OBJECTIVE MEASURES OF SPACIOUSNESS AND ENVELOPMENT

DAVID GRIESINGER

Lexicon, 3 Oak Park, Bedford, MA 01730
dg@lexicon.com

This paper will discuss current knowledge of spaciousness and envelopment (assumed to be the same perception) with particular attention to measurement techniques that might be applicable to virtual audio systems. We find that all frequencies contribute to the perception of envelopment. At frequencies below 2000Hz envelopment arises from fluctuations in the Interaural Time Delay (ITD) and the Interaural Intensity Difference (IID) at the ears of the listener. Fluctuations that follow the ends of sounds by at least 160ms are the most effective. These fluctuations can be produced by decorrelated sound sources on opposite sides of the listener. The optimum angle for such sources varies with frequency. For frequencies below 700Hz sources should be at 90 degrees from the front. Above this frequency the optimum angle moves toward the medial plane. For broadband signals above 2000Hz, spectral cues characteristic of sound sources behind the listener increase the perceived envelopment. The optimum angle for broadband sound sources with most of their energy above 2000Hz is approximately 150 degrees from the front. We conclude that multiple sound sources (either real or virtual) are necessary for optimal envelopment, consisting of at least a pair of sources at 90 degrees from the front, and a pair at 150 degrees. The effectiveness of these sources will depend on the shape and reflectivity of the room. For measuring the envelopment delivered by the loudspeaker/room system, the Diffuse field Transfer Function (DTF) is recommended.

INTRODUCTION

Envelopment and spaciousness are practically the "raison d'etre" of 3D audio systems. We have long known that in the absence of room reflections standard stereo playback produces a flat, frontal sound. The goal of the first virtual audio systems was to widen the sound stage by creating phantom sound sources at +-90 degrees from the front. Two channel music played through these systems can be highly enveloping. As virtual systems improved, the goal expanded to reproducing multichannel music with phantom sound sources further behind the listener – at +- 120 degrees, or even +- 140 degrees. Depending on the type of music and how it was recorded such sources were sometimes more successful at producing envelopment than virtual sources at the side. How can the increased envelopment be explained? How can we optimize the recording technique to exploit the new phantom sources? How can we measure the differences in envelopment produced by different virtual systems?

The answer to all these questions can be found in the mechanisms used by human hearing for perceiving spatial sounds. We need to develop measures that mimic human binaural processing. In this paper we will first examine some of the past measures for spaciousness. We will then present a hypothesis about how binaural spatial analysis is achieved. Several new measures will be presented, along with preliminary data on their performance in practice.

1. OVERVIEW OF PAST MEASURES OF SPACIOUSNESS AND ENVELOPMENT

One problem with previous work on spaciousness and envelopment is that the two terms are difficult to define. In most literature the term "spaciousness" has been associated with the Apparent Source Width, or ASW. Unfortunately such an association is not in keeping with the popular meaning of the word. The American Heritage Dictionary defines "spacious" as "1. providing or having much space or room," and "2. vast in space or scope, 'a spacious view.'" Thus in the English language a concert hall can be spacious, the reverberation of an oboe can be spacious, but the sonic image of an oboe cannot be spacious. Furthermore we have found that the broadening of a source image occurs more easily in small rooms than in large rooms. We suggest that the association between "spaciousness" and ASW should be abandoned.

Where we wish to describe source broadening, we will simply use the term ASW.

Barron described a perception he called spatial impression, or SI, with a quote from Marshall, "The sensation of spatial impression corresponds to the difference between feeling 'inside' the music and looking 'at' it, as through a window." This description would seem to apply to the sense of envelopment, the perception of being enveloped by the music. With this definition SI and envelopment seem to be the same thing. Although the perception of being surrounded by the music does not necessarily relate directly to the perception of space, music listeners are probably aware that the surround impression does not naturally occur in small rooms. To experienced listeners the perception of envelopment is one of the joys of a large concert space – and it seems appropriate to associate the word "spaciousness" with this perception. We will treat the terms spaciousness and envelopment as synonymous.

How do we measure envelopment? The most popular current measures were developed for concert hall use. They attempt to quantify the envelopment perception for a concert listener in a particular seat by analyzing an impulse response measured from a single point on stage to the listener. It is not clear how this type of measure can be adapted to the world of virtual audio, but we will try.

Barron proposed measuring SI by the ratio of the lateral sound energy (measured through the sound velocity with a figure of eight microphone) to the total sound energy (measured with an omnidirectional microphone) in the first 80 milliseconds of an impulse response. The microphones are calibrated to give the same output on-axis in a free field. He called this measure the Lateral Fraction, or LF.

$$LF = \frac{\int_0^{80\,ms} lateralvel\,ocity(t)^2\,dt}{\int_0^{80\,ms} totalpress\,ure(t)^2\,dt} \qquad (1)$$

His proposal led to a firm association in the literature between early lateral reflections and spatial impression – and between spaciousness and ASW. Other authors have followed suit. For example, Hidaka and Beranek associate spaciousness with the Interaural Cross Correlation (IACC) of the first 80ms of a binaural impulse response. They called their measure IACCe. Morimoto has a similar measure.

$$IACC = \max_{(\tau->+-1ms)} \int_0^{\infty} \frac{Il(t-\tau)Ir(t)}{Il(t-\tau)^2 + Ir(t)^2} dt$$

The conventional view -- which relates spaciousness with ASW and early lateral reflections --makes several predictions that violate common observations. One of these is that small rooms (living rooms and offices) should sound spacious. For example, my office has dimensions of about 3.5m x 3.2m x 2.5m. It has a measured reverberation radius of 0.5m. Colleagues who converse with me typically sit 1m to 1.5 meters distant. In spite of having a substantial quantity of early lateral reflections, the room does not sound "spacious" at all! Although there is a pronounced perception of being in an enclosed space, the reflected energy is perceived as neither enveloping or belonging to a large open space. ASW and the desirable spatial aspects of musical acoustics - whether we call them spaciousness, envelopment, or reverberance - appear to be independent. We argue here that ASW is not a spatial impression at all – it is an observable property of the source image, but does not directly imply a space.

Recently it has become obvious that although LF and IACCe may be related to ASW, they do not correlate well with envelopment. Bradley and Souloudre propose that envelopment be measured by the lateral hall gain LG. LG is a measure of the absolute strength of the lateral sound energy 80ms and more after the direct sound. As with Barron's LF, the lateral velocity is determined from an impulse response measured with a figure of eight microphone. But this time we measure it at a standard distance from the source, 10 meters. The total pressure is again measured with an omnidirectional microphone, but this time in an anechoic chamber, and the result is adjusted for an effective distance of 10 meters. (We are assuming the source is omnidirectional.)

$$LG = \frac{\int_{80\,ms}^{\infty} lateralvelocity_{10\,m}(t)^2 dt}{\int_{0\,ms}^{\infty} totalanechoicpressure_{10\,m}(t)^2\,dt} \qquad (2)$$

The proposal has considerable merit – although the author believes that 160ms is a better choice as a beginning time for the integration.

