Hearing Spatial Detail in Stereo Recordings

(Hören von räumlichem Detail bei Stereo Aufnahmen)

Siegfried Linkwitz

Linkwitz Lab, sl@linkwitzlab.com

Abstract

In natural hearing we sense direction, distance and surrounding space of a sound source. In stereo playback over loudspeakers at +/-30^0 we perceive real and phantom sources. Auditory cues for source direction, distance and surroundings are present, but are often corrupted by the loudspeakers themselves, by their placement in the room and by room reflections due to their radiation pattern. Loudspeakers and room can be perceptually hidden from attention. Reflections must be sufficiently delayed and mimic the direct sound by having the same spectral content. What remains is a 3-dimensional phantom auditory scene in front of the listener, behind the loudspeakers. Such scene can be studied with confidence to determine spatial plausibility and believability of a stereophonic recording/mix. Loudspeakers with constant directivity, such as omni, dipole or cardioid, are necessary to realize this benefit.

1. Introduction

The recording and reproduction of sound in stereophonic format has been practiced for many decades. In its simplest implementation two microphones pick up the sound at one location in a specific environment. The two electrical signal streams are stored on a CD and played back at a later time, in a different environment, over two loudspeakers. A listener perceives sounds coming from the loudspeakers and from between the loudspeakers. The room exerts a strong influence upon the clarity and tonal balance that is heard. More than two microphones will have been used in most recordings and their electrical output signal streams will have been mixed down to two channels. The process of placing a multiplicity of microphones in specific locations and then combining their output signal streams, as the program material demands, has become an art-form. It is guided by the sensibilities and expectations of producer and consumer. Recording decisions are based upon what is heard from the monitor loudspeakers in their acoustic environment. Typically there is little resemblance between a recording studio and the playback room of any consumer of the recording, Fig. 1.

Sound always exists in a space [1]. We hear both the direct sound from a source and the response of the space to that source via multitudes of reflections. As a species we have developed the ability to distinguish direct sound and reflections generated by one source from those of another source, which may exist in the same or in a different location. Our brain continually analyzes the streams of air pressure variations at the eardrums for patterns that match memory, or could be new, and must be responded to by motion. We have memory of the natural Gestalt of sources and even in different environments, or when we receive few cues at the eardrums, do we recognize the source. [2] – [5]

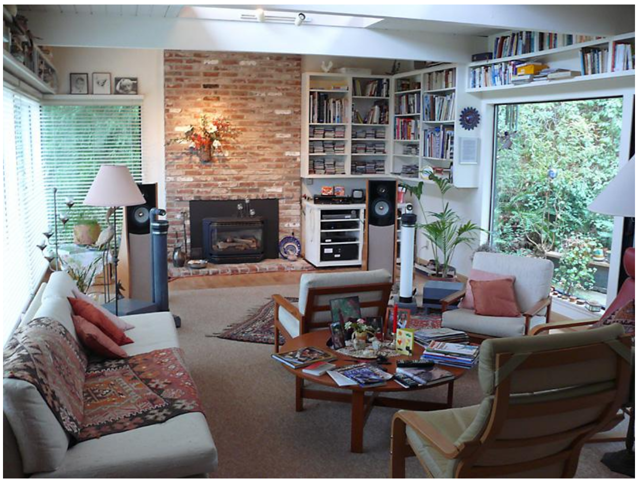


Figure 1: My living and listening room.
No special acoustic treatment is provided, just the normal "stuff of life".

In the transmission of sound from an acoustic event to the microphones, to loudspeakers and listener many natural cues are lost and artificial, misleading cues have been introduced. In particular, the spatial relationship of sources to each other and their environment is rarely recorded in a natural way, or believably reproduced over the typical loudspeaker setup in a typical room. Microphones have been capable for a long time to sample the full spectrum and dynamics of audible sound. Loudspeakers can be readily built that cover the full spectrum of sound and with dynamics greater than enough to overload an acoustically small space. How to record and how to reproduce spatial detail in stereo will be discussed below from the perspective of a loudspeaker designer. I listen to a wide range of music, though mostly classical, in my own living room, Fig.1. I believe that many recording engineers and also many loudspeaker designers are not aware of the wonderful illusion that we can create with simple two-channel reproduction, when we minimize misleading data streams at the ears and cooperate with natural hearing processes in the brain. Stereo in its purest form is about recording real sources and reproducing them over two loudspeakers in a room as phantom sources, without hearing room and loudspeakers, while generating a believable illusion of sound sources in their spatial context. Timbre, localization and spaciousness of the illusion are essential contributors to a satisfying auditory experience. They should be preserved from recording to reproduction.

2. A systematic approach to stereo

This will be a view of stereo, as if I had the assignment to record an acoustic event like I heard it and then to reproduce it in my mind as simply as possible at a later time. The design solution should be practical, convenient and cost effective.

2.1. Binaural recording and playback

2.1.1. Recording sound pressure streams at the eardrums

The first decision in my assignment will be to select the location from where I hear the acoustic event. If I were to record a concert in a symphony hall I would find "the best seat". A listener to my recording would have an acoustically familiar experience, assuming he/she is used to hear the orchestra from an audience perspective, because he knows already the Gestalt of such events.

I would record the sound pressure variations at my ear drums while facing the orchestra and holding my head as still as possible [6]. Playback would be over headphones that are equalized to recreate at my eardrums the exact air pressure variations from the recording session, Fig. 2. Such equalization is not trivial but can be accomplished with a DSP engine.

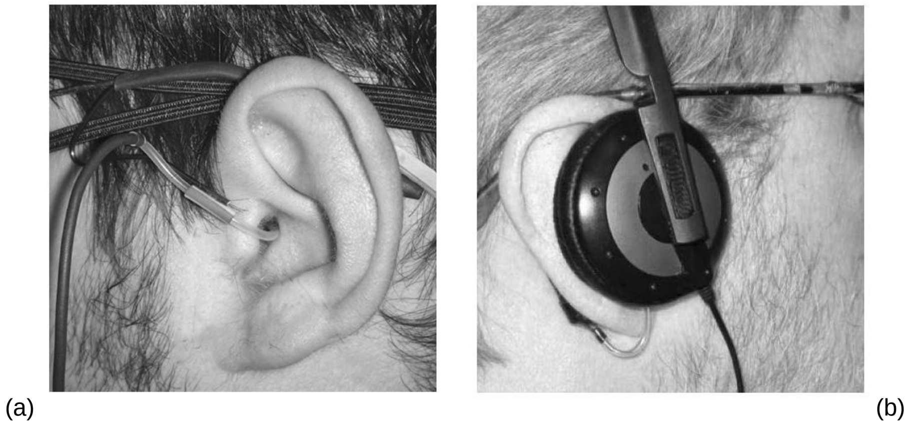


Figure 2: (a) Recording the sound pressure streams at the eardrums using a small diameter, flexible tube with a small microphone capsule at its remote end. (b) Headphone reproduction after the signal transmission chain has been equalized using the same tube microphone.

