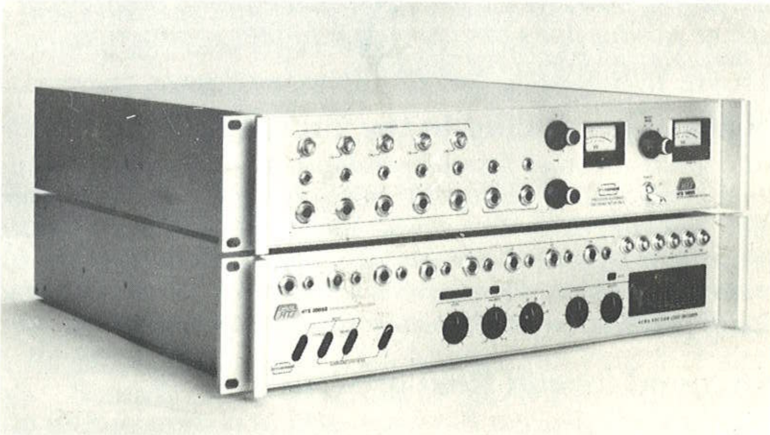
SHURE HTS®

Stereosurround®

A Better Way to Produce Stereo Sound

A Guide to the Stereosurround® Audio Production Format



Stereosurround Q & A

Answers to a number of important questions about the Stereosurround Audio Production Format.

Stereosurround - Three Common Misconceptions

Brief discussions on misconceptions about the capabilities, applications, and benefits of the Stereosurround Audio Production Format.

Music, Broadcast & Video Production Applications of the Stereosurround Audio Production Format

A technical description of the Stereosurround Audio Production Format based on papers presented at the Audio Engineering Society and the National Association of Broadcasters.

Q. Stereosurround A.

A Better Way to Produce Stereo Sound

What is the Stereosurround® Audio Production Format?

The Stereosurround Audio Production Format is a practical, Dolby Stereo® compatible 4-2-4 matrix encode/decode process derived from motion picture audio techniques that allows the generation of a more realistic sound field using four discrete-like reproduction channels that can be transmitted or stored on two conventional audio channels.

How many consumers can enjoy productions in the Stereosurround format?

Today, there are over 3 million Stereosurround format-compatible decoding devices in consumer homes. These range from inexpensive add-on devices for existing stereo systems, to television receivers with built-in surround sound, to elaborate component Home Theater systems that rival the best motion picture theaters.

What are the benefits for consumers who have a two speaker stereo system, or single speaker monaural system?

Productions in the Stereosurround format offer conventional stereo listeners a wider, fuller sound image than is possible with conventional stereo production techniques. Stereosurround format productions also assure a higher production quality for every listener including those with monaural systems.

What are the costs of producing in the Stereosurround format?

Assuming that Stereosurround format production equipment is being added to an existing two channel production facility, the costs are quite low considering the sonic benefit to be gained. Typical costs range from about $12,000 to $20,000 depending upon the sophistication of the production monitoring amplifer/loudspeaker system. Once the production equipment is in place, there are no additional production costs unique to this format. No additional production staff or microphones are required.

Is the Stereosurround format field proven?

YES! The Stereosurround process has been used in a wide variety of live, live-to-tape, and post-production programs. These include two complete seasons of Chicago Cubs baseball on Superstation WGN, Super Bowl XXIV, NBC-TV's Saturday Night Live and Nasty Boys series, the 1989 and 1990 Grammy Awards programs, the NCAA Final Four Playoffs on CBS-TV, and many other sports and entertainment events and corporate productions for such companies as Coca-Cola, Volkswagen, and McDonald's. Stereosurround was used in FM Superstation WFMT's broadcast of the Grant Park Symphony's Fourth of July Celebration aired live from Chicago. In the recording studio, it has been embraced by New York's Power Station and was recently used in Telarc's production of the "Spies'" latest compact disc release. In all of these diverse applications, Stereosurround has proven itself to be reliable and straightforward to use while at the same time significantly adding to the excitement of the audio production.

How strong is the future for the Stereosurround production format?

The Stereosurround format has been developed to allow the production of both audio-only and audio-for-video programs with significantly improved spatial reality. It is compatible with a variety of existing multi-channel decoding hardware (Dolby Stereo, Dolby Surround®, Ultra Stereo®) as well as two speaker stereo and single speaker monaural playback equipment. Because two channel audio storage and transmission systems will be the dominant consumer format for many years to come, this system offers a much better approach to creating more realistic sound fields with four-channel reproduction than is possible with any two speaker stereo system while retaining complete compatibility with two speaker and single speaker reproducing systems.

Stereosurround®

A Better Way to Produce Stereo Sound

Stereosurround®
Three Common Misconceptions

1.

Two speaker stereo is a pure process that cannot be significantly improved.

Not True. *Stereo from the Greek word* stereos *means solid, firm, or three-dimensional. When applied to audio recording and reproduction systems, it means realistic or lifelike. From the perspective of a typical consumer, however, the term stereo has come to mean an audio program or system designed for two loudspeaker channels, left and right. The fact that just two reproduction channels are used is an accepted compromise due to the widespread availability of two transmission and storage channels.*

Early psychoacoustical studies of audio recording and reproducing systems indicated that three reproduction channels could greatly improve audio realism. Further, the addition of at least one rear channel was found to provide a more robust and convincing sense of spatial reality. By adding a front center and surround channel to a conventional stereo system, stable front stage images could be created for a larger listening space while also creating interior or all-around ambience images. Sounds could be made to move not only across the front of the sound stage, but between front and back.

