

Spaciousness and envelopment in musical acoustics

David Griesinger
Lexicon 100 Beaver Street
Waltham, MA 02154

Abstract:

Conventional wisdom holds that spaciousness and envelopment are caused by lateral sound energy in rooms, and that it is the early arriving lateral energy which is most responsible. However small rooms often have many early lateral reflections, but by common definition small rooms are not spacious. This paper (briefly) describes a series of experiments into the perception of spaciousness and envelopment. The perceptions are found to be related most commonly to the lateral (diffuse) energy in halls at least 50ms after the ends of notes (the background reverberation) and less often but importantly to the properties of the sound field as the notes are held. Experiments with orchestral music at high reverberant level indicate that it is the very late >300ms reflected energy which is most responsible for spaciousness. A measure for spaciousness - Lateral Early Decay Time (LEDT) is suggested, and results of this measure in several halls are given. A good match between the new measure and subjective impressions of the halls is found.

Introduction

When we started studying hall acoustics about 10 years ago we assumed that spaciousness, envelopment, and reverberance were all essentially the same thing. However the literature on the subject identified spaciousness and envelopment with lateral reflections, particularly reflections arriving before the first 80ms. This literature was supported by simple laboratory experiments by ourselves and many others, testing the effects of single reflections at different delays and angles. Yet several mysteries remained. First, a series of measurements into an electronically enhanced hall showed there was little change in the early reflected energy when the enhancement was switched on, even though the change in subjective impression was dramatic. Second, by common observation small rooms are not spacious, even though they can have very strong early lateral reflections. Third, in a concert it is the middles and ends of notes which are spacious, never the attacks. Fourth, the spaciousness or hall impression in a large hall with speech or solo music seemed independent of the source distance, even though the impulse response changed a great deal. Something was fishy with the conventional view of spaciousness.

One difficulty turned out to be a matter of definition. "Spacious" in the American Heritage Dictionary is defined as "1. providing or having much space or room," and "2. vast in space or scope, 'a spacious view.'" In English a concert hall can be spacious, the soundfield of an oboe can be spacious, but an oboe cannot. In spite of this the majority of research into spaciousness equates the term with the apparent source width (ASW). This supposed equivalence has caused a great deal of trouble - and will not be followed here. Where we use the term spaciousness or spaciousness we give it the common meaning - that the sound field gives the impression of a large and enveloping space. Thus spaciousness and envelopment are similar impressions for the purpose of this paper. We want to distinguish between sounds which have a large source width, sounds which have an acoustic impression associated with them which is largely frontal, and sounds where an acoustic impression surrounds the listener. In our view only sounds of the last type are spacious. We will use Barron's term "spatial impression" (SI) to refer to an acoustic impression, whether it is spacious or not.

We started a continuing series of experiments into the perception of spaciousness and how this perception can be increased through electronic enhancement or careful hall design. We decided to it was necessary to understand the physics used by the ear and brain to detect lateral direction, and how the brain interprets fluctuations in the apparent direction due to reflected energy. Although this work was exciting and seemed original, it did not explain the lack of spaciousness in small rooms, and a satisfactory measure seemed elusive.

To arrive at a more satisfactory solution we found we had to combine the work on lateral localization with two other concepts. The first of these involves the properties of musical sounds. Musical sounds very seldom resemble the impulse response we typically measure. In fact, the impulse response is misleading when we try to understand spaciousness and ASW. The ear must cope with sound events as they arrive, and these are ordinarily not impulses. With speech the sounds are phonemes, with music they are notes. Such events have onsets, pitch, timbre, loudness, and endings. When sounds are strung together to form music or speech there are spaces between notes and phonemes. ASW by definition is a property of horizontal localization, and localization depends on the attacks of sounds. To understand ASW we must look at the properties of sound onsets and how the ear uses them to determine localization. But ASW and spatial impression are different. To understand SI we must look at how acoustics alters the pressure at our eardrums during the spaces between sounds, and also during the middle and ends of sounds. We must look at how the brain organizes sound events into related streams. Our goal is to find how the ear interprets the alterations, particularly how the ear interprets fluctuations in the interaural intensity difference (IID) and the interaural intensity difference (ITD.) The impulse response by itself is only a tool to this end.

The second key concept is the sonic background. We found in (18) that the perception of reverberance in music and speech arises from a brain function involved with the perception of distance. The brain processes incoming sound into a foreground stream - the part which holds the information content of the signal - and a background stream. The background stream can contain many different parts - in a forest for example it might contain wind or water noise, as well as general chirps of animals or airplanes. It is perceived by our brains as continuous, and the loudness of foreground sound is compared to the background as one measure of the source distance. In a reverberant environment the background is the reverberation. When a musical note ends, there is a time delay between the ending of the note and the start of the loudness integration for background. The time depends on the individual. It is at least 50ms long, but the sensitivity to the background increases as delay increases, up to a maximum at about 160ms. Where the musical material which creates the background is not highly masking the loudness of the background is absolute - it does not depend on the loudness of the foreground. This explains the observation that hall impression on speech or solo music can seem independent of the distance to the source. With orchestral music, the time delay before reverberance is particularly audible after the end of a note is further increased by several factors, including the tendency of music to mask its own reverberation. We have found that for most music, 350ms is an appropriate time delay.

The final piece of the puzzle was to realize that envelopment and spaciousness are in fact closely related to reverberance. If you can't hear the reverberance you do not perceive the hall as spacious. However, it is the lateral or diffuse component of the reverberation which matters. Reverberation from the front or overhead is not perceived as spacious. From this data a new measure both for reverberance and for spaciousness is suggested - the lateral EDT, or LEDT

Laboratory experiments with multiple reflections

For several years we have been investigating the perception of reflections and reverberation using headphones. The apparatus consists of sound sources of pink noise, pink spectrum clicks, speech, and music. These sources pass through a 48dB/octave variable band pass filter, and then to a commercial digital signal processor. The processor includes a program "Binaural Simulator" which adds reflections and reverberation to the incoming sound.

In these experiments the incoming sound is monaural. It is directed equally to both ears to form the direct sound image. Lateral reflections are simulated by adjustable attenuators and delays. The reflections are lateralized through a combination of interaural delay and a low pass filter. The output of one attenuator and delay is directed to one earphone, and an additional delay and a single pole low pass filter is added before the reflection is connected to the other ear. An additional delay of 750us and a low pass of 4kHz is effective in lateralizing the reflection. Using this simulator the delay and level of up to six reflections, 3 on each side of the listener, can be adjusted in real time. The simulator also allows reverberation with many different time profiles to be added to the sound (26,27).

Lateral reflections with band limited pink noise

We have used this simulator many times over the last few years to experiment with both reflections and reverberation. When noise is used as a source, and the bandwidth is either broadband or includes only frequencies above 300Hz, the results closely follow previous work by Keet, Barron, Schubert, etc. Lateral reflections with delays greater than about 10ms produce a sense of surround which is distinct from the central image. The threshold for the surround impression is about 20dB below the direct level, and when multiple reflections are used it is the total energy which determines both threshold and loudness. Thus when you combine several reflections each of which is below audibility the result may be audible if the total energy is greater than -20dB. With noise there is very little change in the spatial properties of the sound as the reflection delay is increased beyond 10ms. The impression is always of a sharp central source in the presence of a surround field. The impression is constant until a single reflection is within 3dB of the direct energy, at which point the apparent width increases and the position of the direct sound shifts toward the reflection. ASW is a poor description of the effects of reflected energy on a noise signal above 300Hz. The impression is of a sharp sound source in the presence of a surround. When noise is limited to frequencies below 300Hz the results are similar, but more delay is required for the surround effect.

Discussion of noise experiments

It is our belief that the surround impression depends on a fluctuation in the interaural intensity difference (IID) and the interaural time difference (ITD). The fluctuation is the result of interference between the direct sound and the reflected sound at the two ears. (21) expresses these fluctuations as a fluctuations in the “pseudoangle” the position a source would have from the measured IID and ITD if the ear were able to follow rapid movement. In the case of the noise experiments described above the fluctuation in both the IID and the ITD pseudoangles are small until the reflected energy is within 3dB of the direct energy. As a result, the apparent width of the source stays sharp, broadening only when the total reflected energy becomes large.

With this understanding of the origin of spaciousness it becomes possible to predict how different types of continuous signals and source angles will be perceived. We must only consider how the reflected sound interferes with the direct sound to cause fluctuations in the ITD and the IID. A simple argument predicts the observed dependence of SI with source angle, and the observed dependence of SI on vibrato and tremolo in musical tones. See (21). The observation that with pseudorandom noise spaciousness requires a longer time delay at low frequencies than at middle and high frequencies is provocative and not entirely understood at this time. A similar increase in the needed delay for spaciousness at low frequencies was noted by Schultz (42).

Noise signals are perceptually continuous - they form a single (although lengthy) event. The surround can be perceived separately - it can have a different timbre for example, and it can have different spatial properties. However the surround and the direct signal are bound together. For noise from 300Hz to 2000Hz, if the level is varied the spatial properties remain the same. ASW (if any) is independent of the loudness of the sound, and so is the impression of the surround. However when the noise includes significant low frequencies, as the level increases the spaciousness increases, sometimes dramatically. Hidaka and Beranek (30) explain this effect through the well known increase in threshold of audibility as frequency decreases. If the reflected energy is below audibility it will not be heard, and it will not produce spaciousness. The surround associated with noise signals is the spatial impression we refer to as “continuous spatial impression”, or CSI. It can be enveloping when it is strong, but is not always “spacious”. It may be relatively independent of the loudness of the source, particularly at frequencies > 300Hz. CSI is depicted visually in figure 1b.

Experiments with band limited music and speech as sound sources.