2.1.2. Hearing the reproduced eardrum signals

What will I hear when my eardrum signals from the recording session have been reproduced? My brain will create from the signal streams an Auditory Scene that has an extremely high degree of similarity with what I heard originally. When I close my eyes I may think that I am back in the concert hall. The tonality of the auditory scene is the same. Instrument sounds and noises are coming from the directions that I remember, though may seem to be closer. But there are significant differences in the behavior of the reproduced Gestalt. The auditory scene does not change when I move my head. It moves, tilts and walks with me without changing Auditory Perspective. I cannot turn my head towards the noisy visitor from the recording session. Nor can I tune him out as I did then. The eminently important ability to adjust my Auditory Horizon has been lost. Thus, I may now become aware of air conditioner noise that I did not remember hearing and possibly other intrusions into the auditory scene. I may notice the absence of tactile sensations.

2.1.3. *Other individuals hearing my eardrum signals*

What are other persons likely hear when they listen to my eardrum equalized recording using the same headphones? First of all they have not experienced the auditory scene of the recording session, though they may be very familiar with such Gestalt. They are likely to be impressed by the realism and believability of what they hear. But their brain will also receive cues that are misleading to the degree that my head shape, pinna, ear canal and torso differ from theirs. Thus the tonality of their auditory scene may be different at high frequencies. Source directions could be somewhat changed but, most significantly, source distances will be foreshortened. Out-of-head localization will be difficult in front, though relatively easy to the sides, rear and top. Even if a corresponding visual scene is presented simultaneously to the auditory scene, the Auditory Distances seem shorter than the visual distances [7].

2.1.4. *A head-tracking demonstration*

The reduction of Auditory Distance between listener and sources within the auditory scene is a normal perception, whenever the signals at the eardrums are independent of head movement. In the limit, when proper auditory cues are missing, the auditory scene will be placed inside the head, between the ears. I have experienced a demonstration that illustrated the importance of head turning to the perception of distance [8], [9]: A dual mono presentation of a jazz quartet over headphones produced a monaural auditory scene between my ears. A movement tracking device on top of my head was turned on, like in Fig. 3. A few left-to-right turns of my head and suddenly the auditory scene had moved clear across the room to the door that I was looking at. It was still of the same size and timbre as perceived initially, but now at a large Auditory Distance, floating in the room in front of the distant door. I turned a full 360 degrees and the AS was perceived from all angles as being in front of the door across the room. Now as I was looking towards the distant auditory scene the head tracking device was turned off. Nothing happened. The auditory scene was still over there. A left-to-right turn of my head and still nothing happened. The jazz quartet continued to play over there. A few more movements of my head and suddenly I perceived the auditory scene sliding towards me, accelerating and locking into my head between the ears.

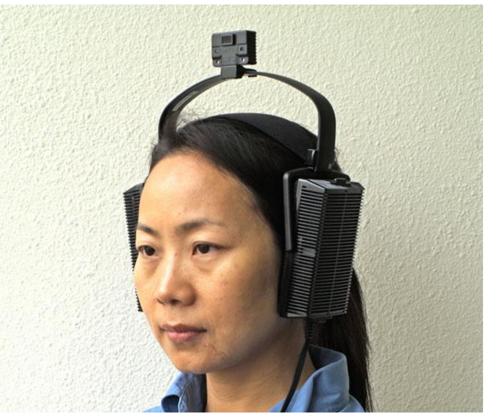


Figure 3: Example of a motion tracker on top of the head band (Smyth Research)

What had happened here is that first the frequency response of the headphone signals had been continuously changing according to the direction that I was facing. The initial system calibration was for the door across the room. As I moved my brain became trained by consistent directional cues for the changing auditory scene, which it memorized. Thus, upon turning off the head tracking device, nothing happened initially. Only after further left-to-right turns, which repeatedly produced inconsistent auditory cues, did the brain give up and place the auditory scene between the ears, because it could no longer locate the auditory scene outside the head.

2.1.5. Head diffraction and directional hearing

The human head diffracts a sound-wave depending upon its frequency and angle of incidence. At low frequencies and long wavelengths the head is a small obstacle and only the arrival time and phase difference between the signals at the eardrums is of usefulness for directional hearing (ITD). At high frequencies, where the wavelengths are small compared to the head size, it can block the soundwaves producing level differences at the eardrums that vary with the angle of wave incidence (ILD). Furthermore the head shape and the particular details of the pinna shape and the ear canal act as filters, which shape the spectrum of the incident wave before it reaches the eardrum. Directional hearing ability is weak in the transitional range between 700 Hz and 3 kHz. Much of our directional hearing can be described by the Head-Related-Transfer-Function. This function depends to varying degree upon the individual anatomy. It changes for close source distances and differs to some extent between free-field and diffuse-field sound incidence. The HRTF behavior is inconsistent with the observation that the tonality of a sound that I hear is largely independent of the angle of sound incidence. For example, a person is talking to me and I turn my head 90 degrees. The ear drum signal spectra change dramatically, yet the person sounds essentially the same [10].

The head-tracking demonstration showed me the importance of head movement to find the auditory distance and direction for the auditory scene in my head relative to the outside world. Head-tracking is essential for creating an Artificial Reality and is sometimes used in arcade video games. If we could add head-tracking to the binaural recording and reproduction process, then we would have a system that works essentially independent of any individual’s anatomy. It would become easy for the brain to overcome remaining misleading cues in order to experience a believable auditory scene, especially when combined with visual cues. Only tactile sensations would be lacking, unless they were recorded and reproduced also.

2.1.6. Trade-offs in binaural recording and playback

A binaural system is capable of exceptional tonal and spatial fidelity, if the system response is calibrated for a specific individual.. Head-tracking is less needed. Source distance cues are usually present in the signal streams, in their envelope and amplitude. Though normally we turn our head towards a new sound to find or confirm the direction and distance from which it is coming.

What has been described so far is not a very practical system and must be simplified. Inevitably this will lead to compromised performance. First of all, the recording can be made with an artificial head and torso in place of a live person, Fig. 4(a). The shape and dimensions are the average of a large number of individuals. The mechanical behavior of the construction material for the outer ear matches that of a real ear. Microphones are coupled to the eardrum locations. Recordings may be played back using supra-aural headphones, which are acoustically open and less subject to ear cavity resonances. Their frequency response is nominally flat, but that varies with individual users. It is relatively easy, though, to equalize them and to remove the worst resonances. Listening to a recording reveals a highly realistic auditory scene. Tonality is correct if the broad ear canal resonance has been equalized with a notch filter. Sound sources are perceived in their real angular directions, forwards, to the side and above. There is a real sense of continuous space in which the sounds occur. But there is a problem, which to me makes headphone-listening an unsatisfactory experience: Distances to the sources are foreshortened and frontal sources are located inside my head.

