Loudness Variation When Downmixing

The loudness of surround format programs is generally assessed assuming they are reproduced in surround. Unfortunately, as explained in the technical note, program loudness may change considerably when reproduced in stereo. This shift should not be confused with the overall gain change necessary to prevent clipping at the decoder output. We are concerned with inter-program differences when reproducing surround programs in stereo or mono.

Problems can arise when program loudness is normalized assuming reproduction in one format yet the programs are reproduced in a different format. For programs normalized in surround, a listener with surround reproduction would hear uniform playback loudness. A listener with a two-channel system would hear loudness variations between programs.

This note explains the origin of loudness changes when downmixing. Theoretical worst-case loudness differences are developed using a mathematical model of the downmix process and the standard loudness measurement algorithm. Measurements on an assortment of material are presented to give a perspective on the severity of the problem.

This investigation was prompted by the recognition of several theoretical discrepancies when applying the BS.1770 loudness computation to surround vs stereo formats. Not long after beginning this work a customer supplied sample material which exhibited a significant differential, enough to create problems when broadcast. Clearly the problem was not merely theoretical.

Understanding the ITU standard

The basic ITU loudness measurement algorithm is diagrammed in Figure 1. The audio channels (except the LFE) are independently filtered with a low frequency roll-off to simulate the sensitivity of the human ear and a high frequency shelf to simulate head diffraction effects. The combined response of these filters is referred to as "K weighting".

Surround channels are given a 1.5 dB boost to account for the relative gain provided by their position on each side of the listener. The power in each channel is summed to obtain the power in the entire signal. This power is averaged over a specified time period to obtain a metric for loudness. By assessing loudness separately for each channel the algorithm simulates perceived loudness in a space where reverberant mixing predominates. Readings are reported in LKFS (Loudness, K-weighted, relative to Full Scale) which may be thought of as "loudness dBFS".

A listeners' real-time perception of loudness is best described with a moderate averaging time. In R128 the EBU recommends a running 3 second average which they call "Short-Term" loudness (abbreviated "S"). They also define a momentary loudness (abbreviated "M") using a 400 ms averaging time which offers ballistics similar to a VU meter. They also define a set of "integrated" measurements where the results are gated with a Start/Stop control to allow selection of the audio segment to be measured.

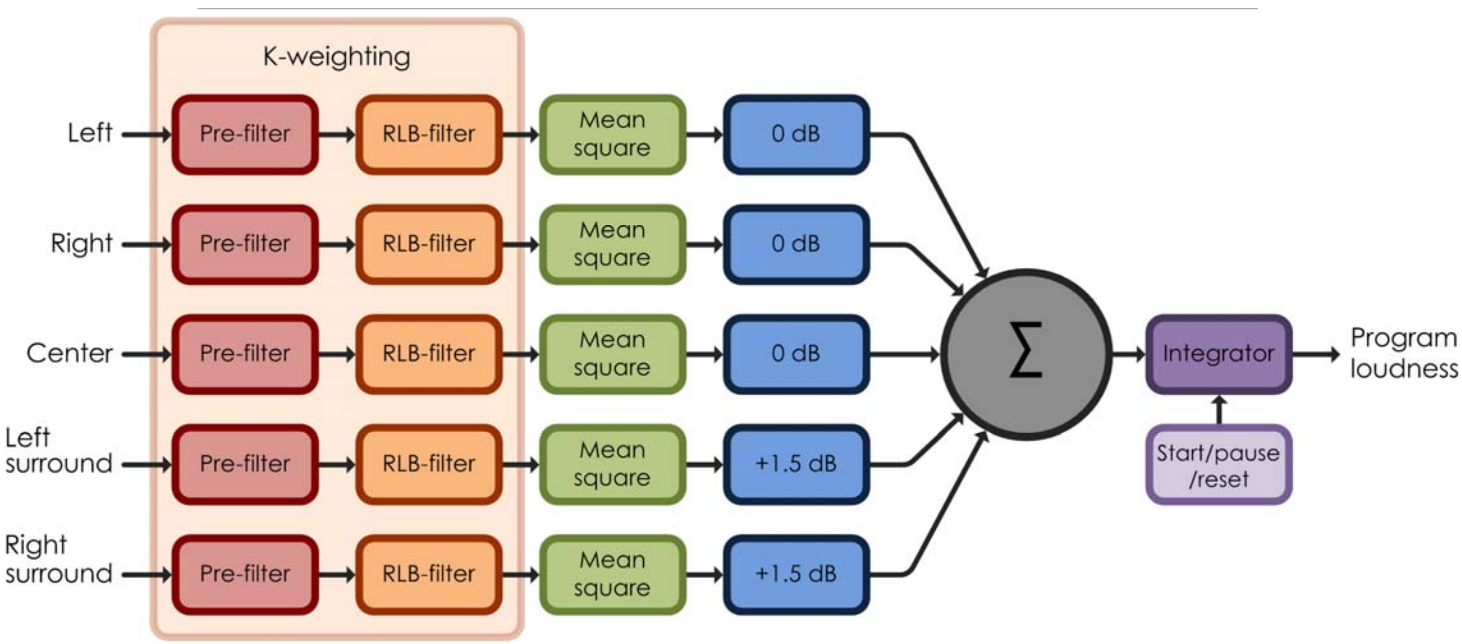

Figure 1 BS1770 Block Diagram

Downmixing

Despite the transmission of a surround signal, most viewers will hear the program reproduced in stereo, or even in mono. Although the percentage of viewers with access to surround playback is increasing steadily it is unlikely that the majority of viewers will ever be listening in surround. The practical aspects of where consumers situate televisions (think kitchen, bedroom, laundry room, etc.) dictate that the surround playback format will always remain the minority. However, the surround presentation in a living room or family room may well be the preferred location for viewing prime time or critical (think sports and movies) content.