An advantage of an easily adjusted experimental apparatus is that sometimes you find something unexpected while just playing around. We made such a discovery while trying to determine the reflection

delays and amplitudes needed to broaden the sound image above 300Hz. We used speech as a source. It was found that reflections with delays less than 20ms can broaden the source - but the degree of the broadening depended on the rise time of the particular phoneme. See figure 2. In general if a strong lateral reflection arrived during the rise time of the phoneme the image was broadened, otherwise it was unaffected. This was the expected result - we believe ASW depends on the repeatability of the ITD during event onsets, both across third octave bands and between different elements of a foreground stream. There is considerable neurological evidence for the importance of the rise time of sound events. It is possible to find nerves which fire only on the rise of events, with a firing frequency which depends on the rise time. That there would be strong evolutionary pressure to use the onset of a sound event to determine localization seems obvious. The ITD of the onset is unlikely to be contaminated by reflections.

However, while moving the delays around we noticed that there was a dramatic change in the surround impression when one or more of the delays exceeded 50ms. With speech a single reflection with less than 50ms of delay does not produce a sense of surround. Although a spatial impression is produced the apparent placement of this impression is frontal - closely associated with the direct sound but not changing the apparent width of the source very much. As the delay time is increased an abrupt change in the impression occurs at about 50ms. Single reflections with a greater delay begin to become separable from the direct sound. With a single strong reflection one is aware of a separate sound event, and can sometimes localize it independently. With multiple reflections arranged semi uniformly from 10ms to more than 55ms discrete echoes are not audible, and the impression becomes that of a sharp source in the presence of a fully enveloping surround. When all the reflections are less than 50ms the spatial impression is frontal. When the sound source has slow note onsets - such as legato strings - this reflection pattern also produces considerable source broadening (ASW). However the spatial impression remains frontal.

Table 1: Two sets of reflections: set A is frontal, set B is spacious.

Set A:	1L	2L	3L	4R	5R	6R
levels	-10dB	-10dB	-10dB	-10dB	-10dB	-10dB
delays	11.39	24.71	41.26	5.49	22.79	47.92
Set B:						
levels	-10dB	-10dB	-10dB	-10dB	-10dB	-10dB
delays	11.39	24.71	54.99	5.49	22.79	59.32

This observation was tested with 6 listeners, 5 of whom heard the effect strongly. The sound source was speech limited to 300Hz to 2000Hz. One subject (af) said “Set A sounds like a voice in the presence of a public address system which is located in the front of a room. Set B sounds like you are in an airport, and the speakers are all around you.” The subjects were also asked to start with set A and increase the delay of the last reflection until the sound shifted from frontal to surrounding. The five subjects who could hear the effect selected delays between 52 and 60ms. The subjects were then retested with chamber music as a sound source. All found hearing the difference between the two sets more difficult, but claimed the difference did persist. All listeners reported that with noise as a source, set A and B were spatially identical. One of the subjects then increased the average delay of the reflections, and found that the same surround impression could be generated with ~6dB less total reflected energy when the reflections varied in delay from 35 to 100ms, when compared to set B. With set A this subject said the “frontal” localization of the SI was preserved until the direct sound was about 10dB less in energy than the total reflected energy. Below this level ASW was very high. The sound seemed to come from opposite sides of the head, but the effect was not spacious.

Discussion of speech experiments

To understand the speech experiments we must consider the ability of the hearing mechanism to sort incoming sounds into groups. Sorting requires a certain minimum separation in both time and frequency. The original meaning of the Haas effect is related to there being a minimum time between two sounds before the ear can ascribe them to separate sound events. If two sounds arrive within 50ms of each other they are almost always combined into one event by the ear. Since the direct sound is strongly localized, and

the early reflected energy is bound to it perceptually, the surround effect which would otherwise be created by the fluctuating IID and ITD during the first 50ms is inhibited. A spatial effect is perceived, but it is frontal and not particularly strong. This is the spatial impression we called EHS in an earlier paper (25), and we refer to it as ESI in this paper. It is depicted in figure 1c. This is the spatial impression of small rooms.

Small rooms generate most of their reflected energy within the first 50ms. Thus the spaciousness they generate is ESI, and not the spaciousness associated with the sonic background. A very reverberant small room does produce a sonic background, but the impression is not very spacious. ESI has its uses. Both Beranek (4,5) and the author (25) note that without some energy arriving before 50ms the sound in a hall is too sharp and direct. The author finds there is an optimum level for the very early energy of about -6dB relative to the direct sound. This is also true in recorded music. When a microphone is close to a performer a “too focused” sound can result. Leakage to additional (stereo) microphones at about a -6dB level, or deliberately arranging lateral reflections can eliminate the problem.

The sonic background

In previous work we (18) studied the loudness of reverberation, and discovered that the results could be explained by postulating that the brain contains a “background detector” a neural circuit which keeps track of the loudness of the background between sound events. Such a detector is useful in determining the distance of a sound source. That work indicated that a certain minimum time must elapse between the end of a foreground event and the beginning of the loudness integration which determines the background level. In our experiments with speech and music at low reverberant level this time depended on the individual, but averaged about 160ms. These experiments strongly indicate that what we perceive as spaciousness depends on the background detection process. However, at least with anechoic speech strong reflections are able to be counted as background with as little as 55ms of delay. Reflections arriving earlier are counted as part of the direct sound, and their directional content is inhibited. As delay increases the ease with which the reflections can contribute to the background increases, and their directional content contributes to background spatial impression, or BSI. BSI is depicted in figure 1a.

Reflections which contribute to ESI bind with the foreground sound to form a new sound. Typically both timbre and decay shape are audibly changed. A consequence of this binding is that the effect of the reflections is independent of the loudness of the sound source. Loud playing and soft playing are identically affected. Beranek’s “intimacy” is not level dependent. The opposite is true of BSI. Subjects in (18) could match the loudness of the background impression with better than a 2dB repeatability. This ability to match the background loudness was absolute - it did not depend on the loudness of the foreground events. With the tone burst experiments described in (24) and later in this paper, the background impression is that of a steady tone, the loudness of which can be very easily matched to another background or to a test tone. This is not true of the foreground events, which are of indeterminate loudness. In a headphone experiment with speech where the level of the reflections is held constant while the level of the direct sound is varied the surround impression is independent of the direct level. The spaciousness from BSI thus does not depend on the direct to reverberant ratio, but on the absolute loudness of the background. This observation is not true when there is considerable masking, as we will see later.

It should be emphasized that the background impression is a property of a sound stream - a series of related sound events strung together in time. Although both reverberation and spaciousness can be perceived from a single note or phoneme, the impression is much stronger with a string of notes or phonemes. Under these conditions the background is perceived as a constant sound, separate from the foreground stream.

Ends of sound events

It is not possible to separate input sounds into a foreground and a background stream without knowing that a foreground event has ended. However here we are on more difficult ground. There is no neurological evidence for an “end of event” detector, and yet some such detector must exist, or the background would be

inseparable. At some level the brain must make the decision that a foreground event has ended. Just how and when can be determined by experiment.

We built a digital amplitude modulator at the input to the binaural synthesis program to probe the separation process. With the modulator tone bursts with controlled rise time, hold time, and fall time could be generated. The fall could also be broken into two sections, each with a separate fall time. figures 3 - 7 show some of the waveforms which could be generated.

The modulation of figure 3 has a fast rise time - less than 10ms total. This is similar to some phonemes. With this modulation sonic images are sharp whenever reflections are $> 10\text{ms}$ delayed. The burst does not decay to zero - there is a constant section 18dB lower than the peak. When you listen to an organ tone with the modulation of figure 3 and no added reflections the loudness of the background is determined by the level of this constant part of the modulation. Although the background is clearly audible, it is not enveloping. It appears to be centered, occupying the same image space as the foreground. As we add lateral reflections the image changes. If the reflections are $< 50\text{ms}$ in delay the background appears frontal, but diffuse. If some of the reflections are $> 50\text{ms}$ the result can be highly spacious.

The high spaciousness is caused by the large variation in the IID and ITD in the space between the notes, where the direct sound is low and the reflections can have a maximum effect. Since the end of the note event is clear, the background loudness integration can start as soon as possible after the note end, making the effective loudness of the background high. If we try the same reflection pattern with the modulation in figure 4 the loudness of the background is lower. One must reduce the level of the reflections in figure 5 by about 6dB to achieve the same loudness with the modulation of figure 4.

These and many similar experiments using both reflections and reverberation (see reference 43) suggest that: High spaciousness requires large fluctuations in the IID and ITD during the background sound. The loudness of the background, and the amount of spaciousness depends on the ease with which the background and the foreground can be separated. Where the sound events have a slow fall, more time must elapse before the background loudness integration can begin, and during this time the reverberant energy will have decayed.

Effect of early reflections on spaciousness

The understanding that spaciousness is (for the most part) an aspect of the sonic background requires us to reevaluate the effect of early reflections. We performed many experiments with the modulator where multiple lateral reflections with delays from 10ms to 100ms were added to a constant (and spacious) reverberation. We have not found a single case where adding the reflections causes the spaciousness to increase. In general, the spaciousness decreases, sometimes dramatically. This result can be explained through the separation process. If we have a series of notes such as in figure 3, but with approximately 200ms length and with 200ms of silence between notes, adding a 1 sec RT reverberation at a total energy of about -5dB produces high spaciousness. During the silence between notes the reverberation decays at 0.06dB/ms. If we add early reflections, whether they are lateral or not, the most likely result will be that the apparent length of the note will lengthen, and the space between them will shorten. Since the actual reverberation is decaying, by the time we detect it as background it will be weaker. Additionally, the space between notes gets shorter - allowing less time for integrating the background before the next note masks it.

