STEREO MICROPHONE TECHNIQUES...... *Are*

Now that most of the technical aspects of sound recording and reproduction have reached an advanced state, it is my contention that the quality of the source material has become the major limiting factor. For acoustically derived musical performances the microphone technique used to capture the event stereophonically has a profound influence on the attributes of the reproduced sound. I believe that spaced-microphone recording techniques are fundamentally flawed, although highly regarded in some quarters, and that coincident-microphone recordings are the correct way to go. The "air" and "ambience" and "depth" so valued in spaced-microphone recordings are shown to be largely the artifacts of phasiness due to the microphone spacing and not acoustic ambience at all. To support my claims I first define my terms and then review briefly the historical emergence of these two techniques in the 1930s, together with the relevant aspects of stereophonic imaging theory. A series of simple experiments is described which it is hoped the reader will use to demonstrate to his own satisfaction certain basic properties of stereophonic systems. Conclusions are drawn regarding the different characteristics of spaced-microphone and coincident-microphone recordings and a plea is made for more critical and knowledgeable evaluation. It is pointed out that stereophony is inherently incapable of fully natural imaging, and that the Ambisonic surround-sound system is the proper extension to multichannel reproduction. The paper is deliberately nonmathematical in the hope that it will provoke lively thought, discussion, and experimentation among those engaged in sound recording. A mathematical appendix is provided for those readers interested in understanding the reasoning underlying my conclusions. A tape comparing simultaneous coincident- and spaced-microphone recordings is available so that readers can assess for themselves some of the effects discussed. (See p. 733 for ordering details.)

the Purists Wrong?*

by STANLEY P. LIPSHITZ

INTRODUCTION

The last few years have seen a dramatic improvement in our ability to accurately record, distribute, and reproduce musical signals, and the benefits of this digital audio technology are now available to consumers in their homes. Concomitant improvements have taken place in microphone and loudspeaker design as well, and the technical capability of the complete recording/distribution/reproduction chain is impressive. Yet it is my contention that many recording producers and engineers are currently making recordings which are a disaster from the stereophonic point of view. What is on the master tape is now laid bare without the masking effects of the earlier technology, and what the consumer can now hear is frequently unpleasant. It has become fashionable to praise as "purist" those stereophonic recordings made with the basic minimum of just two or three microphones. I happen to believe that this trend is in the right direction, but I also believe very strongly that fewer does not always mean better, that not all purist microphone techniques are equally good or valid, and that some are just downright wrong and misguided. I shall try to make a strong case for the use of single-point (that is, coincident) stereophonic microphone techniques in preference to widely spaced microphone configurations.

I am aware that I am treading on dangerous ground here, in that an aesthetic judgment is called for when attempting to rate a stereophonic recording as "good" or "bad." Clearly not everyone will agree on the criteria to be used nor on their relative importance in the overall ratings. Nevertheless, I maintain that it is possible to formulate overall objectives against which different recording techniques can be judged. Even if you do not agree with my conclusions, I do hope that you will be stimulated into experimenting and listening critically to the results of your experiments with these objectives in mind. I feel that the source material is now the weakest link in the chain from the artist to the listener, and that improvement here requires an enlightened reassessment of what goes on in the process of capturing the original sound and reproducing it through two loudspeakers.

WHAT IS STEREO?

Stereo arose from experiments which took place simultaneously but independently at Bell Telephone Laboratories in the United States [1]–[3][1] and at what became Electric and Musical Industries Ltd. (E.M.I.) in Great Britain [4]–[10] in the early 1930s. I shall have more to say about these experiments shortly, but first I must define what I mean by stereo. Perhaps the Bell engineers had the best term for it. They called it "auditory perspective" (but see [11]).

First, stereophonic sound must be clearly distinguished from binaural sound. Undoubtedly the most perfect re-creation of an acoustic event for a listener is in principle obtained by a system that can re-create at his or her eardrums the same acoustic pressures as occurred originally. This is, however, not as easy as it sounds. It involves recording the signals at the ears of the auditor, or at the ears of a dummy head and torso with pinnae and ear canals identical to those of the auditor, and reproducing those signals in such a way (usually through headphones) that the same acoustic pressures occur at the eardrum position as would have occurred naturally. This is the basis of the so-called binaural recording/reproducing system. Only two channels are required. The important point is that these two signals must be independently and *separately* conveyed to the two ears of the listener—the right ear must not hear the left-ear signal, and vice versa. Possibly the earliest binaural demonstrations took place at the Paris Opera in 1881 where Clement Ader had set up 10 spaced telephone transmitters on the stage linked pairwise to individual receivers in the Palace of Industry several kilometers away [11]. If a listener held adjacent receivers to his ears he heard the sounds picked up by a pair of widely spaced microphones, which provided a binaural effect. I shall not discuss binaural stereo further here.

What the Bell and E.M.I. researchers understood by stereo is the same as what we understand by stereo today, and this is quite distinct from binaural. In stereo sound reproduction using two loudspeakers, each ear hears *both* loudspeakers, and is in-

[1] The references cited do not represent a complete bibliography, but have been carefully chosen from the very large number of published works. A more extensive list of references will be found in the works cited.

* Presented at the 78th Convention of the Audio Engineering Society, Anaheim, California, 1985 May 3–6.

tended to do so. This is a fundamental point, the misunderstanding of which has led some companies to misguidedly design loudspeakers which attempt to cancel this interaural crosstalk in the belief that this will lead to more accurate stereo reproduction. This is not so. Such loudspeakers, to the extent that they can succeed in their goals, are suited for binaural rather than for stereo reproduction. The two types of signals are not interchangeable. Stereo is predicated on the existence of this "interaural crosstalk." The problems in stereo are those of capturing suitable signals to be fed to the two loudspeakers, given that each ear hears both loudspeakers, and choosing the loudspeakers' polar patterns and geometrical configuration to widen the "stereo seat" if so desired. We shall discuss the former aspect in some detail. The problem of freeing the listener from the "stereo seat," by enlarging the region within which the image remains reasonably free from distortion, is in my view a reproduction-related question rather than one bearing directly upon the recording technique. A number of schemes have been proposed for achieving this goal [12]–[17].

Stereo derives from the Greek word "stereos" meaning solid. Webster's dictionary defines stereophonic as "giving a three-dimensional effect of auditory perspective." This is an acoustic analog of the objectives of stereoscopic photography, although it should be clearly realized that the *techniques* of stereoscopy are more analogous to those of binaural than to those of stereophonic sound reproduction. The aim is to provide the illusion of a solid, three-dimensional acoustic image in the mind of the listener. This image should have both lateral (left–right) precision as well as a realistic sense of depth or distance to the sound sources. Now, unlike stereoscopy, where distance perception is largely a function of parallax (that is, differences between the images seen by the two eyes), in auditory terms distance is largely sensed by the relationship between the direct and the reverberant sounds arriving at each ear. Whereas each eye has the ability to localize an image in two dimensions, it is principally through the use of *both* ears that we can do this aurally. So the analogy must not be pushed too far. Nevertheless, a vague, imprecise stereophonic image may be likened to an improperly overlapped and aligned stereoscopic image, in that detail becomes blurred as regards both lateral image positioning and depth perspective.

> *Many engineers aim only to produce a spread of sound images which seems pleasant*

I consider such blurring to be a defect, although I will admit that some people like soft-focus lenses. What perturbs me greatly is that many people see this defect as a positive virtue. I agree with Leakey [14, p. 240] that, rather than accurate imaging, many engineers "aim only to produce a spread of sound images which seems pleasant. Often it is the case that the more 'ethereal' the sound images appear, then the better the system is appreciated. Such systems can be regarded, however, only as attempts at pseudo-stereophony, even though two channels might be employed." I hope to show below that one reason is the general confusion between ambience in a recording, which can enhance the stereo image, and phasiness, which can only serve to degrade it. But more of that later. For the moment I am trying to define the objectives of a stereophonic system. Stereo is inherently limited by the availability of only two transmission channels. One can do better and set higher goals if one has more than two channels available, and this will also be discussed briefly. I believe that the best that stereo can do is to provide a credible illusion that between and beyond the pair of loudspeakers there exists another acoustic environment within which the musicians are located and performing. I am considering only musical events which have a genuine acoustic origin within an acoustic setting, a situation that pertains to all classical musical performances and a great many others besides. I accept that not all music originates under such conditions. If it does not, the compilation of a stereo image is a synthetic exercise with no possible reference back to an acoustic original, and I will not be considering this situation further here.

It should be noted that my definition of stereo is neither a "you are there" nor a "they are here" one. It is closer to the former than the latter, since for most of us it is ludicrous to consider having a full symphony orchestra present in the listening room. However, it is not possible, with two channels alone, to fully place the listener *into* another imaginary environment. All the direct sounds from the loudspeakers reach him from the front quadrant only, a simplification which prevents his being placed *within* the other environment. The best one can do is to provide the illusion that the end of the listening room between and beyond the loudspeakers has been removed, enabling the listener to "listen into" the original recording venue as if it were situated behind the loudspeakers. He is still aware of the acoustic presence of his own listening room superimposed upon that of the recording environment, and with only two channels one cannot afford to make the listening room anechoic without unpleasant consequences. It probably is desirable, though, to minimize reflections from the loudspeaker end of the room, as is done frequently now in monitoring studios, so as to allow the listener to perceive more clearly the recorded acoustic of the performing space. The latter will normally be dominant due to the fact that the performance space is invariably much larger than the listening room.

If more than two transmission channels are available, one can do much better. Using three or more transmission channels in a properly designed surround-sound system one can hope to place the listener *within* the ambience of the recording locale. Since such a system can re-create ambient sound arrivals from all around the listener, it is then possible (and indeed, even desirable) to suppress the effects of the listening room. For such reproduction systems (for example, Ambisonics) an acoustically dead listening room would be preferable. It is my belief that as more sophisticated reproduction systems become available, the correct trend will be toward more anechoic listening environments.

It goes without saying that no stereo system can capture and correctly reproduce direct or ambient sounds picked up from outside the frontal quadrant. Directional distortions are inherent in any attempt to reproduce surround sound over only two loudspeakers. Different stereo systems will perform differently in this respect, each having its own directional peculiarities for sounds originating from the remaining three quadrants.

HISTORICAL PERSPECTIVE

My definition of stereo appears to be compatible with the concept as envisaged by the original inventors at Bell Laboratories and at E.M.I., not that they gave any explicit definition. Indeed, most of what we know about the experiments conducted by Alan Blumlein and his coworkers at E.M.I. is contained in his patents from that time [4]–[8], the earliest of which has justifiably become a classic in the field. A good survey of Blumlein's contributions to stereophony will be found in [9], [10]. As regards Harvey Fletcher and his coworkers at Bell, we have their own detailed accounts to go by [1]–[3]. The two teams followed quite distinct paths.

The scientists at Bell started from the idea of re-creating the original *macroscopic* acoustic wavefront within the listening environment. They envisaged a "curtain of microphones" in front of the performers, each microphone connected by means of its own transmission channel to the corresponding loudspeaker in a "curtain of loudspeakers" in the listening room. The wavefront at the curtain of microphones would be recreated at the curtain of loudspeakers and propagate into the listening room as if the performers had been present

Alan Blumlein

behind the curtain.[2] Unfortunately this hypothetical scheme would not correctly handle those reverberant acoustic wavefronts which travel in the direction from the listener's side of the curtain toward the performer's side, but this is a deficiency of all stereophonic systems anyway. Since the Bell concept is one of a large-scale wavefront reconstruction, it would not require that a listener be located in a specific "stereo seat." The information rate needed for a successful large-scale wavefront reconstruction scheme is vast, and beyond the realms of possibility at present. So a simplification to a finite and small number of transmission channels is called for.

The Bell workers considered two or three channels to be the practical maximum, and conducted their experiments using this number. Two or three is a far cry from infinity, however, and having made this simplification, one should really ask oneself whether a two- or three-channel wavefront reconstruction scheme makes any sense. The answer, as we shall see, is no. It is interesting to take note of the fact that the Bell scientists tried all possible combinations of two or three microphones with two or three transmission channels and two or three loudspeakers for reproduction. These experiments included the idea of a center-bridged microphone in a two-channel system, this being a configuration often highly praised among purists. Another point worth noting is that the Bell listening setup was relatively distant from the loudspeakers, the total included angle subtended at the listener being only on the order of 35°. This is about half of what one would tend to use nowadays, and certainly must have influenced the weighting they placed on lateral imaging accuracy or the lack thereof.

Blumlein, on the other hand, realized from the outset that with only two transmission channels a macroscopic wavefront reconstruction scheme in the listening room was not possible, and thus invoked psychoacoustic criteria in designing his system. His aim was to produce acoustic signals in a limited region around the head of a listener in the stereo seat, which would lead to the formation of an accurate virtual image of the source position. In effect, Blumlein's system set out to re-create on a *microscopic* scale the wavefronts that the Bell people had hoped to generate macroscopically. This is feasible and indeed practical. Blumlein's scheme relied on the realization that simple level differences at the loudspeakers would create both level *and* phase differences at the listener's ears because of the fact that each ear hears both loudspeakers and the ears are laterally separated [4]. By suitably choosing these level differences it is possible to produce stable images between the loudspeakers. This procedure is, for obvious reasons, called "intensity stereo." We would nowadays produce the required level encoding by the use of a coincident pair of directional microphones, classically the "Blumlein pair" of 90°-angled figure-of-eight microphones. Blumlein indeed patented both this arrangement as well as the sum–difference or M–S (mid–side) schemes for deriving these loudspeaker feed

[2] A theoretically better concept would be to surround the preferred listening position in the performing space with a multiplicity of outwardly oriented microphones, each connected by its own transmission channel to the corresponding loudspeaker in a similar configuration of inward-pointing loudspeakers surrounding the actual listening position in the listening room.

signals [6], [7] once directional microphones became available—specifically the RCA ribbon microphone of the early 1930s. In his initial experiments [4], however, in the absence of other than omnidirectional microphones, he was forced to derive a directional pickup pattern by subtracting and re-equalizing the outputs of two closely spaced pressure microphones, a process he called "shuffling."

