General overview of spatial impression, envelopment, localization, and externalization

David Griesinger

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

Abstract

This paper will provide an overview of the current knowledge on the problems of spatial impression, envelopment, and localization in enclosed spaces. The emphasis will be on small spaces, but the approach draws heavily on concert acoustic studies. The literature on this subject has been contradictory and controversial. In the last few years a consensus has been developing that envelopment is primarily determined by lateral reflected energy arriving at least 80ms after the direct sound. Localization is traditionally linked to the "apparent source width" (ASW) which has been associated with the interaural correlation (IACC) of the first 80ms of the impulse response of a space. Although ASW and IACC may be useful in concert hall measurement, they do not work well in small spaces. This paper will present two new measures, the Diffuse Field Transfer function (DFT), and the Average Interaural Time Delay (AITD). The DFT is a measure of envelopment, and is useful both in small rooms and large rooms. The AITD is a measure for "externalization" a sonic property unique to small rooms. A closely related measure, the Net Interaural Time Delay (NITD), is useful in understanding localization in small spaces.

1. INTRODUCTION

In the best concert halls and opera houses low frequency sounds envelop the listeners. Although one is aware that the attack of the kettledrums comes from the stage or the pit, the ring of the drum and the rumble of the bass drum come from all around the hall. The bass viols and the cellos have the same property, particularly when they play pizzicato. One of the joys of an organ concert is hearing the bass swirl around the cathedral when a pedal note is held. When the acoustics produce envelopment music has a living quality that is highly prized by conductors and players. When recorded music is played through loudspeakers, envelopment can often seem adequate at frequencies above 1000Hz, but poor at lower frequencies. In fact, many recording engineers seem to be unaware that low frequency envelopment is either possible or desirable.

Envelopment at higher frequencies can also play unexpected tricks. Normally the sound image from a conventional stereo system stays fixed between the two loudspeakers. But this is not always the case. Occasionally sound seem to surround the listener, even in a non reflective room.

Listening rooms also suffer from a perceptual anomaly that has no obvious counterpart in performance spaces. Low frequency instruments in popular music, such as the kick drum and the bass guitar, are almost always perceived as coming from inside the head. This perception does not occur in the concert venue, even when these instruments are amplified. This "in the head" localization is unique to recorded music. It is always perceived as artificial by the author. In this paper we will use the word "externalization" to describe this perceptual property.

Both externalization and envelopment depend strongly on the recording technique, but they appear to be independent of each other. In rooms where low frequency envelopment is perceptible, low frequency instruments in classical music are often perceived as external, while low frequencies in popular music are often "in the head".

Both envelopment and externalization are highly dependent on properties of the room. Years ago we noticed that it is possible to perceive low frequency envelopment in some home listening rooms, and not in others. In [43] we attempted to study envelopment through measurements of localization. We noted that in many listening rooms it was possible to localize low frequencies

to a particular loudspeaker, but phantom images were unstable. Panning low frequencies between two loudspeakers did not yield the same positional dependence as is noted at higher frequencies. The phantom image tended to pull to the center of the listener's head – and be judged as closer to the center than it was intended to be.

In [43] we found that the apparent position of the low frequency sound could be brought more into alignment with the high frequency sound if the separation at low frequencies was increased electronically. The circuit, dubbed a "spatial equalizer" has become popular with many engineers. However the primary virtue of the spatial equalizer for these engineers turned out not to be improved localization, but enhanced envelopment. The spatial equalizer works by increasing the left minus right (L-R) component of the sound below 300Hz. In most listening rooms the improvement in localization is subtle, but the improvement in envelopment is obvious.

In the succeeding years we noticed that this circuit is completely inaudible in some rooms. Ironically, in most sound mixing studios the circuit is inaudible because the antiphase component of the low frequency sound cannot be heard. These rooms typically have a high degree of symmetry and carefully control the low frequency reverberation time. The widespread use of such rooms for sound mixing has had at least two undesirable effects. These rooms tend to make professional engineers unaware that low frequencies can be enveloping. These rooms also encourage the use of microphone techniques that enhance imaging at the expense of envelopment. For example, recording techniques that utilize only closely spaced omnidirectional microphones (such as most binaural techniques) produce excellent imaging with earphones, but poor low frequency envelopment with loudspeakers. If your room does not permit you to hear envelopment, you will not know what you are missing.

The externalization of low frequencies is even more mysterious. Typical home stereo systems often externalize low frequencies, whereas symmetrical listening spaces are the most likely to sound artificial. These rooms are often described by their owners as possessing "tight" low frequency imaging. To my ears the low frequencies are centered, but they are unlike anything one would hear in a concert.

So we have at least three mysteries to untangle. First, why do some rooms support low frequency envelopment, and what can be done to provide it in rooms that do not? Second, why do the kick drum and the bass guitar almost always end up banging away inside your head and what can we do to get them out? Third, why do some recordings sound enveloping even when you listen to them through two front loudspeakers in a relatively non reflective room?