Experiments by Bill Gardner and the author have shown that for source material such as speech or solo instrumental music, using reverberant levels typical of stage performance by a soloist, the impression of reverberance and envelopment depends on the absolute strength of the late reflected energy, and not the direct to reverberant ratio.

The reverberation matching experiments used solo music as a sound source. We smoothed impulse responses of equal reverberant loudness with a 160ms window, and then plotted them on the same scale. We found that they all tended to cross at about 160ms. This led us to propose a measure for the reverberant loudness for solo musicians – the amount of self support one hears while playing. This measure we called Running Reverberation, or RR160. If p(t) is an impulse response,

$$RR160 = \frac{\int_{160ms}^{320ms} p(t)^2\,dt}{\int_{0}^{160ms} p(t)^2\,dt} \qquad (3)$$

We repeated the experiments using orchestral music and reverberant levels typical of the audience areas in a concert hall. The results were quite different.

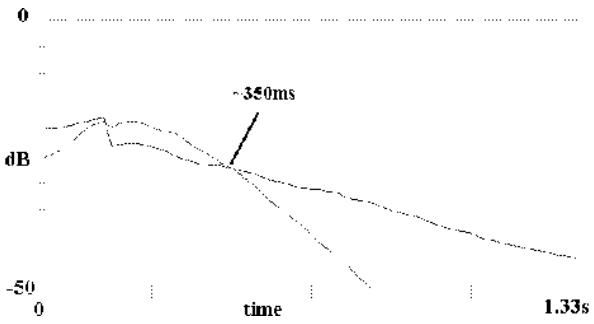

Figure 1: Two impulse responses of equal reverberant loudness for orchestral music. Both have been smoothed by a 160ms sliding window. Notice they cross at 350ms from the direct sound. For a two second reverberation time this is equivalent to 10dB of decay.

Figure one shows what happens when we plot two equal loudness impulse responses after performing a 160ms smoothing. They cross at about 350ms. These experiments have been repeated with many impulse curves, with similar results. It seems that the reverberant level at times at late as 350ms after the notes end is important to human perception of orchestral envelopment.

1.2 IACC

LF and LG do not explicitly depend on frequency. Low frequencies and high frequencies can both be measured with omnidirectional and figure of eight microphones. In any case the optimal sound direction for envelopment is 90 degrees. It is well known through experiments by Ando and others that the optimal angle for spatial impression of all types depends on frequency, with frequencies above 700Hz requiring sound sources closer to the medial plane. Ando, Blauert, Schroeder and others proposed using the InterAural Cross Correlation as a measure. Ando has shown that IACC and ASW correlate well, and that above 700Hz the optimal angle for reflected energy to produce ASW also produces the lowest values of IACC.

The author has shown that IACC is a measure for the ratio between the medial sound energy and the lateral sound energy. IACCe can be easily converted to a measure of the ratio between the lateral energy and the total energy in the first 80ms of an impulse response. There are two major differences between IACCe and LF. Both these differences result from using a dummy head instead of a figure of eight microphone to find the lateral energy. Cross correlation of the dummy head signals yields a measure that is quite close to human perception for frequencies 500Hz and above. Unfortunately the same is not true at low frequencies. Cross correlation is sensitive to phase and amplitude differences between the dummy head signals. Below 300Hz these differences become small, and IACC approaches one regardless of the perceptual properties of the room.

If we strongly believed in the accuracy of the IACC as a measure of spatial properties, we would expect that both ASW and spaciousness would always be low below 300Hz. The opposite is the case. In fact, in concert halls and opera houses envelopment can sometimes be primarily a low frequency phenomenon. The difference between an outstanding hall and an average one is often the presence of adequate low frequency envelopment.

1.1 IAD

The author has suggested a simpler measure – the InterAural Difference or IAD - that is useful in concert hall research. The IAD may also be useful in virtual audio. The IAD is the ratio in dB between the square of the equalized difference signal from a dummy head, and the sum of the squares of the two individual ear signals. The IAD can be calculated from an impulse response, or can be found as a continuous function of music signals. For example, if L and R are the left and right signals from a dummy head, then

$$IAD = 10 * \log_{10}\left(\frac{(eq(L(t)-R(t)))^2}{L(t)^2+R(t)^2}\right) \quad (4)$$

The equalization of the difference signal consists of a 6dB per octave boost below 300Hz. The equalization compensates for the reduction in interaural difference below 300Hz in a diffuse sound field. Ideally the equalization should be adjusted so the IAD has the value zero at all frequencies in a large reverberant space.

It is frequently useful in concert hall research to calculate the IAD in octave bands, smoothed with a 40ms window that is moved through a measured impulse response. This will show the lateral energy of an impulse response as a function of time, and this may give insight into where the reflected energy in a hall is coming from as a function of time and frequency.

1.3 RT, EDT, and LEDT

For many years acousticians have characterized concert halls by their reverberation time, RT. RT originally was measured by measuring the time it takes a sound to decay 60dB. RT can be calculated from an impulse response by measuring the slope of the impulse response after backwards integration.

A somewhat improved measure for reverberance and envelopment is the Early Decay Time, or EDT. EDT was originally defined as the time it takes a sound to decay 10dB. In this form the measure can be quite useful, as the first 10dB of decay is affected by many factors that influence the apparent loudness of the running reverberation.

Unfortunately Schroeder suggested measuring EDT from
an impulse response with a linear regression over the first 10dB of decay. The suggestion has not been helpful, as it eliminates most of the difference between the EDT and RT. Our work suggests that the EDT measure could be made more universal by replacing the –10dB limit of the EDT measurement by a 350ms limit. We recommend simply finding the slope of a line drawn from the peak of a Schroeder integral to a point 350ms down the decay curve. The new measure would more accurately reflect the sound of halls that were not close to 2 seconds of RT.

RT and EDT do not include any spatial attributes of the sound. As a further improvement to EDT we have suggested using the backwards integration of the IAD in the measure instead of the backwards integration of the sound pressure. The new measure is called the Lateral Early Decay Time, or LEDT. To measure LEDT we draw a line from the peak of the backward integration of the pressure to a point on the backwards integration of the IAD at 350ms. The slope of this line is the LEDT. Notice that in general the integration of the IAD is below the integration of the pressure – the IAD and the pressure are only equal in a diffuse field. Thus values of the LEDT tend to be less than values for the EDT, and the more the reverberation comes from the front the lower the value of LEDT will be.