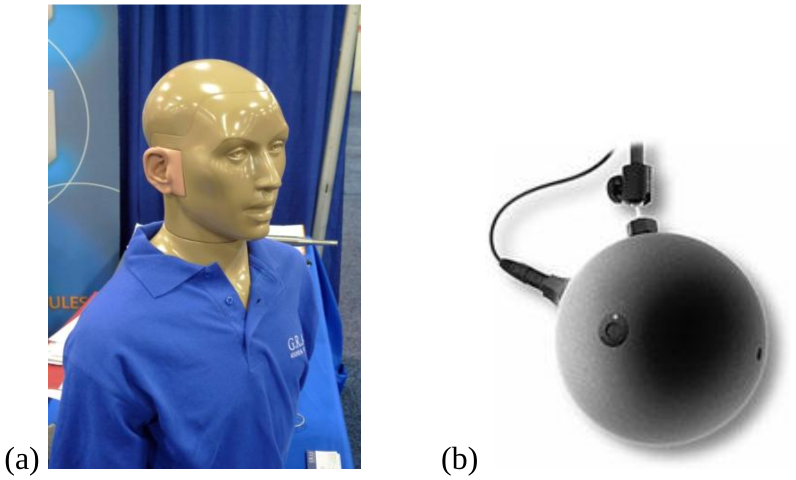


Figure 4: Artificial head (a) and sphere microphone (b)

Binaural recording can be simplified while preserving correct tonal and directional cues by placing the microphones at the entrance to the ear canal, which they block. This removes the ear canal frequency response from the recording but preserves the spectral cues that the pinna imparts at higher frequencies upon sounds from different directions.

2.1.7. *The sphere microphone*

In a further simplification the artificial head is replaced by a sphere of 17.5 cm diameter, with microphones flush mounted to where the ear canal opening should be or at 180^0 to each other, Fig. 4(b). Now, without a pinna, there is no differentiation between sounds arriving from the front or rear hemispheres. Intensity and particularly timing differences between the two microphone signals still carry a strong resemblance to a real head. Excellent natural sounding recordings can be made with a sphere microphone not just for headphones but also for loudspeaker reproduction. I am always amazed by the amount of realistic information that the brain is able to extract from sound streams, even when many tonal and directional cues are missing.

2.2. Loudspeaker presentation of a stereo recording

2.2.1. Localizing sound from a single loudspeaker

Stereo recordings are played back over two loudspeakers in rooms of various sizes and shapes. Let us first listen to a single loudspeaker in a room, playing back, for example, the same single microphone recording of a jazz quartet that I had heard in the head-tracking demonstration, Fig. 5.

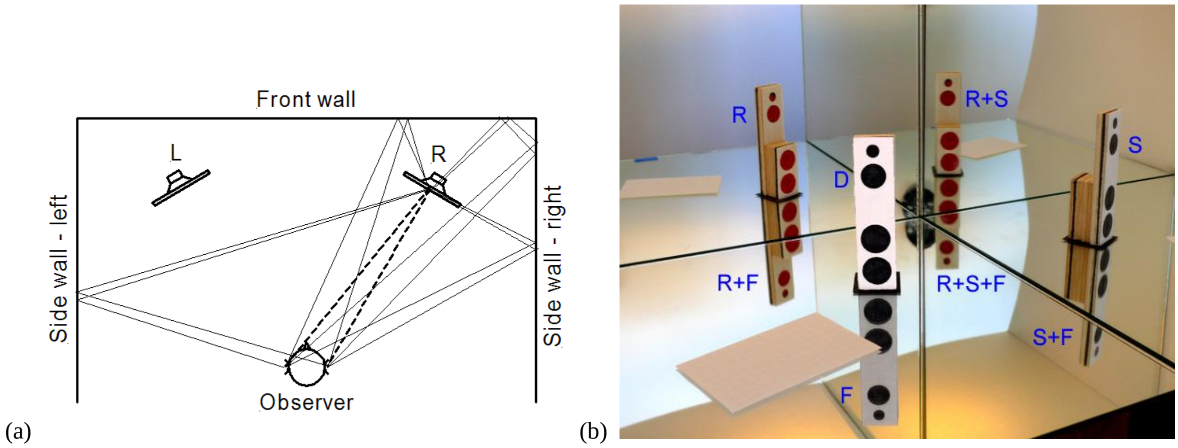


Figure 5: Direct sound and reflections from a mono loudspeaker in a room (a). Model of a dipole loudspeaker near a room corner and its reflected images (b).

The sound unmistakably originates from the single loudspeaker in the room, outside of my head. Even if I cannot see the loudspeaker I can localize it, especially when it is free-standing in the room. The distance from me to the jazz quartet sound is the same as to the loudspeakers and possibly more. As I hear the reverberation of the instrument sounds in the recording venue, I have a sense of space surrounding the instruments and also of depth. If the group had been recorded in an anechoic chamber, I would hear them as sounding very dry. The width of the presentation is too narrow for the quartet, but would be just right if the loudspeakers were small and reproduced a single voice. The acoustic center of the loudspeaker should be at the height of the listener's ears for the sound to appear coming from the front and not from above or below. We recognize source height and distance by the floor reflection and other cues.

2.2.2. Frequency response of the mono loudspeaker

What should be the frequency response of the loudspeaker? If I want to accurately reproduce what the microphone picked up, then the frequency response must be flat and cover at least the same frequency range as the microphone did. Certainly this must be the case for the signal travelling directly from the loudspeaker to my ears. If I do not want the sound to change, as I sit or stand in different parts of the room, then the loudspeaker must radiate uniformly, with the same flat frequency response, into all directions. The loudspeaker must be an acoustically small source to exhibit such behavior. A pulsating sphere, a monopole, is "the mother of all loudspeakers". An omni-directional loudspeaker is the practical implementation of such monopole source. Uniform radiation into all directions will cause a multitude of room reflections and reverberation of the radiated signal, Fig. 6. The response to a loud handclap reveals much about the room's acoustic Gestalt. An impulse radiated from the monopole is used to describe the room mathematically in time and frequency domains. I am convinced that we perform such room acoustic analysis pre-consciously, upon entering a new space, using whatever sounds are present. It is a survival mechanism that allows us to quickly determine direction and distance of surprising sounds amongst a multitude of reflections and other sounds. We hear the monaural loudspeaker source and find that it is located in a reverberant space. We adjust our acoustic horizon to what requires attention. Thus we can also hear the spatiality in the monaural recording if it contains cues from its venue. For source localization we perceptually differentiate between direct and reflected sounds depending on their strength and delay. We mostly ignore the multiplicity of images of the mono loudspeaker in the room surfaces, and thereby move the room beyond the auditory horizon. [11] – [14]

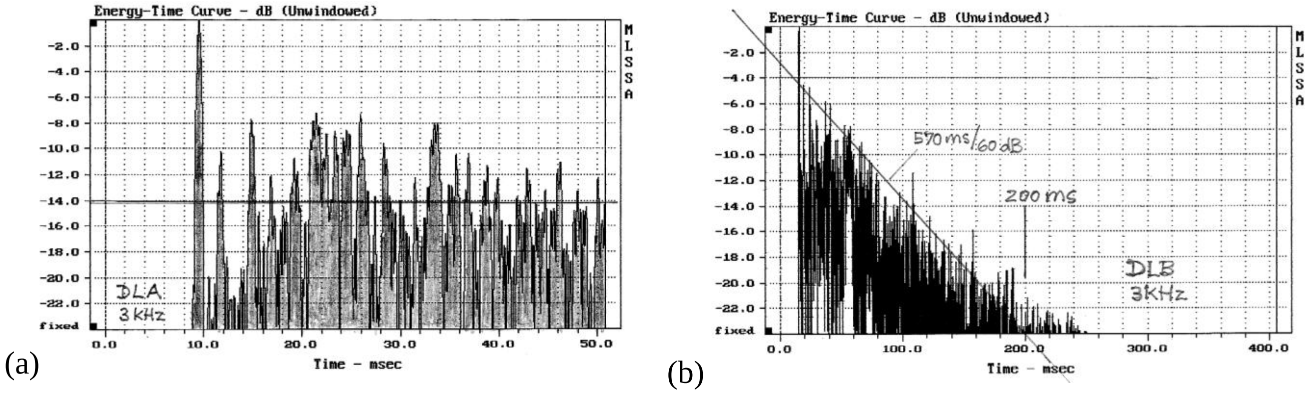


Figure 6: Direct and reflected signals for a 3 kHz toneburst from the left dipole loudspeaker in Fig.1 and Fig.14. (a) At location A during 50 ms. (b) At location B during 400 ms. [14]

2.2.3. *Realism of the center phantom source*

Now let me add a second loudspeaker and play the jazz quartet in dual mono. The two loudspeakers and listener are arranged in the form of an equilateral triangle. The loudspeakers are seen at +/-30^0 from the center axis, Fig. 7.