It is fascinating to realize that Blumlein's fundamental 1931 patent [4] contained within it the basis of our current 45–45° stereo disk system. In fact, both the E.M.I. and the Bell work was way ahead of its time and did not see commercial realization until the mid 1950s with the advent of stereophonic tapes [10] and then disks.

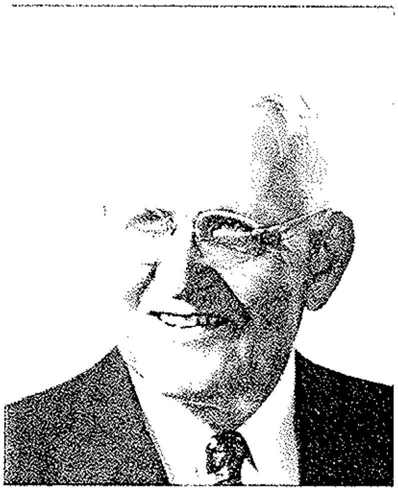
To understand the substantial differences in behavior between the Bell spaced-microphone and the Blumlein coincident-microphone stereo arrangements, it is necessary to discuss the relevant auditory theory.

THE PSYCHOACOUSTICS OF HEARING

I shall consider only image localization in the horizontal plane since we are primarily interested in stereo. In normal hearing, when listening live to a single acoustic source ("natural" hearing), each ear receives only a single copy of the direct sound from the source. In stereo reproduction each ear hears two copies of the sound, once from the left-hand loudspeaker and once from the right-hand loudspeaker. This is a fundamental but essential difference, and it is what enables the two fixed loudspeakers to create the illusion of an image separated from their actual positions when fed with suitable signals. So we must consider stereo hearing as distinct from natural hearing and actually quite unnatural—it is in fact an artificial creation. To understand it, we first recap the situation in natural hearing. An excellent discussion of spatial hearing together with an extensive list of references is given in Blauert [18, chaps. 2 and 3].

Natural Hearing

When listening live, the signals at the two eardrums differ in time of arrival, level, and spectral content, and these differences depend on the source position. The time of arrival difference is due to the physical spacing of our ears and cannot exceed about 630 μs in real life. This corresponds to a path-length difference of about 210 mm, and this in turn represents a half-wavelength at a frequency of around 800 Hz. It thus follows that for frequencies below about 800 Hz there is an unambiguous phase relationship between the two ear signals and the source direction, the ear nearer the source having the leading phase. This is what I shall call the *low-frequency regime*. This phase difference is frequency dependent; in fact it varies linearly with frequency since it represents a pure time delay. At frequencies above about 800 Hz the interaural phase shift can exceed 180°, and so the ability (on periodic or steady-state signals) to discern which ear's signal is leading and which is lagging is lost. So, clearly, interaural phase relationships would appear to be useful cues only at low frequencies.



Harvey Fletcher

There is a further reason for this, which is probably not fortuitous but represents the kind of compatible adaptations that organisms seem to acquire by evolution. The transmission mechanism from the inner ear to the brain is by neural impulses generated by the hair cells in the cochlea. These hair cells can fire at rates of only about 1000 pulses per second, and then only for brief periods. (For an excellent general survey of the physiology of the hearing mechanism see Schroeder [19].) At low frequencies the acoustic input waveform can produce a sympathetic modulation of this neural firing rate during the course of a single cycle, and so the ear has to some extent a waveform-following ability. This would enable the discrimination of interaural phase differences at low frequencies, and this seems to be the case. However, it follows that at frequencies well above 800 Hz the ability to follow the input waveform will be lost and hence that phase relationships (both within each ear's signal as well as between the signals at the two ears) will become irrelevant. This is found to be so, and can easily be demonstrated. (See Experiments below.) Above about 1.6 kHz this ability is essentially lost. We thus see that in the low-frequency regime the phase-sensing ability ties in with the physical spacing of our ears to enable directional localization of live sounds in natural hearing. In the octave between 800 and 1600 Hz our localizing ability is not good, but a further mechanism based on interaural level differences comes to the rescue at frequencies above about 1.6 kHz, a region which I shall call the *high-frequency regime*. Moreover, for impulsive signals the time-of-arrival difference can still be used for unambiguous image localization in the horizontal plane (front or back).

Indeed, the diffracting and occulting effects of the head and outer ears (pinnae) produce very significant level differences at the two ears for sources which are not in the median plane in this high-frequency region where their dimensions are comparable to the acoustic wavelength. These interaural level differences are strongly dependent on both source position and frequency and also on the individual listener's head size and pinna shape, and can exceed 20 dB for certain combinations of direction and frequency. They represent the second main source of directional information in natural hearing, but play a secondary role to that of the interaural time differences. These frequency-dependent level differences produce modifications to the spectrum of the source signal which are used as a further cue, especially for localizing sounds which lie above or below the horizontal plane and for providing front–back discrimination ability. The level and time-of-arrival differences reinforce each other: a source to the left is louder in the left ear as well as reaching it first. Conversely, if the two ears are furnished with coherent signals such that the left-ear signal is louder and/or earlier than that reaching the right ear, an image toward the left will be formed, and the location of this binaural image will vary monotonically as a function of both the level and the time differences.

So far I have based the discussion largely on the behavior of our hearing system for periodic or sinusoidal source signals. These are characteristic of the sustained portion of musical notes but take no account of the transient-type signals produced by many percussive instruments, nor of the fact that much speech and music is in reality a *modulated* periodic signal the envelope of which varies slowly relative to the carrier signal itself. For impulsive and modulated types of signals it appears that the time-of-arrival difference at the two ears is the primary cue [18]. This is, of course, unambiguous for all frontal source directions in the horizontal plane and, as we have seen, does not exceed about 630 μs naturally.

Stereo Hearing

The two basic mechanisms which we have at our disposal to synthesize image localization in the front quadrant between the two stereo loudspeakers in the decidedly unnatural situation represented by stereo hearing are thus interaural time (or, in the low-frequency regime, phase) differences and interaural level differences. The only major question is how to produce at the listener's ears from the *two* source loudspeakers such differences as will be interpreted by the hearing mechanism as representing a credible image between the loudspeakers. Bear in mind that in stereo each ear hears *two* signals, one from the left-hand loudspeaker and one from the right-hand loudspeaker. The level and time (or phase) differences at the listener's ears are *not* the same as those at the loudspeakers. It is the former that matter as regards image localization, but the latter which will be determined by the relationship between the performer and the microphones. It is important that, as far as possible, these two loudspeaker signals combine at the listener's ears to produce cues which are compatible with natural hearing. We can to a large extent analyze the listener's response on the basis of how he responds in natural hearing to the signals arriving at his ears, as discussed above. It is thus necessary to analyze how the two loudspeaker signals combine at the ears of the listener. These loudspeaker signals may be thought of as differing in level and/or timing (that is, phase). We shall discuss the properties of spaced-microphone and coincident-microphone signals and how they relate to these psychoacoustic criteria in the next section. For the moment we just consider the consequences of different loudspeaker feed signals.

THE CONSEQUENCES FOR STEREO

Intensity Stereo

By means of intensity differences alone between the two loudspeaker signals it is possible to move the perceived image continuously between the two loudspeakers. This is of course well known and is the mode of operation of the standard balance control or pan pot. A level difference on the order of 12–15 dB is needed to move the image fully over to the louder loudspeaker, but this is somewhat dependent on the nature of the signal. In the low-frequency regime *the signal intensities at the two ears remain almost equal*, even though the loudspeaker signals may differ greatly in level. This is due to the lack of significant effects by the head and pinnae in modifying the signals and due to the close proximity of the two ears. The image shift heard is

Bruel & Kjaer microphones are available from:

Arizona, Phoenix
EAR PROFESSIONAL AUDIO
(601) 267-0600

California, Encino
EVERYTHING AUDIO
(213) 995-4175

California, Los Angeles
WESTLAKE AUDIO, INC.
(213) 851-9800

California, Oakland
LEO'S PROFESSIONAL AUDIO
(415) 652-1553

Illinois, Chicago
BRIDGEWATER SOUND
(312) 281-8920

Illinois, Elmhurst
AVC SYSTEMS
(312) 279-6580

Illinois, Pekin
MILAM AUDIO COMPANY
(309) 346-3161

Illinois, Schaumburg
STUDIO-SONICS CORP.
(312) 843-7400

Maryland, Silver Spring
RECORDING CONSULTANTS, INC.
(301) 587-1800

Minnesota, Minneapolis
AVC SYSTEMS
(612) 729-8305

Michigan, Farmington Hills
HY JAMES
(313) 471-0027

New York, New York
MARTIN AUDIO VIDEO CORP.
(212) 541-5900

Ohio, Parma Heights
PHIL REDDISH SOUND
(216) 885-3030

Oregon, Portland
RMS SOUND
(503) 239-0352

Pennsylvania, Philadelphia
TEKCOM CORP.
(215) 627-6700

Tennessee, Nashville
VALLEY AUDIO
(615) 383-4732

Washington, Seattle
RMS SOUND
(206) 362-0491

B K

Bruel & Kjaer Instruments, Inc.

due primarily to the *phase* differences between the ear signals produced when the two loudspeaker signals combine at each ear. Since the ears are separated, the path lengths from each loudspeaker to the two ears differ, and this produces opposing phase shifts in the sum signals at the two ears. Thus, surprising as it may appear, *in the low-frequency regime loudspeaker level differences are translated into interaural phase shifts*, the ear on the side of the louder loudspeaker having the leading phase [10], [12]–[14], [18], [20]–[22]. Indeed, this phase shift varies approximately linearly with frequency, and so represents a pure interaural time difference. This is precisely what happens in natural hearing. The ear on the same side as the source hears a signal leading the far ear in phase, this linear-phase difference being indicative of the fact that there is a pure time-of-arrival difference between the two ear signals. Intensity stereo thus re-creates natural interaural time delays in the low-frequency regime. It is important to appreciate that this result is true, not just for sinusoidal or periodic signals but for any low-frequency signal. The relationship between interloudspeaker level difference and image position is a fairly linear one. Since these results are not well appreciated, I have attempted in the Appendix to give a clear exposition of the situation for those readers who may be interested in a fuller understanding.

In the high-frequency regime interaural phase relationships no longer matter, and indeed would be ambiguous. But again, one can demonstrate that pure interloudspeaker level differences produce lateralization of the apparent image position. Here, head masking and ear diffraction are significant, and both interaural level and phase differences result from such loudspeaker signals, only the former being relevant, however [18]. According to some investigators [10], [13], [22], the apparent image position is slightly farther away from the center for high-frequency signals, and a small correction circuit can help fuse both low- and high-frequency images more perfectly [23].

As regards transient or broad-band signals, theories exist for explaining why good, consistent imaging is produced by interloudspeaker level differences alone. That of Vanderlyn is particularly interesting and noteworthy [21].

From the foregoing we see that a stereo signal with imaging based upon interchannel level differences alone will produce good imaging over loudspeakers. This is the basis of Blumlein's original patent and of all coincident-microphone stereo recording systems ("intensity stereo"). The left- and right-channel microphones are made as spatially coincident as possible to minimize interchannel time delays, and their polar patterns are chosen to produce the necessary interchannel level differences as a function of source position. The ultimate in coincidence is provided by a microphone such as the Calrec Soundfield microphone, whose output feeds correspond to microphones coincident within millimeters to extremely high frequencies. The optimum microphone arrangement is 90°-angled figure-of-eights, whose cosine polar pattern produces the most accurate imaging in the frontal quadrant as well as the most uniform spread of reverberation between the loudspeakers [24]. Other polar patterns are also possible if suitably angled. For example, hypercardioids at around 105° included angle or cardioids at about 131° included angle will also work reasonably well. One crucial requirement is that the microphones' polar patterns be independent of frequency, and this is something that only the very best small-diaphragm microphones approach. I believe that many people who have experimented with coincident-microphone stereophony have been misled by the consequences of poor microphone polar patterns, which often become more directional at high frequencies. This will cause the central images to recede and become less clearly defined in the arrangements described above (often known as X–Y stereo). The alternative of M–S (mid–side) stereo is, with ideal polar patterns, mathematically equivalent to X–Y stereo, but can in practice often produce better central image quality with many available microphones [25]. I cannot discuss these detailed aspects further here, but will merely note that the Soundfield microphone provides the most uniform and coincident polar patterns of any available stereo microphone.

> *One crucial requirement is that the microphones' polar patterns be independent of frequency*

Time-Based Stereo

Having seen that intensity stereo can work well and is compatible with achievable microphone configurations, we look at the second basic possibility, namely, stereo based solely on interchannel time differences. For relatively distant microphone placement, this is the form of the signals that would be produced by closely spaced (say, under 1 m) omnidirectional microphones, for which the path-length differences to the performer produce negligible level differences but significant time differences. (Such an arrangement, even without a dummy head, will produce quite reasonable binaural imaging when reproduced over headphones, but we are primarily concerned with two-loudspeaker stereo reproduction here.) Widely spaced microphones produce both intermicrophone level *and* time differences, of course. The curves of constant time difference (that is, constant path-length difference) are hyperbolas drawn with the two microphones as foci.

