10dB for a continuous note.

The upward solid line shows the actual build-up, and the downward solid line shows the decay from a shorter note – here a 100ms excitation. Note the actual D/R for the short note is only about 6dB.

T10R – the time for the reverberation to rise to 1/10 the final energy – is less in Boston (Fig. 2) than in Amsterdam (Fig. 3), but after about 50ms the curves are nearly identical. (Without the direct sound they sound identical.)

Over the past few years, I have made many posts to the SAC Email Group stating that the level of reflected and reverberant energy is what counts citing RR 150 and RR 350 as a better means of determining the sonic characteristics of a venue than trying to interpret EDT vs RT. This is because RR is measure of the ratio of reflected energy in one portion of the ETC to another. For example RR150 compares the ratio of energy between 0 and 50ms to the energy between 50 and 150ms – in other words, the ratio of energy that is likely to be merged into the foreground stream with the energy that might cause fluctuations in absolute magnitude in the range of vocal formants at the onset of the next phoneme. Likewise, RR350 is a comparison of the energy in the first 50 ms to the energy that is likely to be merged into the background stream. Another advantage to RR is that the time windows can be adjusted to compare any portion of the ETC. With careful manipulation, RR can be used to determine the D/R – but this is dependent on having sufficient ITG. Even so, RR alone cannot provide a good indication of how human neurology might interpret acoustic distance.

Presence is defined as the immediate proximity of someone or something. Unfortunately, the term “Auditory Presence” is very vague, and can be interpreted differently by musicians, acousticians and lay people. We are trying to find a better term for the acoustic sensation of closeness to the sound source(s). In order to stop using an otherwise confusing term, I will refer to this sensation as acoustic involvement until a better term can be found. This is very different from a visually constructed sense of presence where the eyes train the brain where to expect each sound – and this is where we hear it!

I have also mentioned that pitch played an important role in the determining acoustic distance; and that some of the important cues for acoustic distance had nothing to do with binaural imaging. The example that I mentioned used a single omnidirectional microphone with monaural playback over headphones. With the microphone placed in a typical enclosed acoustic volume, the person listening on headphones will be able to reliably identify when the source is close to the microphone, and when the source moves away from the microphone – say from 1.5’ to 10’, and moreover,

when the source appears acoustically close and when it does not.

Further investigation has revealed that the absolute magnitude during the onset of sounds, particularly in the range of vocal formants is an important element in determining acoustic distance, and that fluctuations in pitch and or harmonic coherence are interpreted by our neurology as distance from the sound source. In order to perceive acoustic involvement, we must be able to both aurally localize sound sources, as well as distinguish between multiple sources, nearly all of the time. For source localization to be successful we must be able to hear the direct sound.. The ability to hear all of the notes as well as localize each instrument draws us into the performance. When this happens, the listening experience is engaging. Since harmonic coherence, source localization, and musical clarity correlate well, it is possible to use source localization as a measure of acoustic involvement. This effect can be easily measured as $1/(1-IACC)$ over time to determine the absolute value at the onsets of sound events, and is sensitive to both lateral and medial reflections.

Again from David:

Figure 4 shows the audible fundamental components in the formant frequencies as a function of time. The vertical axis shows the effective D/R ratio at the beginning of two notes from an opera singer in Oslo to the front of the third balcony (fully occupied.) The sound there is often muddy, but the fundamental pitch of this singer came through strongly at the beginning of two notes. There is a strong sense of acoustic intimacy - he seemed to be speaking directly to me, and I liked it.

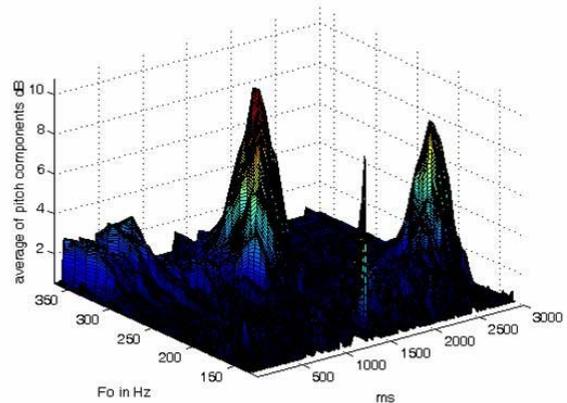


Figure 4

In the following graph we plot $1/(1-IACC)$ as a function of time and third octave band. Note that the IACC peaks at the onset of notes can have quite high values for a brief time. This happens when there is sufficient delay between the direct and the reverberation, and sufficient D/R.

