Preference versus Reference: Listeners as participants in sound reproduction

Durand R. Begault[1]

[1]NASA Ames Research Center, MS 262-2, Moffett Field, CA 94035 USA

ABSTRACT

A basic tenet of sound reproduction is the concept of 'fidelity', meaning as many aspects of the acoustical context of an original recording be reproduced as faithfully as possible to please the 'critical listener'. The point of reference may either be the studio control room or a live venue, and deviation in reproduction may be objectively quantified within limits. More recently the importance of quantifying subjective response so as to build a point of reference based on preference data has been introduced. An alternative model considers the critical listener as a participant in the sound reproduction process, for whom the sensation of 'ideal reproduction' is malleable and that can take many forms. From this standpoint, the taxonomy of attributes such as timbre, level and spatial reproduction fall short of any point of reference in many real-world audio reproduction contexts, and yet the critical listener is satisfied. Furthermore there is no reason to assume that preference is 'time invariant', although this is frequently assumed. Acoustic analyses of loudspeaker playback in typical rooms are suggestive of spatial characteristics that are the result of conscious or unconscious participation by the end user. It is suggested that there are spatial attributes that fall out of this participation that are currently undervalued in current research.

1. INTRODUCTION

Audio research has traditionally been concerned with a *reference* condition, against which the quality of sound reproduction can be evaluated via objective measures. More recently, there has been interest in perceived quality of sound reproduction, and how perceived quality might differ from objective measures. Such investigations give insight into perceptual qualities that might otherwise not be assessed from physical measures, and indicate the *preference* of listeners.

Spatial audio reproduction is widely recognized as an important component of both objective and subjective perspectives of audio reproduction. We can refer to this as *spatial auditory fidelity* or *spatial audio quality*. Rumsey states that "The primary aim of most commercial media production is not true spatial fidelity to some notional original sound field" [1]. However, both objective and subjective approaches to spatial audio reproduction require establishment of a comparative reference, against which quality is measured. This reference, or 'gold standard', can be established based on the intent of the engineer; to what people "like" (*preference*); or on acoustical measures of the "notional original sound field".

This paper discuses two examples in which *preference* differs from a reference condition. First, the idea of 'listeners as participants' addresses the complexity of everyday listening and how it differs from the model of listening used in psychoacoustic experiments, particularly for spatial audio reproduction. Second, the subtleties of preference for sound amongst three cinemas are examined, and how preference differs from what might be expected based solely on objective measures.

2. REFERENCE VERSUS PREFERENCE

A *reference* emphasis focuses on the precision with which an engineer can manipulate a sound to a specific perceived location, i.e., to a prescribed *reference* or calibration point for a listener. This perspective is analogous to the sending of 'messages' in communications theory and the notion of source-medium-receiver transitivity (see [2, 3]). We can then speak of *spatial audio fidelity* as a quantitative variable concerning spatial reproduction *accuracy*.

One approach is to measure localization accuracy of listeners, with reference to an intended target. Such studies are often concerned with what Blauert describes as the location of the 'sound event' versus studies of listener's perception of the 'auditory event'. The auditory event, or what we actually hear, can include percepts in addition to the auditory event's location. "Spatial coincidence of the sound source and the auditory event is certainly a common situation….It also occurs- and not infrequently- that the auditory event appears in a position different from the sound source" [4]. Non-musical auditory displays for workers in aviation, rescue, control communications, etc., have been shown to benefit from a prescriptive, reference-based system concerned with perception of the location of the sound event [5]. To control variability, such auditory displays typically use headphones.

The alternative emphasis on spatial audio quality investigates more fully the perception of the auditory event by the listener, and what attributes they might use to scale preference. Rumsey has summarized current work in reference [1] from this conference. Ultimately, the aim is to determine engineering parameters that allow prediction of what will be perceived as 'correct' or 'enjoyable' spatial sound reproduction.

The origin of this work lies in concert hall research in the 1970s by Schroeder and by Barron and Marshall [6,7]. They established that listeners prefer music listening in rooms that are more spatially decorrelated at the ears compared to rooms that are less so. Interestingly, the main attributes cited by Rumsey in [1] intersect with those attributes cited in concert hall acoustic research: including auditory source width, envelopment, distance ('intimacy', 'strength'), locatedness, and naturalness [8].