For impulsive-type sounds, over a limited range of interloudspeaker time delays of up to maybe ±3 ms, it is found that the image can be moved over the stereo stage. It should be recalled that in natural hearing one never experiences interaural delays of more than 630 μs, and this happens here too, of course. Each ear receives one signal from *each* loudspeaker

If the interloudspeaker delay is sufficiently small relative to the loudspeaker subtended angle (such as under 200 μs for the normal stereo arrangement of 60° subtended angle at the listener), each ear hears the impulse from the loudspeaker on its own side before the crosstalk signal from the opposite loudspeaker arrives. If the delay exceeds this amount, the situation is different and the earlier loudspeaker's signal arrives at both ears before the signal from the delayed loudspeaker arrives at either ear. In spite of this, up to delays on the order of 3 ms a continuous image movement between the loudspeakers has been found to be possible on this basis. But the stereo stage is covered in a rather nonuniform way, such that the half-left and half-right images require only about one quarter of the delay needed to produce full-left and full-right images. This is one contributory factor in producing the "hole in the middle" effect common with spaced-microphone recordings: sources near the center image well away from the center and suffer considerable image broadening. The imaging is also variable from listener to listener, is affected by small head movements, and is quite signal dependent. In other words, it does not appear to be a good basis for stable stereo imaging. If the delay exceeds 3 ms (as can occur with microphone spacings of more than 1 m), the sound appears to come from the earlier loudspeaker only, until one reaches delays sufficiently long that the Haas (precedence) effect no longer holds and one becomes aware of an echo from the delayed loudspeaker. This can happen if the signals are derived from microphones spaced sufficiently far apart (say, more than 15 m apart), but this is not usually the case for realistic spaced-microphone setups.

For nonimpulsive sounds, however, an analysis [18, chap. 3] shows that the interaural signal differences produced by interloudspeaker time delays do not lead to consistent imaging at all. Surprisingly, perhaps, it is found that *in the low-frequency regime pure time delays between the loudspeakers produce only level and polarity differences between the two ear signals*. (See the Appendix for a proof of this result.) Of primary concern is the fact that the ear on the side of the earlier loudspeaker need not receive the louder signal, and indeed at low frequencies does not! So the interaural level differences produced at low frequencies do not always reinforce the imaging produced by impulsive sounds. Sometimes the low-frequency image pulls in the opposite direction from the image of the transient, broadening and smearing the overall image. The interaural polarity reversals set in at frequencies for which the time of arrival difference of the two loudspeaker signals at *each* ear exceeds a half-period. When this occurs in the low-frequency regime, the sound takes on a distinctly "phasey" quality and the image becomes quite anomalous. Otherwise the image is merely blurred to a greater or lesser extent. In the high-frequency regime, interloudspeaker time differences lead to both interaural level and phase differences, but with no consistency from the imaging point of view. In effect this means that if the time delay between the two microphone signals exceeds the interaural delay, the sound becomes phasey and the image disintegrates. This would force the use of microphone spacings of no more than about 200 mm, for which the first interaural polarity reversal would occur at about 1 kHz for normal source positions, in order that the low-frequency imaging anomalies should not drastically degrade the resultant recording. This would, however, introduce intermicrophone delays of only 300 μs maximum for normal source positions, and hence a very narrow image on impulses, in view of the fact that around 3 ms (that is, 10 times as much) is needed for full-width imaging. So one is faced with an impossible dilemma: use close microphone spacing to prevent low-frequency phasiness and get basically monophonic imaging, or use wider spacing for imaging on impulses and tolerate the low-frequency amorphousness. Even then the high-frequency imaging produces a bad hole in the middle since it is quite nonlinear, and this is aggravated by the intermicrophone level differences which now also occur with wide spacing of the microphones, for these tend to pull the images even farther away from the center and into the loudspeakers. Imaging in the sense of my definition would appear impossible in spaced-microphone stereo. These defects are indeed quite familiar to critical listeners.

Imaging in the sense of my definition would appear impossible in spaced-microphone stereo

These facts may be unpalatable to some, but are not surprising. For if we think of sinusoidal signals only, for a constant interloudspeaker time delay (as would be produced, for example, by a performer in a constant position relative to two spaced microphones), the two loudspeaker signals reaching the stereo seat go in and out of phase purely as a function of frequency. For frequencies at which the delay represents a multiple of the period, these loudspeaker signals add (constructive interference), while for frequencies for which the delay is an odd multiple of the half-period, the loudspeaker signals precisely cancel (destructive interference). At the listener's ears the signals go in and out of phase as well as rising and falling in amplitude as the frequency is varied. We have a comb-filtering interference effect at each of the ears. The combing occurs at a higher rate at the ear on the side of the leading loudspeaker. At each comb null the interaural polarity reverses. The polarity differences are not worrisome in the high-frequency regime, but we know from our earlier discussion that they will destroy low-frequency imaging. The sound becomes phasey at low frequencies, as certain frequency components are in phase at the two ears whereas others are out of phase. Synthetic ambience devices use time delays to produce just such combing. Here our microphones are doing it for us. It thus appears that time-delay-based stereo cannot work well.

Spaced-Microphone Stereo

In most practical situations in which spaced microphones are used, one has *both* level and timing differences in the two channels due to the relatively large intermicrophone spacing (many meters). For sources nearly equidistant from the two microphones (that is, central sources) the level and delay differences are small, while for source positions far from the center very significant level differences and interchannel delays occur. For impulses both level and delay effects pull the image in the same (and correct) direction, and it soon collapses into the nearer loudspeaker in consequence. For steady-state signals, as we have seen, the imaging is totally anomalous and frequency dependent for each source position that is not precisely on the centerline. (See the Appendix.) Only for sources well to the sides does the level difference dominate and keep the image stable, but then these images come from near the loudspeakers instead of from between them, as intended. One has what is commonly called a "hole in the middle," a characteristic of two-spaced-microphone stereo. Even closely spaced (that is, comparable to the interaural spacing) omnidirectional microphones do not solve the problem, but do tend to reduce it. As the microphones are moved together, the low-frequency phasiness is reduced and the image tends to concentrate at the center of the stereo stage instead of spreading widely. This may be more acceptable but really is not much more than monophony. Closely spaced omnidirectional microphone recordings are suitable only for binaural listening over headphones. The surprising thing is that some people are recommending this for stereo recording.

I know of no published research showing that the two- or three-channel spaced-microphone arrangements can in any consistent way recreate a semblance of the original wavefront in the listening room, let alone in the vicinity of the listener's head. They cannot and do not produce stereo imaging in the sense in which I have defined it. There is, on the other hand, good evidence to support my contentions [26]. A simple thought experiment is instructive. Imagine two holes (or doorways) in the wall between two adjoining rooms, and it is not difficult to convince oneself that on the basis of the acoustic signals which come through these openings the listener will *not* be able to accurately localize the position of the sound source on the far side of the wall. But these two holes are the "loudspeakers" of our elementary stereo experiment, when viewed from the listener's side of the wall, and the "spaced microphones" when seen from the performer's side. For stereo we are obliged to use spaced loudspeakers, but the fallacy is to conclude from this that these loudspeakers should be fed with signals derived from spaced microphones.

One sign of a good stereo recording/reproduction system in addition to producing stable images is that the loudspeakers are not audible as sources of the sound

A modification, investigated by the researchers at Bell, was to use a three-channel system of spaced microphones. As their data show [3, p. 247], this arrangement is somewhat better than the two-microphone two-channel scheme, as would be expected. Central images are stabilized, but the desired uniform spread across the stereo stage is not produced. Instead of one hole in the middle, we have two smaller anomalous areas half-left and half-right, and as with the two-channel arrangement, the images still tend to cluster around the loudspeakers. One sign of a good stereo recording/reproduction system in addition to producing stable images is that the loudspeakers are not audible as sources of the sound. Indeed one can, with one's eyes closed, ideally not tell precisely where the loudspeakers themselves actually are. A good coincident-microphone recording can achieve this, but not a spaced-microphone recording using only two or three channels. Neither can the use of a center-bridged microphone or a center-bridged loudspeaker in a two-channel system significantly improve its performance in this regard. I should point out that the center-bridged three-microphone two-channel arrangement is often used in current purist-type recordings.

To conclude this section, I remark that the stereo hearing models presented above can be formalized into mathematical models which can then be used to assess the imaging of any proposed stereo system (see [10], [13], [14], [16], [18], [22], [26]–[30] and the Appendix). These models predict the behavior described above. The most convenient description seems to be in terms of the total sound particle velocity vector at the listener's head in the low-frequency regime and the "energy vector" in the high-frequency regime. For a single plane or spherical progressive wave the sound velocity vector (normal to the wavefronts) points away from the source, and so it is to be expected that one should, in the low-frequency regime, localize the source in the direction of the velocity vector in natural hearing. For stereo (or surround-sound) hearing this still appears to be true at low frequencies: even though the two progressive wavefronts coming from the two loudspeakers do *not* combine to produce a *single* progressive wave, the localized direction is nevertheless given by the resultant of their individual velocity vectors [27]. At high frequencies, on the other hand, the energy vector, determined on the basis of power flow from each source past the listener's head, appears to be appropriate [28], [29]. This takes recognition of the ear's phase deafness at high frequencies. A more detailed discussion of these mathematical models is beyond the scope of this paper, but I shall return to them briefly when discussing the fundamental limitations of stereo and their solution.

THE BASIC STEREO SYSTEMS COMPARED

We have investigated the underlying theory of the two opposed stereo microphone systems: coincident versus spaced. I would now like to describe verbally what I hear when listening to carefully made recordings of both types using the most accurate microphones available. Many of these effects have already been predicted and briefly described in the previous section. Other authors have, of course, provided their own assessment of the different configurations; see, for example, [26], [31]–[33], which reach conclusions generally similar to mine. I would also like to discuss some experiments which we recently undertook in order to try to pin down the reasons for some of the audible artifacts that distinguish these two types of recordings.

Coincident Microphones

Blumlein-type crossed figure-of-eight recordings produce wide, stable imaging on the direct sound from the performers. The image does not wander as the musician plays up and down the scale, and is generally quite narrow and unblurred. This presupposes an accurate polar pattern over the frequency range, which fortunately is easier to produce in a figure-of-eight (or omnidirectional) microphone than in a cardioid. Nevertheless, most commercial microphones have polar patterns which depart somewhat from the ideal, especially at high frequencies, and this tends to blur the image. With most commercial figure-of-eights it is necessary to reduce the front-quadrant angle between the microphones to between 80 and 85° in order to compensate somewhat for these aberrations and prevent a slightly recessive central image. The best available examples (Calrec Soundfield and Schoeps MK8) are excellent and work well. Separate capsules should be mounted one above the other and as close as possible. Only a microphone like the Soundfield can make them effectively coincident, and by its very nature ensure frequency-independent polar patterns over the bulk of the frequency range. A correction circuit for the difference (L − R) channel may be worth trying [23].

Of course, figure-of-eights pick up front and back equally (interestingly, their directivity index of 4.8 dB is the same as that of a cardioid), and so Blumlein stereo recordings contain a substantial amount of hall ambience. Being angled at 90°, the random-incidence power pickup is independent of direction in the horizontal plane (because $\sin^2\theta + \cos^2\theta = 1$). This gives the most uniform possible spread of reverberation between the loudspeakers and con-

DO IT YOURSELF

Listed below are brief details of some simple and easy-to-perform experiments which will enable you to confirm for yourself many of the properties of hearing and of stereo reproduction which I have discussed previously. In particular, I very strongly urge you to fit a polarity-reversing switch in the feed to one of your two stereo loudspeakers so that you have it available for assessment purposes on a continuing basis. This is one of the simplest and most useful tools for stereo analysis.

In each experiment, try a variety of musical sounds, both percussive and legato, simple and complex, narrow band and wide band. Observe the two transducer feeds on an X–Y oscilloscope to enhance the understanding of what is being heard and why.

Interchannel Phase Audibility (low frequency and high frequency)

(a) Use a dual mono signal over headphones. Try reversing the polarity of one channel only and note the oppressive effect. Using low-pass filtered material with cutoff around 800 Hz note that the phasiness is almost intolerable. Using high-pass filtered material with cutoff around 1.6 kHz note that the image remains central and polarity reversal is almost undetectable (try harpsichord, cymbals, slapsticks). Sharp-cutoff filters are required for this experiment. This demonstrates properties of binaural hearing.

(b) Perform the same sequence of experiments using normal stereo loudspeakers this time. The same general conclusions still apply under single-channel polarity reversal, since on a mono signal this still creates equal out-of-phase signals at a central listener's ears. This relates to phasiness in stereo listening.

Time and Intensity for Stereo Imaging

(a) Again with a dual mono signal, listening over both headphones and loudspeakers, confirm that interchannel level differences (use the "balance" control) shift the image laterally, with 12–15 dB causing a full-width shift. Understand that on headphones one is listening to a pure interaural *level* difference alone, whereas over loudspeakers the effect is due solely to an interaural *time delay* in the low-frequency regime. (See the Appendix.) This points up the vast difference sequently the most accurate depth impression. Do not forget that it is basically the ratio between the direct and the full-sphere reverberant sounds produced by any source which enables one to perceive and assess distance ("depth") live, and the same is true in recordings. This is one of stereo's great advantages over mono—the reverberation and the direct sound do not both need to come from the same point. Of course, a surround-sound system like Ambisonics which can handle both direct and reverberant sounds accurately is the next major step up to reality.