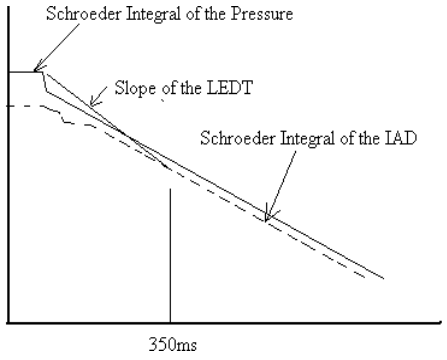

Figure 2: Diagram showing the calculation of the LEDT from the Schroeder integration of the impulse response and the IAD. In this diagram the pressure curve is backwards integration of the rms sum of the two binaural channels, and the IAD is the backwards integration of the equalized difference between the two channels.

2. INTERAURAL FLUCTUATIONS

So far in this paper we have looked at envelopment as a perceptual phenomenon, without attempting to explain how it is detected. We found some years ago that envelopment as a perception depends on fluctuations in the Interaural Time Delay (ITD) and Interaural Intensity Difference (IID). We will not explain this theory in detail here, but we will outline the major results.

First, simple measurements show that small amounts of lateral reflected energy can cause large fluctuations in the IID and the ITD.

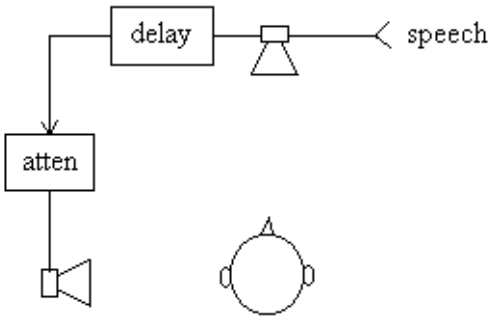

Figure 3: A single reflection at –10dB can produce 6dB fluctuations in the IID.

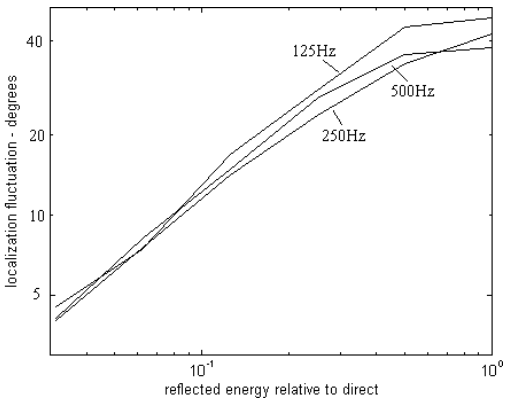

Figure 4: Plot of fluctuation in the IID (expressed as degrees of apparent angle) as a function of the reflected energy for a single reflection at 90 degrees.

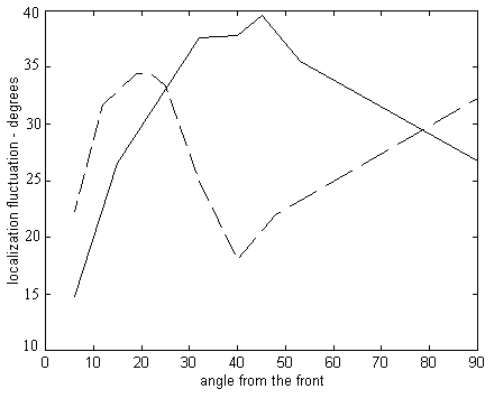

Figure 5: Interaural Fluctuations as a function of frequency and lateral angle. Solid line is 1000Hz, dashed line is 2000Hz

2.1 Minimum time delay needed for fluctuations

Interaural fluctuations arise from interference between lateral reflected energy from different directions, and interference of lateral energy with the direct sound. In the simple case shown in figure 3, where there is only a single lateral reflection, there is a minimum delay needed before significant fluctuations develop. The physics of sound interference shows that the minimum delay needed is inversely proportional to the bandwidth of the signal. The minimum delay is approximately 0.5/bandwidth. For example, a 100Hz bandwidth signal requires a minimum of 5ms delay to develop fluctuations.

The bandwidth of a critical band on the basilar membrane is approximately 100Hz for mid frequencies, so we would expect mid band noise signals to require about 5ms delay before envelopment is created – and this is exactly what we observe.

If we try single reflection experiments using 3kH and higher noise signals, the minimum delay needed is smaller. This is observed also. A critical question then arises – what is the bandwidth of a "typical" music signal? The answer depends both on the type of music and the frequency. A large string section with several different musical lines can produce a signal that is quite wide in frequency – filling a substantial portion of a high frequency critical band. For such a signal the effective bandwidth is that of the critical band filter in the ear.

This is not the case with a bass guitar line. While a particular note is held the bandwidth can be quite low, particularly when the note is played without vibrato. The bandwidth is typically less than one Hz. However we should remember that we are concerned with envelopment here, and envelopment is ideally dependent on the reverberant component of a musical signal. What is the bandwidth of reverberation?

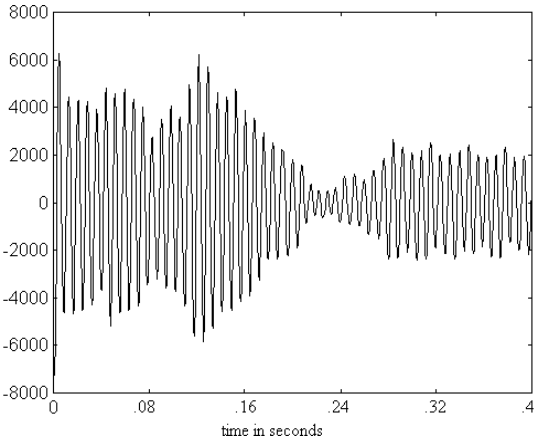

Figure 6: Decay of a 60Hz tone in Boston Symphony Hall.

When a tone decays in a large hall interference produces fluctuations in phase and amplitude. These fluctuations increase the bandwidth of the signal. The broadening depends on many factors, but can be as high as 3Hz.

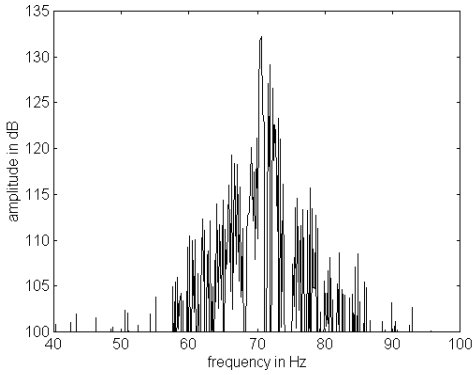

Figure 7: Bandwidth of the decay in figure 6. Note the approximately 3Hz width.