Figures 6 and 7 show waveforms from one of these early reflection experiments. The subjective impression of adding the early reflections is an increase in the apparent length of the note. Spaciousness decreases. In another experiment early reflections in the time range of 10ms to 40ms were added at high level. These reflections strongly increased the loudness of the foreground notes, but did not change the spaciousness. The effect of early reflections in our experiments was to increase the apparent length and loudness of a note - they also gave it an "acoustic" quality - but they do not increase spaciousness, and in many cases they diminish it.

These experiments and others indicate that even when the note ends are entirely masked by reverberation a sonic background is formed. The time delay before the start of the loudness integration becomes longer. A drop in level of 3 or 4dB at a 1dB/ms rate is sufficient to separate the note from the background, but where such a drop is absent the separation occurs anyway, but only after the foreground event decreases in level by about 6dB.

Measuring spaciousness

The most important result of these experiments is that spaciousness is determined primarily by the spatial properties of the sound at least 50ms after the ends of notes, and secondarily by the spatial properties of the sound while the note is held. Together these facts preclude the common association of spaciousness with early lateral reflections. For most music spaciousness is associated with the perception of reverberance, and this perception depends on late reflected energy.

There is however another important result - that spaciousness is not a property which can be determined from the impulse response without assuming a lot about the properties of the music. At the very least we would like to know the length of the notes. Longer notes excite the hall reverberation more, which makes the strength of the reverberance relative to direct sound and early reflections higher. With the higher reverberant level from long notes it is harder to detect the end of the note, and thus the separation process for foreground and background is more difficult. We also need to know the rise time and the fall time of the notes, and the average frequency with which gaps appear in which a background impression can develop. Beranek (6) mentions that orchestras change the way they play to compensate for different halls. Deliberately slow fall times will compensate for a lack of hall reverberance by filling the spaces between notes, but the work in this paper shows that slow fall times and small gaps reduce the perception of spaciousness. The thickness of the orchestration matters also. Sparsely orchestrated music, such as Mozart, makes both reverberance and spaciousness more apparent. Bruckner, with continuous sound and dense orchestration, needs a lot more reverberant level to be equally spacious. BSI also depends on the total loudness of the sound.

So we have a dilemma. To measure spaciousness the way the ear hears it we should not use the impulse response, but analyze binaural tapes of particular pieces of music. Early work by Schroeder did this, although the measure used, IACC, is sensitive to a mix of the continuous form of spaciousness (CSI) and early spatial impression (ESI). IACC is also insensitive to spatial properties of sound below 300Hz, which is just where CSI becomes most important. We could improve on this work with a more sophisticated model of the hearing process, and the experiments in this paper are directed to this end. However in the mean time we have practical problems to solve, rooms to compare and improve. We need a temporary measure which tells us roughly what we need to know, using an impulse response (collected any way we can) as the input. The measures proposed below must be seen in this light. They are not intended to work perfectly, but they may be more useful than other simple measures.

The parsing between foreground and background affects the perception of both spaciousness and reverberance. We have studied the perception of reverberance for several years. (24-27). We have developed two measures for it. One, running reverberance (RR), matches data in (18) which was taken under conditions of high direct sound - conditions encountered by solo musicians while practicing or on stage. The other, early decay time at 350ms (EDT350), used data where the direct energy was low - similar to the sounds heard by audience in a concert hall. These two cases differ greatly in the ease with which the sound can be separated into a foreground and a background stream. Where the reverberation is high the ends of notes become quite difficult to distinguish, particularly if the note is held a long time. Typically in a hall the Schroeder integral of the impulse response in a majority of the seats drops less than 2dB when the direct sound stops, indicating that at the end of a long note will not be clearly identified. As the note becomes shorter the end becomes clearer, but in any case we are often unable to determine the precise time a note has ended. It is still possible to separate a foreground and a background stream - but the time needed from the actual end of the note to the maximum sensitivity of the background perception becomes longer. Exactly how much longer needs to be determined, but it may be substantial. (It takes 200ms for sound to decay 6dB with a reverberation time of 2 seconds.)

In (18) we determine that when the ends of notes are very clear - such as when a solo musician is listening to the reverberation of his own instrument on stage - the sense of self support (reverberance) is given by the energy in a time window from 160 to 320ms, divided by the energy of the direct sound.

Where $P(t)$ is the pressure impulse response as a function of time:

$$RR = RR160 = 10 * \log_{10} \left(\frac{\int_{160ms}^{320ms} p(t)^2 dt}{\int_0^{160ms} p(t)^2 dt} \right)$$

RR is directly relevant to the design of concert spaces. It helps answer the question of how loud a rear wall reflection should be to provide self support. It also suggests that multiple reflections are likely to give just as much support as a single reflection, and give less disturbing echo. We typically measure RR at a fixed distance - about 0.5M. The measure is quite similar to what Gade calls "support", but we assume the early energy is not heard. (The inhibition time of 160ms was also found in work on stage support by S. Nakamura as reported at the 14th ICA.) RR might be improved by taking into account the direction of the reflected energy.

For the audience in the hall the ends of notes are much less clear. In a series of pair matching experiments we found that the best measure for reverberance in a hall (where the late reverberation is assumed to be spatially diffuse) is given by the level of the Schroeder integral approximately 350ms after the direct sound. (26). We express this level as an early decay time (EDT350)

RR and EDT350 are very similar measures. They differ in two ways. RR assumes that self support is excited by relatively short notes. Although the impulse response is used directly in the calculation, it is integrated in 160ms blocks. This time was chosen as a compromise between fast music and speech. The time delay before onset of the background audibility - also 160ms in this case, came from the data in (18). EDT350 uses the Schroeder integral in the calculation, not the window integrated impulse response. It thus is measuring the properties of notes which have been held a significant fraction of the reverberation time. The 350ms used in EDT350 came directly from pair matching data with orchestral and chamber music. The time chosen is not without precedence. Jordan's definition of EDT suggests using -10dB of decay as the lower measurement point, and this is the same as 330ms with a 2 second RT. Schroeder suggested using -15dB, which would be 500ms at 2 seconds RT, or 350ms with a 1.4s RT. Thus our 350ms time limit is similar to the way EDT has been measured for many years. Note also we define EDT through only two points on the Schroeder integral - the peak (the loudness of a held note) - and a point later in time representing the loudness of the reverberant energy. Currently many researchers follow Schroeder's suggestion and determine the EDT by a curve fit to the early Schroeder integral. This procedure underestimates the importance of the foreground loudness, and is not recommended. When the direct sound is high curve fitting is not very repeatable when experimental properties such as sample rate, etc. are changed.

As previously mentioned the increased inhibition time in EDT350 over RR may be a result of the difficulty in determining the ends of notes. Another process is probably also at work. Continuous music masks its own reverberation. (26) and (29) report on a computer program which can determine the frequency of gaps in the music where a background effect can be heard. The program found that orchestral music was highly sensitive to the reverberant level. Small increases in level caused a 3 or 4 fold increase in reverberant audibility. Increasing reverberation time and pre-delay increase the probability that reverberation will be unmasked, which would tend to increase the optimum time in the EDT measure over the 160ms we found for RR.

Lateral early decay time - LEDT

When the EDT350 measure was first proposed we were under the impression that by the time 350ms had elapsed the reverberation in all halls was spatially diffuse. Since that time we have measured several halls (opera houses) where this is not the case. We propose a modification to the EDT350 measure to take into account the spatial properties of these halls. The new measure is LEDT, the lateral early decay time. It is a binaural measure, calculated from a binaural impulse response. It depends on the equalized interaural difference - the IAD. The IAD is found in the time domain from a binaural measurement by subtracting the left and the right impulse responses, and then equalizing the result with a low frequency boost filter of 6dB per octave starting at 400Hz. Figure 8. The IAD can be found in the frequency domain by a phase sensitive subtraction of the FFTs of the left and right impulse responses, and then multiplying the amplitudes by the values shown in figure 8, which is plotted for 1/3 octave bands. This calculation is simple and fast in a computer based measurement system, and our system has included it for some years. With this equalization the IAD below about 500Hz is identical to the response of a figure of eight microphone, but the method is self-calibrating. You get both the omni response and a figure of eight response, correctly equalized and balanced, from one binaural measurement. (Two small omni microphones are also quite inexpensive compared to a high quality velocity microphone.) Above 500Hz the IAD follows the true directional response of the head better than a figure of eight microphone.

The IAD is a measure of the diffuseness of a space as a function of time. In a diffuse concert hall such as Boston Symphony Hall, if we plot the IAD for a particular octave band on the same graph as the rms sum of the left and right channels we find the two decays overlap except near time zero, where the direct sound dominates the sum curve, and is missing from the difference curve.

To find the LEDT we calculate the Schroeder integral of the impulse response twice, once for the rms sum of the two ears, and once for the IAD. The EDT350 is then found, but using the peak of the sum integral and the level of the IAD integral at 350ms. In a spatially diffuse hall the EDT350 and the LEDT are equal.

If $S(t)$ is the Schroeder integral of the sum of the squares of the two microphone signals, and $S(0)$ is the peak of $S(t)$, and $SD(350)$ is the value of the Schroeder integral of the frequency equalized difference between the two microphones at 350ms, then:

$$LEDT350 = \frac{60 * 350ms}{(S(0) - SD(350)) * 1000ms / sec}$$

The 40ms integrated sum impulse response and the IAD of one of the non diffuse halls is shown in Figure 9. In this hall the reverberation time of the stage house is longer than the reverberation time of the hall, and the reverberance is mostly frontal. When the impulse response is measured in the unoccupied hall with the curtain down the later energy appears diffuse, but the LEDT value is still low because of the high strength of the direct sound and early reflections. LEDT and EDT values in the unoccupied hall are shown figure 10. By universal agreement, the hall sounds drier and less enveloping than the standard EDT and RT values would predict. The LEDT is a better match to subjective impression.

