Understanding & Verifying Loudness Meters

Richard Cabot and Ian Dennis

This note covers the history of broadcast loudness measurement, the ITU BS1770 standard including the 2011 revisions and other loudness measures proposed by the EBU. It then describes how to verify that a loudness meter meets these requirements.

Some History

Loudness has been a hot topic lately. Loudness issues have been around since the beginning of broadcasting, or at least broadcast advertising, but they were exacerbated by the DTV transition and inconsistent use of the dialnorm metadata. Listener dissatisfaction increased, ultimately culminating in passage of the CALM Act.

In the early years of DTV these problems prompted development of loudness measurement technologies, resulting in the ITU standard BS1770 which describes a fundamental loudness measurement algorithm. This was validated through multiple sets of listening tests on various pieces of program material. It also describes a true-peak meter for determining the peak amplitude expected when a digital audio signal is reproduced in the analog domain or transcoded into another digital format. The ITU BS1770 standard is referenced in the ATSC recommended practice A/85 which describes how broadcasters should insure a satisfactory listening experience for DTV viewers. The ATSC document also states "Users of this RP should apply the current version of ITU-R BS.1770". We will discuss the importance of this below.

When the CALM Act was originally introduced it mandated uniformity of commercial and program loudness. It did not include measurement methods and was so vague that it would have been a nightmare for both broadcasters and the FCC. Fortunately, industry representatives convinced Congress that adoption of the recently-developed A/85 recommended practice would resolve the situation. The legislation was rewritten to mandate that the FCC enforce ATSC A/85, and its successors, through appropriate rule making.

The CALM Act created an obvious opportunity for equipment manufacturers to provide loudness measurement tools. Although a few loudness measurement products were on the market by 2009, many more were introduced at NAB and IBC in 2010, and there are currently over a dozen loudness measurement products of various forms on the market. Each is vying for a piece of the large short-term market created as broadcasters equip their facilities for loudness

At the end of October 2010 the ITU committee which maintains BS1770 accepted (after much negotiation, principally regarding the relative gate threshold described below) changes submitted by the EBU. The result is a significant improvement in the calculation of loudness which makes the measurement much more sensitive to the loud portions of an audio segment. The effect is to prevent advertisers from significantly increasing the loudness of a portion of a commercial by drastically reducing the loudness elsewhere. Consider a hypothetical example where an announcer screams at widely spaced intervals throughout a commercial in an effort to get the viewers' attention. The new method assesses this spot as louder than the original technique. It also removes the explicit requirement to measure the loudness of dialog and instead bases the assessment on all audio content (except for the LFE). This prevents unscrupulous advertisers from circumventing loudness limits by blasting a viewer with non-vocal content, while removing the need for proprietary algorithms which only measured loudness when dialog was present.

This new version of BS1770 will probably be published in early 2011. Recall that ATSC A/85 already specifies that updates to BS1770 automatically apply. Consequently meters in use will need to conform to the revised specification. Most manufacturers of such products have been following the developments in the EBU and ITU and have upgraded their software to accommodate the change. Unfortunately there is no assurance that they have implemented the changes and, equally important, that they have implemented them correctly.

As a user, how do you assess whether a meter you are considering meets the revised specification? It's not as easy as testing a VU meter or a PPM. The EBU technical recommendations include basic test signals for verifying compliance with the specified measurement algorithms. Unfortunately these test signals are not comprehensive and some list expected results which are incorrect.

We will describe a suite of tests developed specifically to check every aspect of a meter's design. These tests also give diagnostic information about any implementation issues that exist. They are available at no charge from the Qualis Audio website www.qualisaudio.com .

Our test suite was developed by crafting signals whose parameters change dynamically so as to stress individual portions of the measurement in isolation. Each test can then maximize its sensitivity to the specific implementation errors it was designed to detect. The signals were passed through mathematical models of the algorithm and also through models with intentional implementation errors. The signals were optimized to give the largest difference between readings obtained by the correct model and those obtained by incorrect implementations.

