An Image Model Theory for Stereophonic Sound

2869 (A-3)

Gary C. Eickmeier
Independent Industrial Designer
Lakeland, FL

**Presented at
the 87th Convention
1989 October 18-21
New York**

AUDIO
ES
®

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

AN IMAGE MODEL THEORY FOR STEREOPHONIC SOUND

Gary C. Eickmeier

independent industrial designer

A new theory for stereophonic sound is presented as background for a loudspeaker currently under development. The theory uses image modeling to visualize all sound fields present in the live performance: direct, early reflected, and reverberant. The reproduced sound will be most like the original when the image model of the reproduction soundfield matches that of the live sound. The model enhances understanding of perception of the spatial qualities of reproduced sound; is a basis for loudspeaker design, positioning, and acoustical requirements of the room; and extends to techniques for the original recording.

INTRODUCTION

By now most of the technical aspects of sound recording and reproduction are well in hand. We have sufficient accuracy and flexibility in our transducers, amplifiers, and storage media to do most anything we want, and yet more often than not the sound still lacks the realism of a live performance. It is my belief that the most fundamental aspect of the process, stereophonic theory itself, is not yet complete and correct. This can lead to incorrect practice at both ends of the process (recording and reproduction), false goals as to what can be achieved with stereophonic sound, and research paths that turn out to be dead ends when trying to improve the situation.

This paper will introduce a new theory for stereophonic sound. It is presented as a model which can be used to study all aspects of the recording and reproduction process. It is based on a concept from architectural acoustics called image modeling, which is simply a way of studying the incident angles, intensities, and so on of a sound source and all acoustic images of that source in a given acoustic space. Instead of drawing ray traces for the reflected sound, you draw the virtual images of the source as additional sources behind the reflecting surfaces [1,2]. This technique gives a more visual understanding of what it is that we are actually hearing. To be more specific, what is meant by the image model is not just the positions and intensities of the virtual images, but all of the spatial, spectral, and temporal characteristics that are essential (audible) components of the model.

It is possible to make a drawing of the image model of musical instruments in a concert hall or loudspeakers in a playback room. Image model theory, then, can be stated as follows: The reproduction will be most like the real thing when the image model of the reproduction soundfield comes as close as possible to that of the original. This seemingly innocuous statement is in fact a big change in the way that we think about the process. Stereophonic sound is seen as a large scale, acoustical, field-type reproducing system in which all sound fields present in the original are physically reconstructed in the playback acoustic. The reproduction is seen as a 3-dimensional model of live sound, as opposed to a wavefront or portaling process. The major change in practice is that reflected sound is incorporated into the construction of the stereo image.

We will see the implications of this as we proceed, but first a review of the history of auditory perspective is in order, to better illustrate how the new theory differs.

REVIEW

The first and most basic type of auditory perspective was actually binaural. This paper has nothing to do with binaural sound, but since both Harry F. Olson and William B. Snow cautioned us that the terms binaural and stereophonic tend to be continually confused with each other, perhaps a brief definition would be good.

Olson has defined the two systems as follows: "Binaural is a closed-circuit type of sound reproducing system in which two microphones, used to pick up the original sound, are each connected to two independent corresponding transducing channels which in turn are coupled to two independent corresponding telephone receivers worn by the listener. A stereophonic sound reproducing system is a field type sound reproducing system in which two or more microphones, used to pick up the original sound, are each coupled to a corresponding number of independent transducing channels which in turn are each coupled to a corresponding number of loudspeakers arranged in substantial geometrical correspondence to that of the microphones."[3] Snow comments that "It has been aptly said that the binaural system transports the listener to the original scene, whereas the stereophonic system transports the sound source to the listener's room."[4]

It would be worthwhile at this point to emphasize the difference between a closed-circuit type system (binaural) and a field type system (stereophonic) as used in the above system definitions. Binaural reproduction is meant to isolate the listener from his actual acoustic environment and to isolate the two channels from each other by presenting the sound directly to the ears by means of headphones. The recording was made with a binaural head placed in the "best seat in the house," where it could hear both the music and the complete acoustic environment from that position. The system requires each ear to receive exactly the same signals that impinged upon the binaural head during recording. With the stereophonic system the microphones are placed much closer to the orchestra, recording the sound that exists in the region closer to the instruments rather than that which would occur near a typical listening position. This means that the recording contains correspondingly less of the concert hall acoustic and more of the direct and early reflected sound from the region of the proscenium. The recording is reproduced on loudspeakers which are placed at a distance from the listener, the entire process taking place in an acoustic space which is different from that in which the recording was made. The resultant sound depends on the new acoustic surroundings to impart acoustical qualities required for good sound [5].