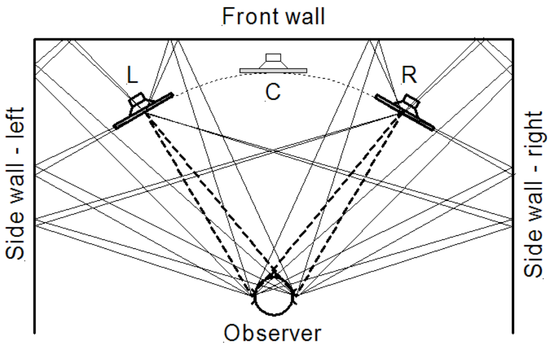


Figure 7: Monaural phantom source between two loudspeakers

I hear the jazz quartet in front of me, though less focused than when coming from the single loudspeaker. Moving my head a small distance to the left readily moves the quartet into the left loudspeaker. A movement to the right moves the quartet into the right loudspeaker. Turning my head left-to-right neither affects the positioning of the phantom source nor its tonality. It is perceived like a real source even though the eardrum signals do not change in the exactly same way as those from a real source would, like from a center loudspeaker, in front of me when I turn my head. The brain is working with a multiplicity of misleading cues and makes sense of them. Thus the distance of the phantom source is based on distance cues that the left and right loudspeakers provide themselves by their direct signals and room reflections. These cues place the phantom source slightly behind a line connecting the loudspeakers. I have observed occasionally that highly directional loudspeakers in a highly absorptive environment with low reflections can produce a center phantom source, a female voice that floats in front of the line between the loudspeakers. In an anechoic room the center phantom source can even manifest inside the head, which mimics headphone listening.

Clearly then, two loudspeakers can produce a perceptual event that has no precedence in nature and evolution. The brain adapts to the situation by comparing the novel auditory cues to familiar ones. Two identical signal streams arriving from symmetrically placed sources at both ears can only mean that there is a sound source half-way between the loudspeakers, even when we cannot detect direct signals coming from that direction. We can distinguish, though, the phantom center source from a real source, a center loudspeaker. The phantom source is less focused and has a different tonality. The reason is acoustic cross-talk between left and right loudspeaker signals at the ears and secondly, spectral coloration due to a 30 degree angle of sound incidence versus 0 degree for the frontal source. Monaural pink noise clearly shows a comb filtering effect that occurs slightly to the left and right of the “sweet spot”. It has been called the fatal flaw of stereo, which would be correct to say if we only listened to unnatural test signals [15]. Since we know the Gestalt of human voice, violin, piano, etc. the brain fills in and that flaw is rarely noticed.

2.2.4. *Crosstalk cancellation*

The difference between a phantom center and a center loudspeaker can be corrected with a cross-talk cancelling circuit in the electrical signal path. The left loudspeaker produces a desired signal at the left ear and an undesired signal at the right ear. The right loudspeaker channel is programmed to produce a desired signal at the right ear and simultaneously a signal that cancels the undesired signal from the left loudspeaker at the right ear. Since this combined signal is also transmitted to some degree to the left ear it will be cancelled by a corresponding signal from the left loudspeaker, and so on. Due to phase shift errors this process only works over a small volume around the “sweet spot”. Other locations in the room receive signals that sound unnatural, Fig. 8.

The cross-talk issue can also be resolved mechanically. The two loudspeakers are moved as close together as possible. A large sheet of plywood is placed between the loudspeakers extending to the nose of the listener in front of and at some distance from the loudspeakers. The large panel blocks signals from the left loudspeaker reaching the right ear and vice versa for the right loudspeaker. This setup behaves essentially like listening to headphones, but from a distance and without in-the-head localization. [16] [17]

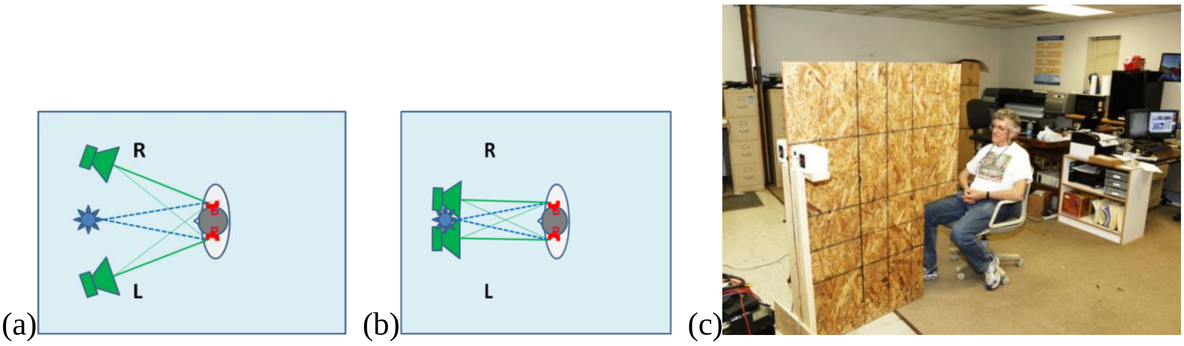


Figure 8: Crosstalk cancellation. Normal stereo setup (a). “Stereo Dipole” setup for crosstalk cancellation and a wide phantom sound stage (b). Setup with mechanical barrier, which replaces the electrical compensation of the loudspeaker drive signals (c)

Electrical or mechanical cross-talk cancellation schemes have not found wide adoption in the market. Many audiophiles are quite willing to live with the flaws of stereo, particularly since the brain can accommodate them knowing the Gestalt of natural acoustic sources and events.

2.2.5. *Frequency response of the stereo loudspeakers*

The sound streams at the eardrums are spectrally colored by the angle of sound incidence due to diffraction, as described by the measured HRTF. The center phantom source perception is based upon a 30^0 horizontal HRTF, while a real center source would have a 0^0 HRTF. Thus it would seem that the loudspeaker frequency response should be corrected to emulate a 0^0 sound incidence.

