

Ambiophonics, 2nd Edition

Replacing Stereophonics to Achieve Concert-Hall Realism

By **Ralph Glasgal**, Founder Ambiophonics Institute, Rockleigh, New Jersey www.ambiophonics.org

Dedication

To Johann Sebastian Bach, Richard Wagner, Gustav Mahler, and Lauritz Melchior, without whom we would never have bothered and to Manfred Schroeder, Don Keele Jr., Bob Carver, Yoichi Ando, and Ole Kirkeby without whose collective research we would never have succeeded.

Introduction

Ambiophonics: Recreating the Concert Hall Experience at Home

There are essentially only two ways for music lovers to enjoy music performed for them on traditional acoustic instruments. One is by going to a concert hall or other auditorium, and the second is by being elsewhere and playing the radio/TV/internet or a recording. This book and the techniques it describes are dedicated to helping you make the remote music-listening experience as audibly exciting as the live experience. Those audiophiles who share the dream of recreating a concert-hall sound field in their home, and who constantly strive to create a sense of "you-are-there," we have christened "ambiophiles". We call the science and technology used to create such an acoustic illusion "Ambiophonics".

Defining the Problem

Barry Willis wrote in *Stereophile Magazine* (August, 1994), "The idea that any musical event can be reproduced accurately through a two-channel home-audio system in a room that in no way resembles the space in which the original event took place is ludicrous."

Mr. Willis was absolutely correct in this when he wrote those words, but is much less so now, because Ambiophonics successfully works and its purpose is precisely to make the home-audio room resemble the space in which the original event took place. He goes on, "At present even the best discrete multichannel surround systems can offer only an illusion of being there." Experienced ambiophiles (a rare breed) would agree, but would also point out that surround sound is deliberately designed to produce the illusion of "they-are-here-around-you" which, while exciting for movies, is always going to be the antithesis of "being there". Finally, thoroughly despondent, Mr. Willis writes, "Tonal accuracy is the best that can be hoped for in a traditional audio system; true spatial accuracy will never happen. Audio products should come bearing this disclaimer: WARNING: IMAGE PRESENTED IS LESS THAN LIFELIKE."

The rest of us need not despair. Thirty years of experiments have been devoted to demonstrating that "lifelike" can happen, and with exceptional fidelity to the original, from just two standard LP/CD/SACD/DVD channels. As hard as it may be to believe, Ambiophonics works better with two recorded channels than other techniques such as surround 5.1 can do with multi-mic, multi-channel media such as Blue Ray or DVD where music is concerned. Yes, the Ambiophonic method described in

this book may not always precisely duplicate a particular hall, but it can create a hall and a vibrant stage that could exist architecturally, that rings true, and is lifelike enough to mimic a good seat at a live musical event.

Traditional Audiophile Articles of Faith

Many, if not most, serious audio enthusiasts presently believe that it is possible to achieve a solid stage image that may even extend beyond the loudspeaker positions, by employing the usual arrangement where two loudspeakers and the listener form something close to an equilateral triangle. They have faith that the perfect loudspeaker, amplifier, CD, LP, or 96/24 DVD player, and special cables will produce that wide, sharp imaging, stage depth, and ambient clarity that we all seek. Many also believe that audiophile-grade equipment, properly selected and tweaked, combined with signal path minimalism is more likely than simple acoustic listening room treatments to produce a higher fidelity sound field with enhanced width and depth. Some audio hobbyists prefer to listen primarily to small ensemble "they-are-here" small jazz-combo sounds such as found in the Chesky catalog and thus have no need or even desire to achieve a realistic orchestral or operatic sound field. They feel strongly that large scale symphonic or operatic classical sound reproduction is not what the high end should concern itself with and this view is reinforced at hi-fi shows and showrooms where almost all demos use recordings of small combos, often consisting of just a voice, a guitar and a little percussion. Many devoted home listeners also hold that the rear hall reverberation captured by the recording microphones is being properly reproduced when it comes, together with the direct sounds, from the front loudspeakers.

A new breed of video-age audiophile is convinced that hall ambience, extracted from specially encoded or directly from multichannel recordings and steered or fed to two or even four surround speakers can achieve the "you-are-there" illusion. This latter group is at odds with those who hold that any such processing or non-minimal microphone techniques is anathema.

Considering these prevailing and conflicting conceptions and misconceptions, it is remarkable how good, and even exciting, a sound can be produced by such ad-hoc but still basically stereophonic methods like 5.1. The musical sound generated by products from the overwhelming majority of serious stereo or 5.1 surround reproduction equipment manufacturers is truly first class as far as it goes. But the traditional methods of deploying this superb equipment at home has reached a dead end as far as closing that last yawning gap between perfect but flat fidelity and true spatial realism.

Ambiophonics-the Next Audiophile Stereophile Paradigm

I believe that Ambiophonics not 5.1 or similar surround sound method is the logical successor to Stereophonics. I also believe the majority of serious home music listeners are closet ambiophiles who really want to be in a realistic, electronically created concert hall, church, jazz club, theater or opera house when listening to recorded music at home or in a car, etc. The purpose of this book is to pass on the results of the research and experiments that I and others have performed. Ambiophiles everywhere can take comfort in the fact that it is both theoretically possible, possible in practice, and reasonable in cost to achieve the formerly impossible dream of recreating a "you-are-there" soundfield from standard

unencoded LPs, CDs, MP3s, or DVDs in virtually any properly treated room at home. In Ambiophonic parlance, when we say "real" we mean that an acoustic space of appropriate size and stage width has been created that is realistic enough to fool the human ear-brain system into believing that it is within that space with the performers on stage clearly delineated in front. The nice thing about Ambiophonics and existing two channel recordings is that so-called stereo recordings are not inherently stereophonic. That is, the microphones act somewhat like ears. They don't know that their signals are going to be played back in an untreated room and subjected to crosstalk, pinna angle distortion, and the other ills described below. Thus virtually any two channel recording of acoustic music, unless panned or multi-mic'ed to death, will respond well to Ambiophonic processing and reproduction.

The Ambiophonic techniques described in the following chapters produce a sound stage as wide as that seen by the recording microphones, an early reflection sound pattern that defines the hall size, and the character of the recording space, the listener's position within that hall, and a reverberant field that complements the content of the music and the original recording venue.

Although Ambiophonics does not rely on decoders, matrices or ambience extraction, it does incorporate commercially available PC or other digital signal processors, which are essentially special-purpose computers, to recreate the appropriate ambience signals. It is therefore a prime article of audiophile faith that while such signal generators are always subject to improvement, they have already reached an audiophile level of performance if one uses them Ambiophonically as described in the chapters that follow. It is also not the belief of the author that there is only one fixed way to achieve the Ambiophonic result. But I believe the Ambiophonic principles enumerated below can form a better foundation to build on than now eighty-year old stereo and its unfortunately, closely related, surround-sound technology.

In brief, Ambiophonics uses speaker correction, radical, crosstalk cancelled front channel loudspeaker positioning, computer recreation of real, early reflections and the later reverberant fields, and additional loudspeakers, strategically placed, to create accurately a wide front stage and propagate such ambience. Not every audiophile will be able or willing to do all that I suggest, but as each feature of the Ambiophonic system is implemented the improvement in realism will be easily audible and clearly rewarding.

If any science can be called ancient, acoustics is certainly one of them. The literature on acoustics, concert-hall design and sound recording is so vast that I am prepared to concede in advance that no individual fact or idea in the chapters below has not already appeared, at some time in some journal. I can only hope that the concatenation of all the ideas and devices that define Ambiophonics has some modicum of novelty. While I don't need to credit pioneers as far back as Helmholtz and Berliner, I would like to acknowledge my debt to such relatively recent researchers as W. B. Snow, James Moir, Don Keele Jr, stereo dipole-ist Ole Kirkeby, Manfred R. Schroeder, and his former colleague Yoichi Ando whose ideas on how to build better public concert halls inspired me to adapt his methods to create fine virtual halls for at home concerts.

Preface

The Psychoacoustic Flaws in Both Stereo and 5.1 Music Reproduction and Why Multi-Channel Recording Cannot Correct For Them

In the 21st century, it seems reasonable for videophiles and audiophiles to ask where the bridge from stereo reproduction to the next sonic century is leading or even if there is such a bridge. Stereophonic sound reproduction dates from 1931 and unfortunately as we shall see in this book has serious unredeemable flaws. But it only makes sense to replace it if there is something better that is reasonably practical and of true high-end quality. Fortunately, there is such a paradigm as described in the chapters that follow.

What Is Realism in Sound Reproduction

In this book, realism in staged music sound, game, or movie reproduction is understood to mean the generation of a sound field realistic enough to satisfy any normal ear-brain system that it is in the same space as the performers, that this is a space that could physically exist, and that the sound sources in this space are as full bodied and as easy to locate as at a live event. Realism does not necessarily equate to accuracy. For instance, a recording made in Symphony Hall but reproduced as if it were in Carnegie Hall is still realistic even if inaccurate. In a similar vein, realism achieved carelessly does not always mean perfection. If a full symphony orchestra is recorded in Carnegie Hall but played back as if it were crammed into Carnegie Recital Hall, one may have achieved realism but certainly not perfection. Likewise, as long as localization is as effortless as in real life, the reproduced locations of discrete sound sources might not have precisely the same sometimes exaggerated perspective as at the recording site to meet the standards of realism discussed here. An example of this occurs if a recording site has a stage width of 120 degrees but is played back on a stage that is only 90 degrees wide. What this really means in the context of realism is that the listener has moved back in the reproduced auditorium some twenty rows, from the first row but either stage perspective can be legitimately real. Finally, mere localization of a sound source does not guarantee that such a source will sound real. For example, a piano reproduced entirely via one loudspeaker, as in mono, or by two in stereo is easy to localize but almost never sounds real. The mantra goes, Mere Localization Is No Guarantor of Realism. Interestingly, one can have monophonic realism as when you hear a live orchestra from the last row of the balcony but can't tell (without looking) whether the horns are left, right, or center.

Since most of us are quite familiar with what live music in an auditorium sounds like, we soon realize that something is missing in our stereo systems. What is missing is soundfield completeness and psychoacoustic consistency. One can only achieve realism if all of the ear's hearing mechanisms are simultaneously satisfied without contradictions. If we assume that we know exactly how the ears work, then we could conceivably come up with a sound recording and reproduction system that would be quite realistic. But if we take the position that we don't know all the ear's characteristics or more significantly that we don't know how much they vary from one individual to another or that we don't know the relative importance of the hearing mechanisms we do know about, then the only thing we can do, until a greater understanding dawns, is what Manfred Schroeder suggested over a quarter of a century ago, and deliver to the remote ears an exact replica of what those same ears would have heard if present where and when the sound was originally generated. The old saw that, since we only have two ears, we only need two channels in reproduction has been justly disparaged. I would rephrase this hazy axiom to read, that since humans have only two ear canals, to achieve realism in reproduction, we need only provide the same sound pressure at the entrance to a particular listener's ear canal, even in the presence of head movement, that this same listener would have experienced at his ear canals had he himself been present at the recording session. Fortunately, it does turn out that only two recorded channels are in fact needed for realistic frontal music reproduction (more are actually detrimental) and it is the purpose of this book to show why this is so and how to do it. For music, movies, or games in the

round only four recorded channels are needed. These principles also apply to electronically generated music or sound effects.

This axiom requires that all reproduced, mid, and higher frequency direct or ambient sound come from as close to the correct direction as possible so as to reach the ear canal over a path that traverses the normal pinna structures and head parts. Thus home reproduced hall reverberation should reach the ears from many sideward and rearward locations and the early reflections from a variety of appropriate front, side and rear directions. This is why just the two rear surround speakers of 5.1 can never provide psychoacoustically satisfying hall ambience. Likewise central sound sources should come from straight ahead rather than from two speakers spanning 60 degrees. (A center speaker is no help in this regard as we will show below). Another precept that must be kept in mind is that your pinnae are unique like fingerprints. Using somebody else's pinna or pinna response, unless you get desperate, is not a good audiophile practice. A case in point is the use of dummy head microphones with pinnae. If the sound is reproduced by loudspeakers then all the sounds pass by two pinnae one of which is not even yours, and the result is strange and often in your head. If you listen, using normal pinna compressing earphones or ear buds, then you are listening with someone else's pinnae and there is no proper directional component at higher frequencies. The usual result is that the sound seems to be inside your head. If the dummy head doesn't have molded pinnae, and you listen with earphones, there are no pinnae at all and the sound again seems to be inside your head or strange. You can't fool Mother Nature.

Perfecting Stereo

While there are some widely held hi-end beliefs that may have to give way to psychoacoustic reality, the basic audiophile ideal that two channel recordings can deliver concert-hall caliber musical realism is not that far off the mark. However, having only two recorded channels does not mean being limited to only two playback loudspeakers. I call the coming replacement for today's stereo 'ambio' optimized but uncompromised for the recording and reproduction of frontal acoustic performances such as concerts, operas, movies, and video. By definition, and as substantiated below, where audiophile purity is concerned, multi-channel recording, especially with a center front channel, not only is not needed but is actually psychoacoustically counter productive. The sonic 3D genie cannot be squeezed into the 5.1, 6.1, or 7.1 or 10.2 moving picture surround sound bottle.

There are two basic theoretical technologies that are prime candidates to replace stereo or 5.1 where mass marketing and complex technical concepts should not be (but of course are) major stumbling blocks. One is the wavefront reconstruction method often employing hundreds of microphones and speaker walls or Ambisonics. The Ambisonic wavefront reconstruction method generates the correct sound pressure and sound direction in a region that at least encompasses one listener's head. Both are binaural technology methods that directly duplicate the live experience at each ear. Both technologies aim to deliver to the entrance of your ear canal an accurate replica of the original sound field. The Ambisonic method does have the advantage that it can reproduce direct sound sources from any angle and so is quite well suited to non-concert events or movies. But since the Ambisonic wavefront reconstruction method requires a special microphone, a minimum of three (or better four recording) channels and a very complex decoder, is not as user-friendly as other binaural technologies, and does nothing for the existing library of LPs and CDs it will not be considered further here.

As we shall show, the advantages of a binaural technology method such as Ambiophonics is that only two recorded channels, two front loudspeakers, and a scaleable number of optional ambience speakers are necessary. Although using a single pinnaless dummy head microphone (Ambiophone) works best, this new playback technology does not obsolete the vast library of LPs and CDs; it enhances most of them almost beyond belief. Ambio is also room shape and decorator friendly in that the front speakers can be very close together and thus be placed almost anywhere in a room. Another difference between direct wavefront reconstruction such as Ambisonics and wavefield synthesis, and ambiophonic binaural

field synthesis as in Ambiophonics is that in the latter case one can season the experience by moving one's virtual seat or changing a space, entirely, to suit the music or your taste. As explained in later chapters, this is not logical with 5.1 multi-channel recording systems since to make such changes you would be incurring the expense of a processor to undo the original expense of recording and storing the now superfluous center and rear surround tracks.

I Want To Be Alone, Or, The Listening Mob Fallacy

The concept that dedicated video, game, or music listening in the concert hall, theater, jazz venue, or at home is a group activity is superficial. Yes, there may be 2500 people in the opera house, but while the curtain is up there is, ideally, no interaction between them. Each member of the audience might just as well be sitting alone unless you believe in ESP. Listeners in a concert hall are also restricted as to the size of their sweet spot. They can't slouch to the floor or stand up, their permitted side to side or back to front movement is not extensive and there are plenty of seats in most halls where the sound and the view are not quite so sweet.

At home, how often does the gang come over to sit with you for five hours of *Die Götterdämmerung*? Certainly, serious home listening to classical music and to a lesser extent longer popular genres such as Broadway shows, games, movies, jazz concerts, etc., is sad to say a solitary or at most a two couple pursuit. Of course we all want to demonstrate our great reproduction systems to friends and family, but since these sessions usually last just a few minutes, one can show off the system to one or two people at a time and after everyone has heard it, at its best, from the sweet spot, the party can go on.

The point here is that it is difficult enough to correct the inherent defects of stereo and create a concert-hall caliber soundfield at home without making compromises in the design in order to unduly enlarge the sweet spot. Note that stereo, Ambisonics, VMAx, 5.1, 7.1, etc. all have listening box limitations that one must live with. In the case of stereo, if one moves towards the loudspeaker one senses a hole in the middle. The stage is sensed as being half to the left and the other half to the right. As one moves back from the speakers the stage becomes narrower and eventually one seems to be listening to just one speaker. If one moves to the side then one soon localizes to the nearest speaker and can clearly hear just one channel. 5.1 has similar problems except that, since the dialog is already more in the center speaker, it remains so even if one is offset to one side, closer, or further away.

By contrast, in Ambiophonics, if one moves closer to the speakers, one hears ordinary stereo. If one moves back, the stage remains wide and very little changes. If you move to the side, you still hear both channels, which is why a center speaker is never needed. This happens because each speaker is fed both its direct signal and a slightly delayed version of the other channel plus the center channel (if present as in 5.1). In Ambiophonics, one can stand, recline, nod, or rotate the head without affecting localization. This is in contrast to using earphones where, if you move your head, the apparent stage moves with you. Headtracking is never needed for loudspeaker binaural.

Why Stereo Can't Deliver Realism Without Some Fixing

By now, every one in the industry has recognized that when a two channel recording is played back through two loudspeakers that form an 60 or 90 degree angle from the listener, that each such speaker communicates with both ears, producing interaural crosstalk. The deleterious effects of this crosstalk at both low and especially the higher frequencies have been greatly underappreciated. For openers, crosstalk is what almost always prevents any sound source from appearing to come from beyond the angular position of the loudspeakers. This result is intuitively obvious, since if we postulate an extreme-right sound source, and can safely ignore the contribution from the left speaker, we can now hear the right speaker by itself, as usual with both ears, and no matter how we turn our heads the sound will always be localized to the right speaker as in any normal hearing situation. However, if we keep the

right speaker sound from getting so easily to the left ear then the brain thinks that the sound must be at a larger angle to the right, well beyond the, say 30 degree position of the loudspeaker, since, as in the concert hall, the lesser sound reaching the left ear is now fully attenuated and delayed by the head and filtered by the left pinna. So, for starters, stereo, because of its crosstalk, inadvertently compresses the width of its own sound stage.

A second, perhaps more serious defect, is also caused by this same crosstalk. For centrally located (mono) sound sources, two almost equally loud acoustic signals reach each ear, (instead of one as in the concert hall) but one of these signals, in the normal stereo listening setup, travels about half a head's width or 300 usec., longer than the sound from the nearer speaker. This produces multiple peaks and nulls in the frequency response at each ear from 1500 Hz up known as comb filtering. Since the nulls are narrow, and are muddied by even later crosstalk coming around the back or over the top of the head, and since the other ear is also getting a similar but not precisely, identical set of peaks and nulls, the ear seldom perceives this comb filtering as a change in timbre; but it can and does perceive these gratuitous dips and peaks as a kind of second, but foreign, pinna function and this causes confusion in the brains mechanism for locating musical transients. Remember, in real halls the ear can hear a one degree shift in angular position, but not if strong comb-filter effects occur in the same 2-10 kHz region where the ear is most sensitive to its own intrapinna convolution effects and interpinna intensity differences. As long as this wrongful interaural crosstalk is allowed to persist, the sound stage will never be as natural or as tactile as it could be and for some people, such listening is fatiguing after awhile and all 60 (or LRC) degree stereo reproduction sounds canned to them.

Pinna-Sensitive Front Speaker Positioning

Just as there are optical illusions, so there are sonic illusions. One can create sonic illusions by using complex filters to create virtual sound sources that float in mid air or rise up in front of you. As with optical illusions some people detect them and some people don't. The most prominent audio illusion is in stereo where phantom images are created between two speakers. You may have observed that most optical illusions are two-dimensional drawings, that imply ephemeral three dimensions. Likewise there is something indistinct about the stereo phantom illusion. This is because the phantom image is largely based on lower frequency interaural cues and barely succeeds in the face of the higher frequency head and pinna localization contradictions. The fact that earphone systems such as Toltec based processors, a host of PC virtual reality systems, SRS, Lexicon, etc. can move images in circles just by manipulating high frequency head and pinna response curves, even if not of great high-end quality, does show that these hearing characteristics are of considerable importance. Thus the direction from which complex sounds with energy over 1500 Hz originate, particularly from the frontal stage, should be as close to correct as possible.

Most stage sounds, particularly soloists and small ensembles, originate in the center twenty degrees or so. Remember that we want to launch sounds as much as possible from the directions they originate. Thus it makes much more sense to move the front channel speakers to where the angle between each of them to the listening position is perhaps ten degrees. This eliminates the pinna processing error for the bulk of the stage. But, of course, if the speakers are so close together, what happens to the separation? The answer is that with the crosstalk eliminated, as is necessary anyway, separation, as in earphone binaural, is no longer dependent on angular speaker spacing.

The Ambiodipole

Crosstalk elimination is not a concept new to just Ambiphonics; but most of the older electronic crosstalk elimination circuits such as those of Lexicon, Carver, Polk etc. assume the stereo triangle and have, therefore, had to make compromises to enlarge the sweet spot size over which they are effective. I would hesitate to class any of them as high-end components, especially as they still promote pinna

position errors. Usually good crosstalk cancellers require complex compensation for the fact that the crosstalk signal being canceled has had to go around the head and over the pinna on its way to the remote ear. Since Carver, Lexicon, etc. don't know what your particular head and pinna are like, they assume an average response and thus can't do a very good job of cancellation at high frequencies. If they try, most listeners experience phasiness, a sort of unease or pressure particularly if they move about. But when the speakers are in front of you there is not much of the head to get in the way and so the head response functions are much simpler, less deleterious if ignored or averaged, and head motions make little difference. Ambidipoles are just now appearing but you can easily achieve an inexpensive and truly high-end result using a simple three foot square six inch thick absorbent panel set on edge at the listening position. You get used to the panel rather quickly and it is a high-end tweak that needs no cables and produces no grunge. Either electronic processors or panels allow complete freedom of head motion without audible effect and afford more squirm room at the listening position than one has in a concert hall. Two people can be accommodated comfortably but usually one needs to be directly behind the other for optimum results, not unlike high-resolution stereo.

Earlier crosstalk cancellation systems were less than satisfactory because they were not recursive. That is, in the earlier systems the unwanted signal from the left speaker at the right ear was cancelled by an inverted signal from the right speaker and that was the end of it; but this right speaker cancellation also reaches the left ear and so one has a new form of crosstalk.

In Ambiophonics, this later crosstalk is cancelled over and over again until its level is inaudible. A comparison can be made with reverberation time in concert halls. Normally, the reverberation time is specified as the time it takes for a sound to decrease by 60 decibels. This implies that the human ear is sensitive to concert hall reverberation at this low a level. Likewise, crosstalk is still deleterious even if its level gets to be quite low after several cycles of successive cancellation. We call this process recursive Ambiophonic crosstalk elimination (RACE).

One can have video with crosstalk cancellation (XTC) but adding a picture can have its misleading side. One reason that so many listeners are impressed with the realism of movie surround sound systems is the presence of the visual image. While the research in this field is not definitive, it stands to reason that a brain preoccupied with processing a fast moving visual image is not going to have too much processing power left over to detect fine nuances of sound. Certainly, if you close your eyes while listening to any system, your sensitivity to the faults of the sound field is heightened. Thus when a seemingly great home theater system is used to play music only, without a picture, the experience is often less than thrilling. Adding a picture to Ambio seems to make fine adjustments to the ambient field much less audible, but one must observe that most people keep their eyes open at concerts and so perhaps an image is desirable to provide the ultimate home musical experience.

Nothing we have done to make the front stage image more realistic and psychoacoustically correct has required any extra recorded channels. I call all these changes to standard stereophonics, Ambiophonics or Ambio. Ambio, does not rely on the fluky phantom image mechanism. But there still remains one further difficulty with the stereo triangle and that is that we need a proper ambient field coming from more directions than just those of our now crosstalk-free, pinna-correct, front speakers.

The Case For Ambience By Reconstruction

Like a federal budget agreement, a method of achieving that air, space, and appropriate concert hall ambience at home, has technical devils in its details. The most obvious suggestion, based on movie and video surround-sound techniques, to just stick the ambient sound on additional DVD multi-channel tracks, on closer examination, just can't do it for hi-enders. The problem with using third, fourth or fifth microphones at or facing the rear of the hall and then recording these signals on a multi-channel DVD, is that these microphones inevitably pick up direct sound which, when played back from the rear or side

speakers, causes crosstalk, pinna angle confusion, and comb filter notching. It is also pinnatically incorrect to have all rear hall ambience coming from just two point sources even if these surround speakers are THX dipoles. Remember, using rear dipoles implies a live listening room, which will thus also increase unwanted early reflections from the front speakers. Additionally, recording hall ambience directly is really not cost effective or necessary. Unlike movies, the acoustical signature of Carnegie Hall (despite its always ongoing renovations) does not change with every measure, so why waste bits recording its very static ambience over and over again? It is much more cost and acoustically effective to measure the hall response once from the best seat (or several) for say five, left, right, and center positions on the stage (If the hall is symmetric, the measurement process is simpler) and either include this data in a preamble on the DVD, store it in your playback system or provide it as part of a DVD-ROM library of the best ambient fields of the world.

The process of combining a frontal, two (I hope) channel recording with the hall impulse response is called convolution and convolution is the job of the ambience regenerator which may be a PC or a special purpose DSP computer or it may be a part of the DVD/CD DAC. The use of ambience reconstruction would obviate the need for DTS or Dolby Digital multi-channel recordings at least where classical music is concerned. Unlike frontal sound, ambience can and should come from as many speakers as one can afford or has room for. Crosstalk, and comb-filtering are not problems with ambient sound sources if these signals are uncorrelated (unrelated closely in time, amplitude, frequency response, duration, etc.) which is normally the case both with concert halls and good ambience convolvers.

An Uphill Political Struggle

The cause of concert-hall early reflection and reverberation tail synthesis by digital signal processors (DSP) in computers or audio products was set back by the late Michael Gerzon, the Oxford Ambisonics pioneer, who wrote in 1974 "Ideally, one would like a surround-sound system (yes, he did use this term in 1974) to recreate exactly, over a reasonable listening area, the original sound field of the concert hall.... Unfortunately, arguments from information theory can be used to show that to recreate a sound field over a two-meter diameter listening area for frequencies up to 20 kHz, one would need 400,000 channels and loudspeakers. These would occupy 8 GHz of bandwidth equivalent to the space used by 1000, 625-line television channels!"

Later, however, Gerzon did not let information theory prevent him from capturing a 98% complete concert-hall sound field using a single coincident array of four microphones. Indeed the complete impulse response of a hall can be measured and stored on one floppy disk by placing an orthogonal array of three microphone pairs at the best seat in the house and launching a test signal from the stage during the recording session or at any time.

Convolution to The Rescue

An audiophile-friendly approach to ambience reconstruction is to derive the surround speaker feeds by convolution of a two channel recording, preferably made using the microphone technique described below, that limits rear hall pickup. The questions to be asked are these:

- How many channels of early reflections with reverberant tails do we really need to reconstruct and where in the listening room should these speakers be placed? How many channels of late reverberation do we absolutely need to generate and where should these speakers be placed?
- How realistic sounding is the software now available to generate these fields?
- What about the problem that each instrument on the stage produces a different set of early reflections?

There may never be a definitive answer to the first question. Just as there is no sure recipe for physical concert hall design, there is no best virtual concert hall specification. But, adjusting the number, placement, and shape of early reflections is easily more audible than changing amplifiers or cables and offers a tweaker delights that can last a lifetime. I can only say that in my own experience, just as there are thousands of real concert halls that differ in spite of being real, so there are thousands of ambience combinations that sound perfectly realistic even if not perfect. How do you get more real than real? Remember, absolute, particular hall parameter accuracy is not essential to achieve realism. By analogy, even if one sits on the side, in the last row of the balcony at Carnegie Hall where the ambience is lopsided, the sonic experience is still real. In my opinion the best software for this purpose is based on impulse response measurements made in actual concert halls as was done by JVC and Yamaha some 10 years ago for consumer products and is being done all the time by acoustical architects tuning auditoriums. Others, such as Dr. Dave Griesinger at Lexicon, create ambience signals using an imaginary model. I am not talking here about professional effects synthesizers that generate artifacts never heard by anybody in any physically existing space. Someday, I presume, we will have a DVD-ROM that contains the ambient parameters of Leo Beranek's 76 greatest concert houses of the world and a simple mouse click will yield a selection. With enough hall impulse responses stored, you could even select a seat and a stage width. (If it's a solo recital one wants only central derived early reflections, if a symphony orchestra, the works, etc.). There are already over 100 impulse responses of concert halls available on the Internet.

While I may not be the best one at executing my own theories, I have gotten startlingly good results using the new convolvers available. It is a rare AES convention that does not describe advances in the state of this art. Another important point is that ambience regeneration is scaleable. As computers get faster, and cheaper and as convolution software gets better, it is easy to upgrade or add more ambience speakers. The hall ambience storage method is also inherently tolerant of speaker type and the precise location or speaker response matter little and are akin to repainting the balcony or curving a wall in the concert hall.