2. ENVELOPMENT IN CONCERT HALLS

The study of envelopment in concert halls has been marked by contradictions between common observation and accepted theory. In [44] we outline a theoretical framework that resolves these discrepancies. The framework has the following major parts:

1. Envelopment at low frequencies is perceived when the interaural time delay fluctuates at a rate of between 3Hz and 20Hz. Above 400Hz fluctuations in the interaural intensity difference (IID) and fluctuations in the ITD are both important.

Below 400Hz the interaural time delay is the principle cue for localizing the horizontal direction (azimuth) of low frequency sounds. In the absence of reflections, the ITD determines azimuth with high accuracy – within a few degrees at frequencies of 500Hz and above. Below 500Hz the accuracy is proportional to the frequency, so that localization to +-20 degrees is still possible in the 63Hz octave band.

Lateral reflected energy causes the ITD – and thus the perceived azimuth – to shift. When the sound source is broad-banded, or consists of a musical tone with vibrato, the shift in ITD becomes a fluctuation. For sources of both speech and music the fluctuation is essentially random (or chaotic) in nature. Fluctuations at rates slower than 3Hz are perceived as source motion. Above this frequency they are perceived as envelopment.

2. Where the sound source consists contains rapid attacks – such as the start of a speech phoneme, or the attack of a musical note – the onset of the sound at the listener is uncorrupted by reflections. In this case the ITD during the attack accurately determines the sound direction, and later fluctuations

produce envelopment. Thus it is possible to have both good localization (a low apparent source width) and high envelopment at the same time. At higher frequencies the IID performs a similar role, and localization is determined by both IID and ITD.

3. For speech or music that is relatively transparent to reverberation, the fluctuations are maximal during the pauses between phonemes or notes. The loudness of the reverberation during these pauses determines the degree of envelopment. The importance of the reverberation during pauses is increased by the formation of the background sound stream. Where the background stream is easily separated from the foreground stream, envelopment is ~6dB stronger for a given direct to reverberant ratio.

4. For thickly orchestrated music where such pauses are rare, the overall amount of fluctuation determines the degree of envelopment. This is the case when the source is relatively continuous, such as massed strings, or a pink noise test signal.

5. In carefully controlled loudspeaker experiments by Morimoto's group in Kobe, it has been found that reflections that come from behind the listener are more enveloping than reflections that come from the front. They have proposed measuring envelopment through the front/back ratio.

Statement number 5 indicates that the envelopment perception does not solely arise from interaural fluctuations. 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 discriminate front sound from rear sound in two ways. At low frequencies the two can be discriminated through small movements of the listener's head, that cause predictable changes in the ITDs. Unfortunately head movements produce no shifts in ITDs when the sound field is largely diffuse – this localization cue works only with sound that is well localized. At high frequencies it is possible to discriminate sound that comes from behind by notches in the frequency spectrum at about 5kHz. These notches are present when sound comes from greater than about 150 degrees from the front.

We conclude 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.

The statements given above outline the relationship between the physical sound field and the perception of envelopment. Within this framework there are several unavoidable difficulties. For example, it is plain from statements three and four that the perceived degree of envelopment will depend both on the loudness of the sound source, and the type of music being played. This effect is easily observed, and it does not make it easy to develop reliable measures. Statement number 5 implies that the spectrum of the source and the spectrum of the reflections will make a significant difference to how envelopment is perceived.

For frequencies below 1000Hz the perception of envelopment depends on maximizing the fluctuation in the interaural time delay of the sound field at the listener. For music that is relatively transparent to reverberation it is the reverberant component of the perceived sound that creates envelopment. For more continuous music, it is the total fluctuation in the ITD that counts.

For frequencies below 1000Hz in small listening rooms the reverberation time is usually sufficiently short that the room itself is unable to develop significant fluctuations in the ITD in the pauses between syllables or notes. Thus creating the perception of envelopment in the background stream depends on how the room acoustics interact with the material (particularly the reverberation) on the recording. The loudspeakers and the room form a transfer system, which ideally can transmit the enveloping properties of the recording to the listener. Our job is to find a way of measuring the effectiveness of this transfer system, and then to find how the transfer can be optimized.

At frequencies above 1000Hz the front/back ratio will have important implications for our choice of loudspeaker positions.

3. ENVELOPMENT AT HIGH FREQUENCIES

In the frequency band between 300Hz and 3000Hz envelopment is determined by a combination of fluctuations in the IID and the ITD. This fact was discovered independently by the author and Blauert. Blauert notes in [7] and [8] that fluctuations in IID have slightly different perceptual properties than fluctuations in ITD. He concludes that different brain structures are involved in the two perceptions. Envelopment at frequencies below 300Hz is determined almost entirely by fluctuations in the ITD. The important virtue of the framework presented in [44] for the perception of envelopment is that the concept of interaural fluctuations takes the study of envelopment out of the realm of psychoacoustics and into the realm of physics. We can model the mechanisms that create these fluctuations, and we can measure them.