Blumlein recordings produce the truest stereo depth perspective of any stereo format. As a corollary, there is usually a preferred replay volume setting at which the reproduced perspective sounds natural. The volume control acts like a "distance control." The most natural setting turns out to represent a realistic reproduction level, often quieter than expected. (It is surprising how often music tends to be reproduced at well above realistic levels. This distorts the spectral balance of a recording.) It should be noted that in the absence of any reverberation (such as under anechoic conditions) the Blumlein signals are none other than sine–cosine pan-potted mono signals, and this applies to the direct sound under reverberant recording conditions. It is the way the Blumlein arrangement handles reverberation as well as the accuracy of this sine–cosine pan-potting on direct sound, which contribute to its great advantages.

A concomitant property of Blumlein recordings is that unpleasant or excessive room reverberation will produce less than ideal results. Of course, one ought not to record in bad halls and demand that the recording not show this fact. Of all available stereo techniques I consider the Blumlein configuration to be the most accurate, both objectively and subjectively. The precision of the image is sometimes considered to be a deficiency with this method. But as I shall explain, the "air around the instruments," so prized by some, is *not* ambience but rather phasiness and the resultant image blur. True ambience should come from the recording locale and not be an artifact produced by defects in the recording technique. True ambient information is valuable, but I can do without synthetic impostors which have nothing to do with ambience at all. Good Blumlein recordings actually contain a surprising amount of ambience, which can usually be revealed, especially for central performers, by listening to the (L−R) difference signal. In fact, the direct-sound suppression for a central source can be so good that in the (L−R) mode one can literally "listen to the hall." A further advantage of the technique, especially important to

EXPERIMENTS

between binaural and stereo listening.

(b) Insert an adjustable digital delay line in the dual mono feed to the right channel, so that it receives the same signal as the left channel but delayed relative to it and with the right channel level adjustable relative to that of the left channel. Ensure that the digital delay is noninverting at low frequencies by checking visually with an asymmetrical waveform like a half-wave-rectified 200-Hz sine wave, for example. [If a variable delay is not available, a digital audio processor with "video in" linked straight through to "video out" can be used to provide about 10 ms of high-quality but fixed delay. Beware of possible polarity inversion, however. For example, the Sony PCM-F1 is polarity inverting, whereas the Sony PCM-701 is not. If the left and right channels are cascaded, either processor provides ≈ 20 ms of delay (noninverted). For reference, the unity-gain setting of the PCM-701 is around 6 on the gain control.] Recall that in natural hearing, interaural delays are always less than about 630 μs. Listen on headphones as well as on loudspeakers. With equal levels set in both channels, increase the delay gradually to, say, 20 ms starting at zero. Note that on loudspeakers the recording begins to sound more ambient and spacious. This is phasiness of the type produced by comb-filter digital reverberators. Transients appear to stay in the left channel once the delay exceeds about 3 ms, but for delays in the range of 0–3 ms the transient image can be positioned between center and full left as a function of the delay. This is the only range over which time-based stereo can work, and then on transients only. Verify that for nontransient low- and high-frequency signals no consistent frequency-independent imaging can be produced over loudspeakers for delays in the range of 0–20 ms. As the delay is increased toward 20 ms, the sustained low-frequency portions of the sound begin to spread across the stage in a pitch-dependent way. Conversely, keep the delay fixed at, say, 10 ms and vary the level of the delayed channel. The same effects are heard. Unless the level of the delayed channel is above that of the undelayed one, the delayed channel is not localized per se. If its level is made too high, or the delay too long, the image "splits" and both sources are heard. This experiment demonstrates the precedence effect and the impossibility of stable stereo imaging based on interloudspeaker time differences, such as are produced by spaced omnidirectional microphones. [Conversely, note that on headphones for delays ≤ 630 μs good imaging is produced. (This, of course, is the range of normal binaural hearing.)] This experiment also demonstrates that interchannel delays produce an ambient, airy, warm, and reverberant quality, all of which are artifacts of the phasiness thus introduced.

(c) Electrically sum the original and the delayed signals and listen in mono to the combing as a function of delay. Keep both signals equal in level. Note the changes in spectral broadcasters, is its mono (L+R) compatibility. Timbral changes due to comb-filter cancellations do not take place with any coincident-microphone technique. If one reverses the polarity of one of the two channels, one produces a most horribly oppressive and phasey sound with no low-frequency imaging at all. It is rather like the ghastly effect produced by reversing the polarity of one earpiece of a headset while listening to mono. The very extent of the effect is indicative of the high degree of coherence between channels in Blumlein and other coincident-microphone systems.

I have made hundreds of coincident-microphone (Blumlein and other configurations) recordings, and the above comments are the summation of my experience. For those who complain of a lack of "air" or "warmth," I would point out that since in stereo reproduction all the recorded sound, both direct and reverberant, comes from the front quadrant, it cannot sound fully natural. The recorded reverberation does not surround the listener as does live reverberation. Only a good surround-sound technique can cure this defect without degrading other aspects of the recording. Note also that artificial reverberation as currently available cannot simulate the acoustics of specific halls. This must be recorded in order to be properly re-created. Reverberation, although highly complex, is not random. There is a correspondence between the direct sound and the reverberation for each source position and each microphone position. Spaced microphones pick up multiple images of this reverberation, whereas a single-point microphone captures it coherently.

Viewing the left- and right-channel signals on an X–Y oscilloscope is an instructive experience as the high degree of coherence is apparent. The trace confines itself largely to the in-phase first and third quadrants, with only the reverberation contributing

DO IT YOURSELF EXPERIMENTS

balance and timbre. A direct A/B comparison with the input signal will make the changes quite apparent. This reflects on the mono compatibility of spaced-microphone recordings as well as the combing such recordings produce at the listener's ears in stereo. Verify that for small delays (<300 μs), such as would be produced by quasi-coincident microphones, the effect is not too objectionable, while for longer delays (as with regularly spaced omnis) the results can be quite bad. If the delay is large enough that the comb tooth spacing is small enough to ensure that more than one tooth falls inside each critical bandwidth of the ear (about one-third octave), the effect may be less objectionable. At low frequencies this requires inordinately long delays and is not useful for stereo.

(d) Listening to double mono over loudspeakers, move your head sideways from the center through a small distance (up to, say, 300 mm). This introduces effectively only interloudspeaker time delays. Notice how the image moves from center, broadening in the process and becoming somewhat phasey. Can you hear the colorations caused by combing at the ears? This becomes very obvious on a double-mono white or pink noise source. This combing is the problem with time-based stereo and is, of course, also a problem with off-center listening to any form of stereo.

Stereo Listening Experiments over Loudspeakers

Listen critically to various types of stereo recordings, of high quality, of coincident, quasi-coincident, and spaced-microphone types. Make your own if necessary. Arrange to be able to reverse the polarity of *one* of the loudspeaker feeds. Watch on an X–Y oscilloscope connected to the loudspeaker feeds. Subtract the two channels and listen to the (L−R) signal to assess the coherence of the images of central sources in the recording. If possible use a double sum–difference type of matrix (as used in M–S recording) to alter the amount of (L−R) signal relative to the (L+R) sum. Note how the presence of too much (L−R) signal causes central images to fall apart. On Blumlein recordings note the "correctness" of unity (L−R) to (L+R) gain ratio. Try the same experiment on two-spaced-omni recordings. Listen to the mono sum signal to assess mono compatibility and combing. Listen from locations away from the stereo seat to assess image changes which occur as one moves laterally in front of the loudspeakers. (Many of these experiments can also be performed using the recordings on the available tape— see p. 733.)

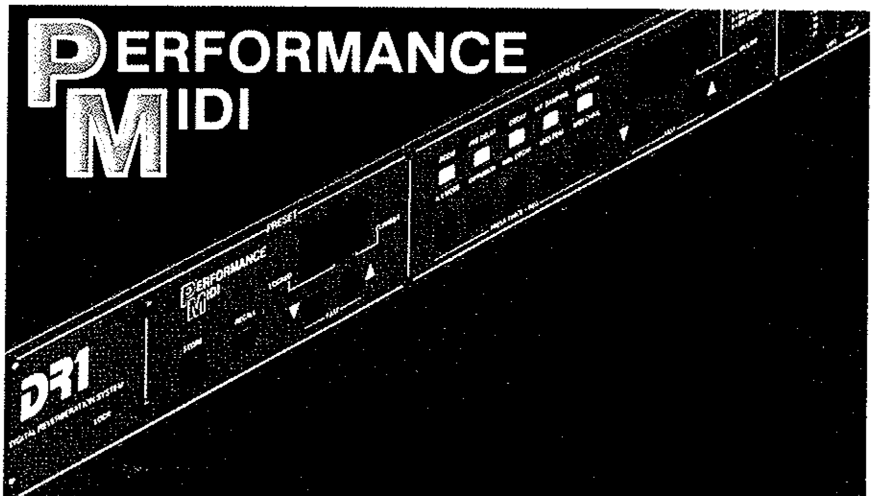
to the second and fourth (out-of-phase) quadrants. Under anechoic conditions, the frontal pickup quadrant of course produces totally in-phase signals.

Changing the patterns used and their angling in X–Y coincident recording, or using the corresponding M–S configurations, enables the stereo stage width and depth, and ambience pickup to be modified relative to Blumlein. This can ameliorate the effects of poor halls and poor microphone polar patterns somewhat, but the spread of ambience between the loudspeakers is not as uniform [24]. I have found 105°-angled hypercardioids (Schoeps MK41s are excellent) preferable to 131°-angled cardioids under less than ideal conditions. This may partly be an artifact of the inadequate polar patterns of most cardioid microphones. All these alternative configurations produce a narrower total stereo stage width than Blumlein, for a given arrangement of performers and microphone position. 90°-angled cardioids, often recommended, produce a very narrow, center-dominant stereo stage. The recommended anglings given above are based on a theoretical +3-dB mono sum for central sources. This is the same as given by 90°-angled figure-of-eights. I shall have more to say about the reasons for this later, but for now I should like to mention that for microphones of less than ideal cardioid or hypercardioid patterns these angles may need to be reduced until a stable, nonrecessed central image is attained.

Quasi-Coincident Microphones

Moving on next to closely spaced pairs of directional microphones, with spacings comparable to the interaural spacing, we come to the ORTF, NOS, and similar schemes. These stand midway between the coincident and the widely spaced arrangements. Because the spacing is comparable to the interaural spacing, the resultant phasiness is largely confined to the high-frequency regime where it is less objectionable and only slightly degrades the intensity-based imaging. In the low-frequency regime the microphones are still essentially coincident. Here they differ from closely spaced omnidirectional configurations in that outwardly angled directional microphones are used. This gives a tremendous advantage in that intensity cues relating to the performers' left–right positioning are captured, something that is absent with closely spaced omnis. So in the low-frequency regime we have a system that works largely on the intensity stereo principle. In the high-frequency regime the interchannel phase differences are aurally acceptable and the microphone directional characteristics still provide intensity-based imaging, but with some phasiness which extends down into the midband to an extent that depends on the actual spacing adopted. On transients the time-based image reinforces the intensity-based image to broaden the stereo stage width. The phasiness and the consequent slight delocalization also contribute to the generation of a somewhat wider but less precise stereo image than would be given by a similarly angled coincident pair.

These attributes can be viewed in one of two ways. If one values lack of phasiness and ultimate image precision (as I do), one would hear this as a slightly degraded form of co-



Our new software installed in your DR1 offers you the most extensive MIDI implementation of any reverb made and includes Performance MIDI!

You now have real time control, in addition to full computer control, preset dump and reload, and master/slave operations allowing you to link two DR1s. Other features standard in ART's new Version 1.2 software include a Flanger/Chorus Room, a Preset Sequences feature and 10 additional new Factory Presets.

For those who demand performance, check out these specifications for our DR1 with Performance MIDI:

PRESETS:	40 FACTORY. 100 USER PROGRAMMABLE!
D/A BANDWIDTH:	16 BIT LINEAR - 14KHz RESPONSE!
UPDATEABLE:	YES! AND, WE DELIVER. VERSION 1.2 NOW AVAILABLE!
REMOTE CONTROL:	YES! FULL CONTROL OF ALL REVERB AND PERFORMANCE MIDI FUNCTIONS!
MIDI:	YES! THE MOST MIDI CONTROLLED REVERB AVAILABLE!
SPECIAL EFFECTS:	YES! UPDATEABLE AND EXPANDABLE!
DECAY TIME:	0.1 TO 25 SECONDS AND DYNAMIC DECAY TIME CAPABILITIES!
PREDELAY:	0.0 TO 200ms. IN 1.0ms. INCREMENTS!
OTHER CONTROLS:	H.F. DAMPING, POSITION AND DIFFUSION!

Simply, if you want the best sounding reverb with the fullest MIDI implementation available including the best sounding PLATES, ROOMS AND HALLS, contact us. We'll continue to impress you with unique products that deliver superior performance!

Call or write today, for more information!

- *Digital Reverberation Systems*
- *Digital Delay Systems*
- *Full Line of Equalizers*
- *Pitch Transposer Packages*
- *Software Update Packages*

PRO-AUDIO NEVER SOUNDED SO GOOD!