Today it is practical to implement the findings of these early studies through the use of sophisticated 4-2-4 (four channels into two channels and then back to four channels) matrix encoding and decoding techniques using existing two channel transmission and storage systems. The resulting programs offer the benefit of significantly greater realism when reproduced over systems with three front channels (left, center, right) and a single rear channel (surround), while at the same time being completely compatible with traditional stereo and mono playback systems.

The term Stereosurround® has been created as an easily understood format identifier for programs based on this technology. Just as mono and stereo serve as format identifiers for programs created for single speaker and two speaker reproduction, Stereosurround identifies programs that have been created for three front channels and one rear channel, while at the same time being compatible with traditional stereo and mono systems. Such proprietary processes as Dolby Stereo® and Ultra Stereo® are Stereosurround-compatible production techniques.

2.

Stereosurround is useful only for motion picture or television audio productions.

Quite the contrary. Our experience in working with many audio mixers and music producers is that the Stereosurround format has a great deal to offer in music recording. Of all audio applications, music mixes benefit the most from the unambiguous stereo imaging and depth of field capabilities offered by Stereosurround production techniques. This is true even when the primary audience will be using two speaker stereo for playback.

There are many benefits for the audio mixer as well. For example, a fundamental difficulty in producing a compatible stereo/mono mix is controlling the relative phase of the program elements. Elements that tend to be out-of-phase are reduced in amplitude in the mono mix and, in some cases, disappear altogether, whereas in the stereo mix they are always present, but difficult to localize. One of the many benefits of this production format is that a Stereosurround monitoring system makes it significantly easier to identify and correct those elements of the mix contributing to such problems. This is possible because all program elements are localized accurately from in front of to behind the listener according to their relative phase. As a part of this process, a Stereosurround encoder provides an effective tool for spatially repositioning these elements to achieve the desired result in both the mono and stereo mix. In addition, extensive production experience has demonstrated that an exciting and compatible Stereosurround program can also be created.

3.

The Stereosurround process will force me to compromise my present successful production techniques.

Not True. This is an understandable concern, particularly when past success has been based on established production values. However, the beauty of the Stereosurround production format is that you do not have to give up these production values in committing to the process. As a matter of fact, once you configure your production facility for the Stereosurround format you have not given up the ability to produce in single speaker mono or traditional two speaker stereo. In other words, the format is such that you can use as much or little of the process as you wish.

Even if you decide not to encode in the Stereosurround audio production format, you should evaluate the Stereosurround decoding of your two speaker stereo production. This will allow you to know what it will sound like on the dramatically growing number of Stereosurround playback systems on which it will be played.

Music, Broadcast & Video Production Applications of the Stereosurround® Audio Production Format

Robert B. Schulein
General Manager, Shure HTS
Evanston, Illinois 60202-3696 U.S.A.

The Stereosurround audio production format is a practical, compatible 4–2–4 matrix encode/decode process that allows the generation of more realistic sound fields using four discrete–like reproduction channels that may be stored or transmitted using two conventional audio channels. A brief technical description of the encoding and decoding process will be presented followed by an analysis of a variety of field and studio production techniques. Specific program examples discussed will include sports, music entertainment, made–for–television dramas, and commercial advertising productions. In addition, the technical robustness of the process will be analyzed with respect to its tolerance to amplitude and interchannel delay errors.

0 INTRODUCTION

The Stereosurround production format has as its roots, techniques developed by the motion picture industry that were designed to add more realism and excitement to the audio-visual entertainment experience. Nearly 50 years ago, the motion picture "Fantasia" was produced using a multichannel sound process (Fantasound) that involved not only three reproduction channels behind the screen but loudspeakers located to the sides and overhead [1]. Since that time, numerous multichannel motion picture formats have been developed with a common objective of creating a more realistic sound field for the listener/viewer. By using at least three front channels, and one rear or side channel (L = Left, C = Center, R = Right, S = Surround) as shown in Figure 1, significant benefits can be gained over conventional two-speaker reproduction.

With the addition of the surround and center channels, stable center images can be created for a large listening space as well as interior or all around ambience images. In addition, sounds can be created that move not only across the front sound stage but between front and back. These additional channels of information contribute significantly to establishing a greater sense of spatial reality resulting in an increased sense of being in a real-world sound field. In spite of the impressive performance of many multichannel systems, practical considerations such as exhibition costs and multiple print inventories have prevented the wide acceptance of any discrete format. What has evolved, however, are two four-channel matrix oriented discrete sounding formats commonly referred to as Dolby Stereo® and Ultra Stereo® based upon the patents of Scheiber [2], [3], [4]. In addition to providing a discrete sounding four-channel production format, these processes have been accepted as both two-speaker stereo and single-speaker mono compatible. Unlike the matrix techniques developed as a part of the consumer quadraphonic products of the early 1970's, these processes use high performance steering logic techniques that allow the creation of unambiguous images over a wide listening/viewing area [5], [6]. These production techniques are in widespread use today in the feature motion picture industry and are commonly used in major film releases.

The matrix techniques associated with the motion picture industry had limited additional applications until the development of new audio and audio/video technologies associated with broadcasting and home entertainment systems. Specifically, the BTSC–MTS Stereo Television System, stereo satellite television services, the VCR with FM (Hi-Fi) audio tracks, the optical video disc with PCM digital audio tracks, and the compact disc all represented new media with robust characteristics capable of accurately processing matrix encoded audio programs. Initially, programming consisted of matrix encoded feature films that were broadcast or transferred to video cassette and video disc for reproduction by consumers. As a result of these productions, a home theater oriented professional and consumer awareness has developed, along with a market for decoding and related audio equipment to accurately reproduce these programs [7]. A particular driving force behind the acceptance and interest in this multichannel audio process has been the fact that a picture is involved and, hence, the listener/viewer is more aware of the spatial realism that can be created. This should be contrasted to the situation that existed with the earlier consumer quadraphonic techniques. These techniques were developed around music-only programs and had a restricted ideal listening space due to the lack of a center channel.