The new BS1770 algorithm operates on multiples of a basic 100ms interval, so readings differ slightly with variations between the start of the measurement and the start of the signal. These reading differences follow a cyclic pattern, with alignments 50ms apart creating maximal difference. Consequently the test signals were evaluated at a reference alignment and at an alignment 50ms delayed, and signal characteristics adjusted to minimize this difference - though sometimes this was in direct conflict with the desire to maximize the sensitivity to implementation errors.

Understanding the ITU standard

The original 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" and is illustrated in Figure 2.

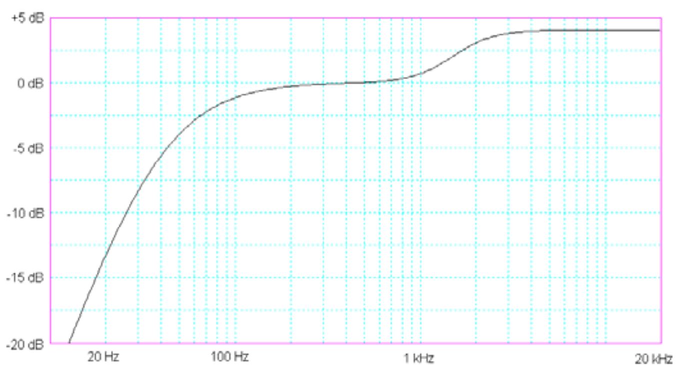

Figure 2 K-weighting filter

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 the entire program to obtain a single number metric for the program loudness. If a "dynamic" indication of loudness is desired, a 3 second moving average is typically used. Readings are reported in LKFS (Loudness, K-weighted, relative to Full Scale) which may be thought of as "loudness dBFS".

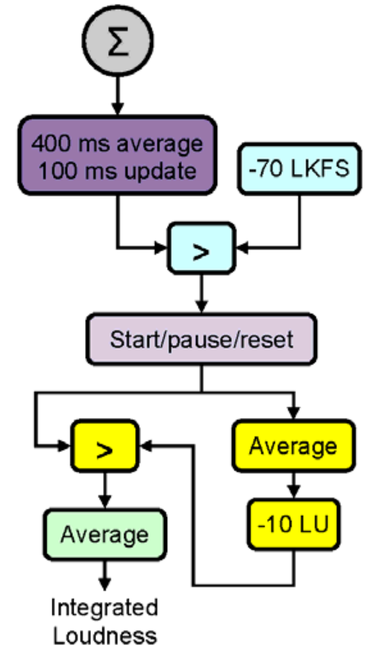

Figure 3 Additional processing in BS1770 rev 2011.

The ATSC recommendation specifies that loudness measurements should focus on dialog or an alternative anchor element. The intent was that the viewer would set the dialog loud enough to be intelligible in their environment, and that maintaining constant dialog loudness would maintain intelligibility. This assumed "well behaved" content (many commercials don't fit this description), and also depended on proprietary loudness measurement technology. In an effort to address these, and other, issues the EBU PLOUD committee revisited BS1770. Their work resulted in the 2011 revision of BS1770.