Applied Research & Technology Inc.
215 Tremont Street
Rochester, New York 14608
(716) 436-2720

incident stereo, which indeed it is. On the other hand if one considers phasiness and its feeling of "airiness" and "ambience" to be a positive characteristic, perhaps in the absence of genuine ambient warmth, one would consider these systems to be a greatly improved form of spaced-omni stereo, which indeed they are too. Listening to the (L−R) signal shows that it has much more high-frequency content than the corresponding coincident configuration. Mono compatibility is still quite good since the close spacing does not produce disastrous combing in the low-frequency range. An X–Y oscilloscope examination confirms the reduced coherence at high frequencies. As with coincident techniques, it therefore follows that a polarity reversal of one channel produces gross phasiness and accompanying image breakup at low frequencies.

This combination of performance characteristics has led to quite widespread use of closely spaced directional microphones as a compromise between the acknowledged virtues of truly coincident microphones and the perceived virtues of spaced omnidirectional microphones. I personally find the latter to be defects and not virtues, and hence prefer true coincidence to this half-way compromise.

Spaced Microphones

Finally we come to the spaced-omni arrangement. As already discussed, the very question of imaging now becomes secondary because there is very little true imaging at all. Central images are quite anomalous and phasey. They wander around with pitch and are very broad. Off-center images tend to cluster around the loudspeakers, which become apparent as sources of sound. As a performer moves from the center to the side, the "image" quickly collapses into the nearer loudspeaker. The tonal quality of instruments is audibly affected by the serious combing which occurs in the low-frequency regime. The sound may be variously described as being more "ambient," "airy," or "warm" than coincident microphone recordings in the same locale. Partly this is due to the difference in the directivity index of the microphones used. The omnidirectional microphone has a higher ratio of reverberant to direct sound than the directional microphone. But I maintain that the difference is largely one of phasiness and that this airiness is being confused with true ambience. Many people, perhaps, have never been confronted with this distinction, and I hope to encourage you to listen more carefully and critically to this aspect.

> *Information is only useful if it can be usefully processed*

A bridged-center microphone fed equally into both left and right channels helps central imaging somewhat, depending on its level, but cannot prevent the anomalies of spaced-microphone stereo. Certainly this arrangement is more acceptable than the use of only two spaced omnis.

The low-frequency phasiness of these arrangements is well known to require careful source placement if successful disk cutting is to be possible. It is often stated that it is the time "information" being picked up which is a valuable asset that coincident microphones miss. I beg to differ. As the previous section and the Appendix should make clear, information is only useful if it can be usefully processed, and this information when reproduced over loudspeakers is misinformation. The cues received at the listener's ears are not compatible with normal human hearing, especially in the low-frequency regime [26].

A further characteristic difference between spaced and coincident systems should be mentioned. With Blumlein stereo, in the absence of special loudspeaker polar behavior to prevent it, the whole stereo image moves over with the listener as he moves laterally away from the stereo seat, but at a greater rate. Thus significant image distortion occurs with significant displacement of the listener from the center. With spaced-omni stereo some images still tend to cluster around the two loudspeakers, even for somewhat noncentral listening positions. Neither system per se treats off-center listeners better. The profound differences lie in how they treat central listeners. The imaging properties of stereo systems for off-center listeners can be assessed quite well using only central sound sources in the recording. The performance on central sources best characterizes the imaging behavior.

Many of these characteristics can be illuminated by a few experiments. Try reversing the polarity on *one* of the channels. In contrast to the coincident and almost-coincident arrangements, one frequently finds that one cannot hear any significant difference. This is especially true of the two-spaced-omni case. Think about what this means. It means that the coherence between channels is very low indeed—there is in fact an almost random-phase relationship between the two channel signals. Doesn't this worry you? Effectively it does not matter much whether your recording/reproduction system is in phase or not, you can hardly tell the difference. An X–Y oscilloscope display will help confirm this feature. The "ball of string" on the screen is distributed equally between both in-phase and out-of-phase quadrants. Reverse the polarity of *one* signal to the oscilloscope and see if you can tell the difference. Listen to the (L−R) signal and compare it with the (L+R) signal. The (L−R) signal is not largely reverberant sound, but rather direct sound. The mono sum (L+R) displays quite clear combing, which distorts tonal qualities. Note that the sound occasionally seems to spread beyond the arc subtended by the loudspeakers, an artifact of the out-of-phase relationship between the signals in the two channels for certain frequencies and source positions. No true stereo system should do this, as stereo cannot produce true images beyond the loudspeaker arc. This is indicative of an error in the recording method. These may be rather damning assertions, but I can support them.

PHASINESS INVESTIGATED

We conducted the following careful experiments to confirm that the interchannel phasiness is the cause of the perceived airiness. Two simultaneous stereo digital recordings were made during live concerts, one using a pair of coincident microphones and the other a pair of spaced microphones. These two recordings can be played back in synchronism so that one can during replay switch between the two and compare aspects of imaging, ambience, airiness, and so on. In order that, as far as possible, only the single parameter of intercapsule spacing distinguished these two recordings, various precautions had to be taken. Four nominally identical Schoeps MK41 hypercardioid capsules were used, matched into two pairs, one the coincident pair and the other the spaced pair. The matching within each pair was extremely close, while between pairs the matching was generally within ± 1 dB across the frequency range. The same type of capsules had to be used for both recordings in order to ensure not only similar tonal balances in each, but also so that the direct-to-reverberant ratio was the same for all four channels.

The coincident pair was angled at an included angle of 105°. In the first experiment the spaced pair was separated by 1 m and angled so that the left microphones of each pair were parallel, as were the right microphones (that is, the spaced pair was also angled 105° outward). This may at first seem strange, but it was essential in order to maintain the direct-to-reverberant pickup as near identical as possible between the two left microphones as a pair and the two right microphones as a pair. Since the microphones were relatively distant from the performers (2–3 m), the relative levels at the microphones were thus maintained nearly identical between pairs. This can be confirmed during replay by comparing the two left microphone signals with each other (and similarly for the two right microphones). There is almost no distinguishable difference between these individual single-microphone signals. The vast differences become audible only when one listens in stereo to the two recordings. All the previously discussed properties manifest themselves, including the more "ambient" sound of the spaced-microphone recording. Because of the controls described, it is primarily in the interchannel phase relationships that the two recordings differ. Left–right level differences are essentially the same for each performer in each of the two recordings. The highly coherent interchannel phase relationships of the coincident pair contrast vividly on an X–Y oscilloscope with the random phase relationships of the spaced pair, as expected, and this difference is thus proven to be the cause of their audible differences. Polarity reversal of one channel of the coincident recording is very unpleasant, but is not greatly audible on the spaced recording.

The second experiment again used a 1-m spacing for the spaced pair, but this time these two microphones were angled straight ahead. With changes in the direct sound pickup now also present, this is not a single-variable experiment like the preceding one, but the same general artifacts are present with of course somewhat less extreme left–right imaging on the spaced pair this time. The phasiness again contributes to the ambient feeling of the spaced recording.

The last experiment moved the spaced pair into a quasi-coincident arrangement with a spacing of 250 mm between capsules and 105° included angle. This is close to the ORTF and NOS configurations. Since angling was again the same between spaced and coincident pairs, this experiment essentially varied only the capsule spacing. This time the comparison is more subtle. The quasi-coincident pair provides quite good low-frequency imaging, with some slight phasiness evident lending a bit of "air" to the sound compared with the fully coincident pair. The loss of imaging precision is noticeable but not disturbing as was the case with the 1-m spacing. Phase coherence is

DEMONSTRATION CASSETTE ORDERING INFORMATION

This cassette is produced on CrO_2 tape, and is available with either Dolby B or Dolby C noise reduction.

Price $10 each
To order, please complete and return with check to: Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-0075, USA.

Name ________________ Dolby B ____ Dolby C ____

Address ________________

____

________________ Zip ________

Check enclosed for $________

EUROPEAN ORDERING INFORMATION

Continental Europe: Ms. Titia K. S. Bakker (c/o PolyGram B.V.), P.O. Box 189, 3740 AD Baarn, The Netherlands. Price: 490 BEF. *(All orders will be invoiced. Invoices will give the amount due in BEF and your local currency, and will indicate a selection of payment methods suitable for various European countries.)*

United Kingdom: AES British Section, Lent Rise Road, Burnham, Slough, SL1 7NY, United Kingdom. Telephone Burnham (STD 062-86) 63725. Price: £6.00.

reasonable and single-channel polarity reversal quite unacceptable.

I believe that these experiments serve to confirm my claims. In order to enable readers, who may not be in a position to conduct similar experiments, to assess these recordings for themselves, a high-quality cassette reduction is being made available at modest cost by the Audio Engineering Society. It contains switched A/B comparisons between the pairs of simultaneous recordings and illustrates not just the phasiness aspect but also many of the properties previously discussed, both good and bad. (See p. 733 for ordering information.)

WHAT'S WRONG WITH STEREO?

A few comments on some additional inherent limitations of stereo are in order, and here I am referring to the best that stereo can do, namely, Blumlein stereo. Having direct sound coming only from the front quadrant is not in itself a severe restriction for most purposes. But having *recorded* reverberation coming only from the front is quite unnatural, and placing the listener within the reverberation is a fundamental benefit of a proper surround-sound system. There is, however, one little-appreciated aspect of the way stereo treats the direct sound which is wrong and can only be remedied by a proper multichannel system.

Most engineers are probably familiar with the fact that the mono sum (L+R) of Blumlein stereo is up 3 dB relative to each channel separately for central sources. This is a consequence of the fact that the proper pressure–velocity relationship is not correctly maintained at the listener's head. In a free-traveling plane wave the sound pressure (a scalar quantity) and sound particle velocity (a vector quantity perpendicular to the wavefronts) are in phase, and their magnitudes are related by the resistive characteristic impedance of air. So the direct sound at a listener's head has this property too. In stereo there are two wavefronts going past the listener's head—one from each loudspeaker. Their individual pressures add in scalar fashion and their velocities in vector fashion to obtain the resultant sound field at each point. This addition does not maintain the proper pressure–velocity relationship, as can easily be seen.

Consider equal in-phase signals at the loudspeakers, as would be produced by a center-front source. At the listener's head the sound pressure is up by a factor of 2 (6 dB) by scalar addition, but the sound velocity is up by only 3-dB (a factor of $\sqrt{2} = 1/\sqrt{2} + 1/\sqrt{2}$) by vector addition if the loudspeakers subtend a 90° angle at the listener. There is thus a 3-dB deficit. This is really indicative of the fact that the two separate wavefronts launched from the stereo loudspeakers are not equivalent to the single wavefront originally coming from the source. The summing error reduces as the angle between the loudspeakers gets smaller, becoming zero when the angle between the loudspeakers is very small. (But this is not useful for stereo.) This possibly explains why stereo setups usually use a subtended angle of less than 90°–60° is more common, leading to a summing error of just over 1 dB.

> *Three and not four channels of information is the optimal number for horizontal surround sound*

This error can be eliminated if one has loudspeakers surrounding the listener, and is correctly handled in the Ambisonic surround-sound system. Since in this system (see, for example, [28]–[30]) both pressure (W) and velocity (X, Y, Z) signals are independently available, suitable loudspeaker feeds can be synthesized so that the correct pressure–velocity relationship is regained at the listener's head. This requires the sympathetic behavior of *all* the loudspeakers for a center-front source, something quite different from earlier quadraphonic systems which, being pairwise pan potted, were no better than stereo in this respect. Ambisonics correctly synthesizes both the zero- and the first-order spherical harmonics of the sound field at the listener's head. The Soundfield microphone enables the W, X, Y, and Z signals to be derived directly from the original acoustic field. A further indication of the profound difference between Ambisonics and quadraphonics lies in the realization that three and not four channels of information is the optimal number for horizontal surround sound. The fourth (Z) channel provides height information.

It would be taking us too far afield to get into a detailed discussion of Ambisonics here. It is a system which I see as the successor to stereo. As I indicated, it reconstructs at the listener's head the original pressure–velocity relationships of the direct sound in the recording locale. In this sense it is a highly accurate *local* wavefront reconstruction scheme. And just to indicate further aspects in which it is right and others are wrong, I point out that the reverberant sound field, like a standing wave pattern, does not have the same unique one-to-one correspondence between pressure and velocity that one finds in a traveling wave. At a pressure node, for instance, the pressure is always zero whereas the velocity is not. Since Ambisonics handles pressure and velocity independently, it can correctly handle both direct and reverberant signals whereas stereo cannot. And herein lies the basis for its extreme naturalness and precision in imaging, depth, and ambience. This is the way of the future.

ACKNOWLEDGMENTS

I would like to thank my colleague John Vanderkooy for many stimulating discussions on this topic and for help with the recording experiments. I also thank Bob Snelgrove of Gerr Electroacoustics for making it possible for us to acquire a Calrec Soundfield microphone. The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

• REFERENCES •

[1] H. Fletcher et al., *Bell Labs. Rec.*, vol. 11, pp. 254–261 (1933 May); vol. 12, pp. 194–213 (1934 Mar.).

[2] H. Fletcher et al., *Elec. Eng.*, vol. 53, pp. 9–32, 214–219 (1934 Jan.).

[3] H. Fletcher et al., *Bell Sys. Tech. J.*, vol. 13, pp. 239–308 (1934 Apr.).