Once we have moved the problem into the realm of physics, we can see that the problem of envelopment has two parts – the nature of the envelopment receiver, and the nature of the sound field surrounding the receiver.

The receiver for envelopment is the human head, the outer and inner ears, and the brain. The soundfield is the combination of direct and reflected energy at the listening position. The head/ear system is the antenna for the perception of envelopment. Like other antenna, this system has directional properties. We can study the directional dependence of the head/ear system for single reflections by convolving test signals with published Head Related Transfer Functions (HRTFs) such as the ones put on the web by Bill Gardner. After the convolution we can detect and measure the interaural fluctuations that result using the software detectors that will be described later in this paper.

This work is yet to be done. However several years ago the author did models of the high frequency behavior of interaural fluctuations using a simple head model. We found that the resulting fluctuations depend on the azimuth of the reflection. For 1000 Hz the optimum angle lies in a cone centered on a line drawn through the listener's head. This cone makes an angle of about 60 degrees from the front of the head. At about 1700Hz the cone includes the standard loudspeaker positions at +- 30 degrees from the

front. This data agrees with subjective experiments into the angular dependence of ASW and spaciousness reported by Ando.

(Ando and Morimoto use the interaural cross correlation (IACC) as one of their measures of the spatial properties of concert hall sound. Morimoto uses the IACC of the first 80ms of a binaural impulse response as a measure. Beranek and Hidaka call this measure (IACCe). For reasons described extensively in [44], the author finds the measure often misleading, particularly in small rooms. However the IACCe does describe the high frequency angular dependence of envelopment quite well.)

Using interaural fluctuations it can also be shown that if there are two sound sources in an anechoic space, and uncorrelated band-limited noise is played through each source, the sense of envelopment will be maximum when both sources are at the optimal angle for the frequency band in question. For frequencies below 700Hz the optimal angle is 90 degrees – the sources must be at the sides of the listener. As the frequency rises, the optimum position moves toward the medial plane.

The angular dependence of envelopment at high frequencies has important consequences for sound reproduction systems. A few years ago the author was playing a stereo tape of applause through a standard two channel system in a relatively dead room. He was very surprised to hear the applause coming from all around him, even though the loudspeakers were clearly in front. Adding a bandpass filter quickly showed that the perception was a simple result of the angular dependence of high frequency envelopment. The 1700Hz band produced a surrounding perception, even though the loudspeakers were at +-30 degrees. Because the applause had significant energy in this band, the result was enveloping. We conclude that stereophonic reproduction works in part because envelopment in a particular frequency band can give an impression of envelopment in all bands, and even with a relatively narrow front loudspeaker pair, envelopment at some frequencies will be high.

Morimoto's front/back ratio becomes important when the reflected sound contains significant energy above 3000Hz. In the author's experience with concert halls and opera houses it is rare that there is significant energy at those

frequencies coming from the rear. Sound from the rear of a hall is often spatially diffused, and has been reflected off many surfaces. Typically each surface takes a toll on the high frequency content. In halls where an electronic enhancement system has been installed it is possible to experiment with the frequency spectrum of the reflected energy, and invariably the sound is better if frequencies above 3000Hz are attenuated. Thus the author is not convinced that the front/back ratio is an important measure for concert hall design.

However Morimoto has shown conclusively that the front/back ratio is important when listening with loudspeakers, and the author can confirm the observation. This issue becomes quite important in the design of surround sound playback systems. Several of the best sound mixers for surround sound have noted that they prefer the rear loudspeakers to be placed +- 150 degrees from the front, and not the more typical +-110 degrees. At 150 degrees these loudspeakers are capable of producing the frequency notches in the HRTFs that indicate rear sound, and speakers at 110 or 120 degrees cannot. The author is convinced that loudspeakers at 150 degrees produce a more exciting surround sound picture. A loud sound effect from 150 degrees from the front is much more exciting than one from 120 degrees.

In addition, one of the most effective uses of a surround sound system for popular music is the presentation of a live performance, where the applause and noise from the audience draws the listener into the experience. Applause has substantial high frequency content, and the front/back ratio becomes important.

However, what is good at high frequencies is not necessarily good at low frequencies. As we will see, for optimal envelopment at low frequencies the surround loudspeakers should be at the sides of the listening area. We might conclude that a single pair of surround loudspeakers is not sufficient – although if we are clever there may be ways around this problem.

4. ENVELOPMENT AT FREQUENCIES BELOW 1000Hz IN SMALL ROOMS – THE DFT

Envelopment below 1000Hz depends on interaural fluctuations, and on the ability of human perception to separate sound into a

foreground and a background stream. Where this separation is possible, it is the spatial properties of the background stream – specifically the amount of interaural fluctuations – that determine the envelopment we perceive. In a small room there is almost never sufficient late reflected energy to contribute to the background perception. The late energy must be supplied by reverberation in the original recording. To predict the degree of envelopment we perceive we must be able to predict the strength of the interaural fluctuations during the reverberant segments of the recording.