Before proceeding further it helps to review the standard way the channels of a surround program are downmixed to make stereo. Prior to the scaling needed to prevent overflow, the general 3/2 downmix equations for an LoRo stereo signal are

Lo = 1.0 * L + clev * C + slev * Ls ;
Ro = 1.0 * R + clev * C + slev * Rs ;

The clev and slev values (the attenuation applied to the center and surround channels respectively) is specified by fields in the Dolby metadata. Although the coefficients are changeable with metadata settings most transmission chains use the default values of -3 dB. The figure below graphically illustrates the process.

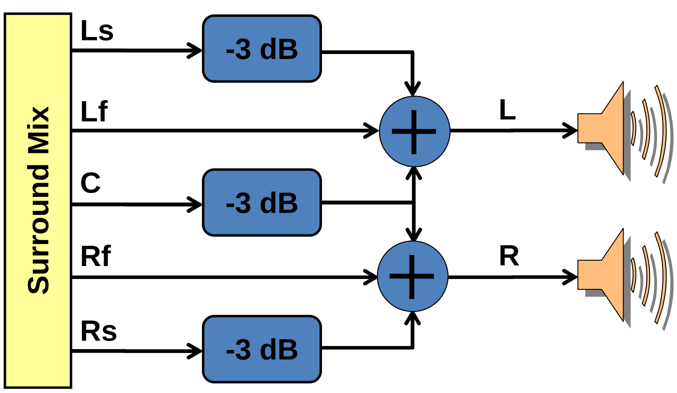

Figure 2 Standard downmixing of surround to stereo

The left front and right front channels are passed straight through to the left and right stereo outputs. The center front channel is mixed into both the left and right stereo outputs with a 3dB attenuation to compensate for the fact that its content will be reproduced through two speakers instead of one. The left and right surround channels are folded into the left and right outputs after an attenuation of 3dB. This attenuation is intended to compensate for the lack of directional separation and the resulting tendency of ambient sounds in the surround channels to mask primary content in the front channels.

Inter-program loudness differences

As cited earlier, when a surround program is downmixed and reproduced in stereo the loudness may change significantly. These changes, if they occur, arise from four different sources.

1. BS1770 specifies a 1.5 dB higher gain for surround channels than for front channels. When surrounds are downmixed they no longer get this additional gain when measured so their contribution to loudness is reduced by 1.5 loudness units.
2. AC3 metadata specifies attenuation of surround channels of 3dB or more when downmixing to stereo. When surrounds are downmixed this attenuation results in a further reduction of their contribution by 3 loudness units.
3. Interchannel phase related cancellation can occur during downmixing which reduces signal amplitude correspondingly reduces the loudness.
4. Loudness is based on summing individual channel powers, yet downmixing linearly combines channels before power computation. This can increase the loudness substantially as explained in the next section.

The first two effects, attenuation applied during downmixing and loudness meter surround channel gain combine to give a reduction in the surround channel contribution to loudness of 4.5 dB when a surround signal is downmixed and reproduced in stereo. Interchannel phase related cancellation would further reduce their contribution.

Loudness increases when downmixing

To understand how loudness can increase during downmixing we will begin with a simple example. When a source appears in only the center front channel the loudness is unchanged when downmixed to stereo.

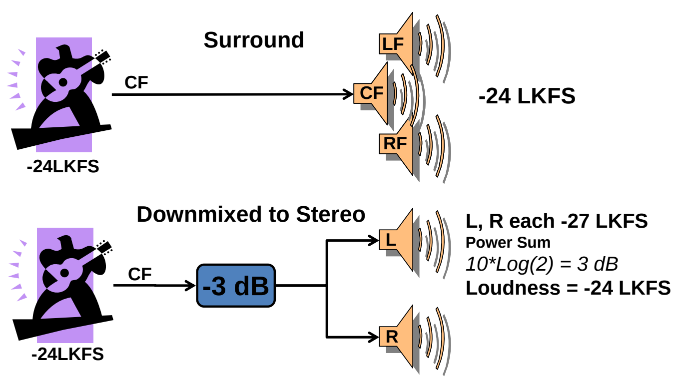

Figure 3 Downmixing a center channel source

Mixing the same source in all three front channels results in a loudness boost of 4.8 LKFS due to the power sum. This can be easily compensated by adjusting the mix level. However, when this is downmixed the channel voltages sum and boost the level by 4.6 dB. When the power in L and R are combined in the viewers room there is another boost of 3 dB.