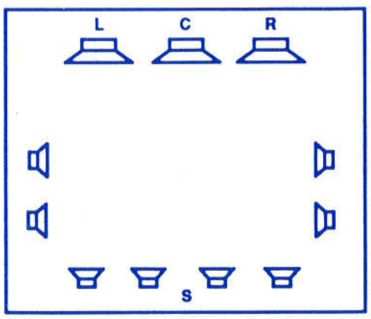


Fig. 1. Loudspeaker locations for a multichannel reproduction system with improved spatial reality

Because of a growing consumer interest in stereo audio with video as well as an increasing number of home theater type playback systems, there is a desire among program producers to create a greater variety of productions with stereo audio in addition to major motion pictures. Specific examples include sports events, musical variety programs, made-for-television dramas, music videos, commercials, and audio-only programs. The Stereosurround production format has been developed for and is currently being used for such productions. It is designed to give maximum performance benefit to a growing number of consumers with sophisticated decoding equipment, while at the same time, providing a compatible production for two-speaker stereo and single-speaker monaural listening. In addition, experience has demonstrated that the Stereosurround production format makes it possible for audio engineers to mix more efficiently and accurately due to the inherent ability of the process to reveal common production errors.

1 THE STEREOSURROUND PRODUCTION FORMAT

1.1 General

The Stereosurround production format uses a 4–2–4 matrix process oriented around a four input, two output encoder, and a two input, four output decoder as shown in Figure 2. In program production, the output of the encoder is recorded or sent to the transmission chain as well as fed to a decoder to monitor the production. Encoded programs are optimized for playback through a decoder with similar characteristics.

The characteristics of the encoder and decoder are based on the matrix portions of the Dolby Stereo and Ultra Stereo production process, but have been refined for modern broadcast and consumer audio storage and transmission systems.

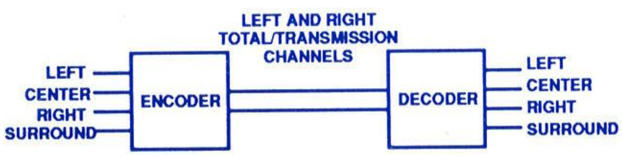


Fig. 2. Stereosurround Encoder/Decoder Block Diagram

The Stereosurround production format has, however, been designed to allow the creation of programs that are highly compatible with existing Shure HTS Acra-Vector logic, Dolby Surround Pro-Logic, and Ultra Stereo® decoding equipment as well as being downward compatible with Dolby Surround® and other decoding systems designed for three front channels and one rear channel. In addition, such productions are compatible with two-speaker stereo and single-speaker monaural reproduction systems.

1.2 Encoding and Decoding—Theory and Performance Capabilities

The matrix portion of the Stereosurround production format is a nonsymmetrical encode/decode process in the sense that the decoding portion involves adaptive matrix modification (steering) logic circuits to make the system perform as if four independent discrete channels are being used. Figure 3 indicates the basic elements of the encoding process in block diagram form.

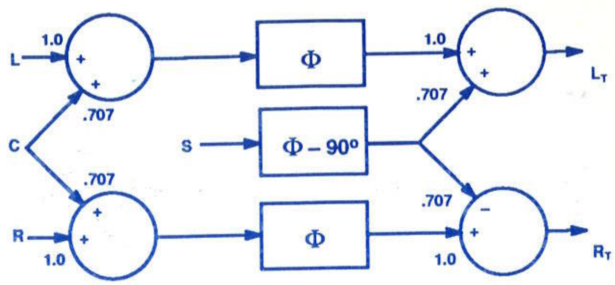


Fig. 3. Simplified Stereosurround Encoder Block Diagram

Mathematically, this process may be described as follows:

$$L_T = L + .707\ (C + jS)$$

$$R_T = R + .707\ (C - jS) \qquad (1)$$

where L_T and R_T represent the left and right total or transmission encoded signals, and L, C, R, and S represent the left, center, right, and surround input signals, respectively. It is important to note that the elements in the block diagram denoted by Φ and Φ – 90° are all-pass networks characterized as having a flat amplitude response, but a frequency dependent phase response of Φ(f) and Φ(f)– 90°. Analysis of the complete set of encode/decode equations reveals that such phase shift functions are required in order to allow for an interior or all around sound mixing capability. The decoding matrix (before directional enhancement or matrix modification) is defined by the four equations:

$$L' = L_T$$

$$R' = R_T \qquad (2)$$

$$C' = .707\ (L_T + R_T)$$

$$S' = .707\ (L_T - R_T)$$

where the prime notation is used to denote that the decoded signals are only approximations to the original L, C, R, and S inputs. The need for matrix enhancement or modification becomes clear when the decoded signals (2) are represented in terms of the original input signals. This can be seen from equations (3), where equations (1) and (2) have been combined to produce:

$$L' = L + .707\ (C + jS)$$

$$R' = R + .707\ (C - jS) \qquad (3)$$

$$C' = C + .707\ (L + R)$$

$$S' = jS + .707\ (L - R)$$

These equations clearly show that each decoded signal contains the original input signal, as well as its spatially adjacent channel's signal reduced in level by 3 dB. This can also be shown graphically as in Figure 4, where the 3 dB arrows indicate the areas of significant crosstalk, and the infinity arrows indicate channels producing no crosstalk. For example, if L was the only input to the encoder, a portion of L (attenuated by 3 dB) would appear in both the center and surround channels, and none of L would appear in the right channel.

