Audio Engineering Society

Convention Paper

Presented at the 127th Convention
2009 October 9–12 New York, NY, USA

Automated Assessment of Surround Sound

Richard C. Cabot

Qualis Audio, Inc., Lake Oswego, Oregon, 97034 USA
richc@qualisaudio.com

ABSTRACT

The design of a real time electronic listener, optimized for surround sound program assessment, is described. Problems commonly encountered in surround audio production and distribution are automatically identified, including stereo/mono downmix compatibility, balance, metadata inconsistencies, channel interchange, loudness, excessive or inadequate level, and the presence of hum. Making measurements which correlate with audibility, displaying the results in a form easily understood by non-audio personnel created numerous design challenges. The technology used to solve these challenges, particularly that of downmix compatibility, will be described.

1. INTRODUCTION

1.1. The Problem

Monitoring audio signals through a broadcast chain has long been a job for humans, skilled in audio, well versed in the potential problems and attentively listening to the program on an accurate reproduction system. Particularly in television broadcast, such people are scarce. The recent explosion of television channels and delivery systems has drastically increased the number of programs to be monitored. The shift to surround sound has added additional failure mechanisms such as front/rear channel reversal and compatibility with stereo and mono reproduction. Economic realities have further constrained both the availability of skilled personnel and the acoustic quality of their monitoring environment while reducing the time available to accomplish the task.

1.2. Core Issues

The issues facing professionals and organizations creating and delivering surround programs which drove the development described here were:

- Mixing and monitoring surround is a far more complex and challenging task than it is for stereo programs. There are many more opportunities for error.
- Budgets, both financial and time, are shrinking.
- Personnel are expensive, skilled personnel are very expensive.

- People get tired and bored. Things don't go wrong often (hopefully) so vigilance is difficult to maintain.

- Record keeping is important for post-mortem analysis and for assessing financial accountability. People hate to keep records.

The obvious solution to these issues is automation, substituting an intelligent device for the overworked, expensive, drudgery avoiding humans. Such a device would free skilled personnel for other, more creative tasks.

1.3. Monitoring Functions

An investigation of the problems typically encountered in surround production and delivery yielded the following "requirements list" that drove the product development:

- Signal path failure or "dead channels"
- Level issues: loudness, clipping, "overs"
- Channel swapping or rearrangement
- Stereo and mono compatibility
- Spatial balance
- LFE compatibility
- Hum
- Metadata errors and inconsistencies

Some of these, such as dead channels, clipping and loudness are straight forward to monitor and the technology to do so is well understood. Others, such as hum or stereo and mono compatibility have to-date required experienced personnel using specialized equipment. Compatibility in particular has required the interpretation of visual displays and a technical understanding of the effects of signal phase on the downmixing process.

2. COMPATIBILITY

2.1. The Compatibility Problem

Most motion pictures and prime-time television are produced in surround today. Being the premier format, the version evaluated when awarding Oscars for sound mixing, mixing engineers understandably put their attention on how their content sounds in surround. Though most theaters will reproduce the content in surround, the eventual release on DVD will not experience the same uniformity of presentation. Indeed, as is the case with sound for digital television, the majority of viewers today will experience the audio in stereo and a nontrivial percentage will hear it in mono.

The conversion of surround to stereo or of stereo to mono involves combining channels together, algebraically summing their waveforms. Antiphase signals will cancel when combined, reducing in level or disappearing completely. This can happen when individual channels are accidently inverted. However, the more insidious situation occurs when just one component in a surround mix appears in multiple channels but shifted in phase. This can easily happen when a single source is picked up by multiple non-coincident microphones. When the outputs of these microphones are combined there will be cancellations and the signal level will be reduced. If this happens to an actors voice dialog can become unintelligible and viewers, advertisers and producers get very upset.

2.2. Downmixing

When surround programs are downmixed to stereo the process follows one of a small number of methods. Consider the most common case, downmixing 5.1 surround to stereo and mono in an ATSC (Dolby Digital) environment.