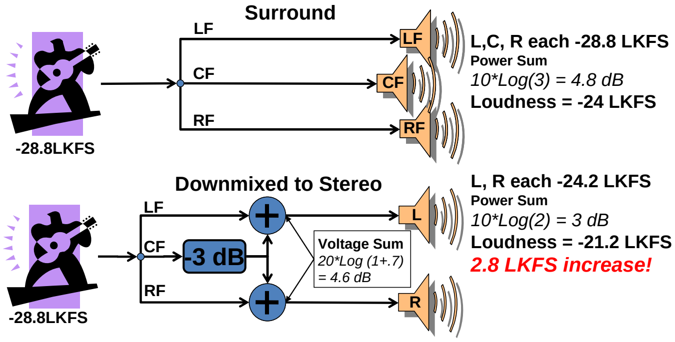

Figure 4 Downmixing a three channel source

The loudness of the stereo reproduction exceeds the surround version by 2.8 LKFS! The ATSC recommended practice (referenced by the CALM Act) specifies a tolerance of +/- 2 LKFS. Since downmixing will boost loudness of 3 speaker mono by 2.8 LKFS, a compliant program can become illegal!

Theoretical analysis

The ITU loudness measurement specification boosts surround channels by 1.5 dB to compensate for head shadow effects. When these channels are downmixed, they are reproduced in front of the listener where they are not subject to gain. Consequently, there is an inherent 1.5 LU difference between surround channels and downmixed surround channels. Further, AC3 metadata specifies attenuation of surround channels of 3dB or more when they are downmixed into left and right stereo channels. Combining these two effects gives an inherent 4.5 LU difference between surround and stereo reproduction of content in the surround channels.

The center front channel is reproduced in front of the listener in both surround and stereo formats. Consequently the ITU standard does not apply a different gain in these two cases. The standard Dolby downmix applies 3 dB loss to the center front channel before mixing it into both the left and right stereo channels. However, since the center front signal appears in both stereo channels it encounters a 3dB gain from occurring twice. This makes the center front channel loudness unchanged when reproduced in stereo.

However, when a signal appears in multiple channels simultaneously downmix loudness behaves differently. Consider the case of the same signal in the center front and left front channels. When loudness is measured in surround the result is proportional to $(LF)^2 + (CF)^2$. When measured in stereo the result is proportional to $(LF + CF)^2$. If the signals are equal amplitude there will be a factor of 2, or 3 dB, difference between surround and stereo reproduction. For signals appearing in multiple channels simultaneously the difference can be even larger.

Interchannel phase related cancellation will further reduce loudness when downmixing. Cancellation can occur between, and affect any of, the channels, not just the surrounds. Insight to this is provided by the Qualis Audio Sentinel downmix compatibility display. If severe enough, loudness will be affected. The result will always be a level reduction, further reducing the loudness of surround programs when reproduced in stereo.

The loudness of surround and stereo presentation of various channel combinations was modeled. The results (for cmix = -3dB, smix = -3dB) with no overall gain scaling of downmix coefficients are summarized below:

	All	3 Fr.	Center	Stereo	Typical	Surr.	Worst
LF Level	0	0	-80	0	-12	-80	-3
CF Level	0	0	0	-80	0	-80	0
RF Level	0	0	-80	0	-12	-80	-3
LS Level	0	-80	-80	-80	-26	0	-9
RS Level	0	-80	-80	-80	-26	0	-9
Level when downmixed, correlated and uncorrelated inputs							
Correlated	7.66	4.65	-3.01	0.00	-0.05	-3.01	4.43
Uncorrelated	3.01	1.76	-3.01	0.00	-2.48	-3.01	0.27
Loudness							
Surround	4.64	1.76	-3.01	0.00	-2.47	1.50	0.72
Correlated	7.66	4.65	-3.01	0.00	-0.05	-3.01	4.43
Uncorrelated	3.01	1.76	-3.01	0.00	-2.48	-3.01	0.27
Loudness of downmix compared to original surround version							
Correlated	3.01	2.88	0.00	0.00	2.41	-4.51	3.72
Uncorrelated	-1.63	0.00	0.00	0.00	-0.02	-4.51	-0.45

The table shows a worst case loudness increase of +3.7 LU with all channels driven by correlated signals, 3 dB larger in the CF and 6 dB softer in the surrounds. The worst case decrease of -4.5 LU occurs with signals in the surrounds and nowhere else. Correlated signals with interchannel phase issues can cause larger decreases if cancellations occur.

Mixing even small amounts of the center front channel into the left and right front channels will significantly increase loudness of the downmix. The figure below shows the downmix loudness differential as a function of the attenuation applied to the center front before mixing it into left and right front.