3. LISTENERS AS PARCIPANTS

In listening experiments, there are several important differences from the true 'ecology' of everyday home listening. Experimental design necessitates minimizing independent variables and confounding influences. Furthermore, it requires the attention of the listener to be 'active': focused only upon the experimental variable(s) of interest. In all other aspects, the listener is 'passive'. They are not participants in the choice of program material, sound reproduction equipment, or in determining any simultaneous activities that they may like to do while listening to music.

Usually, the attitude of the body and head reflects the 19th century model of the concert hall experience: quietly looking directly ahead, actively listening. Generally speaking, the experience of listening in the laboratory is invariant except for the stimuli under control of the experimenter.

What really occurs when people listen to spatial sound at home? In reality, the 'spatial auditory art object' of reproduced music involves a listener who varies the mode of audio reproduction in many ways, who in effect 'participates' in the rendering of spatial qualities. Furthermore, their interaction with the 'art form' is time-variant, involving multiple frames of reference to acoustic waveforms.

The following lists some obvious and some subtle causes for why everyday listening varies significantly from laboratory listening tests.

• *Movement of the head or the body away from the 'sweet spot" during active listening.*

It is well known that everyday localization involves movement of the head and body, partially in an attempt to minimize binaural differences so as to direct the eyes toward a sound source. In the laboratory, it is nearly impossible without a chin rest or bite bar to maintain a time-invariant orientation of the ears to loudspeakers. The head and body move for comfort and to effectively resample the audio environment from multiple perspectives, even when seated in the 'sweet spot'.

• *Moving into and out of one or more sound fields during semi-active listening.*

Many persons out of necessity or desire listen to music from one or more loudspeaker 'systems' while moving about their homes, in the course of performing various activities or chores. A single loudspeaker system can be made audible throughout multiple rooms.

Wireless systems now allow the same program material to be broadcast through multiple loudspeakers in different rooms relatively inexpensively, compared to wired systems. It is likely that listeners will increasingly participate in making 'home soundscapes', where their entire home can be acoustically energized so as to make musical program material intelligible from multiple perspectives.

Research has not yet determined if the collective ensemble of different spatial sound impressions may result in a single global impression. Such an overall impression may be significant to listeners as an attribute in evaluating ‘spatial quality’.

• *Different levels of attention and involvement*

A listener will devote more or less attention to musical material depending on their location, their interest in the program material, and on the demands of their other activities. The appreciation of spatial quality may involve presentations that are preferred for ‘semi-active’ listening. Some listeners prefer sound reproduction to consistently mimic the qualities of background music (akin to Eric Satie’s concept of “furniture music”). This has ceased to be a pejorative term amongst musicians and listeners alike.

Interestingly the background music systems that are ubiquitous in many settings are inherently spatial, but little research has been devoted to spatial quality evaluation for such systems. Spatial sound quality of background music may also be influenced with association with other previous experiences of background music. The Muzak corporation describes on their website the emotional aspects of “audio architecture” as follows:

> Audio Architecture is emotion by design. It is the integration of music voice and sound to create experiences that link customers with companies. Its power lies in its subtlety. It bypasses the resistance of the mind and targets the receptiveness of the heart. When people are made to feel good in, say, a store, they feel good about that store. They like it. Remember it, Go back to it. Audio Architecture builds a bridge to loyalty. And loyalty is what keeps brands alive. [9].

Perhaps some listeners would evaluate spatial audio quality by associating it with enjoyable previous experiences of background music, and seek to incorporate that into their homes. To paraphrase the Muzak quote above, “When people are made to feel good, in say, their, homes, they feel good about that home”; etc..

• *Adjusting playback level*

Listeners not only differ both between and within themselves in their preferred playback levels for different program material: some will actively manipulate levels, e.g., during a film. This causes alterations of timbre, explainable by equal loudness contours, and can easily vary spatial percepts.

Usually, if a button, dial or other sort of interface control over an acoustical parameter is given to a consumer, they are in effect being invited to manipulate that parameter, and in most cases they will.