This revision maintains the same filtering and power

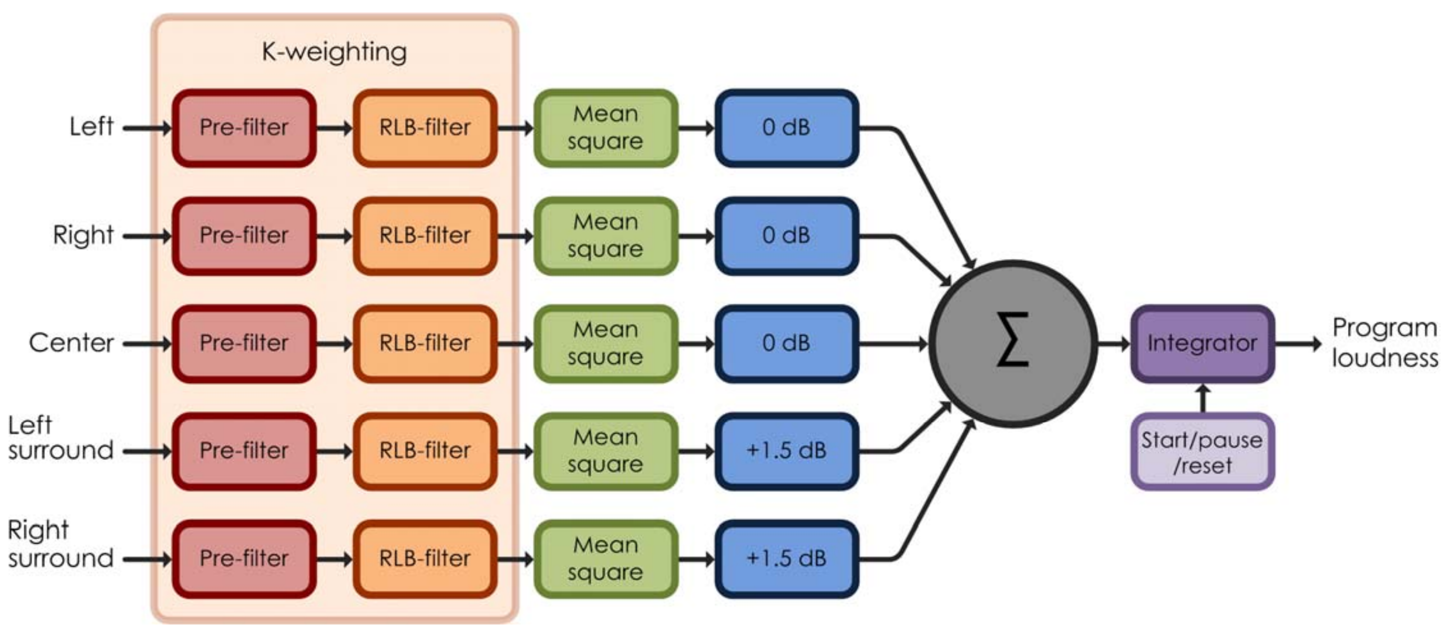

Figure 1 BS1770 Block Diagram

measurement method used in the original standard, but changes the way measurements are averaged and presented. The integrator stage of Figure 1 is replaced with the processing shown in Figure 3. The channel power is summed over 400 ms intervals. These intervals overlap by 75% so a new value is obtained every 100ms. Results are gated with a Start/Stop control to allow selection of the audio segment to be measured. An absolute gate of -70 LKFS is applied, which automatically eliminates lead-in and playout portions of isolated audio segments.

The algorithm focuses on the foreground portion of the audio by a two step averaging procedure (the yellow elements in Figure 3): The 400 ms values are averaged over the entire content being measured; the resulting LKFS value is decreased by 10 and used as a gating threshold. The individual 400 ms measurements which exceed this threshold are averaged to form the final reading called "Integrated" loudness (abbreviated "I"). This "relative gate" focuses the assessment on foreground sounds, the elements which generally dominate viewers' judgments of program loudness.

The EBU unsuccessfully lobbied the ITU to change the units for absolute loudness from LKFS to LUFS. The EBU has stated its intention to use the LUFS designation so these two equivalent units will be in simultaneous use.

The most recent versions of the EBU documents (December 2010) also specify a -8 LU gate instead of the -10 LU gate compromise reached with the ITU. The EBU PLOUD group has stated it will change these documents when the revised version of ITU BS1770 is released. In the interim, it is important to be clear what gate threshold is used when evaluating or using a loudness meter.

Both the ITU and ATSC documents also specify a true-peak meter. This is a device which measures the peak value a digital audio waveform will reach when it is reproduced in the analog domain or when it encounters many forms of digital processing. To understand the problem, recall that digital audio represents a continuous analog signal by a series of samples, taken at regular intervals determined by the sample rate. As Figure 4 illustrates, there is no guarantee that samples will land on the audio waveform peak.

