Surround Sound Beyond 5.1

Inspiration

I have been fortunate to be interested in audio over the past 50 years and have been witness, student and consumer of the progress. From the days of 78 rpm mono records to the LP stereo vinyl and to the advent of digital CD audio and now moving to the brink of multi-channel surround. For natural music recording the surround channels offer improved presentation with regard to imaging parameters such as perception of distance and spatial depth, spatial impression, envelopment (due to reverberation and/or ambient non-located and non-reflected sound, e.g. applause). I, for one, have moved to present image rather than a collection of sounds or instruments.

The term “spatial imaging“ includes both, imaging of spatial impression as well as imaging of spatial depth. In the concert hall the major cues for perception of depth or distance are delay and level (in relation to the direct sound) of early and late reflections (reflection pattern), and of reverberation. There is only some distance information in the direct sound – namely the relative loudness of different musical sections, and the possible presence of some high frequency roll-off due to air absorption. In a recording the directional cues in the direct sound have even less importance than in natural hearing. In a recording the spatial image has more importance than in natural hearing. The imaging of spatial depth is related to the design of indirect sound rather than direct sound. On the other hand, direct sound provides the relevant localization cues with respect to direction

The perceived auditory spatial impression, which is caused by actual or reproduced indirect sound of a room comprises two attributes of the auditory event. The first is ”reverberance” described as a temporal slurring of auditory events caused by late reflections and reverberation. Reverberance can be understood as the perception of a background sound stream which is primarily perceived in the time gaps between foreground sound events. Only in the case of continuous foreground sound streams the image is broadened by the reflected (spatially diffuse) sound. The second is auditory spaciousness, which is identified as characteristic spatial spreading of auditory events due to late reflections. If the events have a short rise-time compared to the arrival of major lateral reflections they will be perceived as sharply localized. If the rise-time is slow the images will be broadened.

Ambisonics was the development of a group of British researchers about 20 years ago. They developed a surround sound system that would enable a musical performance to be captured for replay in a conventional living room in which, as far as possible, the original performance (real concert hall or multi-track mix) would be recreated. The system was named Ambisonics (surround sound). The Ambisonic system records 4 channels of information known as W, X, Y and Z. These channels represent Omni/pressure information, front/back differential, left/right differential and up/down differential. The system was designed from the beginning to enable recordings to be made with a special surround microphone, (the Soundfield Microphone). Ambisonics production equipment generates a four-channel signal, called 'B-format', that embodies all the information in the soundfield, resolved into left/right, front/back, up/down information and an omni- mono sum signal. B Format is the metadata of sound with complete information and can be presented in any way. It is economical requiring only 4 channels.

Interestingly, as early as the mid-seventies, Ambisonics included the capability to record and reproduce height information, which even now, is not part of surround sound practice. It adds to the realism of a surround system as much as rear speakers. I have found that the extension of each dimension extends the emotional involvement in the music, i.e. mono to stereo, stereo to surround and surround to 3D surround.

For replay, the B-format signal is fed to a decoder that derives the number of speaker's outputs for mono/stereo, 5.1 surround or beyond. In this replay system each speaker contains virtually all elements of the recording, but with different relationships. The speakers work together to recreate the acoustic and ambience of the original recording, i.e. sound field reconstruction. My first thoughts are that each of the ambisonic speakers would be at the 45, 135, 225, 315 degree angles. It is well known from stereo that one is unable to generate phantom images between a pair of speakers that have an included angle of greater than 60 degrees. One of the problems of Quad and 5.1 is that there is a hole in the sides and rear. Thus I have chosen 45 degrees as an optimum, adding speakers at 0, 90, 180 and 270 degrees. In order to reproduce the vertical dimension I place an up and down speaker at each location.

After much thought the speaker locations I have chosen are at +/-22.5, +/-67.5, +/-112.5, +/-157.5. This arrangement also gives the best fit of equal distance from the center in a square room. Since the time of arrival is critical in surround this gives you the least time delay between arrival time of all the speaker to the central point. i.e. the radius to each of the speaker from the center should be the same. Having an upper and lower speaker at each location reproduces the vertical component. Thus we have two rings of eight speakers. In essence we move from a line(stereo) to a flat pentagram(5.1 surround) to a cylinder. Since it is difficult to present a sphere I compromised with a cylinder. Early research at Bell Labs showed that more channels gave better reproduction of the sound.

Tools

The tools that I have chosen for my work in B format are the Scope Platform from Creamware. I have chosen this because there are the atoms and modules that one can use to create new devices and an easy GUI interface. I have found this very useful in the creation of B format tools to use in the decoding and manipulation of the soundfield. You can also build devices that will enable you to check out your ideas.

Scope is a combination of hardware and software for audio development. The hardware consists of a PCI board that contains 15 Shark DSP processing chips that installs in a PC computer. It contains 24 channels of ADAT ODI I/O and 2 MIDI cables as well as a word clock input. The ADAT ODI interfaces are 8 channels, 24 bit. It supports sampling frequencies of 32, 44.1, 48 and 96 kHz. There is 32 megabytes of SDRAM on board for memory intensive signal processing. It has 32-bit audio-bus architecture. The software is a graphic user interface that allows you to design audio devices in DSP using the graphic interface.

I have designed and built in Scope 5, 6, 8 speaker layouts for planar surround as well a 12 and 16 speaker layouts/decode circuits for 3D surround. The planar surround consists of speakers placed on a single plane or ring. The 5 speaker decode is designed with speakers at 0 degrees, +/-45 degrees and +/-120 degrees. The 6 speaker decode is designed with speakers placed at 60 degree spacing between the speakers. The 8 speaker decode is designed for speakers with 45 degree separation between the speakers. The 12 speaker decode is two rings of speaker with 60 degree separation. There is an upper ring and a lower ring to give the vertical component of the decode. The 16 speaker decode which I have chosen as my standard is two rings of speakers with 45 degree separation between the speakers with one ring in the up position and one ring in the down position.