The fact is that the brain is not all that sensitive to whether there are 30 early reflections from the right and only 25 from the left or whether they come from 50 degrees instead of the concert-hall ideal (according to Ando) of 55 degrees. If the reverberant field is not precisely diffuse or decays in 1.8 seconds instead of 2.0 seconds, that may only mean you are in Carnegie Hall instead of Symphony Hall. I make no claim to be an authority on setting ambience hall parameters, and I am sure many audiophiles could do better at this game. I now use 2 large area speakers at the sides and rear to provide a reverberation field as diffuse as possible.

Since central early proscenium reflections come from the recording via the main front speakers, these need not be regenerated and, of course, by definition they are natural and are coming from the proper directions. For side, overhead, and rear ambience using the left channel to recreate left leaning early reflections (some of which may end up coming from the right) and the right channel to produce a set of right reflections, the early reflection patterns for different instruments on the stage have enough diversity to exceed the threshold of the brain's reality barrier.

Whither Recording In An Ambiophonic Hi-End World

While audiophiles do not often concern themselves with recording techniques over which they have little control almost any LP or CD made with either coincident or spaced microphones is greatly enhanced by Ambio playback. But one can heighten the accuracy, if not gild the lily of realism, by taking advantage in the microphone arrangement, of the knowledge that, in playback, the rear half and side parts of the hall ambience will be synthesized, that there is no crosstalk, that the front loudspeakers are relatively close together, and that thus listening room reflections are minimized. To make a long story short, exceptionally realistic "You-are-there" recordings can be made by using a head shaped, pinnaless ball

with holes at the ear canal positions to hold the microphones. The Schoeps KFM-6 is a good example of such a microphone even though it is a sphere and an oval would be slightly better. However, for best results, this microphone should be well baffled to prevent most rear hall ambience pickup. KFM-6 recordings are a feature of the PGM label, produced by the late Gabe Wiener who was a staunch advocate of this recording method, first expounded by Guenther Theile. As expected, these PGM recordings are exceptionally lifelike when played back Ambiophonically so as to be free of crosstalk or pinna distortion. The Ambiophone is a microphone array specifically designed to make recordings optimized for Ambio playback.

The reason such a microphone is optimum is that particularly for central sounds the sound rays reach the ears almost as they do in the concert hall. That is, one ray from a central instrument reaches the left ear of the microphone, goes to the left speaker where it is sent straight ahead to the left pinna and ear. The fact that the head response transfer function of the microphone is not the same as the listener's is not significant for central sound sources that don't cross either head. For side sources the microphone ball becomes a substitute for the listener's HRTF but at least there is still only one HRTF and one real pinna in the chain. Perhaps the hardest part of migrating to Ambio will be to convince recording engineers, who are usually rugged individualists, to use microphones and positions that are Ambio compatible.

Law of the First Impression

No matter how many great stereo systems I listen to, they still never have the impact that my first Emory Cook stereo disc had. Likewise, I still compare the multichannel systems I hear now to the mental image of air and presence I retain of the first RCA CD-4 true discrete quad LP of Mahler's 2nd I heard in the early 70's. The moral of this phenomena is that the first time anyone hears a major upgrade in reproduction, particularly when going beyond two speakers for the first time, they are always very favorably impressed. Dissatisfaction with systems like the Hafler arrangement, SQ, Dolby pro-logic etc only set in later. This is the scenario with the new discrete multi-channel format for music as well. At first 5.1 or even 7.1 sounds really exciting and a great contrast to stereo but in the end it fails as a realistic replica of the live music concert-hall experience.

Chapter 1

Ambiophonics, considering music rather than video for the moment, is the logical replacement for stereophonics and a technical methodology which, if adhered to closely, makes it possible to immerse oneself in an exceedingly real acoustic space, sharing it with the music performers on the stage in front of you. Ambiophonics can do this, even using ordinary standard and existing two channel recordings. We will show in the chapters that follow that, as hard as this may be to believe, there is nothing to be gained as far as realism in acoustic music reproduction is concerned by using more than two recorded channels (as opposed to multi-speaker) and that the complex microphone arrangements that multichannel recording implies are actually deleterious and wasteful of bandwidth that could be put to better use. Ambiophonics is like a visit to a concert hall and is for serious listeners who do not often read, talk, eat, knit, or sleep in their home concert halls, any more than they would at a live performance.

Ever since 1881 when Clement Ader ran signals from ten spaced pairs of telephone carbon microphones clustered on the stage of the Paris Opera via phone lines to single telephone receivers in the Palace of Industry that were listened to in pairs, practitioners of the recording arts have been striving to reproduce a musical event taking place at one location and time at another location and time with as little loss in realism as possible. While judgments as to what sounds real and what doesn't may vary from individual

to individual, and there are even some who hold that realism is not the proper concern of audiophiles, such views of our hearing life should not be allowed to slow technical advances in the art of realistic auralization that listeners may then embrace or disdain as they please.

What is Realism in Sound Reproduction?

Realism in staged music sound reproduction will usually be understood to mean the generation of a sound field realistic enough to satisfy any normal ear-brain system that it is in the same space as the performers, that this is a space that could physically exist, and that the sound sources in this space are as full bodied and as easy to locate as in real life. Realism does not necessarily equate to accuracy or perfection. Achieving realism does not mean that one must slavishly recreate the exact space of a particular recording site. For instance, a recording made in Avery Fisher Hall but reproduced as if it were in Carnegie Hall is still realistic, even if inaccurate. While a home reproduction system may not be able to outperform a live concert in a hall the caliber of Boston's Symphony Hall, in many cases the home experience can now exceed a live event in acoustic quality. For example, a recording of an opera made in a smallish studio can now easily be made to sound better at home than it did to most listeners at a crowded recording session. One can also argue that a home version of Symphony Hall, where one is apparently sitting tenth row center, is more involving than the live experience heard from a rear side seat in the balcony with obstructed visual and sonic prospect. In a similar vein, realism does not mean perfection. If a full symphony orchestra is recorded in Carnegie Hall but played back as if it were in Carnegie Recital Hall, one may have achieved realism but certainly not perfection. Likewise, as long as localization is as effortless and as precise as in real life, the reproduced locations of discrete sound sources usually don't have to be exactly in the same positions as at the recording site to meet the standards of realism discussed here. (Virtual Reality applications, by contrast, often require extreme accuracy but realism is not a consideration.) An example of this occurs if a recording site viewed from the microphone has a stage width of 120 degrees but is played back on a stage that seems only 90 degrees wide. What this really means in the context of realism is that the listener has moved back in the reproduced auditorium some fifteen rows, but either stage perspective can be legitimately real. But being able to localize a stage sound source in a stereo or surround multi channel system does not guarantee that such localization will sound real. For example, a soloist's microphone panned by a producer to one loudspeaker is easy to localize but almost never sounds real.

In a similar vein, one can make a case that one can have glorious realism, even without any detailed front stage localization, as long as the ambient field is correct. Anyone who has sat in the last row of the family circle in Carnegie Hall can attest to this. This kind of realism makes it possible to work seeming miracles even with mono recordings.

Reality is in the Ear of the Behearer

While it is always risky to make comparisons between hearing and seeing, I will live dangerously for the moment. If from birth, one were only allowed to view the world via a small black and white TV screen, one could still localize the position of objects on the video screen and could probably function quite well. But those of us with normal sight would know how drab, or I would say unrealistic, such a restricted view of the world actually was. If we now added color to our subject's video screen, the still grossly handicapped (by our standards) viewer would marvel at the previously unimaginable improvement. If we now provided stereoscopic video, our now much less handicapped viewer would wonder how he had ever functioned in the past without depth perception or how he could have regarded the earlier flat monoscopic color images as being realistic. Finally, the day would come when we removed the small video screens and for the first time our optical guinea pig would be able to enjoy peripheral vision and the full resolution, contrast and brightness that the human eye is capable of and fully appreciate the miracle of unrestricted vision. The moral of all this is that only when all the visual sense parameters are provided for, can one enjoy true visual reality and the same is true for sonic reality.

Since most of us are quite familiar with what live music in an auditorium sounds like, we can sense unreality in reproduction quite readily. But in the context of audio reproduction, the progression toward realism is similar to the visual progression above. To make reproduced music sound fully realistic, the ears, like the eyes, must be stimulated in all the ways that the ear-brain system expects. Like the visual example, when we go from mono to stereo to matrix surround to multi-channel discrete, etc. we marvel at each improvement. But since we already know what real concert halls sound like, we soon realize that something is missing. In general, multi-channel recording methods or matrix surround systems (Hafler, SQ, QS, UHJ, Dolby, 5.1, etc.) seem like exciting improvements when first heard by long realism deprived stereo music auditors, but in the end don't sound real. What is usually missing is completeness and sonic consistency. One can only achieve realism if all the ear's expectations are simultaneously satisfied. If we assume that we know exactly how all the mechanisms of the ear work, then we could conceivably come up with a sound recording and reproduction system that would be quite realistic. But if we take the position that we don't know all the ear's characteristics or that we don't know how much they vary from one individual to another or that we don't know the relative importance of the hearing mechanisms we do know about, then the only thing we can do, until a greater understanding dawns, is what Manfred Schroeder suggested over a quarter of a century ago, and deliver to the remote ears a realistic replica of what those same ears would have heard when and where the sound was originally generated.

Four Methods Used to Generate Reality at a Distance

Audio engineers have grappled with the problem of recreating sound fields since the time of Alexander Graham Bell. The classic Bell Labs theory suggests that a curtain, in front of a stage, with an infinite number of ordinary microphones driving a like curtain of remote loudspeakers can produce both an accurate and a realistic replica of a staged musical event and listeners could sit anywhere behind this curtain, move their heads and still hear a realistic sound field. Unfortunately, this method, even if it were economically feasible, does not deliver either accuracy or realism. Such a curtain acts like a lens and changes the direction or focus of the sound waves that impinge on it. Like light waves, sound waves have a directional component that is easily lost in this arrangement either at the microphone, the speaker or both places. Thus each radiating loudspeaker, in practice, represents a new discrete source of sound with uncontrolled directionality, possibly diverting sound meant for oblivion in the ceiling down to the listener and causing other sounds to impinge on the head at odd angles.

Finally this curtain of loudspeakers does not radiate into a concert-hall size listening room and so one would have, say, an opera house stage attached to a listening room not even large enough to hold the elephants in Act 2 of Aida. This lack of opera-house ambience wouldn't by itself make this reproduction system sound unreal, even if the rest of the field were somehow made accurate, but it certainly wouldn't sound perfect. The use of speaker arrays (walls of hundreds of speakers) surrounding a relatively large listening area has been shown to be able to reproduce ambient sound fields with remarkable accuracy. But while this technique may be useful in sound amplification systems in halls, theaters or labs, application to playback in the home seems doubtful. This approach is called Wavefield Synthesis or WFS.

The Binaural Approach

A second more practical and often exciting approach is the binaural one. The idea is that, since we only have two ears, if we record exactly what a listener would hear at the entrance to each ear canal at the recording site and deliver these two signals, intact, to the remote listener's ear canals then both accuracy and realism should be perfectly captured. This concept almost works and could conceivably be perfected, in the very near future, with the help of advanced computer programs, particularly for virtual reality applications involving headsets or near field speakers. The problem is that if a dummy head, complete with modeled ear pinnae and ear canal embedded microphones, is used to make the recording, then the listener must listen with in-the-ear-canal earphones because otherwise the listeners own pinnae would also process the sound and spoil the illusion.

The real conundrum, however, is that the dummy head does not match closely enough any particular human listener's head shape or external ear to avoid the internalization of the sound stage whereby one seems to have a full symphony orchestra (and all of Carnegie Hall) from ear to ear and from nose to nape. Internalization is the inevitable and only logical conclusion a brain can come to when confronted with a sound field not at all processed by the head or pinnae. For how else could a sound have avoided these structures unless it originated inside the skull? If one uses a dummy head without pinnae, then, to avoid internalization, one needs earphones that stand off from the head, say, to the front. But now the direction of ambient sound is incorrect. The original 3D IMAX is an example of this off the ear method, as supplemented with loudspeakers for bass and rear direct sound effects.

The fact that binaural sound via earphones runs into so many difficulties is a powerful indication that average head shadows and individual outer ear convolutions are critically important to our ability to sense sonic reality but as we shall see loudspeaker binaural is an essential element of the Ambiophonic paradigm.

Wavefront Synthesis

A third theoretical method of generating both an accurate and a realistic soundfield is to actually measure the intensity and the direction of motion of the rarefactions and compressions of all the impinging soundwaves at the single best listening position during a concert and then recreate this exact sound wave pattern at the home listening position upon playback. This method is the one expounded by the late Michael Gerzon starting in the early 70's and embodied in the paradigm known as Ambisonics. In Ambisonics, (ignoring height components) a coincident microphone assembly, which is equivalent to three microphones occupying the same point in space, captures the complete representation of the pressure and directionality of all the sound rays at a single point at the recording site. In reproduction, speakers surrounding the listener, produce soundwaves that collectively converge at one point (the center of the listener's head) to form the same rarefactions and compressions, including their directional components, that were heard by the microphone.

In theory, if the reconstructed soundwave is correct in all respects at the center of the head (with the listener's head absent for the moment) then it will also be correct three and one half inches to the right or left of this point at the entrance to the ear canals with the head in place. The major advantage of this technique is that it can encompass front stage sounds, hall ambience and rear direct sounds equally, and that since it is recreating the original sound field (at least at this one point) it does not rely on the quirky phantom image illusion of traditional Blumlein stereo.

The Ambisonic method is not easy to keep accurate at frequencies much over 1500 Hz and thus must and does rely on the apparent ability of the brain to ignore this lack of realistic high frequency localization input and localize on the basis of the easier to reconstitute lower frequency waveforms alone. This would be fine if localization, by itself, equated to realism or we were only concerned with movie surround sound applications.

Other problems with basic (first order) Ambisonics include the fact that it requires at least three recorded channels and therefore can do nothing for the vast library of existing recordings. Back on the technical problem side, one needs to have enough speakers around the listener to provide sufficient diversity in sound direction vectors to fabricate the waveform with exactitude and all these speakers positions, relative to the listener, must be precisely known to the Ambisonic decoder. Likewise the frequency, delay and directional responses of all the speakers must be known or closely controlled for best results and as in many loudspeaker systems the effects of listening room reflections must also be taken into account, or better yet, eliminated. Higher order ambisonics (HOA) require many more media channels and speakers and so is not very useful in a home system context.

As you might imagine, it is quite difficult, particularly as the frequency goes up, to insure that the size of the Ambisonic field at the listening position is large enough to accommodate the head, all the normal motions of the head, the everyday errors in the listener's position, and more than one listener. Those readers who have tried to use the Lexicon panorama mode, the Carver sonic hologram or the Polk SDA speaker system, all designed to correct parts of a simple stereo soundfield at the listener's ear by acoustic cancellation will appreciate how difficult this sort of thing is to do in practice, even when only two speakers are involved.

In my opinion, however, the basic barrier to reality, via any single point waveform reconstruction method, like Ambisonics, is its present inability, as in the earphone binaural case, to accommodate to the effects of the outer ear and the head itself on the shape of the waveform actually reaching the ear canal. For instance, if a wideband soundwave from a left front speaker is supposed to combine with a soundwave from a rear right speaker and a rear center speaker etc. then for those frequencies over say 2500 Hz the left ear pinna will modify the sound from each such speaker quite differently than expected by the equations of the decoder, with the result that the waveform will be altered in a way that is quite individual and essentially impossible for any practical decoder to control. The result is good low frequency localization but poor or non-existent pinna localization. Unfortunately, as documented below, mere localization, lacking consistency, as is unfortunately the case in stereo, 5.1 surround sound or Ambisonics is no guarantor of realism. Indeed, if a system must sacrifice a localization mechanism, let it be the lowest frequency one.

Ambiophonics

The fourth approach, that I am aware of, I have called Ambiophonics. Ambiophonics assumes that there are more localization mechanisms than are dreamed of in the previous philosophies and strives to satisfy them all, even the unknown ones. The advantage of focusing on sonic reality is that this reality is achievable today, is reasonable in cost, and is applicable to existing LPs, CDs, DVDs, movies, games, in homes, cars, PCs, etc.

One basic element in Ambiophonic theory, in the case of music, is that it is best not to record rear and side concert-hall ambience or try to extract it later from a difference signal or recreate it via waveform reconstruction, but to regenerate the ambient part of the field using real, stored concert hall, data to generate early reflections and reverberant tail signals using the new generation of digital signal processors. The variety and accuracy of such synthesized ambient fields is limited only by the skill of programmers and data gatherers, and the speed and size of the computers used. Thus, in time, any wanted degree of concert hall design perfection could be achieved. A library of the worlds great halls may be used to fabricate the ambient field as has already been done in the pioneering JVC XP-A1010. The number of speakers needed for ambience generation does not need to exceed six or eight (although Tomlinson Holman of THX fame is now up to ten and I usually go with 16) and is comparable to Ambisonics or 7.1 surround sound in this regard. But even more speakers could be used as this ambience recovery method, called convolution, is completely scaleable and the quality and location of these speakers is not critical.

Ambiophonics is less limited as to the number of listeners who can share the best experience at the same time than stereo, 5.1 or most implementations of other methods using a similar number of speakers but Ambiophonics is certainly not suited to group listening. However, like a non-ideal seat in a concert hall one has a marked sense of space anywhere in the room while the orchestra is playing somewhere over there.

The other basic tenet of Ambiophonics is similar to Ambisonics and that is to recreate at the listening position an exact replica of the original pressure soundwave. Ambiophonics does this by transporting you to the sound source, stage, and hall. In other words, Ambiophonics externalizes the binaural effect,

using, as in the binaural case, just two recorded channels but with two front stage reproducing loudspeakers and eight or so ambience loudspeakers in place of earphones. Ambiophonics generates stage image widths up to almost 180 degrees with an accuracy and realism that far exceeds that of any other 2 channel or even multi channel recording scheme.

Psychoacoustic Fundamentals Related to Realism in Reproduced Sound

The question is how to achieve realistic sound with the psychoacoustic knowledge at hand or suspected. For starters, the fact that separated front loudspeakers can produce centrally located phantom images between themselves is a psychoacoustic fluke akin to an optical illusion that has no purpose or counterpart in nature and is a poor substitute for natural frontal localization. Any reproduction method that relies on stimulating phantom images, and this includes not only stereo but most versions of surround sound, can never achieve realism even if they achieve localization. Realism cannot be obtained merely by adding surround ambience to frontal phantom localization. Ambisonics, earphone binaural, and Ambiophonics do not employ the phantom image mechanism to provide the front stage localization and therefore, in theory, should all sound more realistic than stereo and, in fact, almost always do.

The optimized Ambiophonic microphone arrangement discussed later could make this approach to realism even more effective, but I am happy to report that Ambiophonics works quite well with most of the microphone setups used in classical music, video, or audiophile caliber jazz recordings. Adding home-generated ambience, provides the peripheral sound vision to perfect the experience.

Since our method is to just give the ears everything they need to get real, it is not essential to prove that the pinna are more important than some other part of the hearing mechanism, but the plain fact is that they are. To me it seems inconceivable that anyone could assume that the pinna are vestigial or less sensitive in their frequency domain than the other ear structures are in theirs. As a hunter-gatherer animal, it would be of the utmost importance to sense the direction of a breaking twig, a snake's hiss, an elephant's trumpet, a bird's call, the rustle of game etc. and probably of less importance to sense the lower frequency direction of thunder, the sigh of the wind, or the direction of drums. The size of the human head clearly shows the bias of nature in having humans extra sensitive to sounds over 700 Hz.

Look at your ears. The extreme non-linear complexity of the outer ear structures, and their small dimensions defies mathematical definition and clearly implies that their exact function is too complex and too individual to understand, much less fool, except in half-baked ways. The convolutions and cavities of the ear are so many and so varied so as to make sure that their high frequency response is as jagged as possible and as distinctive a function of the direction of sound incidence as possible. The idea is that no matter what high frequencies a sound consists of or from what direction a transient sound comes from, the pinnae and head together or even a single pinna alone will produce a distinctive pattern that the brain can learn to recognize in order to say this sound comes from over there.

The outer ear is essentially a mechanical converter that maps sound arrival directions to preassigned frequency response patterns. There is also no purpose in having the ability to hear frequencies over 10 kHz, say, if they cannot aid in localization. The dimensions of the pinna structures and the measurements by Moller, strongly suggest, if not yet prove, that the pinna do function for this purpose even in the highest octave. Moller's curves of the pinna and head functions with frequency and direction are so complex that the patterns are largely unresolvable and very difficult to measure using live subjects. Again, it doesn't matter whether we know exactly how anyone's ears work as long as we don't introduce psychoacoustic anomalies or compromise on the delivery of frequency response, dynamic range, loudness, low distortion, and especially source and ambience directionality, during reproduction.

Basics of Concert Hall Psychoacoustics

In order to produce a concert-hall sound field or any other sonic experience in the home without actually building a concert hall, we need to know what the ear requires at the minimum for accepting a sound field as real. Knowing this, it is then possible to look for ways to accomplish this feat in a small space and within a budget, without compromising the reality of the aural illusion. While not everything is known about how the ear perceives distance, horizontal and vertical angular position, hall enclosure size and type, and maybe absolute polarity, enough is known to allow Ambiophonics to create a variety of sound fields suited to different types of music or drama that are real enough to be accepted as such by the ear-brain system.

In general the only parts of the hearing mechanism that concern us specifically are the ear pinnae and the existence of two ears separated by a head. Even without consulting the hundreds of papers on this subject, it is clear that the pinnae are designed to modify the frequency response of sound waves as a function of the direction from which the sound comes. It is also clear that no two individuals have ear pinnae that are identically shaped. But to give a general idea of what one person's pinna does in the horizontal plane: for a sound coming from directly in front, the frequency response at the ear canal entrance, measured with a tiny microphone inserted into the ear canal, is essentially flat up to 1000 Hertz. For most people, the response then rises as the rear of the pinna interdicts sound and reflects it additively into the ear canal. A broad 11 dB peak in the response is reached at about 3000 Hz after which the response drops off to minus 10 dB at 10 kHz and then begins to rise again. A response spread such as this of 21 dB in the treble region is quite substantial, and if a loudspeaker had this kind of response it would get very poor reviews indeed. It is also easy to see that differences in individual pinnae are not easy to correct with tone controls or equalizers. For a sound coming from the side to the near ear, a slow rise in response starts at 200 Hz, reaches 15 dB at 2500 Hz, drops to 1 dB at 5 kHz, rises to 12dB at about 7 kHz and then drops to 4 dB at about 10 kHz. (after Henrik Moller et al) This side response is quite different from the dead ahead response and indicates that we are very sensitive to the direction from which sounds originate even if we listen with only one ear. For sounds directly rearward, the pinnae cause a dropoff of 23 dB between 2500 Hz and 10 kHz. Other radically different frequency responses occur for sounds coming from above or below. The pinnae seem to be entirely responsible for our sense of center-front sound source height.

What this means for realistic sound reproduction is that whatever sound we generate must come to the listening position from the proper direction. In theory, it would be possible to modify the pinna frequency response of say ceiling reflections to mimic side reflections, but such an equalizer would have to be readjusted for each human being. It is much easier to place the ambient loudspeakers around the listener and feed the appropriate signals to them, as described in later chapters. These pinnae effects also explain why launching, deliberately or inadvertently, recorded rear reverberant hall sounds from the main front loudspeakers, (or proscenium stage ambience from rear speakers) in stereo or 5.1 surround systems, does not and cannot sound realistic.

Although a one-eared music lover, using one pinna, can tell the difference between a live performance and a stereo recording (and Ambiophonics works for such an individual) it is two-eared listeners that Ambiophonics can help the most. Two ears can enhance the listening experience in a concert hall (and life in general) only if there are differences between the sounds reaching each ear, at least most of the time. The only differences the sound at one ear compared to that of the other ear can have are differences in intensity, arrival time, two pinna patterns and absolute polarity. In an acoustical concert hall or any real physical space, it is not possible for absolute polarity to be inverted at just one ear and certainly not at just one ear at all frequencies simultaneously. Thus we need to consider what the difference (or lack of difference) between the ears in sound arrival time and intensity (over the frequency region where the pinna do not function) does for listeners at a concert.

It is clear, since the distance between the ears is relatively small, that at very low frequencies there can be no significant intensity difference, regardless of where a low-bass sound originates. At the other, very high frequency extreme, the head is an effective barrier to sounds coming from the side and, therefore, intensity differences provide the strongest non-pinna related directional cues. At the higher bass frequencies the brain can begin to use arrival time differences to locate a sound. At higher frequencies in the 500 to 1500 Hz region, both time and intensity differences play a role, until as the frequency continues to rise only pinna pattern intensity differences matter. Finally, the sensitivity of the ear to the arrival time of sharp transients is often cited as a hearing parameter but this is just a different way of describing the mechanisms cited above.

There is one more relevant psychoacoustic characteristic of the binaural hearing mechanism which does relate to intensity and arrival time. This is the ability of the ear-brain system to focus on one particular sound source out of many. Most of us can, if we wish, pick out just one voice or instrument in a quartet, or in the classic example, overhear one conversation at a noisy cocktail party. This focusing ability is strong in live three-dimensional concert situations and weak when trying to distinguish one voice in a monophonic recording of Gregorian chant. The relevance to Ambiophonics is that if you can generate a concert-hall stage and sound field real enough to fool the brain, the ability to focus does appear. At a live concert, distractions such as coughing, subway rumble, and program rattling are much less obtrusive because one can focus on the stage and the music. Likewise at home, such distractions as needle scratch, tape hiss, hum, cable idiosyncrasies, amplifier defects, and domestic noises become easier to ignore if you are immersed in Ambiophonic atmosphere. This concentration effect is particularly startling when playing CD transfers of noisy Caruso acoustic-era recordings.

The Ambiophonic Playback System

Ambiophonics was developed to provide audiophiles, record collectors, equipment manufacturers, and, eventually, recording engineers with a clear, understandable recipe for generating realistic music or movie surround sound fields, consistently and repeatedly, either from the vast library of existing two channel recordings or from new multi-channel media made, hopefully, even more realistic by keeping Ambiophonic principles in mind.

The basic home elements required, if the ultimate in realism is desired, are as follows:

1. Loudspeaker crosstalk avoidance. For reasons discussed in a later chapter, the front main left and right loudspeaker sounds must be kept acoustically isolated to their respective ears at the listening position or positions. This may be done using the Ambidipole software discussed in later chapters. The two front speakers are moved to a position almost directly in front of the listeners. This is an advantage over standard 60-degree stereo since the speakers are as easy to locate and as noncritical in this regard as monophonic sound reproduction was before the coming of stereo.
2. Side hall reverberation. Left and right side reverberant signals must be recreated and reproduced through loudspeakers placed roughly to the right and left of the listening area.
3. Rear hall reverberation. Left and right rear hall reverberation signals must similarly emanate from two or more speakers behind or elevated behind the listening position.
4. Speaker correction. While not mandatory, a greater sense of direct sound realism can be achieved if the front loudspeakers are truly identical in frequency response and sensitivity.

The technical reasons for these requirements are discussed here and in the chapters that follow. It is hoped that once the physics and the psychoacoustic laws are understood that the reader may be able to think of better ways to achieve the same end. Ambiophonics was not developed in a day and the reader may not want to implement the entire Ambiophonic system at one time. But each element in the system, when implemented, does result in an appreciable audible improvement.

What Ambiophonics Specifically Achieves

If you employ the techniques described in the chapters below, you will produce a rock-solid sound stage that consistently extends far beyond the right and left positions of the closely spaced front loudspeakers. You will find that even with the main left and right loudspeakers directly in front of you, there is not only no compromise in the perceived stage width or depth, but a substantial improvement over 60 degree stereo or 5.1 surround with virtually any recording or file. You will also see that recreated hall ambience, if launched from the correct direction by well-situated loudspeakers will yield the sense that you are in a hall similar to that in which the recording was made.

Since two-eared listening is more vibrant than one-eared listening, sound fields that differ at each ear in intensity or arrival time are more exciting, and in concert halls add spatial interest to the event. Thus when we come to consider home-concert-hall/home theater design, it is not enough to just maintain the separation of the front left and right channels; it is also necessary to ensure the diversity of all the signals launched into the home listening space. Correlation is the opposite of diversity, and in the next chapter we will consider the significance of the correlation factors of both music and auditoriums so that we can have sound as realistic as possible.