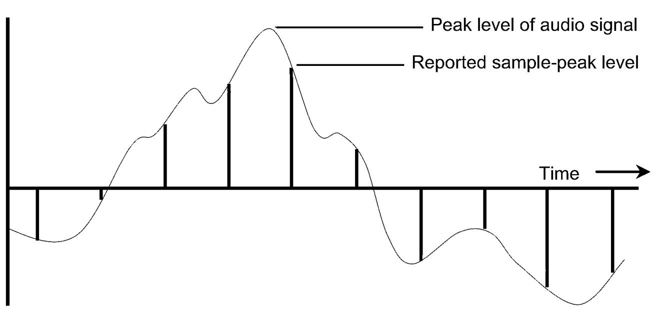

Figure 4 The need for true-peak indication (fATSC A/85)

However these samples do represent the underlying audio waveform and when it is reconstructed the peak will be restored. This peak can also occur when the samples subjected to many types of processing - anything which introduces phase shift or time offset – such as sample rate conversion, filtering or delay. If this happens in the digital domain the new samples may clip, even if the original samples did not reach digital full scale. Since many peak meters merely display the maximum audio sample, they incorrectly gauge the system headroom.

Additional EBU Loudness Descriptors

The EBU recommendation introduces other measures which are still under consideration by the ITU. Intended to assist mixers and program personnel in creating and characterizing content, their acceptance by the ITU is unlikely to impact CALM Act requirements. However, given their potential usefulness in production it is helpful to understand them.

The EBU specifies "Momentary" loudness (abbreviated "M") as the stream of 400ms measurements which drive the gating mechanisms described earlier. When displayed on a meter they look much like a VU display since the 400 ms averaging time is close to the 300 ms of a classic VU meter. Watching this display during a mix helps a mix engineer estimate the program loudness of live productions. The "Maximum Momentary" loudness is specified by the EBU as the largest 400 ms measurement during the measurement time. The listeners' perception of loudness is best described with a longer averaging time. The EBU recommends a running 3 second average which they call "Short-Term" loudness (abbreviated "S").

The EBU also define a measurement called "Loudness Range" (abbreviated LRA). This is derived from the Short-Term loudness using a relative gating process similar to that described above but with the gate set at -20. The LRA is the span from the 10% to 95% points on the distribution of Short-Term loudness values that pass the relative gate. The LRA is descriptive of the program material dynamic range. Using the 95% point allows occasional extremely loud events while the 10% point ignores modest silent intervals during the program. If the LRA exceeds about 15 it is likely that viewers will be unable to find a single volume control setting appropriate for the entire program.

Loudness Meter Evaluation

The test suite described here checks BS1770 compliance of any loudness meter. It may be downloaded as a set of wave files and a selector menu from www.qualisaudio.com. The tests and expected results are described below. New tests will be added as they are developed.

The selection menu is shown in Figure 5. A pdf file containing this menu is packaged with the waveform files. If they are unzipped into a common directory, clicking on the links within the figure will launch play the files. Alternately, the files may be played directly from standard audio playback software. The full suite currently consists of 12 tests. All tests except Test 2 (it is 5.1 Dolby Digital encoded) are stereo signals and should be applied to the LF and RF channels of a surround loudness meter. All tests except the first three comprise a 1 kHz sinewave at varying amplitudes. When the expected result is a range rather than a specific target, this is due to the 100ms alignment uncertainty.

Cautions regarding EBU tech 3341

The December 2010 versions of the EBU documents still specify a -8 LU gate instead of the -10 LU gate compromise reached with the ITU.

This suite was developed and verified using a -10 LU relative gate. The expected results will differ if a -8 LU relative gate is used.