Recording engineers know that for best results the reverberation in a two channel recording should be uncorrelated – completely different in the left and right channels. It is the job of the loudspeaker/room system to cause the listener to have adequate interaural fluctuations when this condition occurs. The loudspeaker/room system is acting as a transfer system, transferring the decorrelation in the recording to the listener's ears. We need a measure for how effectively this transfer works.

Finding a test signal for measuring envelopment:

This problem turned out to be much more difficult than expected. Much of the difficulty lies is choosing a test signal that will adequately represent music. As mentioned in the section on concert hall acoustics, the perception of envelopment is highly influenced by the presence of gaps in the music that allow reverberation to be heard. In this study we will assume that such a gap has already occurred – we are trying to model the transfer function of the reverberation within that gap.

The correlation time of musical signals:

However music has another very interesting property that is highly relevant to this study. Music generally consists of notes – segments of sound that have a recognizable pitch. Although the notes may be rich in harmonics, the fundamental frequency is often steady.

If we autocorrelate a musical signal with itself we will see that over a certain length of time the autocorrelation function is strongly non-zero. That is, as long as a note continues there will be a strong fluctuation in the autocorrelation function. Ando has published material that

found that the average length of the non-zero region of the autocorrelation function was related to the type of music being played, with modern serial music having a short correlation time, and romantic symphonies having a long correlation time. Such a result would be expected from the average length of the notes in the various musical styles, as well as the presence in many types of modern music of percussive sounds with no well established pitch. Ando discovered that as the correlation length of a musical example increased, the ideal reverberation time of a performance space also increased. We will see that this result can be predicted from the properties of the DFT.

When we started working for a measure of envelopment in rooms we used band limited pink noise as a test signal. Reference [44] contains several experiments and observations of the spatial properties of band limited pink noise. The correlation length of band limited noise depends on the bandwidth chosen (and to some degree on the filter type.) For our first work we choose to use a bandwidth approximately equal to the width of a critical band in the human ear. The results showed that this was not a good choice. Although the results appeared to be accurately reflect the envelopment of noise signals in small rooms, they did not predict the envelopment of musical signals.

As an experiment, we measured the frequency width and phase fluctuation in the reverberation from musical notes of various lengths in a real concert hall (Boston Symphony Hall.) The resulting reverberation had high coherence. We were able to use the reverberation in Boston as a test signal in calculating the DFT, and we found that to achieve a similar results with noise we had to use a filter bandwidth of 1-2Hz. A disadvantage of such a narrow frequency is that the DFT in a small room is often highly frequency dependent, and to find an average value of envelopment one must separately calculate the DFT at many different frequencies. Unfortunately there appears to be no short-cut. We must calculate our measure using a narrow band test signal.

The reason we need a narrow band test signal is obvious in hindsight. Small rooms have relatively short reverberation times, often less than 0.5 seconds. The time constant of such a space (the time it takes the sound to decay by $1/e$) is the reverberation time divided by 7, or 70 milliseconds. If the music (or the reverberation from the music) has a correlation length that is significantly longer than this time constant the room does not generate interaural fluctuations directly. This means that a single driver in the room will not produce interaural fluctuations at the ears of a listener. The room can however detract from the transfer of fluctuations from multiple drivers, and this is the effect we wish to measure.

A detector for envelopment at the listening position

A detector for envelopment is also a difficult problem. We had hoped to be able to use a measure developed for concert halls, the interaural difference, or IAD. To make a long story short – it doesn't work. We were in fact unable to find a proxy measure that could even duplicate the perceived envelopment in an anechoic space, let alone the envelopment in a reflective space. For example, it is obvious that a single sound source in an anechoic space is incapable of producing envelopment. Our measure must show that this is the case. By similar reasoning, two sources driven by highly correlated material should also give near-zero envelopment in an anechoic space, and the measure should reflect this. The IAD fails on both counts.

(As luck would have it, the IACC measure does distinguish between echoic and anechoic spaces. However – again making a long story short – the IACC fails to predict observed results at frequencies below 200Hz.)

We found it was necessary to go back to first principles. The hypothesis in [44] predicts that the perception of envelopment at low frequencies depends on fluctuations in the ITD. There is no shortcut – to measure envelopment we must convolve the binaural impulse response from each sound source with an independent signal, and then measure the fluctuations in the ITD that result. In practice this is difficult. Evolution has had millions of years to perfect methods of extracting ITD information from noisy signals at the eardrums. We had to develop an algorithm in software that make this extraction. Our current measure still needs some work, but it seems to give useful results. Using this algorithm and a narrow band noise signal as a test probe we developed our measure for

envelopment in small rooms. We call it the DFT, or diffuse field transfer function.

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. 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 from independent filtered noise signals. Various frequencies and bandwidths can be tried, depending on the correlation time of the musical signal of interest.