A common-sense way of stating all this is that with binaural we are recording and reproducing ear signals, whereas with stereo we are reproducing the orchestra itself, and the soundstage surrounding it, on a macroscopic scale in the playback room. The channels will blend with each other acoustically, and both ears are free to hear both (or all) speakers. We are neither isolating the channels from each other (at the ears) nor the listener from the playback room. Rather, we are using the acoustics of the playback room to reconsuruct the sound field around him. The listener then experiences the total sound field, rather than having individual channels piped to his ears. The confusion between the two systems arises when people begin to believe that the object of "accurate" stereophonic reproduction is to get the sound that went into the microphones straight to the listener´s ears, or (worse) that the sound from each speaker should be heard by only the respective ear, with all interaural crosstalk eliminated.

The binaural system is straightforward and uncontroversial enough, but the stereophonic system, even at the most fundamental theoretical level, is still undeveloped. To be more specific, lateralization has been discussed in great detail and related unequivocally to the intensity or time difference of the direct sound from two loudspeakers. Localization and spatial impression have not. Entire classes of auditory cues necessary for localization in enclosed spaces, as discussed by Benade, Moulton and Ferralli, have been omitted from reproduction with the usual stereo arrangement [6,7]. The body of knowledge from architectural acoustics as to what causes good sound in a concert hall is temporarily sidestepped when we attempt to reproduce the original sound field as a high direct field from two point sources, with no explanation as to how we are supposed to regain those qualities [8,2,9, p348]. Many maverick products exist in the marketplace to address some of these problems, but none of them is as yet a part of an overriding stereo theory in any scientific or engineering sense.

It is difficult, therefore, to review the present state of stereophonic theory because there is no single theory, written down in so many words in textbook form. However, from the many articles that have been written and from the practice of stereo certain trends can be inferred.

From the Bell labs "curtain of sound" theory stereophonic sound is seen as a wavefront that can be approximated by two or three speakers (and corresponding microphones) [3,4]. From Blumlein, stereo is thought to be a microscopic (as opposed to macroscopic, or large scale) process wherein a coincident pair of microphones encode the direction of all arriving sounds by means of intensity, and the loudspeakers deliver those signals to the listener´s ears so that the original sound field might be perceived at the listener´s head [10,11].

The trend to note with both of these versions is that stereo is thought to operate as a sort of windowing or portaling process wherein the sound that was recorded is simply being relayed to the listener by the reproduction chain. Stereophonic sound is thought to be a "trick" that attempts to fool the ears into hearing all audible spatial properties of live sound strictly by means of lateralization - like looking through a portal into another acoustic space. The degree of success of the illusion is thought to depend on the "accuracy" of the system, and the status of stereo theory as we know it today can be thought of as a search for greater and greater accuracy.

Notice also that the above descriptions are strictly two-dimensional processes. The theories are based only on the direct sound radiated from a pair

or a line of speakers. They are "blind" to the effects of loudspeaker radiation pattern, positioning, and room acoustics. We started with the system definition as a field type system, reproduced in a real acoustic space by loudspeakers, but as far as the explanation of how it works goes, the playback room might as well not exist, and nowhere do we find reflected sound incorporated as part of stereo theory.

AN ANALOGY

The best way to illustrate this highly conceptual problem is with an analogy.

Many people have used the "brick wall" analogy - that stereo is something like punching out two holes in a brick wall separating you from the performance. Some writers widen the two holes and join them together, some claim that their systems knock down the entire wall, but we are always witnessing the sound through a large portal, standing on the outside looking in.

That´s a good starting point, and a nice, simple analogy to make the desired point, but let´s take it one step further. Imagine your listening room plunked down in the middle of Symphony Hall with you in it. We´re going to punch out first two holes (or a portal) in front of us, to "let the music in." Then, the surround sound devotees will punch out some more holes in the rear and perhaps side walls, to let all the ambience in. Under the "accuracy" banner, we say that when the reproduction chain gets good enough, the sound will be indistinguishable from this punched-out shell of a room, with nothing between you and the music but air.

The caution at this point is that this would all be very fine thinking except that, no matter how many channels we have, we will never quite make it all the way because, in this analogy, we must remember that the sound can get into the imaginary room, but it can´t get out, and so the sound still bounces around the listening room with the time between reflections of the smaller space.

The main point of this section, however, is that this is not a good analogy at all.