Chapter 2

Concert-Hall Sound Characteristics

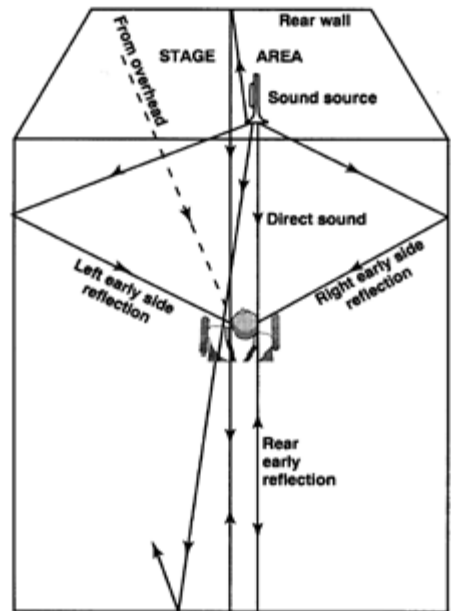
In order to recreate a realistic concert-hall or opera-house sound field at home, it is necessary to know what makes a great music auditorium sound the way it does. Literally hundreds of papers and books have been written on this subject, and while physical concert hall design is now largely based on computer simulation and known acoustic principles, there is still a lot of subjective opinion and art involved. This is also the case in creating a domestic concert hall or a domestic home movie theater. Again the Ambiophonic principles discussed below can be applied to electronic music, games, virtual reality, video, etc.

Concert-hall listeners, not too far back in the auditorium, can usually detect left-to-right angular position of musicians on the stage, can sense depth or the distance they are from the performer, can sense height if say a chorus is elevated on risers, can sense the size of the space they are sitting in, and sense its liveness. Some people can also sense where they are in such a space and what is behind them. When listening to recorded music at home, we want our system to provide us with the same sonic clues that the concert hall provides to its patrons present in the hall during a performance.

In this chapter we explore what makes a hall sound both real and good, so that we can determine which features of a hall we must absolutely duplicate at home in order to fool our ears into thinking that we are in a concert-hall space that is palpably real. We also need to know enough about hall parameters so that we can optimize the ambience controls of our domestic concert just as we do our stereo volume, balance, tone controls, etc.

Concert-hall listeners, not too far back in the auditorium, usually can detect left-to-right angular position of musicians on the stage, can sense depth or the distance they are from the performer, can sense height if say a chorus is elevated on risers, can sense the size of the space they are sitting in, and sense its liveness. Some people can also sense where they are in such a space and what is behind them. When listening to recorded music at home, we want our system to provide us with the same sonic clues that the concert hall provides to its patrons present in the hall during a performance.

In this chapter we explore what makes a hall sound both real and good, so that we can determine which features of a hall we must absolutely duplicate at home in order to fool our ears into thinking that we are in a concert-hall space that is palpably real. We also need to know enough about hall parameters so that we can optimize the ambience controls of our domestic concert just as we do our stereo volume, balance, tone controls, etc.



Direct Sound and Proscenium Reflections

First, for a listener in the audience, there must be an unobstructed path for direct sound to travel from the stage to the listener's ears. This direct sound is then followed by early reflections from the back wall of the stage, the side walls of the stage, the ceiling and, to a lesser extent, the floor of the stage. These first or early reflections come at the listener from roughly the same quadrasphere as the direct sound, i.e., the front 150 degrees or so. Depending on the depth, width, and height of the stage, and its sound reflectivity, these early proscenium reflections arrive from 10 to 300 milliseconds after the direct sound and are fairly strong.

Sound-Signal Correlation

At this point we must introduce the concept of sound correlation. A piece of music, on paper, such as an organ fugue, has a correlation value that represents how the present sound relates to the previously heard sound. The extent of this self or internal structural correlation, called autocorrelation, depends only on the score and the length of time over which correlation is looked for. The intrinsic autocorrelation value of the music, when it is performed, will be then also be modified by the amplitude, delay, angle of incidence and number of reflections experienced. Correlation factors go from 0 to 1 where 1 means the next sound is completely predictable and 0 means there is absolutely no relationship between one note or transient and the next or even no relationship between the beginning of a note and the end of it.

We are also very concerned with the correlation between the sounds reaching the right and left ears. This correlation factor is called Interaural Cross Correlation (IACC). The existence of IACCs less than 1 makes stereophonic and binaural perception possible. Thus, there are autocorrelation factors that describe the signals impinging on a single ear and there are the interaural cross-correlation factors that describe the sound differences between the ears.

An example of simple autocorrelation properties is the round "Row, Row, Row Your Boat as sung by two voices outdoors. If we look at the sound over just the short period of time it takes one voice to sing "Row, Row," and the other voice to sing "Merrily, Merrily," the voices will appear to be entirely uncorrelated. But if we look at the relationship over a period of minutes, we would discover a higher

value of autocorrelation since each voice eventually sings exactly what the other voice has just sung. If one voice is a tenor and one a soprano, this correlation is weakened, and if the tenor sings out of tune, softly, in French, and is indoors in the next room, the correlation factor begins to approach zero. Most people would prefer to hear such a performance with an autocorrelation factor higher than zero but still much less than 1. A "1" would imply that the tenor and soprano were singing precisely the same notes and words at the same time, in the same room milieu, and in the same vocal range.

Autocorrelation and Musical Sounds

Different types of music have different autocorrelation values when looked at through a window of three seconds or longer. For example, an organ playing in a cathedral will have a significantly larger value than a solo guitar playing outdoors. The reason all this is pertinent to concert-hall sound is that the autocorrelation value of music determines the type of ambient field that will make it sound best. Thus a concert hall may be well designed for orchestral music but be a horror for a string quartet. The advantage of the home concert hall is that, unlike the real hall, we can, if we wish, adjust our home hall to suit the autocorrelation value of each musical selection.

Significance of the Hall IACC

While hall reverberation characteristics are the key factor in coping with autocorrelation problems, it is really the interaural cross-correlation value particularly of the early reflected sounds that largely determines the quality of a concert hall and provides the best aural clues to hall presence. In the concert-hall ambience world, the IACC value largely represents what happens in the milliseconds after the arrival of a direct sound sample. Hall design research has shown that the IACC should be kept as small as possible (greatest signal difference between the ears for as long as possible) for the most pleasing concert-hall sound. This should come as no surprise to audiophiles who have always believed in maintaining as much left-right signal separation as possible.

To quote Professor Yoichi Ando, (*Concert Hall Acoustics*, Springer Verlag, 1945), "The IACC depends mainly on the directions from which the early reflections arrive at the listener and on their amplitude. IACC measurements show a minimum at a sound source angle of 55 degrees to the median plane." To translate this, the average person's ears and head are so constructed that a sound coming from 55 degrees to the right of the nose, impinging on the right ear, will not produce a very good replica of itself at the left ear due to time delay, frequency distortion and sound attenuation caused by the ear pinna shape and head obstruction. The IACC value for this condition is typically .36, which is a remarkably good separation for such a situation.

Ando points out that 90 degrees is not better because the almost identical paths around the head (front and back) double the leakage and, therefore, do not decrease the IACC effectively, particularly for frequencies higher than 500Hz.

By contrast, if an early reflection or any sound arrives from straight ahead, the IACC equals one since both ears hear almost exactly the same sound at the same time, and this is desirable for the direct sound from sources directly in front of the listener. That is, the direct frontal sounds should be more correlated than any reflective signals that follow in the first 100 milliseconds or so. As reflections bounce around the hall, the IACC of the reverberant field increases. The rate at which this inter-ear similarity increases determines how good a concert hall sounds when a piece of music with a particular autocorrelation value is being performed. That is why a pipe organ sounds better in a church than in a disco.

The lesson to be learned from all this correlation stuff is that early reflections in the home listening room should have as much left-right signal separation as the recording or ambience processing allows and that many early reflections (but not later reverberant tails) should come from the region around 55 degrees.

More on Early Reflections

Some front proscenium reflections in the concert hall come from above. However, such vertical reflections strike the pinna of both ears from pretty much the same angle with the same amplitude and at the same time. Thus these reflections are highly correlated at the ears and, therefore have little effect in adding to the spatial interest of a concert hall. In our discussions of domestic concert halls, we will, therefore, assume that early reflections from above are of lesser importance or can be safely ignored and indeed, experiments with raising front reflection speakers overhead show this to be optional.

To quote Ando again on early reflections: "The time delay between the first and second early reflections should be 0.8 of the delay between the direct sound and the first reflection." That is, later reflections should be closer together. "If the first reflection is of the same amplitude and frequency response as the direct sound, then the preferred initial time delay is found to be identical to the time delay at which the envelope of the autocorrelation function (coherence of the direct sound) decays to a value of 0.1 of its initial value." Ando found that first reflection delays of from 30 to 130 ms. were preferred, with the exact listener preference proportional to the duration of the autocorrelation function or the average or the average time over which the music is related to itself most strongly. That is, listeners prefer later initial reflections for organ music or a Brahms symphony and earlier ones for a Mozart violin sonata. Such a preference is perhaps intuitively obvious: for most organ music, if the first reflection arrived too soon, it would be ineffective, since the same direct note would probably still be sounding. We will make use of these rules of thumb when it comes time to set the early-reflection parameters for a given recording in our reconstituted concert hall.

We can all agree that different types of music sound best in different types of halls. For instance, symphony orchestras usually sound good in concert halls, string quartets sound better in salons or recital halls, and organs are more at home in churches or cathedrals. While one could use room treatment, panes, etc. to construct a home listening room that could very accurately mimic Carnegie Hall, this room would not be appropriate for a listener whose record collection also includes jazz, opera, madrigals, lieder and solo piano. Any home music theater must be capable of adapting quickly to each type of music being played. Fortunately the convolution technique described in later chapters makes this possible if one knows how halls work so that one can then operate the convolver intelligently.

Reverberation

After the frontal early reflections come the rear, ceiling, and rearward side reflections and reflections of these reflections from the proscenium and all the other hall surfaces. Once these reflections are so close together that the ear or even measuring instruments cannot distinguish them they are called collectively "reverberation" and form a reverberant field. The reverberant field has many parameters that concert hall designers tinker with and that we will be able to season to taste at home. They are the sound level at the onset of the reverberant field, its density, its frequency response and such response changes with time, its angles of incidence, its diffuseness (i.e., its directionality versus intensity), its rate of decay, and its interaural cross correlation. Combinations of these reverberant train parameters allow a listener to perceive the liveness and, to some extent, with the help of the early reflections, the volume of the structure.

The reverberation preferences of concert-goers are again dependent on program material. Chamber music, jazz combos and string symphonies usually sound better with shorter reverberation times. (For the record, the official definition of reverberation time is the time it takes for the sound pressure of a single impulse to fall by 60 dB or to one-millionth of its initial strength.) Large choral works and organ recitals usually benefit from longer reverberation times, with opera stagings somewhere in between. In numerical terms, reverberation times range from over 3 seconds for cathedrals to 1 to 2 seconds for opera houses and concert halls to .5 to 1 second for recital halls or bars. Since the home listener may

perhaps have a wide-ranging music or video collection, we must take care to see that the home concert hall can be quickly optimized for the specific recording being played.

Depth Perception

The ears' ability to detect distance is not as good as that of the eyes'. Depth localization depends on large values of the interaural level difference for nearby sources and for more distant sounds on a hazy feeling for absolute loudness, timbre differences with distance (such as high frequency roll-off), time-of-arrival differences between direct and reflected sound and, if indoors, the ratio of direct to reflected sound. The first four of these factors are easily captured on recordings directly by microphones or can be manipulated by recording engineers, using delay compensation for spot microphones. Since Ambiphonic playback recovers more ILD than the stereo triangle, depth perception is enhanced. The use of surround speakers producing concert-hall ambience also enhances the feeling of depth.

The fourth depth localization factor is sometimes difficult to preserve directly on a disc. If a recording is made outdoors or with microphones that do not pick up many reflections or much hall reverb, then any ambience added later during reproduction will affect all sound source positions equally. For example, increasing the level of the reverberant field makes the listener feel he is further back in the auditorium rather than increasing the distance between the front and rear instruments.

However, as a practical matter, I do not sense any loss of depth perception in my own domestic concert hall. This may be because most recordings are not dry enough to make the effect audible. But more likely, in the average live concert hall, the stage and its shell are so reflective that the direct sound of all instruments, whether located at the front or the back, has about the same ratio of direct-to-reflected sound. This front-to-back stage depth, as opposed to average distance to the stage, particularly for a balcony listener, is not easy to perceive in the typical hall. Also, in some recordings, multiple spot microphones are placed so close to their sound sources that almost no difference in the ratio of direct-to-reflected sound of any instrument is actually recorded. To compensate for this, ambience pickup is then relegated to other remotely placed microphones, so again all instruments recede together. In the home reproduction system, as in the concert hall, it is unlikely that any lack of differential depth perception will actually disturb the illusion of being there.

Chapter 3

Understanding Sound Fields

Human hearing using two ears is called binaural and was developed by evolution. Binaural sound is what most of us listen to all the time. Audiophiles sometimes think of binaural sound as a recording made with a dummy head and played back through earphones. This is a poor imitation of the real thing and is not what we will mean when we refer to the binaural hearing mechanism in this book. Stereophonic sound, by contrast, is a sonic illusion, akin to optical illusions, and simply one of several non-binaural man-made methods of recreating a remote or recorded sound field in a completely different place and time. Stereophonic sound fields are almost always auditioned by binaural listeners whose brains must then reconcile the lack of a binaural field with the presence of a stereophonic one and like optical illusions stereophonic sonic illusions are not always stable and almost never realistic sounding. The commonplace (but misnamed) stereophonic recordings that normally consist of two full-range unencoded, discrete channels, one left and one right are (despite adjustments by recording engineers based on what they hear using studio stereo monitors) not inherently stereophonic and therefore need not suffer the ills that playback via the stereo triangle engenders. That is, the microphones don't know that the sound they pick

up is going to be played back via two widely spaced loudspeakers and thus none of the imperfections of the stereo triangle discussed below apply to the recording before it is played back.

Although we later describe an Ambiophonically optimized recording microphone arrangement, almost any mic setup used to produce two channel recordings works reasonably well when reproduced Ambiophonically. Indeed one of the basic premises of this book and the technology it describes is that the usual two-channel recorded program material contains sufficient information to allow accurate simulation of a binaural concert-hall experience. This is indeed fortunate since it allows the existing library of LPs and CDs to be reproduced with unprecedented realism and shows that multi-channel mic'ing and recording methods, where music is concerned, are actually counter productive according to the tenets of binaural technology. That as few as two channels should be more than adequate can be intuitively understood by simply stating that if we deliver the exact sound required to simulate a live performance at the entrance to each ear canal, then since we only have two ear canals, we should only need to generate two such sound fields. The questions are why existing stereophonic and earphone binaural recording techniques fall short, and what can be done to make up for these shortcomings at least where music reproduction is concerned.

Monophonic Sound

Before the advent of stereo recording we had single-channel or monophonic recordings. Most recordings were made by using one or more microphones and mixing their outputs together before cutting the record, filming the sound track, or making a tape. Such a monophonic recording, if reproduced by two loudspeakers, can be thought of as a special case of stereophonic sound reproduction. It is the case where a sound is the same at both ears and the interaural cross-correlation factor of the sound is 1. In a concert hall, such a signal coming from the stage is sensed as coming from that stage regardless of which direction a concert goer faces. Let us now consider a listener in the balcony of a large hall during a live concert. For this listener, the angle that the stage subtends is very small. Both ears get essentially the same signal, the direct sound from the stage is weak because of distance, and the hall ambience is strong and both are largely the same at each ear. Thus, the players seem to be remote, but still front and center. However, the balcony listener is enveloped in a completely realistic but mostly monophonic reverberant field and therefore hardly notices that his ability to localize left and right sounds is minimal. The lesson we want to draw from this is that mono recordings can be made to sound quite realistic in the home concert hall if you don't mind the impression of sitting further back in the auditorium. The same applies to recordings of solo instruments such as the piano or a singer standing in the curve of the piano.

The reproduction of single central mono or panned sources via two spaced front loudspeakers is also prone to exactly the same crosstalk effects that result from stereophonic reproduction, but, fortunately the solution is the same (see below) for both mono and two channel recordings. To summarize, it is possible to have realism without separation, via a combination of true hall ambience with a corrected front stage and this is one of the main tenets of the Ambiophonic method.

The Stereophonic Illusion

There is a slightly flawed theory, still quoted quite often, that a perfect replica of a given concert-hall sound field can always be produced by putting an infinite number of stage-facing microphones at the front of the stage, all the way up to the ceiling. After being stored on a recorder with an infinite number of channels, this recording can then be played back through an infinite number of point-source loudspeakers, each placed exactly as its corresponding microphone was placed. But the performance replication of such a wall would not be perfect because the loudspeakers would not radiate sound with the same directional characteristics as the sound impinging on the microphone and the final result would also be impacted by the quality of the room into which all these speakers were radiating, but at least the stage would be wide, have depth, and be realistic sounding. As the number of microphones and speakers

is reduced, the quality of the sound field being simulated suffers. By the time we are down to two channels height cues have certainly been lost and instead of a stage that is audible from anywhere in the room we find that sources on the stage are now only localizable if we listen along a line equidistant from the last two remaining speakers and face them. While there are many two channel speaker arrangements possible, the most popular two-channel reproduction method is the stereophonic technique of reproducing two-channel recordings through two loudspeakers with the listener and the two speakers forming an equilateral or wider isosceles triangle.

Stereo takes advantage of one rather unnatural psychoacoustic illusion, which is that as a recorded sound source moves on the stage from the left to the right, and as the playback signal likewise shifts from the left speaker to the right speaker, most listeners hear a virtual sound or phantom sound image move from one speaker position to the other. Compared to real life hearing, this phantom audible illusion does not move linearly and there is a tendency for the sound to jump to the speaker location as the sound moves to the side. If identical sounds come from each speaker, (the monophonic case above) then most central listeners hear a phantom sound that hangs in the air at the halfway point on the line between the loudspeakers. Just as there are some individuals who cannot see optical illusions, so there are a few individuals who cannot hear phantom images. Just as optical illusions are just that-illusions that no sighted person would confuse with a true three-dimensional object, so phantom stereo illusions could never be confused with a truly binaural sound field.

Nevertheless, for some 70 years this illusion of frontal separation and space is so pleasing to most listeners that stereophonic reproduction has remained the standard music reproduction technique ever since Alan Dower Blumlein applied for his patent at the end of 1931. (See Appendix) The illusion created by stereo reproduction techniques is far from perfect, even if the highest grade of audiophile caliber reproducing and recording equipment is used. The first problem is that the image of the stage width is confined to the arc that the listener sees looking from one speaker to the other. Occasionally, an out-of-phase sound from the opposite loudspeaker, an accidental room reflection, or a recording site anomaly will make an instrument appear to come from beyond the speaker position. These images, however, are almost always ephemeral and often not reproducible. Thus, in non-Ambiophonic systems, in order to get a useful stage width with stable left-right localization, the loudspeakers must be placed at a wide enough angle to mimic the angular proportions of a concert hall or theater stage but not so wide that the phantom center image illusion collapses.

With most stereo systems, there is a "sweet spot" at the point of the triangle where the listening is best. This, unfortunately, is what we are faced with when only two front channels (or three for that matter) are available. The "sweet spot" is also a characteristic of Ambiophonic reproduction although the spot is somewhat larger and less critical in the case of Ambiophonics. It is difficult enough to recreate concert-hall sounds from two discrete recorded channels (and even harder using multi channels) for one or two listeners in the home, without trying to do it for a whole room full of people.

Basically several listeners can listen to Ambiophonics at the same time but they have to be one behind the other. Consider the following. In stereo if you move too close to the speakers you get a hole in the middle. If you move back you get mono. If you move to the side you mostly hear just one channel. In general, out-of-the-sweet-spot stereo is tolerated by most everyone since it is clearly not truly realistic when you are at the sweet spot you don't feel you are missing much when you are off center. With Ambiophonics if you move too close to the speakers you get stereo, if you move back you still get the same wide stage until you bump into the rear wall. If you move offside, you get normal mono sound since both channels are present in both speakers and this is good for movie dialog even if there is no center speaker. With Ambiophonics, you can recline, nod, lean, and rotate your head, stand, etc. There are similar advantages for Ambiophonics versus 5.1, one prime advantage being that no center speaker is ever needed.

If you use the two optional rear speakers then offside 5.1 listeners will distinctly hear both the rear and front stages in their proper locations. For most movies this works better in the home than in the movie theater. Most 5.1 systems cannot really reproduce a rear stage of direct sound effects, but Ambiphonics does this even for offside listeners. In 5.1 if you are off-center and back a bit you will likely just localize to one of the rear surround speakers. If you have four speaker (two Ambiodipoles) then, if you like, one can listen facing front and one can listen facing the rear sitting back to back if you are playing two channel media.

Stereophonic Crosstalk

By far the major defect of stereophonic reproduction is caused by the presence of crosstalk at the listener's ears generated by the loudspeakers. Again, the crosstalk is an artifact of stereophonic reproduction and is not present in the recording. We will show that eliminating this crosstalk widens the stereo soundstage way beyond the narrow position of the loudspeakers, eliminates spurious frequency response peaks and dips (comb filter effects), and allows the speakers to be moved much closer together eliminating the need for phantom imaging or a center channel.

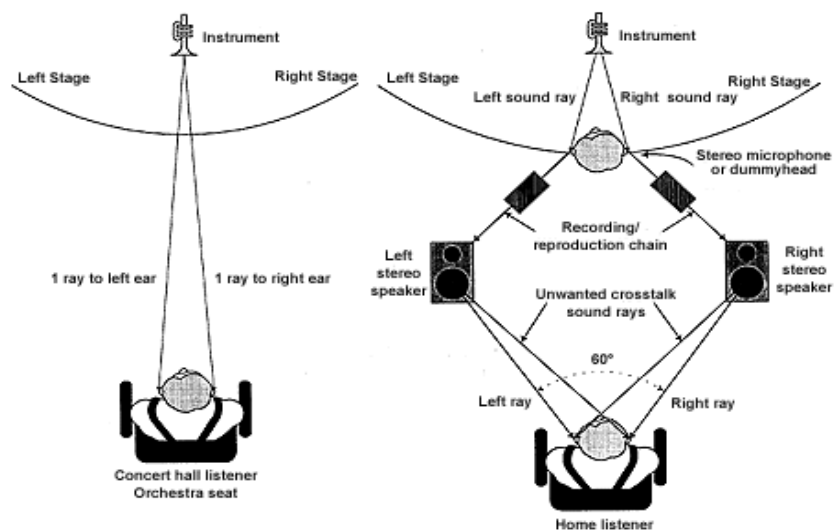


Figure 1: Comparison of live concert hall listening geometry with home stereophonic listening practice showing the additional unwanted crosstalk sound rays impinging on the pinna from too large an angle, thereby, causing unrealistic playback artifacts.

In a concert hall, direct sound rays from a centrally located instrument reach each ear simultaneously: one ray per ear (see Figure 1). By contrast, for a centrally located recorded sound source, reproduced in stereo, identical rays come from the right and left speakers to the right and left ears, but a second pair of uninvited, only slightly attenuated, longer, right and left speaker rays also passes around the nose to the left and right ears (see Figure 1). The problem is that these unwanted rays, which cross in front of the eyes and diffract around the back and top of the head, are delayed by the extra distance they travel across the head. At its greatest, this distance is just under 7 inches. For a middling distance of say 3 1/2 inches, it takes sound one-quarter of a millisecond to do this. A quarter of a millisecond is half the period and, therefore, half the wavelength of a 2000 Hz tone.

When two signals, one direct and one a half-wavelength delayed, but of similar amplitude, meet at the ear, cancellation will occur. At 4000 Hz the delay is one full wavelength and the sounds will add. Thus at frequencies from the octave above middle C and up, all sounds add or subtract at the ears to a greater or lesser degree, depending on the recorded sound source position, the angle to the speakers, the listener's

head position, nose size and shape, head size, differing path lengths around the head, and other geometrical considerations. Note that if the sound source at the recording studio or the listener at home moves a few feet or inches to the left or right, a whole new pattern of additions and subtractions at different frequencies will assault the listener. This interference phenomena is called comb filtering, and largely explains why many critical listeners are so sensitive to small adjustments in stereo listening or speaker position, and to relatively minute playback system electrical and acoustical delay or attenuation. Bock and Keele measured comb filter nulls as deep as 15 dB for the 60-degree stereo loudspeaker setup. Note that for extreme side images the comb-filter effect is minimal.

Thus the acoustical frequency response of a normal stereo setup actually depends on the angular position of the original instrument or singer. As indicated above, it is fascinating that these frequency response anomalies are not clearly audible as changes in tone but rather manifest themselves as imprecisions in imaging and a sense that the music is canned. But it is possible to hear the change in timbre caused by comb filtering. Simply play pink noise from a test CD over your stereo system and rotate the balance control from hard left to hard right. As the image of the noise passes thru the center one can clearly hear a drop in the treble loudness of the noise and a distinct change in its character. Alternatively one can walk normally from the left to right and hear the change in the noise as one passes through the center area.

Note that a phase shift change between channels of only a few degrees can shift a stereo crosstalk comb filter null by hundreds of Hertz. Even a small, one-degree phase shift change between the left and right channels at 2000 Hz will cause a shift of 71 Hertz in the position of a crosstalk null or peak. Crosstalk comb-filter patterns are thus a function of any asymmetry in amplifier output impedances or delays, differential delays in cables, or differential speaker time delay by virtue of their positions relative to the listening position or their impedance networks. For instance, a vacuum-tube driven left midrange speaker can interact with an overlapping right tweeter to produce interaural crosstalk peaks and nulls that are otherwise not present in the solid-state amplifier case. Such patterns may be audible to some individuals. Any changes in the interaural crosstalk pattern are interpreted by the brain as a spatial artifact such as more or less depth, air, or hollowness. Of course, any change in listener position, or speaker location causes similar shifts in the crosstalk peaks and nulls and further complicates equipment comparisons by ear in stereo or surround sound. The irregular directional and largely unpredictable frequency response of the standard stereophonic 60 to 90 degree listening arrangement would never be accepted in an amplifier, a speaker, or a cable. Why such a basic listening system defect continues to be so universally tolerated and studiously ignored is difficult to fathom.

The binaural perception of directional cues depends on both the relative loudness of sound and the relative time of arrival of sound at each ear. Which mechanism predominates depends on the frequency. Unfortunately, since these delay and stereophonic comb-filter artifacts have an effect extending from below 800 Hz on up, they very seriously impact on both mechanisms and thus impair the ability of the listener to detect angular position with lifelike ease. It is also these crossing rays that limit stereo and surround sound imaging to the line between the two front speakers. (See below) If we are to achieve anything close to concert-hall realism, we must eliminate these crosstalk effects and provide a directionally correct single ray to each ear. But first we will need to present evidence of the extraordinary sensitivity of the ear pinna to such comb filter patterns.

Imaging Beyond the Speaker Positions

A major problem with stereophonic crosstalk is that it limits the apparent stage width. For sound sources that originate, say, at 90 degrees far to the right of the right microphone, we can temporarily ignore the left channel microphone pickup. Then in the stereophonic listening setup, the right speaker will send unobstructed sound to the right ear and a somewhat modified version of the same sound to the left ear. The ear-brain naturally localizes this everyday sound situation to the speaker position itself instead of to

the 90 degrees the data is indicating. Thus, no matter how low the left channel volume is, the recorded image can never extend beyond the right speaker in standard stereo reproduction (see Figure 2).

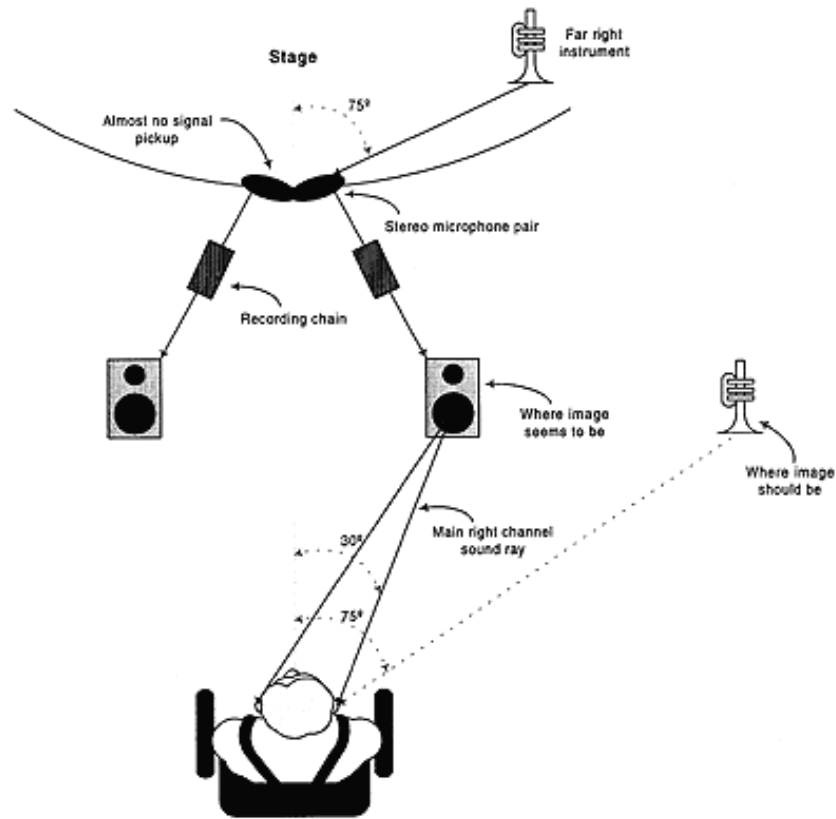


Figure 2: Stereophonic crosstalk limits the sound stage and alters localization of original sound image.

