



White Paper

Using the ITU BS.1770 and CBS Loudness Meters To Measure Loudness Controller Performance

By Robert Orban
Chief Engineer, Orban
February 10, 2010

The ATSC has recently released a Recommended Practice: *Techniques for Establishing and Maintaining Audio Loudness for Digital Television (A/85:2009)*. This specifies use of a long-term loudness meter based on the ITU BS.1770 algorithm for assessing and setting the loudness of DTV broadcasts.

For many years, Orban has used the Jones & Torick loudness controller and loudness measuring technology¹ in its products for loudness control of sound for picture. Developed after 15 years of psychoacoustic research at CBS Laboratories, the CBS loudness controller accurately estimates the amount of perceived loudness in a given piece of program material. If the loudness exceeds a preset threshold, the controller automatically reduces it to that threshold. The CBS algorithm has proven its effectiveness by processing millions of hours of on-air programming and greatly reducing viewer complaints caused by loud commercials.

Since first licensing and using the CBS algorithm in its Optimod-TV 8182 back in the early '80s, Orban has continually refined and developed this technology. In the last

¹ Jones, Bronwyn L.; Torick, Emil L., "A New Loudness Indicator for Use in Broadcasting," J. SMPTE September 1981, pp. 772-777.

quarter century, audio processors from Orban and CRL using the CBS loudness controller have processed millions of hours of on-air television programming — an unsurpassed track record that no other subjective loudness controller technology can claim.

Because the ATSC recommends the BS.1770 algorithm, many engineers facing the problem of controlling broadcast loudness have wondered how the CBS and BS.1770 technologies compare. The purpose of this paper is to present, using both meters, comparative measurements of the output of Orban's current audio processors² that use our latest refinement of the CBS loudness controller technology.³

Test Setup

A stereo recording of approximately 30 minutes of unprocessed audio from the output of the master control of a San Francisco network station was applied to the 2.0 processing chain of an Optimod-Surround 8685 processor, set for normal operation using its TV 5B GEN PURPOSE preset. The digital output of the processor was applied to the digital input of an Orban 1101 soundcard, which was adjusted to pass the audio without further processing and to apply it to an Orban software-based loudness meter⁴ that simultaneously computes the BS.1770 loudness and CBS loudness. The first 750 seconds of the program material were a daytime drama with commercial and promotional breaks, while the remainder was local news, also with commercial and promotional breaks.

BS.1770 meter was adjusted to produce a 10-second integration window in which, per the BS.1770 standard, all data are equally weighted. The CBS Loudness Gain control was set to +5.5 dB. Data were logged every 10 seconds and included the maximum meter indication produced by both the BS.1770 and CBS meters in each 10-second interval. This produced a total of 163 data points, which were imported into a scientific plotting application (PSI Plot).

² Optimod-Surround 8585 and 8685, Optimod 6300 (with version 2.0 and higher software), and Optimod-PC 1101 and 1101E (with version 2.0 and higher software).

³ For a discussion of the CBS and BS.1770 technologies, see <http://orban.com/meter/Technology.html>. The ATSC A/85:2009 document also discusses the BS.1770 algorithm.

⁴ This software is available for free download at <http://orban.com/meter/>. However, the free version of the software does not support the logging functions that the author employed to acquire data for the figures in this paper.