Though the gain factors may take on one of three different values, specified by the metadata, the most commonly used equations are:

$$L = LF + CF/2 + LS/1.4 \quad (1)$$

$$R = RF + CF/2 + RS/1.4 \quad (2)$$

Mono is derived by summing the left and right, giving

$$M = LF + RF + CF + LS/1.4 + RS/1.4 \quad (3)$$

Note that in each case an overall attenuation is applied to maintain peak levels at unity gain to prevent clipping.

The result of these equations is that a center channel signal, the typical location for main dialog, is summed into the left and right channels with 6dB of attenuation.

Consider the effect of an attenuated version of this dialog being routed to the left and/or right front channel but shifted in phase. This could easily result from this

dialog spilling into microphones picking up other actors or positioned to record environmental ambience. Alternately, a signal containing spillover could be mixed to a phantom center, perhaps in an effort to get reverberation on the dialog at issue. This spillover would be expected to occur at a reduced level relative to the primary pickup. However, the center front is attenuated 6dB in the downmix process, putting it much closer in level to the spillover signal. When these combine the dialog would be attenuated, the cancellation becoming more severe due to the attenuation introduced by the downmix equations.

2.3. Lissajous Displays

To better understand the problems of monitoring downmix problems in surround lets begin by reviewing the traditional techniques used when downmixing from stereo to mono.

Mono compatibility has traditionally been monitored with a Lissajous display. The left and right channels drive the vertical and horizontal channels of an oscilloscope. Equipment specifically designed for audio monitoring typically will rotate the display counterclockwise by 45 degrees to make the left channel appear as a diagonal line tilting toward the upper left and the right channel appear as a line tilting toward the upper right.

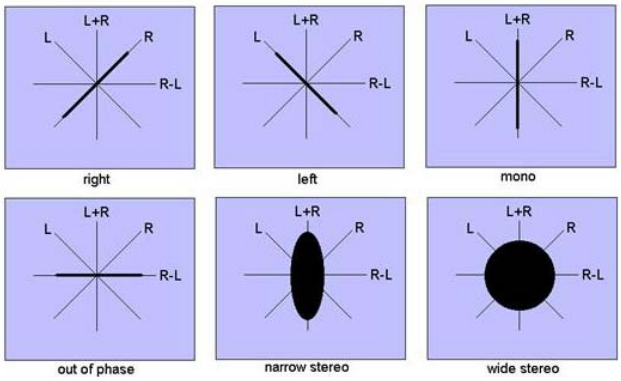

Figure 1 Stereo Lissajous displays from Brice [1]

Interpretation of such a display (we consider the rotated version) is moderately simple, following some basic rules:

- Vertical line: mono, OK
- Horizontal line: mono, BAD
- Vertical "football": stereo, OK
- Horizontal "football", stereo, BAD
- Round ball, stereo, probably OK

With time, most mixing engineers learn to understand the characteristic shapes and spot the signs of trouble. This does presume that they are audio knowledgeable, or want to be, and are actually watching the display when problems occur.

Many manufacturers have taken the graphical display out of the picture (pun intended) by using "correlation" meters. These multiply the left and right channels together and average the result, creating an indicator that is positive when the channels are in-phase and negative when they are out-of-phase. This is usually normalized by the channel levels, creating an indicator scaled between +/-1. A good stereo signal will hover near zero, a good mono signal will be positive. Indications that go very negative represent problem content which will cancel when reproduced in mono.

2.4. Multichannel Lissajous Displays

Now consider the case of surround program monitoring using Lissajous or correlation displays. The first problem in monitoring surround audio compatibility with correlation or Lissajous displays is the sheer number of channel pairs involved. Ignoring the LFE channel for now, a 5.1 program contains 10 channel pairs. A 6.1 program has 15 channel pairs and a 7.1 program has 21. This is illustrated in Figure 2 below.

Many commercial products only analyze neighboring pairs, shown in blue. Others add the LF/RF channel pair, shown in green. The author is unaware of any which display the diagonal channel pairs, shown in purple. Even with this simplification there are 5 or 6 channel pairs to display.

The challenge facing the user is watching that many correlation meters or Lissajous patterns at the same time. With one exception, vendors of such tools have used various schemes to pack these displays onto a single XY display.