Many people, especially audiophiles, have the impression that the recording contains a perfect image of the performance as witnessed from the best seat in the house. This may be true with binaural, but stereophonic is a very much different process. The problem with the above analogy is that it pictures the sound as having been "witnessed", or recorded, from the vantage point of the listener in the room suspended in the middle of the concert hall. This is not the case. What we have done is dispatched the microphones up to the orchestra, recorded the musicians and the soundstage surrounding them, and brought back the sound to be played again from entirely within our room, not from outside with holes punched in the walls so we can hear it. This is quite a different thing, and it forces us for the first time to think of our listening room not as a nuisance variable but as the performing space itself. For better or for worse, the room must be thought of as an integral part of the sound, to be used to construct the same sort of spatial patterns that existed in the real concert hall, rather than fought with sound killing materials. I believe that this is for the better, because, once we reconstruct the sound fields in our playback room, all of the characteristics of live sound can be present, making the sound real and not a trick. The stereophonic recording can be thought of as a concentrate, to be mixed with the playback acoustic in a way that models the reproduction after the real thing. Although we must inevitably hear some of the listening room along with the "flavor" of the recorded acoustic, the realism can be stunning.

THE REAL THING

In order to construct a realistic image model in our playback space, we need to first examine the pertinent characteristics of the real thing.

Let's imagine an experiment with which a team of researchers wishes to measure all of the temporal, spectral, and spatial characteristics of "the real thing" under controlled conditions. The team goes into a typical, good sounding concert hall with a battery of test equipment and begins. The measurements will all take place at the best seat in the house - say, fifth row center. For the source, we will use "idealized" musical instruments in the form of three omnidirectional loudspeakers with flat frequency response throughout the spectrum. We will place them at stage left, center, and right, to represent the three extremes, stereophonically speaking (Fig. 1). By knowing exactly what is coming out of the source, we can tell exactly what happens to the sound by the time it reaches the listening position.

If we send appropriate test signals through these sources, what will we measure and hear about all of the characteristics of the sound that is incident upon the listener?

TEMPORAL: Temporally speaking, we can hear the size of the concert hall by means of the time between reflections [8], and we can hear the reverberation time of the hall. The direct sound is the first to arrive at our ears (and test instruments), followed in short order by the early reflections from the front and sides of the hall, and finally the reverberation "tail", or full diffuse field as the reverberation time dies out [12].

SPECTRAL: The first thing we notice is that the frequency response of our perfectly flat test sources is not heard as anything near flat at the listening position. Bass frequencies are reinforced by their own early reflections, to the extent that they are equally loud throughout the concert hall, except for minor local maxima and minima due to cancellations and standing waves. High frequencies are attenuated by absorption and diffusion (scattering) as they reflect from the various surfaces throughout the hall, and by simple inverse square attenuation due to distance from the source. The end result is that the frequency response of our flat source takes on what is commonly referred to as a room curve, with humped low end up to about 1k and a gradually falling high end beyond that [13].

SPATIAL: In the present context this characteristic refers simply to the incident angles from which the sound arrives at the listener. It has nothing to do with the temporal characteristic, but to understand it better it is necessary to divide up the arriving sound into its three main temporal stages. These are the first arrival (direct) sound, the early reflections, and the full reverberant, or diffuse field. We can look at these three fields in terms of their loudness, direction of arrival, and frequency balance, or contribution to our total impression of the timbre of the instruments. We measure the direct sound as having a flat frequency response, as coming straight from the source (defining the lateral localization of the sources), but as being a very small part, volume wise, of the total sound incident upon the listener [8,12,13]. As for the early reflected sound, its spatial nature can best be understood by drawing an image model of the three sources, showing the positions of the virtual images and therefore the perceived directions from which their primary reflections arrive. It is very important to note that this early reflected sound comes from different points in space from the direct sound, and it is the relationship between the source and all of its acoustic images that enables us to better localize sounds in a reverberant space than in a free field [6]. All of the images taken together enrich the sound and help to define the size and shape of the soundstage on which the instruments are playing. Spatial impression is the psychoacoustic term used to describe the broadness of this total sound image as perceived at the listening position [9]. It is very important to the enjoyment of the sound of a good concert hall and to the realism of a reproducing system. Spectrally speaking, the early reflected sound is close to the original, but more diffuse, and rolled off above 10k [5]. The full reverberant field is virtually omnidirectional, or non-directional, coming at the listener from 360 degrees around him [8,2,12,13]. It is a dynamic process, being composed of a build-up of the longer duration, more steady-state tones, as opposed to transients, and its frequency balance is dominated by midrange on down. This is the quality which causes the perception of a "sweetening" or musicality to the sound [8,14,6].

The sum total of all this is that the sound always seems to be coming from the sources themselves, because of the precedence effect of the high frequency transients carried by the direct sound. The early reflected sound lends the spatial impression, giving us clues as to the depth and width of the surfaces surrounding the sources - the soundstage size. The subjective impression of frequency balance, especially with live musical instruments, comes almost totally from the reverberant field, including the early reflected sound.

The real thing has tremendous physical size, enormous power, and spectral, temporal, and spatial characteristics as indicated above. This is what we are trying to reproduce.