4. Convolve each binaural impulse response with a different band filtered noise 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. Measure the DFT as a function of the receiver position in the room under test.

The most difficult part of this process is 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.)

The current code uses noise signals with 10000 samples each. This is about 3.7 seconds at the sample rate of 2756Hz. We are interested in random fluctuations in a signal in the 3Hz to 17Hz band – with typical maximum energy around 5Hz. Needless to say, the accuracy of such a measurement given a 3.5 second duration is not high. The DFT we get from this measurement has a semi-random variation of about 2dB, with occasional values as many as 3dB different from what was expected. Thus the DFT we present here is somewhat noisy. We could improve the accuracy by increasing the length of the noise signals – gaining the usual improvement by a factor of 1.4 for each doubling of the signal length.

5. CALIBTATION OF THE DFT MODEL

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

5a. Calibration of the DFT for noise signals

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. 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. We need to choose the bandwidth.

In an effort to calibrate the DFT detector, 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 [44] 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 depends on the delay of the reflection. 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, with relatively low sound pressure. A delay of 13ms produces a nearly monaural impression, with little or no envelopment at all. As the frequency 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 combination of delays chosen. This observation has pronounced implications for concert hall design. 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!

5b. 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 1-2Hz bandwidth. We will show some results based on narrow band noise.

5c. 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 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. Which situation is optimal? 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 [44]. 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. Our conclusion is that a diffuse field is not optimal at low frequencies. We can calibrate our DFT detector by measuring its value when it is exactly between two noise sources in a simulated anechoic space. The value we get is frequency dependent, but for the 63Hz octave band it is about 0.24ms. In the following paper we will use this measure to study the envelopment at low frequencies in listening rooms.

In conclusion we believe the DFT is a useful model for how human hearing perceives envelopment when the sound source is band limited pink noise. It is not clear at this time whether significantly different results would be obtained with other source signals.

6. ITD AND EXTERNALIZATION

The "in the head" perception also arises from the behavior of the ITD. When we are outdoors or in a large space, a sound source at the side produces ITDs of about 0.75ms at the listener's head. We easily perceive from this ITD that the source is external and at the side. When the sound source is in the medial plane (directly in front, overhead, or behind) the ITD is much lower. We perceive the sound as centered. We can usually tell if the source is directly in front or behind us. When this is possible, we perceive the sound as external.

It is well known that our ability to discriminate between rear sound and frontal sound at low frequencies depends on our ability to move the head. In natural hearing small movements of the head produce predictable changes in the ITD. Small head movements localize and externalize a source in the medial plane.

As we will see, in a small room the situation is quite different. For medial sources (or for a phantom image of a center source) a listener will experience a low ITD regardless of how the head is rotated. It is not possible to determine if the sound is from the front or the back.   The perception becomes internal and artificial – a perception peculiar to recorded music.

It is not necessary that the sound source be localized to a particular direction for externalization to occur. If the ITD changes randomly – not in synchrony with head movements – externalization is still successful. The resulting sound is perceived as external and enveloping. Our job in this paper is to find a measure that describes the degree to which low frequencies are externalized in a particular sound field – ideally from a modeled or measured binaural impulse response. We can then use the measure to find how externalization can be optimized.

7. BINAURAL IMPULSE RESPONSE AND ITD

The Fourrier transform of the impulse response yields both pressure and phase as a function of frequency. To study the case where there is more than one driver active at the same time we find it convenient to first find the impulse response from each source separately. The phase of the transform from each source is adjusted if

necessary, and then the transforms are added to find the total pressure and phase. The transform of a binaural impulse response is found by separately calculating the transforms for each ear. At low frequencies we assume the head can be approximated by two omnidirectional receivers, separated laterally by 25cm. We do not attempt to model diffraction around the head, which is minimal at the frequencies of interest.

The ITD as a function of frequency can be found by finding the phase difference between the transforms of the left and right binaural responses. The phase difference is then converted to a time difference by dividing by frequency.

It is useful before going further to examine the perceptual meaning of the impulse response and its Fourrier transform. Both are mathematical concepts. It is not obvious that either should have any relationship to what we hear with music. The impulse response is the response of the room to a pistol shot or small explosion. These are fortunately rare in everyday life. Nonetheless there is a considerable body of literature that attempts to relate musical perceptions to direct features of the impulse response.

The transform of the impulse response describes the steady state response of the room to single frequency sinusoids. These are plausibly musical. In a small rooms after the time it takes for the sound to travel across the room a few times the pressure is nearly at the steady state value. Most musical notes are longer than this, and often the attacks of low frequency instruments are slow enough that the room can be modeled by the steady state condition. However each frequency in the transform represents only a single musical note. We would like to know the response of the room over a range of notes. We need a measure for average pressure and average phase, where the average is over frequency.