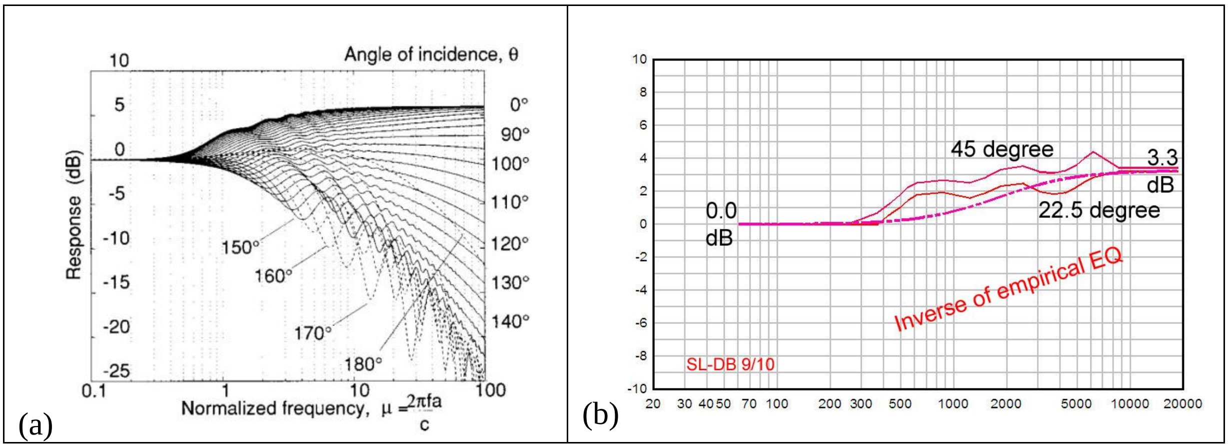


Figure 9: Sound pressure at a fixed point on a rigid sphere for different angles of sound incidence (a). Sound pressure at 22.5^0 and 45^0 incidence relative to 0^0 incidence for a 17.5 cm sphere (b). The dotted curve is the inverse of the loudspeaker equalization.

The HRTF is a complicated function and not easily manipulated, but it contains strong general trends, which can be studied on a spherical model of the human head. For a 17.5 cm diameter rigid sphere and comparing a 30^0 incidence to a 0^0 incidence, we observe that the sound pressure at high frequencies is 3.3 dB higher than at low frequencies for a point on the sphere where the ear canal entrance would be. [18] – [21] The Sphere-RTF has a broad and irregular transition region between 400 Hz and 7 kHz, Fig.9. Empirically, a simple -3.3 dB RC shelving-lowpass filter centered at 1.8 kHz provided the right amount of high frequency attenuation to an otherwise flat and near constant directivity dipolar loudspeaker, in order not to sound overly bright and to increase focus and depth of the phantom source [22]. The particular filter was found to work optimally for this loudspeaker in several different rooms and for different listeners. Both commercial recordings and known test recordings confirmed the desirability of the equalization for greater naturalness of the auditory scene.

2.2.6. *Controlling phantom source placement*

The monaural phantom source can be panned between left and right loudspeakers by changing either the relative volume level of the loudspeakers or by adding delay to one or the other channel, Fig. 10. The Duplex Theory of directional hearing, in which inter-aural time differences (ITD) dominate at low frequencies and level differences (ILD) at high frequencies, partially explains the phantom source behavior. Also, level panning affects the tonality of the phantom source, because the HRTF amplitude changes at the ears. In delay panning only the phase is shifted at the ears of a centrally positioned listener. Source location panning is the standard method for producing a recording where many microphones have been used to record individual instruments, voices or groups. This technique captures acoustic sub-spaces around individual sources, which cannot be readily combined into a natural and believable sense of space around the performers. Also depth is usually missing as phantom sources are layered upon each other. [23] – [25]

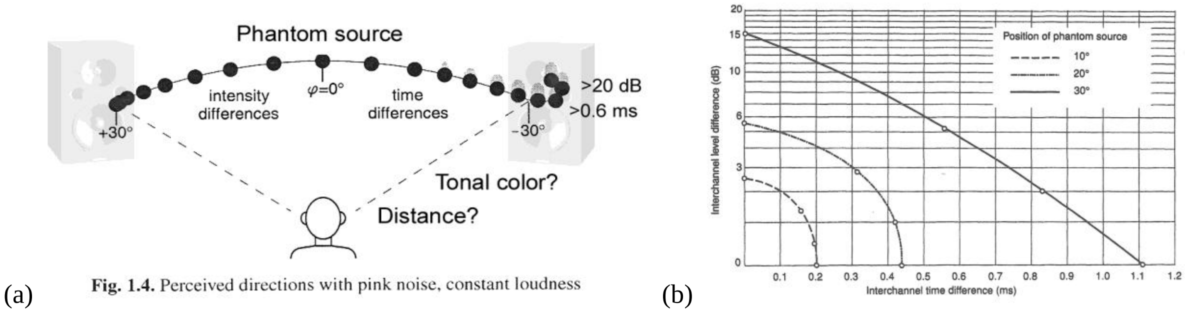


Figure 10: Amplitude or delay panning of a source to any position between two loudspeakers (a). Amplitude or time differences required to pan a source to 10^0, 20^0 and 30^0 off-center (b).

2.3. The room and its effects upon the stereo presentation

Stereo loudspeakers are typically listened to in rooms, not outdoors. Loudspeakers are usually designed for a flat on-axis frequency response, but when measured in a room at the listening position and over a 50 ms time window, they rarely exhibit a flat response. The reason is, of course, the reverberation of the loudspeaker emissions as sound is reflected from surfaces and objects in the room and as modal resonances build up. Often it is naively assumed that the response should be simply equalized to flat at the listening position using a DSP device, but this ignores the processing capabilities of our brain. We are very familiar with detecting and drawing information from sounds in reflective spaces and must therefore be careful how we interfere, if we want to hear a natural sounding auditory scene with minimal brain processing of artifacts and therefore untiringly. Reflections can be useful to stereo sound reproduction, if they have common place characteristics. We like neither an anechoic chamber nor a reverberation chamber for listening to music or voice. Between those two extremes and more towards the live end around RT60 = 500 ms, lies the optimum, but there is more to reflections than their reverberation time, particularly in acoustically small spaces like most listening rooms. [26], [27]

2.3.1. *Hiding the room perceptually*

A single loudspeaker for monaural reproduction should be omni-directional so that it sounds the same from every location in the room. Not only the direct signal from the loudspeaker exhibits a flat frequency response at every location, but the reflections and the reverberant soundfield are also essentially frequency independent, being copies of the direct sound. The response in a 50 ms time window, as measured with a microphone at a specific point in space, will not be flat due to interference between direct, reflected and standing sound waves.

If two stereo loudspeakers with frequency independent radiation pattern are placed in the room such that they are at least 1 m distant from large adjacent surfaces and objects, then the reflected sounds will be more than 6 ms delayed compared to the direct sound at the “sweet spot” for listening. It appears that such time gap is sufficient for the brain to lift the auditory scene stimulated by the direct left and right loudspeaker signal streams from the streams of room signals that also impinge upon the ears. The loudspeaker and listener configuration must be symmetrical with respect to left and right room boundaries. Surfaces behind the loudspeakers, in front of the listener, should be primarily diffusive. The space behind the listener should be primarily absorptive. Side wall reflections should not be absorbed. If the room talks back in the same voice, with similar spectral content to the direct loudspeaker sound, then the room is of background interest and we adjust our acoustic horizon to focus on the auditory scene formed by the direct sound and its spatial information content. Thus we hear the recording venue, develop an auditory image of it as individual instruments or groups illuminate the venue space and it responds by reverberation. The venue space can sound larger and have more depth than the listening room in which we sit.