3. THREE TYPES OF ENVELOPMENT

One of the reasons for the confusion about the perceptual meaning of spaciousness is that spatial perceptions depend strongly on the type of sound used as a stimulus. In our research we have identified three separate spatial impressions, only two of which could be called enveloping. These perceptions are Continuous Spatial Impression (CSI), Early Spatial Impression (ESI), and Background Spatial Impression, (BSI).

Continuous spatial impression (CSI) results when lateral reflected energy interferes with a continuous sound source, such as pink noise. The reflections that give rise to CSI can have any time delay greater than about 10ms. CSI is fully enveloping – sound seems to come from all around the listener, even if there is only a single lateral reflection. CSI also depends on the ratio of the medial sound to the lateral sound. It is independent of the absolute loudness of the source.

Early Spatial Impression (ESI) results from lateral reflected energy that arrives within 50ms of the end of an impulsive sound, or a sound that consists of discrete notes. ESI is not enveloping – the spatial impression is that of a small room, and the room sound seems to come from the general direction of the sound source. Perceptually the room sound is bound to the source itself. Like CSI, ESI depends on the ratio between the medial and the lateral sound. It does not depend on the source strength.

Background Spatial Impression (BSI) arises when the source consists of a series of short notes or a sequence of phones from speech. The brain organizes the notes or phones into a foreground perceptual stream – the words of a single speaker, or the notes in a melody. Where there are more than one series of notes – a musical duet, or two people talking at once, the brain assigns each series to a separate foreground stream. Energy that arrives in the spaces between the phones or notes is assigned to a single stream – the sonic background.

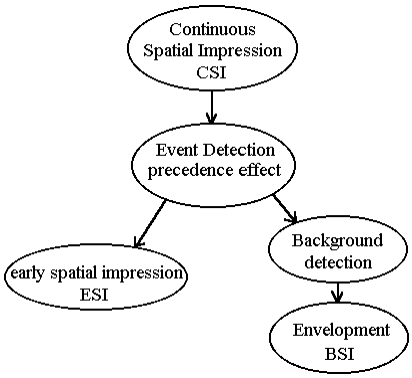

Figure 8: The neural process of event detection causes the fluctuations to be split into two types – early and late. Early fluctuations give rise to ESI, late ones to BSI.

Sounds that arrive within 50ms of the ends of the notes are bound to the notes themselves, and form part of the foreground stream. If these sounds have a spatial direction different from the direct sound the perception of ESI occurs. Sounds that arrive just a little later than 50ms after the ends of notes are difficult to hear. There is a perceptual inhibition that occurs in the background stream just after 50ms. This inhibition is

gradually released, such that background sounds arriving 150ms or so after the ends of the notes are strongly audible. If these sounds are sufficiently spatially diffuse they will be perceived as BSI – and they will be enveloping.

The strength of the BSI perception is absolute – it depends on the amount of spatially diffuse reverberant energy, and thus depends on the strength of the sound source. The louder the music is played the greater the perception of BSI will be. At sound levels typical of classical music the strength of the CSI perception depends on the ratio of the direct sound to the reverberant sound – and is always less than the strength of the BSI perception. The difference is level dependent, and is typically about 6dB. This is another way of saying that lightly scored music or solo music is likely to sound more reverberant than thickly scored music, even if the thickly scored music is louder.

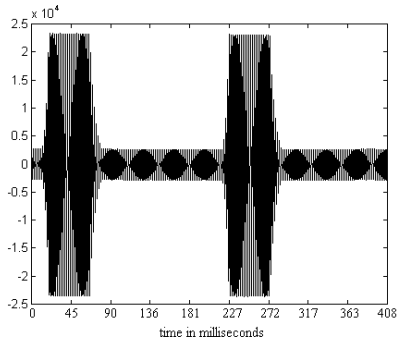

Figure 9: Consider a series of tone bursts as a signal.

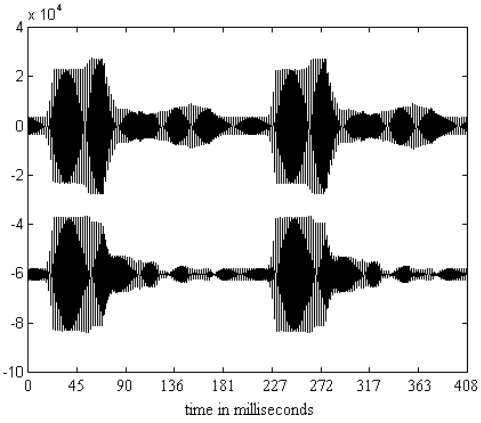

Figure 10: When we add reverberation the beginning of each burst is uncorrupted, and there are large interaural fluctuations in the spaces between the notes.

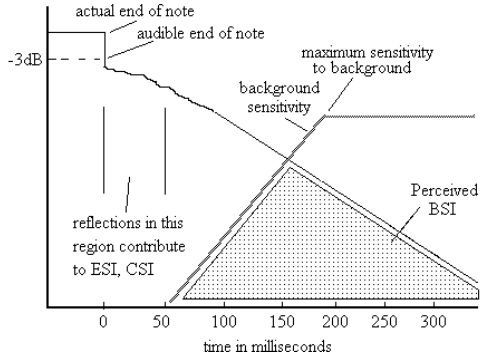

Figure 11: The separation process assigns the early fluctuations to the note itself. After a period of inhibition the later fluctuations are assigned to the sonic background.

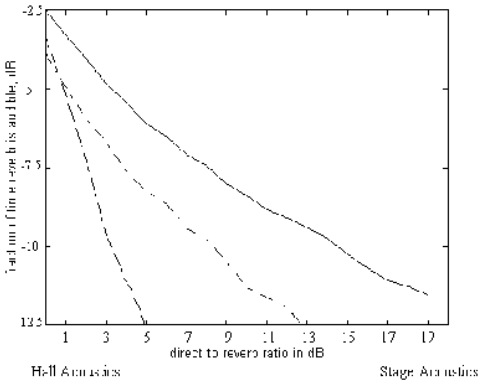

Figure 12: Musical masking for three different types of music. Solid line – solo recorder. Dot – dash – Chamber music. Dashed line – orchestral music. Notice that for orchestral music the slope is four to one – a 1dB change in direct to reverberant ratio makes a 4dB change in reverberant audibility.

Figure 12 shows that reverberant audibility also depends on the transparency of the music to reverberation. When the music is thickly scored, with very few places where reverberation can be heard, the audibility of the reverberation, and thus the perception of envelopment, becomes a strong function of the direct to reverberant ratio.

4.0 SPATIAL IMPRESSION IN SMALL ROOMS

For measuring or understanding virtual audio systems the distinction between BSI, ESI and CSI is important. We must separately consider how envelopment arises with continuous sounds and with music and speech. We must pay particular attention to the behavior of a virtual system to reverberation. Reverberation must be

reproduced with real or virtual spatial diffusion – particularly in the horizontal plane.

