9 Jan 2010

The Siegfried Challenge

David Clark

At the past few AES Conventions, Siegfried Linkwitz has presented papers claiming to have achieved subjectively **near-perfect sound reproduction** in an ordinary room with stereo source and speakers. In his latest paper he challenges the AES to study this illusion scientifically to **determine optimum playback parameters**. The AES does not directly do this kind of thing, but there might be members or their companies who would take it on.

I ran the idea past the SMWTMS hi-fi club and got enough interest to go forward. Siegfried is now on board with loaning a pair of his **Orion speakers** and helping with the design of the experiments. Competitive products and approaches will be included as well. Listening test sessions are planned to start late January 2010 and continue through March.

The basic concept of these experiments presumes that signal transmission, speaker distortion and frequency response can be sufficiently good that they need not be issues. The goal, rarely attained, is the creation of an **Auditory Scene** (AS) that is very close to that heard at the live performance. Listening room acoustics and loudspeaker's interaction by directivity, off-axis frequency response and positioning are the variables to be studied.

The initial testing will be narrow in scope:

- Live, unamplified music and speech
- Commercial recording techniques
- Living room acoustics and speaker playback
- Two-channel stereo (upmix and multi-speaker playback may be tested)
- Listeners have recent live music experience
- Listeners have ability to evaluate using AS parameters, i.e. locations of phantom images

Two stages of testing are planned; a verification of the **Existence** of a plausible AS illusion and the **Optimization** of the illusion. The Existence test can only be sighted and will rely on auditory memory of live events in the subject's experience. The tests will include the Linkwitz recommended speakers, positioning and acoustics. Optimization testing will involve repositioning and exchanging speakers behind an opaque screen and will be double blind. Optimization listening will be a comparison to a fixed reference in order to establish an anchor point and allow rapid comparison switching.

In conjunction with the Existence testing, we will have the **opportunity to introduce alternatives** to the Linkwitz Labs, Orion speakers. Also, we plan to use moderate-size box-speakers for the fixed reference in the Optimization tests.

We are looking for listeners who can spend a few hours of critical listening at David Clark's house. We are also looking for the loan of best-quality speakers from manufacturers. These would be evaluated, behind the screen, as alternatives to the Orion speakers. Please contact me at dlc@dlcdesignaudio.com. More information is available at http://www.provide.net/~djcarlst/smwtms.htm

Done and to do list

1. Define achievement of success in creating the illusion of the original sound so that we can test for it in the Existence Test. The concept of Auditory Scene (AS) has helped in this. Plausibility of the AS when compared to recent memory of a similar acoustic event rather than absolute accuracy is required. It is recognized that this is highly subjective as it depends on listener attitude and experience. Listeners will be required to pass a qualification test to participate.
2. Design Existence Test. “Behind the curtain” listening evaluation of Linkwitz Orion speakers (and others) in the recommended acoustic environment. “Is Auditory Scene recreated?”
3. Design Optimization Test. Double-blind comparison of fixed reference speaker pair to different arrangements of Orion pair (or others, all behind visually opaque screen). “Is the AS better with reference speakers or the unknown speakers?”
4. Define test results that will support the hypothesis that a plausible AS can be created.
5. Define test results that will find what parameters are important to the best AS and determine their values for the present listening room.
6. Select short segments of program material. Start with selected LiT disk tracks.
7. Invite listeners
8. Invite alternative speakers and signal processing
9. Prepare listening room, switching equipment, associated equipment and test forms
10. Run tests over a two-month period
11. Analyze and publish results

BACKGROUND INFORMATION

What is stereo?

Here, “stereo” is considered to be a two-channel audio process for playback of a recording to communicate aspects of an original acoustic event to a listener. We will consider only speaker playback in a room. The “original” may be entirely synthesized, but we will concentrate on acoustic events captured for playback.

Early audio engineers discovered that two audio channels gave a **vastly improved** sense of ensemble size and arrangement of instruments in space. “Stereophonic” means, solid sound, in Greek. Optimization of recording microphones and speaker/room characteristics has been empirical rather than based on psychoacoustic theory. Other than demonstrations and lab experiments, the first stereo was in movie theaters in the late ‘30s.