8. NORMALIZED AVERAGE PRESSURE (NAP), AND AVERAGE ITD (AITD)

To find the average pressure we could simply sum the pressure squared over the range of interest to find the total pressure in the frequency band. The summation assumes that each frequency was present in the steady state – which is not particularly likely – but this measure of

average pressure is useful nevertheless. To make the measure closer to what we hear we weight the sum by -3dB per octave, so each 1/3 octave band gives an equal contribution to the sum.

We start with a binaural impulse response, either measured or calculated. That is, a separate impulse response for the left ear and the right ear of a dummy head.

$IL(t)$ = impulse response for the left ear
$IR(t)$ = impulse response for the right ear
Sum = sum of the frequency bins over the frequencies of interest

2. $FFTL(f) = fft(IL(t))$
3. $FFTR(f) = fft(IR(t))$

7. Average Pressure = sqrt(sum((magnitude $FFTL(f)^2$)/f)

This average pressure contains a frequency factor, which we would like to eliminate when we plot pressure as a function of frequency. We can define a normalized average pressure:

8. Normalized Average Pressure = Average Pressure / (sum (1/sqrt(f))

For low frequencies we assume that the magnitude of FFTL is approximately equal to the magnitude of FFTR.

We suggested above that the ITD at a particular frequency could be found by comparing the phase of the left and right parts of a binaural response. The problem is that this phase difference is independent of the pressure at that frequency. It is possible – in fact it is nearly always the case – that the frequencies where the phase difference is the largest are the ones for which the pressure is the lowest. We are not likely to take much notice of a large ITD if the pressure is low. Unlike a lateral figure of eight microphone, which gives a maximum output at the nulls of a lateral standing wave, we cannot determine the ITD of a sound we cannot hear. If we stand at such a null, we hear nothing.

It would be possible to develop a model that determines ITD the same way the ear does, by first separating the sound into critical bands, and then finding the timing of zero crossings for each ear. See [44.] As we will see later, such a model is quite expensive computationally. Since we

want to calculate the AITD at many frequencies and at many points in a room, there is a strong incentive to develop an efficient model.

To make an average ITD that is closer to what we actually hear we weight the ITD with the pressure. ITDs with low pressure are given low weighting, and ITDs with high pressure are weighted more strongly. We then normalize the resulting sum by dividing by the average pressure over the same frequency range.

Given this description the method of finding the AITD follows directly:

5. Phase_angle(f) = angle(FFTL)-angle(FFTR)
6. ITD(f) = Phase_angle(f) /(2*pi*f)

To find the AITD we multiply the absolute value of ITD(f) by the absolute value of the pressure, and sum over frequency. We also apply a –3dB per octave weighting to give equal weight to all bands. The sum is then normalized by dividing by the weighted sum of the absolute pressure over the same frequency range.

6. AITD = sum(magnitude(FFTL(f))*abs(ITD(f))/(sqrt(f))/(average pressure)

It is also possible to define a NET ITD (NITD) – one which preserves the sign of the ITD in the sum. The NITD is always less than the AITD, but it indicates our ability to determine the direction of a sound source, not just its externalization.

7. NITD = sum(magnitude(FFTL(f))*ITD(f)/(sqrt(f))/(average pressure)

Note that for simplicity we use only the magnitude of FFTL(f) to represent the pressure at both ears. We could average the magnitudes of FFTL and FFTR, but there seems to be no reason to do so in practice.

9. AITD AND NITD – WHAT DO THEY MEAN?

The NITD is a measure that expresses the apparent direction of a sound source in a given frequency band. Not every frequency within the band will localize to the same degree. In fact, due to standing waves some frequencies may localize to the wrong direction. The NITD

averages out these differences, giving us a lower value than we would get using individual frequencies. It does reflect the average localization of a narrow band noise source. We are interested in externalization of the sound, and externalization can take place even in the absence of correct localization. Thus for externalization the average absolute value of the ITD, the AITD is a better measure.

At low frequencies in a free field – with no reflections – AITD and NITD are equal, and independent of frequency. Their value depends on the angle between the listener and the source, and can be easily calculated.

If source_angle is the angle between the source and a line drawn between the listener's ears (listener facing forward):

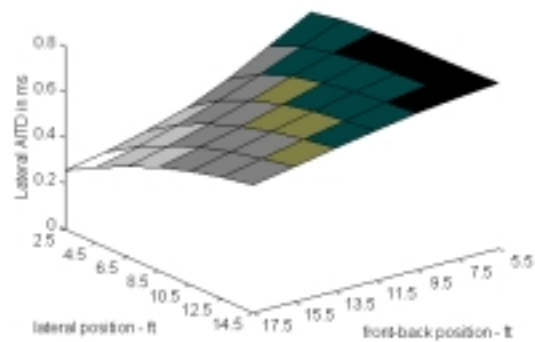Lateral AITD = 0.75ms*cos(source_angle)

Figure 1: The Lateral AITD in the center of a 17'x23' anechoic space with a driver in the upper left corner (at .1',.1'). Where the value is high, the sound source will be localized to the side of the listener.