Figure 5 shows a different singer measured at the same location. In this solo aria, the singer is not able to reach the balcony with the same strength. The fundamental pitches are

not well defined, and the singer seems muddy feels (acoustically) far away. This performance was not engaging.

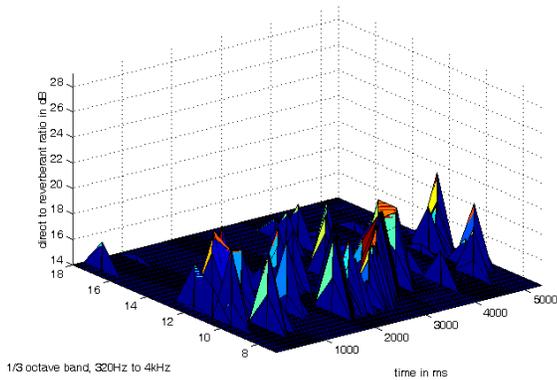


Figure 5

Figure 6 shows the number of times per second that a solo violin can be localized from row 4 of a small shoebox hall. It also shows the perceived azimuth of the violin. As can be seen, the localization – achieved at the onsets of notes – is quite good, and the azimuth, ~10 degrees to the left of center, is accurate.

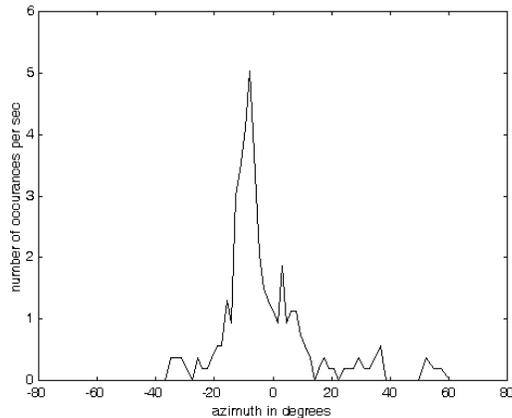


Figure 6

Figure 7 shows a plot of the same data for the violin as a function of (inverse) azimuth, and the third octave frequency band. As can be seen, for this instrument the principle localization components come at about 1300Hz. Interestingly, Human ability to detect azimuth, as shown in the threshold data, is maximum at this frequency.

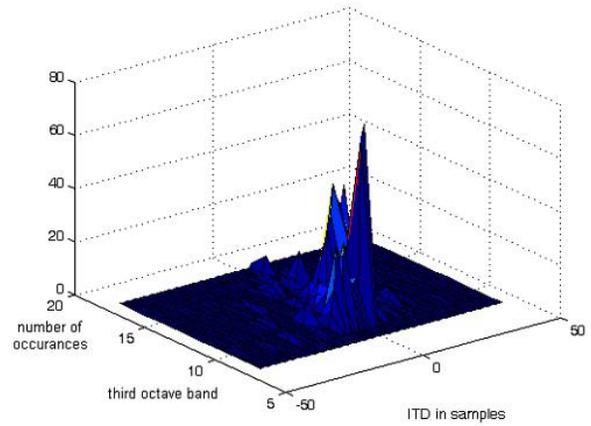


Figure 7

In Figure 8, $1/(1-IACC)$ is plotted as a function of time and third octave band. Note that the IACC peaks at the onset of notes can have quite high values for a brief time. This happens when there is sufficient delay between the direct and the reverberation, and sufficient D/R.