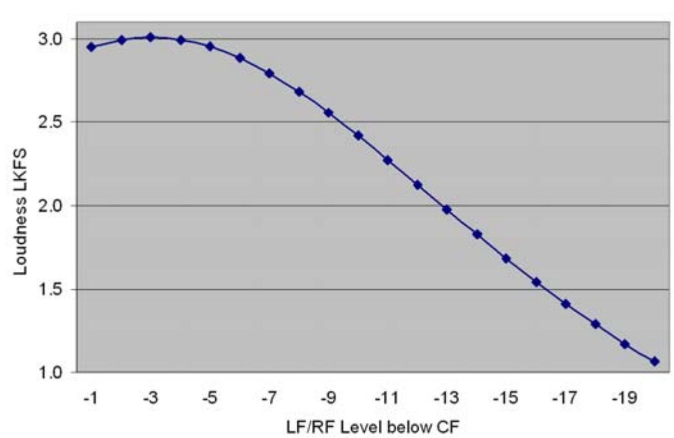

Figure 5 Effect on downmix of bleeding CF into LF & RF

When the center front is mixed into left and right front with a 12 dB reduction in level it increases the loudness of a downmix by more than 2 dB. With center front attenuated 20 dB before mixing into left and right front there is still more than a 1 dB increase in the loudness of the downmix.

Program material is unlikely to reach the extremes of loudness variation when downmixed, but it is common for it to reach more modest, but still significant shifts. The column labeled "Typical" represents levels seen on the first disc of the Ken Burns documentary series "Prohibition". Measurements on a variety of real world signals, given in the following figure, shows variations more than half these amounts.

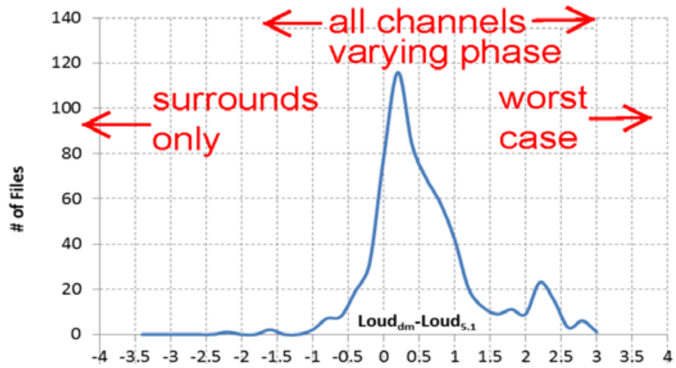

Figure 6 Typical loudness shifts after downmixing measured on program material compared to theoretical worst cases

This data was taken by Scott G. Norcross and Michel C. Lavoie of the Communications Research Centre in Canada. The loudness of multichannel audio programs recorded off-air and from a selection of DVDs was measured in accordance to ITU-R BS.1770-2. These values were then compared to those measured on the same programs after downmixing. Where it was possible, the measurements were broken down into individual chapters of DVD content to provide a more detailed view of any changes that were occurring. The differences observed for 626 multichannel files is shown as a histogram.

Most content shows a deviation of less than 1 LU. However, approximately 20% of the content measured shows larger differences than that. Approximately 10% is more than 2 LU divergent.

The red arrows show the theoretical extremes presented earlier, excluding the effects of phase cancellation.

To accurately assess loudness changes resulting from downmixing it is necessary to measure the downmixed signal.

Qualis Audio loudness meter implementation

The Qualis Audio Sentinel measures all parameters specified in ITU BS-1770 and EBU R128. This includes the momentary (400 ms), short-term (3 s) and integrated (averaged across the entire clip) metrics. It will do this both on the surround program and simultaneously on a downmixed version of the surround program. The short-term measurement best tracks listener perception of loudness and provides assessment of loudness trends in real time. The difference between the two short term loudness measures is also computed and displayed with the surround loudness as shown in Figures 4 & 5.

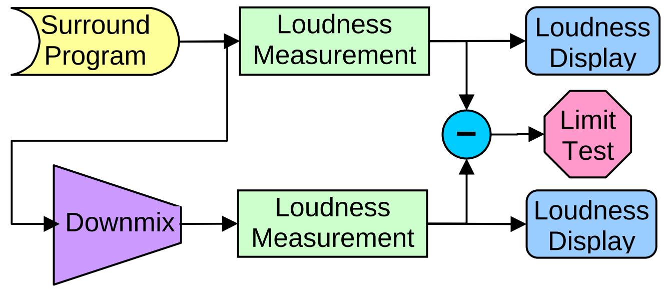

Figure 7 The downmix loudness measurement architecture

The measurements from the two loudness meters (surround and downmix) are each shown on a novel dual bargraph display, giving a complete loudness picture. The short-term downmix loudness is duplicated on the surround bargraph using a dual arrow style indicator. This simplifies reading the difference between the surround and downmix short-term loudness. The real time meter implementation is shown in the following figure.