A **monophonic** audio chain captures pressure variations at a single point in space and converts this to an analogous voltage variation over time or into data representing the same thing. Even if multiple microphones are used, a single amplitude over time, two-dimensional signal results. On playback, a speaker reproduces this as pressure variation where it interacts with the acoustics of the listening room and eventually creates two slightly differing pressure variations at the eardrums. The result can be pleasant: instruments can be identified and their individual melodic lines followed. The size of the acoustic space can be identified and a sense of distance for each instrument can be conveyed. People enjoyed monophonic sound for many years in radio, phonographs, movies and television.

Stereo adds perceived width and spaciousness to the presentation. Instruments can be evenly distributed within the angle defined by the two speakers. This greatly improves the ability to concentrate on a single instrument or group at will. Width enhances clarity and definition. In addition the sensation of spaciousness of the recorded venue is greatly improved. Instead of detecting size from mono reverberation cues, one senses the space as though actually in it.

Physically, the stereo reproduction does not resemble the original acoustic event. Instead of multiple sound sources, each having direct and indirect acoustic paths to the listener, there are only two real sound sources, the speakers. For the centerline listener, a centered auditory event at playback is the result of the two separated sources producing exactly the same sound at exactly the same time. This is **impossible** in the natural acoustic world. Slight variations in timing and amplitude between the stereo speakers steer the auditory event from side to side. Any real world reflection is too delayed in time to cause this illusion. It is agreed that human hearing suppresses reflections after the direct sound arrival and thereby gains the ability to identify the direction of the true source. It seems that presentation of this physically impossible sound to the listener evokes an illusion of a single sound source localized somewhere between the two speakers.

Perceived width of the acoustic event is produced in a second way: time and amplitude differences much larger than those occurring at the two ears of a listener are captured in the recording and presented at playback. Although the difference is **diluted by each ear hearing each speaker**, the result is a somewhat crude mechanism for hearing left on the left and right on the right. It does serve to stretch the perceived width from one speaker to the other.

Our unconscious ability to suppress reflections in the room we are in not only helps localize the first arriving sound, but indirectly, it helps us localize source distance in this room. We unconsciously measure direct to reverberant ratio using the reflections that we supposedly suppressed. From the evolutionary standpoint, we have converted useless room reverberation information into useful source distance information. One result of this is that we become unaware of reverberation of the room we are in. I call this **Auditory Dereverberation** and it is very important to sound reproduction because otherwise we would hear the reverberation of the recording venue overlaid with the local acoustics. A second, less favorable, result is that we tend to localize at the distance of the true sources of the sound, the speakers.

It seems that we ignore the speakers as sources of sound if they do not have an **acoustic signature** that calls attention to them and if they both are playing recorded sounds of another space. It is as though auditory dereverberation is applied to the true sources of sound just as it is to the true listening environment. By acoustic signature I mean highly irregular frequency response or diffraction (which may be aspects of the same problem), distortion or noise. It also seems, from years of observation, that proximity to reflecting surfaces results in amplitude and timing of reflections that calls attention to the speakers. Performance standards and placement rules can be derived from these observations.

Another factor that I have observed to aid naturalness of directional localization, but at some sacrifice to distance localization, is a **strong diffuse sound field** in the listening room. The diffuse sound tends to widen images and lessen the negative effects of large differences between the recorded channels as well as soften the precision imaging found on the centerline. Distance resolution is usually compressed as well. This can be a perfect match to the auditory scene found in a good balcony seat in a concert hall. On the other hand, a lot of diffusion may make a close perspective with sharply defined images impossible even for one on the centerline.

Spaciousness is heard when delayed versions of the sound differ at the listener's two ears. This difference can be found in the recorded reverberation (which is always delayed from the direct sound). The playback speakers' angular spacing, in itself, preserves a portion of the recorded difference giving perception of spaciousness. Listening room reflections complement this effect by further randomizing pressure differences at the ears by diffuse reflections. Adding the diffuse room reflections transforms spaciousness heard as originating in front of the listener to having a more **enveloping quality**. The spaciousness hierarchy of perception is as follows: **1.** Mono reverberation defines size of recorded space and allows distance localization. **2.** Stereo anechoic reverberation adds spaciousness. **3.** Stereo with diffuse listening room reverberation adds envelopment.

In summary, stereo is audio recording and playback using two channels. When done with care, it adds ensemble width, directional localization and spaciousness to monophonic attributes. It addresses perceptual requirements by substituting physically practical mechanisms rather than attempting point for point recreation of the original sound field. "Recreation of the Auditory Scene" is a phrase that captures successful stereo. Physical acoustic accuracy is impractical and unnecessary.