[4] A. D. Blumlein, British Patent 394 325 (application 1931 Dec. 14; granted 1933 June 14); reprinted in *J. Audio Eng. Soc.*, vol. 6, pp. 91–98, 130 (1958 Apr.).

[5] A. D. Blumlein, British Patent 429 022 (application 1933 Oct. 23; granted 1935 May 23).

[6] A. D. Blumlein, British Patent 429 054 (application 1934 Feb. 10; granted 1935 May 23).

[7] A. D. Blumlein, British Patent 456 444 (application 1935 Feb. 7; granted 1936 Nov. 9).

[8] B. J. Benzimra, "A. D. Blumlein—An Electronics Genius," *Electron. & Power*, vol. 13, pp. 218–224 (1967 June).

[9] A. Hope, "Blumlein," *Hi-Fi News & Rec. Rev.*, vol. 22, pp. 57–61, 115–117 (1977 Aug., Sept.).

[10] H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn, "The 'Stereosonic' Recording and Reproducing System," *IRE Trans. Audio*, vol. AU-5, pp. 96–111 (1957 July-Aug.); reprinted in *J. Audio Eng. Soc.*, vol. 6, pp. 102–117 (1958 Apr.).

[11] "The Telephone at the Paris Opera," *Sci. Amer.*, vol. 45, pp. 422–423 (1881 Dec.); reprinted in *J. Audio Eng. Soc.*, vol. 29, pp. 368–372 (1981 May).

[12] F. H. Brittain and D. M. Leakey, "Two-Channel Stereophonic Sound Systems," *Wireless World*, vol. 62, pp. 206–210, 331–334 (1956 May, July).

[13] D. M. Leakey, "Some Measurements on the Effects of Interchannel Intensity and Time Differences in Two Channel Sound Systems," *J. Acoust. Soc. Am.*, vol. 31, pp. 977–986 (1959 July).

[14] D. M. Leakey, "Further Thoughts on Stereophonic Sound Systems," *Wireless World*, vol. 66, pp. 154–160, 238–240 (1960 Apr., May).

[15] J. Crabbe, "Broadening the Stereo Seat," *Hi-Fi News & Rec. Rev.* vol. 24, pp. 64–68 (1979 June), pp. 74–79 (1979 July), pp. 81–83 (1979 Sept.).

[16] J. M. Kates, "Optimum Loudspeaker Directional Patterns," *J. Audio Eng. Soc.*, vol. 28, pp. 787–794 (1980 Nov.).

[17] M. F. Davis, U.S. Patent 4 503 553 (application 1983 June 3; granted 1985 Mar. 5).

[18] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, J. S. Allen, Transl. (MIT Press, Cambridge, MA, 1983).

[19] M. R. Schroeder, "Models of Hearing," *Proc. IEEE*, vol. 63, pp. 1332–1350 (1975 Sept.).

[20] G. F. Dutton, "The Assessment of Two-Channel Stereophonic Reproduction Performance in Studio Monitor Rooms, Living Rooms and Small Theatres," *J. Audio Eng. Soc.*, vol. 10, pp. 98–105 (1962 Apr.).

[21] P. B. Vanderlyn, "Auditory Cues in Stereophony," *Wireless World*, vol. 85, pp. 55–60 (1979 Sept.).

[22] J. C. Bennett, K. Barker, and F. O. Edeko, "A New Approach to the Assessment of Stereophonic Sound System Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 314–321 (1985 May).

[23] P. B. Vanderlyn, British Patent 781 186 (application 1954 Aug. 18; granted 1957 Aug. 14).

[24] M. A. Gerzon, "Stabilising Stereo Images," *Studio Sound*, vol. 16, pp. 60–64 (1974 Dec.).

[25] W. L. Dooley and R. D. Streicher, "M–S Stereo: A Powerful Technique for Working in Stereo," *J. Audio Eng. Soc.*, vol. 30, pp. 707–718 (1982 Oct.).

[26] B. Bernfeld and B. Smith, "Computer-Aided Model of Stereophonic Systems," presented at the 59th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 26, p. 380 (1978 May), preprint 1321.

[27] Y. Makita, "On the Directional Localisation of Sound in the Stereophonic Sound Field," *EBU Rev.*, pt. A, no. 73, pp. 102–108 (1962 June).

[28] M. A. Gerzon, "Surround-Sound Psychoacoustics," *Wireless World*, vol. 80, pp. 483–486 (1974 Dec.).

[29] M. A. Gerzon, "The Optimum Choice of Surround Sound Encoding Specification," presented at the 56th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 25, pp. 332, 334 (1977 May), preprint 1199.

[30] P. B. Fellgett and M. A. Gerzon, "Ambisonics, Parts 1, 2," *Studio Sound*, vol. 17, pp. 20–22, 40 (1975 Aug.), pp. 24–30 (1975 Aug.); Correction: *ibid.*, p. 60 (1975 Oct.).

[31] C. Ceoen, "Comparative Stereophonic Listening Tests," *J. Audio Eng. Soc.*, vol. 20, pp. 19–27 (1972 Jan./Feb.).

[32] B. Bartlett, "Stereo Microphone Technique," *db Mag.*, vol. 13, pp. 34–46 (1979 Dec.). Also: Letter to the Editor, *ibid.*, vol. 14, p. 8 (1980 Apr.).

[33] R. D. Streicher and W. L. Dooley, "Basic Stereo Microphone Perspectives—A Review," *J. Audio Eng. Soc.*, vol. 33, pp. 548–556 (1985 July/Aug.).

GOOD LISTENERS.

If you're fanatical about the quality of your audio recording or reproduction, let Selco's superb VU Meters, PPM Indicators, and Audio Level Meters listen in.

No mistake, Selco's VU's and PPM's are truly professional instruments. VU Meters meet the full requirements of American National Standard ANS C16.5-1954. PPM Indicators conform to specification BS 4297:1968, with models available meeting European and BBC specs.

Several different styles in each type of meter gives you a wide choice of sizes, mounting arrangements, and scale lengths, too.

How pleasant that these outstanding instruments are priced as low as 36.88 each in OEM quantities.

And if you have slightly less demanding applications, Selco's reliable, well made Audio Level Meters may be your answer. Same quality taut band meter movement, similar handsome cases and scales. But prices as low as 15.50 each in quantity.

Request Selco's full color brochure on the entire line. Or give us a call. We're good listeners, too.

selco PRODUCTS CO.

7580 Stage Rd. • Buena Park, CA 90621 • Phone (213) 921-0681, (800) 25-SELCO • Telex 655457.

AES BOOTH 806

APPENDIX

STEREO LOCALIZATION AT LOW FREQUENCIES

The aim of this appendix is to prove some of the assertions made in the section "The Psychoacoustics of Hearing" regarding the nature of the interaural signals in stereo listening. Similar conclusions can be found in [4], [10], [13], [14], [16], [18], [22], [27], to which we refer the reader.

Mathematical Preliminaries

Referring to Fig. 1, let l represent the interloudspeaker distance, h the interaural distance (assumed to be much less than l: $h << l$), and d the perpendicular distance of the listener in the stereo seat from the line joining the loudspeakers, which subtend an angle $2\theta_0$ at his position. We also consider only the low-frequency regime in which head-shadowing and pinna effects are negligible, and so we can ignore level differences between the signals produced by the left loudspeaker at the left ear (P_{LL}) and at the right ear (P_{LR}), and similarly for the right loudspeaker signals P_{RL} and P_{RR}. We cannot, however, ignore the relative time delay τ_h between P_{LL} and P_{LR} or between P_{RR} and P_{RL} in computing the total acoustic pressure signals P_L and P_R produced at his ears. For reasons of symmetry which will greatly simplify the description of events, we measure these time delays relative to the nominal time of arrival of the signals at the center of the listener's head; that is, P_{LL} and P_{RR} are advanced by $\tau_h/2$ and P_{LR} and P_{RL} are delayed by $\tau_h/2$ relative to these nominal signals.

We now consider both loudspeakers to be driven from a common source signal, but with level differences and relative time delays. Denote by L and R the levels of the left and right loudspeaker signals, respectively. (Unless otherwise stated, we shall take both L and R to be positive; a negative value would represent a polarity reversal in that channel.) Furthermore, suppose that there is a time difference τ_l between the two loudspeaker signals (left leading right if $\tau_l > 0$, and conversely if $\tau_l < 0$). Again for symmetry reasons let us consider the left loudspeaker signal to be advanced by $\tau_l/2$ relative to the input signal, and the right loudspeaker signal to be delayed by $\tau_l/2$ relative to the input signal.

It is simplest to consider the description of events in the frequency domain rather than in the time domain. Relative to the input signal, the left loudspeaker's acoustic output has transfer function $L \exp(j\omega\tau_l/2)$ and the right loudspeaker's output has transfer function $R \exp(-j\omega\tau_l/2)$, where $\omega = 2\pi f$ is the radian frequency. To within a common (and therefore irrelevant) level factor and time delay representing the transmission from the loudspeakers to the nominal center of the

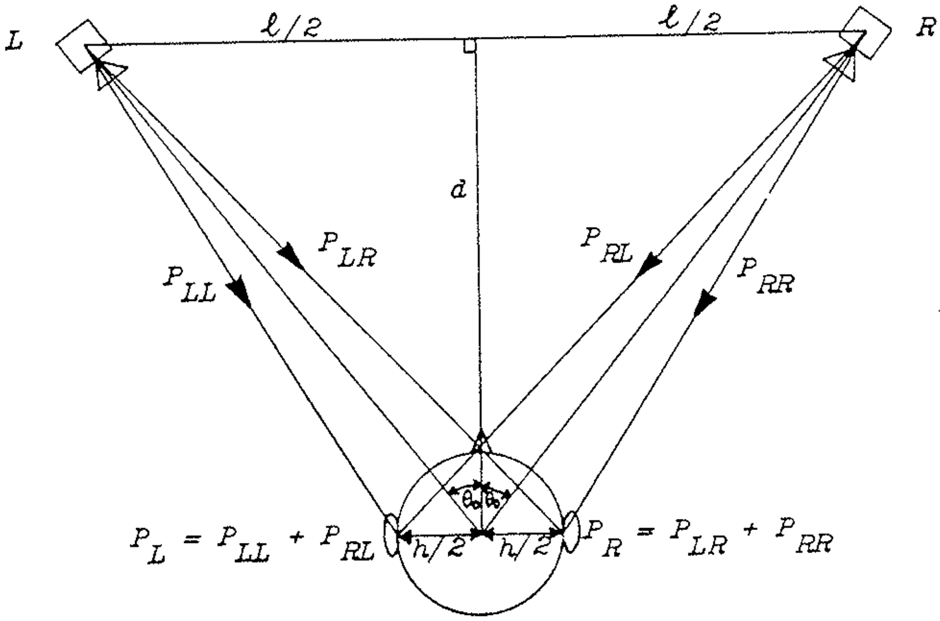


Fig. 1. Geometric relationship betwen loudspeakers (subtending an angle $2\theta_0$) and central listener, showing the component pressure signals generated by each loudspeaker at each ear.

listener's head, the acoustic pressures produced at the ears by each loudspeaker independently are

$$P_{LL} = L \exp\left[j\omega \frac{\tau_l + \tau_h}{2}\right] , \qquad P_{LR} = L \exp\left[j\omega \frac{\tau_l - \tau_h}{2}\right]$$
$$P_{RL} = R \exp\left[-j\omega \frac{\tau_l + \tau_h}{2}\right] , \qquad P_{RR} = R \exp\left[-j\omega \frac{\tau_l - \tau_h}{2}\right] \tag{1}$$

and hence the total pressure at each ear is given by

$$P_L = P_{LL} + P_{RL} = L \exp\left[j\omega \frac{\tau_l + \tau_h}{2}\right] + R \exp\left[-j\omega \frac{\tau_l + \tau_h}{2}\right]$$
$$= (L + R) \cos\left[\omega \frac{\tau_l + \tau_h}{2}\right] + j(L - R) \sin\left[\omega \frac{\tau_l + \tau_h}{2}\right] \tag{2a}$$

$$P_R = P_{LR} + P_{RR} = L \exp\left[j\omega \frac{\tau_l - \tau_h}{2}\right] + R \exp\left[-j\omega \frac{\tau_l - \tau_h}{2}\right]$$
$$= (L + R) \cos\left[\omega \frac{\tau_l - \tau_h}{2}\right] + j(L - R) \sin\left[\omega \frac{\tau_l - \tau_h}{2}\right] . \tag{2b}$$

From Eqs. (2) we can compute the resultant interaural level ratio $|P_L/P_R|$ and phase angle $\arg(P_L/P_R)$;

$$\left|\frac{P_L}{P_R}\right|^2 = \frac{(L + R)^2 \cos^2\left[\omega \frac{\tau_l + \tau_h}{2}\right] + (L - R)^2 \sin^2\left[\omega \frac{\tau_l + \tau_h}{2}\right]}{(L + R)^2 \cos^2\left[\omega \frac{\tau_l - \tau_h}{2}\right] + (L - R)^2 \sin^2\left[\omega \frac{\tau_l - \tau_h}{2}\right]} \tag{3a}$$

$$\arg\left[\frac{P_L}{P_R}\right] = \tan^{-1}\left[\frac{L - R}{L + R} \tan\left(\omega \frac{\tau_l + \tau_h}{2}\right)\right] - \tan^{-1}\left[\frac{L - R}{L + R} \tan\left(\omega \frac{\tau_l - \tau_h}{2}\right)\right] \tag{3b}$$

where we note that

$$-1 \leq \frac{L - R}{L + R} \leq 1 . \tag{3c}$$

All our conclusions will follow from an examination of Eqs. (1)–(3).