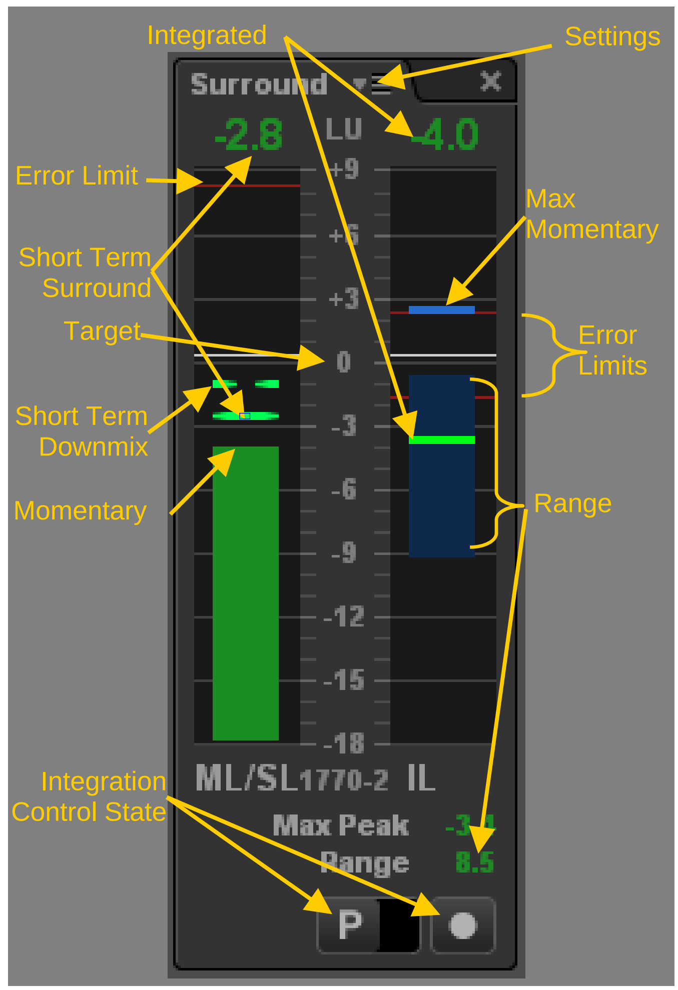

Figure 8 Qualis Audio loudness meter panel displays

The short term loudness is shown numerically and as a horizontal diamond shaped element on the left bar graph while the momentary loudness is mapped to the bar. The right bar-graph shows the integrated loudness both in numeric and graphical form along with other integrated parameters defined in EBU R128 including the loudness range, the maximum true peak, the maximum momentary loudness, and the maximum short term loudness. These are controlled with the same gate as the integrated loudness.

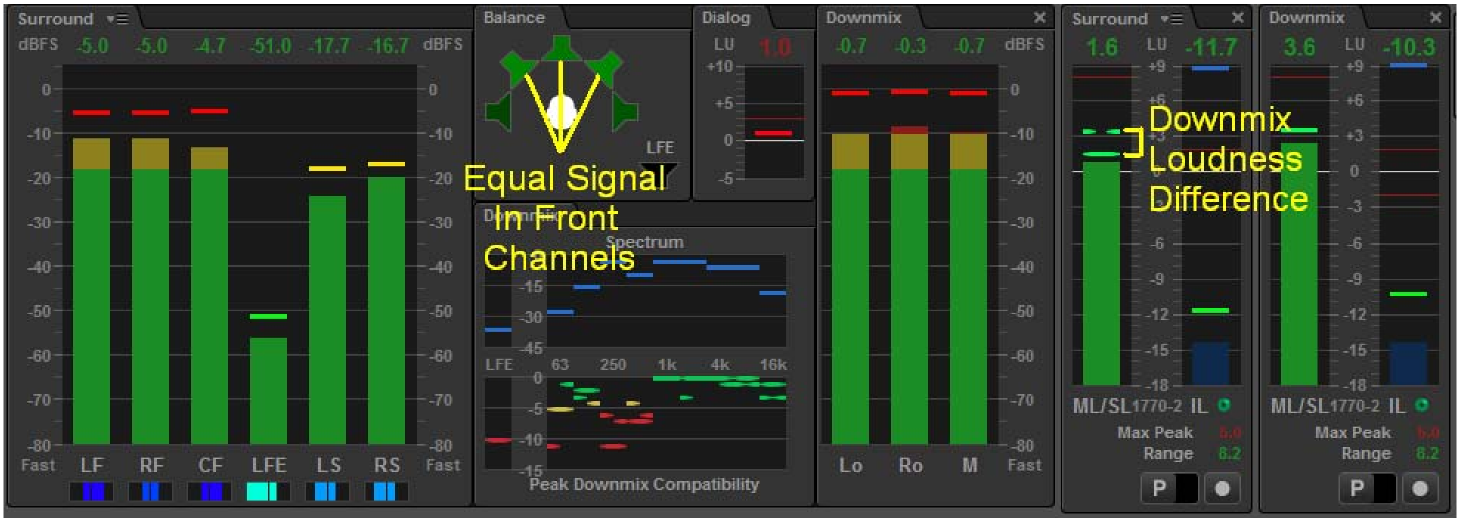

Figure 9a Measurements at cursor location “H” in timeline below, showing higher loudness when downmixed compared to surround