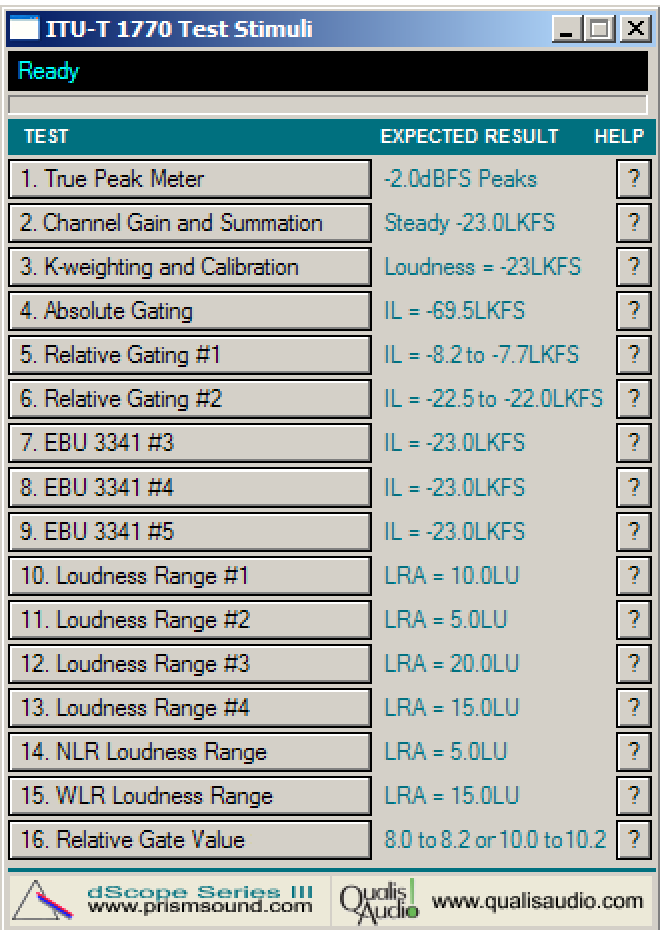

Figure 5 Test Suite Menu

Test 1 True-Peak meter interpolation

Test 1 checks the accuracy of the True-Peak meter. The initial waveform is a 1/8 sample rate, -6 dBFS sinewave whose samples occur at the peaks. After 3 seconds, the frequency changes for one cycle to ¼ sample rate and the amplitude increases to -2 dBFS. The samples are chosen to occur 45° off the sinewave peaks as in Figure 6a below.

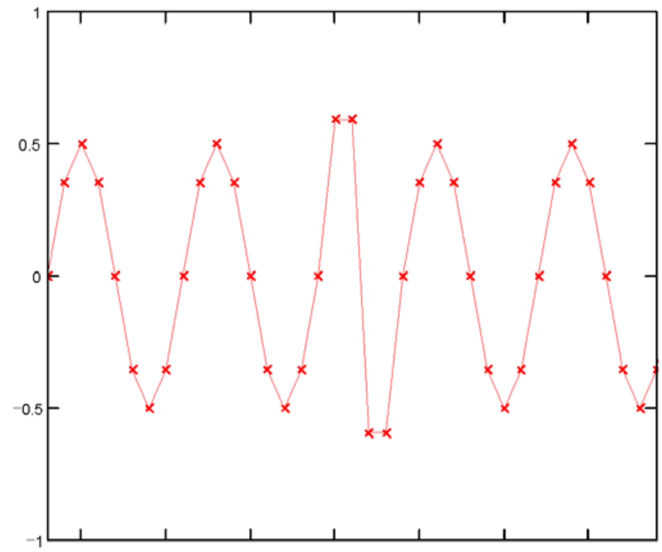

Figure 6a true peak waveform samples

When this waveform is properly interpolated, as would occur when it is reproduced in the analog domain, the waveform of Figure 6b results.

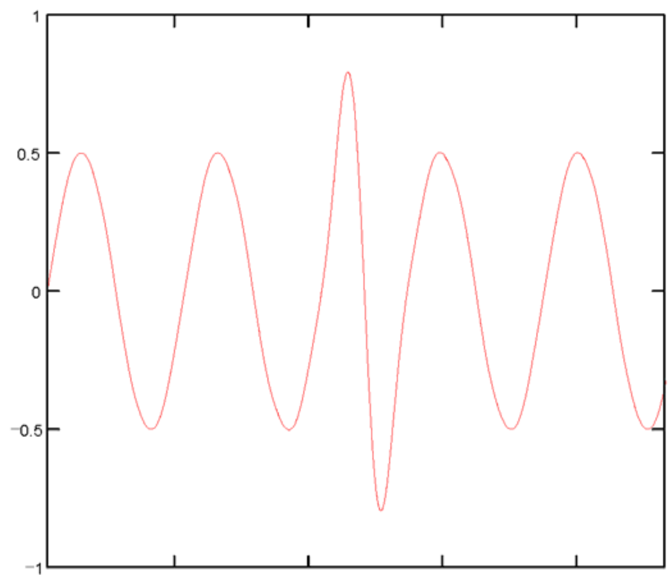

Figure 6b true peak waveform after interpolation

All peak meters will initially read -6 dBFS when measuring this waveform. After 3 seconds a properly interpolating meter will show -2 dBFS. Non-interpolating meters will read -5 dBFS.