If, however, the right speaker sound ray crossing over to reach the left ear could be blocked or attenuated, then at least the low and mid frequency sound could be localized to the extreme right, well beyond the speaker position and just where the recording microphones said the source was located. (High frequency localization is discussed in the next chapter.) Remember, the microphones don't know that the playback will be in stereo with crosstalk and therefore it is not the recording setup that limits stage width. Clearly, eliminating the extra sound ray results in wide spectacular imaging even from existing two channel media. In Chapter five we will discuss two methods of eliminating crosstalk as well as doing away with the stereo triangle altogether.

In the case of live hearing, a sound at the extreme side produces an interaural time difference (ITD) at the ears of about 700-microseconds. If such a recording is played via the stereo triangle the maximum ITD that can be produced is about 220-microseconds. This is because sound coming from a speaker at the 30-degree angle does not have to pass across the entire head and so is not delayed the full 700-microseconds. The shorter ITD is interpreted by the brain as a shallower angle, 30 degrees, for this source.

Similarly, if the microphones detect a channel level difference (ILD) of say 10 dB indicating a source far to the side, then, when reproduced by stereo loudspeakers, this difference will be reduced by half or so since the louder speaker also can see the far ear and thus increase the sound level at that ear reducing the ILD so much that the sound that the mics heard at the far side is now no more than 30 degrees off center.

Loudspeaker Out-of-Phase Effects

In stereo systems it is necessary for the right and left main speakers to be in phase or better expressed be of the same polarity. Phase in this case means that if identical electrical signals are applied to each speaker, the speakers will both generate a rarefaction, or both generate a compression in response to a simultaneous input pulse. When a monophonic recording is played through a pair of out-of-phase loudspeakers, the sound at the ears lacks bass, the phantom center image is not present, and a hazy, undefined sound field seems to extend far beyond the speakers to the extreme sides and sometimes even rearward. Similar effects, only slightly less pronounced, are also present using two channel sources. These subjective effects can be better comprehended now that we understand all about stereo crosstalk. It is clear that equal but out-of-phase very low frequency signals, with wavelengths much longer than the width of the head will always arrive unattenuated and 180 degrees out-of phase at either ear and therefore will always largely cancel. This factor accounts for the thinness of the mono or central (L+R) stereo sound. At somewhat higher frequencies the cancellation is not total. The left ear hears pure left signal from the left speaker that is reduced only somewhat by the now slightly delayed and thus only partially out-of-phase crosstalk from the right speaker. Similarly, at that same instant the right ear is hearing a reduced but pure right-speaker sound that is similar in amplitude but not identical to the pure left-ear sound because the resultant sounds are still out-of-phase. We know that a midrange frequency sound heard only in the right ear seems to come from the extreme right and a sound heard only in the left ear seems to come from the extreme left. This phenomenon is still operative even if the two sounds that come from the sides are identical in amplitude and timbre. Thus, one can easily hear two identical bells as separate left and right sound sources. If, however, we exchange the bells for pink noise, then we can hear the noise only as separate sources when they are not precisely in step (uncorrelated). Since our signals are out of phase they are not identical in time or highly auto correlated and therefore audible as separate entities. Thus, the inadvertent crosstalk elimination caused by out-of-phase speakers that occurs at mid frequencies widens the perceived sound field. As the frequency increases, instead of simple canceling, the comb-filtering effect predominates and the position of the images becomes frequency, and therefore program, dependent, changing so rapidly that no listener can sort out this hodgepodge of constantly shifting side images. Most listeners describe this effect as diffuse, unfocussed or phasy.

In general, mechanical or software crosstalk elimination is not fully effective or needed at very low bass frequencies and so the bass out-of-phase thinness effect, while much reduced, remains if speakers are out-of-phased. In Ambiophonics, the audibility of the out-of-phase effect is much reduced. The stage image still extends from the speakers outward when the recording calls for this. That is, sound sources at the extreme right and left image just as they do when the speakers are in-phase. This makes sense, since we are, listening to one sound source with one ear. To repeat. In the out-of-phase case, for most of the frequency range, each ear is hearing a signal that is distinctive because the signals are of opposite polarity and, therefore the ear localizes each sound as originating from beyond their respective speakers. A phantom center image does not form and the infamous hole-in-the-middle appears. In the out-of-phase Ambiophonic case the speakers are very close together. Therefore, the middle hole is almost nonexistent and the bottom line is that, except for extreme bass response, front speaker phasing or other timing anomalies are more critical in stereo than in Ambiophonics.

Absolute Polarity

When an instrument produces a sound, the sound consists of a series of alternating rarefactions and compressions of air. The sonic signatures of such acoustic musical instruments are determined by the pressure and spacing of these rarefactions and compressions. Electronic recording and reproduction have now made it possible to turn rarefactions into compressions and vice-versa. The significance of this to the problem of establishing a home concert hall is not entirely clear. But a few people seem to be able to hear a difference between correct and incorrect polarity. Therefore, care should be taken that all amplifiers, speakers and ambience sources, taken together, do not invert. Since acoustic reflectors in

concert halls do not invert polarity, the key early reflections, at least, should not be inverted accidentally in home reproduction either and should be delivered to the ears with the same polarity as the direct sound which is, one hopes, also of the correct absolute polarity. If you cannot tell one polarity from the other in your own system, don't despair. For a few people, polarity is only audible when special test signals are used. One possible reason for difficulty in this regard is the nature of many instruments. A listener to the left of a violinist hears one polarity, while a listener to the right hears the other polarity, assuming the string is vibrating in the same plane as the ears of both listeners. But no matter where you stand around a trumpet you get the same polarity. The inverted polarity sound in this case is inside the trumpet. Indeed it has been reported that test subjects are more likely to hear polarity differences where wind instruments are involved.

On balance, one would have to say that it does not pay to agonize over the absolute polarity effect unless you are certain that you or your friends are sensitive to it.

Chapter 4

Pinna Power

The fluted, rather grotesque, protuberances that extend out from each ear canal are called pinna. The importance of satisfying one's pinnae by reproducing sound fields that complement their complex nature cannot be exaggerated. Like fingerprints, no two individuals have exactly identical ear pinna. Thought to be vestigial, even as late as the mid 20th century, the intricacy which characterizes these structures would suggest that their function must not only be very important to the hearing mechanism but also that their working must be of a very complex, personal and sensitive nature. For audiophiles in search of more realistic sound reproduction, an understanding of how the pinna head interact with stereophonic, Ambiophonic. or surround-sound fields is of importance since at the present time a major mismatch exists. Repairing the discrepancy between what the present recording and playback methods deliver and what the human ear pinna expect and require is the last major psychoacoustic barrier to be overcome, both in hi-fi music reproduction and in the frenetic PC/games/multi-media field.

We wish to duplicate, remotely, the normal biological binaural listening experience a listener would have had at a specific location in that original space. As live or rock electronic music enthusiasts, we are first concerned with the recreation of horizontal-staged-acoustic/electronic, usually musical, events recorded in enclosed spaces such as concert halls, opera houses, pop venues, etc., and where the best listening position is centered, fixed, and usually close to the stage. I have called this two-channel subset of the broader 360-degree movie requirement Ambiophonics because it is both related to and a suitable replacement for stereophonics. Another way of stating a major goal of Ambiophonics and describing a still, unsolved problem of virtual reality or surround auralization is the externalization of the binaural earphone effect. In brief, this means duplicating the full, everyday binaural hearing experience, either via earphones, without having the sound field appear to be within one's head, or via loudspeakers, without losing either binaural's directional clarity or the "cocktail party" effect whereby one can focus on a particular conversation despite noise or other voices. So far this goal has eluded those researchers trying to externalize the binaural effect over a full sphere or circle, but it can be done using Ambiophonic methods for the front half of the horizontal plane and using Panambiophonics for the full circle in the horizontal plane.

Pinnae as Direction Finders

It is intuitively obvious, as mathematicians are fond of observing, that duplicating the binaural effect at home, simply involves presenting at the entrance of the home ear canal an exact replica of what the same ear canal would have been presented with at the live music event. But to get to the entrance of the ear canal, almost all sound higher in frequency than about 1.0 kHz must first interact with the surface of a pinna. Each pinna of your ear is in essence your own personal high frequency direction finder. The pinna of my ear produces a quite different (and undoubtedly superior) series of nulls and peaks than does yours. The sound that finally makes it to the entrance of the ear canal, in the kilohertz region, is subject to severe attenuation or boost, depending on the angle from which the sound originates as well as on its exact frequency. Additionally, sounds that come from the remote side of the head are subject to additional delay and filtering by the head and this likewise very individual head plus pinna characteristic is called the Head-Related Transfer Function or HRTF. In this book I will try to distinguish between the functions of one pinna alone, both pinna working together, the HRTF without any pinna effects, and finally the whole enchilada which is understood to include the shadowing, reflection, and diffraction due to the head, and all the resonances and delays in the pinna cavities, particularly the large bowl known as the concha.

The effects of the head and torso become measurable starting at frequencies around 500 Hz with the pinna becoming extremely active over 1500 Hz. Because the many peaks and nulls of the HRTF are very close together and sometimes very narrow it is exceedingly difficult to make measurements using human subjects, and not every bit of fine structure can be captured, particularly at the higher frequencies where the interference pattern is very hard to resolve.

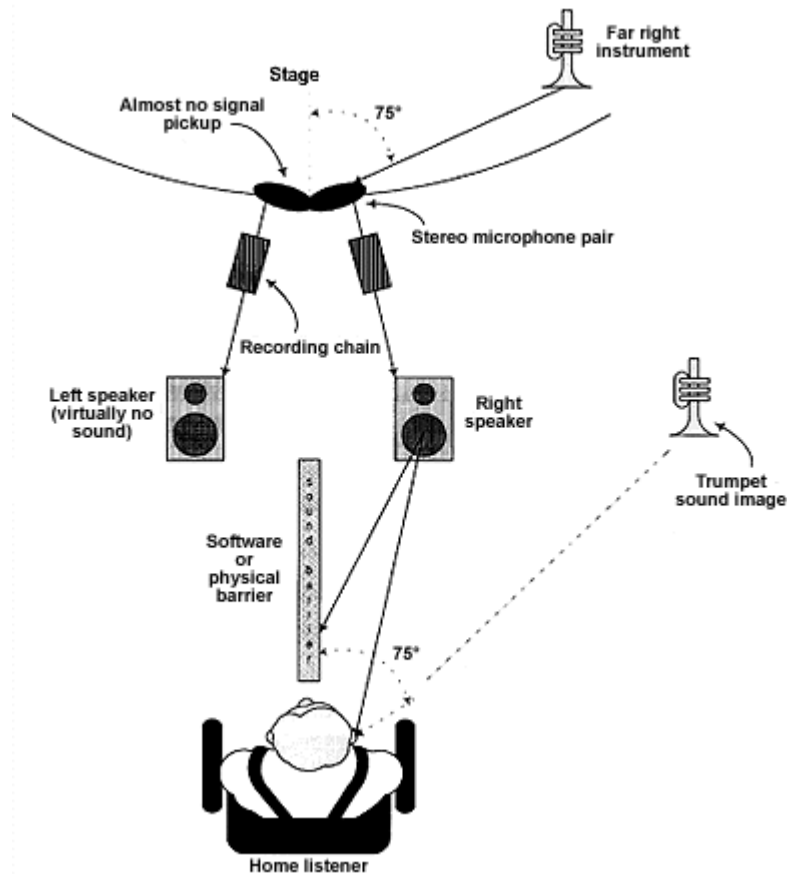


Figure 3: Image in Ambiphonic system matches recording perspective because a signal reaching just one ear sounds as though it is coming from the side.

Figure 4.1 shows a series of measurements recorded by Ronald Aarts made using a small microphone placed right at the entrance to the ear canals for several subjects. As the sound source moves about the head both the variety and the complexity of the response is plainly evident. One can also see the obvious variation between different auditors. Note that when the sound source is at the far side of the head the curves include the head shadowing frequency response. Because the peaks or nulls are so narrow and also because a null at one ear is likely to be something else at the other ear, we do not hear these dips as changes in timbre or a loss or boost of treble response, but, as we shall see, the brain relies on these otherwise inaudible serrations to determine angular position with phenomenal accuracy.

Much research has been devoted to trying to find an average pinna response curve and an average shadow HRTF that could be used to generate virtual reality sound fields for military and commercial use in computer simulations, games, etc. So far no average pinna-HRTF emulation program has been found that satisfies more than a minority of listeners and none of these efforts is up to audiophile standards. Remember that a solution to this problem must take into account the fact that each of us has a different pattern of sound transference around, over and under the head, as well as differing pinna.

The moral of all this is that if you are interested in exciting, realistic sound reproduction of concert hall music, it does not pay to try to fool your pinna. If a sound source on a stage is in the center, then when that sound is recorded and reproduced at home it had better come from speakers that are reasonably straight ahead and not from nearby walls, surround or Ambisonic speakers. The traditional equilateral stereophonic listening triangle is quite deficient in this regard. It causes ear-brain image processing confusion for central sound sources because although both ears get the same full range signal telling the brain that the source is directly ahead, the pinnae are simultaneously reporting that there are higher frequency sound sources at 30-degrees to the left and at 30-degrees to the right. All listeners will hear a center image under these conditions, which is why stereophonic reproduction has lasted 70 years so far, but almost no one would confuse this center image with the real thing. Unfortunately, a recorded discrete center channel and speaker as in 5.1 home theater surround sound is of little help in this regard. We will see later that such a solution has its own problems and is an unnecessary expense that also does nothing for the existing unencoded two-channel recorded library.

Testing Your Single Pinna Power

A very simple experiment demonstrates the ability of a single pinna to sense direction in the front horizontal plane at higher frequencies. Set up a metronome or have someone tap a glass, run water, or shake a rattle about ten feet directly in front of you. Close your eyes and locate the sound source using both ears. Now, keeping your eyes closed, block one ear as completely as possible and estimate how far the apparent position of the sound has moved in the direction of the still-open ear. Most audio practitioners would expect that a sound that is only heard in the right ear would seem to come from the extreme right, but you will find that in this experiment the shift is seldom more than 5 degrees, and if you have great pinnae the source may not move at all. A variation of this experiment is to spin around with your eyes closed and then see how close you come to locating the sound source. The single pinna directional detecting system is stronger than the interaural intensity effect for things such as clicks and explains why one-eared individuals can still detect sound source positions.

Another moral of this experiment is that for most people, over the higher audible frequency range, which includes most musical transients and harmonics, the one-eared pinna/head directional sense is easily a match for the interaural or two-eared-intensity-time difference localization mechanism. Therefore, all

recorded music signals, including direct sound, early reflections, and reverberation had better come from directions that please the pinnae, if you want your brain to accept the listening experience as real.

If you now switch to a fuller range music source, such as a small radio, and repeat the experiment above you will likely hear a greater image shift, since the external ear and head are less important to sound localization as the sound gets down to 400 Hz or so. Even the best stereo systems that seemingly have great localization based on lower frequency interaural time and intensity cues, still sound naggingly unrealistic because of the conflict between the interaural and the intraaural localization mechanisms inherent in the old fashioned stereo triangle.

The Department of the Interior

Eliminate the outer ears, and all the sound will appear to originate inside your head. Do you doubt this? Then open your mouth and hum or sing with your mouth open. You will hear this sound coming from the lip area. Now put both hands or fingers within your ears and the sound will jump up into the middle of your skull. Every child has tried this at one time except maybe you. What the effect illustrates is that in the complete absence of pinna the brain makes the only perfectly logical decision it can based on the sonic facts. That is, that the sound must originate from a point on the brain side of the eardrum, for how otherwise could the sound have avoided being modified by the pinna.

Now while listening to running water or other transient rich sound, bring the flat palms of your hands to within a half-inch of both your ears. You will hear the character of the sound change, usually in a manner that makes the sound seem closer to you. The presence of the additional mass and enclosed air trapped between your palm and ear interferes with the resonances in the cavities of the pinna and changes what you think you hear.

These effects, are why it is so difficult to get a natural externalized sound image using earphones. In-the-ear-canal phones, while quite realistic compared to stereo, are especially prone to producing very pronounced internalization. Again, it does not pay to fool pinna nature and that is why the Ambiophonic method limits itself to using loudspeakers.

I Am Not Alone

Martin D. Wilde, in his paper, "Temporal Localization Cues and Their Role in Auditory Perception" AES Preprint 3798, Oct., 1993 states:

"There has been much discussion in the literature whether human localization ability is primarily a monaural or binaural phenomena. But interaural differences cannot explain such things as effective monaural localization. However, the recognition and selection of unique monaural pinna delay encodings can account for such observed behavior. This is not to say that localization is solely a monaural phenomenon. It is probably more the case that the brain identifies and makes estimates of a sound's location for each ear's input alone and then combines the monaural results with some higher-order binaural processor."

Again, any reproduction system that does not take into account the sensitivity of the pinna to the direction of music incidence will not sound natural or realistic. Two-eared localization is not superior to one-eared localization, they must both agree at all frequencies for realistic concert hall music reproduction.

Pinna and Phantom Images at the Sides

A phantom front center image can be generated by feeding identical in-phase signals to speakers at the front left and front right of a forward facing listener. Despite the inferiority of the phantom illusion, the surround sound crowd would be ecstatic if they could pan as good a phantom image, to the side, in a similar way, by feeding in-phase signals just to a right front and a right rear speaker pair. Unfortunately, phantom images cannot be panned this way between side speakers. The reason realistic phantom side images are difficult to generate is that we are largely dealing with a one-eared hearing situation. Let us assume that for a right side sound only negligible sound is reaching the remote left ear. We already know that the only directional sensing mechanism a one-eared person has for higher frequency sound is the pinna convolution mechanism. Thus if a sound comes from a speaker at 45 degrees to the front, the pinna will locate it there. If, at the same time, a similar sound is coming from 45 degrees to the rear, one either hears two discrete sound sources or one speaker predominates and the image hops backward and forward between them. Of course, some sound does leak around the head to the other ear and depending on room reflections, this affects every individual differently and unpredictably. One can also use HRTF processing to position side virtual images but such methods usually do not sound realistic where music, dialog, or sound effects are concerned and such methods cannot help the existing library since some form of encoding is usually required to get any result at all.

Apparent Front Stage Width

The sensitivity of the ears to the direction from which a sound originates, mandates that to achieve realistic Ambiphonic reproduction, all signals in the listening room must originate from directions that will not confuse the ear-brain system. Thus if a concert hall has strong early reflections from 55 degrees (as the best halls should) then the home reproduction system should similarly launch such reflections from approximately this direction. In the same vein, much stage sound, particularly that of soloists, originates in the center twenty degrees or so more often than at the extremes. Thus it makes more sense to move the front-channel speakers to where the angle to the listening position is on the order of 20-degrees instead of the usual 60. This eliminates most of the pinna angular position distortion.

One might suppose that, if a main speaker is in front, that sounds that are meant to image to the extreme sides will suffer from pinna angle distortion and that we will just have traded the central pinna angle error of the stereo triangle for the side pinna angle error of Ambiphonics. But if you look at the curves of Figure 4.1 you will see that at the wider angles beyond say 60 degrees a sound coming from the side has a clear shot at the entrance to the ear canal and thus the pinna curve is relatively flat and therefore minimal. In practice Ambiphonics easily produces easy to listen to images out to 85-degrees either side of center.

It should also be remembered that, in an Ambiphonic sound field, a seemingly narrower stage is simply equivalent to moving back a few rows in the auditorium and so has not proven to be noticeable. In the same vein, the sensitivity of the pinnae to the directions from which any sound comes dictates that reconstructed or recorded early reflections or reverberant tails attributed to the sides or rear of a concert hall should not come to the home ears from the main front speakers.

Pinna Considerations in Binaural or Stereo Recording

The pinna must be taken into account when recordings are made, particularly recordings made with dummy heads. For example, if a dummy-head microphone has molded ear pinnae then such a recording will only sound exceptionally realistic if played back through earphones that fit inside the ear canal. Even then, since each listener's pinnae are different from the ones on the microphone, most listeners will not experience an optimum binaural effect. On the other hand, if the dummy head does not have pinnae, then the recording should either be played back Ambiphonically, using loudspeakers, or through earphones

that stand out in front of the ears far enough to excite the normal pinna effect. (As in the IMAX system, loudspeakers can then be used to provide the lost bass.)

But one must also take into account the head-related effects as well. Thus if one uses a dummy head microphone without pinnae, then listening with stereo spaced loudspeakers would produce side image distortion, due to the doubled shadow induced ITD and ILD due to transmission around, over and under both the microphone head and the listener's head.

The Rule Is:

In any recording/reproduction chain there should be only one set of Pinnae and it better be yours and only one but at least one head shadow which need not necessarily be yours.

Normal two channel recordings, LP or CD or DVD are not inherently stereo. No recording engineer takes into account the crosstalk and the pinna response errors in reproduction when microphones are selected and spaced. Panning equations used to shift sonic images, likewise, seldom consider the full extent of HRTF effects. This is fortunate since the existing library of recordings is thus not obsoleted in the slightest where Ambiophonic reproduction and the pinna are concerned.

Pinna Foolery or Feet of Klayman

Arnold Klayman (SRS, NuReality) (and many other companies) has gamely tackled the essentially intractable problem of manipulating parts of a stereo signal to suit the angular sensitivity of the pinna, while still restricting himself to just two loudspeakers. To do this, he first attempts to extract those ambient signals in the recording that should reasonably be coming to the listening position from the side or rear sides. There is really no hi-fi way to do this, but let us assume, for argument's sake, that the difference signal (l-r) is good enough for this purpose, particularly after some Klayman equalization, delay and level manipulation. This extracted ambient information, usually mostly mono by now, must then be passed through a filter circuit that represents the side pinna response for an average ear. Since this pinna-corrected ambience signal is to be launched from the main front speakers, along with the direct sound, these modified ambience signals are further corrected by subtracting the front pinna response from them. The fact that all this legerdemain produces an effect that many listeners find pleasing is an indication that the pinnae have been seriously impoverished by Blumlein stereo for far too long, and is a tribute to Klayman's extraordinary perseverance and ingenuity.

While Klayman's and other similar boxes cost relatively little and are definitely better than doing nothing at all about pinna distortion, any method that relies on average pinna response or, like matrixed forms of surround sound, or attempts to separate early reflections, reverberant fields or extreme side signals from standard or matrixed stereo recordings of music is doomed to only minor success. The Klayman approach must also consider that an average HRTF is also required and should be used when launching side images from the front speakers. Someday we will all be able to get our own personal pinna and HRTF responses measured and stored on CD-ROM for use in Klayman type-synthesizers, but until then, the bottom line, for audiophiles, is that the only way to minimize pinna and head-induced image distortion is to give the pinnae what they are listening for. This means launching all signals as much as is feasible from the directions nature intended and requires that pure ambient signals such as early reflections and hall reverberation (uncontaminated with direct sound) come from additional speakers, appropriately located. It implies that recorded ambient signals, inadvertently coming from the front channels, are not so strong that the rear hall reverb coming strongly from up front causes subconscious confusion. (Most CDs and LPs are fine in this regard but would be improved by a more Ambiophonic recording style.)

Two-Eared Pinnae Effects

So far we have been considering single ear and head response effects. Now we want to examine the even more dramatic contribution of both pinnae and the head, jointly, to the interaural hearing mechanism that gives us such an accurate ability to sense horizontal angular position. William B. Snow, a one-time Bell Telephone Labs researcher, in 1953, and James Moir of CBS in Audio Magazine, in 1952, reported that for impulsive clicks or speech and, by extension, music, differences in horizontal angular position as small as one degree could be perceived. For a source only one degree off dead ahead we are talking about an arrival-time difference between the ears of only about ten microseconds and an intensity difference just before reaching the ears so small as not to merit serious consideration. Moir went even further and showed that with the sound source indoors (even at a distance of 55 feet!), and using sounds limited to the frequency band over 3000 Hz, that the angular localization got even better, approaching half a degree. It appears that when it comes to the localization of sounds like music, the ear is only slightly less sensitive than the eyes in the front horizontal plane.

It is not a coincidence that the ear is most accurate in sensing position in the high treble range, for this is the same region where we find the extreme gyrations in peaks and nulls due to pinna shape and head diffraction. This is also the frequency region where interaural intensity differences have long been claimed to govern binaural perception. However, it is not the simple amplitude difference in sound arriving at the outer ears that matters, but the difference in the sound at the entrance to the ear canal after pinna convolution.

Going even further, at frequencies in excess of 2000 Hz it is not the average intensity that matters but the differences in the pattern of nulls and peaks between the ears that allow the two-eared person to locate sounds better than the one-eared individual. Remember that at these higher audible frequencies, direct sounds bouncing off the various surfaces of the pinna add and subtract at the entrance to the ear canal. This random and almost unplotable concatenation of hills and deep valleys is further complicated by later but identical sound that arrives from hall wall reflections or from over, under, the front of, or the back of the head. This pattern of peaks and nulls is radically different at each ear canal and thus the difference signal between the ears is a very leveraged function of both frequency and source position. In their action a pair of pinnae are exquisitely sensitive mechanical amplifiers that convert small changes in incident sound angles to dramatic changes in the fixed unique, picket fence, patterns that each individual's brain has learned to associate with a particular direction.

Another way of describing this process is to say that the pinna converts small differences in the angle of sound incidence into large changes in the shape of complex waveforms by inducing large shifts in the amplitude and even the polarity of the sinewave components of such waveforms. (Martin D. Wilde, see above, also posits that the pinna generate differential delays or what amount to micro reflections or echoes of the sound reaching the ear and that the brain is also adept at recognizing these echo patterns and using them to determine position. Since such temporal artifacts would be on the order of a few microseconds it seems unlikely that the brain actually makes use of this time delay data.)

Angular Perception at Higher Frequencies

To put the astonishing sensitivity of the ear in perspective, a movement of one degree in the vicinity of the median plane (the vertical plane bisecting the nose) corresponds to a differential change in arrival time at the ears of only 8 microseconds. Eight microseconds can be compared to a frequency of 120,000Hz or a phase shift of 15 degrees at 5kHz. I think we can all agree that the ear-brain system could not possibly be responding to such differences directly. But when we are dealing with music that is rich in high-frequency components, a shift of only a few microseconds can cause a radical shift in the frequency location, depths, and heights of the myriad peaks and nulls generated by the pinnae in conjunction with the HRTF. To repeat, it is clear that very large amplitude changes extending over a wide band of

frequencies at each ear and between the ears can and do occur for small source or head movements. It is these gross changes in the fine structure of the interference pattern that allow the ear to be so sensitive to source position.

Thus, just considering frequencies below 10kHz, at least one null of 30db is possible for most people at even shallow source angles, for the ear facing the sound source. Peaks of as much as 10db are also common. The response of the ear on the far side of the head is more irregular since it depends on head, nose and torso shapes as well as pinna convolution. One can easily see that a relatively minute shift in the position of a sound source could cause a null at one ear to become a peak while at the same time a peak at the other ear becomes a null resulting in an interaural intensity shift of 40db! When we deal with broadband sounds such as musical transients, tens of peaks may become nulls at each ear and vice versa, resulting in a radical change in the response pattern, which the brain then interprets as position or realism rather than as timbre.

In setting up a home listening system, it is not possible to achieve a realistic concert hall sound field unless the cues provided by the pinnae at the higher frequencies match the cues being provided by the lower frequencies of the music. When the pinna cues don't match the interaural low frequency amplitude and delay cues, the brain decides that the music is canned or that the reproduction lacks depth, precision, presence, and palpability or is vague, phasey, and diffuse. But even after insuring that our pinnae are being properly serviced, other problems are inherent in the old stereo or new multi-channel surround-sound paradigms. We must still consider and eliminate the psychoacoustic confusion that always arises when there are two or three widely spaced front loudspeakers delivering information about a stage position but erroneously communicating with both pinnae and both ear canals. We must deal with non-pinna induced comb-filter effects and the stage-width limitations still inherent in these modalities even after 80-years. But this is a subject for the next chapter.