Stereo is perhaps best suited for recreating the AS of an orchestra in a concert hall. The original AS has a limited ensemble width and limited directional localization of individual instruments. The hall reverberation makes the instruments louder and blended while giving the AS an enveloping character often described as "three dimensional." This AS can be captured and reproduced using two channels to an acceptable degree for critical, yet accepting listeners. The listener's "**Willing suspension of disbelief**" is a requirement for AS recreation.

Auditory Scene

In discussions with Siegfried Linkwitz (SL) on what constitutes “accurate illusion,” the phrase Auditory Scene (AS) has come up. It seems useful to adopt the language of AS to describe this subjective experience. Much of the science behind AS comes from **Prof. Albert Bregman** of McGill University, http://webpages.mcgill.ca/staff/Group2/abregm1/web/ Bregman has studied AS formation from the perspective of understanding how the auditory system organizes sound into patterns. He calls this Auditory Scene Analysis. We are interested in how we can replicate a given concert-hall AS by recording and reproducing it. We will settle for a plausible AS.

An auditory scene is the array of auditory events that our ear-brain system creates from the pressure vs. time inputs to the eardrums. The mechanism is completely different from the visual scene we perceive through our eyes. The analogy should be taken no farther than the name “scene.”

The **eye-brain system** is a massively parallel system that maps nerve cells at the retina to corresponding locations in the brain to produce a visual scene of great spatial resolution. The receptor cells, however, have poor color resolution. (Green can be perceived from either pure green or a mixture of pure blue and pure yellow.)

The **ear-brain system** is a serial system that analyses the pressure variations into 30 frequency bands. (Bass mixed with treble does not result in the perception of midrange; it is heard as a bass pitch and a treble pitch.) The AS is not high in spatial resolution compared to the visual. Nevertheless, the ear is a very powerful identification mechanism and simultaneous auditory events can be analyzed and grouped.

Two eyes working together give us improved depth perception and two ears working together give us improved directional perception. The two systems working together give us impressive awareness capability. Usually attention is focused on the visual sense with hearing in the background confirming and sharpening the events. When there is no light, hearing becomes dominant, but we tend to retain the “scene” as our map of what is going on. This is our auditory scene. It is strongly affected by **visual history and expectations**, not just sounds.

Creating an AS from two pressure variation inputs is **not trivial**. Consider individual notes played sequentially on a piano. The first note is an auditory event, but why is it one event rather than separate events for the fundamental and each harmonic? The answer:

1. The harmonics are all harmonically related
2. All harmonics started at the same time
3. All harmonics came from the same direction
4. Visual input and recognition of the sound, bias towards integrating into a single event

Now we add the sequence of notes that forms a melody. Why is this heard as an “**auditory stream**” of information from the piano rather than a sequence of auditory events? Answer:

1. Similarity of harmonic amplitude structure of new and old notes
2. Similarity of onsets
3. Same direction and distance as the earlier parts of the stream
4. Notes are not too close together in time

Now we add a second stream of information, a singing voice. How are the two **segregated** into streaming auditory events? Answer:

1. They may not be segregated if they are synchronized in time

2. Dissimilarity of onsets
3. Dissimilarity of direction and distance
4. Streams cross each other in frequency
5. Dissimilarity of harmonic structures

Now we could add chords to the piano playing and add more instruments. Still, we are able to segregate into many individual streams. In practice, composers and performers choose to elicit both segregation and integration of streams for artistic effect. All of the integrated and segregated auditory events, together with awareness of the acoustic environment, when the events suddenly stop comprise the AS. This AS can simplify when we relax our attention, or can become complex as we shift our **auditory focus** from one stream to another.

Stereo (or even monophonic) reproduction is capable of rendering a plausible or accurate AS if certain conditions are met (**The Existence test**). Subjectively, they are:

1. The speakers do not seem to be the source of sound
2. The listening room acoustics are not audible as such
3. Record/reproduction chain artifacts are sufficiently low
4. Recording captures timbre and direction of instruments and acoustic effect of the space
5. There is defocusing of critical centerline listening at playback

We would like to quantify these requirements through a series of subjective experiments. To do this, listeners must be able to identify and rate each of the 5 issues for every presentation in the experiments.