Test 2 Channel gain and power summation

Test 2 is a variation on EBU Tech 3341 Test Case 6. This signal is Dolby Digital encoded and stimulates all channels simultaneously, including the LFE. Power summation is checked by using sinewaves of slightly different frequencies. A compliant meter will read -23 LKFS. If the meter sums in the time-domain, rather than summing powers, the reading will cycle. If the meter includes the LFE, the reading will be too high. The channels are driven at the levels and frequencies specified in the table below

Channel	Frequency	Amplitude
LF	999.61 Hz	-28 dBFS
RF	1000.39 Hz	-28 dBFS
C	1000.00 Hz	-24 dBFS
LFE	100.00 Hz	-15 dBFS
LS	1000.39 Hz	-30 dBFS
RS	999.61 Hz	-30 dBFS

Test 3 Calibration and weighting filter response

Test 3 checks the weighting filter response at six frequencies: 25 Hz, 100 Hz, 500 Hz, 1 k, Hz 2 kHz and 10 kHz using sinewaves of varying amplitudes. The waveform begins with a 5 second tone at 1 kHz so the meter reading can reach a stable value. It then steps through each of the listed frequencies at 3 second intervals and returns to 1 kHz. The sequence used is:

Time	Frequency	Amplitude
0-5 s	1 k Hz	-22.99 dBFS
5-8 s	10 k Hz	-26.33 dBFS
8-11 s	2 k Hz	-25.35 dBFS
11-14 s	1 k Hz	-22.99 dBFS
14-17 s	500 Hz	-22.33 dBFS
17-20 s	100 Hz	-21.15 dBFS
20-23 s	25 Hz	-11.92 dBFS
23-26 s	1 k Hz	-22.99 dBFS

This is slow enough to allow reading the response from either the momentary or short-term loudness displays. A meter with a properly implemented filter will give a constant reading of -23 LKFS.

Test 4 Absolute Gate implementation

Test 4 alternates a 1 kHz tone between -69.5 and -90 dBFS to exercise the absolute gating function. The sequence used is:

Time	Amplitude
0.0-0.8 s	-90.0 dBFS
0.8-1.3 s	-69.5 dBFS
1.3-1.8 s	-90.0 dBFS
1.8-2.3 s	-69.5 dBFS
2.3-2.8 s	-90.0 dBFS
2.8-3.3 s	-69.5 dBFS
3.3-4.0 s	-90.0 dBFS

A compliant meter will exclude the -90 dBFS portions from the measurement and will read -69.5 LKFS. A meter which does not implement absolute gating will include the -90 dBFS portions and will read -73 LKFS or less. A meter which does not apply both the relative and absolute gates will read between -71.5 and -72.0 LKFS.

Test 5 Relative Gate implementation

Test 5 steps the amplitude of a 1 kHz tone in increments and durations such that the relative gate threshold will be determined by a single amplitude segment and the other segments should be ignored. The sequence used is:

Time	Amplitude
0.0-0.5 s	-90.0 dBFS
0.5-1.7 s	-23.5 dBFS
1.7-2.3 s	-6.0 dBFS
2.3-3.5 s	-23.5 dBFS
3.5-4.0 s	-90.0 dBFS

The average amplitude is -13.1 to -13.2 dBFS if the -90 dBFS portions are removed by the absolute gate. This creates a relative gate threshold of -23.1 to -23.2 dBFS. The -23.5 dBFS segments should therefore be ignored. A meter which correctly implements relative gating reads -7.7 to -8.2 LKFS; a meter without only absolute gating reads between -13.1 and -13.2 LKFS.