We expected that the lack of spatial diffuseness at low frequencies would also show up in a measurement made with a stop chord or even with continuous music in the same hall. Figure 11 shows such a measurement, made under occupied conditions. Notice the IAD is well below the sum curve at all times, indicating that the bass is frontal. This measurement shows that total diffuseness is low, since even with the sound source in the pit, and the presence of audience seat back absorption, the apparent source of the sound is frontal. The IAD measures what we hear in a musical environment.

Bradley in (16) suggests a measure for spaciousness (LEV) which is quite promising. The lateral hall gain - $LG[80ms-inf]$ - is the ratio of the lateral energy after 80ms measured with a figure of eight microphone, divided by the total energy the source would produce in a free field at 10m. In a hall with exponential decay LG should decrease somewhat with increasing source to receiver distance. If the source distance is D , and the speed of sound is C , the amount of decrease is given by: $60dB * D / (C * RT)$. In a diffuse, reverberant hall this decrease will be small, and LG will be almost independent of source distance. Our

data show the measure might be improved by choosing a longer inhibition time than 80ms for the lateral energy. Another minor criticism is that the figure of eight microphone does not match the directional sensitivity of spatial impression above 500Hz. It might be easier to use the IAD instead of a figure of eight microphone. The results should be very similar, and this would allow direct comparison between the two measures.

LG measures the relative strength of the background impression in different halls, and should strongly correlate with BSI. LG should work well when we compare the sound of the same orchestra in different halls. LG is a measurement of hall gain - a smaller hall (with its louder sound) will sound more spacious if the same orchestra plays in the same way in two halls. However, small orchestras often play in small halls and large orchestras play in large halls. If we compare two performances of equal total loudness we might find the spaciousness in two different halls to be the same, even if one was small and had a high LG and the other was large and had a low LG. LG is not sensitive to the effects of early reflections, which appear in general to reduce, not enhance, spaciousness. However the major difficulty with LG is a practical one: measuring LG requires a calibrated source. LEDT does not depend on source strength. The data can be collected in a few minutes with two small microphones, a DAT recorder, and handclaps, balloons, or a small loudspeaker. If the hall volume can be estimated, the hall gain can be calculated from the RT, and the lateral gain can be estimated from the difference between LEDT and EDT. When the measurement position is within the reverberation radius LEDT will decrease as you approach the source. At greater distances in a hall with exponential decay LEDT will be independent of distance. Strong early reflections will reduce the value of LEDT. For the present if it is possible it seems wise to calculate both LG and LEDT, and to see how they work in practice.

Implications for music performance spaces

As always, optimizing spaciousness in a performance space depends critically on the type of performance. For speech, spaciousness can be quite high with low values of LEDT. However since intelligibility and spaciousness depend on quite different time periods in the impulse response it is possible to have a hall where there is both high clarity and high spaciousness at the same time. Intelligibility and localization depend on how the hall degrades our ability to detect the beginnings of phonemes and notes. Spaciousness and reverberance depend on how the hall effects the ends of notes and phonemes - and energy arriving after 160ms is of primary importance. To achieve high clarity, spaciousness, and reverberance at the same time we must maximize the later energy while keeping the early energy free of bumps in the 50 to 160ms region which would mask the ends of notes or be heard as separate events. Ideally the energy in this time range should be uniform and not too high.

Intelligibility and clarity are properties chiefly associated with frequencies above 500Hz. For these frequencies it seems clear that spaciousness (BSI) and reverberance depend on later reverberant level, and not on the early reflections. Strong early reflections may increase intelligibility, but in these experiments we found that by filling the gaps between notes early reflections may increase the masking of the music and decrease the spaciousness. Thus to be beneficial the early reflections should arrive promptly, and reflections in the 50 to 120ms time range should be well diffused in time and not high in level. Increasing early lateral reflections at the expense of later reverberance is unwise if intelligibility is already high enough that the ends of notes of medium length are already detectable. The fact that spaciousness depends both on the later reverberant level and on the ability to clearly hear the ends of notes helps explain why seats in the middle of a hall are often more spacious than seats further back.

Frequency and level dependence of spaciousness

In our experience the LEDT values at low frequencies (500Hz and below) are particularly important to the perceived spaciousness and envelopment, a result also found by Morimoto (35). In another opera house we installed a system to enhance the reverberance. We found that the effect was optimum (for a Wagner opera) when the reverberant level was 6dB higher below 500Hz than above. See figure 12. Such a curve gave a greatly improved spaciousness on the orchestra without changing the excellent intelligibility of the hall for the singers. It increases LEDT at low frequencies substantially over the values in the unassisted hall. Time

and equipment limitations prevented us from achieving the optimum spatial diffuseness of the low frequency sound in this experiment, as a plot of the LEDT and the EDT in the unoccupied hall clearly shows. For this measurement the system was set for ballet, so the reverberant level is more constant with frequency and the reverberation time is higher. Figure 13. LEDT was very helpful in analyzing the results of this experiment, and in designing the system which will be permanently installed (34).

Music contains at least half of its energy below 500Hz. Figure 14. We contend that spaciousness and reverberance are of great importance for these frequencies. In fact, many well liked halls which would be perceived as dry on the basis of their mid frequency reverberation time are quite pleasant because the low frequency LEDT is high. (The Tanglewood music shed is a good example.) The opposite is also true. Low frequency instruments tend to have long note decays - for example string bass, bass drum, and kettle drum. This makes the separation of sound into a foreground and a background stream difficult, and makes it largely irrelevant what the reverb time of the hall is. However BSI is not the only mechanism available for spaciousness. CSI also works when the total energy at the listener's ears is spatially diffuse. In a concert hall the direct sound is often severely attenuated by seat back absorption. The remaining sound - if it is not entirely from the ceiling and the back wall - is likely to be diffuse. Thus as long as the low frequency energy is strong enough, lateral enough, and delayed $\gg 55\text{ms}$ the low frequencies are likely to be sufficiently enveloping and spacious. Early lateral reflections - which tend to increase diffuseness at the expense of later reverberant level - may be effective at low frequencies and not at high frequencies. But once again there is a complication. To produce spaciousness and not ASW the low frequency energy must have sufficient delay. This may be difficult to achieve with a first order reflection. Since low frequency loudness correlates with RT and EDT, until this situation is better understood LEDT (or LG) seems an appropriate measure.

BSI at all frequencies depends on two properties of the sound source - its absolute level, and the degree to which it masks its own reverberation. When the music is thickly orchestrated and played legato BSI is severely masked. In this case spaciousness is a combination of CSI (most of the time) and BSI in the relatively few gaps where it can be heard. When this occurs spaciousness may be more independent of sound level than for more transparent music. Where BSI is easily heard the spatial impression is proportional to the loudness of the source. This may explain why a hall with a high reverberant level (like the Vienna Grosse Musikvereinsalle or the Zurich Tonhalle) is effective with music like Mozart, but can become overwhelming with louder music. Ideally we would like the reverberant level to depend on the total loudness - a solution which becomes possible with electronic acoustical augmentation. This idea was tried in (34) with good results.

Conclusions

Source width increases when lateral reflections arrive during the rise times of sound events. When the rise time is uncontaminated by reflections the image is sharp. When reflected energy arrives within 50ms of the end of a sound event the spatial impression of a small room is created. This impression is not spacious. Spaciousness and envelopment occur when spatially diffuse reflected energy which is at least 50ms delayed arrives either during note events or in the spaces between events. Spaciousness and envelopment arise from two spatial impressions, one depending on the separation of incoming sound into a foreground and a background, and the other depending on the total spatial diffuseness of the sound. When the foreground consists of a series of connected events with clear endings, such as notes or phonemes, the spatial impression associated with the background impression (the reverberance) is dominant. Under these conditions it is primarily the late (more than 160ms delayed) reflected energy which determines the perceived spaciousness, and the spaciousness depends on the loudness of the source. Early lateral reflections increase the loudness and apparent length of foreground events, but may decrease the spaciousness. With orchestral music the loudness of the background impression and the spaciousness is determined both by the reverberant level and by the degree of musical masking. Lateral Early Decay Time (LEDT) and lateral gain (LG) are suggested as useful measures for spaciousness and reverberance.