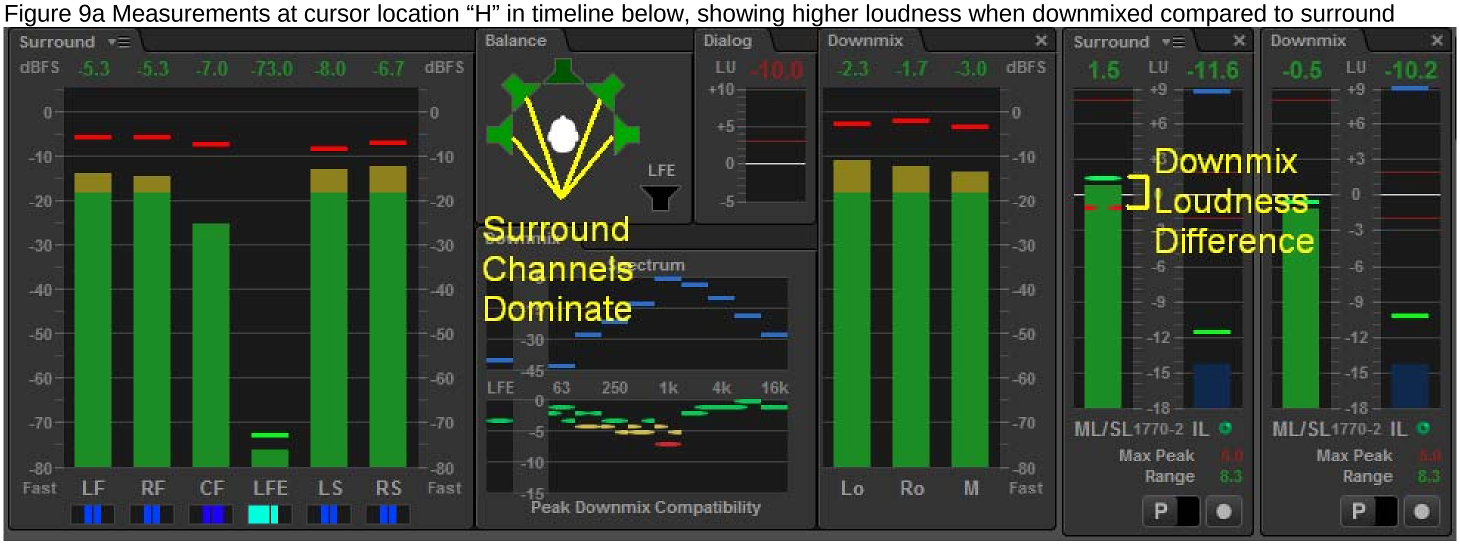

Figure 9b Measurements at cursor location “H” in timeline below, showing lower loudness when downmixed compared to surround

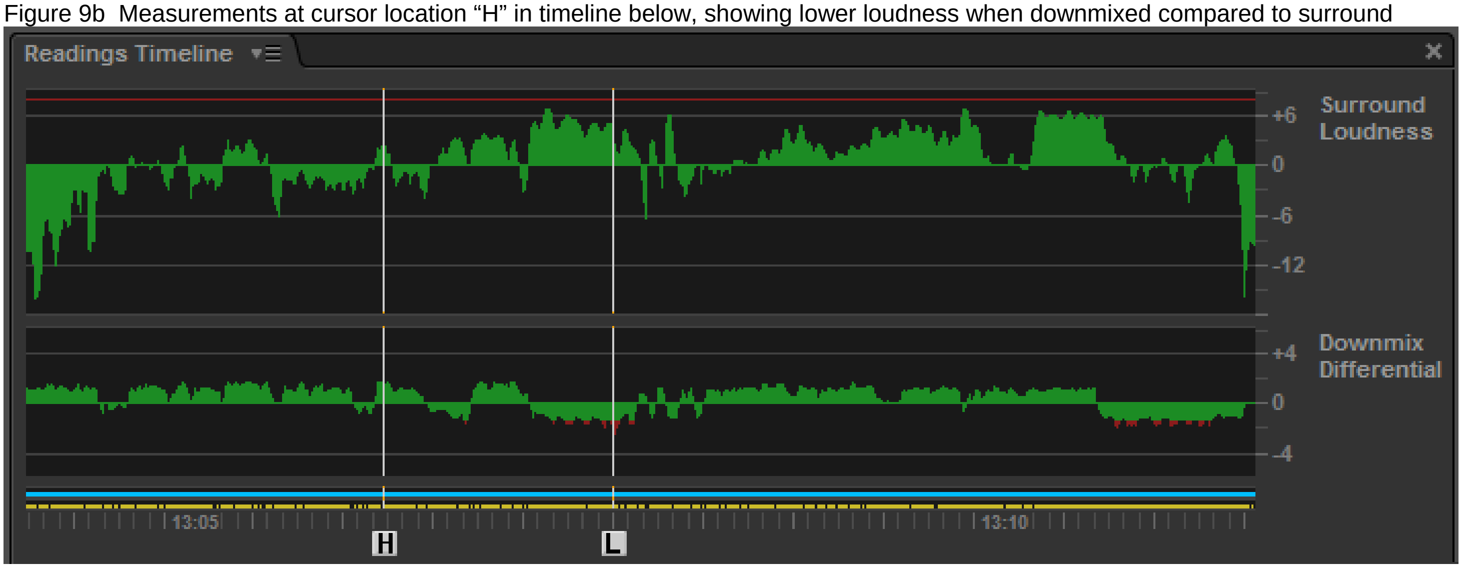

Figure 9c Surround loudness and loudness difference when downmixed vs time (measured on Adele Live at the Royal Albert Hall)

The short-term loudness, downmix loudness and downmix loudness differential measurements may be graphed as a function of time. A practical application of this capability is presented in the next section.