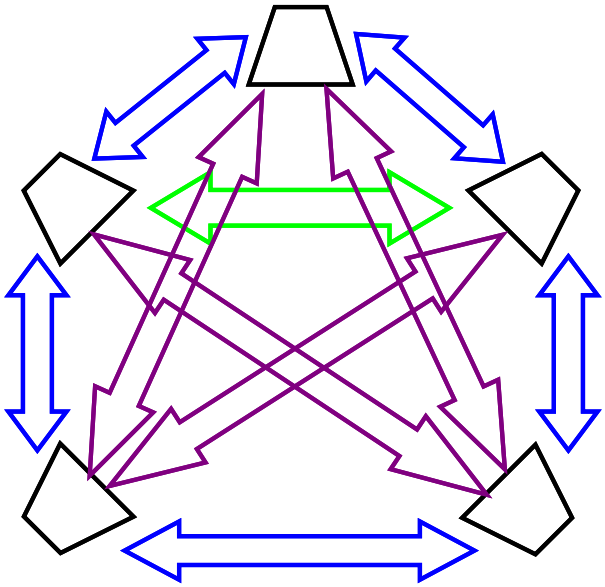

Figure 2 Channel pairs in a surround program

All of these schemes take advantage of the redundancy evident in the four quadrants of the Lissajous display. Since the lower halves of the displays in Figure 1 offer no additional information compared to the upper halves, the display may be truncated or folded at the horizontal axis.

Packing 5 or more of these now truncated displays into a single picture is where the inventive differences between competing displays occur. Some manufacturers use color to get the additional dimensionality required, others use geometric transformations, and some use both. A simple, yet effective, implementation of the geometric transformation approach is found in Brice [1]. Several manufacturers have placed additional indicators alongside, above and below the main multichannel display in an attempt to adequately represent the multiple phase relationships involved.

In the author's opinion, all of these schemes suffer from two fundamental problems.

1. Someone has to be watching the display for it to be useful.

2. The user doesn't really want to know about the phase relationships anyway.

This second point bears elaboration.

2.5. The Compatibility Question

The question the mix engineer really wanted answered is not "What are the phase relationships between all my channels". That's just what they have had the tools to answer since the early days of stereo. With stereo the answer wasn't too complicated to be usable. With surround the answer is generally too complicated to be usable by anyone except a highly experienced audio engineer.

The question most mix engineers really want answered is: "Will it sound the same in stereo and mono as it does in surround?" They mix in surround, know that it sounds the way they want it to, but don't have the time to listen to the whole mix again in stereo, then again in mono.

Driven in part by the realization that existing equipment vendors weren't answering the users question and in part by the desire to have an answer suitable for unattended operation, the author and his colleagues re-examined the problem from basic principles. If the user needs to know how the stereo and mono presentations compare to the original surround mix, that's the comparison to measure.

2.6. The Compatibility Answer

We begin by performing a downmix of the original channels into Left and Right stereo channels using the same downmix equations used by the end-user's reproduction equipment. These are then downmixed to Mono. Armed with these three additional channels the challenge becomes comparing them to the original surround program.

Recall that the fundamental problem is not whether the spatial position of the components will be "correct" in the stereo presentation compared to the surround. Spatial position is entirely irrelevant in the mono case. Rather the biggest concern in motion picture and television reproduction is whether the content will be present at a reasonable approximation to its original level in the surround mix.

This is a highly tractable problem. To solve it we begin by measuring the power spectrum of each of the original surround channels and of the three downmixed channels. This is done in 256 approximately log spaced bands across the 20Hz to 20kHz range. The power spectra of the surround channels are downmixed using

the same equations used to obtain the downmix channels. The result is compared to the power spectra of the downmix channels. They should be equal. If not, it can only be due to phase cancellations in the original downmix operation. Since the power spectra are not phase sensitive their downmix contains all energy present in the original program. The downmix signals are affected by surround channel phasing and represent what is heard by a viewer with stereo or mono reproduction equipment.

When differences between these two versions are found they are grouped on an octave basis and presented to the user. The grouping is performed solely to reduce the volume of data presented and to make the presentation easier to understand.