References and Bibliography

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. Y. Ando and Y. Kurihara "Nonlinear response in evaluating the subjective diffuseness of sound fields" J. Acoust. Soc. Am. 80 (3), Sept. 1986 pp 833-836
3. M. Barron "Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure" J. Sound and Vibration (1981) 77(2), 211, 232
4. L. Beranek "Music, Acoustics and Architecture" John Wiley, 1962
5. L. Beranek "Concert Hall Acoustics - 1992" J. Acoust. Soc. Am., Vol 92, 1. No. 1, July 1992
6. L. Beranek "Concert and Opera Halls How They Sound" Acoustical Society of America, 1996
7. J. Blauert "Zur Tragheit des Richtungshorens bei Laufzeit- und Intensitätsstereophonie" Acustica 23 p287-293 (1970)
8. J. Blauert "On the Lag of Lateralization Caused by Interaural Time and Intensity Differences" Audiology 11:265-270 (1972)
9. J. Blauert "Raumliches Horen" S. Hirzel Verlag, Stuttgart, 1974
10. J. Blauert "Spatial Hearing" MIT Press, Cambridge MA 1983
11. J.S. Bradley "Contemporary Approaches to Evaluation of Auditorium Acoustics" Proc. 8th AES Conference Wash. DC May 1990 pp 59-69
12. J. S. Bradley "Contemporary approaches to evaluating Auditorium Acoustics" Proceedings of the Sabine Conference, MIT June 1994. Available from the Acoustical Society of America, 500 Sunnyside Blvd. Woodbury, NY 11797
13. J.S. Bradley and G.A. Souloudre "Spaciousness judgments of binaurally reproduced sound fields" Ibid. p 1-13
14. J.S. Bradley "Comparisons of IACC and LF Measurements in Halls" 125th meeting of the Acoustical Society of America, Ottawa, Canada, May 1993
15. J.S. Bradley "Pilot Study of Simulated Spaciousness" Meeting of the Acoustical Society of America, May 1993
16. J. S. Bradley, G. A. Souloudre 'Objective measures of Listener Envelopment' J. Acoust Soc. Am. 98, 2590-2597 (1995)
17. J. S. Bradley, G. A. Souloudre 'Listener envelopment: An essential part of good concert hall acoustics' JASA 99 (1) Jan 1996, p 22
18. W. Gardner and D. Griesinger 'Reverberation Level Matching Experiments' Proceedings of the Sabine Conference, MIT June 1994. p263
19. M. A. Gold 'Subjective evaluation of spatial impression: the importance of lateralization' ibid. p97
20. D. Griesinger "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. D. Griesinger "IALF - Binaural Measures of Spatial Impression and Running Reverberance" presented at the 92nd Convention of the AES March 1992 Preprint #3292
22. D. Griesinger "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. D. Griesinger 'Progress in electronically variable acoustics' Proceedings of the Sabine Conference, MIT June 1994.
24. D. Griesinger 'Subjective loudness of running reverberation in halls and stages' Proceedings of the Sabine Conference, MIT June 1994. p89
25. D. Griesinger 'Quantifying Musical Acoustics through Audibility' Knudsen Memorial Lecture, Denver ASA meeting, 1993. Copies available from the author
26. D. Griesinger 'Further investigation into the loudness of running reverberation' Proceedings of the Institute of Acoustics (UK) conference, Feb 10-12 1995.
27. D. Griesinger 'How loud is my reverberation' Audio Engineering Conference, Paris, March 1995. Preprint # 3943
28. D. Griesinger 'Design and performance of multichannel time variant reverberation enhancement systems' Proceedings of the Active 95 Conference, Newport Beach CA, June 1995
29. D. Griesinger 'Optimum reverberant level in halls' proceedings of the International Congress on Acoustics, Trondheim, Norway June 1995

30. T. Hidaka, L. Beranek, T. Okano '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, 988-1007 June 1995
31. J.P. Jullien, E. Kahle, S. Winsberg, O. Warusfel "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
32. E. Kahle 'Validation d'un modele objectif de la perception de la qualite acoustique dans un ensembl de salles de concerts et d'operas' Doctorate Thesis, IRCAM June 1995
33. W. deV. Keet "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
34. A. Kreiger 'Nachhallzeitverlangerung in der Deutschen Staatsoper Berlin' Tonmeister Informationen (TMI) Heft 3/4 Marz April 1996.
35. M. Morimoto, Z. Maekawa "Effects of Low Frequency Components on Auditory Spaciousness" Acustica vol 66(1988) pp 190-196
36. M. Morimoto, C. Posselt "Contribution of Reverberation to Auditory Spaciousness in Concert Halls" J. Acoust. Soc. Jpn (E)10, 2 (1989)
37. M. Morimoto, Z. Maekawa 'Auditory Spaciousness and Envelopment' 13th ICA, Yugoslavia 1989
38. M. Morimoto 'The relation between spatial and cross-correlation measures' 15th ICA Norway 1995
39. P. Schubert 'Die Wahrnehmbarkeit von Einzelruckwurfen bei Musik' PhD thesis, TU Dresden 1967
40. M.R. Schroeder, 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, 1195ff (1974)
41. W. Hartman 'Localization of a sound source in a room' Proceedings of the 8th international conference of the Audio Engineering Society, May 3-6 1990
42. T. J. Schultz, "Acoustics of the Concert Hall" IEEE Spectrum (June 1965) pp 60-62
43. D. Griesinger "Spaciousness and envelopment in performance spaces" submitted to Acta Acustika, May 1996

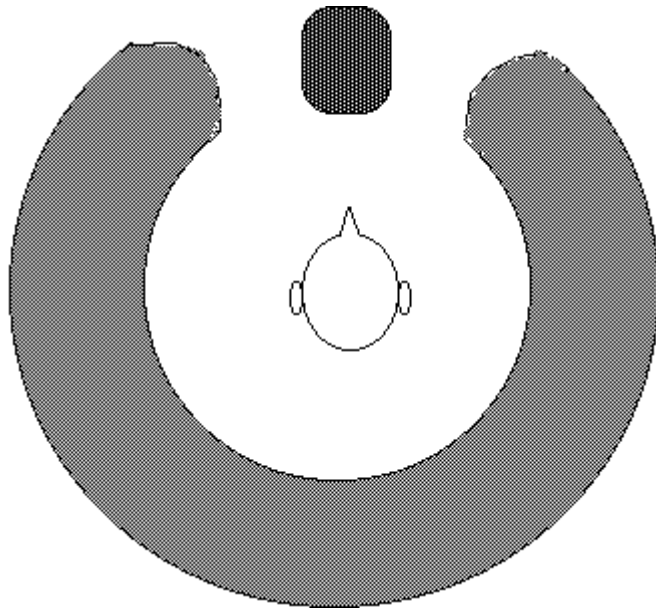


Figure 1a: Background spatial impression (BSI) - Spatial impression from a pulsating source with clear event endings, where diffuse reflected energy is $> 55\text{ms}$ delayed. This sound is heard as a separate perception from the notes which create it. This is the perception associated with the common meaning of spacious and spaciousness. It occurs in sparsely orchestrated music and strongly in solo music which is played *marcato* or *detache*. The strength of the impression depends on the strength of the reflected energy, and not on the ratio of the direct to reflected energy.

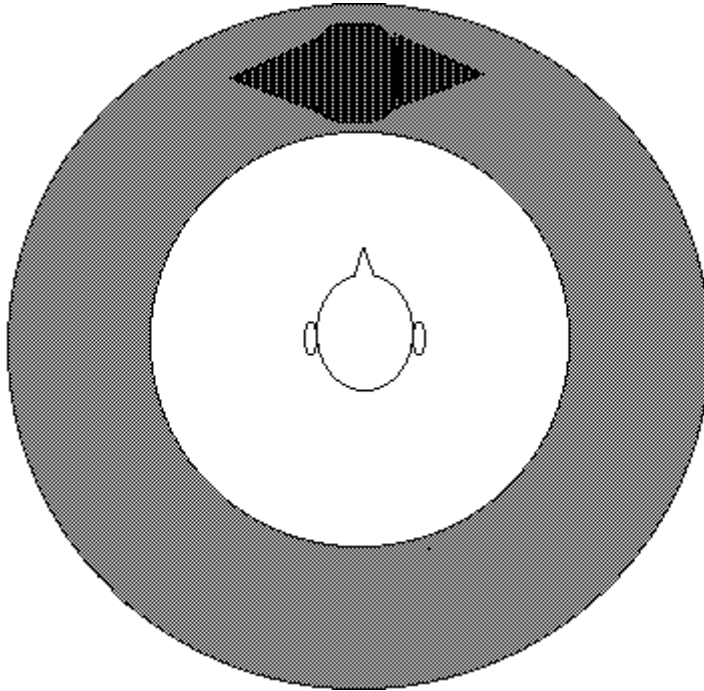


Figure 1b: Continuous spatial impression (CSI) - Spatial impression associated with a continuous source or a source with indistinct event ends. While the sound appears all around the listener, it is not perceived as a separate entity or impression. CSI is heard in thickly orchestrated music and when a soloist plays legato. Although very similar to BSI, CSI is subjectively 4 to 6dB weaker than BSI unless the diffuse reflected energy is equal to or greater than the direct sound. Under these conditions BSI and CSI are difficult to distinguish. CSI can be spacious when BSI is absent or masked. Delays $> 55\text{ms}$ are also required for CSI with any pulsating or musical source. CSI - at least with noise as a source, is not dependent on source loudness as long as the reverberant level is well above the threshold of hearing.

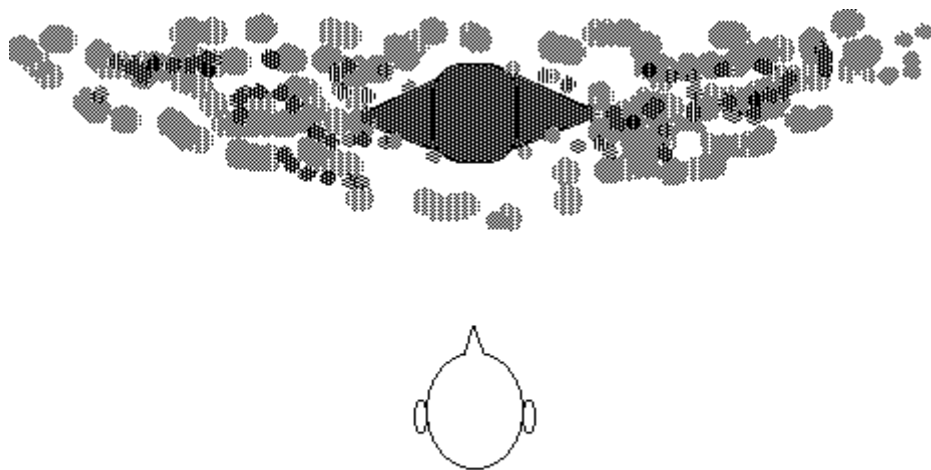


Figure 1c: Early spatial impression (ESI) - a spatial impression from a source which includes discrete events, arising from reflections arriving before 50ms. The spatial impression is primarily frontal. This is the spatial impression of a small room. ESI is not spacious or enveloping. In a large rooms ESI and BSI are separately perceptible. ESI depends on the ratio of direct and reflected energy, and is thus independent of source loudness. Depending on the sharpness of event onsets, diffuse energy in the 2ms to 50ms range will also increase ASW. This is particularly true with legato sources such as romantic strings.