5.0 LF, LG, IACC, EDT AND LEDT AS MEASURES OF BSI

The perception of BSI depends both on the "transparency" of the source to reverberation and on the source loudness. Unfortunately these two source properties interact. Our preliminary experiments with solo music and orchestral music indicate that the time delay needed for the maximum BSI is greater with more complex music. Thus for accurately measuring BSI in the concert hall we need to know the type of music. Of the measures presented so far, only LG is an absolute measure. LG will be higher when the absolute reverberant level in a hall is high. However the measure is not as simple as it might seem. In most halls LG and the normal hall gain will be very similar. Hall gain is measured in a very similar way to LG, but with an omnidirectional microphone. In classical room acoustics hall gain is determined almost entirely by the total amount of absorption in an enclosed space. For this reason hall gain tends to be high in small halls, and low in large halls. In other words the reverberant level is high in small halls, and low in large halls. The direct to reverberant ratio may be more constant – listeners on the average sit further from the source in large halls, so the ratio tends to be constant.

However, small halls typically have shorter reverberation times than large halls. In a small hall much or even most of the reverberant energy falls in the first few hundred milliseconds. Thus it is vital to know at what time delay we start the integration in calculating LG. If we start the integration too soon, small halls will measure more enveloping than they sound. We better do it right, because we know from experience that large halls are typically more enveloping than small ones.

What happens if the room is the size of a home listening room, or a small studio used for mixing music? By definition a small room cannot be spacious – and it is our common observation that small rooms are seldom enveloping. Although small rooms can have a high reverberant level – remember the reverberant level is inversely proportional to the total absorption – the time delay of this reverberation is short.

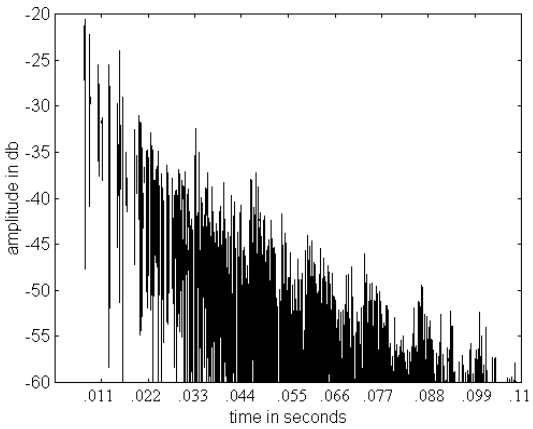

Figure 13: the decay curve of a 12'x15'by 9' room, with surface reflectivity of 0.8. The reverberation time is 0.2 seconds, and the time constant (decay to 1/e) is 30ms. The net delay time of the reflected energy is insufficient to generate interaural fluctuations at mid frequencies. This room will not be enveloping with a single sound source.

When the time constant of the room is small compared to the inverse bandwidth of the music, interaural fluctuations are not created – and envelopment will be low. Envelopment may be larger with a high bandwidth signal, such as pink noise. In the room shown in figure 13 the minimum bandwidth of a signal that will produce CSI is 17Hz, and BSI will not be produced at all.

The delay time in a small room is sufficiently small that the only form of envelopment that can be produced is CSI. The primary spatial impression in these rooms is ESI – the spatial impression of small rooms.

6.0 SPATIAL PROPERTIES OF SMALL ROOMS WITH RECORDED MUSIC AND MULTIPLE DRIVERS

In an anechoic space envelopment can be created by placing loudspeakers at opposite sides of the listener, and driving them with signals that vary in amplitude and phase in a way that can reproduce the interaural fluctuations that would be heard in a large space.

Figure 14: Envelopment can be produced with multiple loudspeakers even in an anechoic space if the loudspeakers are lateral and reproduce signals that fluctuate appropriately.

However the amplitude of the fluctuations produced depends on the sine of the angle of the loudspeaker from the front. Below 700Hz fluctuations will be maximum when the loudspeakers are at the side (90 degrees), but the fluctuations decrease rapidly as the speakers move toward the front.

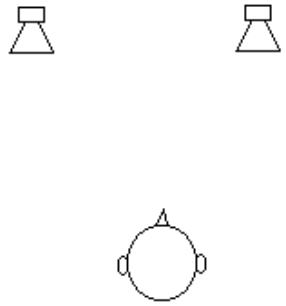

Figure 15: A standard stereo loudspeaker placement produces little envelopment below 700Hz in an anechoic environment because the speakers are not at the opposite sides of the head.

Although a standard stereo system such as the one in Figure 15 produces little envelopment at frequencies below 700Hz, it can work better at higher frequencies. Figure 5 shows the angular dependence of envelopment for a single reflection. It also predicts the optimum angle for two decorrelated loudspeakers. Loudspeakers at +-30 degrees can produce a substantial sensation of envelopment in the 1700Hz band. There are other phenomena that act at higher frequencies that also influence envelopment.

7.0 ENVELOPMENT AT HIGH FREQUENCIES

In carefully controlled loudspeaker experiments Morimoto's group in Kobe found that reflections from behind the listener are more enveloping than reflections from the front. The sound source was solo violin. Morimoto proposed measuring envelopment through the ratio of reflected energy from the front to reflected energy from the rear. He calls this ratio the front/back ratio.

These experiments indicate that where it is possible to discriminate between front and rear sound, rear sound is perceived as more enveloping. We know from the physics of sound perception that it is possible to distinguish front sound from rear sound in two ways.

At low frequencies the two can be distinguished through small movements of the listener's head. Small head movements cause predictable changes in the ITDs. Unfortunately head movements produce no shifts in ITDs when the sound field is largely diffuse.

The primary front/back cue at high frequencies is spectral. Sound that comes from the front of a listener has notches in the frequency response at about 8kHz. As a source moves from the front to the side and then the rear, these notches disappear around +-90 degrees, to be replaced by notches at about 5kHz when the source reaches about +-150 degrees. The solo violin spectrum is particularly rich in sound energy in the region of these directionally dependent spectral anomalies.

We believe that the front/back ratio is primarily important for sound that includes a substantial amount of energy above 3kHz. With such sound sources it is most likely possible to make the front/back discrimination during the separation of sound into a foreground and a background stream. Thus the front/back ratio is likely to contribute to both the continuous form of envelopment, and its perceptually much more important relative, the background perception.