First we should listen for improvement or deterioration of the AS compared to a reference presentation. The best AS will be the same as memory of a live sound or its plausible equivalent. To judge AS credibility, one must mentally examine it fully. This will require **concentration** on changing elements. I suggest the following be judged by the listener:

1. Auditory events (images) arrayed as intended with respect to depth and side-to-side localization
2. Ability to focus on a selected stream and change focus to another stream. (Cocktail party effect)
3. Ability to segregate streams at the will of the listener, not just at the will of the artist
4. Identification of the space as to large or small, reverberant or dead
5. Freedom of head movement without shifting images or sense of pressure

We hypothesize that meeting the set of conditions (not know exactly what they are yet) will result in experiencing all of the attributes of the auditory scene in reproduction. The evaluation form for **Optimization Test** will include the AS judgments above.

Auditory Dereverberation

Let's start with without considering speakers and playback. I define "auditory dereverberation" as the psychoacoustic **acclimatization process** of suppression of reverberation. I mean the real space the listener is in at any time (the "live" space) , not the playback of a recording made in another acoustic space (the "recorded" space). This process can be observed by focusing your attention on what happens as you enter a somewhat reverberant space for the first time. At the moment you enter, you are aware of the acoustics and you tend to look around to connect what you hear with what you can see. For a reverberant space, voices at a distance may be unintelligible at first. After a while, you usually loose awareness of the reverberation and intelligibility improves. Our hearing was developed over millions of years as a **survival mechanism**. Imagine the consequences of entering a cavern and hearing footsteps of a predator as coming from every direction. The survivors keyed in on the first arriving sound, determining its direction, distance and taking appropriate action. Echoes, reflections and reverberation were suppressed in the process.

We are now using the same hearing mechanism for the pleasure of listening to music in reverberant concert halls. We also listen to stereo playback of music from those reverberant concert halls. While we have empirically figured out how to make good recordings for playback (move mikes close and get lots of separation), we have not engineered the process. I believe that understanding dereverberation will lead to better understanding of the record-playback process.

Let's look at what is lost in the dereverberation process and what is gained. Of course, considerable conscious awareness of reverberance is lost, particularly when listening to speech and complicated music. (This can be demonstrated by making a single channel recording using an omnidirectional microphone in the concert hall and playing it back over a single speaker outdoors or in an anechoic environment. The recording will be heard as highly reverberant.) Reverberation makes the **sound louder** even though it is suppressed as such and the ability to localize distance is greatly improved. We may hear the images as less compact, but we can still identify the exact direction. We also experience envelopment of sound, but not of image directions.

I have observed **two extremes of stereo reproduction** from systems using competent components; dominance of direct sound and dominance of room reverberated sound. The first might be called the "High-End" system commonly found at the Consumer Electronics Show. Typically, the black vinyl and tube amplifiers drive a pair of skinny 2-way speakers positioned way out from the walls and a single listener sits exactly on the centerline, quite close to the speakers. What I hear is precise imaging within the angles defined by speaker separation. Depth localization is very good and so is identification of the acoustics of the recording. What bothers me is that much of this illusion collapses, shifts or changes timbre when I move slightly or rotate my head. This creates a sense of pressure, particularly for front-center instruments. Also, while I hear reverberation, I do not sense 360 degree envelopment.

The other extreme is typically found with wide directivity speakers on the other side of a large reverberant room. Dominance of playback reverberation leads to imprecision of imaging in azimuth and depth, in my experience. Also, the room cannot be dereverberated such that the recorded acoustics come through.

Between these extremes, there is a range of personal taste for a more "in your face" experience or a more relaxed and distant perspective. Within this central range, it is my opinion that the Auditory Scene should be dictated far more by the recording than by the playback acoustics.

Simple Dereverberation Experiment

Last Summer, I visited Michigan State professor William Hartmann who is a well known psychoacoustics researcher. I ran my dereverberation theory by him and he noted that some aspects had been confirmed by published experiments. He suggested we do a little experiment right there. His anechoic chamber and reverberation chambers are separated by a large and quite reverberant work area. (I had noted my usual adaptation and dereverberation to this room when I entered.) We made recordings of the professor reading text at a distance of one meter in the work area, then again in the anechoic chamber.

When we played the workroom recording back in the anechoic chamber (over a Minimus 7 speaker on a stand placing it at head height), we heard excess workroom reverberation that was not present when we listened to his voice live in the workroom. The voice sounded artificial and distant even though there was no added reverberation from our anechoic playback room.