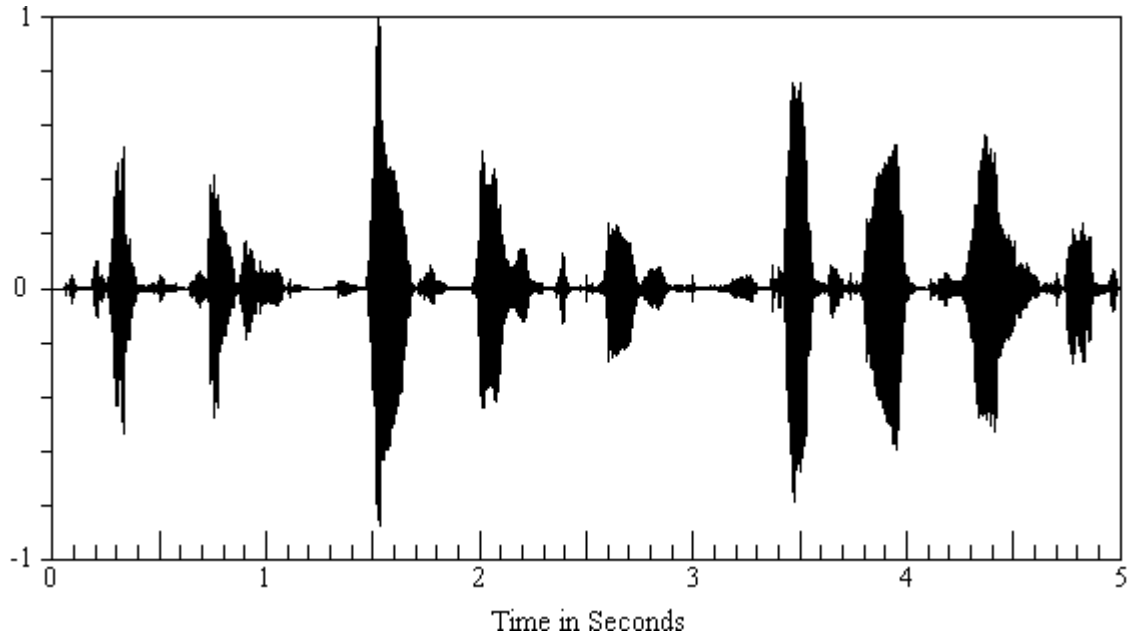


Figure 2: 1kHz octave band filtered speech. (“The course is titled...”) Note the fast rise time of most phonemes, and the clear spaces between phonemes. A lateral reflection during the rise time can broaden the image. Reflections arriving later do not. Lateral reflections in the spaces between the phonemes are spacious if the delay > 50ms.

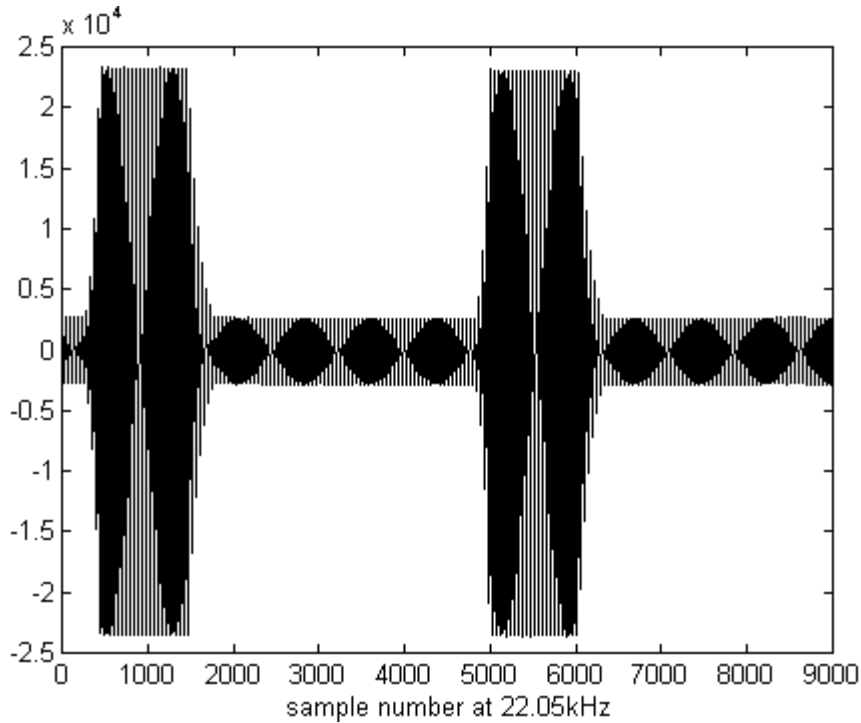


Figure 3: 500Hz tone with 18dB modulation, 2dB/ms rise time, 1dB/ms fall time. This waveform when combined with lateral reflections produces sharp sonic images and high spaciousness (when some of the reflections are > 50ms delayed.)

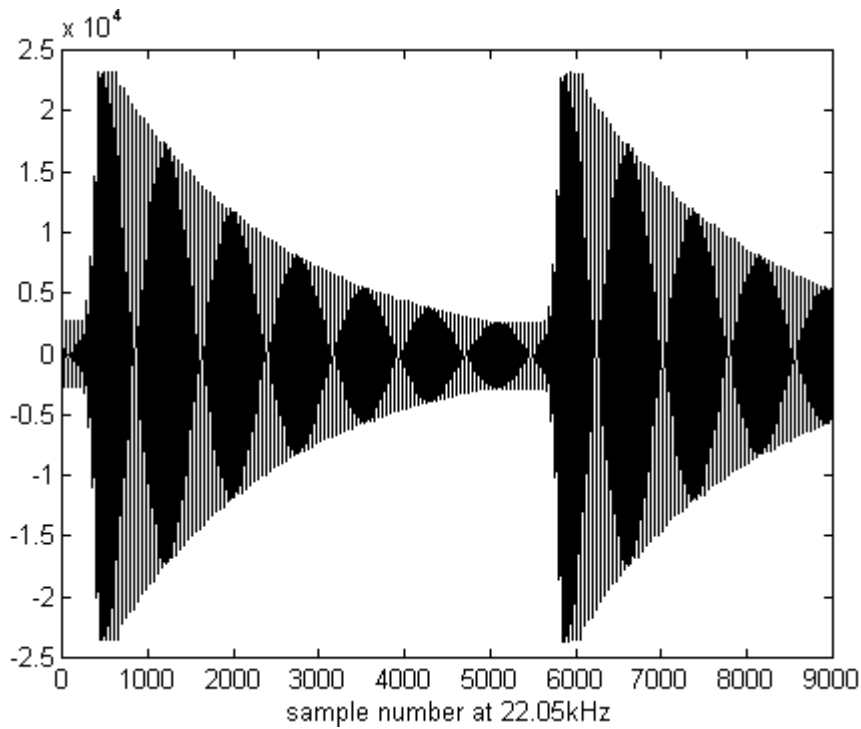


Figure 4: A waveform of similar loudness as figure 3, but with a slow fall time of 0.06dB/ms (1sec RT) This waveform also produces sharp images, but the spaciousness is weaker than with figure 3.

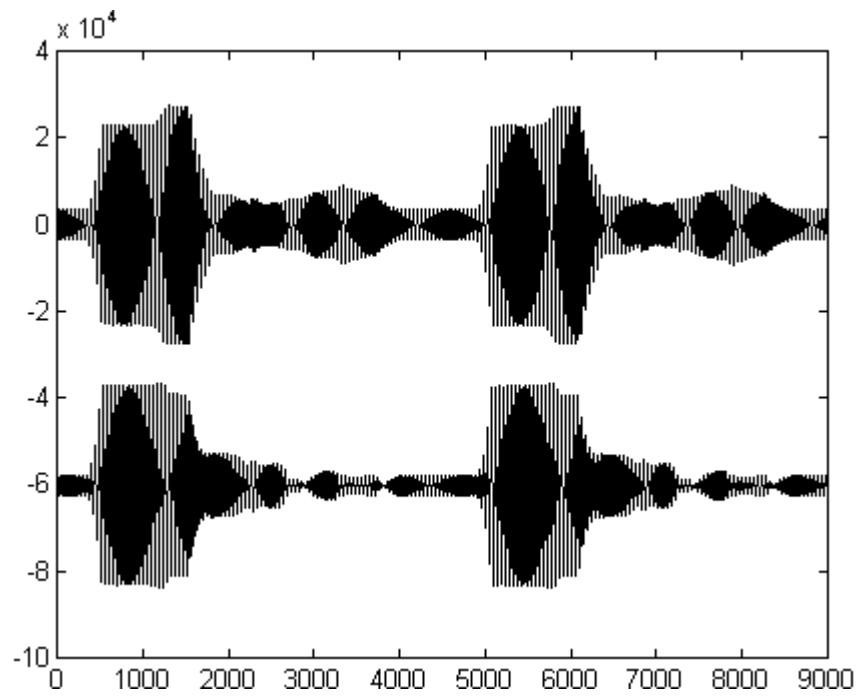


Figure 5: modulation of figure 3 with reflections, some of which are > 50ms. Note the large variation in the IID in the space between the notes.