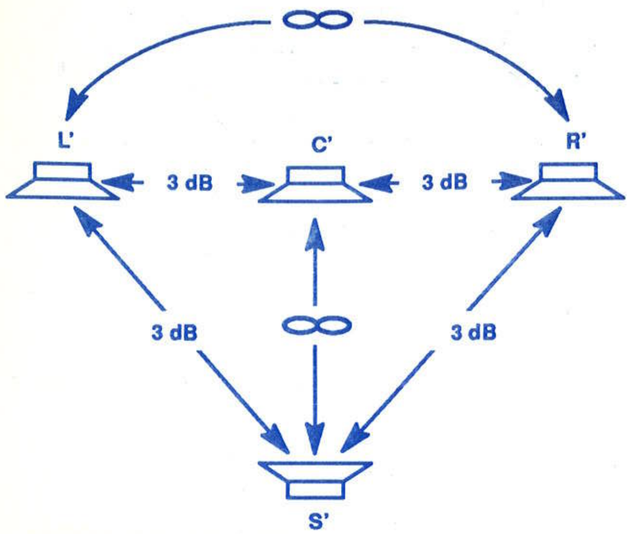


Fig. 4. Graphical representation of 4-2-4 matrix decoding crosstalk prior to directional enhancement

From a perception standpoint, the situation depicted in Figure 4 is a step in the right direction in that all reproduced sounds are most dominant in the intended directions. Subjectively, however, this level of decoding accuracy cannot produce unambiguous localization perception over a wide listening area, particularly with respect to centrally placed images. In order to significantly improve this situation, matrix enhancement circuitry is employed in the decoder. Functionally, this circuitry serves to reduce the undesired crosstalk signal components (3 dB crosstalk elements) as a function of the dominant content of the program. Since it is not possible to achieve full independence of the four channels at all times under all conditions, it is necessary to make system compromises. In the case of the Stereosurround format, compromises have been made with artistic considerations in mind. For example, the integrity of the center channel is most important, and a number of steps are taken to keep center channel information out of the surround channel. On the other hand, the ability of the system to simultaneously enhance a center and right or left discrete image is found not to be as important and, consequently, is not possible without some crosstalk. The decoding action can, however, very quickly alternate between these image locations with a low level of crosstalk giving the perception of very discrete behavior. As a result of these necessary compromises, it is common practice in Stereosurround format productions, to monitor through the encode/decode process so that any localization difficulties resulting from a particular mix can be heard and dealt with. Based on these artistic considerations, the Stereosurround production format can be viewed as a technological art form with the following capabilities:

1. Accurate localization of dialog, music and sound effects across a three-loudspeaker sound stage for listeners seated in a wide listening area.

2. Unambiguous rearward localization of specific dialog, music, and sound effects.

3. The creation of interior sounds which surround the listener with environmental or ambience effects from all directions.

4. Smooth panning capabilities across the front sound stage and between rear, interior, and front locations.

5. Simultaneous reproduction of localized or panned sounds and interior sounds.

From a compatibility standpoint, it can be seen from Equation (1) that a two-speaker reproduction system will accurately decode left and right information with the center channel appearing in both outputs at a reduced level creating a phantom center image. The surround information will also appear in both channels at a reduced level, but with opposite polarity which typically results in a diffuse image. A monophonic system will reproduce the sum of L_T and R_T which will contain all signals with the exception of the surround channel input. Although this may seem to be an unacceptable situation, it is not the case in actual practice. Since hard pans to the surround position are generally not required for artistic reasons, but rather pans to and from the surround position, a great deal of compatibility exists. For such pans, the monophonic listener simply hears the production element get loud then soft, or soft then loud. This generally correlates very well with the screen action in audio/video productions. In practice, a compatible surround image can be created by panning near, but not all the way to the surround position, in which case a portion of the surround element of the mix will appear in the sum or monaural signal. As previously indicated in Equation (3), decoding circuits that do not employ enhancement logic will decode the production, but with reduced separation (location accuracy) between channels.

1.3 Encoding and Decoding—Hardware Considerations

A practical realization of the Stereosurround format encoder described in Figure 3 is shown in block diagram form in Figure 5. Two groups of fixed-position inputs in addition to a panable input have been found to be useful for a variety of productions due to the limitations of many mixing situations. In addition, the panable input allows the creation of a custom mix bus position that might otherwise be difficult to create with available mixing controls. Precision all-pass encoding networks are required in order to guarantee acceptable (greater than 40 dB) encoded channel separation. Of particular importance are center to surround, and left to right crosstalk performance. An 80 Hz low frequency shuffling circuit has been included, on a switchable basis, for those applications where low-frequency out-of-polarity or difference information must be avoided. Functionally, this circuit attenuates $L_T - R_T$ information below 80 Hz without affecting the $L_T + R_T$ component. As an aid to system setup, a metering circuit and input trim controls are provided to allow the calibration of all inputs.

All inputs on the encoder are full bandwidth including the surround channel. In practice, however, the decoded surround channel is often limited to approximately 7 kHz by the characteristics of typical surround channel consumer circuitry. This is due in part to the bandwidth of cost effective time delay circuits employed in the surround channel signal path whose primary function is to put the listener in the acoustic center of the listening space, and to increase perceived center-to-surround channel separation.