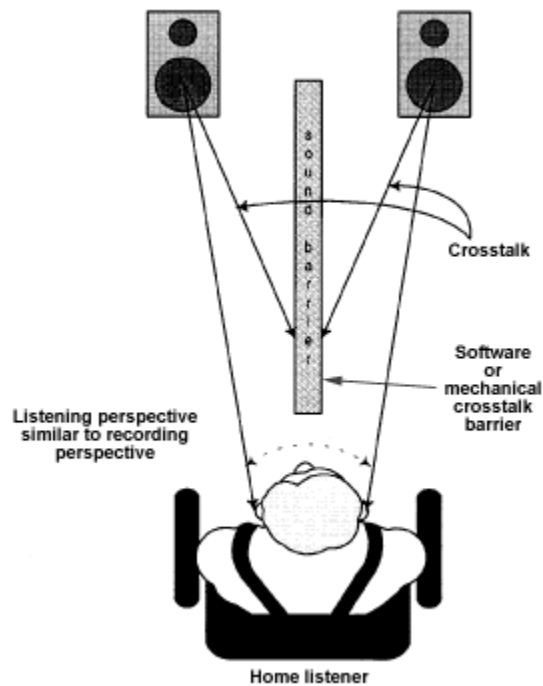


Figure 4: Ambiphonic main front channel listening arrangement eliminating crosstalk and mimicking microphone view

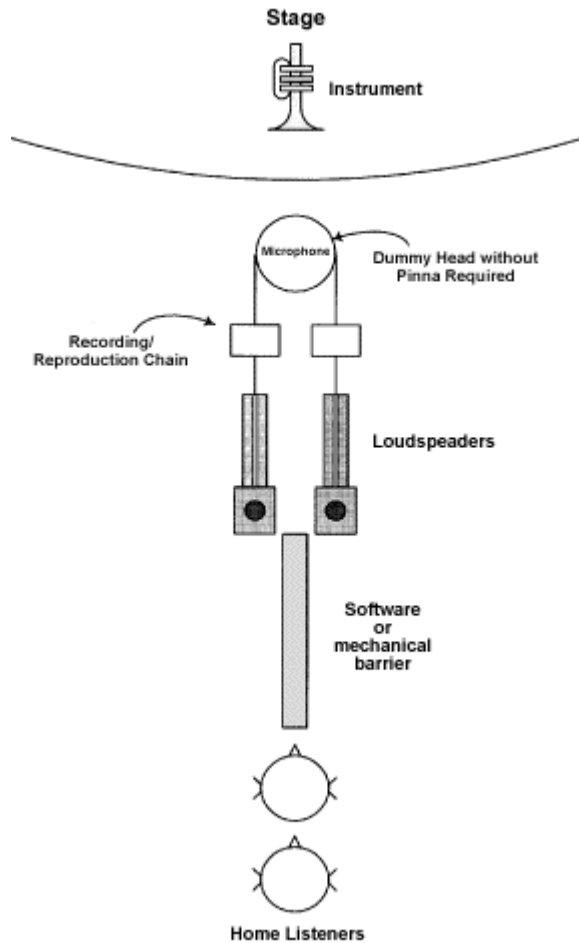


Figure 5: Ideal Ambiophonics recording/reproduction chain.

Chapter 5

A Listening Room For Ambiophonics

Originally, I thought that, as in stereophonics, listening room reflections would be deleterious to Ambiophonics. But this is not the case. In Ambiophonics room treatment or correction is not normally required except for the lowbass region where, as in stereo, room modes are a problem. However, speaker response correction is highly desirable. The reasons for these conclusions are discussed below.

The three main pillars upon which Ambiophonics is constructed are the software or mechanical Ambiodipole, surround ambience convolution, and speaker correction. One can enjoy ambiophonics listening to two little PC speakers so obviously it is not necessary to have a special room. Fortunately, turning a family room, spare bedroom or rec room into an acoustically viable environment for a quality domestic concert hall or surround experience need not require a big budget, a building permit, or even a single carpenter. The trick is to understand what factors might degrade Ambiophonic realism and then do something about them. Again, although room treatment or correction is not essential to an

Ambiophonic experience, the technology involved will be described below for those who opt to create the ideal listening environment.

In stereophonic reproduction where the speakers are far apart, room reflections have a delay comparable to the direct sound and so do effect localization and interfere with the functions of the pinna. In Ambiophonics however the speakers are quite close together and one usually sits closer to them. In this case the early reflections from the room are somewhat later than the direct sound and so the effects of room reflections are minimized and akin to the early reflections one would hear in a concert hall from other seats and heads. Furthermore, if one uses surround speakers to provide ambience derived from hall impulse responses, then the even longer delayed signals from these surround speakers swamp any short delay room reflections. This does not apply to the low bass where room modes can effect Ambiophonic reproduction as much as they do stereophonic. However, having the front speakers so close together makes it easier to do bass management using the room response correction software and processors now available from several sources.

While two speaker Ambiophonics is already a major improvement over the stereo triangle, the ultimate in reproduced realism depends on the collaboration between crosstalk free front and rear speaker pairs and additional surround speakers (if the performance is situated in a large space rather than the home listening room or small studio). In general, a truly exciting binaural hearing experience is the result of the ear-brain system not having to labor at resolving conflicting sets of acoustic cues: the concert hall (as presented by the playback system) on the one hand, and the local playback environment on the other. The less adulterated the set of cues, the more persuasive the experience. Again, in most Ambiophonic systems the effect of the local environment is much less audible than is the case in stereo or 5.1.

The possible causes of acoustic disappointment are many but, happily, experience shows that most home media rooms suffer from insufficient absorption. For realistic concert-hall like reproduction we must eliminate any characteristics of the home listening room that modify the rear hall and front stage we are going to create via convolution and the Ambidipole. For Ambiophonic purposes it is only necessary to get the reverberation time of the room down to about 0.2 seconds, which is far from the .01 of an anechoic chamber. Remember that in a real concert hall there are some short early reflections from nearby seats or people and so a completely dead environment at home is not a requirement. However the trick is to deaden a room over the entire audible frequency range and this requires different techniques in the treble and in the bass. In brief we will see that treble reduction of early room reflections is best done using inexpensive wall treatment while bass reflections and room modes are best tamed with electronic speaker/room response correction systems. Again this is gilding the lily and except for the bass response, the improvement is likely to be inaudible to most listeners.

Reflections

Sounds arrive at a listener's ears from many directions: from sources themselves (the speakers) and from walls and objects that reflect sound toward the listener, much as mirrors reflect light. Because reflected sounds must travel further, they arrive at the listener after the direct sound with an altered frequency response and loudness level. The brain interprets these reflections differently, depending on which direction they come from, on how much later they arrive, how they are tonally changed, and how much louder or softer they are. (Curiously, reflected sounds can sometimes be louder than the direct sound in small rooms if they take two or more paths to the listener-say from the ceiling, floor and a side wall-and if the path lengths are the same so that they are additive. A reflected sound that follows the direct sound by less than about one-fiftieth of a second is perceptually fused with the direct sound, i.e., the brain generally cannot distinguish the two as separate acoustic events. But despite this, uncontrolled, strong, and very early reflections (0 to 20 msec) make a mess of perceived tonal quality and wreak havoc with stereophonic (but seldom Ambiophonic) imaging. Reflections arriving somewhat

later are interpreted as room ambience. Reflections trailing the direct sound by more than about one-fifteenth of a second can be heard as discrete echoes or more likely as reverberation. Shorter echoes can be particularly offensive if the room concentrates or focuses such sound. Concave room features, in general, such as bay windows, are frequent culprits and should be avoided if high-quality acoustic results are intended.

Room surfaces have three primary acoustical properties-absorption, reflection, and diffusion (a complex form of reflection)-but only absorption is of real use in the cause of eliminating audible room reflections at the listening position. Couches, carpets, cabinets, bookcases and other furnishings all contribute to a room's reflection patterns, albeit usually in unplanned and acoustically erratic ways. For example, carpeting on a concrete or hardwood floor soaks up a fair amount of treble energy, but allows bass to bounce right back into the room. Large closed glass windows typically reflect middle and high frequencies back into the room, but let bass pass right through. A bookcase might absorb highs, scatter (diffuse) mids, and ignore the bass altogether. Thus, a room for Ambiophonic listening should, if one is pursuing this option, be treated for a real reduction of wideband reflections.

More Evil That Rooms Do

While the ideal Ambiophonic loudspeaker would aim its sound only toward the listeners, most loudspeakers spread their output, to some extent, like floodlights illuminating both people and surroundings. A speaker firing directly at the listener will also direct sound sideways, up and down, even backwards. In a typical untreated room, this "unaimed" energy hits a wall or cabinet and bounces back toward the listener only a split second after the direct sound. Think of these delayed versions as the acoustical cousins of multi-path "ghosts" on a TV screen. But this is true of most all acoustical environments and the brain can cope with this if the field is binaural but not if it is stereophonic.

Thinking Ambiophonically, it is necessary to understand the following. The average untreated living room has a reverberation time of about six-tenths of a second. Since a recital hall could have a reverberation time of as little as eight-tenths of a second, and even a concert hall can be in the 1.5 second range, the typical home listening room reverberation time is surprisingly significant compared to the halls in which music is performed. Let us assume that we are playing a recording of a large choral work that includes a normal ratio of direct sound to hall reverberant pickup. When such a recording is played stereophonically in a typically small, live, home environment, the direct sound stimulates the listening room to produce a reverberant field that tells the brain that the performance is in a room that is small and bright. But then the recorded reverberant field reaches the ears and tells the brain that the room is large and acoustically warm. When you add to this the comb filtering and pinna effects due to the spurious directional early room reflections that further confuse the brain, it is no wonder that stereo playback of recordings of larger musical groupings never seem to be realistic no matter how much we tweak our systems.

In the Ambiophonic case, one normally has surround speakers generating ambience with reverberation times of around 2 seconds so if the room reverb time is reduced just by half to say .3 seconds the room effect becomes inaudible. Additionally with the front speakers close together one can be a bit closer and the ratio of direct sound to room reflections increases making the room effects on localization that much less significant.

Soaking It Up

Absorbers are devices designed to soak up sound. Most absorbers work by converting acoustical energy into thermal energy. Typically they do this by forcing sound waves through a dense maze of small fibers that rub together to produce friction and heat. Carpet, soft furnishings drapes and even clothing can

provide useful absorption in the treble and upper midrange, where you'll find female vocals, violins, trumpets, flutes, cymbals, and other high pitched sounds.

Acousticians refer to special sound-soaking materials like fiberglass batts as frictional absorbers, or more colloquially, "fuzz". Generally, the thicker and denser the fuzz, the more effectively it traps sound. A dense, two-inch thick fiberglass panel mounted directly on a wall absorbs nearly 100% of sound incident upon it in the range from 500 Hz (about one octave above middle C on the piano) up to 20,000 Hz, the approximate upper limit of human hearing. To absorb much energy below 500 Hz requires a significantly thicker panel, usually 4 inches, or an air gap of a foot or two between the panel and the wall. Either way, using fuzz to soak up the lower midrange and bass requires considerable space. Devices such as resonating tubes, i.e. fuzz surrounding a tall tubular cavity are only marginally effective in the bass region. As we shall see the expense of such bass absorbing devices is much better invested in a computerized speaker/bass correction system.

Splayed Walls for Fanatics

If building a new listening room or remodeling an existing room. It is possible to splay both of the side walls and front and rear walls. The walls should lean outward at an angle of five degrees or more as they increase in height. The conventional wisdom has been that eliminating parallel surfaces is not worthwhile since the behavior of such a room in the bass frequency region is unpredictable in advance and hard to measure after the fact. But bass standing waves are not the only problem one must find a solution to and bass correction systems handle bass without difficulty even if the walls are splayed.

For upper midrange and high-frequency sounds the soundwaves coming from floor-standing loudspeakers will be reflected, as light would, in an upward direction. As these rays go from wall to wall they must go up to the ceiling before they can return to ear level. Hopefully, in making this longer up-and-down trip, they will lose significant energy and also fall beyond the critical 20-millisecond early reflection time zone. This is essentially a benign form of diffusion, which largely avoids diffusing sound to the listening position. In general, splaying the walls can make the absorption treatment of the walls and floor a little less critical.

Reverberation Time

The amount of absorption that should be placed in a room varies according to the room's size. All things being equal, a big room sounds more live than a small one, requiring more absorption to bring it down to the same level of acoustical merit. This quality is expressed as reverberation time: the amount of time it takes for a sound in a room to drop 60 decibels in level from the moment the source stops producing sound. The shorter the reverberation time, or T60 as it is called, the dryer the room sounds. In general, a dedicated Ambiophonic listening room should be quite dead with a reverb time of .3 seconds or less. Because it is derived by averaging the time it takes sound to decay by 60 decibels across a broad segment of the audible spectrum, describing a room with a single reverb time figure is often misleading. A poorly designed room might boast a textbook-perfect average of T60, yet sound disjointed and unpleasant because some frequencies die out quickly while others linger on and on. Ideally the T60 in any one-third-octave band between 250 and 4,000 Hz should not deviate from the average T60 by more than 25%. Translated into frequency-response terms familiar to audiophiles, this ensures that the room's reverberant sound energy is flat within about a decibel or so throughout the most sensitive range of human hearing.

One challenge lies in controlling reverberation in the bass frequencies where T60 figures might easily be triple or quadruple that in the midrange. Unfortunately, the use of fuzzy coated tubes, Helmholtz resonators and other well advertised gimmicks are largely ineffective. But if left unaddressed, the lack of low-frequency absorption can create an annoying unevenness in the reverberation characteristic of a

home theater, media room or Ambiophonic home concert hall. We shall see below that electronic room correction systems are the answer to the Ambiophile's prayer.

The Weight of The Sabines

The Sabine is the unit of sound absorption and it is computed by multiplying the area of an absorbing surface in square feet by its absorption coefficient. The absorption coefficient is simply the fraction of sound that is absorbed by the material at a particular frequency or over a band of frequencies. Thus a window open to the outside swallows up any sound that passes through it and, therefore, has the highest possible absorption coefficient of one. If the open window is one-foot square, its total sound absorption is one Sabine. Ten square feet of 4-inch thick fiberglass could absorb some 9.5 Sabines at 500 Hz and higher, but only about 7 Sabines at 100 Hz. A 660-cubic-foot room (10x14x19) would need approximately 700 Sabines of absorption to get down to a reverberation time of .2 seconds. Using 4-inch fiber wall panels, the area requiring padding would be in excess of 700 square feet, or about half the surface of the room allowing for the small absorption contributed by other surfaces such as rugs, drapes and furniture.

Tacking up 100 square feet of fuzz on each sidewall yields the same absorptive value and produces the same T60 as moving the fuzz to the front and rear walls. However, the quality of the stereo sound you hear, even the intelligibility of music and dialog, could differ dramatically. Absorption is best deployed on the ceiling and the front portions of the side walls, where they prevent sound from the main front stereo speakers from bouncing into the listening area a split second after the arrival of the direct, speaker-to-listener sound. If left untreated, these reflective surfaces allow strong early reflections to disrupt tonal balance and imaging, and scramble the often subtle aural cues in stereo that give music and soundtracks their texture and life. However, the Ambiophonic situation is much less critical and one can just treat as many surfaces as one can bear to treat. As with most things in life, compromises may be necessary. Remember, even if your listening room is not Ambiophonically perfect, neither are most concert halls.

Background Noise

Sabine noted that halls exhibit the same basic sonic behavior at very low sound levels as at very high ones. If you are an active concertgoer, you may have noticed that concert halls show their distinctive sonic personalities even during those hushed moments when the maestro mounts the podium and raises his baton. Recreating in a residential setting the characteristic sound of a real hall begins with getting that "silence" right. Unfortunately, the typical home is neither designed nor constructed to allow the Ambiophile to hear the desirable level of sonic detail. If you turn off your playback system, shut the windows and door, and just listen to your listening room for a few minutes with eyes closed, you'll be aware of how much noise is there. Acousticians have developed a sort of numerical shorthand to describe background noise levels. Known as "noise criteria" (NC curves, and usually specified in increments of 5, from NC-70 (extremely noisy) down to NC-15 (very quiet). These curves are weighted to account for the fact that the ear is less sensitive to low frequencies than to high. The curves' numerical designations are arrived at by taking the arithmetic average of sound pressure levels at 1 kHz, 2 kHz, and 4 kHz. A useful target for a purpose built Ambiophonic listening room is NC-20; a spec often encountered in the design of professional recording studios. NC-35 would be the minimum standard for a legitimate Ambiophonic experience.

Unlike treating a room to reduce reflections, keeping outside noises outside is probably a job for an outside contractor as major structural alterations involving gypsum, studs and concrete are often required.

Bass Behavior

One of the most universally vexing problems of the home audio experience, stereo or Ambio, the fact that residentially sized rooms give erratic support to low-frequency sounds. When a particular bass note's wavelength precisely fits a major room dimension, the note is strongly reinforced or cancelled in a phenomenon called a standing wave. Bass will boom or fade depending on where one is in the room and the frequency involved. The room also exaggerates or cancels any higher harmonics of these low bass frequencies. However, as the absorption properties of the room begin to take their toll this standing wave effect fades. Basically, standing waves are due to the fact that most rooms simply can't attenuate bass reflections enough to prevent them from interfering with themselves over several rebounds. Or another way of stating the same thing is to observe that the T60 bass reverberation time of most small rooms is much larger than the treble T60 and that the density of this home tail is much greater than that found in the concert hall.

Eliminating bass modes is the subject of much quackery. There are magic room dimension ratios, which help a little, and there are the resonant boxes and tubes for room corners which help a little. But even a room that is painstakingly dimensioned and equipped with tubes galore to provide the smoothest possible distribution of low frequency modes will seem bass-boomy in some places, weak in others and about right somewhere else. Fortunately new Digital Signal Processing logic has come to the rescue to solve this problem with singular success and relatively low cost and simplicity.

Room Correction Systems

In essence we want the bass response of the room to be correct at the usual listening area of the room. It really doesn't matter much what is happening in the corners or behind us when we are not sitting there. So let us temporarily set up a microphone at the listening position or even several such adjacent positions and measure the bass characteristics of the room and the speakers (and the amplifiers for that matter) at that point. Once we know what the room and the speakers are doing to the bass we can get a digital-computing engine to correct any errors in bass response in both amplitude and time. A room correction system is essentially a very fine parametric equalizer able to control amplitude at any bass frequency (or treble for that matter) with a resolution of 2 Hz or better. The most exciting feature of an RCS is its ability to measure both the speaker response and the effects of the room on this response and do something about them. Once the peccadilloes of the speaker and room are known (by launching a series of test impulses through the system to the microphone and getting the impulse response of the setup), the room correction software can then calculate the fine grained amplitude and delay equalizer settings needed to eliminate them.

The methodology of measuring the impulse response rather than the frequency response has a tremendous advantage over conventional steady-state-tone measuring methods. Say one measured the bass loudspeaker/room response using a sinewave oscillator and a microphone attached to a meter. Then, using the resultant curve to set a conventional equalizer feeding the speaker, one would assume that a flat bass response would be achieved. Wrong! Music, in particular, consists mainly of transients. Thus, if a standing wave in the room causes say a loss of 10 dB at 100 Hz at the listening position and we apply a 10 dB boost at the speaker, then a brief but audible 10 dB peak will be heard until the standing wave room response catches up to cancel that peak. It is not the frequency response of the speaker/room system that needs to be corrected but the transient response. While improvements in this kind of room correction system will come at an increasing rate, present systems can only cancel early reflections within a period of one wavelength of the frequency involved. Thus, a reflection off the rear wall from ten feet behind the microphone will be delayed about 20 ms. This delay corresponds to the period of a frequency of about 50 Hz. Thus a typical room correction system will not be able to deal accurately with the components in this reflection at frequencies above this. On the other hand, bass corrections are quite effective for near reflections coming from the floor, ceiling or walls. It is providential that the electronic room correction systems work best where conventional absorption treatments work worst.

It is inevitable that room correction modules will be included not only in Ambiophonic processors but in stereo and video control centers as well. It is anticipated that before too long Ambiophonic processors will appear that include room correction, Ambiodipole software and the real hall convolution computer.

Chapter 6

Ambiophonic Loudspeakers - Ambipoles, Ambiostats, and Surrstats

In an Ambiophonic system there are two different functions the loudspeakers must perform. The first one is to generate the phantom-image/comb-filtering free front stage and the second one is to reproduce the surrounding concert hall ambience. I call each speaker an Ambipole (or Ambiostat if electrostatic) and a pair of speakers that generate the front or rear stage an Ambiodipole. The other speakers that provide early reflections and reverberation tails are called Surrstats. While I will describe the ideal loudspeaker for each Ambiophonic purpose, the ultimate choice for audiophiles will, as always in stereo or home theater, be determined by their, budget, space, and what they already own that can be adapted to this purpose.

We wish to apply the rules of good concert-hall design to the choice of home concert-hall loudspeaker characteristics and speaker placement. Let us assume that we have available the high quality software-generated hall ambience signals described in detail in Chapter 8. Let us also assume that our listening room is treated well enough to eliminate the bulk of the counterproductive listening room reflections using absorption panels and hopefully an electronic speaker/bass correction system as described in Chapter 5. Furthermore, let us also assume that we will be using the software or barrier Ambiodipole arrangement for the left and right front channel speakers which are separated by a 20 degree angle directly in front of the listener or listeners as described in more detail in Chapter 7.

There is one general characteristic that applies to all the loudspeakers used in a domestic concert hall: all speakers should be as focused or coherent as possible, so as to reduce the number and level of slightly delayed sound rays. Since RACE crosstalk cancellation depends on the speakers being identical and symmetrical, speakers with multi-drivers and complex crossovers may be difficult to make identical. Conversely full range electrostatic loudspeakers are easy to control this way. Small satellite speakers behave like flashlights and are thus well suited for use as low cost Ambiodipoles.

The 5.1 Home Theater Conundrum

Since so many of my readers are devotees of video home theater and its ad-hoc arrangement of two surround speakers placed at the rear sides, I think it would be best to first discuss the shortcomings of this arrangement before proceeding to describe something more realistic and scientifically based. The home theater movie people recommend two dipole speakers placed on edge so that the acoustic null such speakers produce is facing the listening position. Dipole speakers are speakers that radiate sound equally loud from opposite sides. Additionally, these sounds are of opposite polarity and so cancel where they collide in a room. Some dissenters argue that monopole, that is direct, single polarity, radiators, are better. Either type of rear surround speakers may be reasonable some of the time for movie and video sound reproduction. However, where classical music, jazz, etc. is concerned 5.1 has ignored some serious and seemingly insoluble acoustic problems common to both of these rear surround speaker types. The use of dipole speakers assumes that the listening room is quite live, because otherwise, the dipoles would be relatively inaudible. But a live home theater room means that the direct sound from the front speakers will be reflected, in spades, from all these nearby surfaces, especially since there are three of them up front emitting direct sound. To add insult to degradation,

their reflections cannot be thoroughly eliminated by room treatment with absorbers or diffusers, if the dipoles are to function properly. In movies, these spurious early reflections only slightly impede our ability to locate dialog and sound effects because of the precedence effect and because the brain has no preconceived notions of the acoustic spaces the rapidly changing scenes are supposed to be set in. In contrast, in classical music reproduction, these early direct sound home theater wall reflections produce cues indicating the hall is small while the recording and the brain say the hall must be large. The brain usually resolves this contradiction by deciding that the music is canned. This is one of the many reasons realism, as opposed to mere localization, in both stereo or multi-channel 5.1 music reproduction, is such an elusive goal.

But if one forgoes dipole surrounds and uses directional rear speakers in conjunction with room treatment, the front stereo or 5.1 three-speaker stage improves, but music reproduction still sounds unrealistic. This effect in 5.1 is due to the fact that all the rear half-hall ambient sound is coming from two discrete speaker locations where as in a real hall the sound comes equal in power (but not in detail) from all directions (diffuse field). Even if the recording is made so perfectly that no direct stage sound is emanating from these rear speakers there is still no concert hall in the world that delivers all its early reflections and reverberant tails from two small side spots. The resultant pinna angle error added to the abnormally low interaural cross correlation factor, signals the brain that something is rotten in the state of Dolmark and again the result for 5.1 classical music reproduction is disappointment for any experienced audiophile or concert goer. Already this point is being conceded by products that now offer 7.1 or 12.2. Even conceding that monopoles with room treatment are better for music than dipoles without it, the problem remains that making such multi-channel recordings or trying to extract hall ambience to feed more than two surrounds (without including erroneous proscenium direct sound or frontal early reflections) is easier said than done. The answer is full surround convolution as discussed in chapter 8.

The Front Speakers

In an Ambiphonic system the front speakers should be placed almost directly in front of the listener with each speaker aimed at the listening area. (See next chapter). For best results the front main speaker pair, the Ambiodipole should be as directional as possible. In, theory the ideal speaker for this purpose would behave like a flashlight, with a sound beam emanating from a single point at ear level and the rest of the room in deep shadow. The more focused an Ambiopole is, the more effective the software is.

The front speakers used should be capable of reaching concert-hall volume. The normal speaker selection criteria of good frequency response, low distortion, reasonable time coherence and affordable price naturally still apply. Since the use of a speaker/bass correction DSP can correct most speaker response anomalies, one can choose the front speakers based primarily on their radiation patterns. An Ambiphonic speaker, designed by Soundlab, called an Ambiostat, can be used in pairs to form a virtually perfect Ambiopole. One such model is a six-foot by three-foot vertical panel that is slightly curved in the horizontal direction. The behavior of such panels as vertical line sources makes the job of crosstalk elimination that much easier.

Front Early-Reflection Loudspeakers

At least one pair of the early-reflection (but not later reverb) speakers should be placed about the critical plus or minus 55-degree angle to the listening position. This angle is where the ear is most sensitive to such spatial cues. Of course if many surround speakers are available then they should be spaced in whatever way is most convenient or specified by the hall convolver. The ideal speaker for this purpose is one that radiates to the listener from as large an area as possible just as concert-hall walls do. Large electrostatic or ribbon loudspeakers are excellent in this application especially if they can be

turned to a horizontal position. If they are dipoles don't forget to put sound absorbing material behind them. A useful property of such large-area full-range sound radiators is that they provide significant diffusion without the need for physical diffusion panels. Ideally, one wants all surround speakers in an Ambiophonic system to cover as wide a horizontal arc as possible. This corresponds to the situation in a real concert hall, where the predominant early reflections arrive from many different side directions because the originating sound sources are spread out on the stage and have various angles of incidence and, therefore, reflection. In the home environment, the computer reconstructed early reflections are the same for all the right-channel instruments, the same for all the left channel instruments, and the same for all the center instruments. This moderate lack of precise spreading of the apparent early reflections would seem to detract from the concert-hall ideal. But just as the perfect Philharmonic Hall has yet to be built, so our home room may be real but not 100% ideal. By mounting speakers on their sides or by leaning tall speakers, at say a 45-degree angle, the ambient signals arrive at the listening position with a greater diversity of direction and delay. As discussed below, the reverberant field needs to be as diffuse as possible. Therefore, to the extent that either recorded reverberation or recreated reverb is present at these side-rear loudspeakers, there is an additional benefit to being wide and as horizontal as possible, providing both vertical and horizontal dispersion.

Soundlab has produced a speaker called a Surrstat which is essentially an Ambiostat turned on its side. Being slightly concave toward the listener it delivers ambient sound most efficiently over a wide angle without. Its rear wave should be absorbed, although the convex side, radiating over a wider angle, reduces, on average, the intensity of the resulting reflections impinging on the listening position in most rooms. Eventually, distributed mode loudspeakers, that are essentially flat panels that radiate equally from both sides of their entire surface, will be quite useful in this application.

Side and Rear Reverberation Loudspeakers

Side and rear speaker pairs are fed with some early reflections but largely uncorrelated reverberation tails. Since in a concert hall, various reverb tails reach the listener from virtually all directions, the ideal speaker would be a set of thin squares, which could be hung on all the walls. I find, however, as above, that large electrostatics or ribbon speakers do an excellent job, particularly if they can be mounted horizontally. One could also use multiple small, inexpensive box speakers arranged on pedestals around the rear half of the room. Again, in theory, each reverberation sound source should have its own independent reverberation computer but the Japanese have shown that such speaker walls can easily fool the ear-brain system even when some of the reverberation tail speakers are correlated. Incidentally, there is no reason why Ambiophonic surround speakers need to be matched if they can still be reasonably set to the sound level required.

Since the rear reverberant field often has a strong vertical component coming from the auditorium balconies and ceiling, we have found it advantageous, but by no means critical, to use one pair of rear speakers elevated as much as possible. These sometimes provide a richer simulation and a better match to concert-hall design theory but "better real" is not more real than "real" and this suggestion is, perhaps, gilding the lily unless the measured hall response specifically includes elevation data.

Chapter 7

Ambience Convolution

One precept of Ambiophonics is that for music one should be sure to surround ordinary two channel discs with a fully directional reverberant field. For existing recordings, one uses the techniques

described below to produce hall sounds for surround speakers. The Ambiophonic alternative is to use the Ambiphone (described elsewhere) to record the rear half of the hall and then feed this to a rear Ambidipole so as to generate the rear half circle of hall sound. But this method requires four channel media, new recordings, a special microphone and does nothing for the huge existing library of CDs, LPs, etc.