We can also define a Medial AITD – by simply rotating the listener's head by 90 degrees in azimuth and calculating the AITD again. In a free field – assuming the source angle has not changed:
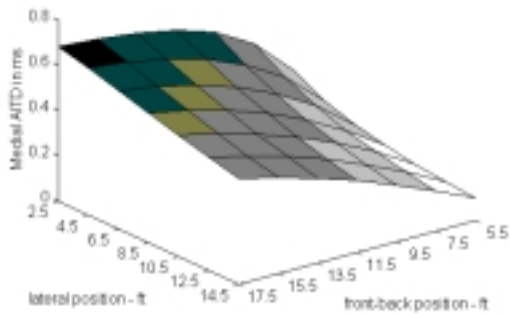
Figure 2: The Medial AITD in the center of a 17'x23' anechoic space with a driver in the upper left corner at .1', .1' Note that where the lateral AITD is low, the medial AITD is high.

It makes sense to define a Total AITD, which is simply the RMS sum of the lateral and the medial AITD. Since cos^2 + sin^2 is one, in a free field. Total AITD = 0.75ms.
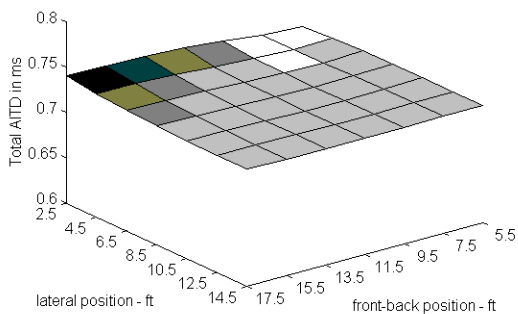

Figure 3: The Total AITD (the RMS sum of the lateral and medial AITDs) in the same space as figures 2 and 3. Note that the driver can be accurately localized everywhere, as we would expect in an anechoic environment. The differences from 0.75ms are due to sampling errors in the model.

The total AITD is a measure of how easily a sound source can be localized. In a free field it has the constant value of 0.75, which means the source can be localized to its true direction with full accuracy. In a reflective room the total AITD is almost always less than this value, as standing waves tend to reduce the ITD at the listener. Where the total AITD is significantly less than 0.75 the sound source will be "in the head".

At present it is not clear how close the AITD must be to 0.75ms for complete externalization. In our informal experience a value of 0.3ms or lower will almost always sound internal. Values of 0.4ms or so can sometimes be internal and sometimes be external. Clarifying this question will require further experiments. However the AITD is useful as a measure. We want to find strategies of speaker placement and recording technique that bring the AITD as close as possible to the free field value. Since in most cases the listener does not move the head very much, at low frequencies we are mostly interested in the lateral AITD.

Models that use these measures to study the properties of listening rooms will be presented in the next paper.

REFERENCES

1. Ando, Y. and Singh, P.K. and Kurihara, Y., Subjective diffuseness of sound field as a function of the horizontal reflection angle to listeners. Preprint received by the author from Dr. Ando
2. Ando, Y. and Kurihara, K., Nonlinear response in evaluating the subjective diffuseness of sound fields. J. Acoust. Soc. Am. 80 [1986], 3 pp 833-836
3. Barron, M., Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure. J. Sound and Vibration 77(2) [1981] pp 211, 232
4. Beranek, L., Music, Acoustics and Architecture. John Wiley, 1962
5. Beranek, L., Concert Hall Acoustics - 1992. J. Acoust. Soc. Am. 92, [1992]
6. Beranek, L., Concert and Opera Halls How They Sound. Acoustical Society of America, 1996
7. Blauert, J., Zur Tragheit des Richtungshorens bei Laufzeit- und Intensitatsstereophonie. Acustica 23 [1970] p287-293
8. Blauert, J., On the Lag of Lateralization Caused by Interaural Time and Intensity Differences. Audiology 11 [1972] pp265-270
9. Blauert, J., Raumliches Horen. S. Hirzel Verlag, Stuttgart, 1974
10. Blauert, J., Spatial Hearing. MIT Press, Cambridge MA 1983
11. Bradley, J.S., Contemporary Approaches to Evaluation of Auditorium

Acoustics. Proc. 8th AES Conference Wash. DC May 1990 pp 59-69

12.   Bradley, J.S., Contemporary approaches to evaluating Auditorium Acoustics. Proceedings of the Sabine Conference, MIT June 1994.