2.3.2. *Hearing the room*

An example will illustrate our ability to tune out the room. Record from the “sweet spot” the sound of your stereo loudspeakers in your room while listening to a CD track using microphones as in Fig. 11, [28]. Next play this recording back through the loudspeakers and compare the reproduction to what you heard before.

You will hear the loudspeakers in your room, similar to hearing an orchestra in a concert hall on a CD. You no longer can remove your room from the presentation as you did when you listened to the CD track initially. You may also note noises in your room that you were not aware of, or how spectrally colored the loudspeakers are. You now hear how the microphones have sampled spatial information for two fixed points, which is insufficient for the brain to fully reconstruct a 3-dimensional auditory scene. Head-tracking would have been needed.

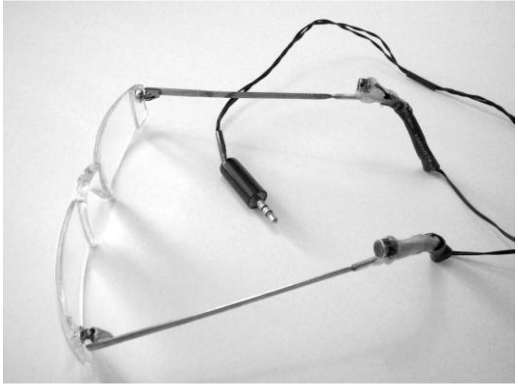


Figure 11: Recording with microphones on the sides of the head, excluding the pinna.

2.3.3. *Hiding the stereo loudspeakers perceptually*

In order to hide the room behind the acoustic horizon the loudspeakers must first illuminate the room uniformly at all frequencies. This can only be accomplished with an acoustically small source. [29] [30] The radiation pattern could be omni-directional (monopolar), dipolar or cardioid, but there are practical considerations, which favor a dipole [31]. The typical figure-of-eight radiation pattern of a dipole can be maintained down to the lowest frequencies. The directionality reduces the excitation of room resonance modes. The total power radiated into the room is 4.8 dB less for a dipole than for a monopole at the identical on-axis sound pressure level. The room is less engaged because the intensity of illumination is reduced compared to the monopole, yet spectrally neutral, Fig. 12. A dipole loudspeaker employs an open baffle, thus there is no energy storage inside an enclosure and re-radiation of airborne energy through resonant vibrations of the enclosure walls and through the driver cone. The typical box loudspeaker, which is omni-directional at low frequencies and increasingly forward radiating as frequency increases, often suffers from box radiation issues. Radiation pattern and box radiation are responsible for the generic loudspeaker sound that most people are familiar with and expect to hear. Box loudspeakers often fails to reproduce the open sound of natural acoustic sources and events.

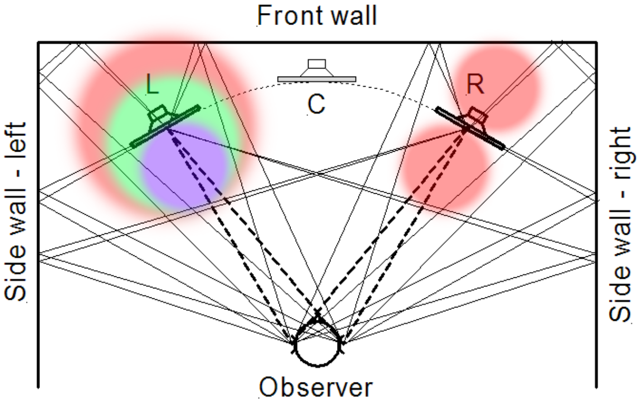


Figure 12: A typical box loudspeaker L is omni-directional at low frequencies and becomes increasingly forward directional at higher frequencies. A dipole loudspeaker R can have a frequency independent radiation pattern.

For the loudspeaker not to stand out in the experienced auditory scene it must not add sounds of its own, whether due to energy storage or intermodulation distortion. It must be able to handle the softest and loudest passages effortlessly, while generating near realistic sound pressure levels.

2.3.4. *Phantom source distance & space*

The loudspeakers are hidden when the real signals from left and right loudspeakers have merged with the phantom sources in a spatial continuum that is not hard-bounded by the loudspeakers. Spectrally similar room reflections allow the auditory scene to be dominated by the direct loudspeaker signals, which carry information about the recorded acoustic event in its spatial context. The auditory scene becomes 3-dimensional, located behind the loudspeakers, with depth, width and even some height. In many cases it can only be a miniaturization of reality, as heard from a distance. Indeed the volume control acts as the auditory scene’s size and distance control and its program specific setting is critical to obtain a believable auditory scene, one that creates full enjoyment.

2.4. Recording for stereo

A loudspeaker radiation pattern and room setup that fully exploits the potential in the stereo format brings out the desire to hear recordings, which capture the musical instruments and voices in their natural setting, and as one might have been experienced when attending the concert performance. Two signal channels are available for transporting the recording from the venue to the living room. Two microphones should suffice to capture a natural and spatially believable auditory scene [32], but some musical instruments may not be rendered with sufficient clarity, requiring additional microphones, Fig. 13. The output from the added microphones must be mixed into left and right channels to add phantom sources in phantom locations between and behind left and right stereo loudspeakers.