THE REPRODUCTION

It is now postulated that if any of the above characteristics of the model for live sound is not reconstructed on playback, then the reproduction will sound different from the original. We then propose that realistic reconstruction will only happen if a combination of direct and reflected sound is deployed in the playback room so that the image model of the reproduction soundfield comes as close as possible to that of the original (Fig. 2). The practical realization of the model involves the design of the radiation pattern of the speakers, the positioning of the speakers in the room, and the acoustical properties of the room.

If the reproduction is more correctly seen as a model of the real thing, then we will begin with a room of good physical size and having dimensions and acoustical properties similar to those of a good concert hall. A rectangular room with a high ceiling and normal reflectivity (not too live and not too dead) is best, with the speakers firing down the long dimension.

The radiation pattern and room positioning of the speakers will be chosen so that the geometrical patterns and resultant intensities of all of the images is similar to the live model. It turns out that a greater proportion of reflected sound than direct sound is required in order to more closely mimic the spatial "shape" of the live soundfield, because of the closeness of the speakers to the listener, compared with the distance to the instruments in a concert hall [8].

It remains only to describe the type of sound that results from this technique.

The immediate impression is that the sound does not seem to be coming from the speakers. The speakers behave more like projectors, suspending a very holographic, 3-dimensional aerial image of the sound in the region between the speakers and the reflecting surfaces [14]. The soundstage extends from wall to wall, with individual instrumental images possible from slightly left of the left speaker to slightly right of the right speaker, and taking on a depth from a foot or two behind the speakers to whatever psychoacoustically induced depth has been recorded. Imaging is stable and very even from left to right in both the direct and early reflected domains. Spaciousness of the early reflected sound and pinpointedness of individual images are modulated by the recording, with soundstage size and closeness of the instruments varying from tight and dry to cavernous. Recorded early reflections can seem to actually be coming from the appropriate wall of the listening room.

The impression of an ensemble of performers surrounded by a spacious soundstage strongly suggests that there is a decoding process occuring which causes the direct and early reflected sound that was recorded to become presented as such on playback. The observation is this: If the recording is a very close-miked, dry sound containing little reverberation, the only effect of the reflected sound on playback is a harmless image shift toward the reflecting surfaces. Individual images are still well focused, they simply form themselves slightly behind the plane of the speakers, and they seem to be right in the room with you. If the recording does contain reverberation, the auditory event maintains its position and focus, but there is also a spatial broadening effect as described by Blauert for a single delayed reflection [9]. The situation is not exactly the same, but the effect is. The only explanation that can be offered at this point is the law of the first wavefront [9]. When the balances are just right between direct and reflected in the playback image model, the precedence of the recorded direct sound sets it off from the loudness of the recorded early reflected sound. The effect is similar to that experienced with a delay-based surround system. The spatial nature of the sound changes, but the frontal image remains intact. With the image model, the room is acting as a sound field synthesizer as described in Blauert [9, pp282-285], or as Borish's auditorium simulator [2] and giving us the complete pattern of direct and reflected images as required by Benade for live sound [6] and many others for reproduced sound [8,15,2,7]. One thing that is certain is that, with a true stereophonic recording, the virtual images of the instruments have a greater stereo separation than the instruments themselves, and this may be what we are hearing. With the image model, we literally "set the stage" for this effect to come through on playback, whereas with conventional stereo from direct firing speakers the recorded early reflected sound is spatially compressed and made to come from the same limited set of incident angles as the direct sound, defined by the separation between the speakers (Fig. 3).

There is also a blending effect occuring between the recorded acoustic and that of the listening room. In image model theory, the contribution of the listening room is not a detriment, but rather should be considered absolutely essential to the realism that only stereophonic sound can have. Many of the functions of the concert hall described for the live image model above are now taken on by the listening room. These include the high frequency rolloff, spatial characteristics, and development of the full reverberant field in the most natural way possible. The reverberant field can also be supported by surround speakers in smaller rooms.

The cumulative result of all of the imaging and blending of the reflected sound is that the listener is immersed, or enveloped, in the soundfield the same way he would be at a live event. The closer miked instruments have a "they are here" presence, while the recorded acoustic gives the room the flavor, or ambience, of the concert hall or studio in which the recording was made.

IMPLICATIONS

The practical implications of the image model are presented in four areas:

Loudspeaker Design: If the function of a loudspeaker is seen as being an image model projector, then the radiation pattern can be designed by working backwards from the desired image model to the required radiation pattern (Fig. 4). It turns out that the pattern resulting from this approach should also be very good at time/intensity trading as the listener moves laterally across the room, yielding very stable images in both the direct and early reflected domains. Additionally, such a speaker could be used three-across for larger rooms. It would also be good to have a variable direct to reflected ratio, to adjust the image to the room.