In summary, the algorithm is

- Compute the downmixed spectrum: $\Sigma X(f)^2$,
- Compute the spectrum of the downmix: $(\Sigma X(f))^2$
- Report the difference as a function of frequency

The difference is shown in dB reduction from the original level as a function of frequency. The loss in each of the two stereo downmix channels is shown as right and left facing arrowheads, respectively. The mono downmix is shown on the same graph with a diamond shape. If all three are at the same dB value the result is a rectangular shape. The resulting display is shown in the lower half of Figure 3.

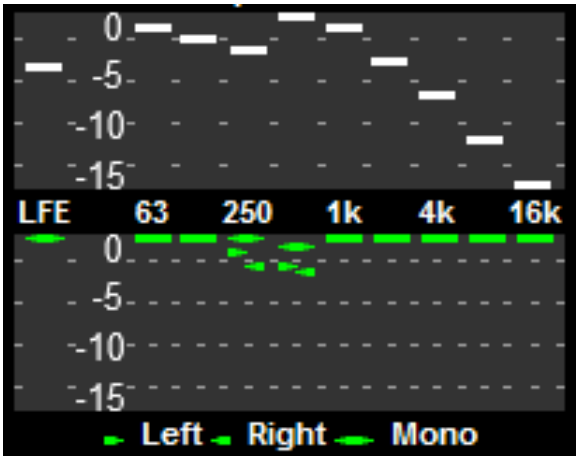

Figure 3 Compatibility Display

The total spectral energy vs. frequency (the sum of all surround channel spectra, excluding the LFE) is displayed above the compatibility graph. This simplifies assessment of the significance of any signal loss, since low level signals are presumably less important and higher losses may be tolerable.

The frequency detail in the compatibility display also aids in assessing the type of content lost in cancellation. If the octaves associated with voice are attenuated it is likely that dialog is affected. Low frequencies are typically associated with sound effects and loss there during stereo reproduction may be more tolerable or even desirable. High frequencies are also associated with effects and may also represent ambience. Again, their attenuation in stereo or mono reproduction is typically of lower concern than loss of dialog.

2.7. Thresholding

Since the original goal was to make a device which would automatically detect problems, the compatibility measurement must be tested, not just displayed. Since users will have differing opinions of what constitutes a problem, there are several selectable parameters used in defining an "error".

The degree of cancellation required to qualify as an error is selectable in 1dB steps from -1dB to -15dB. The frequency range over which this comparison is made is similarly selectable. The comparison may begin at the 63Hz, 125Hz, 250Hz or 500Hz octave band and end at the 2kHz, 4kHz, 8kHz or 16kHz octave band. Since these are octave centers the analysis will extend another 1.4 times lower and higher in frequency, respectively. For example, settings of 500Hz and 2kHz will result in analysis from 350Hz to 2.8kHz, just covering the voice band.

As with any subjective assessment, duration must be considered. Suppose a program contains a brief instant, perhaps due to shifting positions of actors relative to microphones, that there is excessive signal cancellation. This is unlikely to significantly affect dialog or to be noticed by viewers. However, if such cancellation lasted for 30 seconds it most likely would. Consequently the compatibility assessment includes a user selectable duration threshold of 1, 3, 10 or 30 seconds.

Returning to the mix engineers question: "Will it sound the same in stereo and mono as it does in surround?" The measurement described above answers: "These frequencies will drop XdB in stereo, these frequencies will drop YdB in mono". The user decides how many

dB is acceptable, over what frequency range and for how long.

2.8. LFE and Surround Compatibility

Observant readers will notice an additional column at the extreme left labeled LFE. Existing downmix implementations always omit the LFE channel. Whether this is advisable is not really open for discussion, it isn't done and the user isn't given any control over it. This implies that there aren't any compatibility issues with the LFE channel, but that conclusion is wrong.

A problem rarely discussed in the literature concerns the compatibility of the LFE channel with the overall surround mix. The limited size of typical home reproduction environments will result in pressure summation of the surround and LFE channels at the user's listening location. A few manufacturers of reproduction equipment have recognized this and provide control over phase in the bass management crossover region.

When producing content the mix engineer must keep this pressure summation in mind when assessing the balance of LFE in the mix. To assist this assessment an additional downmix compatibility measurement is performed. Using the same technique described earlier for stereo and mono compatibility, the effect of including the LFE on the mono mix is measured. The analysis is restricted to frequencies between 20Hz and 250Hz. Though irrelevant to the mono listener, it represents the audible difference between hearing the full mix in a large space such as a theater and in a small space such as a home environment. The spectrum bar above it represents the level in the LFE channel.