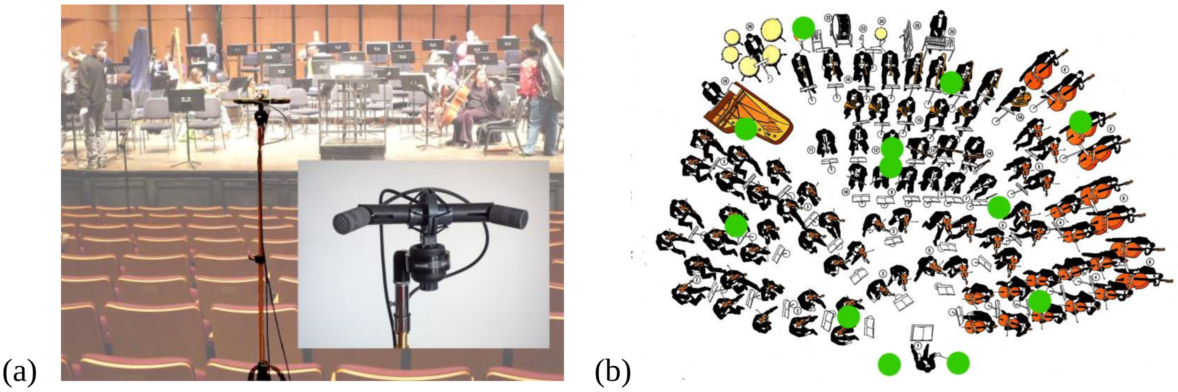
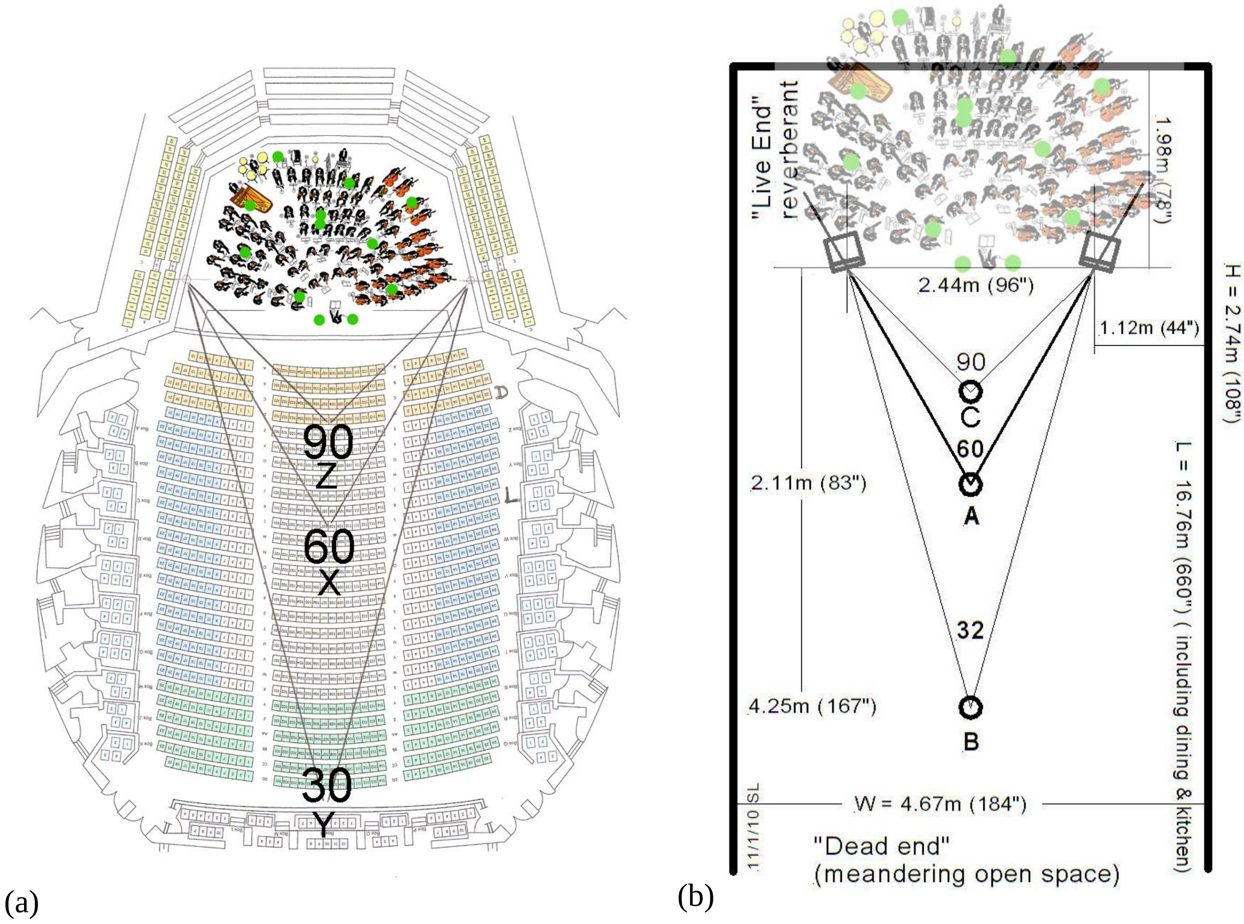


Figure 13: Recording with two microphones from an audience perspective (a). Recording multiple unfamiliar perspectives of an orchestra, which are down-mixed to two channels (b).

Spatial integration of the added sources into the auditory scene that was established by the two primary microphones is a difficult task, if the recording engineer works with monitor loudspeakers and a setup that cannot show him what is happening and what a consumer with a state-of-the-art stereo system will hear in his living room. I must assume that monitoring the recording has its weaknesses; otherwise I cannot believe that the spatial incongruity that I often hear in commercial recordings has been intentional.

2.4.1. *Perspective and distance of the reproduced Auditory Scene*

When loudspeakers are designed properly and set up in a room in such a way that both, loudspeakers and room, are no longer recognized in the auditory scene, then it becomes possible to make meaningful comparisons between the auditory scene experienced at a live concert and its reproduction in the living room. Such comparisons are important to me when I evaluate loudspeakers, their setup and room influences. Seen from the “sweet spot” ‘A’ in my room the two loudspeakers subtend a 60^0 angle, Fig. 14. The phantom source will essentially occupy this angle in front of me and it will not be closer than the distance to the real sources of sound, the two loudspeakers. In the concert hall the musicians, the sources of sound, are seen at a 60^0 angle from seat ‘X’. Let us make a recording from this location with small omni-directional microphones placed on each side of the head, Fig. 11. Later, left and right microphone signals are fed to left and right loudspeakers without any further processing and listened to from location A. The signal streams from the two loudspeakers contain ILD and ITD cues for angular positions within the auditory scene, as well as some distance, elevation and HRTF cues.



(a) (b)

Figure 14: The 60^0 angle of perspective in a concert hall (a) and in a living room (b). The distances to the orchestra are about 10 times larger than to the loudspeakers.

The rear hemisphere and much of the side sounds that the microphones picked up will now be heard coming from the front, in addition to the orchestra. Thus there will be more hall sound in the reproduced auditory scene then was probably heard live. The direct-to-reverberant sound ratio has been reduced. The ear-brain perceptual processor performs a running correlation between left and right eardrum signal streams to locate sound sources in phantom space and to recognize its spatial attributes of width, depth, height and continuity. This is an immediate learning process as different sound sources illuminate parts of that space.

Amplitude and timing differences between the signals streams at the ears, their amplitude envelopes and spectrum are decoded in the brain to place phantom sources in locations between the loudspeakers. Stream content that cannot be mapped between the loudspeakers is lumped into left and right loudspeakers as monaural sources. It is not clear how the auditory scene is affected by this.

Distance and size of the auditory scene are playback volume dependent. The auditory scene moves closer with increasing volume, it becomes larger and more detailed. There is a limit to the maximum volume setting, when the perceived distance to the auditory scene becomes incongruous with its loudness and size. Listening from location ‘B’ will not increase the maximum acceptable volume setting significantly. It will merge the auditory scene into the room with a loss of imaging precision and detail. Listening from location ‘C’ is more immersive into the auditory scene and actually my preferred seat for this type of head-related recording.

In the concert hall location ‘Y’ and the last four rows of chairs are underneath the balcony. From here the orchestra and hall are seen as through a very wide and open window. I am aware of being in a smaller space with a low ceiling listening out into a larger space. This is also audible when listening to a recording from this area of the hall. In general, a sense of height is captured with this simple recording technique, which adds to the realism of the auditory scene reproduced at ‘A’. Location ‘Z’ and closer to the orchestra produce a very unbalanced auditory scene because musicians are visually hidden and distances to them vary greatly. Commercial recordings in this hall typically use a large number of microphones hanging above different sections and instruments of the orchestra. A widely spaced pair above row ‘D’ picks up hall sound. Recordings here have won prestigious awards.