In some applications, it may be desirable to use a companded noise reduction system within the encoding and decoding process other than with the L_T and R_T signals. Such noise reduction may be desired due to a particular limitation or characteristic of the storage or transmission medium. This is commonly done in the Dolby Stereo and Dolby Surround processes with the surround channel, due to limitations in the optical film recording process. In this case, a modified B-type noise reduc-

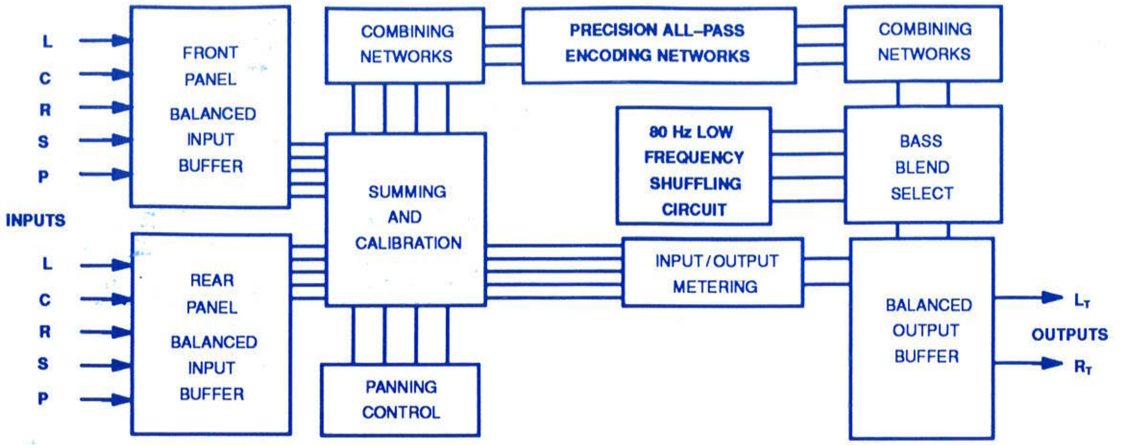


Fig. 5. Stereosurround Encoder System Signal Processing

tion encoder is placed in the surround input of the encoder and a complementary decoder in the surround output of the decoder. If a discrete multichannel system was being used, this would indeed be a logical application of the noise reduction process. Due to the fact, however, that a matrix system is being used, proper companding will not occur because the output of the noise reduction encoder is not what is fed to the decoder. A more technically accurate solution would be to use noise reduction on the difference signal after the encoding process. This could be accomplished by deriving both $L_T + R_T$ and $L_T - R_T$ from the encoder outputs, sending $L_T - R_T$ through the desired noise reduction encoder and then generating a new L_T and R_T for transmission or storage. Similarly, prior to decoding, $L_T + R_T$ and $L_T - R_T$ would be generated and $L_T - R_T$ processed through the complementary noise reduction decoder and finally, L_T and R_T would be regenerated for multichannel decoding. The MTS stereo television system uses a variation of this technique, where sum and difference signals are transmitted instead of left and right, and the difference signal is companded to optimize performance given the restricted dynamic range of its transmission channel.

A system block diagram of a Stereosurround format decoder is shown in Figure 6. The initial sum and difference matrix decoding generates the L', C', R', and S' signals as described in equations (3) prior to processing by the "Directional Enhancement Signal Modification" circuitry. It should be noted that the enhancement process involves a decision-making side chain that does not process any of the actual output audio signals, but rather controls linear sum and difference circuits in the "Directional Enhancement Signal Modification" block where the L", C", R", and S" signals are generated. This approach to decoding serves to significantly reduce the amount of circuitry by which the actual signal is processed. An additional element of the process is the derivation of a low frequency (below 80 Hz) or subwoofer output which is not processed by the directional enhancement circuitry. Since the position of low frequency information below 80 Hz is difficult to localize in rooms, this output serves in effect as a nondirectional fifth channel suitable for creating dramatic impact in productions. A further benefit of this channel is that it allows the use of relatively small loudspeaker enclosures for the L, C, R, and S reproduction locations.

Due to the fact that listeners seldom sit in the acoustic center of the loudspeaker array shown previously in Figure 1, and are typically biased towards the surround loudspeakers, a digital time delay is used to, in effect, move the surround loudspeakers farther back into the listening space. An additional benefit

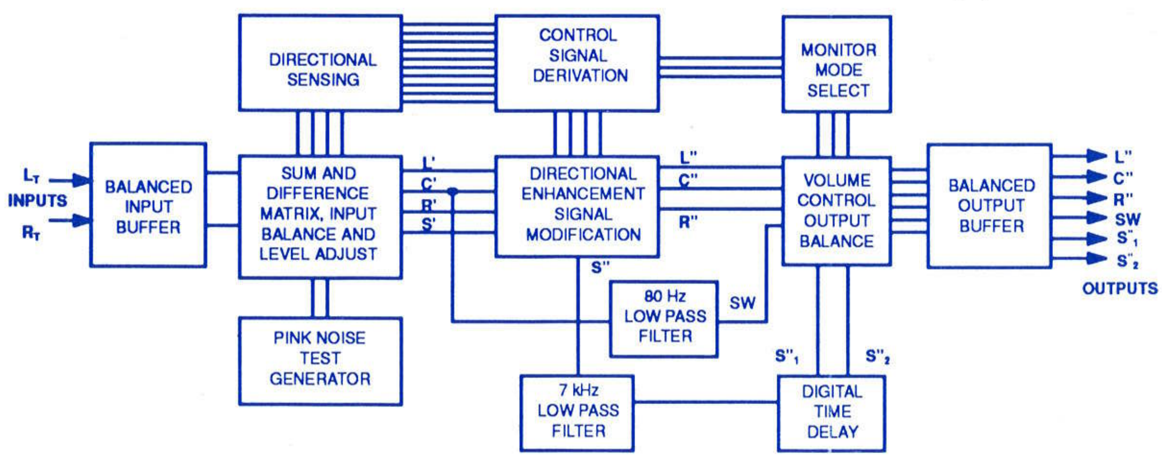


Fig. 6. Stereosurround Decoder System Signal Processing

of this form of signal processing is to further reduce the perception of any crosstalk from the center channel into the surround channel by virtue of the precedence effect. In addition, since some listening environments are long and narrow, two separate surround outputs can be made available with different selectable time delays to more accurately accommodate a larger number of listeners. Finally, a monitor mode selector is provided to allow productions to be monitored using Stereosurround format decoding with and without directional enhancement, as well as using two-speaker stereo and single-speaker monaural reproduction systems. Figure 7 is included to show the design characteristics of an actual encoder (top) and decoder (bottom).