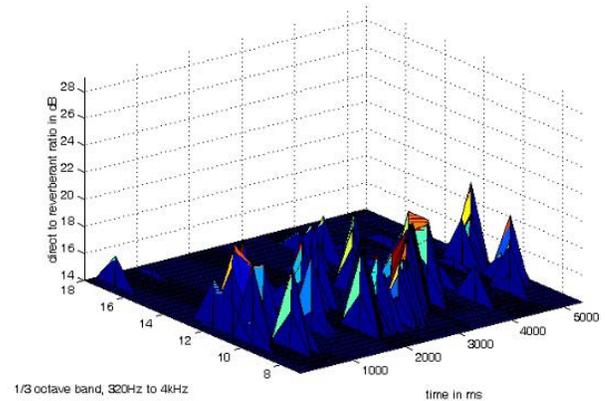


Figure 8

The graph in Figure 9 is for a solo violin in row 11 of the same hall. The sound here is unclear, and the localization of the violin is poor. As can be seen, the number of localizations per second is low. Perhaps more tellingly, the azimuth detected seems random. This is really just noise, and is perceived as such.



Installed Cable Checker

The [Rat Sniffer](#) has a send and receive unit that allows installed cables to be tested for typical fault conditions, displaying the wiring status on 3 LEDs.

"All green" means good. Anything else indicates a fault and will require further investigation. The Rat Sniffer is available from [Audio Control Industrial](#).

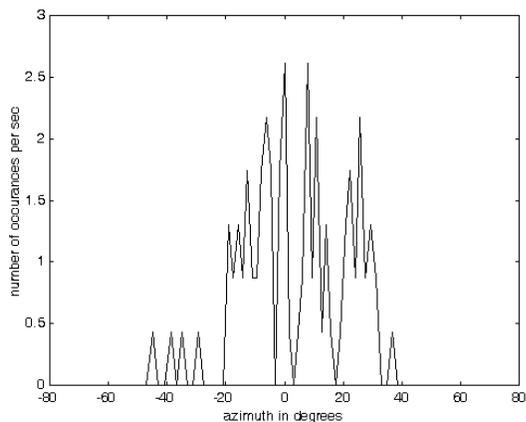


Figure 9

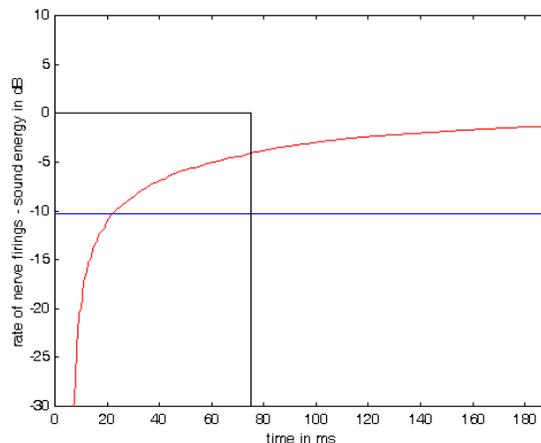


Figure 11

Another important consideration is that we humans perceive an impulse and resultant reflected energy that we capture as an IR as concurrent sound event. In other words we hear the note, click, sweep, etc. – we do not hear the IR. Figure 10 shows a graph of the ipsilateral impulse response from spatially diffuse exponentially decaying white noise with an onset time of 5ms and an RT of 1 second. The direct component is less than 5ms long.

The nerve firings for the direct component of this note last for the duration of the sound event. The nerve firings from the reverberant component build steadily after the onset of the direct sound, and then start to decay. The graph in Figure 11 shows the rate of nerve firings for the direct sound in blue, and the reverberant sound in red.

The black rectangle indicates the 75 ms. time window over which these two rates will be integrated. The RT-60 for this example is two seconds.

In Figure 12, the RT60 is reduced to one second. Note the amount of reverberant signal in the time window has increased substantially – thus raising threshold of localization.