3. Conclusions

Stereo recording, reproduction in a room and hearing must be treated at as a continuum in order to obtain optimum auditory results. Stereo is based upon and relies upon our natural hearing processes, which are capable of creating a believable auditory scene in the mind, even when the air pressure signal streams at the eardrums do not represent the physics of a naturally familiar acoustic event. We must cooperate by hiding loudspeakers and room from perception by proper design and by recording sound as it exists in space naturally. Such a stereo system can be used to recreate art and to create art to its fullest, because the recording engineer/producer knows the outcome. Stereo in its purest form is about recording real sources and reproducing them over two loudspeakers in a room as phantom sources, without hearing room and loudspeakers, while generating a believable illusion of sound sources in their spatial context. Timbre, localization and spaciousness are essential contributors to a satisfying auditory experience and should be preserved from recording to reproduction.

4. Acknowledgements

Audio has been a lifelong interest and a journey on which I learned much from my colleagues at Hewlett-Packard Co., Russ Riley, Lyman Miller and Brian Elliott, when we designed and built all sorts of audio gear in our spare time. Loudspeakers covered a similar range of wavelengths as the microwave test equipment that we developed for HP. Laurie Fincham purchased one of the early HP Fourier Analyzers for KEF and thereafter became my snail-mail discussion partner for loudspeaker design issues. Don Barringer, then recording engineer for the US Marine Band, was always looking for more accurate monitor loudspeakers and built what I had designed for my own use at home. Today he serves as "my other pair of ears across the country", as we refine stereo reproduction for our own use. Writing and talking about our observations hopefully invites others to take a fresh look at industry and marketing practices and to move beyond the current paradigm.

5. References

[1] Blesser, B., Salter, L.-R.: "Spaces Speak, Are You Listening? – Experiencing Aural Architecture", *The MIT Press, 2007*

[2] Bregman, A. S.: "Auditory Scene Analysis – The Perceptual Organization of Sound", *The MIT Press, 1999*

[3] Jens Blauert, J.: "Spatial Hearing – The Psychophysics of Human Sound Localization", *The MIT Press, 1997*

[4] Peter Damaske, P.: "Acoustics and Hearing", *Springer, 2008*

[5] Levitin, D. J.: "This Is Your Brain on Music – The Science of a Human Obsession", *Dutton, 2000*

[6] Griesinger, D.: "Frequency Response Adaptation in Binaural Hearing", *126th AES Convention, Munich 2009, Preprint 7768, http://www.davidgriesinger.com/*

[7] Schulein, R.: "Binaural Audio Technology – History, Current Practice, and Emerging Trends", *Master Class 2, AES 125th Convention, San Francisco, USA, 2008 October 2-5*

[8] Duda, R. O.: Head-tracking, personal demonstration, *AES 125th Convention, San Francisco, USA, 2008 October 2-5*

[9] Mackensen, P., Fruhmann, M., Thanner, M., Theile, G., Horbach, U., Karamustafaoglu, A.: "Head-Tracker Based Auralizartion Systems: Additional Consideration of Vertical Head Movements", *108th AES Convention, Paris, 2002, Preprint 5135*

[10] Theile, G.: “On the Naturalness of Two-Channel Stereo Sound”, *JAES, Vol.39, No. 10, 1991 October*

[11] Hartmann, W. M.: “Listening in a Room and the Precedence Effect”, *Chapter 10 in “Binaural and Spatial Hearing in Real and Virtual Environments”, R. Gilkey and T. Anderson (Eds.), Lawrence Erlbaum Associates, Hillsdale, NJ, 1997*

[12] Yost, W. A.: “Perceptual Models for Auditory Localization”, *in “The Perception of Reproduced Sound”, The Proceedings of the AES 12th International Conference, Copenhagen, Denmark, 1993*

[13] Yost, W. A.: “The Cocktail Party Problem: Forty Years Later”, *Chapter 17 in “Binaural and Spatial Hearing in Real and Virtual Environments” R. Gilkey and T. Anderson (Eds.), Lawrence Erlbaum Associates, Hillsdale, NJ, 1997*

[14] Linkwitz, S. “Room Reflections Misunderstood?”, *AES 123rd Convention, New York, USA, 2007 October 5-8, (Preprint 7162*

[15] Toole, F. E.: “Sound Reproduction – Loudspeakers and Rooms”, *Focal Press, 2008*

[16] Bock, T. M., Keele, D. B.: “The effect of inter-aural crosstalk on stereo reproduction & Monitoring by the use of a physical barrier”, *AES 81st Convention, Los Angeles, USA, 1986 November 12-16, (Preprint 2420A & 2420B)*

[17] University of Southampton, Institute of Sound and Vibration Research, “Stereo Dipole”, http://www.isvr.soton.ac.uk/FDAG/VAP/html/sd.html

[18] Algazi, V. R., Avendano, C., Duda, R. O.: “Estimation of a Spherical-Head Model from Anthropometry”, *J. Audio Eng. Soc., Vol 49, No 6, 2001 June*

[19] Skudrzyk, E.: “The Foundations of Acoustics”, *page 396, Springer 1971*

[20] Shaw, E. A. G.:”Acoustical Features of the Human External Ear”, *Chapter 2 in “Binaural and Spatial Hearing in Real and Virtual Environments”, R. Gilkey and T. Anderson (Eds.), Lawrence Erlbaum Associates, Hillsdale, NJ, 1997*

[21] Duda, R. O., Martens, W. L.: “Range dependence of the response of a spherical head model”, *J. Acoust. Soc. Am. 104 (5), November 1998*

[22] Linkwitz, S.: “ORION-3.1”, *http://www.linkwitzlab.com/orion-rev3.htm*

[23] Michael Williams, M.: “Microphone Arrays for Stereo and Multichannel Sound Recording”, *Editrice Il Rostro, 2004, http://www.posthorn.com/Micarray_1.html*

[24] Wittek, H., Theile, G.: "The Recording angle – Based on Localization Curves", *112th AES Convention, Munich 2002, http://hauptmikrofon.de/HW/AES112_Wittek_Theile.PDF*

[25] Rumsey, F.: "Spatial Audio", *Focal Press, 2005*

[26] Linkwitz, S.: "The Challenge to Find the Optimum Radiation Pattern and Placement of Stereo Loudspeakers in a Room for the Creation of Phantom Sources and Simultaneous Masking of Real Sources", *127th AES Convention, New York, 2009, Preprint 7959, http://www.linkwitzlab.com/publications.htm, #27*

[27] Griesinger, D.: "The Importance of the Direct to Reverberant Ratio in the Perception of Distance, Localization, Clarity and Envelopment", *126th AES Convention, Munich 2009, Preprint 7724, http://www.davidgriesinger.com/*

[28] Linkwitz, S.: "Microphone", http://www.linkwitzlab.com/sys_test.htm#Mic

[29] Harz, H., Koesters, H.: "Ein neuer Gesichtspunkt fuer die Entwicklung von Lautsprechern?", *Technische Hausmitteilungen des NWDR, Jahrgang 3, Nr. 12, Dezember 1951*

[30] Loudspeaker, PLUTO-2.1, *http://www.linkwitzlab.com/Pluto/Pluto-2.1.htm*

[31] Loudspeaker, ORION, *http://www.linkwitzlab.com/orion_challenge.htm*

[32] Schoeps ORTF Stereo Microphone *http://www.schoeps.de/en/products/mstc64u and application http://www.schoeps.de/en/applications/showroom*