We note that if only one loudspeaker is driven (that is, $L = 0$ or $R = 0$), Eqs. (3) correctly predict that $|P_L| \equiv |P_R|$ and $\arg(P_L/P_R) = \mp \tau_h\omega$, a linear phase–frequency relationship. This corresponds to a pure interaural time delay of τ_h, as should occur when listening to a single source (that is, loudspeaker) only. Moreover, when both loudspeakers are driven equally (that is, $L = R$ and $\tau_l = 0$), Eqs. (3) show that $|P_L| \equiv |P_R|$ and $\arg(P_L/P_R) = 0$, as would occur for a real center-front source: double mono produces a central image. In order to predict more general imaging we need to know the relationship between source direction and interaural time delay for single sources. As Fig. 2 shows, for a relatively distant source (as long as we can ignore head diffraction) the desired relationship is

$$\tau = \frac{h}{c} \sin\theta , \tag{4}$$

where θ is the source angle relative to center front, and c is the speed of sound. In particular, for each loudspeaker we have

$$\tau_h = \frac{h}{c}\sin\theta_0 = \frac{hl}{c\sqrt{l^2 + 4d^2}} \qquad (5)$$

provided $h << l$.

To position images between the loudspeakers we have at our disposal the two parameters L/R and τ_l, the interloudspeaker level ratio and time delay, respectively. Of course, assuming that we can indeed produce an image in this way, it will still be necessary to determine what sort of microphone arrangement will capture left- and right-channel signals possessing the necessary L/R and τ_l for each source position. However, it is clear that if Eqs. (3) can be made to reduce to $|P_L/P_R| = 1$ and $\arg(P_L/P_R) = \tau\omega$ for some τ, then this combination of L/R and τ_l will fool the hearing system into believing that it is listening not to two separate sources (the loudspeakers), but to a genuine image in the direction θ predicted by Eq. (4). That is, we will have succeeded in producing credible images not just at center front ($\theta = 0$) and at the two loudspeaker positions ($\theta = \pm\theta_0$), but also at intermediate angles. We shall see that this can be achieved by coincident-microphone techniques, but not by spaced-microphone stereo.

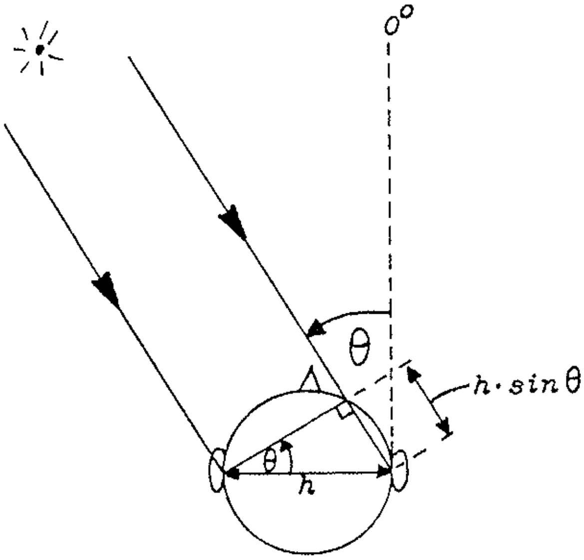


Fig. 2. Geometric relationship when listening to a single off-center source at angle θ.

Before examining these special cases, it will aid understanding to represent the general situation of Eqs. (1)–(3) by means of phasor diagrams for each ear. This is done in Fig. 3. The individual component phasors P_{LL} and P_{RL}, which sum to P_L, and the phasors P_{LR} and P_{RR}, which yield P_R, are shown. The length of P_{LL} equals that of P_{LR}, and the length of P_{RL} equals that of P_{RR}, but these two lengths are not in general equal to each other. In addition, the phasors P_{LL} and P_{RL} rotate in opposite directions as a function of frequency at the rate $(\tau_l + \tau_h)/2$, whereas the pair P_{LR} and P_{RR} also rotate in opposite directions but at the generally different rate $(\tau_l - \tau_h)/2$. Indeed, if $\tau_l = \pm\tau_h$, one of these two pairs of phasors stops rotating entirely, and the acoustic sum at this ear then has constant magnitude and phase, independent of frequency. That at the other ear is still frequency dependent, however. The sum phasors P_L and P_R thus generally differ both in length (interaural level difference) and in phase angle (interaural phase difference).

We must now investigate these differences, and we do so by first examining some special cases.

Intensity Stereo ($\tau_l = 0$)

This case is the basis of Blumlein's original patent [4] and its successors. Being spatially coincident, the microphone (and hence the loudspeaker) outputs differ only in relative level L/R, and so we specialize Eqs. (1)–(3) by putting $\tau_l = 0$. As a consequence, all the component phasors P_{LL}, P_{LR}, P_{RL}, and P_{RR} now rotate at the same rate $\tau_h/2$, and hence, as shown in Fig. 4, the sum pha-

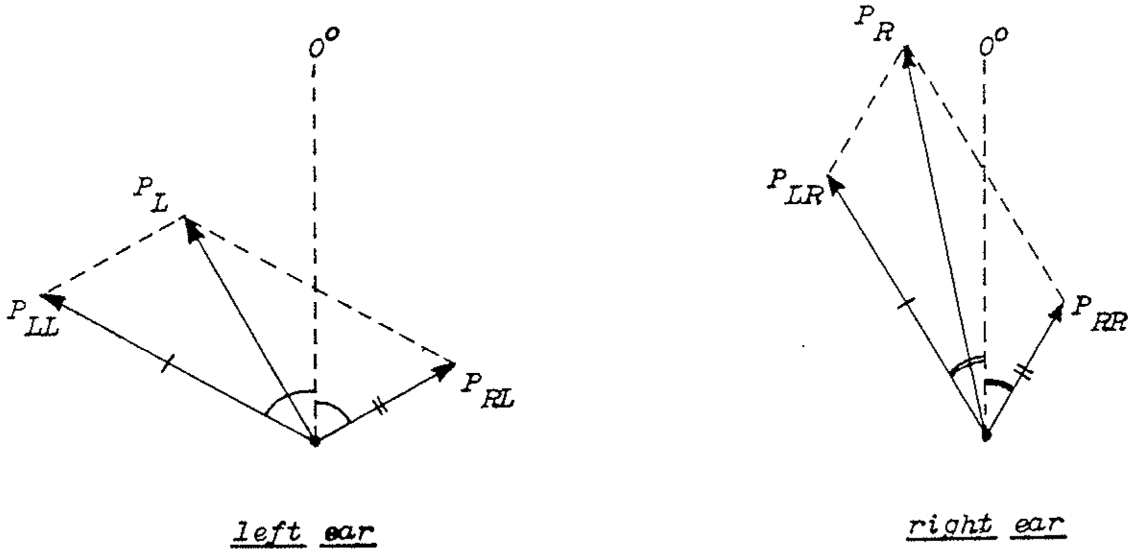


Fig. 3. Phasor diagrams at each ear in the general case of Eqs. (1)–(3), showing how the resultant phasors P_L and P_R are formed.

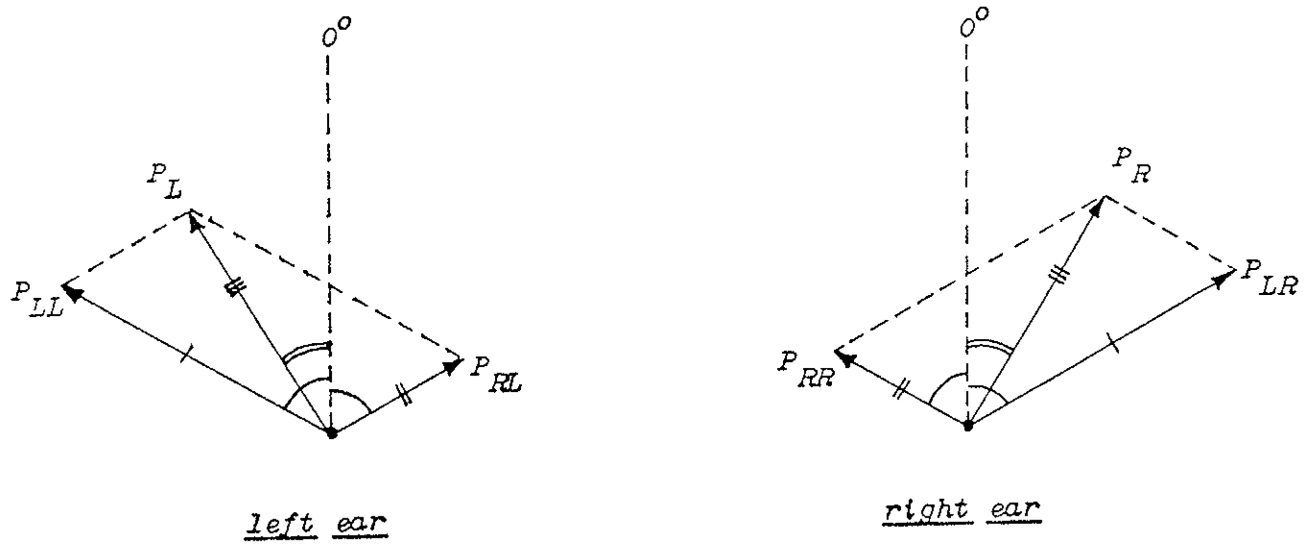


Fig. 4. Special case of Fig. 3 for intensity stereo: $\tau_l \equiv 0$. The phasors P_L and P_R are always symmetrical about the 0° line and of equal length.

sors P_L and P_R always have the same length and rotate in opposite directions at the same rate:

$$|P_L| \equiv |P_R| , \qquad \arg(P_L) \equiv -\arg(P_R) \tag{6a}$$

$$\arg\left[\frac{P_L}{P_R}\right] = 2\tan^{-1}\left[\frac{L-R}{L+R}\tan\left(\frac{\omega\tau_h}{2}\right)\right]$$

$$\approx \frac{L-R}{L+R}\tau_h\omega \ , \ \text{for } \tau_h\omega < 1 \ . \tag{6b}$$

These equations show that, in spite of the level differences at the two loudspeakers, the pressures at the ears are always equal in magnitude at all frequencies as long as head diffraction can be ignored. The interloudspeaker level differences produce an interaural phase difference which, in the low-frequency regime, varies approximately linearly with frequency ω, thus corresponding to an almost constant interaural time delay τ given by

$$\tau = \frac{L-R}{L+R}\tau_h \ . \tag{7}$$

This delay τ as a fraction of the across-the-head delay τ_h is determined by the interloudspeaker level ratio L/R. The approximation $\tau_h\omega < 1$ used in Eq. (6b) is valid to good precision over the whole of the low-frequency regime. To see this, consider an equilateral loudspeaker–listener relationship with l = 3 m and θ_0 = 30°. Then using h = 140 mm and c = 341 m/s, Eq. (5) gives τ_h = 205 μs, and so $\tau_h\omega < 1$ for all frequencies up to 775 Hz. Under these conditions the approximate expression given in Eq. (6b) is accurate to better than 10% for frequencies up to 775 Hz (and to better than 20% up to 1165 Hz).

We thus find the remarkable result that over the whole of the low-frequency regime intensity stereo produces interaural signals which correspond precisely with what occurs naturally when listening to a single real sound source, as discussed earlier. Indeed, combining Eqs. (4), (5), and (7) we can predict that the image produced will appear at off-center angle θ given by

$$\sin\theta = \frac{L-R}{L+R}\sin\theta_0 \ . \tag{8}$$

This is the well-known stereophonic law of sines. Its significance is easy to understand. Since $-1 \leq (L-R)/(L+R) \leq 1$, it follows that $-\theta_0 \leq \theta \leq \theta_0$, that is, the image can be moved smoothly between the loudspeaker positions by varying the relative loudspeaker levels L and R. To complete the chain of reasoning, we show that this is precisely what the Blumlein crossed figure-of-eight microphone arrangement does (and so, for that matter, does a sine/cosine pan pot). As Fig. 5 shows, a source at angle φ from center front produces microphone outputs which depend on φ as follows:

$$L = A\cos\left(\frac{\pi}{4} - \varphi\right) ,$$

$$R = A\cos\left(\frac{\pi}{4} + \varphi\right) \tag{9a}$$

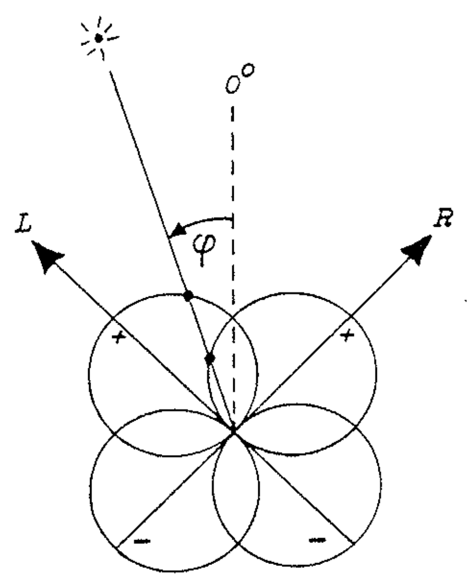


Fig. 5. "Blumlein" crossed figure-of-eight microphones responding to a source at angle φ off center.

and so

$$L + R = \sqrt{2}\, A \cos \varphi \ ,$$

$$L - R = \sqrt{2}\, A \sin \varphi \qquad (9b)$$

(which incidentally are the M and S signals, respectively, of the equivalent forward-pointing/sideways-pointing figure-of-eight M–S microphone arrangement), that is,

$$\frac{L - R}{L + R} = \tan \varphi \ . \qquad (10)$$

Substituting Eq. (10) into Eq. (8) gives the Blumlein "mapping"

$$\sin \theta = \sin \theta_0 \tan \varphi \qquad (11)$$

between source angle φ and perceived image angle θ in the low-frequency regime. The front quadrant $-\pi/4 \leqslant \varphi \leqslant \pi/4$ in the source space is thus mapped smoothly but not linearly onto the stereo stage between the two loudspeakers: $-\theta_0 \leqslant \theta \leqslant \theta_0$. The angular error in relative image position is, however, small.