8.0 AN OBJECTIVE MEASURE FOR ENVELOPMENT AT LOW FREQUENCIES IN SMALL ROOMS – THE DFT

Although we show above that a standard stereo system is capable of producing envelopment at high frequencies, it is clearly not able to do it in an anechoic environment at low frequencies. What happens in a typical room? Clearly room modes will alter the situation. We need a way of quantifying the degree of envelopment produced by various loudspeaker positions in various rooms. Ideally we would like a method that quantifies the envelopment that is likely to be produced from measured or calculated binaural impulse response. A simple measure has proved

elusive. However the theory of interaural fluctuations is sufficiently advanced that we can develop a method directly from it. The method involves a substantial amount of calculation, but appears to give quite sensible answers. We call the method the Diffuse Field Transfer function, or DFT.

It is clear from the previous discussion that if a small room is to produce envelopment (BSI or CSI) the fluctuations in the IID and ITD at the ears of the listener must be transmitted in some way from the recorded signal to the ears of the listener. We will assume that the recorded material contains fluctuations in phase and amplitude that are characteristic of the recording environment. We want to know how the room transfers this information to the ears of the listener – thus the origin of the name DFT.

Here are some of the requirements of the DFT:

1. For a monaural source the DFT should be low whenever the inverse bandwidth of the musical signal is large compared to the time constant of the room. In practice this means the DFT should give a low value for a monaural source in nearly any listening room.
2. The DFT should be low for any two channel system that runs the loudspeakers in phase. (Two loudspeakers with any constant phase relationship are equivalent to a single source.)
3. The DFT should rise when there are at least two sound sources that reproduce signals that have fluctuating amplitude and phase.
4. The DFT calculated in an anechoic environment should be a maximum when the two loudspeakers are at opposite sides of the listener, and decrease to zero proportionally to the sine of the angle of the loudspeakers from the front.

The process of finding the diffuse field transfer function can be summarized:

1. Calculate (or measure) separate binaural impulse responses for each loudspeaker position to a particular listener position. When resampling a calculated response a high sample rate must be chosen to maintain timing accuracy. In our experiments 176400Hz is an adequate sample rate.

2. Low-pass filter each impulse response and resample at 11025Hz, and then do it again, ending with a sample rate of 2756Hz. This sample rate is adequate for the frequencies of interest, and low enough that the convolutions do not take too much time.

3. Create a test signals that accurately model music. Initially we used independent filtered noise signals. Various frequencies and bandwidths can be tried, depending on the correlation time of the musical signal of interest. More recently we have been using the sum and difference of sinusoidal sweep signals with a fixed frequency offset. These signals seem to give more predictable results, and intrinsically average over a band of frequencies.

4. Convolve each binaural impulse response with the stereo music signal, and sum the resulting convolutions to derive the pressure at each ear.

5. Extract the ITD from the two ear signals by comparing the positive zero-crossing time of each cycle.

6. Average the ITDs thus extracted to find the running average ITD. The averaging process weights each ITD by the instantaneous pressure amplitude. In other words, ITDs where the amplitudes at the two ears is high count more strongly in the average than ITDs where the amplitude is low.

7. Sum the running average ITD and divide by the length to find the average ITD and the apparent azimuth of the sound source.

8. Subtract the average value from the running average ITD to extract the interaural fluctuations.

9. Filter the result with a 3Hz to 17Hz bandpass filter to find the fluctuations that produce envelopment.

10. Measure the strength of these fluctuations by finding the average absolute value of the fluctuations. The number which results is the Diffuse Field Transfer function, or DFT.

11. Calibrate the DFT by using the system to measure the fluctuations induced by two decorrelated sources at +- 90 degrees in an anechoic environment.

12. Measure the DFT as a function of the receiver position in the room under test.

The most difficult part of this process turned out to be building the ITD detector in software. The detector must be robust. The signals at the ears are noise signals – in many places the amplitude is low, and the zero crossings can be highly confused. Our detector

should use very simple elements – just timers and filters – to do the job. It must be very difficult to confuse. The design of this detector is beyond the scope of this paper. Persons interested in its design, or in the Matlab code for the whole DFT measurement apparatus, should contact the author. In the current version of the code, it takes about 15 seconds to find the DFT at a single receiver position, so a 7x7 array of positions can be calculated in about 12 minutes (on a laptop with a 150MHz Pentium.)

We wrote a simple image model for rectangular rooms using the Matlab language. Although an image model is not accurate in general for frequencies where the wavelength is large compared to the room dimensions, it can be shown that in a rectangular room where all the surfaces have equal reflectivity the model gives accurate results. In practice even if the surfaces do not have equal reflectivity the simple model seems to do well.

To model the head we simply use two point receivers separated by 25cm. This separation is about right for frequencies below 150Hz. We could improve the model by using measured HRTF's instead of our simple model. It is not clear if the benefit would be worth the considerable pain.

8.1 Two swept sine signals as a musical signal

The musical signal is difficult to model. In the original work we used a narrow band noise signal. To model the low frequency reverberation from musical instruments we chose a bandwidth of 3Hz. The choice of bandwidth is quite important. If we use a bandwidth of more than about 20Hz we find that even small listening rooms can be enveloping with a monaural source. (Listening tests with octave band pink noise found that broadband noise signals through a single loudspeaker are enveloping, whereas musical signals are not.) The problem with using a noise signal is that the measure becomes not very reliable. One must use many seconds of noise for an accurate measurement, and this is computationally expensive. The measurement is also only valid for a particular frequency. Ideally one would like a measure that averaged over a range of frequencies.

We have recently replaced the noise signal with a signal constructed from two sinusoidal sweeps. For example, if we want to make a measurement in the 60-70Hz range we generate a sweep from 60Hz to 70Hz, taking about 10 seconds to complete the sweep. We then generate a second sweep, going from 63Hz to 73Hz in the same time. We then take the sum and the difference of these two sweeps. The sum signal is applied to one of the two sound sources, and the difference to the other. The two signals combine in the room to produce interaural fluctuations that model the DFT over the ~60Hz to 70Hz range.

8.2 Calibration of the DFT

With a single driver in an anechoic space the DFT should be low. In practice, in spite of our limited length of noise or the limited sine sweep, the values we get are at least 40dB less than the maximum values with two drivers. Thus our detector passes this test.

There are two major adjustments to the detector that we must set as best as we can to emulate the properties of human hearing. The most important of these is the bandwidth we choose for the noise signal (or the frequency offset between the two sine sweeps.) When we want to study the envelopment of noise signals in a room we would like a noise signal with the bandwidth and filter shape of a single critical band on the basilar membrane. In the Matlab code we use a sixth order elliptical bandpass filter, similar to the ones in a sound level meter. Let us assume we want to measure the DFT of broadband signals. What bandwidth do we need?

In an effort to answer this question a series of experiments on the envelopment of low frequency noise signals was performed. It is possible to probe the properties of the human envelopment detector through experiments with single lateral reflections, and with multiple lateral reflections. We are interested in how the envelopment impression depends on the delay of single reflections, or the combination of delays in multiple reflections. The apparatus described in [1] was used, with continuous band filtered pink noise as a source. The results were highly interesting.