Again, one of the main precepts of Ambiophonic theory is that where music recording is concerned, it is counter productive to record concert hall ambience during a recording session using microphones and then waste DVD/SACD/MP3/.wav/5.1 bandwidth delivering this defective ambience to the home listener. To understand why this is so, it is necessary to review what we know about how concert halls, opera houses, recital halls, churches, recording studios and rock pavilions operate. A concert hall, theater, or other auditorium is essentially an analog computer. What this hall computer does is operate on (convolve) each ray of direct sound originating on the stage to transform it in amplitude, frequency response, and direction before delivering it to a given seat in the audience area as hall ambience. (In a good hall, without obstructions, we can assume that the original direct sound reaches most of the seats without passing through the analog computer of the hall.) If we consider every seat in the hall, the number of such equations is almost infinitely large but for our purposes we can assume that we are only interested in what this computer is delivering to one or two of the best seats in the house from the left, right and central areas of the stage.

If we now put a measuring device at this best seat and launch a series of test signals from say three positions on the stage it should be possible to determine the most significant equations used by this concert hall computer to deliver ambient sound to this area. Indeed, this is not only possible but can now be done with such finesse that it obsoletes every other method of recording or delivering surround sound for music to the home listener.

The equations that a hall uses to deliver sound to an audience are usually invariant for the duration of not only that performance but over the lifetime of the hall barring serious renovations. Once the equations of a hall are known, there is little point in measuring them every time that space is used to make a recording. We ignore here the slight variations in hall responses depending on the size of the audience present when the hall is measured. I should add that the latest methods of measuring hall responses make it possible to measure halls with the audience there without making them too uncomfortable or straining their patience. There are some who believe that hall impulse responses will soon be measurable while a concert is in progress. Unfortunately in the case of movies where the scene changes frequently, this method of surround sound generation is not feasible. However there is a viable alternative to consider later.

Why Recording Hall Ambience Directly For Surround Speakers Using Microphones Is Not Possible to High-Fidelity Standards

In a concert hall, early reflections and reverberation tails reach a listener from all directions. But in good halls this ambience is not the same in all directions. That is, there is a strong interaural directional component present that interacts with the shadowing function of the head and the pinna structures to allow the hall to be appreciated in all its glory by concertgoers. At home it is necessary to deliver as many of these hall elements as possible without compromise as to these directional ambience components. If the direction from which hall sound comes were not important then reverberation could simply be fed into the front stereo speakers and no surround ambience speakers would be required. But after seventy years of the stereo triangle era, it is clear that doing this can never sound realistic.

Of course, it is laughable to think that the two or even three surround speakers of the 5.1/6.1 Dolby/DTS/Bluray arrangement could deliver a reasonable replica of what a concert hall does. But even if we ignore this issue for the moment, how do we get the 5.1 signals required to drive the two

surround speakers or the three if a centered rear speaker is used? The recording engineer needs to set up two or three microphones in the hall for the express purpose of generating signals for these speakers. But where in the hall should he place these extra microphones? Answer comes there none. But worse than this ad hoc decision is the fact that most microphones are not very directional. Thus if a pressure microphone is used it will, say, pick up all the early reflections and reverberation tails coming from the ceiling the sides and the rear and lump them all together to later come out of a surround speaker whose location at home and radiation pattern is anybody's guess. Cardioids and velocity microphones are more directional but which way should they point? Invariably, proscenium early reflections will end up coming from the side or even worse the rear and ceiling ambience will be arriving from ear level, etc. Mixing several mic's together does not solve the problem. Of course, many surround tracks are made without benefit of any microphones (using the Lex in record producing parlance) because of these and cost problems. We will see below that before too much longer the virtues of deriving the surround channels from hall impulse measurements rather than microphones will be quite apparent to all music. if not video, recording engineers. Another issue is that this ambience, being recorded willy-nilly somewhere in the hall, does not represent the reverb one would be hearing at the best seat in the hall or indeed any seat unless the ambience microphones are all quite close together about that seat.

Some Ancient History for Skeptics

Once one decides that hall ambience is indeed needed to perfect the reproduction of 2 channel (or 5.1 for that matter) recordings so as to produce a "you are in a concert hall" experience, there are only two ways to go. One is to pick a fine concert hall, construct a model of it at home and put two loudspeakers on its stage. That this technique does work was demonstrated conclusively several times in Carnegie Hall and Carnegie Recital Hall in the 1950's by Gilbert Briggs of Wharfedale Loudspeakers, and most notably by Ed Vilchur, the founder of Acoustic Research. I attended live-versus-recorded presentations by both these gentlemen in New York and not only could I not tell when the live musicians ceased playing and the recording took over, but almost on one else in the sold-out house could either, judging from the gasps and buzz in the audience when the string quartet players finally put down their bows and the music played on. The fact that such an illusion could be created with low-powered vacuum-tube amplifiers and excellent but still relatively primitive loudspeakers, should have tipped us off to the fact that ambience is essentially everything, and equipment quality relatively insignificant where realism is concerned.

It is possible, if impractical, to construct a smallish room that would closely mimic the ambience of Carnegie Hall, at least in the central listening area. The use of modern diffusers, absorbers, and ceiling and floor treatments could produce the reverberation time, reverberant-field frequency response and even the early reflection pattern of any good concert hall. It would then be possible to play recordings in such a room to excellent effect. The advantages of this approach include the fact that such a room would also be excellent for live music soirees as well.

The disadvantages of this approach, for the reproduction of recorded music, are several and instructive. The costs of designing, constructing and tuning such a room are beyond the reach of those of us not direct descendants of Andrew Carnegie. One would also lose the flexibility of being in other acoustic settings such as churches or recital halls. Both Briggs and Vilchur used their own recordings, carefully made to avoid any recording-site hall coloration. Finally the problem of stereo signal crosstalk would remain for most listening positions. In the Briggs Carnegie Hall demonstration, (which, I believe used mono recordings) most listeners in this very large hall were exposed mainly to the reverberant field and their visual senses substituted for any missing or weak directional sound cues.

Characteristics of an Ambient Field

Basically, the only things you can do to a sound wave, launched in an enclosed space, are attenuate it, usually as a function of frequency, or change its direction. Absorption is a form of extreme attenuation. But sound loses intensity merely by traveling a distance through air. A characteristic of attenuation is that it is almost always frequency sensitive, with higher frequencies usually rolling off more than lower frequencies, in air, with distance, or in sound absorbing material. Sound changes direction whenever it encounters an obstruction—usually by reflection as light does (specular reflection), or by diffraction, which is a process by which sound waves sort of ooze around obstacles. As in attenuation, reflection and diffraction are frequency sensitive, with higher frequencies usually being easier to steer or control. Thus every space, but especially a concert hall, can be described acoustically in terms of its attenuation characteristics and its three-dimensional reflectivity pattern as a function of frequency, direct sound-source position, time, and listener-seat location. Our problem is then to either measure these functions in the real halls we like and recreate them via surround speakers in our listening room or design a pleasing but entirely new hall in software that may not exist physically. Both of these approaches are possible using the early JVC or SONY hardware convolvers or the software methods discussed below. It is also always possible to start with a real hall and modify it to taste as you listen to your favorite music.

We need to be able to create any kind of acoustical signature we like within our treated listening room. We have to be smart enough to invent a hall ambience processor that can generate any field, we or the recording engineer want. There is no reasonable alternative to using a special-purpose computer to generate the early reflections and reverberation trains. The only major issue still to be resolved is who should control or own the convolver: the record producer, or the home audiophile. But we need more technical background to decide this issue.

Early Reflection Parameters

To produce a realistic group of early reflections, a computer or digital signal processor needs to recreate and vary the following parameters separately for the left and right stage sounds. These items determine how big the hall is, what its shape is (such as rectangular, fan or low ceilinged), how large the proscenium is, etc.

- The delay between the direct sound and the arrival of its first reflection
- The delay of the second and subsequent early reflections and their density
- The frequency response of these discrete early reflections
- The initial amplitude and rate of amplitude loss for the subsequent reflections of these very early reflections
- The source of each reflection: front, side, rear, left, right, up, down, etc.

Normally these parameters are measured in real halls, churches and opera houses and then stored in memory. If the stored reflection patterns are not pleasing, then they can always be modified to taste. Tweaking such parameters can be a lifetime occupation, as it is with some famous concert halls that are forever being tinkered with.

Reverberation Tail Parameters

After the early reflections become so dense and weakened that the ear is no longer sensitive to their individual arrival times, the reverberant characteristics of the space become evident. The reverberant parameters that need to be recreated by a convolver separately for the left and right signals include:

- Reverberation decay envelope for high frequencies
- Reverberation decay envelope for low frequencies
- Frequency responses for the front, side, rear, overhead, etc. tails with time
- Density of the reverberant field
- Directional characteristics of the reverberant tails

If early reflections persist for a relatively long time before the reverberant field begins, then the space will be perceived as live and possibly large. If the reverberation time is long then the hall will seem live, or if very long, cathedral-like. High-frequency rolloff in the reverberant field also makes the hall seem larger. The directional distribution of the reflections and the reverberant echoes help listeners determine the shape of the space and their position in it.

Again, rather than attempt to program all this from theoretical scratch, it is more practical and likely desirable to measure several good existing halls and store the results. The Japanese, and JVC, Yamaha, and Sony in particular were the pioneers in doing just this. The JVC XP-A1010 Digital Acoustics Processor, circa 1989, seemingly the first really commercially produced convolver, (abandoned in haste when 5.1 movie surround sound took over) stored within its memory the key parameters of fifteen actual halls including six symphony halls of various shapes and sizes, an opera house, a recital hall, a church, a cathedral, two jazz clubs, a gymnasium, a rock pavilion, and a stadium. The Sony professional convolver was the first of a later generation to appear. Sony produced four CD-ROMs each storing some eight impulse responses of the great halls and other enclosed spaces of Europe, Japan and America.

Impulse responses and convolution are techniques that have been proven indispensable in designing new halls that work the first time a note is played in them. The new concert hall of the Tokyo Opera City was designed using computer simulations and a one tenth scale model that allowed Leo Beranek and Takahiko Yanagisawa to hear what the hall would sound like before it was built. They could hear how the sound changes with the location of a seat in the hall, or with the addition of a diffusion cloud, or changes in the shape of the hall, etc. Such hall characteristics as intimacy, clarity, spaciousness, bass ratio, could then be adjusted to match the characteristics found desirable in existing great halls.

However, audiophiles can have an advantage that architects can only dream about. Our halls are not cast in diffuser wood. For if great halls can be simulated to such perfection using convolvers and auralization then why build the hall physically? The hall we simulate on our home computer should sound every bit as good as the one being constructed or better since we can vary our at-home halls to better suit the music being played or just to suit our mood. Perhaps we can even make a hall within our home that sounds better than any Leo Beranek could convolve and then construct.

Adjusting Ambience Parameters for Ambiophonic Listening

To play a recording Ambiophonically, using a convolver, one first consults the recording booklet or jacket to see what acoustic space it was recorded in. Was it a studio, a church, a concert hall, an opera house, a recital hall, a theater, etc. Good recordings include frontal proscenium early reflections and reverberation that naturally should come from the front main speakers. Therefore for best results it is desirable to select that hall if it is in your library or use a hall that sounds as much like the recorded hall as possible. You can do this quickly with a little practice by listening to the main front channels with the surrounds switched off, and estimating the reverberation time of the hall, which in most concert halls or opera houses is from one-and-one half to three seconds. Then estimate other hall characteristics such as liveness, and capacity. You then select the stored hall that best matches your research or assumptions. You can also program your guesses directly, bringing up the surround speaker volumes one at a time to the levels that sound most realistic. Such settings can, of course be stored and recalled at any time.

Convolvers can also be told to compensate for the fact that, some of the time, recorded hall reverberation is being re-reverberated and that some rear ambience is coming from the front speakers. When I first started experimenting with the Ambiophonic method I thought this erroneous reverb might be a serious drawback as far as playing existing recordings was concerned. However, it is easy to see why this is not the case. The small amount of extra rear reverb coming from the main front speakers is quite overwhelmed by the ambience from the hopefully many surround speakers. Also, it is not unusual for a physical hall to re-reflect rear ambience from the proscenium. All that this extra frontal reverb means is that the hall is a little bit livelier than the impulse response suggests. Since this is an easily adjusted parameter it can be corrected for if anyone really hears this effect.

If the recording has reverb mixed into the direct sound, as most recordings do, the convolver will convolve this ambience as if it were a direct sound signal, generating additional ambience. What does this really mean however? It simply means the convolved hall now has a longer reverberation time than we meant to set and that the decay at the end of the tail is not as steep. In physical terms it means the hall has had an additional diffusion cloud installed. This is also an easily corrected condition but, even if left uncompensated for, it seldom is audible even by golden eared audiophiles. The convolver adjustment process becomes instinctive after a while and usually takes less than a minute. Compulsive tweekers could, of course, make ambience parameter adjustment their life's work as there are numerous ways to control volume, delay, hall type, decay and frequency response characteristics for each surround speaker individually and each direct sound channel of which there are hopefully only two or four as in Panambio. The saving grace, which prevents tweak insanity is that once the ambience sounds real and reasonably suits the music and the recording, maybe it can still be improved, but real is real. I have found that minor adjustments seem to change only my perceived position in the hall.

Someday Ambiophonic recordings for the audiophile market will be made without significant recorded rear hall sound, the recommended hall parameters will be printed on the label and the CD or DVD will contain coding to automatically operate the convolver. As part of the research for this book, I listened to hundreds of recordings, both LP and CD. To paraphrase Will Rogers, I never met a classical recording (jazz is too easy) I couldn't work wonders with. The most exciting discovery was that monophonic LPs (or CD versions) even from the 20's could be made to sound exceptionally realistic in an Ambiophonic room. The reason for this seems to be that many early mono recordings, particularly acoustics, have very little recorded room reverberation, making it easier to create a realistic sound field to place them into. Also, the absence of a stereo effect in the presence of well-tailored hall ambience tells the ear/brain system that the source is distant. Thus, for large mono ensemble sound sources the listener appears to be in the balcony of a large hall-but balcony or not, real is real.

Because of the cocktail party effect, needle scratch or frequency-response aberrations become minor distractions, and Caruso, Toscanini, or Melchior never sounded so thrilling or three-dimensional before-and the Caruso recordings are over 100 years old.

Measuring Real Concert Hall Ambient Fields

Only three convolvers worthy of the name have ever been commercially available. All are Japanese, one from JVC, one from Sony and one from Yamaha. Although the JVC unit (like the others) is no longer available its technology is still of paramount importance. A group of researchers in 1987 at the Victor Company of Japan (JVC) headed by Yoshio Yamazaki and including Hideki Tachibana, Masayuki Morimoto, Yoshio Hirasawa, and Junichi Maekawa, developed what they called a symmetrical Six-point Sound Field Analysis Method for measuring the acoustic characteristics of a concert hall. In their measurement method, an array of six microphones is placed at a good seat in the hall and a series of test impulses is launched from one or more points on the stage. All six microphones are omnidirectional and are arranged in three pairs. The microphones in each pair are spaced about six inches apart. One pair of microphones straddles the mounting pole horizontally, left to right, one

mounts front to back in the same plane and one pair sits up and down. The center points or origins of each microphone pair are coincident. The impulse, or test patterns launched from the front stage, that each of these microphones hears, then goes to a computer which produces a list of all the discrete early reflections detected by the array, including their time of arrival, their amplitude and their direction of origin.

That such an array can detect all this information is not too hard to understand. For example, any impulse coming from center rear will hit the vertical pair of microphones and the left-right pair of horizontal microphones simultaneously. The front to-back pair will experience the maximum possible back-to-front delay of .4 milliseconds. Thus when the computer detects such a situation it records that a center rear reflection has been received. Likewise a direct impulse from overhead will only produce a time delay in the vertical pair of microphones and a reflection from the side will only show delay in the left-to-right pair. No matter what angle a reflection arrives from, its amplitude and direction can be computed and stored.

In a real concert hall many reflections may be arriving simultaneously, so how did the gentlemen from Japan sort them out? First, each reflection of say a particular impulse generates a signal in all six microphones. All six signals, attributable to a single source, will have essentially the same peak amplitude since the microphones are so close together. Thus any unequal peaks indicate a collision of two or more reflections. Second, the times it takes for a sound to go from one microphone of a pair past the mounting pole to the other microphone of the same pair are identical for all the pairs. Thus all three-microphone pairs should record peaks that are symmetrical in time about the same origin, but with three different spacings depending on the angle of travel. Thus unequal delay to and from the origin indicates an impulse collision. Finally, the ratios of these three delays define the angle to the reflection source, and it happens that for such an orthogonal array, the sum of the three cosines squared of the angle to the impulse source to each axis will add up to one. These three characteristics of the impulses detected by the microphone array represent three simultaneous equations which, when solved, allow a computer to distinguish between two or even three simultaneous or very closely arriving reflections. Since this measuring technique is relatively portable, the JVC team was able to make accurate measurements of halls like the large and small Concertgebouw of Amsterdam, the Alte Oper in Frankfurt, the Beethovenhalle in Bonn, the Philharmoniehalle in Munich, the Staatsoper in Vienna and the Koln Cathedral.

Unfortunately all this brilliant pioneering effort was abruptly subverted when surround sound video systems became the preoccupation of the Japanese establishment. However, JVC did make a few hundred convolvers before the ax fell and these proved that every recording engineer could and should have such an array and PC at any recording session. The engineer could then pick the best listening seat for the array, measure the hall response and later, enter the stored results directly onto a CD or DVD for later loading into the home ambience convolver, probably a PC of some type with lots of DSP power. See Chapter 9 for a discussion of Ambiophones and Ambiophonic recording suggestions.

Sony Decides a Convolver Is Essential If Surround SACD Is to Flourish

Both the DVD-A and Sony's competing format, SACD, are very high resolution, music only, formats seemingly attractive to only the high-end audiophile market. With the addition of multichannel surround capability, however, a wider, more lucrative, audience could be found for these video-less technologies. It is thus clear to everybody in the industry that the future of both systems depends on being able to provide music in a multichannel surround/ambient format. Apparently, Sony decided that unless they provided a means for the industry to make surround music recordings with the same high quality as the SACD disc itself that their investment would be lost. Their problem remained, however, as indicated earlier, that no one knows how to make music surround recordings using microphones that sound realistic or pleasing enough to attract a mass market or even a niche audiophile market segment. The

DVD-A group always assumed that Ambisonics would fill this requirement but Sony decided on a more realistic solution: the Sampling Digital Reverberator, DRE-S777.

A Rose by Any Other Name Is Still a Convolver

The Sony DRE-S777 was not made for home audiophile consumers. It was not a sampling digital reverberator; it was a stored hall convolver. Sony Electronics, Inc., Broadcast and Professional Company made it, for professional recording engineers. It was not user friendly. It was Sony's position, that hall convolution should be the province of the SACD producer and not the home listener. The idea was that the recording engineer should just make the best two channel stereo recording he could and then fabricate as many surround channels as he felt was desirable using the stored halls in the DRE-S777. Superficially, this seems like a good idea. It spares the recording engineer the onerous and expensive burden of placing ambient microphones in a hall about which he or she knows very little and for which there is no basis in the mathematics of acoustics for doing so. With a DRE-S777, after the session is over, the producer can go back to his studio and try out different hall ambience combinations and generate as many surround channels as the standard will allow.

Since, as of this writing, the only market is for 5.1 speaker arrangements, he is unlikely to configure more than two channels for surround speakers. Of course, if you don't like the hall the producer has picked or you want more than two surround ambience speakers, Sony was not interested in your problems. The advantage of having the convolver under listener, rather than engineer control, is that since the producer doesn't have to waste DVD/SACD bandwidth on ambience, he can provide direct sound for additional rear and side speakers where the composer has sanctioned such a practice. Indeed such rear or side direct sound channels can share in the ambience of the front stage since the convolver can easily accommodate such an option. The DRE-S777 was priced at five figures and so was not affordable by most home listeners. Sony produced four CD-ROMS containing the impulse responses of great halls and churches in Europe, Japan and America. One DRE could output four surround channels in real time. That is, it could convolve the left input to produce two ambient surround signals and the right channel to produce yet two more different ambient surround signals. Sony used digital signal processing chips that could process 256,000 events in the life of each input music sample. This was long enough to handle the reverberation of even the largest cathedrals.

Four surround channels are nice but eight or more is even better. The sound from four DRE-S777 reproducing a symphony orchestra embraced in the ambience of the Konzerthaus, Berlin, via 16 surround speakers is overwhelming. But now with modern PC processors a single PC can convolve ambience signals for sixteen or more surround speakers.

Sony's Impulse Response Measuring Method

The usual way to measure impulse responses is to put a relatively small group of microphones at a desirable location and then aim pulses at them from various positions on the stage. In Ambisonics a coincident microphone with one omnidirectional microphone and three collocated figure eight microphones is used. However, the extraction of the ambient data using the Ambisonic approach is quite difficult compared to the six-microphone method used by JVC described above. The SONY approach was clearly related to the fact that it was the professional recording division that had been involved in this development. A preoccupation with the necessities of the 5.1 speaker arrangement was clearly in evidence. Sony used up to ten fairly widely spaced microphones to record impulse test patterns from left, right and center stage speakers. Not all ten microphones were used in every hall but when all are present five of the microphones are omnidirectional and five are Cardioids. The omnis form a rectangle with one of their number in the center. The rectangles vary with the hall but are typically 18 feet wide by 15 feet deep. While there does not seem to be any mathematical foundation for this arrangement, one can put surround speakers at the same positions around the home listening position

and the loudspeaker will output the same ambience toward the listener from this location that the microphone picked up.

Unfortunately, a lot of directional information is theoretically lost using this technique compared to the JVC method. For example an early reflection coming from the rear in the hall will be aimed to the listening position from the rear/side 45-degree direction. Perhaps for this reason there is another rectangular array of five cardioid microphones. Cardioids are directional to the extent that they pick up mostly from the half sphere they face. The cardioids are arranged in a 5.1 rectangular pattern with three in front and two at the rear side corners. The cardioids are aimed at the four corners of the halls with the fifth one pointing directly front. They form a rectangle about 9 feet wide by 4 feet front to back. In this case, if a speaker is placed at say 45 degrees to the left side of the home listener and is fed the ambience picked up by this microphone from the left front hall arc, the directionality of the ambient field will be reasonably accurate. For best results one should convolve this ambience response with both the right and the left stage signals and perhaps use two speakers at the left front location or mix them together if this is more convenient. It is not clear to me why the spacing between the microphones used was so large. This appears to be a habit related to the way recording engineers have been trying to record ambience in the last few years. But with ten microphone locations to choose from it is hard to go too far wrong especially if you can afford to use all of them. Variety is the spice of ambience where concert halls are concerned. One can also argue that the ambient field at the center of either of these rectangles is probably not much different from the field near the edges since the halls are so large in comparison to the rectangles.

Some Noise Is Good Noise

When you are in a concert hall or church and the music stops, you are still in a concert hall. Even with your eyes closed you can sense a sort of ambient ambience, a murmur or acoustic dither that even without an audience present tells you what kind of acoustic space you occupy. By contrast, in the Ambiophonic hall, when the music stops you are abruptly transported from a lively exciting space to a rather dead, sounding listening room. For CDs with many silent bands between short selections, this effect can be somewhat disconcerting. Perhaps in the future, recording engineers will avoid such quiet periods.

Chapter 8

Ambipoles and Ambiophones

Ambiophonics combines several technologies to produce realistic sound fields and actually does it optimally via two-channel recording media where most classical/jazz/pop music is concerned. The technologies are convolution for hall ambience, speaker correction, front loudspeaker crosstalk and pinna angle error elimination, and an optional but superior recording microphone design and placement. The basic goal of Ambiophonics is to recreate at the home listening position an exact replica of the original concert hall sound field. Ambiophonics does this by transporting the sound sources, the stage, and hall ambience to the listening room. In other words, Ambiophonics delivers an externalized binaural effect, using, as in the binaural case, just two recorded channels but with two front-stage-reproducing loudspeakers and eight or so ambience loudspeakers in place of earphones. Ambiophonics generates stage image widths of up to 180 degrees with an accuracy and realism that far exceeds that of any other 2 channel or multi-channel recording/reproducing scheme. We will now discuss how to reproduce the front stage of a two channel recording without exposing our ears to comb filtering, phantom imaging or

major errors in the angle of sound incidence on the pinna and how best to make recordings that take advantage of Ambiophonic binaural reproduction technology.

Making Good on the Promise of Binaural Technology

Since we have only two ears, it seems reasonable that only two signals should need to be recorded. Indeed it was Blumlein's original idea that he could externalize the earphone binaural effect using spaced loudspeakers and some novel microphone arrangements. But once you give up earphones for stereo loudspeakers, the interaural-crosstalk and the arbitrary speaker angle destroy the almost perfect, but internalized (within the skull), binaural frontal stage image and with all the stereo hall ambience now coming entirely from the front, the hall ambience sounds unnatural. Binaural theory says that if you sit in the concert hall with small microphones in your ear canal, record the concert, and then later play it back with in-the-ear canal earphones you will experience an almost perfect "you are there" recreation. The only flaw in this method would be that when you moved your head, while listening or recording, the reproduced stage would rotate unrealistically.

But let us consider, briefly, why this recording method can otherwise produce an awesome reality. First of all, the sound from the stage and the hall during such a personal binaural recording reaches your ear canal (and the imbedded microphones) after being filtered by your pinna and your head shape. Since the playback earphones we are using are an in-the-ear-canal type the sound only passes through the pinna or around the head once. Also the pinna used to make the recording are your own, not those on some dummy head carved in wood or plastic. The two channels are kept separate throughout and the left ear playback earphone signal never leaks into the right ear or vice-versa. Thus we can state one of the basic rules of realistic binaural recording technology. In any binaural recording or reproduction chain there should be one and only one pinna function and it must be your own. There must also be one and only one head shadowing entity but in this case whose head it is is not critical. That the head shadowing function is not as individual as the pinna function can be understood when one realizes that sound passes around the head over the top, under the chin, around the back, and varies as the head is tilted or rotated. Thus the brain is not overly sensitive to the exact shape of a particular head or the exact frequency response of the head shadowing function, within reason.

So let us see how we can make use of this knowledge. Let us assume that we have a two-channel recording made using a dummy head that has no pinna. This dual microphone is sitting fifth row center. Its signals are then recorded and played back over two loudspeakers directly in front of the home listener. Let us assume for the moment that these loudspeakers are like laser beams so that their sound is aimed precisely at the proper ear. In this case the listener hears what the corresponding microphone hears and the sound impacts his own pinna with very little incident angle error for central stage sources. For stage sources that are more to the side, the listener hears the head response transfer function of the microphone head and for most humans this is quite realistic. But now the home listener can rotate his head and the image is stable just as if he were in the concert hall. So this technique is not only equal to but superior to the earphone method considered above. There is a pinna angle error for stage sources toward the extreme left and right but fortunately these are the angles where direct sound has a more or less clear shot at getting to the ear canal directly without extreme pinna filtering and also where nature has compensated for the decrease in pinna sensitivity by making the interaural head shadowing most pronounced providing strong and natural horizontal plane localization. In practice, both IMAX and Ambiophonics easily demonstrate that this binaural technology is exceptionally realistic and does produce wide front stages that even allow the cocktail party effect to be in evidence.

Ambidipoles

Now the question is how to make a pair of center front speakers behave like sound lasers. There are two possibilities. One is to put a physical wall or panel in front of the listener. This wall extends to within a

foot or so of the listener's head and keeps the left speaker from radiating to the right ear and vice-versa. This technique works perfectly and if you are an audiophile and want absolute fidelity without cables or extra processing this is a very inexpensive way to go. You can try it first with a mattress on end, if you want to experiment and have some fun. While I appreciate that the use of a barrier will never find universal acceptance, an understanding of how it works is necessary to an appreciation of what a software version of such a crosstalk avoidance system should accomplish. You can make a barrier out of sound absorbing panels with a cutout at the end of it so that it is possible to sit comfortably at the end of it. The thickness of the barrier is not critical, but should be about six to eight inches wide so that when a listener is seated their right eye cannot see the left speaker and vice versa. The wall extending back toward the space between the speakers is, preferably, made with sound absorbing material. This panel can be thought of as a collimator for most sound except the low bass. It eliminates all stray rays from the right that might be heading left and those from the left that might be heading right. A panel such as this is very effective in dampening higher frequency room reflections since it absorbs rays coming from both room sides. The use of an outdoor reflective barrier to eliminate stereophonic crosstalk was described in 1986 by Timothy Bock and Don Keele Jr. at the 81st Audio Engineering Society Convention. While Ambiophonics uses an absorbent barrier, their results are still largely pertinent. They determined that a listener could be further back from the end of the barrier if the barrier was wider, the speakers closer together, and the listener further from the speakers. Stated as an equation:

$$L=X(H+T)D$$

Where, in inches, L is the maximum distance a listener's head can be from the barrier, X is the distance from the listening end of the barrier to the position of the speakers, D is the distance between the centers of the speakers H is the distance between the ears, and T is the thickness of the barrier. For a worst case scenario of a six-inch head, a six-inch thick barrier, an eight-foot distance to the speakers, and a speaker separation of three feet (too much) a listener could be as much as 32 inches, almost three feet from the end of the barrier. Thus the use of a barrier does not in any way make listening uncomfortable or claustrophobic. Our own Ambiophonic barrier geometry allows one to be four feet from the end of the barrier, but at the far end of this range one's head must be more precisely centered. With a four-foot space, two in-line listeners can enjoy the enhanced angular image separation at the same time and indeed the front listener acts as a continuation of the barrier for the second listener. If in doubt about the spacing, the eyeball method is very conservative. As long as no part of the opposite loudspeaker is visible from one eye, excellent separation is guaranteed. Sitting too close to the barrier is not only unpleasant but results in a loss of high-frequency response if the barrier is as wide as the head and absorptive.