• *Loudspeaker Placement*

More often than not, listeners will place loudspeakers in locations non-amenable to ideal spatial reproduction from the perspective of the recording engineer. Sometimes this is dictated either by available space or for keeping pets and children out of reach. But some listeners prefer to have loudspeakers in non-standard locations.

Locating loudspeakers near the ceiling and out of the direct path of the listener can lend a diffuse spatial quality to the sound that is desirable, mimicking the acoustical qualities of a club or public gathering place. Some modern music styles are quite amenable to such qualities, having the sound come from ‘everywhere’, spatially differentiated yet not identifiable from a specific location. The same principle is used for some types of surround sound loudspeakers.

• *Choosing loudspeakers for a specific, non-linear acoustical effect*

Some listeners prefer small, self-contained sound systems over a loudspeaker system. These systems allow a more intimate listening experience as a result of lower sound levels and reduced loudspeaker size and extent. Private or public listening can be effected using a portable boom box. With these systems, stereo reproduction cues are audible but certainly not ‘correct’ with regards to a reference.

There are also some individuals who play monaural recordings through a single loudspeaker, who play guitar recordings through open-backed tube guitar amplifiers, prefer their Victrola to anything else, or play 45 rpm records through their Wurlitzer Jukebox. A specific acoustic non-linearity may be desirable because it is the point-of-reference from which the recording was originally heard. These are the more extreme versions of listeners participating in the formation of the acoustical event.

* * *

The above list does not include the effect of variability in room acoustics, which has been reported on extensively (e.g., [10]), and is analyzed for cinemas in Section 4. Although there have been some published surveys on consumer use of sound systems, little research has looked into how these factors interact, leading to a contemporary *zeitgeist* of music listening, and how to incorporate these factors into a model that might create future directions for predicting spatial sound quality. These variables can be summarized as follows:

First, the acoustical experience of listeners is time-varying, and involves multiple perspectives from which an aggregate internal preference scale may exist. Listeners experience and probably enjoy multiple, differing frames of reference for their overall assessment of timbre and spatial sound quality at different times within a relatively short period of time.

Second, people actively introduce time-varying factors into everyday music listening experiences. For some listeners, the ability to vary the nature of sound reproduction may be nearly as important as program material itself. The recording engineer and the participating listener work together in the realization of reproduced music as 'spatial auditory art object'.

Clearly, any type of reference 'gold standard' based on a passive fixed listener is not applicable to everyday listening. These experiences are highly differentiated from the context of listening experiments. Future research might examine more closely how listeners interact with sound systems, to understand more 'ecologically valid' aspects of spatial sound quality that go beyond a model of listening based on the 19th century concert hall.

4. ANALYSIS OF SINGLE-CHANNEL DATA FROM THREE CINEMAS

The following analysis illustrates how an ideal 'reference' condition for acoustical sound reproduction of spatial effects is at odds with preference. Three cinemas located in the same urban area were analyzed in terms of preference and room acoustics. These will be referred to anonymously as cinemas A, B and C. They are characterized as follows:

A. Historical cinema, built late 1920s, raked seating: mid-band reverberation 1.2 s; approx. 1100 seats

B. Historical cinema II, built early 1930s, raked seating, with balcony: mid-band reverberation 1.0 s; approximately 1300 seats

C. Modern cinema, built in 1990s, with stadium seating: mid-band reverberation 0.3 s; approximately 300 seats.

All three cinemas have comparable sound systems, with modern loudspeakers and amplifiers from the same manufacturer. All three cinemas are also comparable in terms of satisfying modern criteria for film sound systems (THX, etc.). The analyses that follow were based on the center channel cluster, measured at 10 feet from directly in front, and at center seating locations 1/3 and 2/3 back from the screen to the rear wall.

Based on objective data, cinema C would best comply with a reference 'standard'. The low reverberation time is due to the presence of a significant amount of acoustical absorption and its relatively smaller volume. It allows spatial effects to be unhindered by room reverberation or early reflections, and speech intelligibility is excellent. Cinemas A and B, by contrast, have significant early reflections and longer reverberation times. Speech intelligibility is more problematic as a result.