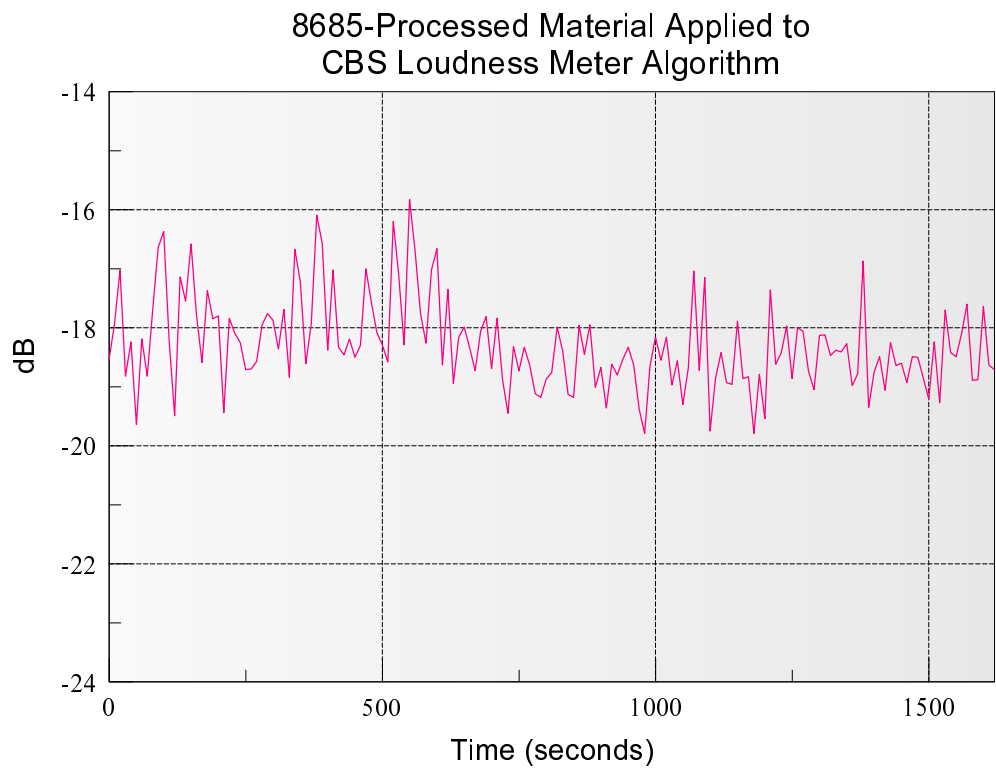
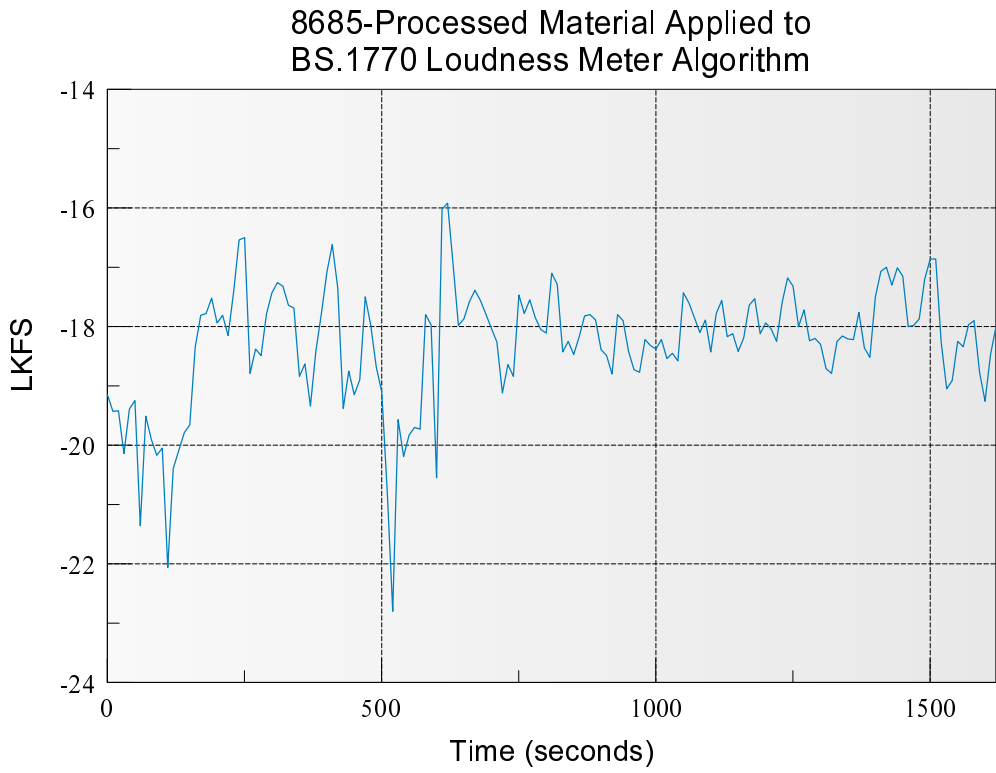