Test 6 Relative Gate implementation

Test 6 checks the relative gate implementation. The tone sequence used is:

Time	Amplitude
0.0-0.8 s	-90 dBFS
0.8-1.3 s	-36 dBFS
1.3-1.8 s	-20 dBFS
1.8-2.3 s	-36 dBFS
2.3-2.8 s	-20 dBFS
2.8-3.3 s	-36 dBFS
3.3-4.0 s	-90 dBFS

The average amplitude is -24.3 to -24.5 dBFS if the -90 dBFS portions are removed by the absolute gate. This results in a relative gate threshold of -34.3 to -34.5 dBFS which will exclude the -36 dBFS segments. A compliant meter reads between -22.0 and -22.5 LKFS, whereas non-compliant meters read between -24.3 and -25.5 LKFS

Tests 7 - 9 modified EBU tech 3341 tests 3 - 5

The expected results listed in EBU Tech 3341 Table 1 for test signals 3, 4 and 5 are slightly off for both -8 and -10 values of relative gate. The discrepancy is because the values in tech 3341 do not take into account the three or four 400 ms measurements which straddle the amplitude transitions in the test signal. The quantity depends on the time alignment of the waveform and the meter integration. Figure 7 illustrates the transition effect, and the variable number of packets involved, for a change from -26 to -20 dBFS.

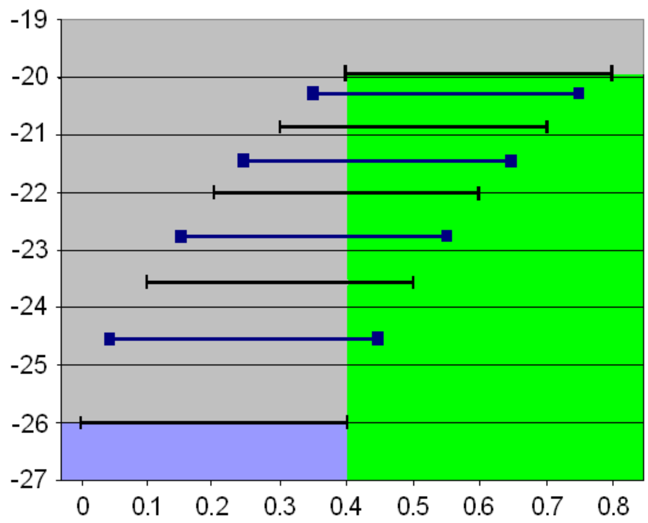

Figure 7 400 ms integration across an amplitude step

The black horizontal lines represent 5 consecutive packets, 3 spanning the transition. The 4 blue lines with square ends illustrate the behavior when measurement packets do not align precisely with the transition. In this example the offset is 50 ms.

The values listed in EBU tech 3341-2010 and the correct values are given in the table below.

EBU Test	Results (LUFS)		
	Listed	-8 Gate	-10 Gate
3	-23.0	-23.06 to -23.08	-23.06 to -23.08
4	-23.0	-23.05 to -23.06	-23.06 to -23.09
5	-23.0	-23.01 to -23.02	-23.02 to -23.03

The range given for the correct values reflects the variability introduced by relative timing between the waveform and the meter. The bias in the results consumes almost the entire +/- 0.1 LU tolerance. The bias can be eliminated by changing the test 3 & 4 -23.0 dBFS input to -22.925 dBFS and the -20.0 dBFS input to -19.965 dBFS in test 5.

Test 10 - 12 Loudness Range calculation

Tests 10, 11 and 12 are implementations of EBU Tech 3342 Test Cases 1, 2 and 3 which evaluate basic loudness range meter operation. Each test consists of a 1 kHz sinewave which spends 20 seconds at each of two amplitudes. The loudness range meter should read the difference between these amplitudes. They are

Test	Amplitude 1	Amplitude 2	Reading
10	-20 dBFS	-30 dBFS	10 LU
11	-20 dBFS	-15 dBFS	5 LU
12	-40 dBFS	-20 dBFS	20 LU

Test 13 Loudness Range relative gate

This implements EBU Tech 3342 Test Case 4 which evaluates loudness range meter gating. It switches a sinewave between several amplitudes to test relative gating in the loudness range algorithm. The sequence is:

Amplitude	Duration
-50 dBFS	20 s
-35 dBFS	20 s
-20 dBFS	20 s
-35 dBFS	20 s
-50 dBFS	20 s

The average amplitude is -26.8 dBFS which will create a relative gate threshold of -46.8 dBFS. The -50 dBFS segments should therefore be ignored. A compliant meter will read 15 LU. If gating is not implemented the meter will read 30 LU.