A practical example

A segment from the DVD “Adele, Live At The Royal Albert Hall” was measured using a Qualis Audio Sentinel surround audio monitor. This disc makes an excellent example because it demonstrates each of the effects discussed above. A portion of the results are shown in Figure 9 a-c. The display in Figure 3a is typical of most of the concert. The vocalist appears in all three front channels simultaneously. When downmixed, the channels reinforce each other, resulting in an approximately 2 LU loudness increase.

However, during the song “Someone Like You” she asks the audience to sing the refrain with her. As shown in Figure 3b, surround channel loudness then exceeds the front channel loudness. Because of the 4.5 dB attenuation when surrounds are downmixed the downmix loudness is lower than the surround loudness by approximately 2 LU.

The short-term downmix loudness differences described above were measured across the entire Adele concert. The results are shown as a histogram in Figure 4.

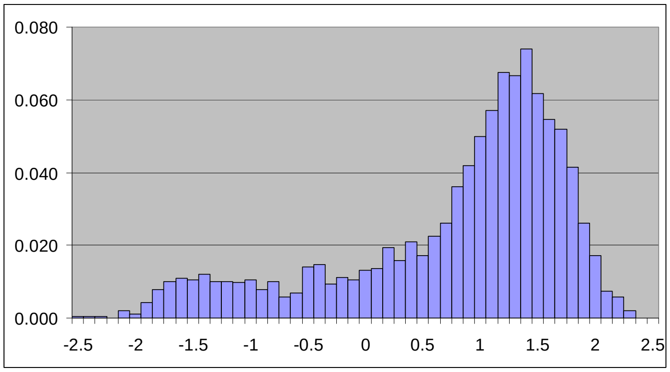

Figure 10 Downmix loudness histogram for Adele concert DVD

The downmix shows a dominant loudness increase of approximately 1.4 LU, driven by the vocal portions of the concert. However, the large number of measurements demonstrating a reduction in loudness when downmixed brings the average for the entire concert down to only a 0.8 LU increase. .

Summary

The ITU BS.1770 loudness measurement standard is widely used to assess the loudness of surround programs for digital television. However, characteristics of the measurement standard along with aspects of the downmix process can cause loudness to shift when some surround programs are reproduced in stereo. This shift can be as much as 3 loudness units of increase when the program contains content which has been mixed into all three front channels simultaneously and almost 4 loudness units when the same content is mixed into all five channels. Conversely, heavy use of the surround channels or interchannel phase problems can create more than 4 loudness units decrease when content is downmixed. Meters exist which will catch these conditions, allowing the user to make adjustments during mixing to prevent these problems.

Appendix: Downmixing Details

The following is excerpted from ATSC standard A/52 which defines Dolby Digital (AC3) coding. Discussion of surround formats other than 5.1 has been omitted. The text below is identical in both A/52A and A/52B.

7.8 Downmixing

In many reproduction systems, the number of loudspeakers will not match the number of encoded audio channels. In order to reproduce the complete audio program, downmixing is required. It is important that downmixing be standardized so that program providers can be confident of how their program will be reproduced over systems with various numbers of loudspeakers. With standardized downmixing equations, program producers can monitor how the downmixed version will sound and make any alterations necessary so that acceptable results are achieved for all listeners. The program provider can make use of the cmixlev and smixlev syntactical elements in order to affect the relative balance of center and surround channels with respect to the left and right channels.

Downmixing of the lfe channel is optional. An ideal downmix would have the lfe channel reproduce at an acoustic level of +10 dB with respect to the left and right channels. Since the inclusion of this channel is optional, any downmix coefficient may be used in practice. Care should be taken to assure that loudspeakers are not overdriven by the full scale low frequency content of the lfe channel.

7.8.1 General Downmix Procedure

The following pseudo code describes how to arrive at un-normalized downmix coefficients. In a practical implementation it may be necessary to then normalize the downmix coefficients in order to prevent any possibility of overload. Normalization is achieved by attenuating all downmix coefficients equally, such that the sum of coefficients used to create any single output channel never exceeds 1.

(pseudo-code omitted)

The actual coefficients used for downmixing will affect the absolute level of the center channel. If dialogue level is to be established with absolute SPL calibration, this should be taken into account.

7.8.2 Downmixing Into Two Channels

Let L, C, R, Ls, Rs refer to the 5 discrete channels which are to be mixed down to 2 channels. Two types of downmix should be provided: downmix to an LtRt matrix surround encoded stereo pair; and downmix to a conventional stereo signal, LoRo. The downmixed stereo signal (LoRo, or LtRt) may be further mixed to mono, M, by a simple summation of the 2 channels. If the LtRt downmix is combined to mono, the surround information will be lost. The LoRo downmix is preferred when a mono signal is desired. Downmix coefficients shall have relative accuracy of at least ±0.25 dB.