Figure 1: Peak Output of the CBS and BS.1770 Loudness Meter in Each 10-second Interval as a Function of Time

Results

Figure 1 shows the results as a function of time for both meters. The peak CBS readings fit within a ± 2 dB window. The BS.1770 readings also fit within a ± 2 LKFS⁵ window except for three short intervals. These intervals correspond to dialog without background music and in the author's opinion illustrate a weakness in the version of the BS.1770 algorithm current at this writing. Television dialog contains inter-syllable pauses, pauses for breath, and sometimes pauses for dramatic effect. The ATSC Recommended Practice describes the problem as follows⁶:

Because the measurement of loudness per BS.1770 is an integrated measurement, quiet passages tend to lower the measured value. To avoid this, the integration may be paused during quiet passages. Automatic triggering, hold and resume, generally called "gating", is being studied by some organizations including the ITU-R, and gating may be added to BS.1770 in the future. Some equipment may offer gating as a feature. As yet, there are no standards for gated loudness measurements. Users should utilize the current version of BS.1770 for measurements.

The popular Dolby LM100 Loudness Meter⁷ in its current revision uses the same Leq(RLB) algorithm as BS.1770, but adds gating to eliminate non-speech material, including silence. The author has used the Dolby LM100 to measure the output of the Orban 8685 with a wide variety of speech material, and has observed that this material is almost always controlled within a ± 1 dB window as measured on the LM100. In the author's opinion, this demonstrates the benefits of a gated measurement. Moreover, the author believes it is unwise to rely on an ungated BS.1770 measurement to set the on-air loudness of unadorned dialog because this can cause the dialog to be too loud with respect to other material.

Histograms

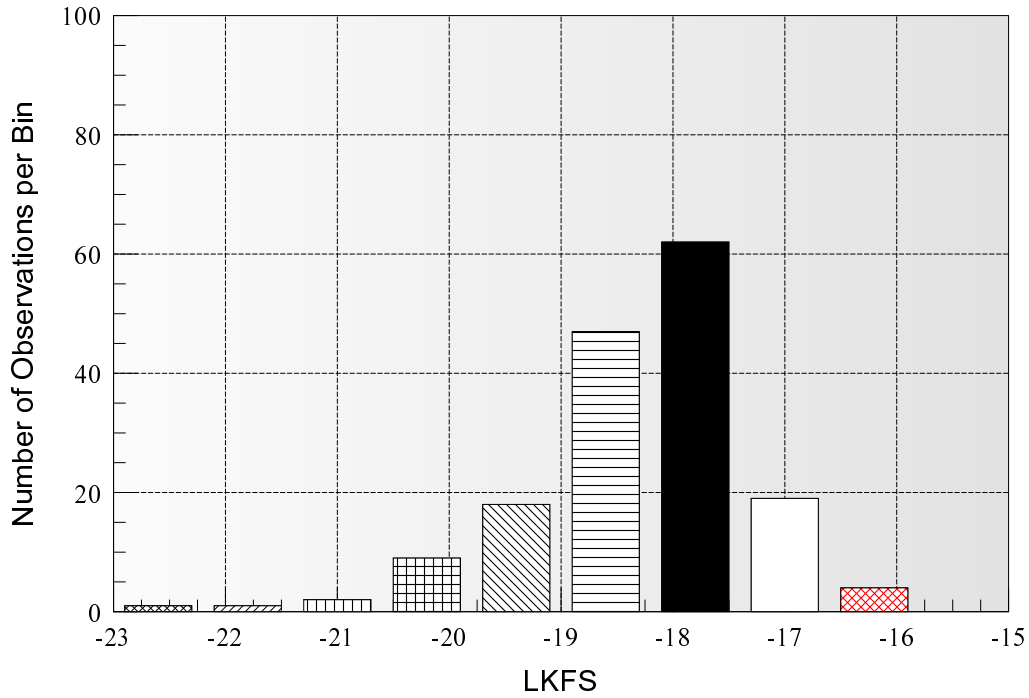
The measurements can also be presented as histograms (Figure 2) that sort the magnitude of the measurements into 1 dB or 1 LKFS-wide slices and show the number of measurements that fit into each of these slices. The histogram thus provides an effective portrayal of the consistency of the loudness — when the data are tightly clustered within a few bins, this indicates that the loudness is more consistent than it would be if the data were spread out into a larger number of bins.