Despite these facts, the rank order of preference for these cinemas in terms of sound quality places C last. Interviews with laypersons and film professionals indicated the order of preference as: B – A – C. Cinemas A and B are appreciated for their visual qualities and their historical nature, which can certainly bias the rating over the modern cinema C. The longer reverberation time of historical cinemas ('movie palaces') likely also conveys a desirable nostalgic characteristic.

There was also a strong preference for cinema B over cinema A, primarily because of perceived speech intelligibility. Cinema B also was reported by some to have a more 'natural' sounding cinema reverberation. Finally, differences were never mentioned or noticed in spatial quality for the conveyance of 'surround' or other effects. Although cinema C would clearly be most effective for conveying spatial effects *accurately* with reference to where the film was mixed, this was never mentioned as a component of overall quality. The following analyses focus on the objective causes for these preference differences.

Subjective speech intelligibility was measured objectively in each cinema by obtaining the Speech Transmission Index value. Measured values can range from 0-1.0, where 0.3 is "bad", 0.3-0.45 is "poor", 0.45-0.6 is "fair", 0.6- 0.75 is "good" and 0.75-1.0 is "excellent". The values are primarily affected by signal level versus noise (reverberation and background noise level). Figure 1 left, shows the measured data for STI from two seating locations in each theatre. In terms of rank order from worst to best, the STI values range from 0.68–0.7 in cinema A, or "good", while from 0.75–0.81 in the cinema B ("good" to "excellent"), and from 0.91–0.94 in cinema C ("excellent").

Early Decay Time (EDT) is similar to reverberation time but calculates the time for sound to decay 10 decibels in the reverberant field. Unlike reverberation time, it is also a sensitive measure of energy from unabsorbed early reflections, i.e., delayed sounds that arrive first to a listener from untreated surfaces. It is also a good measure of ‘running reverberation’, which is the sound decay heard during continuous music or speech. It is apparent that cinema A has the highest levels of early-reflected sound that would decrease the apparent signal to noise ratio and intelligibility.

Figure 2 illustrates the effect of EDT on the running reverberation time at the location 2/3 back from the screen. Cinema C is clearly superior, but the differences between cinemas A and B are not so evident, despite differences in the aural character of the reverberation.

Further analysis using narrow frequency ‘boink’ pulses (5 cycles of a ramped sine wave) indicate that a "gap" in the time domain from about 25 to 60 milliseconds in theater A that is not present in theater B (Figure 3). At a location directly in front of the loudspeaker cluster, there is 5–15 dB greater early reflected sound energy in the first 400 ms (Figure 4). These factors were likely responsible for the different character of the reverberation, and may have ultimately influenced preference for cinema B over cinema A.

These analyses illustrate two points regarding spatial quality assessment. First, it is useful to consider the effects of room acoustics independent of the spatial distribution of reflections. Although we are better at localizing sound with two ears, there is a great deal of spatial information that can be assessed with one ear. Even over a telephone with low acoustic fidelity, we simultaneously receive information about the sound source, and the acoustical space occupied by it.

This acoustical space articulated by a sound source has been defined as the ‘environmental context’ in [3] and includes background sounds, reflections and absorption, both indoors and outdoors. The perceived environmental context is an additional factor that might apply to assessments of the ‘naturalness’ of spatial sound quality. Cinema C may have simply been less preferred than cinemas A and B because listeners like the ‘non-linearity’ of older ‘movie palaces’.

By listening and analyzing sounds using a single microphone, we eliminate binaural localization cues from spatial hearing and are forced to focus on the interaction of the environmental context with the sound source. Analytically, we hear more of what the environmental context is doing to the sound because of the elimination of the binaural masking advantage. The time pattern of reflections and their effect on speech intelligibility is more obvious. Perhaps it is not surprising that speech intelligibility appears to be more important than spatial qualities having to do with localization.

Increasingly, modern film mixes have more layers of music and sound effects competing for the listeners' attention. The music and sound effects reduce the dialog signal to noise level, making intelligibility more difficult. This is compounded by artificial reverberation of speech in conjunction with the natural reverberation of the room, making it still more difficult.