There are indicators used in the decode unit which consists of concentric rings of green/signal lights and red/overload lights concentrically placed in their correct location. There is a master volume control and a front/rear bias control.

I have also designed a B format manipulation tool that allows one to be able to rotate the azimuth +/-180 degrees. You will also be able to tumble the soundfield +/- 90 degrees. The amount of gain in the front/rear differential signal, the left/right differential signal and the up/down differential signal is controlled by volume controls. This will allow you to set the apparent spread of the sound source.

There are several psychoacoustic phenomena that need to be considered in this methodology of recording. The first is shelf filters added to the B format decode. Our ears are more easily confused by very high frequency intensity difference localization and are rolled off. Also our ears prefer additional low frequency separation and Greisinger believes this adds to the envelopment that we enjoy in surround sound. There are adjustable shelf filters to accommodate these ideas.

I am planning a mixer that will allow the mixing of B format and other sources together and also a mixer that will allow the mixing of B format environment plus a stereo pair that will be added to the frontal speakers.

The Process

To give you some insight I would like to share my experience in surround sound. My initial experience was listening to music I recorded in B format (to allow later adjustment of ambience) decoded to 4 channels and found it so much superior to stereo including greater emotional involvement in the music. Without surround the music became less interesting and less dynamic and less detailed. Only a short while later I added the center front channel also decoded from the B format. I wanted more! I wanted to decode the vertical dimension on reproduction. I made a special 10 channel tape and at the AES in San Francisco 2 years ago demonstrated a 5 channel surround with vertical dimension, i.e. a dual 5.1 system. There was no volume control of the speakers but we were able to assess the sound and were amazed at the realism, lack of ‘listening to speakers’, and a large sweet spot. One thought is that you hear a single speaker as an entity only when it produces loudness equal to the loudness in the room. 1 in 16 means you can move closer before you hear a single speaker as an isolated speaker.

The related psychoacoustic principles should be understood as phenomena of spatial hearing governed by specific laws and thus requiring suitable types, configurations and locations of microphones, as well as distinct handling of delay, inter-channel correlation and level balancing of direct/indirect sound. I have found that I am recording two elements and that they are the performance and the environment. This concept of recording is in agreement with the philosophy of Gunther Thiele in his paper that he published (available on the Internet) on correct psychoacoustic recording. He too records using the concept of recording the performance and the hall. The philosophy of Ralph Glasgal who has designed ambiphonic reproduction using a stereo sound source and adding the environment by convolution or artificial reverberation is also in agreement with the concept I am now using.

The dominant cue to create distance, spatial depth and spaciousness is the natural pattern of early reflection in the region of 15-50 ms. In addition with reverberation the spatial impression can be perceived. Furthermore, reverberation can generate perception of reverberance and envelopment.

Moreover, in combination with reverberation, real spatial impression can be produced by reproducing reflections and reverberation through loudspeakers outside the frontal stereophonic imaging area, notably to the side of, and behind, the listening area. The stereophonic quality changes form a simulated to a real impression of spatial depth if the lateral reflections are delivered by surround loudspeakers and actually arrive at the listener from later directions. Experimental results reported show that the amount of spatial impression depends on the angle of sound incidence of lateral reflections. The results are of practical importance because they demonstrate that reflections from the side are the most effective way of achieving spaciousness. In contrast, early reflections in the median plane are disadvantageous.

The environment is recorded using the soundfield microphone. The soundfield microphone provides the basis of the recorded environment. By decoding the soundfield into 3D Surround the environment can be correctly presented. By adding the performance to this you will be able to balance the direct sound of the performance with the ambient sound of the environment in many different venues. I find that you need to correct the time alignment so that the arrival time of the direct signal from the performance and the environment
Is the same.

The recording of the performance can be done in several known manners with directional stereo microphones or close miced instruments or a combination of the two. Basically this is a stereo recording and it is probably better to use a stereo technique such as X/Y cardiod, M/S, ORTF or OCT with spot microphones. We should be able to get a good recording of an orchestra or group by this method. I use this to present a stable frontal image that will be added to the environment. This recording of the performance should include little of the environment.

The Future

I believe that the use of B format plus stereo is a universal format and with appropriate playback you can present the stereo image in the front and then decode the B format into whatever format you wish to listen in. It would also make your record collection obsolescent proof and you could have your sound reproduction advance with whatever your means at the moment. You can record 6 tracks of MLP material on a DVD and listen to that material in whatever format you desire. You can also return to the stereo master and add the B format surround by convolution and a consumer could then purchase this and always keep his investment. As a consumer you also would have the ability to adjust the direct/reverberant sound for you particular desires or acoustic space. This could now become a universal format.

In my next generation decoders I plan to have B format input as well as two channel inputs (decoded a two-track stereo) to be added to the frontal speakers with decode to 3D surround in 12 or 16 speakers. This will serve to create a stable center image as well as giving you the direct sound that can be balanced against the reverberant sound. The soundstage will be wide covering the front of the room. The will make the storage be 4 channels of B format material and 2 channels of stereo material. I also believe the mixer will be a hybrid producing both a B format channel as well as a stereo channel.

Thomas T. Chen, M.D.
Custom Recording/Studio C
2220 Broadridge Way
Stockton, CA 95209
(209) 477-5130
Fax (209) 473-6496