Test 14, 15 Loudness Range Statistical Processing

Tests 14 and 15 are sinewave-based alternatives to EBU Tech 3342 Test Cases 5 and 6. The EBU versions are implemented with program clips. Sinewaves allow a controlled exercise of the gating function and an exact prediction of the correct loudness range. Program clips require evaluation with a reference meter.

Test 14 implements a narrow loudness range (NLR) signal and 15 implements a wide loudness range (WLR) signal. Each uses a 1 kHz sinewave whose amplitude changes in the sequence below.

Test 14	Test 15	Duration
-50 dBFS	-50 dBFS	20 s
-40 dBFS	-40 dBFS	3 s
-25 dBFS	-35 dBFS	23 s
-20 dBFS	-20 dBFS	23 s
-15 dBFS	-15 dBFS	2 s
-20 dBFS	-20 dBFS	23 s
-25 dBFS	-35 dBFS	23 s
-40 dBFS	-40 dBFS	3 s
-50 dBFS	-50 dBFS	20 s
	Total	140 s

The -40 dBFS amplitude occurs for 6 seconds and the -15 dBFS amplitude for 2 seconds. These short durations test the 10% and 95% statistical processing defined in the loudness range algorithm.

Test 14 should measure 5 LU while test 15 should measure 15 LU.

Test 16 Relative Gate Value

The older EBU documents call for a -8 dB relative gate while the ITU BS1770-2011 and the new EBU documents will specify a -10 dB relative gate. This test is designed to determine which gate value is implemented in the meter under test.

Time	Amplitude
0.0-0.5 s	-90.0 dBFS
0.5-1.4 s	-19.9 dBFS
1.4-2.7 s	-7.1 dBFS
2.7-3.5 s	-19.9 dBFS
3.5-4.0 s	-90.0 dBFS

If the relative gate is -8 this signal will give a loudness reading between -8 and -8.2. A -10 dB gate will give a reading between -10 and -10.2.

Summary

This test suite is available at no charge from the Qualis Audio website www.qualisaudio.com. If a loudness meter gives the expected results for each of the tests above there is a very high likelihood that the implementation is compliant with the latest version of BS1770.

These tests were developed in collaboration with Prism Sound. An automation script which controls generation of equivalent test signals by their dScope III platform instead of the wave files described here is available from their web site www.prismsound.com

About the authors

Richard Cabot is the CTO of Qualis Audio and was previously CTO of Audio Precision. He was Chair of the AES digital audio measurement committee and led development of the AES-17 standard.

Ian Dennis is Technical Director of Prism Sound and, as Vice-Chair of the AES digital audio measurement committee, wrote the document which became the true-peak meter specification in BS1770.

References

ATSC Recommendation A/85: “Techniques for Establishing and Maintaining Audio Loudness for Digital Television” November 2009 http://www.atsc.org/cms/standards/a_85-2009.pdf

EBU Recommendation R128: “Loudness normalisation and permitted maximum level of audio signals” http://tech.ebu.ch/docs/r/r128.pdf

EBU Tech 3341: “Loudness Metering: ‘EBU Mode’ metering to supplement loudness normalisation in accordance with EBU R 128” December 2010 http://tech.ebu.ch/docs/tech/tech3341.pdf

EBU Tech 3342: “Loudness Range: A descriptor to supplement loudness normalisation in accordance with EBU R128” December 2010 http://tech.ebu.ch/docs/tech/tech3342.pdf

The EBU R128 test material is available from http://tech.ebu.ch/webdav/site/tech/shared/testmaterial/ebu-loudness-test-setv01.zip

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

We're Listening