It is interesting to note that Eq. (11) predicts that sources somewhat outside the front quadrant will tend to image outside the angle subtended by the loudspeakers, and this is found to be correct. But one must be careful in applying Eq. (11) far outside the front (in-phase) quadrant, for up until now we have assumed that L and R are both positive, but now one of L or R will be negative in the out-of-phase side quadrants, and the assumption $|(L - R)/(L + R)| \leqslant 1$, upon which Eq. (11) is based, breaks down.

The 3-dB center-front level error is also easily deduced from Eqs. (2) and (9) in the Blumlein case $\tau_l = 0$, for we find that the individual ear pressures as a function of source angle φ are given by

$$\begin{aligned} |P_L|^2 &= |P_R|^2 \\ &= 2A^2\left[\cos^2\varphi \cos^2\left(\frac{\omega \tau_h}{2}\right) \right.\\ &\qquad \left. + \sin^2\varphi \sin^2\left(\frac{\omega \tau_h}{2}\right)\right] \end{aligned} \qquad (12)$$

and so, in the low-frequency regime $\tau_h\omega << 1$, we have

$$|P_L| = |P_R| \approx \sqrt{2} A \cos \varphi \ . \qquad (13)$$

So a source at $\varphi = 0°$ is up in level by a factor of $\sqrt{2}$ relative to full-left or full-right sources at $\varphi = \pm 45°$.

In the high-frequency regime our assumptions of no head diffraction and $\tau_h \omega < 1$ break down and more sophisticated treatments are necessary, but even here it is found that stable images can be produced by the Blumlein intensity stereo technique.

Time Differences Only ($L \equiv R$)

A fair approximation of the situation pertaining to spaced omnidirectional microphones placed relatively distantly from the sound sources is that the sound pressure levels at the two microphones are equal (that is, $L \equiv R$) and that the perceived effects are due principally to path-length differences to the two microphones (that is, $\tau_l \neq 0$). Referring to Eqs. (1)–(3) we see that now

$$P_L = 2L \cos \left[\omega \frac{\tau_l + \tau_h}{2}\right] \qquad (14a)$$

$$P_R = 2L \cos \left[\omega \frac{\tau_l - \tau_h}{2}\right] \qquad (14b)$$

and so

$$|P_L| = 2L \left|\cos\left(\omega \frac{\tau_l + \tau_h}{2}\right)\right| \qquad (15a)$$

$$|P_R| = 2L \left|\cos\left(\omega \frac{\tau_l - \tau_h}{2}\right)\right| \qquad (15b)$$

$$\arg \left[\frac{P_L}{P_R}\right] = n\pi \ , \qquad (15c)$$

$$n = \ldots, -2, -1, 0, 1, 2, \ldots$$

at all frequencies, as long as head diffraction effects can be ignored. This implies that the pressure levels at the ears vary cyclically with frequency, and at different rates $|\tau_l \pm \tau_h|/2$ at the two ears, and that these two signals are always either precisely in phase or in antiphase. This is clearly shown by the phasor diagrams of Fig. 6, since the equality of the lengths $|P_{LL}| = |P_{RL}| = |P_{LR}| = |P_{RR}|$ now ensures that P_L and P_R are alternately parallel and antiparallel as a function of frequency. Furthermore, whenever $\omega = (2n + 1)\pi/(\tau_l + \tau_h)$, we have a total pressure null at the left ear (that is, $P_L = 0$), and similarly at the right ear (that is, $P_R = 0$) whenever $\omega = (2n + 1)\pi/(\tau_l - \tau_h)$, $n = \ldots, -2, -1, 0, 1, 2, \ldots$. We thus have two comb filters, one at each ear, each with a different comb tooth spacing. These signals rise and fall in level and flip polarity at different rates for each source position τ_l. This is what I call "total phasiness." In natural hearing one never has a situation in which the direct sounds from single sources produce signals at the two ears which are in total antiphase over a whole range of frequencies. The interaural phase–frequency relationship of Eq. (15c) is a staircase function with treads at multiples of 180°, whereas in natural hearing (and also in intensity stereo in the low-frequency regime) one has a linear relationship. Such antiphase signals are quite unlocalizable in the low-frequency regime (unless the one ear signal is very much larger than the other, when the sound appears to come from one of the two sides, which is quite unnatural and unintended), and even at high frequencies where interaural phase does not matter, we have no image formed as apparent sound direction swings from side to side with frequency as a function of the cyclic variation of $|P_L|$ and $|P_R|$.

Up until the first pressure null [that is, over the frequency range $0 \leqslant \omega < \pi/(|\tau_l| + \tau_h)$] the first ear signals are in phase, and one might hope that over this narrow frequency range some credible image might be formed. At very low frequencies $P_L = P_R$ and a center-front (monophonic) image is formed irrespective of the value of τ_l. But for $\tau_l > 0$, for example (that is, source to the left of center), the image moves from the center (at $\omega = 0$) to the right as ω increases, since $P_L < P_R$ over this range. It might be as well to put some numbers on these effects. For an intermicrophone spacing of, say, 1 m, the first pressure null, and hence the start of bad phasiness, can occur as low in frequency as 160 Hz. It is, of course, even worse for wider spacings.

It is sometimes claimed that if omnidirectional microphone spacings comparable to the ear spacing are used, good stereo imaging can be obtained, but this is clearly false. It is true that this reduces the maximum value of τ_l that can occur, and so raises the frequency of the first pressure null, but it leaves one with an essentially center-front monophonic image over most of the frequency range below this. The limiting case, of course, is "coincident omnis."

Experiment does, however, show that imaging *is* possible on transients over the approximate range -3 ms $< \tau_l < 3$ ms. This is beyond the scope of the theory presented here. These images are relatively broad and very listener and source material dependent. Only to a limited extent can the transient imaging negate the disastrous effects of phasiness over the bulk of the frequency spectrum. One cannot produce a genuine stereo image by interloudspeaker time differences alone.

With widely spaced microphones one has both level and time differences in the captured signals, and so finally it remains for us to examine the general case to see whether the poor results of the case presented in this section can be improved upon.

Level and Time Differences

We now return to the general Eqs. (1)–(3), illustrated by the phasor diagrams of Fig. 3, where now we generally have $L \neq R$ and $\tau_l \neq 0$. A simplification of the formula for $\arg(P_L/P_R)$ in Eq. (3b), as in Eq. (6b), is still possible, but now over a very much smaller frequency range:

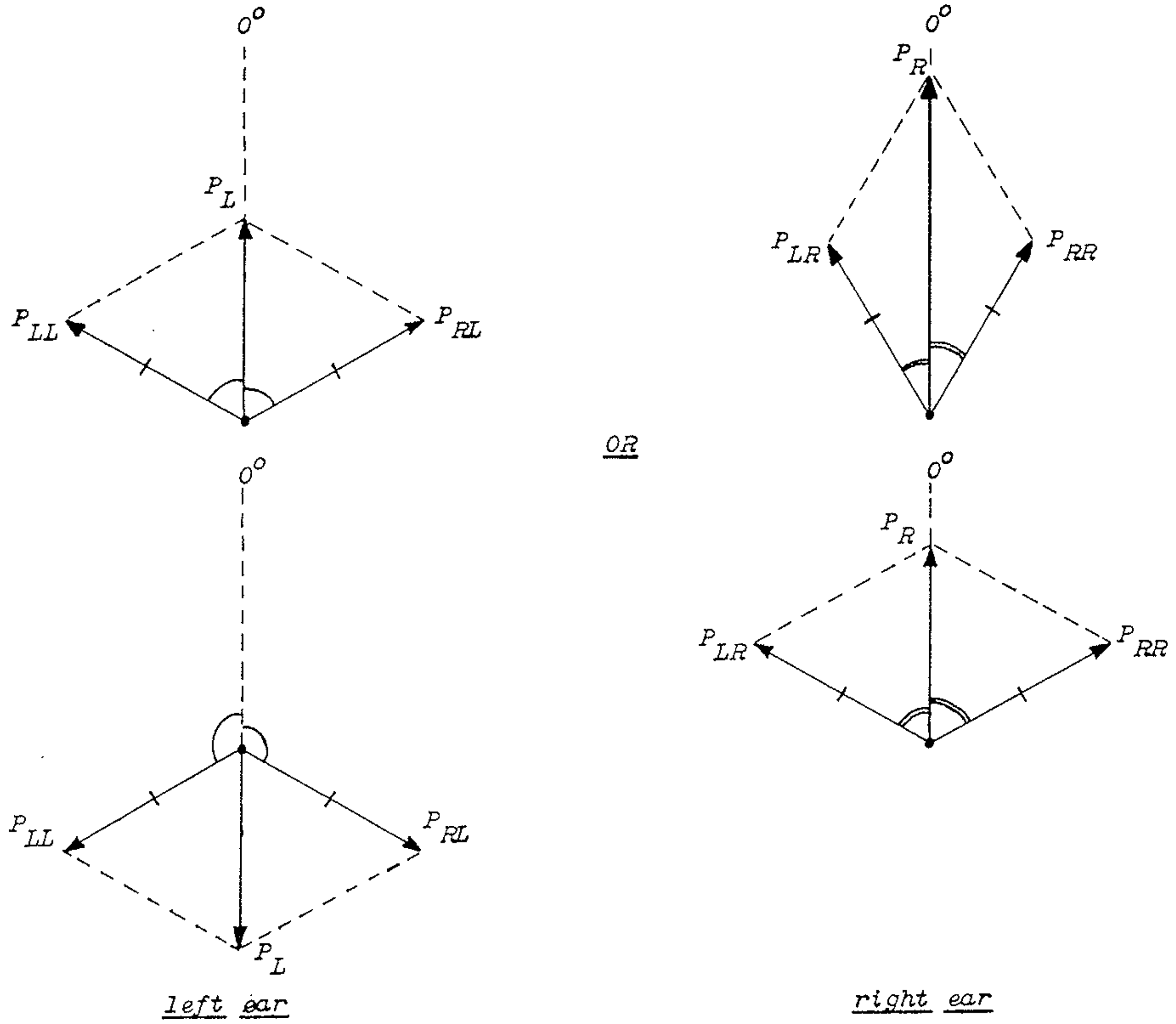


Fig. 6. Special case of Fig. 3 for time-based stereo. Phasors P_L and P_R are always either parallel or antiparallel.

$$\begin{aligned}\arg\left[\frac{P_L}{P_R}\right] &\approx \frac{L-R}{L+R}\,\frac{\tau_l+\tau_h}{2}\,\omega \\ &\quad - \frac{L-R}{L+R}\,\frac{\tau_l-\tau_h}{2}\,\omega \\ &= \frac{L-R}{L+R}\,\tau_h\omega\ , \\ &\quad \text{for } (|\tau_l| + \tau_h)\omega < 1\ .\end{aligned} \tag{16}$$

It is interesting to see that the *same* approximately linear-phase relationship as for Blumlein stereo still holds, but over a frequency range which progressively reduces as $|\tau_l|$ increases. In other words, interloudspeaker level differences still produce the same interaural time-delay differences as we found before, but now generally only over a small portion of the low-frequency regime. For example, a modest value of $\tau_l = 1$ ms would produce a linear-phase relationship only up to about 130 Hz, and $\tau_l = 3$ ms only up to about 50 Hz. Thus even moderate intermicrophone spacings on the order of 1 m or so can be considered to result in coincident-pair behavior only over a very restricted portion of the low-frequency range. And even then, significant level differences are necessary to produce image movement away from center. For example, the ORTF configuration with spacing of 170 mm can image on the basis of intermicrophone level differences only up to around 250 Hz, that is, only over one-third of the low-frequency regime. It may seem surprising that to first order the intermicrophone delay τ_l does not contribute to the interaural delay exhibited in Eq. (16). For frequencies above the linear-phase range, the interaural phase ceases to bear any coherent relationship with natural hearing. When this occurs over the bulk of the low-frequency regime where interaural phase is significant, the image disintegrates and becomes phasey.

Considering next the interaural level differences, by Eqs. (2) we have in the general case

$$|P_L|^2 = (L+R)^2\cos^2\left[\omega\,\frac{\tau_l+\tau_h}{2}\right] + (L-R)^2\sin^2\left[\omega\,\frac{\tau_l+\tau_h}{2}\right] \tag{17a}$$

$$|P_R|^2 = (L+R)^2\cos^2\left[\omega\,\frac{\tau_l-\tau_h}{2}\right] + (L-R)^2\sin^2\left[\omega\,\frac{\tau_l-\tau_h}{2}\right]\ . \tag{17b}$$