Loudspeaker Positioning: Image model theory shows quite graphically that loudspeakers should be positioned for imaging, not for frequency response. The positioning is done according to specific geometrical rules, the purpose of which is to effect the evenness of the imaging of the reflected sound (Fig. 5). Mis-positioning speakers that have a high reflected sound component can cause uneven clusterings of images, leading to a hole in the middle or stretched soloists.

Room Acoustics: This area has already been touched upon above. Further research needs to be done on the reflecting surfaces around the speakers. It is possible that specular reflectivity would yield even more dramatic 3-dimensionality and focus for the entire soundstage and all images within it.

Recording: The major point here is that if the virtual images are not present in the recording they will not be available for playback. This means that the recording engineer is miking not just the musical instruments, but the early reflected sound as well (Fig. 6). The process might be described as close miking the soundstage. Also, since the image model is an orthographic projection in the listening room from relatively widely spaced loudspeakers, this suggests that spaced omnis are more correct than coincident microphones. The microphones will only be able to capture the correct intensity and timing relationships between the instruments and their virtual images if they are positioned near those images.

CONCLUSION

A new theory for stereophonic sound has been introduced which sees the reproduction as a model of live sound. Reflected sound has been incorporated into basic stereo theory, not as ambience enhancement but as an integral part of the frontal sound image. Making an image model drawing of a given sound system gives insight into the sound we hear from that system, and modeling live sound gives direction for improving the reproduction. In total, the model gives insight into all aspects of the recording and reproduction process. I believe that it has high correlation with observation about the sound of various types of reproducing systems, and gives theoretical validity to the type of sound that most people prefer.

This paper has been presented in the spirit of a full disclosure of the principles behind a loudspeaker currently under development. The hope is that the theory will help to solve the last remaining problems in the search for more realistic reproduction of music - from which we can all benefit.

REFERENCES

1. Berkovitz, Robert, "Sound and Your Listening Room," Stereo Review, Aug 1983.

2. Borish, Jeffrey, "An Auditorium Simulator for Home Use," presented at 74th AES Convention, New York, Oct 1983.

3. Sunier, John, The Story of Stereo, 1881 -, Gernsback library, New York, 1960.

4. Snow, William B., "Basic Principles of Stereophonic Sound," Journal of the SMPTE, Nov 1953.

5. Clark, David L., "Measuring Audible Effects of Time Delays in Listening Rooms," presented at 74th AES convention, New York, Oct 1983.

6. Benade, A.H., "From Instrument to Ear in a Room: Direct or Via Recording," Journal of the Audio Engineering Society 33:4, April, 1985, pp 218-233.

7. Moulton, David, and Ferralli, Michael, "The Localization of Phantom Images in an Omnidirectional Stereophonic Loudspeaker System," presented at 81st AES convention, Los Angeles, November 1986.

8. Bose, Amar G., "Sound Recording and Reproduction," Technology Review, June, July/August 1973

9. Blauert, Jens, Spatial Hearing, The MIT Press, 1983.

10. British Patent Specification 394,325. Alan Dower Blumlein. Journal of the Audio Engineering Society, vol. 6, no. 2, April, 1958.

11. Lipshitz, Stanley P, "Stereo Microphone Techniques - Are the Purists Wrong?", Journal of the Audio Engineering Society, vol 34 no. 9, September, 1986.

12. Everest, F. Alton, The Master Handbook of Acoustics, Tab Books, 1981.

13. Davis, Don and Carolyn, Sound System Engineering, Howard W. Sams & co, 1982.

14. Vaughan, Denis, "How We Hear Direction," Audio, December 1983.

15. Davis, Mark, "The Concepts and Technology Behind the DBX Soundfield One Speaker System," presented at 74th AES convention, New York, Oct 1983.