⁵ In the BS.1770 standard, measured loudness is reported as LKFS. A unit of LKFS is the same measure as a decibel. A -15 LKFS program can be made to match the loudness of a quieter -22 LKFS program by attenuating it by 7 dB.

⁶ ATSC A/85:2009, Section 5.2, p. 15

⁷ <http://www.dolby.com/professional/products/broadcast/test-and-measurement/lm100.html>

BS.1770 Histogram
 Integration Time = 10 sec
 Uniformly Sampled every 10 Seconds



CBS Loudness Meter Histogram
 Maximum Meter Indication in Each 10-Second Period

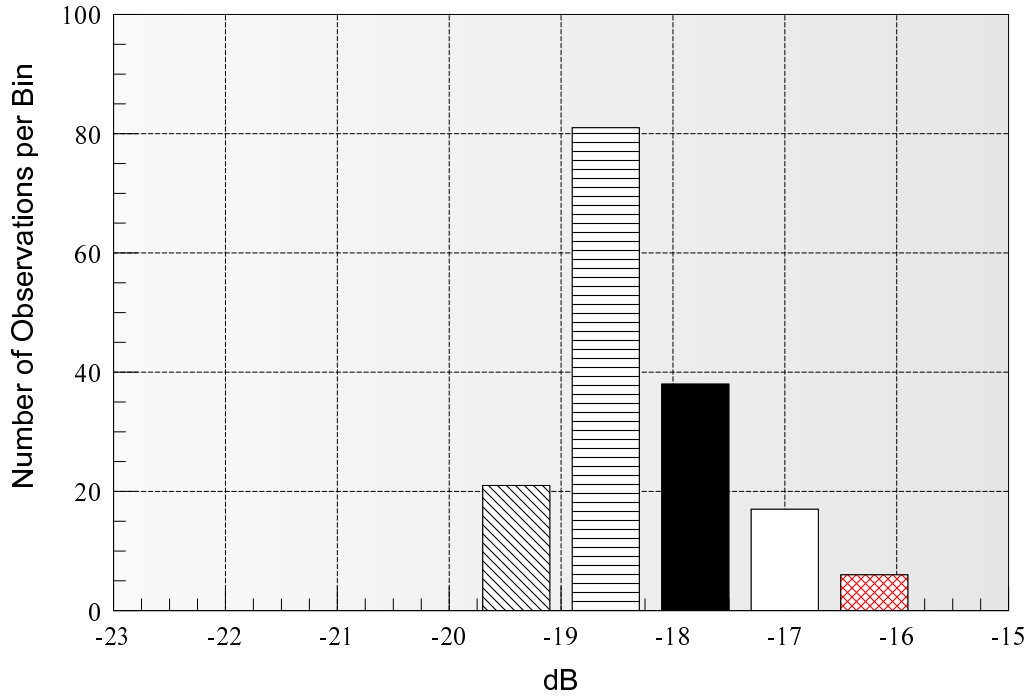


Figure 2: Histograms sorting loudness measurements into 1 dB bins.

Both histograms show the loudness is well controlled, although the details are different. Both the BS.1770 and CBS measurements indicate that most of the data points are in a ± 1 dB window. However, the BS.1770 has low probability outliers that are absent in the CBS graph. These outliers represent the aforementioned periods of unadorned dialog that measure low because the BS.1770 standard does not yet specify silence gating.

Studies indicating that BS.1770 is inaccurate at very low frequencies

In addition to its lack of silence gating, another weakness of BS.1770 is that, unlike the CBS loudness controller and meter as implemented in Orban products, the BS.1770 algorithm does not take into account the effect of the LFE channel, for good reason. Nacross and Lavoie⁸ recently tried to extend the BS.1770 algorithm to include the LFE channel by summing the K-weighted LFE channel’s power into the current BS.1770 algorithm, where the gain is weighted for the fact that LFE channel receives a 10 dB gain boost on playback, per Dolby’s standards. This modified BS.1770 algorithm failed to agree with the judgments of a subjective listening panel unless a 10 dB attenuation “fudge factor” was applied to the LFE channel prior to its power summation with the other channels. Nacross and Lavoie concluded:

A problem exists however, should ITU-R BS.1770 be modified to simply include an attenuated version of the LFE channel. Because the LFE channel receives a 10 dB boost on playback, the low-frequencies on this channel would contribute differently to a loudness measure if they were moved to one of the other main channels, even though the perceived loudness would not appreciably change. This suggests that while LFE content does contribute to the perceived loudness, Equation (2)⁹ does not sufficiently predict how that content should be included.