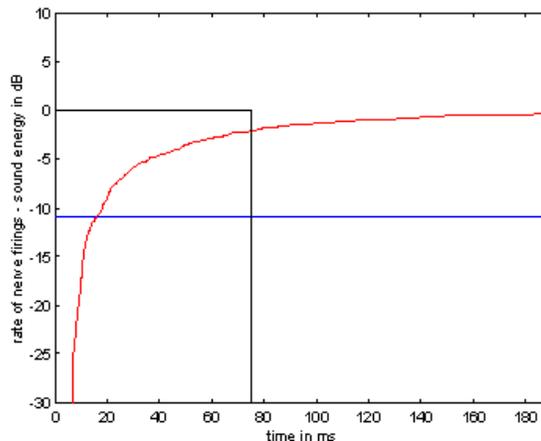


Figure 12

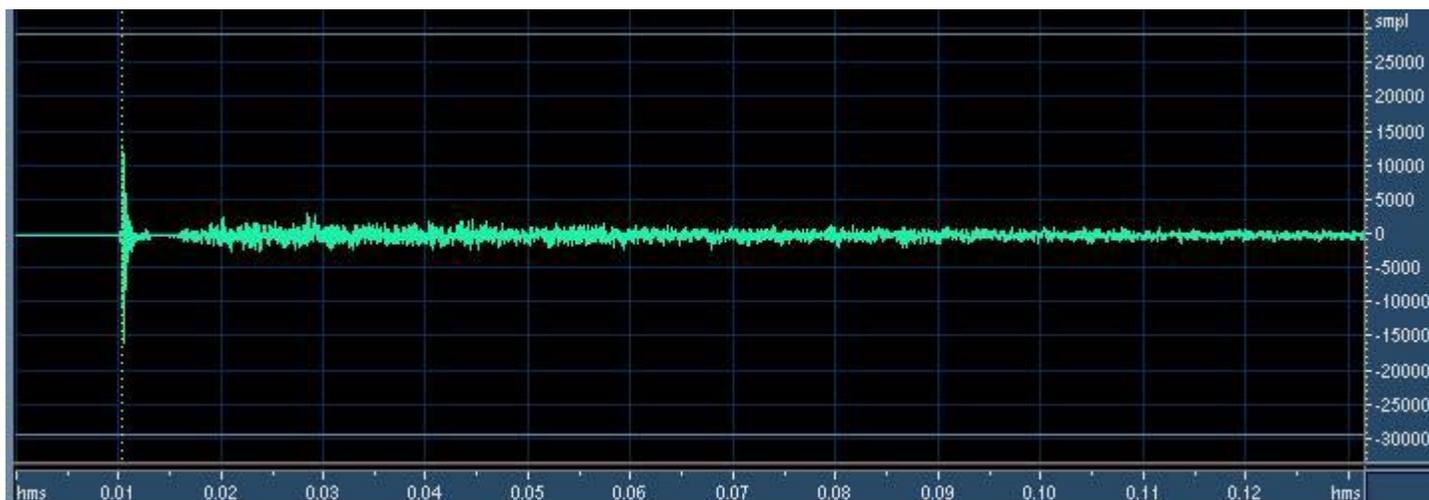


Figure 10

As the scale of an enclosed volume changes, all of the properties relating to the reflected sound change as well, including the relationships of ITG and D/R. The properties of the sound source, as well as our sense of hearing and neurological response, however, are fixed – and therein lies the rub.

For smaller spaces, the ramifications are significant. If you simply alter the scale of the model of Boston Symphony Hall by 1/2, you create a 600 or so seat venue.