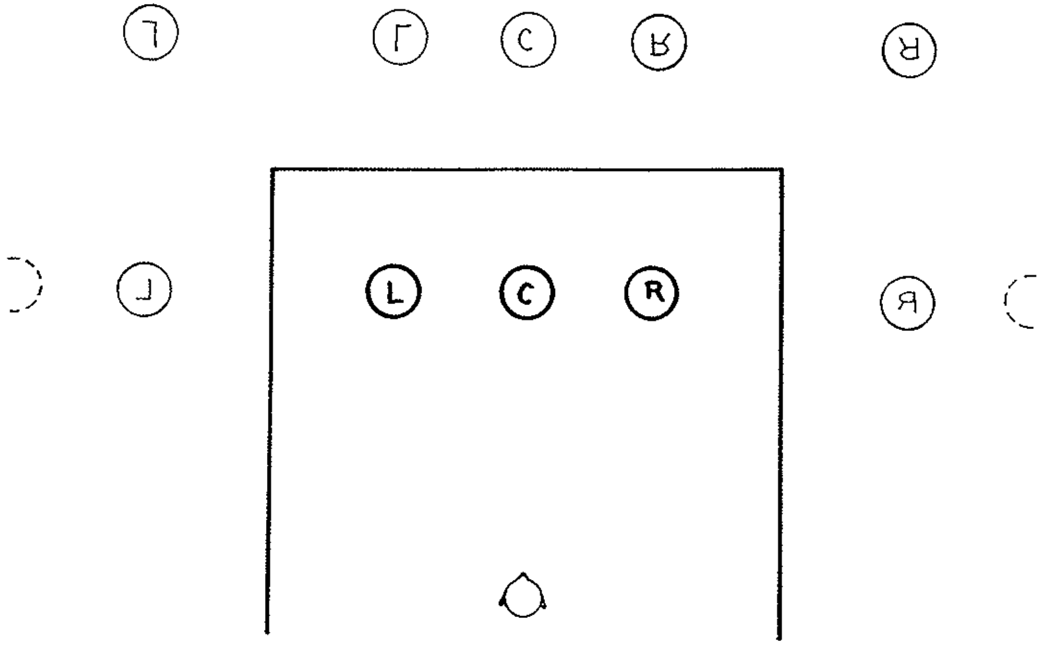


FIGURE 1: IMAGE MODEL OF THE REAL THING

Three omnidirectional loudspeakers are placed on the stage of a concert hall, and test signals are run through them. The easiest way to visualize the spatial characteristic is to make a drawing of the plan view of the sources and all of the acoustic images reflected from the walls nearest to those sources.

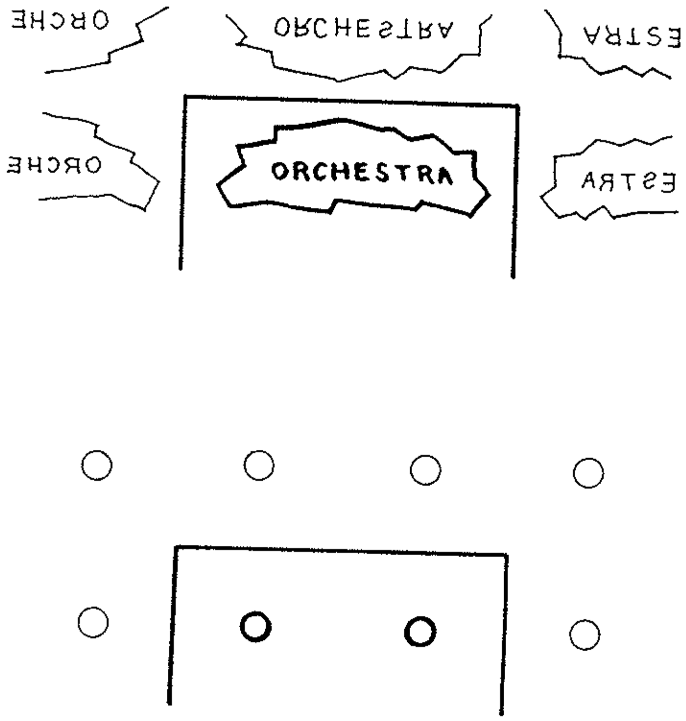


FIGURE 2: COMPARISON OF REPRODUCTION IMAGE MODEL TO LIVE SOUND

The spatial "shape" of the sound that we actually are hearing is a function of the positioning and intensities of all acoustic images with respect to each other. In the reproduction, speaker positions and direct to reflected ratios are adjusted to better approximate the live model.