These examples indicate that, while preference may be articulated in terms of speech intelligibility, there are factors related to the character of the reverberation that take precedence over ‘optimal’ speech intelligibility. Analysis of single- channel data, the equivalent of one-ear spatial hearing, suggests that the presence of an environmental context is more desirable than a relatively ‘dry’ room for cinema sound, despite intelligibility measures.

5. SUMMARY

This examination of everyday listening experiences in the home and of the acoustics of cinemas suggests a future development of sound quality research that may be more ecologically based. Listeners participate in the formation of the auditory ‘art object’ that is the result of a spatial sound mix more than is currently recognized. Listeners as well as professional recording engineers adapt very well to multiple frames of reference for spatial hearing. Finally, evaluation of spatial sound quality may be based on cognitive association with a specific environmental context, or a range of environmental contexts. Perhaps future audio systems will allow listeners to participate more actively in spatial rendering than they already do.

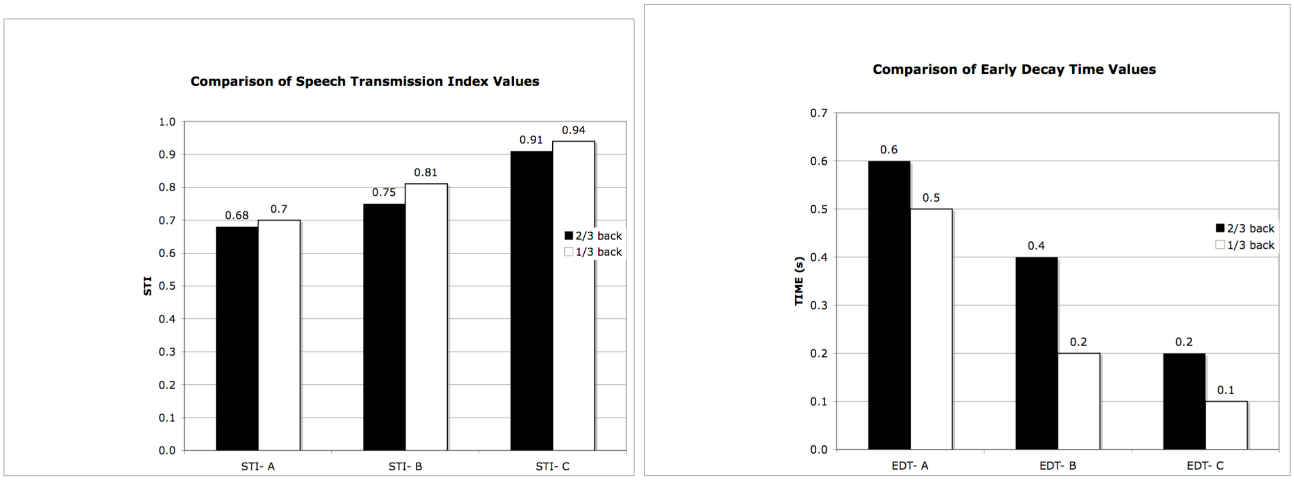

Figure 1. Comparison between measured cinemas of objective speech intelligibility (left) and of Early Decay Time (EDT) (right). Filled bars – center seat, located approximately 2/3 back into the house from the screen; Open bars-center seat, located approximately 1/3 back into the house from the screen.