However, the mainstream Ambiophonic way is to use software and a computer or digital signal processing component to eliminate the crosstalk. I call a pair of speakers that use the public domain software that we have developed to do this, an Ambiodipole. First, although most speakers can be used to form an Ambiodipole, it is best if the speakers chosen are very directional and well matched. A slightly concave electrostatic panel (called an Ambiostat) can actually focus sound well enough that it almost behaves like the laser we have hypothesized. Obviously, if the speakers are focused and time aligned, the software can do its job much better. What the recursive crosstalk cancelling software does is generate slightly delayed reversed polarity signals for the speakers to cancel the crosstalk acoustically before it reaches the ear canal. The cancellation is an infinite series process since the crosstalk caused by the cancellation signal also produces crosstalk, which must then be cancelled and so on. If the Ambipoles were widely spaced, then the crosstalk would have to go around the head and the correction signals would be very difficult to calculate since they would be affected by head position and pinna shape. Thus the front speaker pair should be closer together with about 20 degrees between them so that both the main front speakers emit directly to their outside ears.

Just as it is obvious that a barrier will work better with close together speakers, since speaker proximity makes it easier for the barrier to shadow the appropriate ear, so crosstalk software works better if the speakers are closer together. Ambiodipoles do have a sweet spot limitation although in my experience the sweet spot is larger than that of most well focused stereo or 5.1 systems. In theory if the Ambiodipoles, used to form an Ambiodipole, are constructed as panels with a special curved shape then it is possible to enlarge the sweet spot enough to accommodate two or even three listeners. But such a speaker has yet to be constructed.

In stereo if you move back along the median line between the speakers, the stage narrows and becomes mono. If you move forward, you get a hole in the middle and just hear two speakers, one on either side. If you move offside you normally localize to one speaker and so hear mostly just one channel. Similar effects plague 5.1 which is why a center speaker is used to keep the dialog clearly audible. In Ambiphonics, if one moves very close to the speakers, one hears normal stereo instead of Ambio. If one moves back until one hits the rear wall nothing much happens. One can recline, stand, nod, rotate the head, etc. without ill effect. If one moves sideways, one still hears both channels, clearly so a center speaker is never required. Basically, Ambio has a larger listening area than stereo, but when one is not centered one feels deprived in a way not apparent in stereo. In PanAmbio versus 5.1 there is a similar advantage for PanAmbio, in that off center viewers cannot localize to a surround speaker. Also the front and rear stages or sound effects are clearly audible in PanAmbio no matter where you sit.

Ambiodipole Software

Over the years many versions of crosstalk cancelling software and hardware have been promulgated. Among those best known are hardware devices from Lexicon and Carver (Sonic Holography) and software programs from The University of Southampton (the Stereodipole) and The University of Parma. In general all these early attempts had serious flaws that made them unrealistic, phasey, or unstable. Among the flaws was trying to do crosstalk cancellation using speakers still arranged in the stereo triangle. This is doomed to failure because now the amount of crosstalk depends on what happens as the sound from each speaker crosses the head generating the crosstalk that one must cancel. With a wide speaker angle the attitude and shape of the head will change the crosstalk making it difficult to know what the crosstalk actually is. With the speakers close together there is virtually no significant change in the crosstalk as long as the head is between the speakers. Some such systems used an average HRTF (head response transfer functions) to compensate for the head shadow but this almost never works since nobody is average. Other pioneers put the speakers close together but still used HRTFs mostly to get the proximity effect of a bee buzzing close to the ear. But again, in general, the use of HRTFs is counterproductive and not necessary for normal music or movie sound reproduction.

But the most basic flaw in the early crosstalk cancellation methods was that they were not recursive. That is when you cancel crosstalk you must also cancel the crosstalk due to the signal that cancelled the original crosstalk and this process must be continued to inaudibility. As far as is known to this author, the Ambiphonic program known as RACE (Recursive Ambiphonic Crosstalk Eliminator) was the first fully recursive XTC program to run both in PCs and hi-fi components.

The Ambiphonic Institute's published RACE equations were used by Robin Miller, of Filmmaker, and Angelo Farina of The University of Parma to develop a version of RACE that could be used in programs like AudioMulch or in VST plugins to drive speaker pairs called Ambiodipoles. This new software is designed so that almost all two-channel recordings can benefit from being reproduced Ambiphonically. RACE includes adjustments so that one can select one that makes a particular recording sound most realistic. It is hoped that those reading this book will be able to purchase either Ambiphonic system

components or PCs that can run this software as well as the software for hall convolution and speaker correction.

Bass Localization

Since Ambiophonics is a binaural based system, it does not provide the Blumlein loudspeaker crosstalk signal that amplifies low frequency ILD cues for those recordings made with a coincident microphone arrangement such as the Soundfield or crossed figure eight mics in the M/S (mid-side Blumlein configuration). (See the Appendix A for a detailed analysis of the Blumlein patent and technology.) However, it should be understood that at very low bass frequencies, RACE loses its effectiveness allowing increasing crosstalk as the frequency declines and therefore amplifying LF phase cues for coincident microphone recordings. This is basically a non-issue. Remember that the ear's ability to localize bass frequencies at 80 Hz and below is virtually non-existent. The pinna certainly has no capability in this frequency range and the head is too small to attenuate signals with wavelengths measured in tens of feet. Thus the only localization method available to the brain at very low frequencies is the few degrees of phase shift between the ears. There is no evidence that the brain can detect such small phase shifts and thus worrying about crosstalk elimination at very low frequencies to improve front stage imaging is not productive. Also, at very low frequencies the power required to produce crosstalk cancellation becomes excessive and since it is not necessary RACE automatically avoids low bass crosstalk cancellation.

The Ambiophone

Once we know that playback will be Ambiophonic, the question arises as to whether there is an ideal recording method that can take advantage of the fact that surround ambience will be derived via convolution, that the Ambidipole will eliminate crosstalk and avoid phantom imaging. But I still want to emphasize that although Ambiophone microphone arrangements can make the Ambiophonic approach to realism even more effective, Ambiophonics works quite well with most of the microphone setups used in classical music or audiophile caliber jazz recordings. One can heighten the accuracy, if not gild the lily of realism, of an Ambiophonic reproduction system by taking advantage, in the microphone arrangement, of the knowledge that in playback, the rear/side half of the hall ambience is convolved, that a stage can go out to 180 degrees.

Earlier we considered the binaural model where microphones are inserted in the ear canal of an ideally situated listener. But now the situation is different. We are going to reproduce the hall ambience by convolution so we do not want our binaural listener to pick up any hall ambience from the rear, the extreme sides, or the ceiling. So let us put sound absorbing material just behind his head and above him as well so that he has a sonic view of only the stage in front of him. Now we know that upon reproduction the Ambidipole speaker sound will pass by his pinna on the way to the eardrum. Thus we do not want any pinna at the recording site. Thus the human listener is excused from the recording site and we are left with a pair of baffled head spaced omni or cardioid microphones sitting at the best seat in the house. But the rule stated earlier said there must be at least one and only one head shadow in the recording/reproduction chain and so, since the home listener is directly in front of the Ambidipole, it is up to the Ambiophone to provide a head shadow. So let us put a head shaped oval between the two microphones at this best seat in the house. So our Ambiophone boils down to an oval shaped two capsule assembly baffled to the rear and above comfortably ensconced at the best seat in the house or studio.

Nothing New Under the Sun

After completing the above derivation of the ideal Ambiophone, I began to search for recordings that played back realistically Ambiophonically to see if they had anything consistent or unusual about them.

Not being a recording engineer or a microphone aficionado, it took me awhile to notice that many of the best CDs in my collection were made with something called a Schoeps KFM-6. A picture of this microphone in a PGM Recordings promotional flyer showed a head sized but spherical ball with two omnidirectional microphones one recessed on each side of the ball where ear canals would be if we had an exactly round head. The PGM flyer also included a reference to a paper by Guenther Theile describing the microphone, entitled On the Naturalness of Two-Channel Stereo Sound, J. Audio Eng. Soc., Vol. 39, No. 10, 1991 OCT. Although Theile would probably object to my characterization of his microphone, his design is essentially a simplified dummy head without external ears. He states, "It is found that simulation of depth and space are lacking when coincident microphone and panpot techniques are applied. To obtain optimum simulation of spatial perspective it is important for two loudspeaker signals to have interaural correlation that is as natural as possible.....Music recordings confirm that the sphere microphone combines favorable imaging characteristics with regard to spatial perspective accuracy of localization and sound color....." Later he states "The coincident microphone signal, which does not provide any head-specific interaural signal differences, fails not only in generating a head-referred presentation of the authentic spatial impression and depth, but also in generating a loudspeaker-referred simulation of the spatial impression and depth.....it is important that, as far as possible, the two loudspeaker signals contain natural interaural attributes rather than the resultant listener's ear signals in the playback room."

What Theile did not appreciate is that, for signals coming from the side, the sphere acts as sort of filter for the shorter wavelengths just as the head does. When this side sound comes from side stereo speakers the listener's head again acts as a filter resulting in HRTF squared. The solution, of course, is to use the software Ambidipole and listen to the Theile sphere without the second head response function. Theile also "generates artificial reflections and reverberation from spot-microphone signals." He uses the word artificial in the sense that the spot microphone signals will be coming from the front stereo loudspeakers instead of from the rear, the sides, or overhead. While Theile's results rest as much on empirical subjective opinion as they do on psychoacoustic precepts, they certainly are consistent with the premises of Ambiophonics both in recording and reproduction. Making new recordings using the Schoeps KFM-6 version of the Theile Sphere and evaluating existing recordings made with this microphone show that the theory is correct since such recordings yield exceptionally realistic front stages with normal concert-hall perspectives and proscenium ambience.

Realistic Reproduction of Depth and Perspective

It is axiomatic that a realistic music reproduction system should render depth as accurately as possible. Fortunately, front stage distance cues are easier to record and/or recreate realistically than most other parameters of the concert-hall sound field. Assuming that the recording microphones are placed at a reasonable distance from the front of the stage, then the high frequency roll-off due to distance and the general attenuation of sound with distance remain viable distance cues in the recording. Depth of discrete stage sound sources is, however, more strongly evidenced in concert-halls by the amplitude and delay of the early reflections and the ear finds it easier to sense this depth if there is a diversity of such reflections. The Ambiophonic crosstalk cancellation feature also enhances depth perception since depth perception of close by sources is enhanced when the range of ILD and ITD is greater.

In Ambiophonics, convolved early reflections from the surround speakers make the stage as a whole seem more interesting, but it is only the recorded early reflections coming from the front speakers that provide the reflections that allow depth differentiation between individual instruments. This is why anechoic recordings sound so flat when played back stereophonically or even Ambiophonically, despite the presence of an added ambient field. In ordinary stereo, depth perception will suffer if early side and rear hall reflections wrap around to the front speakers or in the anechoic case, are completely missing. Since it is easy to make Ambiophonic recordings that include just proscenium ambience, why not do so and save on convolver processing power and preserve, undistorted, the depth perception cues?

There remains the issue of perspective, however. When making a live performance recording of an opera or a symphony orchestra the recording microphones are likely to be far enough away from the sound sources to produce an image at home that is not so close as to be claustrophobic. There are many recordings, however, that produce a sense of being at or just behind the conductor's podium. This effect does not necessarily impact realism but you must like to sit in the front row to be comfortable with this perspective. Turning down the volume and adding ambience can compensate for this, but with a loss in realism. This problem becomes more serious in the case of solo piano recordings or small Jazz combos. For example, if a microphone pair is placed three feet from an eight foot piano, then that piano is going to be an overwhelming close-up presence in the listening room and a "They-Are-Here" instead of a "You Are There" effect is unavoidable. This will be very realistic especially with the Ambiodipole, but adding real hall ambience doesn't help much since the direct sound is so overwhelming. The major problem with this type of recording is that you have to like having these people so close in a small home listening room. You may notice that demonstrators of high resolution playback systems in show rooms or at shows, overwhelmingly, use small ensemble, solo guitar, single vocalist etc., close mic'ed, recordings to demonstrate the lifelike qualities of their products and that these demonstrations are mostly of the "They Are Here" variety.

These depth and perspective problems are easily solved by simply placing an Ambiophone at a seat that has a reasonable view of the performers.

Chapter 9

Surround Ambiophonic Recording and Reproduction

Note: Chapter 9 is in the form of an Audio Engineering Society paper. It may reiterate some of the points made in early chapters, but in a different context.

Abstract

Ambiophonics, Panorambiophonics, and Periambiophonics are related surround sound paradigms that reliably deliver up to full 360-degree spherical localization for both direct and ambient sound via two, four, or six DVD/SACD/MLP/DTS/Dolby/ADAT/etc. coding/media channels. They reproduce old or new, standard, 2, 4, 6, or ITU 5.1-channel music discs with unprecedented spatial realism and binaural-like localization accuracy via direct sound radiating front/rear/overhead stage-producing Ambiodipoles and virtually any desired number of ambience surround speakers. Alternatively, superior acoustic recordings can be made using the described Ambiophones (or using convolvers, if fabricated) to capture images of startling depth and presence for music in the round, 3D movie sound tracks, virtual reality, or electronic music soundscapes. Six-channel Periambiophonics adds elevated direct sound to the fully spherical hall ambience vectors already provided by basic Ambiophonics which drives essentially any number of hall ambience speakers regardless of their positions. All the versions of surround Ambiophonics easily deliver a you-are-there, psychoacoustically correct, home listening experience, via home theater media, albeit best limited to just a few listeners.

Introduction

Ambiophonics is a comprehensive sound recording/reproduction methodology, that like or unlike Stereophonics, Ambisonics, THX 5.1, or Wavefield Synthesis, prescribes hardware/software that scrupulously insures that the well known tenets of human binaural hearing (see Appendix B) are

rigorously catered to so as to achieve psychoacoustic and physiological verisimilitude for a normal group of home listeners/viewers who seek and value "you-are-there" realism. Ambiophonics combines crosstalk-free speaker pairs (Ambiodipoles), surround speaker ambience derived from measured hall impulse responses via a convolver (Ambiovolver) and room/speaker correction/treatment to generate a binaurally correct sound field similar to wavefield synthesis. Ambiophonics creates a concert hall stage and hall from just two media channels as found on CDs, MP3s, LPs, etc. feeding a single Ambiodipole. Panorambiophonics requires four media channels as provided by multichannel DVDs or SACDs each pair feeding its own Ambiodipole. Periambiophonics uses six media channels as in DVD-A, DTS-EX, etc. feeding three Ambiodipoles. In each type of system additional hall ambience surround speakers may also be driven via a single Ambiovolver and this is strongly recommended where music is concerned.

A single Ambiodipole in front easily produces a stage of more than 160-degrees in width. A single Ambiodipole to the rear of the listener produces a similar rear stage width. A remarkable property of the Ambiodipole software (RACE) we have developed is that when both front and rear Ambiodipoles are working together, they blend and the front and rear stages widen to a full 180-degrees. Thus, 360-degrees of horizontal localization becomes easily attainable for recordings made with Ambiphones or synthesized. A third or even more Ambiodipoles can be elevated over the front and/or rear Ambiodipoles to add full width stages high in the air and again there is vertical fill between the stages although the extent of this phenomenon has yet to be fully investigated. The most basic Ambiophonic system is meant to allow previously recorded two channel media such as CDs, MP3s, and LPs to be reproduced without the well known limitations of the traditional 60-degree stereo triangle (see Appendix B), to deliver an uncompromised full width direct sound stage from two center-front speakers (an Ambiodipole) and to provide real diffuse but still directional hall ambience to almost any number or location of surround speakers including elevated speakers.

It became obvious in the early development of Ambiophonics that existing stereo microphone techniques could be revised to produce better two channel recordings. Thus, on the recording side, the Ambiphone, a novel, baffled microphone arrangement, takes advantage, when recording, of the knowledge that the playback will be Ambiophonic. (not via the stereo equilateral triangle although Ambiphone recordings are actually backward compatible and sound quite normal in standard stereo) The Ambiphone also assumes that both the amplitude and the directional attributes of the early reflections and reverberant tails of the hall will be properly directed to the appropriate frontal Ambiodipole and surround speakers. (Indeed, this is possible even in the case of non-Ambiphone recordings if the recorded or added reverb, unfortunately mixed into the direct frontal sound, is not too intrusive.)

After a brief review of the basics, this paper is devoted to advanced versions of Ambiophonics which take into account the 5.1, 6.0, 7.1, Dolby/THX and DTS coding/media/speaker arrangements. Standard 5.1 discs may also be played Panorambiophonically (described in detail below) in a manner analogous to the Ambiophonic playback of ostensibly stereo CDs or LPs, but, in this case, using front and rear Ambiodipoles and surround ambience speakers driven by a hall impulse response Ambiovolver. Most 5.1 movie and music DVDs or music SACDs reproduce exceptionally well this way especially when compared with the ITU 5.1 standard speaker arrangement. Panorambiophonics, described below, uses four channel coding/media such as Dolby, DTS, SACD, or DVD-A to deliver an easily localizable 360-degree direct sound stage as in movies, or, for concerts, a very wide front stage that, if in a hall, automatically includes horizontal 360 degree hall ambience. A four channel recording mic, the Panorambiophone, has been designed to make such recordings. Only four speakers (two Ambiodipoles) are used in Panorambiophonic reproduction to reproduce all horizontal plane direct sound and horizontal hall ambience with full circle normal binaural physiology localization. Where the direct sound recording has been made in a dry or small studio, it is possible to enhance the reproduction of these front and rear direct sound fields by adding ambience surround speakers driven by an hall Ambiovolver as in standard Ambiophonics.

Periambiophonics adds a third elevated Ambiopole to Panorambiophonics to provide for a full direct sound stage in all dimensions including some height. The elevated Ambiodipole can be used for direct sound reproduction or ambience. In the latter case this allows a concert-hall direct sound performances to be recreated in a home with just three speaker pairs and no surrounds. Using three direct sound Ambiodipoles allows movies, virtual reality, games and soundscapes to sound more like the live experience. Furthermore, Periambiophonics can combine six-channel Periambiophone recording, and the front, rear, and elevated Ambiodipoles, with an Ambiovolver to add virtually any desired number of surround speakers so as to deliver physiological verisimilitude of a concert hall experience that also includes rear or overhead direct sound sources to a home listener via standard DVD/SACD media. Clearly, both Panorambiophonics and Periambiophonics are well suited to capture, create and reproduce 3D electronic music or virtual reality projects. This paper reviews the theory, techniques, and features, of the hardware and software required to make these various kinds of Pan/Peri/Ambiophonic recordings and to reproduce these as well as stereo CDs and the various multichannel surround media.

Review of Basic Ambiophonics (Fig. 1)

The simplest form of Ambiophonics is meant for the playback of ordinary stereo CDs, LPs, SACDs, MP3s, cassettes, stereo TV, etc. In stereo, the front stage is created between the speakers, in Ambiophonics the stage is created from the speakers outward and so can be much wider. The Ambiodipole speaker pair form an angle to the listener of from twenty to thirty degrees. Listeners can sit anywhere along the line between the speakers and can stand or recline, turn their heads, lean etc.

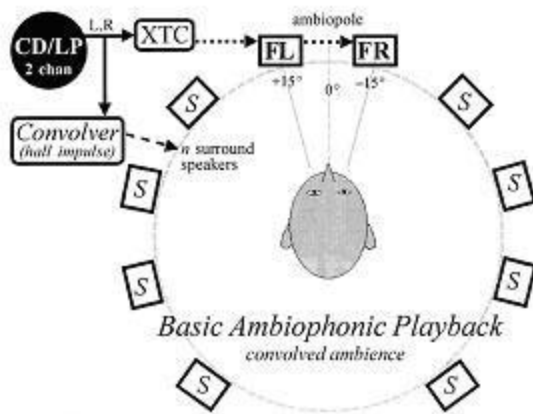


Figure 1.

The scattered surround speakers are fed hall ambience signals calculated for both the left and right channels by a computer which we call an Ambiovolver. The Ambiovolver has stored within it the impulse responses of some of the great halls, churches, and auditoriums of the world and more such hall signatures are being accumulated all the time. One simply selects the hall best suited to the recording or the actual hall where the recording was made. The Ambiovolver is told the location of all the surround speakers in the room and it then generates the appropriate reflections and feeds them to a surround speaker that can then mimic a concert hall wall. In this way the levels, frequency responses, and the directionalities of the reverberant field are maintained. I have driven up to 24 surround speakers this way and, while clearly overkill, the results are gratifying. This is in contrast to normal 5.1 practice where recorded hall ambience whether from front, rear, overhead or the side is lumped together and launched from just two surround speakers. The attached references describe Ambiovolver design, hall impulse response measuring procedures and hall acoustic properties. It is often desirable to keep the listening room early reflection characteristics under control. Absorptive panels are quite effective. However, since

the direct sound speakers are so close together and aimed forward, they are easier to position than for stereo or 5.1. Bad room acoustics are actually less of an issue in Ambiophonics than in stereo. The early reflections and the late reverberant generated by an Ambiovolver normally swamp the listening room acoustics.

Two Channel Ambiophonic Recording

While for many people, with large CD or LP collections, basic Ambiophonics will sound as good as they wish, others will find enjoyment in the improvement that can be achieved by making recordings specifically meant to be played back over Ambiophonic systems. The Ambiophone recording microphone assembly was designed to make this feasible. Basically it is a head shaped ball with two omnidirectional microphones mounted flush where the ear canals would be. The microphone is baffled. That is, it faces forward and is shielded from sound originating from overhead, the rear, or the extreme sides.



Figure 2. An Ambiophone. Two more mics are in the head behind the baffle to pick up the rear stage.

The microphone is placed first to fifth row center depending on taste. The perspective one hears during reproduction is the same as if one were at the mic position during the recording session. The usual considerations of hall radius or ratios of direct to reverberant sound do not apply here since the mic is baffled. Since all hall ambience will be generated from this or other hall impulse responses, it is not necessary to actually record hall reverb during the recording session. The Ambiophone must also collect horizontal frontal or proscenium ambience since this indirect sound should emerge from the Ambiodipole with the direct stage sound. The head shape of the Ambiophone provides the Interaural Level Difference for sounds from the stage sides. The Ambiodipole, being centered in front of the home listener, does not provide this and the rule is there must be a head shadow in the system somewhere. The Ambiophone captures both correct ILD and ITD compared to coincident microphone techniques, spaced omnis, spot mic mixing, etc. The Schoeps KFM-6 turns out to be a good match for an Ambiophone, if baffled during use.

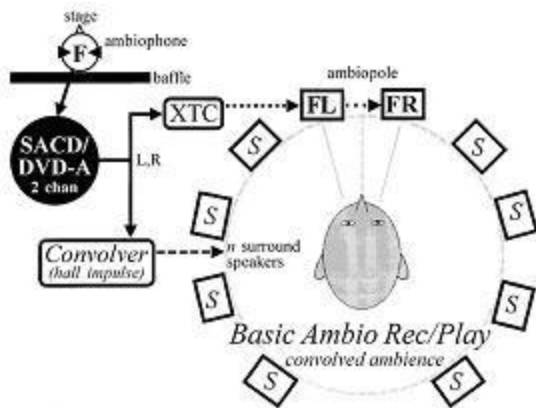


Figure 3. Two Channel Ambiophonic System

Where classical music, reproduced in the home environment is concerned, two channel Ambiophonic recording and reproduction should satisfy even the most golden-eared audiophiles. Ambiophonics appears to entirely swamp the digital sampling differences between 2 channel media such as CD, SACD and DVD-A. It would be an interesting double blind project to see if the different media can actually be distinguished when Ambiophonic conditions prevail during both recording and playback.

Ambio Playback of 5.1 Discs (Fig. 4)

Home theater surround movies or music recordings can be played back Ambiophonically rather than stereophonically in a manner analogous to the playback of CDs and LPs, discussed above. The psychoacoustic disadvantages of the LCR reproduction scheme are reviewed in Appendix B. The left and right frontal 5.1 channels can be fed directly to a crosstalk canceller and thence to an Ambidipole. An Ambidipole also functions as a center speaker it is an easy matter to tell the player to mix the center channel signal into the left and right frontal pair. In Ambiophonics, a center speaker is never required and is even detrimental. Certainly, it is easier to set up the front part of a home theater system using just two center speakers 20 or 30 degrees apart so as not to stand in front of the TV screen than setting up three speakers that must be equidistant and spaced symmetrically. Also, for home TV, viewers like to be centered so the major supposed advantage of the LCR arrangement seems of limited value.

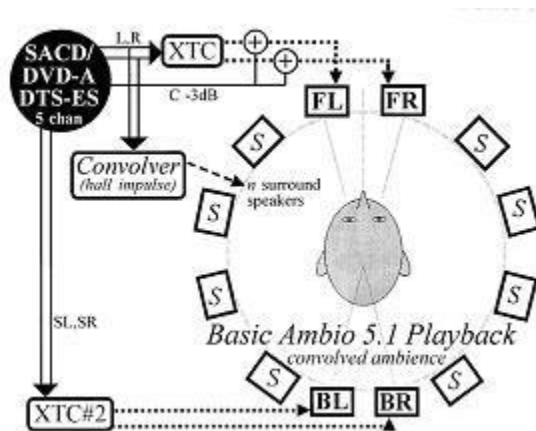


Figure 4. Ambiphonic Playback of 5.1 Material

The two rear surround channels go to a second crosstalk canceller and a rear Ambiodipole. For many movies this arrangement produces a rear stage with excellent localization up to 180 degrees in the rear. Of course, this applies only to movies that were recorded in stereo in the rear, not just dual mono or fabricated sound effects, or ambience. Music DVDs often include real hall ambience captured during the performance and again, if not mono, can provide an ambient field spread across the rear. While not ideal, since ceiling rear and frontal ambience comes from this horizontal rear arc, this effect is better than the standard ITU plus/minus 110-degree arrangement whose properties are discussed in the Appendix. Better yet, for much 5.1 music, where there are no instruments or vocals in the rear, the rear surround channels can well be ignored and the more natural ambience generated by the Ambiovolver can be used in its place now free from the constraints of the ITU surround speaker position mandate. It is also possible to use both the Ambiovolver and the rear Ambiodipole simultaneously. For those who want applause coming from the rear, this arrangement works well for both live music and movie sound effects.

Panorambiophonics

If a four channel medium such as SACD, DVD, or DTS-CD is available, then it is possible to record direct sound sources over 360-degrees in the horizontal plane. A special microphone assembly which we have called a Panorambiophone (or Panambiophone and Panambio for short) is used to capture signals appropriate for reproduction via one front Ambiodipole and one rear Ambiodipole. Front and rear Ambiodipoles do merge seamlessly. We have already demonstrated to hundreds of visitors that the combination of the Panambiophone and the two Ambiodipoles does indeed allow normal binaural localization over the full circle including the 90-degree positions at the extreme sides.

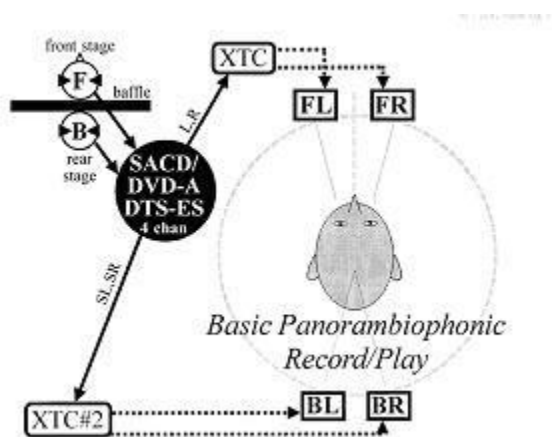


Figure 5. Panorambiophonics Delivers Full Circle Localization via 4 Channels and Speakers

The Panambiophone consists of two Ambiphones placed one behind the other but still both facing in the forward direction. The two head shaped balls must be placed nose to baffle to nape since if there is too much separation the differential delay of slightly off 90-degree direct sound sources will cause comb filtering when reproduced. A baffle between the two Ambiphones insures that the front stage is picked up mainly by the front half head and the rear stage is mainly heard by the rear half head. For concert music both half heads should be protected from overhead reflections. The Panambiophone, like the Ambiphone is placed at the best seat in the house or at the center of the sonic action. In reproduction, two crosstalk cancellers feed two Ambiodipoles and listeners should sit, stand, or recline on the line between them for best effect. Off-line seating still yields interesting front/back localization but the exact angles are unpredictable. During symphonic recording, the rear Ambiphone picks up the rear half

horizontal hall ambience while the front Ambiphone automatically captures the frontal half of direct and ambient sound. Thus, one can have a reasonable you-are-there 360 degree ambience sound experience with just four speakers. This methodology should be compared with the random difficulties encountered using other ambience pickup microphone arrangements such as IRT, Fukada, ORTF, Williams, Decca, etc.