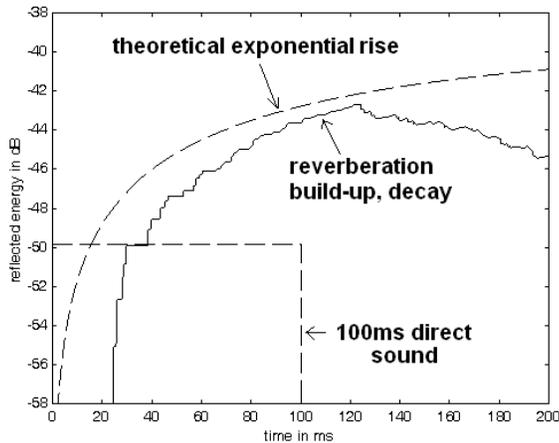


Figure 13 - Boston Full Scale

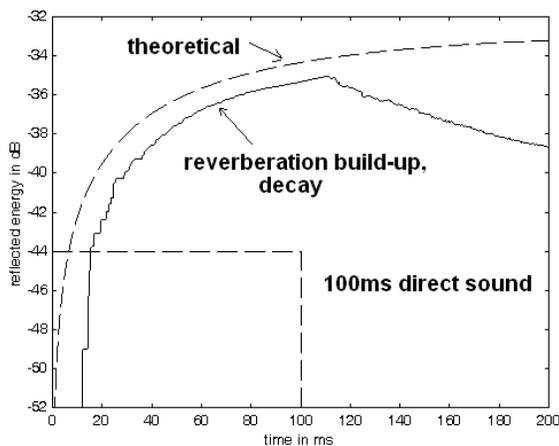
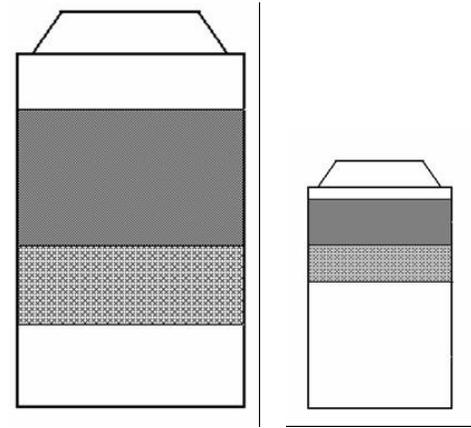
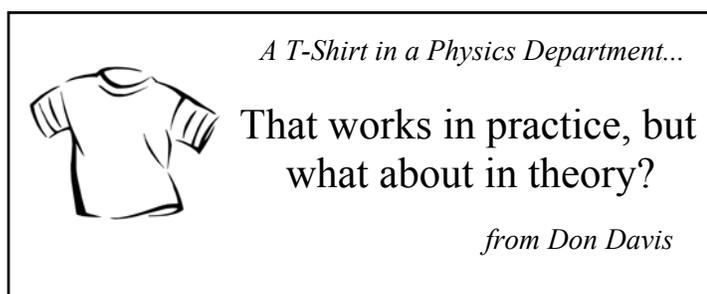


Figure 14 - Boston Half Scale

The reverberation time is reduced by half, as is the ITG. Most significantly, the D/R is reduced from -6db to -8.5db. Even though the reverberation time is shorter, the level of the reverberation is stronger, and hence, the critical distance is shorter. In the full size model of Boston Symphony Hall, the direct sound is distinct in nearly 50% of the seats. In the smaller scaled model, this is reduced to 30%.



Dark Gray = Above Threshold
Gray = Near Threshold
White = Below Threshold

Figure 15. Localization and onset enhancement thresholds for seats

Figures 13-15 provide a glimpse of the problem. In a smaller space the listening experience can change significantly over a relatively short distance – like two rows of seats. If the perception of acoustic involvement is an important factor in determining the quality of sound, can we quantify it?

Using the concept of how the brain interprets nerve firings for direct and reverberant sound, we can derive an equation that expresses the ease of perceiving the direction of direct sound as a decibel value. With the previous simple assumptions, we propose the threshold for detection would be 0, and clear localization would occur at a localization value of +3dB.

Where D is the window width $\approx 0.1s$, and S is a scale factor

$$S = 20 \text{ MAX} \left(\int_{.005}^{\infty} (10 * \log_{10} (p(t)^2)) dt \right)$$

S sets a possible signal/noise ratio for nerve firings.
Localizability in dB =

$$S + 1.5 + \int_0^{.005} (10 * \log_{10} (p(t)^2)) dt - (1/D) * \int_{.005}^D \text{POS}(S + \int_{.005}^{\tau} 10 * \log_{10} ((p(t)^2) dt) d\tau$$

POS limits the possible large negative values for the log. The scale factor S and the window width D interact to set the slope of the threshold as a function of added time delay.

Testing the equations:

The Localization Equation was developed and tested using binaural impulse response generated using known HRTF²s.