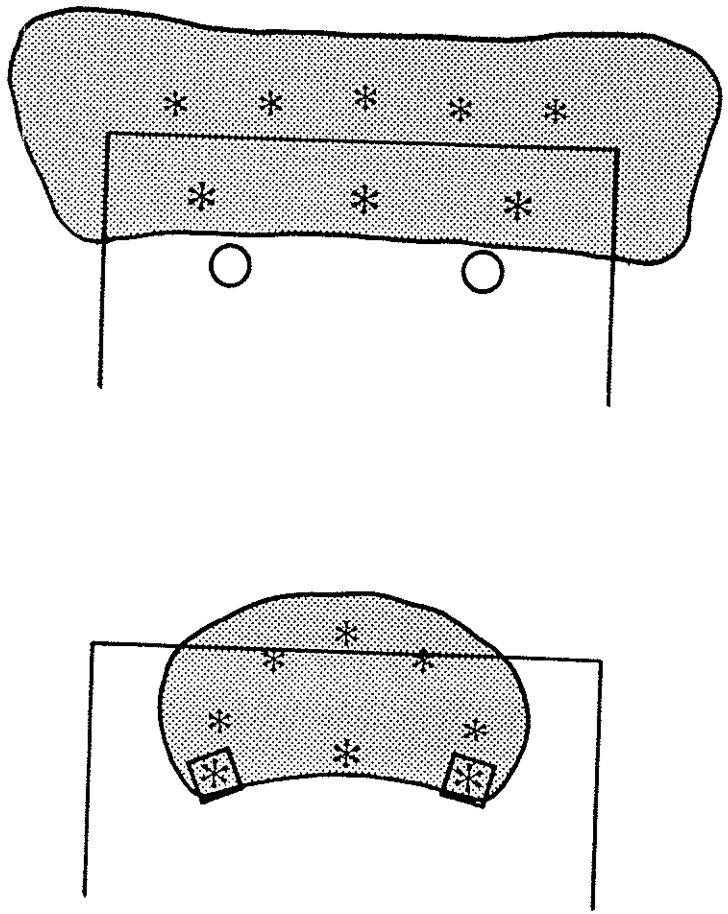


FIGURE 3: IMAGING COMPARISON, OMNI VS DIRECTIONAL SPEAKERS

These are approximate, typical, psychoacoustic impressions of imaging possible with omni (above) and direct firing speakers (below). Shaded area is soundstage size, symbols are individual images. Based on author's observations.

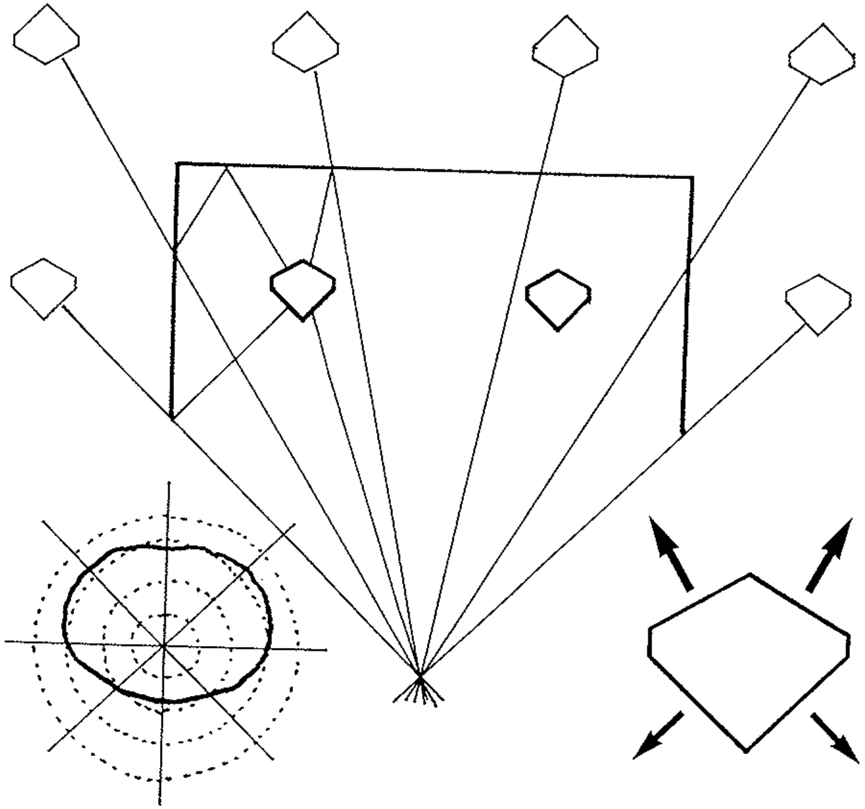
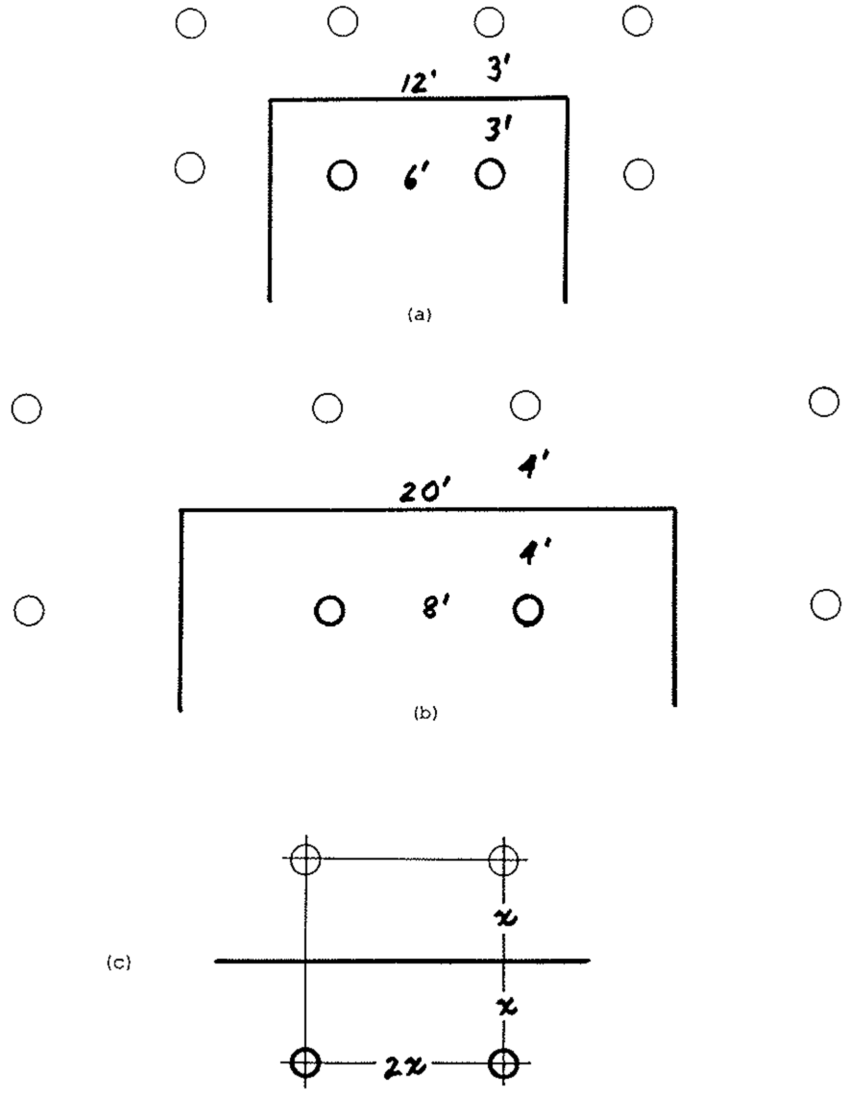