3. SIGNAL ISSUES

3.1. Loudness

A serious problem for any broadcaster is maintaining consistent loudness. If the program is excessively loud the user will reach for the remote control to turn it down, or off. If soft enough that dialog isn't easily intelligible they will similarly reach for the remote to turn it up. If they are motivated to adjust volume very frequently they get rapidly get dissatisfied. Of course the biggest fear among broadcasters is that once the remote is in hand the viewer will simply change channels.

Recent work in the ITU has standardized a new measurement of loudness, BS1770. A diagram of the basic measurement algorithm is shown in Figure 4 below.

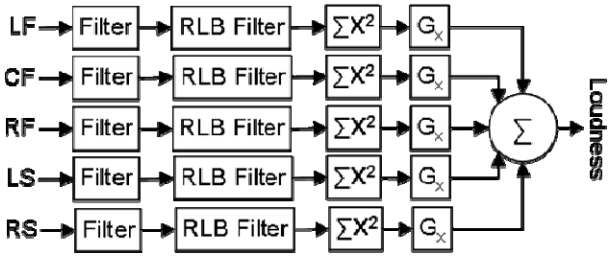

Figure 4 Loudness Computation

The Sentinel implements the measurement diagrammed above, calling it program loudness. It also measures a subset of channels, either the CF by itself or, if CF is not present, the sum of LF and RF. This second measurement is called dialog loudness. It may be used to set dialnorm metadata or compare existing dialnorm values to the actual program. If comparing measured dialnorm to metadata the instrument will warn if the values differ by a user specified amount.

3.2. Clipping

The ITU BS1770 standard also describes methods of measuring true peak level. The signal is upsampled by 4 to a 192kHz rate, assuming 48kHz input, and the peak value measured. This measurement is performed for all 8 input channels as well as the three downmix channels: L, R, and M. The peak levels are displayed, logged and compared to a user defined threshold. If the threshold is exceeded a clipping error is generated.

3.3. Metering

The 8 surround channels and the downmix channels are metered with the users choice of metering standard. These are displayed in real time at 12 readings per second on any Javascript enabled browser. The maximum and minimum metered values over a 1.2 second period are also logged

3.4. Dead Channels

Problems occasionally occur that kill one or more channels in a surround program. Which channel is affected, whether the signal is completely absent or highly attenuated, and the program content itself all

impact the audibility of the problem. Even more important is the attention of facility personnel and the monitoring equipment in use.

Though missing surrounds may be readily apparent to a home viewer with surround reproduction equipment, the typical broadcast in-rack speaker system will completely mask the problem. The only indication to a broadcaster is likely to be low meter readings.

Though missing front channels, particularly center front, should be immediately apparent this again presumes that someone is listening or watching the meters.

The Sentinel tests the individual channel loudness measures which comprise the program loudness measure discussed earlier to detect dead channels. The user may select a minimum level and duration for the front channels and a separate minimum level and duration for the surround channels. This allows tighter limits on the front channels, recognizing the likelihood that surrounds will often be relatively silent in typical program material.

3.5. Mains Hum

Whether surround, stereo or mono, audio programs often inadvertently pick up mains hum. This is detected using three high-Q bandpass filters applied to each channel, one at the mains frequency, one at second harmonic and one at third harmonic. The outputs are rectified, summed and nonlinearly filtered. The proprietary nonlinear filtering helps remove the effects of program material which may happen to occur at the filter frequencies. If the resulting level exceeds a user defined threshold for a minimum duration, a hum error results.

4. CHANNEL INTERCHANGE

4.1. Surround Transport Formats

Surround programs are typically carried in one of several forms: Dolby-E, SDI or pairwise on multiple AES-3 links. Particularly when using multiple AES-3 links, though also possible when using the other methods, channels may be interchanged. The normal pairing of 5.1 programs places LF with RF, CF with LFE and LS with RS. If cables are inadvertently swapped or routing switchers misprogrammed, these can be rerouted with obvious ill effect.