First, (with a single subject) we found that the envelopment from a single lateral reflection oscillates as the delay is changed. There is an interference effect. For frequencies in the 63Hz octave band a single lateral reflection at 5.5ms delay produces a very wide and enveloping sound field, and it takes very little reflected energy to do it. A delay of 13ms produces a nearly monaural impression, with little or no envelopment at all. As the delay rises the envelopment goes through one more cycle, becoming first super wide, and then somewhat less wide. Beyond 20ms all delays sound about the same.

This interference behavior arises from easily calculated cancellation between the direct sound and the reflection. Such interference is not possible when the delays are greater than the coherence time of the noise

signal, and this depends on its bandwidth. Thus the properties of the basilar membrane filters can be studied through the interference effect. We find that at least at 63Hz the basilar membrane can be modeled by an elliptical filter of one octave width. If the basilar membrane was significantly sharper than one octave at 63Hz, we would expect the interference effect to extend to greater delays. If you use a ½ octave filter in the DFT detector you find that the interference effect with a single reflection does extend to higher delays. A one octave filter at 63Hz seems about right.

As an aside – we also found to our surprise that when multiple reflections are used the envelopment and the DFT depend strongly on the particular combination of delays chosen. This observation has pronounced implications for this type of experiment. It appears that when there are multiple lateral reflections the delay times of the reflections relative to each other matter a lot. In fact, once a pattern has been set, the delay of this pattern relative to the direct sound can be varied with no change in envelopment. Some patterns are highly enveloping (greater than a diffuse field) and some are not enveloping at all. It seems that it is not just the total energy in the early reflections that is spatially important! If the delay times are varied randomly at about a one second rate this problem is avoided, and this is the approach we have taken in studying multiple reflections. In a concert hall, where the details of the arrival times of different lateral reflections is different for every instrument the variability of different delay patterns is not likely to be noticed.

8.4 Bandwidth of the test signal for music:

When we wish to calculate the DFT for music the test signal must have a longer correlation length than for noise. As mentioned above, we can use real reverberation as a test signal, or we can emulate the reverberation with a noise signal of 1-3Hz bandwidth. We will show some results based on narrow band noise. Recent work using sweeps appears to be promising, but is not included in this paper.

8.5 Other considerations in the DFT measurement system:

Another physiological variable in the DFT detector is the time constant used in the running ITD filter. Without this filter the ITD detector is hopelessly inaccurate, so there is a considerable reason to include it. Here we simply guess. A time constant of about 50ms seems to work well. In practice the filter is implemented with a variable time constant. The TC is 50ms for strong signals, and rises linearly as the signal amplitude falls. This amplitude dependence keeps zero crossings at low amplitudes (which tend to be very noisy) from affecting the running average very much.

The 3Hz to 17Hz bandwidth used for the fluctuations is also a bit of a guess. It is based on measurements made with amplitude modulated and phase modulated pure tones. The bottom line is that this bandwidth gives quite reasonable results, so it seems a good choice for now.

The raw output of the DFT measurement is a number that represents the average of the absolute value of the interaural fluctuations. It is expressed in milliseconds. What is the meaning of this number? How large should it be when the envelopment is "just right", and is it possible for it to be too high?

We might think we could calibrate the DFT by using impulse responses measured in concert halls. This method has some advantages, but is probably not what we want to do. The DFT depends ultimately on the relationship between the strength of the lateral sound field in the region of the listener with the medial sound field. In a true diffuse field the medial signal will dominate the lateral signal, since the medial field includes both the front/back direction and the up/down direction. In a concert hall this situation is altered by the floor reflection. At 63Hz the floor reflection enhances the lateral direction by canceling the vertical sound waves. At 128Hz the opposite happens, with the vertical sound being enhanced by the reflection. The floor reflection does in fact change how spacious a hall sounds – I have been able during rehearsals to compare the envelopment standing and sitting, and the standing position is more spacious. In any case, why should we expect that a concert hall – even a very good one – would be optimal?

Once again experiments were performed to measure the envelopment using the apparatus of [1]. We found that below about 200Hz the most pleasing overall sensation of envelopment occurs when two uncorrelated noise sources are on opposite sides of a listener in an anechoic space. (Above this frequency the envelopment from such an array seems too wide, and a more diffuse soundfield is preferred.) This implies that below 200Hz the optimum value of the DFT is similar to the values we would measure in a concert hall at 63Hz, where the floor reflection enhances the lateral component of the sound. The DFT of the same hall at 125Hz is NOT preferred! Our conclusion is that a diffuse field (without the floor reflection) is not optimal for envelopment at low frequencies. We decided to calibrate our DFT detector

by measuring the DFT for a detector exactly between two noise sources in an anechoic space.

Figures 16 to 22 show various uses of the DFT to calculate the envelopment in rooms. The captions speak for themselves.

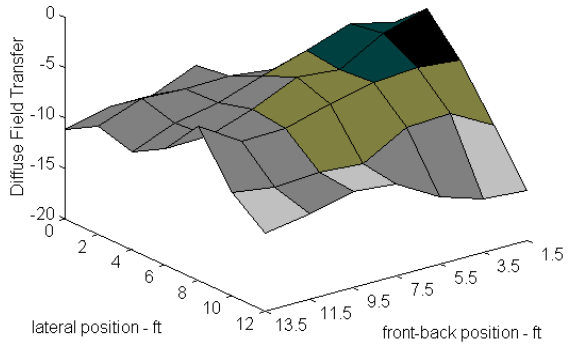

Figure 16: DFT (using 3Hz bandwidth noise at 63Hz) from two loudspeakers in an anechoic space. Note the DFT is largest directly between the two loudspeakers and reduces in level as the listener moves away.

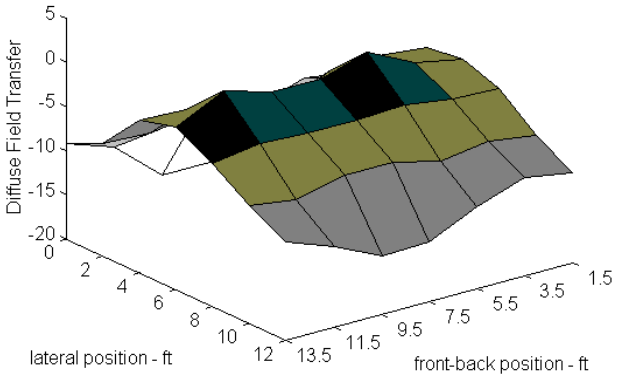

Figure 17: The same, but with the drivers at the side of the space, at 7.5 feet from the front.