FIGURE 4: LOUDSPEAKER DESIGN

Working backward from the desired image model to the required radiation pattern produces a speaker that launches mid and high frequencies along axes angled approximately 30° to the rear and 45° to the front. Resultant radiation pattern should feature good time/intensity trading for both direct and early reflected sound, with very even image intensities all the way across the soundstage.



(a)

(b)

(c)

FIGURE 5: LOUDSPEAKER POSITIONING

For rooms smaller than 16 ft wide (a), the "rule of quarters" maximizes soundstage size and spaciousness. For larger rooms (b), the author has observed an apparent practical limit of 4 ft out and 8 ft apart. The most important rule (c) is to maintain a square aspect ratio between the actual speakers and the images immediately behind them. This provides perfect centerfill.

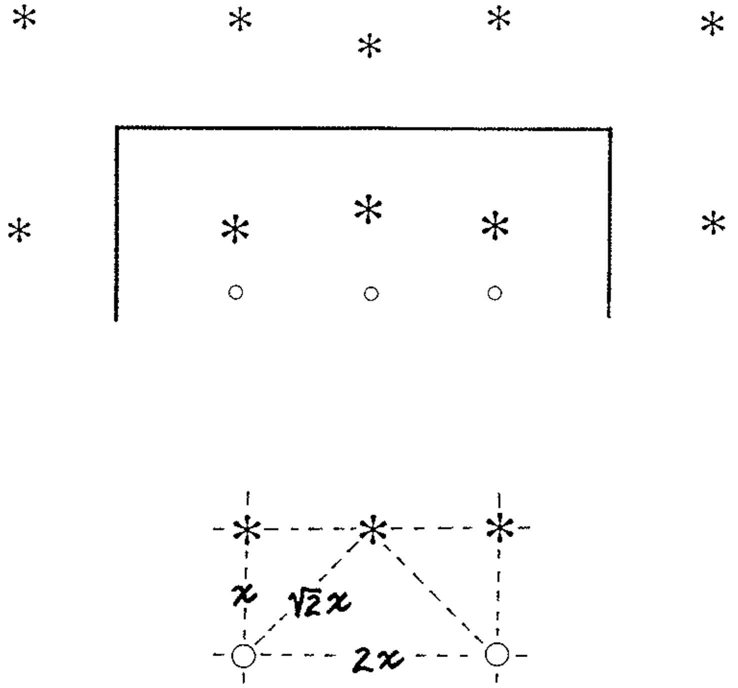


FIGURE 6: RECORDING

The "big picture", above, shows that there are actually two stereophonic recordings being made at the same time - the musical instruments and their acoustic images. Microphones and loudspeakers will be placed in similar geometric arrangements in their respective acoustic spaces. The smaller drawing, below, illustrates that there is a certain limiting distance for the closeness of the line of microphones to the nearest instruments, in order to retain equal loudness of all individual images across the front of the orchestra. Hopefully, this will also be no closer than the critical distance, at which direct and reverberant sound are about equal, so that we can capture a good balance of the total sound power radiated by the instruments.