13.   Bradley, J.S. and Soulodre G.A., Spaciousness judgments of binaurally reproduced sound fields. Ibid. p 1-13

14.   Bradley, J.S., Comparisons of IACC and LF Measurements in Halls. 125th meeting of the Acoustical Society of America, Ottawa, Canada, May 1993

15.   Bradley, J.S., Pilot Study of Simulated Spaciousness. Meeting of the Acoustical Society of America, May 1993

16.   Bradley, J.S. and Souloudre, G.A., Objective measures of Listener Envelopment. J. Acoust Soc. Am. 98 [1995] pp 2590-2597

17.   Bradley, J.S. and Souloudre, G.A., Listener envelopment: An essential part of good concert hall acoustics. JASA 99 [1996] p 22

18.   Gardner, W. and Griesinger, D., Reverberation Level Matching Experiments' Proceedings of the Sabine Conference, MIT June 1994. p263

19.   Gold, M.A., Subjective evaluation of spatial impression: the importance of lateralization. ibid. p97

20.   Griesinger, D. Measures of Spatial Impression and Reverberance based on the Physiology of Human Hearing. Proceedings of the 11th International AES Conference May 1992 p114 - 145

21.   Griesinger, D., IALF - Binaural Measures of Spatial Impression and Running Reverberance. Presented at the 92nd Convention of the AES March 1992 Preprint #3292

22.   Griesinger, D., Room Impression, Reverberance, and Warmth in Rooms and Halls. Presented at the 93rd Audio Eng. Soc. convention in San Francisco, Nov. 1992. AES preprint #3383

23.   Griesinger, D., Progress in electronically variable acoustics. Proceedings of the Sabine Conference, MIT June 1994.

24.   Griesinger, D., Subjective loudness of running reverberation in halls and stages. Proceedings of the Sabine Conference, MIT June 1994. p89

25.   Griesinger, D., Quantifying Musical Acoustics through Audibility. Knudsen Memorial Lecture, Denver ASA meeting, 1993.

26.   Griesinger, D., Further investigation into the loudness of running reverberation. Proceedings of the Institute of Acoustics (UK) conference, Feb 10-12 1995.

27.   Griesinger, D., How loud is my reverberation. Audio Engineering Conference, Paris, March 1995. Preprint # 3943

28.   Griesinger, D., Design and performance of multichannel time variant reverberation enhancement systems. Proceedings of the Active 95 Conference, Newport Beach CA, June 1995

29.   Griesinger, D., 'Optimum reverberant level in halls' proceedings of the International Congress on Acoustics, Trondheim, Norway June 1995

30.   Hartman, W., Localization of a sound source in a room. Proceedings of the 8th international conference of the Audio Engineering Society, May 3-6 1990

31.   Hidaka, T. and Beranek, L. and Okano, T., Interaural cross correlation (IACC), lateral fraction (LF), and sound energy level (G) as partial measures of acoustical quality in concert halls. J. Acoust Soc. Am. 98 [1995] pp 988-1007

32.   Jullien, J.P. and Kahle, E. and Winsberg, S. and Warusfel, O., Some Results on the Objective Characterization of Room Acoustical Quality in Both Laboratory and Real Environments Proc. Inst. of Acoustics Vol. 14 part 2 (1992). presented at the Institute of Acoustics conference in Birmingham, England, May 1992

33.   Kahle, E., Validation d'un modele objectif de la perception de la qualite acoustique dans un ensemble de salles de concerts et d'operas' Doctorate Thesis, IRCAM June 1995

34.   deV. Keet, W., The Influence of Early Lateral Reflections on the Spatial Impression. The 6th International Congress on Acoustics, Tokyo, Japan, Aug 21-28 1968. pp E53 to E56

35.   Kreiger, A., Nachhallzeitverlangerung in der Deutschen Staatsoper Berlin. Tonmeister Informationen (TMI) Heft 3/4 Marz April 1996.

36.   Morimoto, M. and Maekawa, Z., Effects of Low Frequency Components on Auditory Spaciousness Acustica 66 [1988] pp 190-196

37.   Morimoto, M. and Posselt, C., Contribution of Reverberation to Auditory Spaciousness in Concert Halls. J. Acoust. Soc. Jpn (E)10 [1989] 2

38.   Morimoto, M. and Maekawa, Z., Auditory Spaciousness and Envelopment. 13th ICA, Yugoslavia 1989

39.   Morimoto, M., The relation between spatial and cross-correlation measures 15th ICA Norway 1995

40.   Schroeder, M.R. and Gottlieb, D. and Siebrasse, K.F., Comparative study of European

concert halls: Correlation of subjective preference with geometric and acoustic parameters. J. Acoust. Soc. Am. 56 [1974] pp 1195ff

41.   Schubert, P., Die Wahrnehmbarkeit von Einzelruckwurfen bei Musik. PhD thesis, TU Dresden 1967

42.   Schultz, T.J., Acoustics of the Concert Hall" IEEE Spectrum (June 1965) pp 60-62

43.   Griesinger, D. "Spaciousness and Localization in Listening Rooms and their Effects on the Recording Technique" JAES v34 #4 p255-268 April, 1986

44.   Griesinger, D. "The Psychoacoustics of Apparent Source Width, Spaciousness and Envelopment in Performance Spaces" Acta Acustica Vol. 83 (1997) 721-731