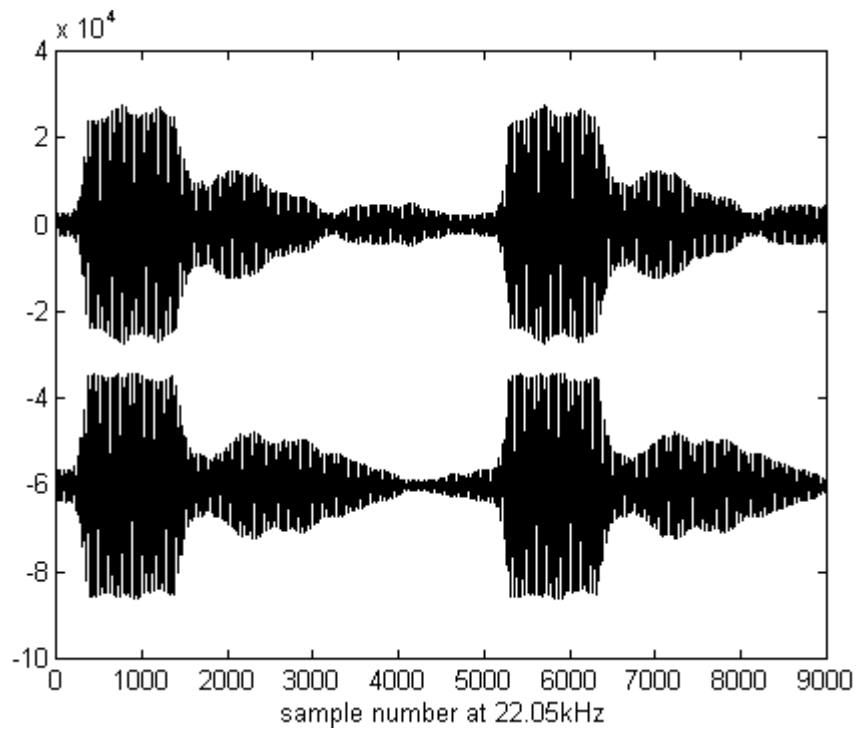


Figure 6: Left and right waveforms for the inhibition experiment. 0.5s RT, no early reflections. The sound is highly spacious. Note the large IID differences in the space between the notes.

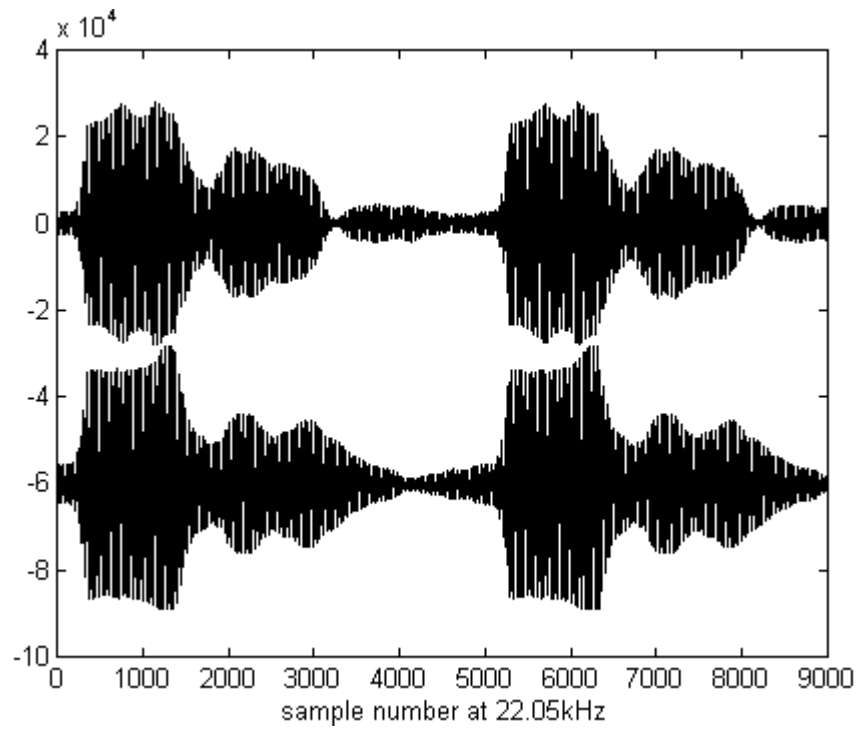


Figure 7: The same waveforms as figure 6, but with early reflections. The sound is less spacious. The reflections are strong enough to inhibit the detection of the end of the note. The auditory length of the note is extended by the reflections, and the apparent reverberation and spaciousness reduced.

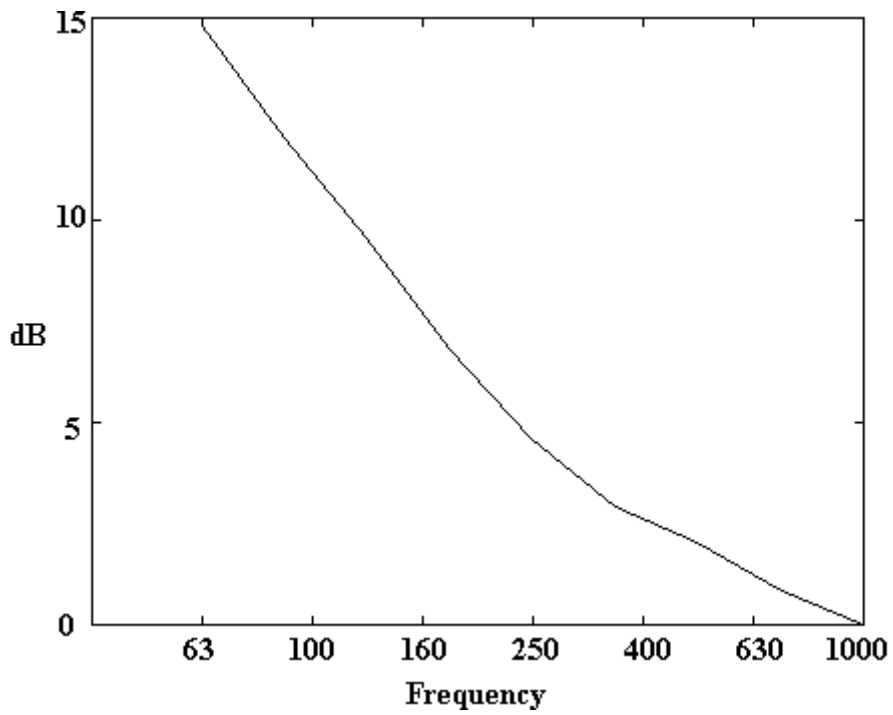


Figure 8: equalizer response for IAD calculation - plotted in 1/3 octave bands

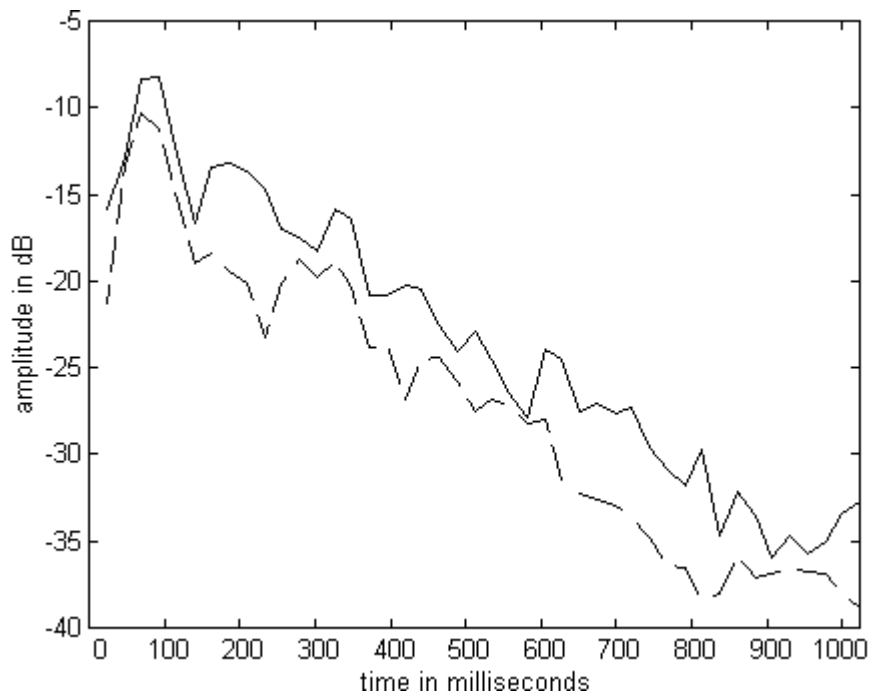


Figure 9: rms sum and IAD for the binaural impulse response of an opera house. Source on stage, microphone in center of stalls. 500Hz octave band