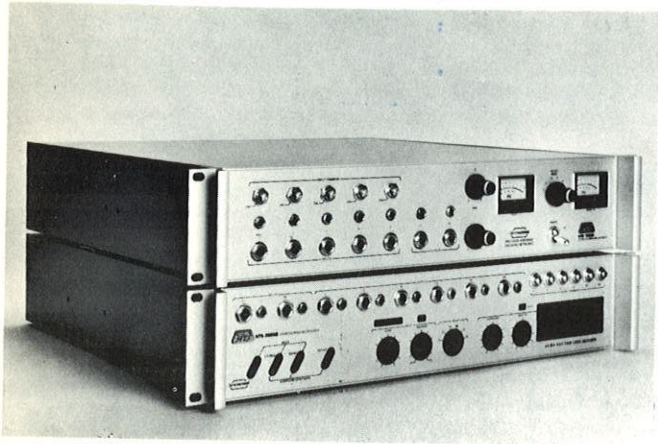


Fig. 7. Stereosurround Encoder (top) and Decoder (bottom)

1.4 Production Considerations

The Stereosurround production format is ideally suited for a post-production environment because it allows the mixing engineer to refine the mix so as to obtain maximum benefit from the process. Excellent performance is also obtainable from a live or live-to-tape production environment provided the audio elements are of a predictable nature, and known mixing compatibility problems are avoided.

A basic element of any production is the establishment of a monitoring environment that will allow the mixing engineer to properly audition the Stereosurround mix. Such a setup is shown in Figure 8. Ideally, all three front loudspeakers should be the same, or at least have very similar tonal characteristics, and should be located at the same listening height. The center loudspeaker should be as close as possible to the video monitor (if used) and located just above or below the screen. Because of the center channel, the left and right loudspeakers can be more widely spaced than for conventional two-speaker stereo productions. As a consequence, it may be desirable to audition the two-speaker stereo compatibility of a Stereosurround production with the left and right loudspeakers moved inward.

Even though a single surround channel is produced in this format, it is highly desirable to reproduce it with a group of loudspeakers so as to create a diffuse sound field. Four loudspeakers are shown in the figure; however, two are frequently used in small production environments. The speakers should be positioned so that the mixer is free to move around with no significant change in the perceived level of the surround channel.

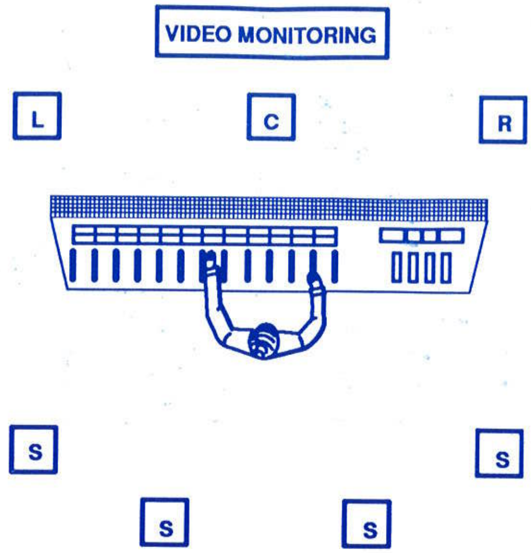


Fig. 8. Stereosurround Loudspeaker Monitoring Configuration

Best results are generally obtained with loudspeakers located high off the floor (5 to 8 feet) so as to reduce the effect of other listeners disrupting the sound field as they move near an individual loudspeaker. A practical check on the surround reproduction channel is to move around the desired listening space, checking for level uniformity while reproducing a pink noise source.

Since productions using the Stereosurround format are capable of generating sound fields that vary between being excessively reverberant to nearly anechoic, an ideal listening environment should be as acoustically absorptive as practical. As an example, excellent results have been obtained in control room environments with midband reverberation times varying between .25 and .3 seconds. Finally, when completing any mix or starting any live production, the monitor system should be objectively or subjectively balanced, using a test generator that alternately cycles a band limited random noise source between the L, C, R, and S decoder outputs.

Figures 9 and 10 indicate typical equipment block diagrams for post-production and live or live-to-tape production environments. A common element of both situations is that of monitoring the encoded signal, as it goes to tape or transmission, through a decoder at the production site. The primary difference between these two production environments is the source of input program material and the need for time code and synchronization between audio and video recording equipment.