Advanced Panambio

If the Panambiphone is used in the concert hall during a live performance, the ambient field cannot be captured with precision because the overhead ambience is either mixed in with the horizontal components or excluded if the microphone is shielded from the ceiling. Likewise in circular direct sound recordings made in studios or on location, it is difficult to include a realistic ambient field as part of the four channel medium. Thus, there are good reasons to use an impulse response, an Ambiovolver, and surround speakers to enhance the performance of basic Panambio.

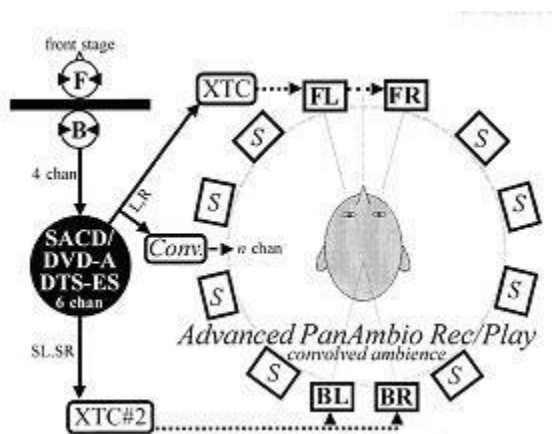


Figure 6. Panorambiophonics Including Both Circular Direct Sound and Hall Ambience

An example of this is Robin Miller's Panorambiophonic recording of an outdoor parade. The bands pass by in front and the crowd's shouts and claps come from behind and from the sides. But when you add impulse response derived reflections from the surrounding office buildings to the mix the scene becomes that much more vivid. Another example is Robin Miller's studio recording of country music with a boisterous audience present. The Ambiodipoles take care of all the direct, front and rear, tunes and shouts but the Ambiovolver transports the whole scene to Nashville.

Periambiophonics

Periambiophonics adds height to the Ambiphonic mix. This requires another pair of media channels but ADAT, DVD-A and DTS-ES are examples of commercially available systems capable of delivering sufficient data to create close to full upper hemisphere soundscapes. Another baffled Ambiphone head is used to capture an elevated front stage and this signal pair needs to be fed to a crosstalk canceller and an elevated Ambiodipole. If one has four pairs of media channels available, then the two front stages and the two rear stages can produce virtually anything desired. To date only frontal and rear elevated stage merging has been tested and Periambiophonics is still a work in progress. The real issue is whether there is a viable commercial home market for such a direct sound technology.

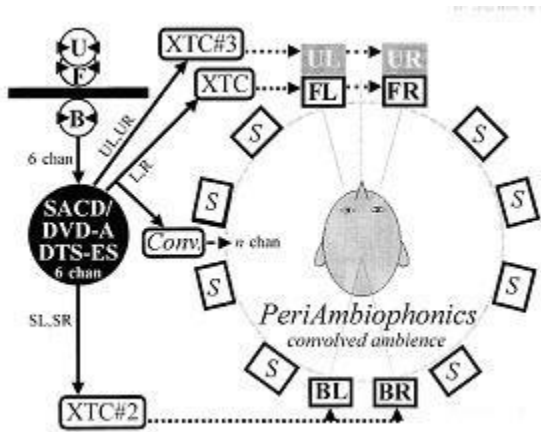


Figure 7. Periambiophonics Provides An Elevated Stage

Figure 7 also shows that convolved ambience can provide periphonic envelopment. As discussed above all the various Ambiophonic methods can employ the Ambiovolver to produce signals for surround speakers at any azimuth or elevation if the impulse response used has been taken in three dimensions.

Ambisonics + Ambiophonics

In the absence of a three ball Ambiophone or impulse response expertise, it is possible to make live recordings using a single Ambiophone facing forward to catch the stage and an Ambisonic WXYZ coincident microphone just behind it to only record hall ambience in all dimensions. The Ambisonic array must be baffled to prevent it from picking up frontal direct sound. Six media channels are also required to store this version of periphonic sound. Instead of an Ambiovolver an Ambisonic decoder is required to deliver the ambience to the surround speakers. Normally, Ambisonic surround speakers must be symmetrically arranged about the listening spot. However, when only ambient sound is being handled Ambisonically, the requirements are less stringent. Also the more speakers used the better the results.

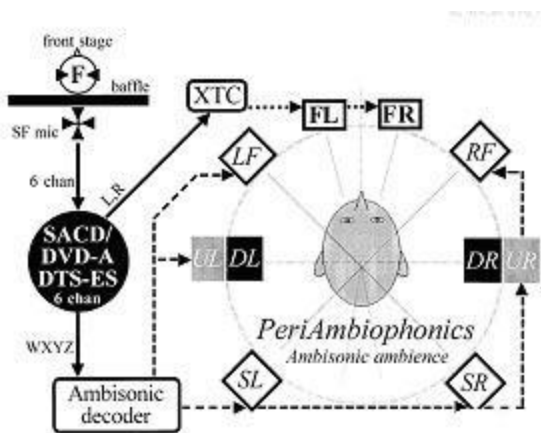


Figure 8. Ambisonic Techniques Can Be Used to Provide Surround Ambience With Height

Since Ambisonic technology and some hardware and software has been available since the 70s, this route may be more attractive to researchers than the Ambiovolver approach. The advantage of the Ambiovolver however, is that hall ambience need not be recorded during the performance. There is also

the complexity of the various Ambisonic four-capsule microphones and control units. Ambisonic techniques are also often used to capture hall impulse responses.

Conclusion

It took 25 years for stereophonics to seriously begin to replace monophonics. It is likely that a similar period will be required for Ambiophonics to replace or at least supplement stereophonics and its twin brother 5.1. But the development of digital signal processors and algorithms able to process digital audio in real time, without audible distortion or noise, has now made it feasible and practical for music/movie lovers to enjoy and recording engineers to deliver greater physiological verisimilitude in music and video recording. Recordings already made with the various varieties of Ambiophones can be demonstrated to all who are interested or doubting. Ambiophonics provides binaural realism and a normal stage perspective when reproduced via one or more Ambipoles. Ambiovolver driven surround speakers easily provide surround ambience without requiring media bandwidth or recording session bother. Ambiophonic recordings should need no spot microphone support, panning algorithms, artificial reflections, or HRTF manipulation and consume just two media channels for classical music or four or six if rear and/or overhead sound stages are desired. Best of all, not only is Ambiophonic reproduction of existing CDs and LPs superior to stereo triangle reproduction but Ambiophonic surround reproduction of 5.1 DVDs and SACD is also psychoacoustically superior and easier to implement than the ITU 5.1 speaker arrangement. However, Ambiophonics is for domesticity and is not suitable for large group listening applications.

Acknowledgements

Major contributions to advancing this technology and making Ambiophonics a living reality have come from Angelo Farina, Anders Torger, Robin Miller, and Enrico Armelloni. They represent Italy, Sweden, and the USA. Their support has been unstinting.

References

- [1] Ralph Glasgal. *Ambiophonics*, 2nd Edition. www.ambiophonics.org
- [2] William B. Snow. *Basic Principles of Stereophonic Sound*. *Stereophonic Techniques*, An AES Anthology 1986.
- [3] James Moir. *Stereophonic Reproduction*. *Stereophonic Techniques*, An AES Anthology 1986.
- [4] Jens Blauert. *Spatial Hearing*, 1997 Edition, MIT Press. www.mitpress.mit.edu.
- [5] Angelo Farina, Enrico Armelloni, Ralph Glasgal. *Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music*. *Proceedings of the AES 19th International Conference*, 2001.
- [6] Ralph Glasgal. *The Ambiophone. Derivation of a Recording Methodology Optimized for Ambiophonic Reproduction*. *Proceedings AES 19th Conference on Surround Sound Techniques*, 2001 www.ambiophonics.org
- [7] Ralph Glasgal. *Ambiophonics, Achieving Physiological Realism in Music Recording and Reproduction*. Preprint 5426 AES 111th Convention Dec. 2001
- [8] Jens Blauert. *Spatial Hearing*, 1997 Edition, MIT Press. www.mitpress.mit.edu.
- [9] Kirkeby, Nelson, Hamada. *The Stereo Dipole*. *J. Audio Engineering Society*, May 1998.
- [10] T. M. Bock, D. B. Keele. *The Effects of Interaural Crosstalk on Stereo Reproduction*. 1986 AES Preprints 2420A & B.
- [11] G.Theile. *On the Naturalness of Two-Channel Stereo Sound*. *J. Audio Eng. Society*, Oct. 1991.
- [12] Masayuki Morimoto, Kazuhiro Iida, Kimihiro Sakagami. *The Role of Reflections from Behind the Listener in Spatial Impression* 1999. www.elsevier.com/locate/apacoust
- [13] A.H. Marshall, M. Barron. *Spatial Responsiveness in Concert Halls and the Origins of Spatial Impression*, 1999. www.elsevier.com/locate/apacoust
- [14] Anders Torger, Angelo Farina. *Real-Time Partitioned Convolution for Ambiophonics Surround Sound*. *Mohonk, IEEE Conference Proceedings*, Oct. 2001.
- [15] Jose Xavier Lopez. *PC-Based, Real-Time, Multichannel Convolver for Ambiophonic Reproduction*. *Proceedings, AES 19th Conference on Surround Sound Techniques*, 2001
- [16] Anders Torger. *Ambiovolver Instruction Manual Link*. www.ambiophonics.org
- [17] R Miller, "Transforming Ambiophonic + Ambisonic 3D Surround Sound to & from ITU 5.1/6.1," Preprint, AES 114th Convention, Amsterdam, March 2003.
- [18] R Miller, *iCompatible PanAmbiophonic 4.1 and PerAmbiophonic 6.1 Surround Sound for Advanced Television , Beyond ITU 5.1, i SMPTE 144th Technical Conf. & Exhibition, Pasadena CA. Oct. 02*
- [19] R Miller, *iContrasting ITU 5.1 and Panor-ambiophonic 4.1 Surround Sound Recording Using OCT and Sphere Microphones, i AES 112th Convention, Munich, May 2002, preprint #5577.*

[20] Robin Miller, *PerAmbiolating 360f: Panorambiophonic 4.0i multichannel evaluation DTS-CD, iPerAmbiolating 360f: Companion ITU 5.0,i multichannel evaluation DTS-CD, with print insert, iSurroundi pages, www.filmaker.com.*
[21] R. Miller, *iAmbiophonic Demonstration: AES 19th, Bavaria 2001i 2-chan. CD with print insert, and PS online Ambio demonstration, iSurroundi pages, www.filmaker.com.*

Appendix A

The Blumlein Conspiracy

On December 14th, 1931 the EMI sound engineer, Alan Dower Blumlein, filed a British Patent Specification 394325 entitled "Improvements in and relating to Sound-transmission, Sound-recording and Sound-reproducing systems." In the usually arcane language common to most patent applications, Blumlein's invention "consists in a system of sound transmission wherein the sound is picked up by a plurality of microphone elements and reproduced by a plurality of loud speakers, comprising two or more directionally sensitive microphones and/or an arrangement of elements in the transmission circuit or circuits whereby the relative loudness of the loud speakers is made dependent upon the direction from which the sounds arrive at the microphones."

Blumlein did not use the word "stereophonic" anywhere in his patent, but he did use the word "binaural." It was well known during the fifty years before Blumlein, that two microphones, spaced the width of the human head, feeding a remote pair of headphones, produced very realistic sound images with solid, stable, directional attributes. The problem was that the sound sources all seemed to lie within ones head or in psychoacoustic parlance, be internalized. What Blumlein sought to do was to externalize this binaural effect using loudspeakers. Externalizing the binaural effect over a full 360-degree sphere is still the Holy Grail of acoustics, particularly among those designing virtual reality video systems that also require an audio counterpart. The now dormant IMAX large screen 3d movie system uses earphones placed about an inch out from in front of the ears as well as speakers behind the screen, behind the audience, and above and below the screen to produce a full (periphonic) acoustic sphere. If home video watchers are prepared to wear earphones as well as have loudspeakers in their home movie theaters this is a very effective technology, but one that is not necessary to realistically reproduce staged musical events as opposed to movies.

Other attempts to externalize the binaural effect over a full sphere or just a circle, include, ambisonics, surround sound and the plethora of computer companies at work generating the virtual reality sound fields for the multimedia applications referred to above. Fortunately, our music problem is, and Blumlein's was, less complex since we need only consider a relatively small part of this sphere and we can assume that all direct music sound sources originate on a single flat stage in front of us (or in electronic scores a flat stage behind as well). In fact, Blumlein's first priority was to provide a better front stage sound for movies shown in theaters.

Blumlein was awarded his patent covering what we now call stereophonic sound reproduction officially on June 14th, 1933. Thus the basic stereo listening triangle is over 75 years old and just as Einstein's theory of relativity eventually refined Newtonian physics, it may be time to reexamine and modify the bedrock concepts upon which Blumlein imaging is based. And what better place to start than with Blumlein himself. Suppose one looked through Newton's treatises and

found cryptic comments by Newton hinting that he knew his laws of matter, acceleration and gravity were not fully accurate at very high velocities and masses. We would then be justified in concluding that Newton had some insight into relativity but chose not to confound his contemporaries who had enough to deal with in distinguishing between mass and weight and who in any case found his formulas were always accurate enough to do jobs like getting rockets off the ground. Newton's laws still work very well today despite relativity if you are not too fussy. So it is with Blumlein. Blumlein's patent is salted with innuendoes and hints of things that should come.

Blumlein knew that his reproduction method using two widely spaced loudspeakers was flawed, but the improvement in sound reproduction over mono was so apparent that there was no need to point out in detail its theoretical imperfections, and in any case he wanted his patent to be awarded and his invention used. However, he seemingly felt compelled to indicate to his technical posterity that he really did know precisely what was right and what was wrong with the stereophonic reproduction method he was proposing. (On the recording side, he had fewer problems and proposed the coincident stereo microphone and what we now call the Blumlein shuffler, both concepts later elaborated on in Ambisonics.) Thus in a paragraph discussing the difference between low frequency phase differences and high frequency intensity differences in providing directional cues, he writes "It can be shown, however, that phase differences necessary at the ears for low frequency directional sensation are not produced solely by phase differences at two loudspeakers (both of which communicate with both ears) (parentheses Blumlein's) but that intensity differences at the speakers are necessary to give an effect of phase difference".

What Blumlein was doing here was indicating that an unavoidable defect could be a virtue in one case. That is, he could not prevent both loudspeakers from having equal access to both ears at low frequencies, (or also having a less predictable access at all higher frequencies), so he came up with a recommended coincident microphone arrangement that counted on this low frequency loudspeaker crosstalk to provide for localization in the relatively narrow low frequency band where the ear can localize only on the basis of interaural phase differences. Thus crosstalk became a necessary evil in the coincident microphoning case. What Blumlein was really saying was that if your microphones produce signals at low frequencies that don't have any phase differences, (as is the case with any coincident microphones) then the loudspeaker crosstalk could save the day but at a cost in higher frequency intensity based localization that Blumlein himself was aware of but could not fully appreciate because of the limited frequency response of the equipment he had to work with.

The way the loudspeaker crosstalk helps in the low frequency case is as follows. At low frequencies it can be assumed that any sound from one speaker will produce the same sound pressure at both ears since the head is not an effective barrier to long wavelength sounds. But the signal will be slightly delayed in getting to the more remote ear. If now there is a second loudspeaker emitting the same low frequency signal, then when this second pair of soundwaves meets the first pair it will combine with the first pair to form a new soundwave. When two waveforms, that have the same shape, but differ in amplitude and also have a fixed time delay between them, are added together, the result is a new wave shifted in phase. At one ear the louder signal combines with the delayed softer signal. At the other ear the softer signal combines with the delayed louder signal. The results are identical amplitudes but different phase shifts at each ear and thus an interaural phase difference between the ears is created that is proportional to the original intensity difference between the microphones.

Of course, if you use a more common, non-coincident microphone technique, such as a head spaced array, this crosstalk can cause localization blurring. That Blumlein understood that this unavoidable crosstalk caused imaging

problems at higher frequencies is clear from some of the other quotes below. He clearly seemed preoccupied with this issue as he prepared his text. In point of fact, we know today that this loudspeaker communication with both ears makes it impossible for standard stereo or its surround sound relatives to create a fully realistic and lifelike stage image. But wait. There is much more to be gleaned from Blumlein. Blumlein's hints to his audiophile posterity continue with "the sense of direction of the apparent sound source will only be conveyed to a listener for the full frequency range for positions lying between the loudspeakers" Thus Blumlein certainly understood that the width of the stage he could create with loudspeakers was limited by crosstalk to the space between those loudspeakers, a serious defect, but one that was not crucial to Blumlein since he was largely concerned with widely spaced loudspeakers in large movie theaters or halls that had fairly narrow screens or stages in comparison to the depth of the theater.

In the context of a patent application however, this is not the sort of observation one would ordinarily include. It is easy to understand why the maximum width of the stereophonic sound image is limited to the angle the speakers subtend at the listening position. Let us assume that a single sound source such as a trumpet is located stage right at 80-degrees. Let us further assume that under these circumstances the sound reaching the left microphone in a stereo recording setup is negligible and therefore no audible sound comes from the left speaker during playback. The trumpet sound blares forth from the right loudspeaker at normal intensity. If the right speaker is at the usual 30-degree angle from the centerline of the normal stereo playback triangle, then the trumpet will appear to be sounding from that position instead of from 80-degrees. This is of course the everyday real life situation where we can easily locate the source of any discrete sound that reaches both ears without impediment.

Many of us have, however, heard recordings of stereo systems that do sometimes produce images that come from beyond the speakers and some audiophiles believe that if they could only get perfect recordings, speakers, cables and electronics, the image would open out. Blumlein was also loath to admit defeat on this point. He writes "but if it is desired to convey the impression that the sound source has moved to a position beyond the space between the loudspeakers the modifying networks may be arranged to reverse the phase of that loudspeaker remote from which the source is desired to appear, and this will suffice to convey the desired impression for the low frequency sound." (hang on to that word "low") This suggestion makes sense in a particular movie scene where you could briefly reverse the phase of one speaker to move dialog or a sound effect off screen, but we know that leaving one speaker out-of-phase all the time does not work for music reproduction via the stereo triangle.

What Blumlein was suggesting is a primitive form of logic steering thus foreshadowing Dolby Pro-Logic. But he has explained why sometimes images do appear beyond the position of the loudspeakers. Any inadvertent phase reversal of a spot microphone in the recording mix or an out-of-phase driver, or a large phase shift in the crossover network of a three or four way loudspeaker system or a reflection from the wall behind a dipole loudspeaker can convince even experienced listeners that wider stages can be achieved, somehow, using normal stereo technology. Unfortunately, logic steering, surround coding and even multi-channel recording methods cannot achieve the binaural ideal that Blumlein was striving for.

So far, Blumlein himself has told us that the stereophonic reproduction method has two inherent flaws. There is a third problem that Blumlein seems to have been aware of because of his use of the word "low" in the last quote. This is the image position distortion caused by higher frequency sounds that hit the pinnae from angles that do not correspond to the actual angles of the recorded source. Thus, perhaps Blumlein had trouble moving a birdcall off stage using his phase reversal trick. A related issue is the question of recorded ambience and here Blumlein appears to be struggling with the

problem of reproducing such recorded hall ambience from the proper direction. "The reflected sound waves which arise during recording will be reproduced with a directional sense and will sound more natural than they would with a non-directional system. If difficulties arise in reproduction, they may be overcome by employing a second pair of loudspeakers differently spaced and having a different modifying network from the first pair." While the vocabulary may be a bit different, this is a pretty good description of surround sound or Ambisonics and is also the basic starting point for the ambience and imaging system I have called Ambiophonics.

Appendix B

Human sound localization is possible using three and only three sonic clues (not counting bone conduction)

1. Time, including phase and transient edge, differences between the ears. This ITD includes the precedence effect.
2. Sound level differences between the ears. (ILD)
3. Single and twin eared pinna direction finding effects.

Each of these mechanisms is only effective in a specific frequency range but they overlap and the predominance of one over the other also depends on genetics, the nature of the signal, i.e. sinewave, pink noise, music, or venue, etc.

For a full range complex sound such as music, experienced live, all three mechanisms are always in play and normally agree. By definition such an experience is said to be realistic or, better phrased for the creative and artistic recording fraternity, said to yield guaranteed physiological verisimilitude. If the three mechanisms are not consistent then we often make errors in localization such as in most earphone listening where the interference with the pinna and head shadow usually result in internalization even if the ITD, including some deliberate ILD crosstalk, is perfect.

Before we get to stereophony, let me discuss the relative strengths of the three mechanisms listed above. Snow and Moir in their classic papers showed that localization of complex signals in the pinna range above 1000 Hz was superior by a few degrees, to localization that relied solely on complex lower frequencies. That is, their subjects could localize bands of high frequencies to within one half a degree but only to one or two degrees at lower frequencies. The accuracy of localization, in general, declines with frequency until at 90 Hz or so, as Bose has demonstrated, it goes to zilch. Remember this when we get to discuss crosstalk.

It is important for understanding the workings of Stereophony that you are convinced that all three mechanisms are significant and I would suggest, with Keele, Snow, and Moir, that the Pinnae are first among equals. You should satisfy yourself on some of this by running water in a sink to get a nice complex high frequency source. Close your eyes to avoid bias, block one ear to reduce ILD and ITD, and see if you can localize the water sound with just the one open ear. Point to the sound, open your eyes, and like most people you will be pointing correctly within a degree or so. With both ears you should be right on despite having a signal too high in frequency to have much ITD or ILD. But with two pinnae agreeing and the zero ILD clue, the localization is easily accurate.

Again, if a system like stereo or 5.1 cannot deliver, the ITD, ILD and Pinna cues intact without large errors it cannot ever deliver full localization verisimilitude for signals like music. If the cues are inconsistent, localization may occur but it is fragile, it may vary with the note or instrument played, and such localization is usually accompanied by a sense that the music is canned, lacks depth, presence, etc. Mere localization is no guarantee of fidelity.

Let us now look at the stereo triangle in reproduction and the microphones used to make such recordings and see what happens to the three localization cues. Basically Stereophonics is an audible illusion, like an optical illusion. In an optical illusion the artist uses two dimensional artistic tricks to stimulate the brain into seeing a third dimension, something not

really there. The Blumlein stereo illusion is similar in that most brains perceive a line of sound between two isolated dots of sound. Like optical illusions, where one is always aware that they are not real, one would never confuse the stereophonic illusion with a live binaural experience. For starters, the placement of images on the line is nonlinear as a function of ITD and ILD, and the length of the line is limited to the angle between the speakers. (I know, everyone, including Blumlein, has heard sounds beyond the speakers on occasion but diatribe space is limited.)

I want to get to the ILD/ITD phantom imaging issue involved in this topic. But let us first get the pinna issue tucked away. No matter where you locate a speaker, high frequencies above 1000 Hz can be detected by the pinna and the location of the speaker will be pinpointed unless other competing cues override or confuse this mechanism. In the case of the stereo triangle the pinna and the ILD/ITD agree near the location of the speakers. Thus in 5.1 LCR triple mono sounds fine especially for movie dialog. In stereo, for central sounds, the pinna angle impingement error is overridden by the brain because the ITD and the ILD are consistent with a centered sound illusion since they are equal at each ear. The brain also ignores the bogus head shadow since its coloration and attenuation is symmetrical for central sources and not large enough to destroy the stereo sonic illusion. Likewise, the comb-filtering due to crosstalk, in the pinna frequency region, interferes with the pinna direction finding facility thus forcing the brain to rely on the two remaining lower frequency cues. All these discrepancies are consciously or subconsciously detected by golden ears who spend time and treasure striving to eliminate them and make stereo perfect. Similarly, the urge to perfect 5.1 is now manifest.

Consider just the three front speakers in 5.1. Unless we are talking about three channel mono, we really have two stereo systems side by side. Remember, stereo is a rather fragile illusion. If you listen to your standard equilateral stereo system with your head facing one speaker and the other speaker moved in 30-degrees, you won't be thrilled. The ILD is affected since the head shadows are not the same with one speaker causing virtually no head shadow and the other a 30 degree one. Similarly the pinna functions are quite dissimilar. (In the LCR arrangement the comb-filtering artifacts now are at their worst in two locations at plus and minus 15-degrees instead of just around 0-degrees as in stereo) Thus for equal amplitudes (such as L&C) where a signal is centered at 15 degrees, as in our little experiment, the already freakish stereo illusion is badly strained. Finally, the ITD is still okay and partly accounts for the fact that despite the center speaker there is still a sweet spot in almost all home 5.1 systems. Various and quite ingenious 5.1 recording systems try to compensate for some of these errors but the results are highly subjective and even controversial. It is also probably lucky that in 5.1 recording, it is difficult to avoid an ITD since a coincident main microphone is seldom used in this environment.

Technical Digression for Recording Engineers

Before getting to side imaging, there are some points on the relationship between microphones and reproductive crosstalk that should be elucidated. Whether crosstalk is beneficial or not depends on what frequency range you are talking about and thus what localization method you are relying on. At high frequencies, in the pinna range, stereo speaker crosstalk is obviously not a benefit. There is no way that this unpredictable pattern of peaks and valleys can enhance localization in a stereo or LCR system This is true whether spaced or coincident mics are used.

Stereo crosstalk can cause a phase shift at frequencies below where comb filtering predominates. That is, two sinewave signals with slightly different delays but with comparable amplitudes will combine to form a new sinewave with some different amplitude and phase angle. I maintain that the phase part of this change is inaudible from 90 Hz on down, nonexistent for the central 10-degrees and virtually non-existent for images from the far right or left, and thus of doubtful audibility in between or in LCR systems. Stereo crosstalk cannot create an ITD for transients captured in coincident mic recordings but it can shift the phase of midbass and low bass. But there is no evidence that the small phase shifts of this type are audible or affect localization. If spaced mics are used, then there is an ITD and crosstalk has little deleterious effect but likewise no benefit.

The ILD is a slightly different story. In the low bass, say below 90 Hz, the phase difference between the direct sound and the crosstalk sound is too small (heads are too small) to cause any significant change in phase and thus change amplitude at an ear when the two signals are added together. So regardless of the microphone used, low bass crosstalk is not the issue. Again, I maintain that the very low bass energy at both ears remains almost the same even if the left and right signals are different in amplitude as in coincident mic'ing. As Blumlein observed, as the frequency goes up the path length difference is equivalent to larger phase angles and so, if there is a difference in amplitude, between the speakers, the signals will go up at one ear and down at the other as the signals are combined on each side of the head. Clearly if the phase shift gets to 90-degrees on up this same crosstalk mechanism becomes detrimental. This boost in mid bass

separation is only applicable to phantom stereo images around 15-degrees. In the center there is no crosstalk amplitude asymmetry to take advantage of and at 30-degrees where the speakers are, hopefully, the stereo separation ensures that the crosstalk has little to add to or subtract from.

If spaced microphones are used, the ILD at low frequencies may be minimum especially for omnis. But let us assume that above 90 Hz there is a substantial ILD as well as an ITD. In this case the LF effect of the crosstalk phase change is sort of unpredictable. Again in the 15 degree region there could be enhancement of bass separation but the ITD induced phase shift could counter this. In summary, crosstalk is really only desirable in the case of coincident mic stereo recordings, as Blumlein wrote, and only if restricted to frequencies below 300 Hz or so as I claim.

Surround Sound Localization

Let us consider surround sound localization. Obviously, if a mono signal is placed at 110 degrees it can be localized using pinna, ILD, and ITD even when facing forward. Between the two rear surround speakers you have effectively a stereo pair spanning 140 degrees. In such a situation, if there is a lot of high frequency energy, the pinna will localize to the speakers and it will be difficult for some individuals to hear sound directly behind or in the central rear region. (The new rear surround channel can fix this, but the LCR anomalies as above will then apply.) However, if there is a real ITD and a real ILD between the rear speakers it is theoretically possible to hear a wide stage to the rear as in the frontal stereo illusion. However the crosstalk, and thus the comb-filtering, is extreme at this angle and it starts at a lower frequency thus interfering with the ILD at 800 Hz or lower. If there is an ITD this can help but then the speakers must be properly placed or delay adjusted. Obviously, if 140-degree spacing was a good way to make a stereo stage, front or rear, it would have been done this way long before now.