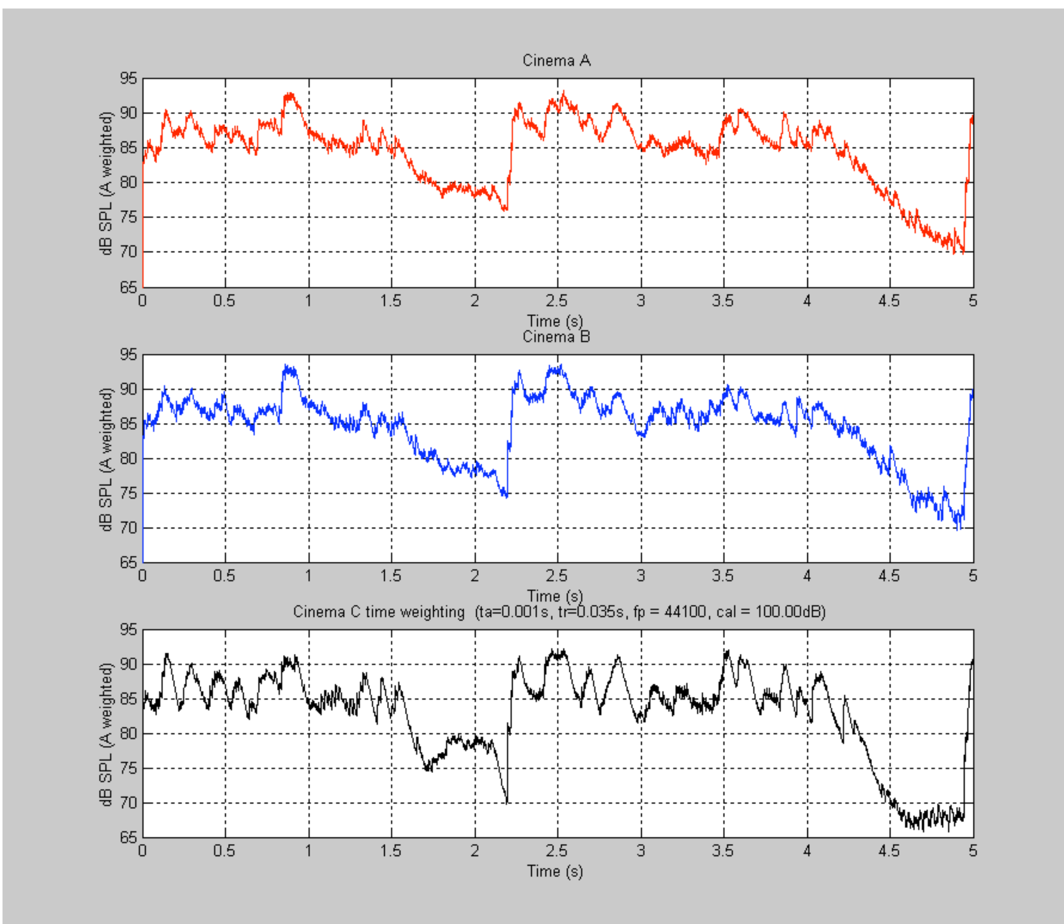

Figure 2. Comparison of 4.5 s of speech across cinemas A, B and C, using a ‘strip chart’ display with .035 s exponential decay (‘impulse’) time weighting. Microphone 2/3 back into house, center

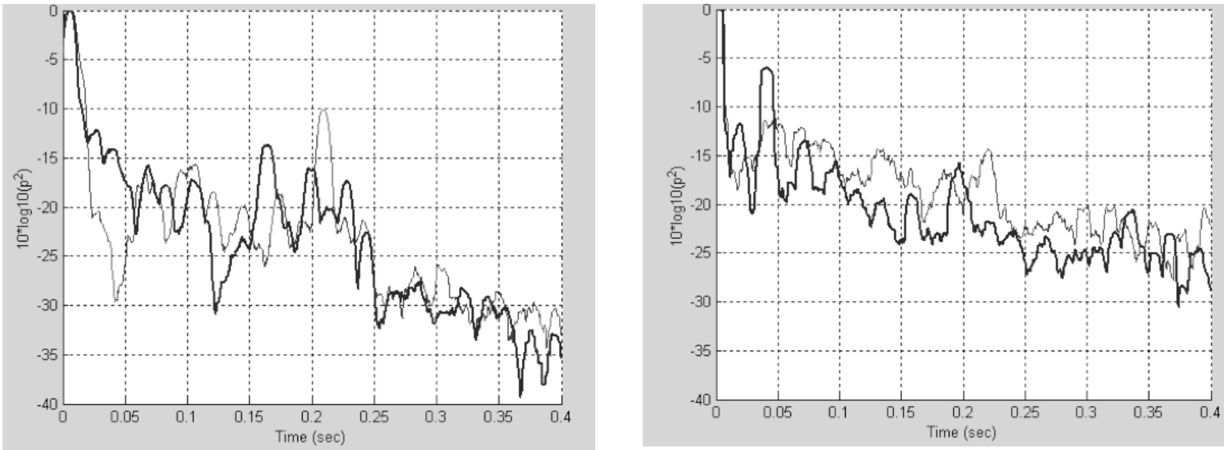

Figure 3. Comparison of narrow frequency “boink” pulse measured at seat 2/3 back in hall. Thick line: cinema B; thin line: cinema A. Left: data for 500 Hz; Right: data for 2 kHz.