Due to the fact that the encoding process is linear, more than one encoder can be used during a production provided that the L_T and R_T outputs of each decoder are summed as a part of the decoding and monitoring process. This capability is of particular value in made-for-television or video productions where there may be a need to alter the production mix at a later time, particularly with respect to additional language releases. In such cases, three encoders are used, one each for dialog, music, and effects (DME), and their outputs recorded on a six-track format. Similarly, in a multi-site live production environment, separate encoders may be used for each location and combined at the production mixer location. In certain live production environments, it may be necessary to delay the L_T/R_T audio mix to properly match the video program. A typi-

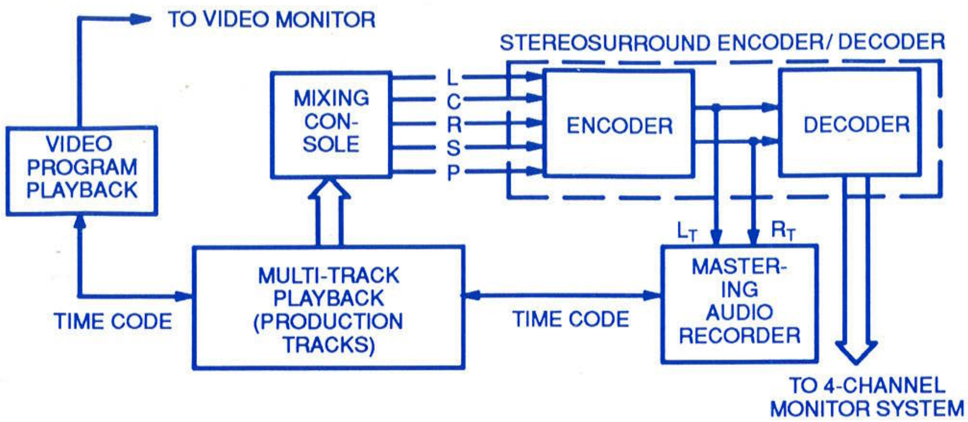


Fig. 9. Stereosurround Post-Production Equipment Block Diagram

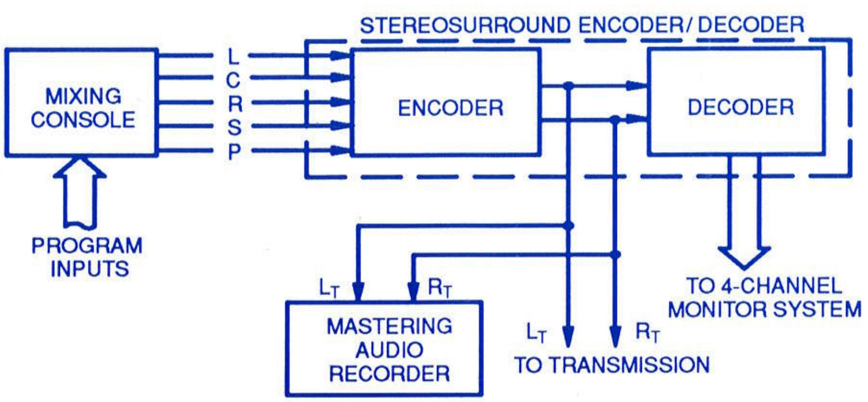


Fig. 10. Stereosurround Live or Live-to-Tape Production Block Diagram

cal situation might involve a pre-recorded audio mix that performers are reacting to as they listen to its playback over a sound reinforcement system. If the acoustic delay due to the distance between the sound system and performers exceeds approximately 40 milliseconds, the audio will appear to lead the video creating an unrealistic result. In order to correct this condition, it is necessary to delay L_T and R_T by precisely the same amount so as to maintain the integrity of the mix. A typical solution is to use a two-channel digital time delay in which a common clock is used to control both channels. Figure 11 is included to show a typical live field-production environment commonly found in mobile audio/video production facilities.

1.5 Mixing Techniques

Developing a mix using the Stereosurround production format is very similar to two-speaker stereo and single-speaker monaural mixing, once microphones have been selected and placed, signal processing has been set, and the configuration of the mixing board and source assignments have been made. In a typical set up, the input buses and outputs of the encoder are connected to the mixer patch bay, such that each input fader can be assigned to or panned between any two input buses. With this condition established, the signal controlled by each fader can then be panned in the "mixing space" to the desired position as the mix is being monitored. Based upon Stereosurround format productions to date, a number of mixing techniques and observations have emerged:

1. Due to the fact that 4–2–4 matrix systems are not discrete, and that localization problems can occur, start by mixing the most dominant sound elements first.

2. When using compression or limiting on dialogue or a vocalist, avoid processing any additional ambient signals through the same compressor/limiter, so as to avoid the dominant sound causing the ambient sounds to pump up and down. This result is particularly bothersome with a multichannel reproduction system.

3. Keep in mind that it is possible to continuously pan signals between the front and surround locations, and that it is not necessary to pan a program element completely to the surround position in order to get a noticeable rearward or surround sense to the mix.

4. When mixing music, start by mixing sound elements directly to the L, C, and R positions and effects to the interior position using separate effects devices for each element. Once the mix has been roughed in, further adjustments should then be considered.

5. If problems occur regarding localization interactions between two or more sound elements, consider panning one

of the signals more toward the interior position, altering its timing, or reducing its amplitude.

6. When using multiple microphones in situations where two or more microphones may pick up the same signal, confirm the polarity of each microphone between its acoustic input and the mixing console output to the encoder. This will greatly reduce the possibility of undesired out-of-polarity information being mixed into the interior or surround position.

7. In many situations where a center oriented vocal element is competing with loud ambience element (sports announcer versus crowd), it is desirable to "widen" the vocal element. This may be accomplished by either panning the element towards the surround position or by using a small amount of stereo synthesis returned to the left and right front positions. Similarly monaural effects associated with sports productions such as baseball "bat cracks" and football player contact sounds can be given more dominance in the final mix.

8. MS and XY stereo microphone techniques can be integrated into Stereosurround format productions provided proper selection of microphone patterns are made. As an example, a popular MS microphone configuration using a cardioid mid pattern and bidirectional side pattern allows the creation of equivalent XY combinations that place sounds properly in the front and rear sound stage. For an in depth discussion of this compatibility, see the reference by Ross [8].

9. Realistic crowd or ambience mixes can be created by using multiple groups of microphones placed so as to sample distinctly different portions of the production environment. Once placed, the outputs of these microphones are then positioned at different points in the "mixing" space by panning between the left or right and surround encoder input buses.