Next was playback of the anechoic recording in the workroom (and comparison to the real Dr. Hartmann). The Minimus 7 was identifiable as a loudspeaker source, but all seemed quite natural with no excess reverberation. Far more natural than the same amount of reverberation heard in the anechoic chamber. Experiment's conclusion: recorded reverberation is far more audible than live reverberation, at least for a simple recording technique.

The practical implication of this is that the superb stereo (or multichannel) recordings we hear did not capture the natural sound at the live event and play it back for us in a natural way in our listening room. The real process is as follows:

1. The original recording is made with microphones very close to the orchestra which gives a high ratio of direct to reverberant sound ratio. (A practice refined by trial and error by recording engineers over decades.)
2. On playback, the performance space cannot be dereverberated so the perceived ratio of direct to reverberant is restored to that of the excellent live seat.
3. The playback environment is dereverberated which makes playback acoustics much less important than previously thought.

26 Dec 2009

Evaluation Form for Auditory Scene Reproduction

The auditory scene (AS) is the mental picture of what is going on around us based on the sound we hear. For live sound, this scene is dominated by sight. Closing our eyes results in an AS that is a result of what we hear and partly what we remember from the visual scene. In sound reproduction, we want an AS illusion that we are listening to the live event. In fact, we are listening to two speakers in a small room. When we close our eyes, we want this reality to be replaced by an accurate AS illusion.

With this testing, we hope to answer the question, Is this possible? Since the AS illusion depends heavily on having experienced the AS of live sound we require that the evaluator has recently attended live music (unamplified) in an appropriate acoustic venue. We expect that the listener was attentive and made notes of the live AS. Also, an attitude of “willing suspension of disbelief” is required.

The form below lists requirements for a successful AS illusion at playback. You are asked to listen to each of the segments at each of two listening locations to form an overall score for each of the attributes. (A reference speaker system known to be poor at generating the AS illusion is provided by the “A” switchbox option. “B” is the system under test.) If you find it more convenient, you can mark each attribute as you go by a vertical line crossing the bad to good rating line. We will average the scores, or you can use a new sheet, mentally average all scores and make a single new mark for each attribute.

Written comments are encouraged.

1. **Speakers disappear** Sound is independent of speakers	0 At speaker	5 Some depth	10 Disappear
2. **Local acoustics** Sound of the room you are Listening in	0 Strong	5 Blended	10 Recorded
3. **Phantom Images** Not looking for precision localization	0 None	5 Uneven	10 Behind and between
4. **Proportion of ambience** “dry” to “wet”	0 None	5 Suppressed	10 Full
5. **Ambience non-localized** Recorded venue reverb	0 At speakers	5 Frontal	10 Surrounding
6. **Freedom of movement** Listener can rotate head or move sideways	0 Lock to centerline	5 Acceptable changes	10 Full

16 Dec 2009

Evaluation Form for Auditory Scene Optimization

The previous auditory scene (AS) testing used what was believed to be the best conditions for reproducing a plausible AS. Now, we wish to see if we can improve upon the source material, playback room, speakers and placement. On the other hand we may find that certain parameters have little effect on creating a plausible AS illusion.

This test asks you to rate qualities of the AS for different playback setups. The first question asks if the basic requirements for an AS illusion are met. These are: speakers disappear as sources of sound, local acoustics suppressed, phantom images between and behind the speakers. The next questions ask you to listen to aspects of the AS that further determine its quality. You are asked to compare each issue to the reproduction by a high-quality reference. This reference will not change and you are asked to rate the configuration under test on a -5 to +5 scale compared to the reference.

Written comments are encouraged.

1. **Basic AS requirements** Speakers disappear. Dereverberation Phantom images	-5	0	5
	AS Lacking	Equals reference	Equals Live
2. **Cocktail Party Effect** Ability to follow different strings at will	-5	0	5
	Poor	Equal	Superior
3. **Close phantom Image** Soloist “in your face when the recording calls for it	-5	0	5
	Distant	Equal	Closer
4. **Distant phantom image** Most recordings intend some Images to be distant	-5	0	5
	Closer	Equal	Distant
5. **Ambience** Should be non-localizable and surrounding	-5	0	5
	Poor	Equal	Superior
6. **Freedom of movement** Listener can rotate head or move sideways	-5	0	5
	Lock to centerline	Equal	Full