- The source position was 15 degrees to the left (and right) of center. Only the ipsilateral channel was analyzed.
- Male speech alternated from left to right with a time gap of 400ms, to allow for complete decay of the reverberation between each word.

- The reverberation was generated using an independent decaying noise signal convolved with each of 54 HRTFs spaced equally around the listening position.

- The HRTFs were equalized so that the azimuth zero elevation zero HRTF was flat from 40Hz to about 4 kHz. The elevation notch at 7.8 kHz was not equalized away, but was left in place.

- Playback was done through headphones equalized to match a loudspeaker placed in front of the listener – again not equalizing the 7.8 kHz notch from the listener’s frontal HRTF of the loudspeaker.

The impulse response was band pass filtered between 1000Hz and 4000Hz before being analyzed for localization.

- If a measured binaural impulse response is used as an input, care should be taken to insure the dummy head is equalized as described above.

- Because of the importance of upward masking in localization, if the low frequencies in the room signal are significantly stronger than the frequency range from 1000 to 4000Hz, localization is likely to be poorer than the equation would predict.

Tests with Speech

A speech signal was convolved with a pair of binaural impulse responses, such that the sound appears to come from +/-15 degrees from the front. The talker counts from one to ten, with each successive number assigned the opposite channel (see Figure 16). Then a fully spatially diffuse reverberation was added, in such a way as the D/R, the RT, and the time delay before the reverberation onset could be varied.

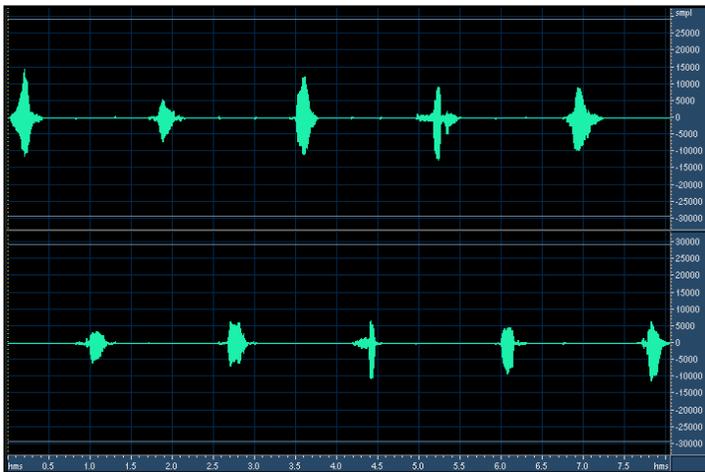


Figure 16 - Speech signal convolved with binaural IRs.

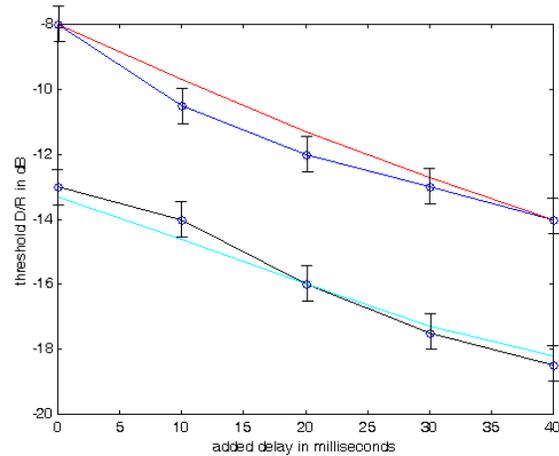


Figure 17 - Localization and onset of enhancement thresholds for seats

Blue – experimental thresholds for the alternating speech with a 1 second reverb time.

Red – the threshold predicted by the localization equation.

Black – experimental thresholds for RT = 2seconds.

Cyan – thresholds predicted by the localization equation.

As you can see from the graph above, the algorithm does a very good job of predicting source localization, and thus, the sensation of acoustic involvement.

When the only requirement is such spaces is for the spoken word, absorption and/or proper implementation of an electro-acoustic reinforcement system can be used to alter the D/R to great effect.