Prior to the scaling needed to prevent overflow, the general 3/2 downmix equations for an LoRo stereo signal are

Lo = 1.0 * L + clev * C + slev * Ls ;

Ro = 1.0 * R + clev * C + slev * Rs ;

If LoRo are subsequently combined for monophonic reproduction, the effective mono downmix equation becomes

M = L + 2 * clev * C + R + slev * Ls + slev * Rs ;

The values of clev and slev are indicated by the cmixlev and surmixlev bit fields in the bsi data ….

Prior to the scaling needed to prevent overflow, the 3/2 downmix equations for an LtRt stereo signal are

Lt = L + 0.707 * C – 0.707 * Ls – 0.707 * Rs ;

Rt = R + 0.707 * C + 0.707 * Ls + 0.707 * Rs ;

The actual coefficients used must be scaled downwards so that arithmetic overflow does not occur if all channels contributing to a downmix signal happen to be at full scale. For each audio coding mode, a different number of channels contribute to the downmix, and a different scaling could be used to prevent overflow. For simplicity, the scaling for the worst case may be used in all cases. This minimizes the number of coefficients required. The worst case scaling occurs when clev and slev are both 0.707. In the case of the LoRo downmix, the sum of the unscaled coefficients is 1 + 0.707 + 0.707 = 2.414, so all coefficients must be multiplied by 1/2.414 = 0.4143 (downwards scaling by 7.65 dB). In the case of the LtRt downmix, the sum of the unscaled coefficients is 1 + 0.707 + 0.707 + 0.707 = 3.121, so all coefficients must be multiplied by 1/3.121, or 0.3204 (downwards scaling by 9.89 dB). …..

(coefficient tables omitted, see table below)

If it is necessary to implement a downmix to mono, a further scaling of 1/2 will have to be applied to the LoRo downmix coefficients to prevent overload of the mono sum of Lo+Ro.

The cmixlev and smixlev values are carried in the AC3 metadata. The original AC3 (Dolby Digital) coding format allowed 3 values for each. A later version allowed finer control over both values as shown in the table below.

Downmix coefficient	cmixlev		smixlev	
	original	enhamced	original	enhamced
1.414 (+3.0 dB)		X		
1.189 (+1.5 dB)		X		
1.000 (0.0 dB)		X		
0.841 (–1.5 dB)		X		X
0.707 (–3.0 dB)	X	X	X	X
0.595 (–4.5 dB)	X	X		X
0.500 (–6.0 dB)	X	X	X	X
0.000 (–inf dB)		X	X	X

The cmixlev and smixlev value do not have the same allowed range in either metadata version. Originally the center channel could be combined into the stereo channels with an attenuation ranging from 3 to 6 dB. The later version adds the ability to boost the center by up to 3 dB or to completely remove it from the downmix. The original metadata set allowed surrounds to be folded into the stereo signal with attenuation of 3 dB, 6 dB or to be completely omitted. The later version allows finer control but always requires at least some attenuation.

Downmix gain scaling

For full-scale sinewaves, the downmix levels will reach values equal to the peak_gain, which can be larger than 0. The time domain downmix is often scaled by an overall gain factor chosen to prevent peaks from ever exceeding full scale. The gain required is computed by summing the downmix coefficients used for each channel added into a downmix channel. For example:

peak_gain = 1.0/(1.0 + cmix + smix) (5.1 case)

peak_gain = 1.0/(1.0 + cmix + smix +0.707*smix) (6.1 case)

peak_gain = 1.0/(1.0 + cmix + 2.0*smix) (7.1 case)

The raw downmix coefficients are multiplied by the overall gain value before use. The Qualis Audio Sentinel provides selection of this gain to allow assessment of levels after downmixing which match the users preferences. The available choices are

- None
- Peak
- Peak – 3dB

Decoder gain scaling

The ATSC standards give the decoder manufacturer discretion in the overall gain applied when downmixing to stereo. The gain used in any specific implementation has a direct impact on the likelihood that a given program will clip at the decoder output. However, the choice of gain has no impact on the relative levels between programs when reproduced in a common format. Aside from the clipping issue, the choice of decoder gain is compensated by a one-time adjustment of the users volume control.

References

Jacobs, Bruce "CALM Down!", HPA Tech Retreat, February 2012

Norcross, Scott G. and Michael C. Lavoie "The Effect of Downmixing on Measured Loudness", AES e–Brief #38, 131st Convention, October 2011

Riedmiller, Jeffrey C. "Dolby Laboratories Recommended Practices For Cable Television Systems", November 2005

Riedmiller, Jeffrey C. "Dolby Digital (AC-3) Overview for Cable", August 2004

ATSC A/52b "Digital Audio Compression Standard (AC-3, E-AC-3)", June 2005

ATSC A/54a "Guide to Use of the ATSC DTV Standard", December 2003

ATSC A/85-2011 "Techniques for Establishing and Maintaining Audio Loudness for Digital Television", July 2011

Qualis Audio, Inc.
PO Box 731
Lake Oswego, OR 97034
503.635.9376
www.qualisaudio.com
info@qualisaudio.com

We're Listening