_____ = rms sum of the left and right ears
 - - - - - = IAD

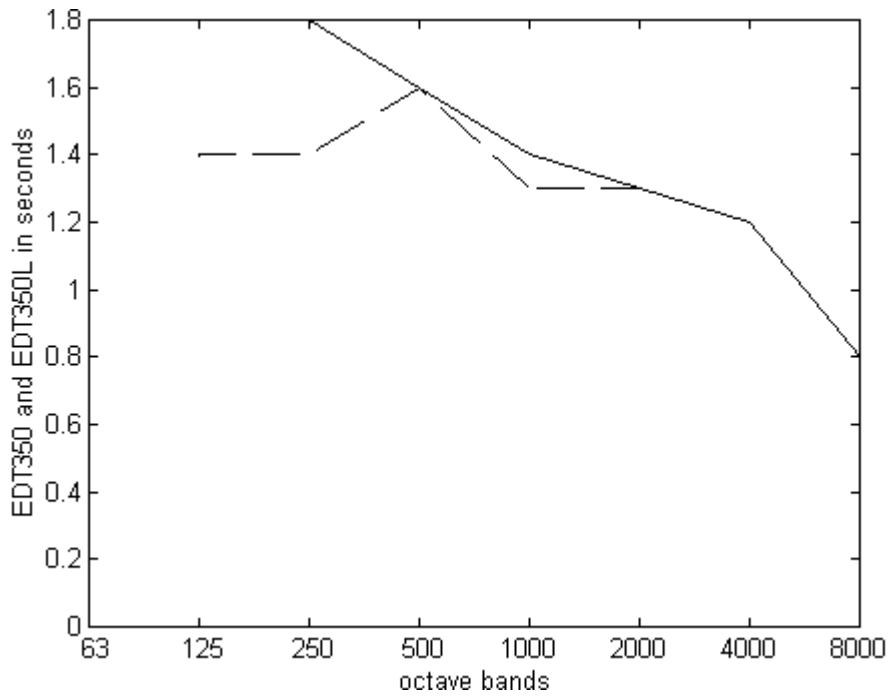


figure10: EDT and LEDT values in the opera house of figure 9, unoccupied

— = EDT

- - - = LEDT

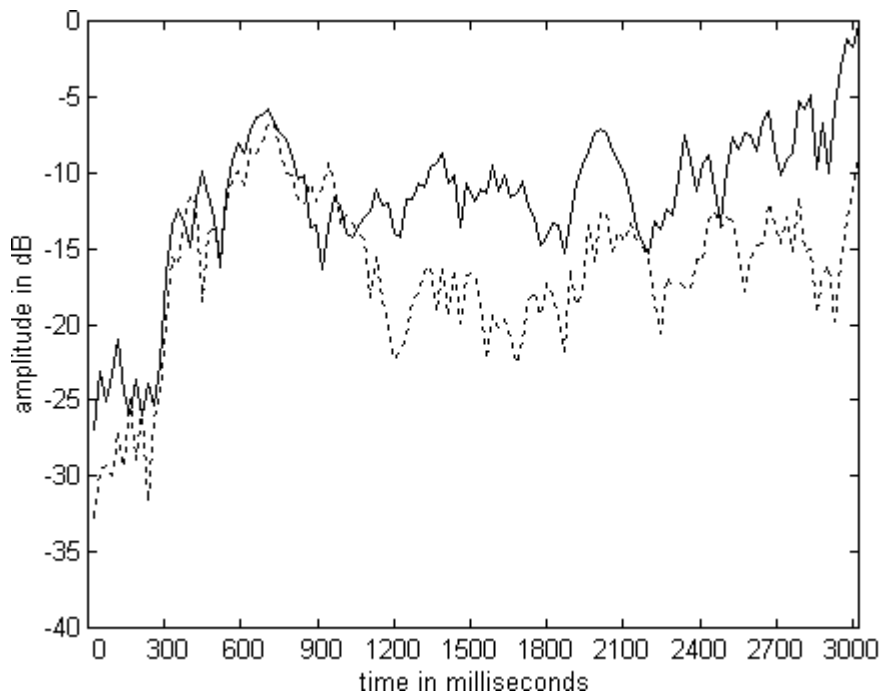


Figure 11: ____Medial and - - - lateral (IAD) sound energy in the opera house of figure 9, from a music segment recorded binaurally in the middle of the stalls. 125Hz octave band

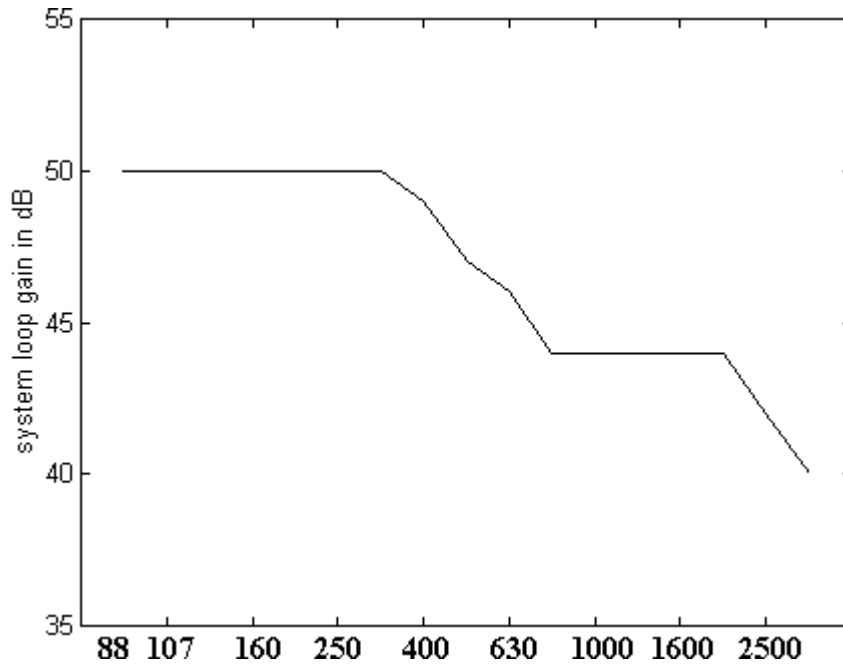


Figure 12: Loop gain of the enhancement system installed in an opera house, plotted versus 1/3 octave band. Note the increase in loop gain below 500Hz. The calibration of the vertical scale is arbitrary.

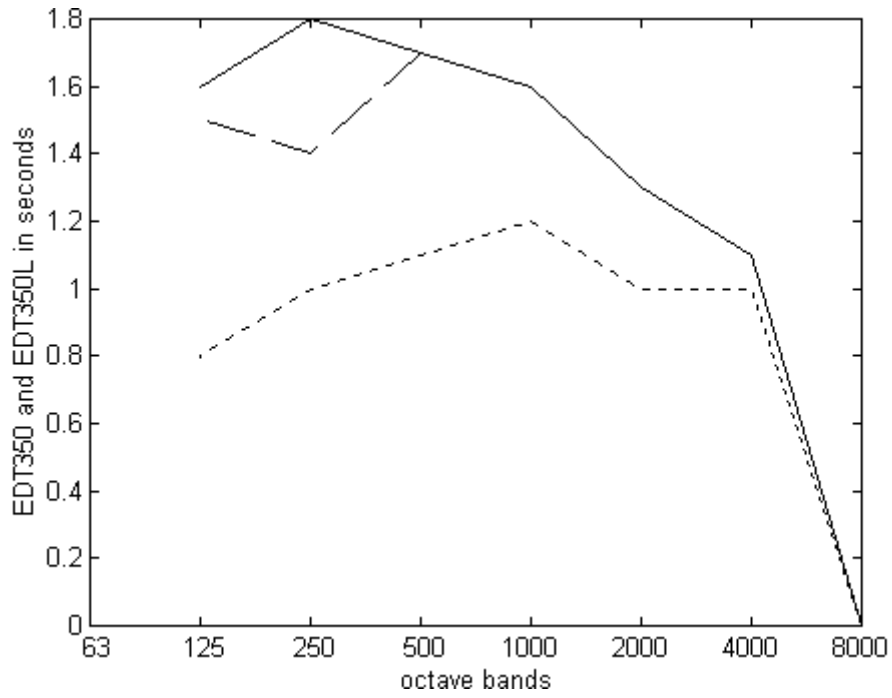


Figure 13: LEDT and EDT values in the opera house of figure 12, but with a flat loop gain. System set for ballet.

- EDT, system on
- - LEDT, system on
- - - LEDT, system off

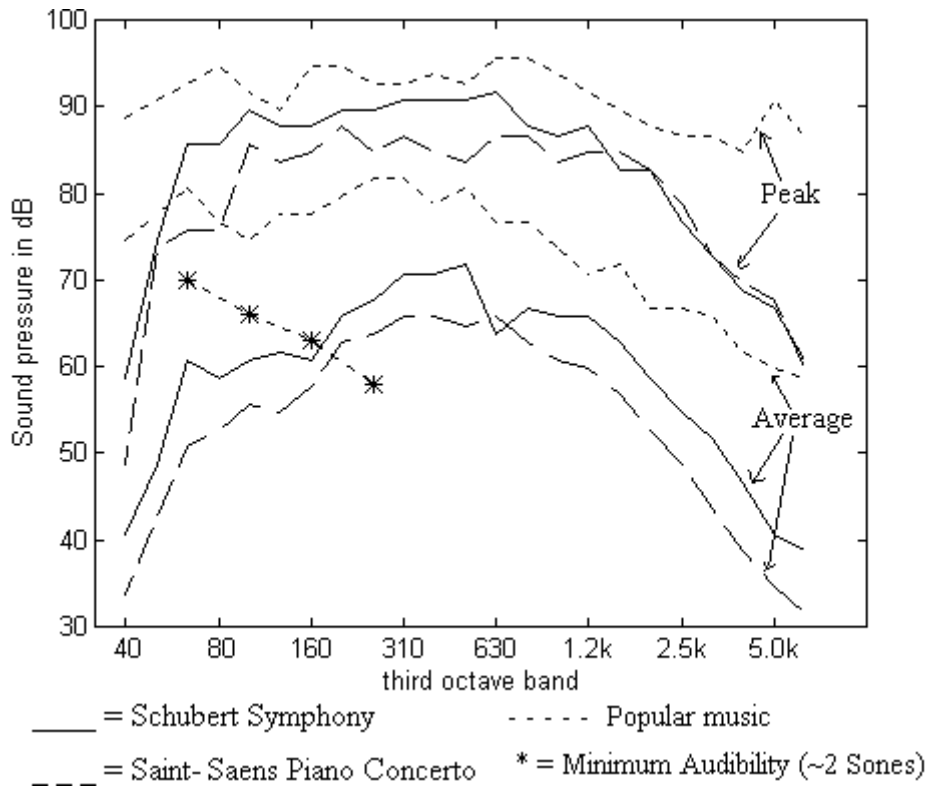


Figure 14: Peak level and average level for three different types of music, plotted with a 2 sone equal loudness curve. 2 sones is not a threshold, but a level below which material at the given frequency is likely to be ignored.