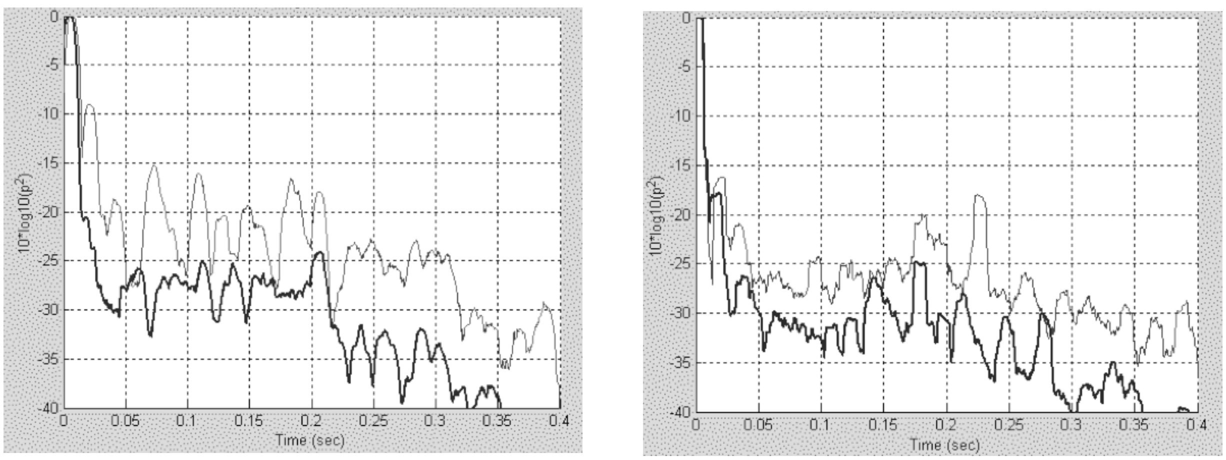

Figure 4. Comparison of narrow frequency “boink” pulse measured 10 feet in front of center loudspeaker cluster. Thick line: cinema B; thin line: cinema A. Left: data for 500 Hz; Right: data for 2 kHz

6. REFERENCES

1. Rumsey, F. (2006) Spatial audio and sensory evaluation techniques- context, history and aims. Presented at *Int. Seminar on Spatial Audio and Sensory Evaluation Techniques, Guilford, UK, 2006 April 6-7* (these proceedings).

2. Moles, A. (1966) *Information Theory and Aesthetic Perception*. Urbana, IL: University of Illinois Press

3. Begault, D. R. (1995) *3-D Sound for Virtual Reality and Multimedia* Cambridge, MA: Academic Press

4. Blauert, J. (1983) *Spatial Hearing. The psychophysics of human sound localization.* Cambridge, MA: MIT Press.

5. Begault, D. R. (1998). Virtual acoustics, aeronautics and communications. *J. Audio Eng. Soc.* **46**, pp. 520-530.

6. Schroeder, M. (1980). Toward Better Acoustics for Concert Halls. *Physics Today* **33**:24-30

7. Barron, M. & Marshall, A. H. (1981). Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *J. Sound Vibration,* **77,** 211-232.

8. Beranek, L. (1996) *Concert and Opera Halls. How they Sound.* Woodbury, NY: Acoustical Society of America.

9. http://www.muzak.com

10. Toole, F. E. (2000) The Acoustics and Psychoacoustics of Loudspeakers and Rooms- The Stereo Past and the Multichannel Future. Presented at *109th AES Convention*, preprint 5201.