For acoustic music, and/or a requirement for both speech and music – the situation is a bit more challenging. Where musical clarity is a problem in small halls, acousticians usually recommend adding early reflections – through a stage shell, side reflectors, etc. Such treatments, however, can reduce the ITG, which serves to increase the perceived acoustic distance to the performers, as well as increase the sense of muddiness. If the client insists on increasing the RT by reducing absorption, the D/R will be decreased further. Carefully adding absorption will increase the D/R and can help to unmask the direct sound. Adding vertical volume can increase the late RT. When it is impossible to add cubic volume, adding absorption to reduce the level of reflected energy from the stage and first lateral reflections to the audience can improve the ITG. Reverberation must be added in a frequency dependent manner that increases RT and only minimally increases RR. This can only be accomplished electronically (this should be no real surprise).

We have been recording performances binaurally for many years. Current technology uses probe microphones at the eardrums. Headphones are then measured at the eardrum using the same microphones. With headphones that lack pinna interference notches the original eardrum pressure can be reproduced exactly. We can use these recordings to make a sonic snapshot of the actual sound in a space. The results can be surprisingly different from our memory of a perfor-

mance. This is due in no small part to the adaptation process that occurs with human neurology. This process occurs at several levels, and involves all of the senses, not just hearing. We acclimate to an environment by sub-consciously disregarding information that we deem to be unimportant, while simultaneously interrogating and focusing on the thing(s) that demand our attention. Many examples of “inattentive blindness” can be found on the University of Illinois Visual Cognition Lab website:

http://viscog.beckman.illinois.edu/djs_lab

We have presented an HD video recording of a string quartet where we alter the audio quality over time from very good to pretty bad at both the AES and ASA. When the visual image is present, no-one catches the discrepancy. With audio playback alone, the changes become evident to almost everyone.

The process of aural acclimation involves the “normalizing” of the data from the critical bands. This can occur from either a visual interpretation of the sound field (I see the violin, so I expect the violin sound to come from that location) or a sonically derived localization map of the soundfield. The formation of the latter takes a bit longer, and is reliant on sufficient occurrence of sound from each independent source. Once established, however, it takes a considerable change in soundfield to reset this condition. Hence the reverberation that we hear at the start of a performance is not what we experience after listening for several minutes. Walk-

ing out of the venue and re-entering a minute or so later will restart the acclimation process.

Results derived from binaural measurements have thus far proven to be of much greater value as it relates to human neurology than the data from B-format made at the same location concurrent with the binaural recording. The B-format recording has greater value for isolating information from a specific direction. Archiving both simultaneously may be the best means of collecting acoustical data. Neither microphone format is new, and a substantial amount of documentation describing the attributes of both already exists. What has changed in recent years is technological advances in hardware, the amount of computational power available for processing data captured by these arrays, and the software available for decoding this information. Hence, topics for future discussion include array basics, description of data that can be garnered, how this differs between arrays and the implications, current implementation(s) and ramifications of the microphone arrays, and sound source topologies.

¹ The Process of Hearing – Barbar - Syn-Aud-Con Newsletter

² W GARDNER ‘Reverberation Level Matching Experiments’ Proceedings of the Sabine Memorial Conference, MIT June 5-7 1994 p263 Available from the Acoustical Society of America

What's In the Box?

No standards exist on loudspeaker wiring or connectors. It can be a minor research project to figure out what's connected to what in a loudspeaker system. A system tech can figure it out with an impedance meter, but even that can require pulling apart connectors, landing leads on miniature contacts, and reading micro-sized numbers from connector housings.

The Cab Driver from Whirlwind is a clever tool for some quick-checks to loudspeaker systems. Four cable pairs can be individually driven with pink noise to determine which pair drives which driver. DC can be momentarily applied to determine cone polarity and to measure DC resistance. The unit is battery powered to facilitate use on top of scaffolds or hanging from a harness. A line level output can be used to drive powered loudspeakers.

Warning - my recently purchased unit was reverse-polarized on the the MDP connector, and it's no easy task to disassemble and correct. Always test the testers!

While I can do all of these things with other tools, I've found the Cab Driver handy enough to allocate it some space in the tool kit. *pb*