4.2. Front/Surround Reversal

Interchanging the front and rear signals will be readily apparent to a listener with full surround monitoring but will be hidden by the downmix process for stereo listeners. The typical broadcast rack speakers have only two channels and as such will mask the problem.

In normal surround programs the front channel level will exceed the surround levels except during brief instances when special effects dictate activity behind the viewer. By continuously comparing the total front channel level to the total surround channel level and, for 7.1 programs, the total rear channel level such situations may be detected. Shifts in the program balance from front to sides or rear which extend over a sufficiently long time period may trigger a front/surround interchange error or a front/rear interchange error. As with other error detection algorithms the duration is user selectable.

4.3. CF/LFE Reversal

If the individual channels in an AES-3 pair are reversed the problem severity depends on which channels are involved. Fortunately the severity also tracks the difficulty in detecting such a reversal.

If the left and right channels are reversed (front, surround or back) in a surround program it is difficult to detect without an understanding of the program visual content. Fortunately the audible result will be noticeable but not catastrophic.

However, the case of CF and LFE reversal will be highly problematic since the viewer's subwoofer is unlikely to reproduce dialog and the viewer's CF speaker is unlikely to reproduce low frequency effects. Furthermore, since LFE is omitted from stereo and mono downmix there will be no dialog for stereo or mono listeners either. Fortunately CF/LFE interchange may be detected by monitoring the relative bandwidth of these two channels. If the LFE bandwidth exceeds the CF bandwidth for a selected duration a CF/LFE interchange error occurs.

To cover the case of non-pairwise signal carriage the LFE bandwidth is compared to that of all other channels whose level exceeds a minimum requirement. Again, if the LFE channel bandwidth is wider than these other active channels an LFE interchange error occurs.

4.4. Wrong Signals

For the case of digital inputs, an ever increasing percentage of the applications, there is other information available which may be tested for problems. All digital audio interconnection formats contain metadata. For example, the AES-3 interface contains information about the signal sample rate, word depth, source and destination, coding format, timecode and channel order.

Also, multichannel carriage formats such as Dolby Digital and Dolby-E contain metadata within their packets. This metadata indicates surround format, time code, word length, etc.

It is also possible to determine information about the signal from the interface hardware or by examining the raw digital audio data. For example the interface hardware can indicate the sample rate of PCM samples it receives. PCM audio data may be examined to determine its active word length.

If the many channels of a surround program are carried on multiple links it is reasonable to expect that the channels are consistent in format and metadata content. If they are not it may be an indication that the signals are not related and some inadvertent routing error occurred upstream of the signals being compared. Flagging these differences allows an operator to verify that the channels are as expected and increases confidence that problems will be detected early.

Channels are defined to be part of a surround program by the format and channel order settings made when connecting the Sentinel. For example only 6 of the 8 inputs may be used for a given program, the other two being assigned as auxiliary channels.

For PCM inputs, comparisons are made between metadata and the measured signal parameters of word length and sample rate. Differences are flagged as program errors. The Sentinel also compares these measurements across all channels in a surround program. Differences are flagged as a program group error.

Similarly, the metadata may be compared across multiple channels in a surround program. Differences are flagged as a group metadata error.

When the signal carried is Dolby Digital or Dolby E encoded the metadata may be compared between the AES-3 or SDI stream carrying the signal and that contained in the coded packet. Differences may be flagged as a coded metadata error.

5. CONCLUSION

The algorithmic design of a device for automatically assessing surround audio programs has been described. Special attention has been given to the previously visual assessment of program balance and downmix compatibility. The result is drastically reduced requirements for operator attention while maintaining adequate quality levels in broadcast and production applications. Commercial details have been omitted but are available online [3]

6. ACKNOWLEDGEMENTS

Matt Ashman provided major contributions to the signal processing algorithms and software described here.

The material described here and other aspects of the Qualis Audio Sentinel are the subject of several patent applications.

Sentinel is a trademark of Qualis Audio.

7. REFERENCES

[1] Brice, Richard: A New Monitoring Tool for 5.1 Audio, www.richardbrice.net/5.1_monitoring_tool.htm

[2] ITU-R BS.1770-1, Algorithms to measure audio programme loudness and true-peak audio level

[3] Sentinel data sheet, www.qualisaudio.com