A recent Australian study may shed light on the failure of BS.1770 when program material contains considerable energy at very low frequencies.¹⁰ The authors used octave-band noise in subjective listening tests with the goal of verifying the K-weighting curve used in BS.1770. The authors state:

⁸ Norcross, Scott G; Lavoie, Michel C., “Investigations on the Inclusion of the LFE Channel in the ITU-R BS.1770-1 Loudness Algorithm,” AES Convention Paper 7829, 127th AES Convention, New York 2009

$$\begin{aligned}
 {}_9 \text{Leq}(w) &= \left[\frac{1}{T} G_{LFE} \int_T^0 \frac{x_w^2}{x_{ref}^2} dt + \sum_i \frac{1}{T} \int_0^T \frac{x_{w,i}^2}{x_{ref}^2} dt \right], dB \\
 i &= L, R, C, L_s, R_s
 \end{aligned}$$

¹⁰ Cabrera, Densil; Dash, Ian; Miranda, Luis, “Multichannel Loudness Listening Test,” AES Convention Paper 7451, 124th AES Convention, Amsterdam 2008

Comparison of the test results with an image of the filter curve currently specified in ITU-R Recommendation BS.1770 (Figure 13) shows good agreement at 250 Hz and above 500 Hz, reasonable agreement at 500 Hz, but marked difference in the bottom two octaves. The relatively good performance of the BS.1770 algorithm in ITU trials suggests that, in partial loudness terms, there was probably not much test content in the 125 Hz band or below. While the existing BS.1770 filter curve is probably a good choice in applications where the program is dominated by speech, and it is certainly an improvement on the A and B curves in that application, it is likely to give significant errors in measuring the loudness of other programs with more partial loudness in the lower frequencies, such as movie soundtracks and popular music. It is therefore desirable to improve on this filter for more general measurement of program loudness.

Conclusions

Several studies have shown that the loudness “comfort range” for typical television listening is +2, -5 dB¹¹. Beyond this range, a viewer is likely to become annoyed, eventually reaching for the remote control to change volume (or worse from the broadcaster’s point of view, to mute a commercial). Whether measured via the CBS or BS.1770 algorithms, the CBS loudness controller algorithm in Orban’s current products effectively controls subjective loudness to much better than this +2, -5 dB window.

The results using BS.1770 metering would be even more consistent if that algorithm employed silence gating (such as that used in the Dolby LM100) to prevent unadorned dialog from reading low compared to music and dialog with substantial background music or effects. The CBS algorithm does not need silence gating because it is a “short-term” loudness measurement that incorporates models of the loudness integration time of human hearing, which the BS.1770 algorithm does not.

Moreover, controlling loudness to a standard such as BS.1770 says nothing about the subjective acceptability of the loudness controller’s action. Automatic loudness controllers can produce all of the well known artifacts of dynamics processing, including noise breathing, spectral inconsistency, gain pumping, and harshness. Improperly designed multiband compressors can reduce dialog intelligibility¹². This is why it is important to carefully assess the audio quality and side effects that an automatic loudness controller produces so that one can choose a device that controls loudness effectively without producing objectionable and unnatural artifacts that can fatigue audiences. All loudness controllers are not created equal even if they produce identical measurements on a loudness meter.

¹¹ ATSC A/85:2009 Annex E, “Loudness Ranges”

¹² Stone, Michael A.; Moore, Brian C. J.; Füllgrabe, Christian; Hinton, Andrew C., “Multichannel Fast-Acting Dynamic Range Compression Hinders Performance by Young, Normal-Hearing Listeners in a Two-Talker Separation Task,” J., AES Volume 57 Issue 7/8 pp. 532-546; July 2009