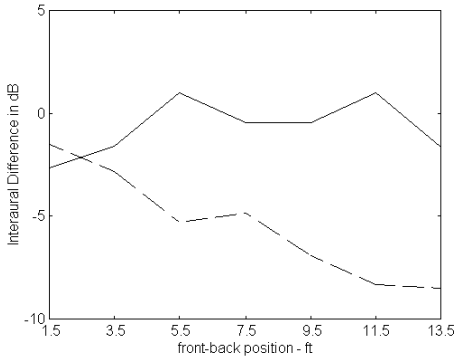

Figure 18: The DFT along the center line for figures 16 and 17. Solid line, drivers at the side. Dashed line, drivers at the front.

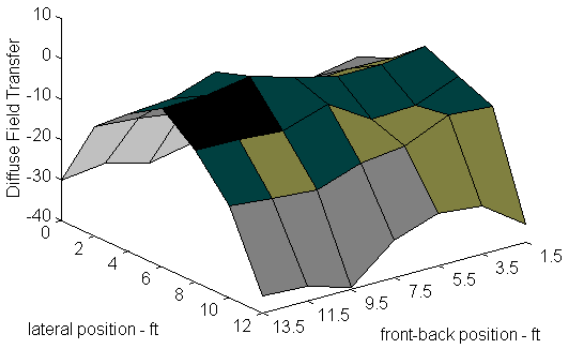

Figure 19: DFT calculated with an octave band noise signal in a 12x15x9ft room, with wall reflectivity 0.8. A single driver was in the front left corner. Note that with an octave band signal this room can produce significant envelopment.

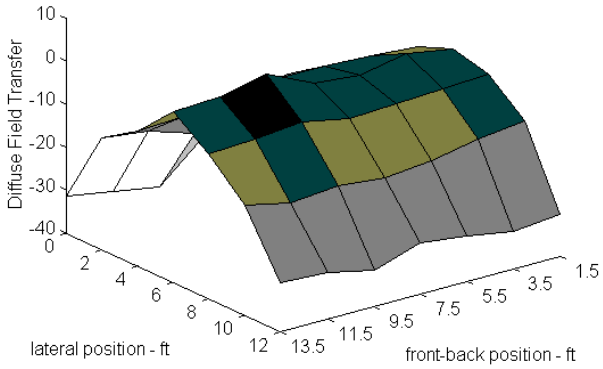

Figure 20: DFT from octave band noise signals in the same room, but with two uncorrelated drivers.

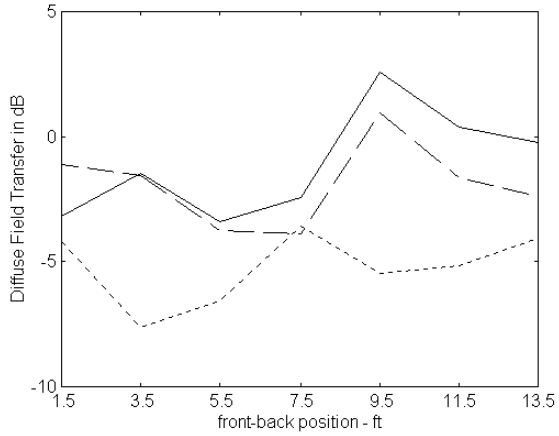

Figure 21: The DFT from octave band noise signals along the center line of the room in figure 20. Solid line is for two drivers, reflectivity 0.8. Dashed line is for two drivers, surface reflectivity 0.6. The solid line is for a single driver, reflectivity 0.6. Notice that as the room becomes drier the envelopment from a single driver goes down dramatically.

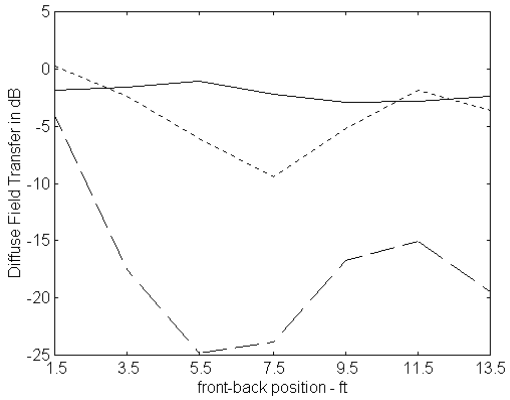

Figure 22: DFT from a 3Hz bandwidth signal in the room of figure 20, with surface reflectivity 0.8. Solid line – two drivers at the sides of the listening area. Dotted line – two drivers in the front of the listening area. Dashed line – a single driver in the front left corner. Note that with a signal that approximates music the envelopment in the room is very low unless two independent drivers are used.

9.0 OBJECTIVE MEASURES OF VIRTUAL SYSTEMS

Virtual systems are generally assumed to be evaluated in an anechoic space. Thus the listening environment can contribute nothing to the impression of envelopment. The envelopment impression must be transferred from the recording to the listener. We must assume that the system is used with signals that contain the fluctuations in phase and amplitude that create envelopment in human hearing. As when we evaluate the envelopment in a stereo system, we are concerned with the ability of the system to transfer these fluctuations to the ears of the listener.

In the absence of listening room reflections it is primarily the success of virtual process in generating good lateral separation at low frequencies that will determine the amount of envelopment we will perceive. Thus lateral separation at low frequencies is probably the best measure of virtual systems for envelopment. The optimum low frequency limit is much lower than most virtual systems are designed to work. In our concert hall work we find that frequencies as low as 60Hz are vital to world class envelopment. A virtual system that provides separation down to 200Hz or 300Hz will not create the envelopment we expect in a great hall.

At higher frequencies we should be concerned with the ability of the system to develop an image at about +-150 degrees behind the listener. A system that can do this well – and maintain good separation between the left rear and the right rear source will provide excellent high frequency envelopment when provided with source signals that contain decorrelated high frequency sound in these two rear channels.

CONCLUSIONS

We have given a overview of the current state of objective measures for envelopment. In spite of some confusion, adequate measures exist. For sound reproduction systems in small rooms the DFT measure may be the only one that makes sense for low frequency musical signals. For virtual systems lateral separation seems to be the best measure – particularly at frequencies below 300Hz. There is some advantage to creating additional sources at high frequencies at +-150 degrees from the front, if there are musical signals that contain decorrelated information above 2kHz, and these signals can be realistically applied to the rear virtual sources.

REFERENCES

1. Griesinger, D. “The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces” Acta Acustica Vol. 83 (1997) 721-731
2. Griesinger, D. “Speaker Placement, Externalization, and Envelopment in Home Listening Rooms” AES preprint 4860

3. Griesinger, D. “General overview of spatial impression, envelopment, localization, and externalization” – presented at the 15th international conference of the AES, Denmark 1998.
4. Griesinger, D. “Spatial impression and Envelopment in Small Rooms” AES preprint 4638.