Fig. 11. Stereosurround Live Sports Production Environment

1.6 Technical Requirements for Recording and Transmission

Stereosurround format productions depend upon the relative amplitude and phase integrity of two channel transmission and storage systems in order to maintain a high level of performance. Variations introduced in either of these parameters during recording, playback, or transmission in essence "re-encode" the production. Fortunately, audio recording and transmission technology has been moving in the direction of improved performance in this regard. Consumer audio formats such as the CD and DAT recorder are extremely robust. Similarly, audio/video formats such as VHS-HiFi, S-VHS-HiFi, Beta-HiFi, and Laser Video Disc technology offer extremely stable relative amplitude and phase performance. In addition, professional digital audio recording systems and cassette based audio/video formats such as D1, D2, Betacam SP, and MII provide a high level of performance. From a broadcast perspective, FM Stereo, and the BSTC-MTS stereo television formats are quite good due, in part, to the need for monaural compatibility.

Those systems offering lesser performance typically involve analog recording techniques where recording and playback azimuth calibration is required to provide good program interchangeability. Errors typically introduced by such devices generally involve phase or interchannel delay with relatively minor amplitude discrepancies. Based upon our laboratory studies, interchannel delay errors of 60 microseconds or less are acceptable. Once errors begin to exceed this value, high frequency components of speech sounds, which are typically encoded for center front reproduction, become surround channel information, and are correctly decoded as such to the detriment of the listening experience.

As far as amplitude errors between channels are concerned, transmission requirements are really no different than those associated with two speaker stereo production techniques. Practically speaking, however, a Stereosurround format monitoring system makes it much easier for an audio mixer to detect a channel imbalance due to the ease by which localization errors can be detected. In addition, crosstalk or channel separation requirements are similar to those associated with conventional two speaker stereo. At least 15 dB channel separation is required for good subjective performance, which is generally not a problem. As channel separation is degraded, the program simply becomes "more monaural."

Looking at the transmission and reception aspects of the MTS System has shown that certain factors can impair the proper transmission and reception of encoded programs. Those problems are typically perceived as localization errors involving front channel leakage into the surround channel, program modulated hum and noise in the surround channel, and excessive difference or surround channel information. A number of factors responsible for these problems are described in detail by Gibson [8] and have been reduced by improvements in transmitter and receiver design. In particular, incidental carrier phase modulation (ICPM) of the video signal into the audio during both transmission and reception can be a problem. Many of these problems are significantly reduced when multipath distortion is reduced or these signals are transmitted over a well-designed cable distribution system.

1.7 Future Improvements

A consideration of fundamental importance in developing the Stereosurround production format has been compatibility with existing consumer and professional two channel reproduction systems while at the same time offering a means of creating a significant improvement in spatial reality. The success and acceptability of the format, consequently, depend heavily on maintaining this compatibility with future system improvements. As an example, the bandwidth of the surround channel is not limited by the encoder, but rather by decoder and delivery system technology. Future programming is expected to take advantage of this fact resulting in productions suitable for playback on both wideband and restricted band surround channels, much the same as audio programs are now produced for both FM and AM broadcast.

2.0 Summary

The Stereosurround audio production format has been described from both a theoretical and practical standpoint. The process has been developed to allow the production of both audio only and audio-for-video programs with significantly improved spatial reality. It is compatible with a variety of existing multichannel decoding hardware as well as two-speaker stereo and single speaker monaural playback equipment. Because two-channel audio storage and transmission systems will be the dominant consumer format for many years to come, it is felt that this format offers a much better approach to creating more realistic sound fields with four-channel reproduction than is possible with any two-speaker stereo system, while retaining complete compatibility with two-speaker and single-speaker reproducing systems.

3.0 Acknowledgement

Of the many individuals involved with the development of the Stereosurround production format, the author would like to particularly thank John Bullock, Mark Gilbert, Stephen Julstrom, and John Owens for their dedication and numerous contributions.

4.0 REFERENCES

[1] L. Blake, "Mixing Dolby Stereo Film Sound," Recording Engineer/ Producer, Vol. 12, p. 68, February, 1981.

[2] P. Scheiber, "Quadrasonic Sound System," U.S. Patent 3,632,886, January 4, 1972.

[3] P. Scheiber, "Multidirectional Sound System," U.S. Patent 3,746,792, July 17, 1973.

[4] P. Scheiber, "Stereophonic Sound System," U.S. Patent 3,959,590, May 25, 1976.

[5] M. Willcocks, "Directional Enhancement System for Quadraphonic Decoders," U.S. Patent 3,944,735, March 16, 1976.

[6] D. Robinson, "CP200–A Comprehensive Cinema Theater Audio Processor," J. SMPTE, Vol. 90, p. 778, September, 1981.

[7] S. Julstrom, "A High Performance Surround Sound Process for Home Video," J. Audio Eng. Soc., Vol. 35, No. 7/8, July/August, 1987.

[8] D. Ross, "A Practical Approach to Applying the MS Stereo Microphone," NAB Broadcast Engineering Conference, March 30April 3, 1990, Atlanta, Georgia.

[9] J. Gibson, "Effects of Receiver Design and Transmission Impairments on Audio Signal Quality in the BTSC System for Multichannel Television Sound," J. Audio Eng. Soc., Vol. 34, No. 9, September, 1986.

SHURE HTS®

Stereosurround

For further information contact:
Shure HTS, 222 Hartrey Avenue, Evanston, IL 60202-3696
1-800-25SHURE • Fax (708) 866 2279

AL1055

Printed in the U.S.A. 5K 9/90