orban White Paper

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

By Robert Orban
Chief Engineer, Orban
revised May 20, 2014

ITU-R BS.1770

In 2009, the ATSC released a Recommended Practice: *Techniques for Establishing and Maintaining Audio Loudness for Digital Television* (A/85:2009). This was later updated as A/85:2013. A/85 specifies use of a long-term loudness meter based on the current version of the ITU BS.1770 algorithm for measuring the loudness of DTV broadcasts.

In December 2011, the FCC adopted rules implementing the CALM Act[1], which, by law, forbids commercials from being louder than non-commercial program material. The new FCC rules incorporated ATSC A/85 (and, by implication, the BS.1770 meter) as an objective means of verifying that the rule was being obeyed.

Because loudness measurement per BS.1770 uniformly integrates all program material, quiet passages tend to lower the measured value. To prevent this, the ITU added gating to the BS.1770 standard, which was revised as BS.1770-2 in March 2011[2]. The gating causes the meter to ignore silence and to integrate only program

[1] The CALM Act applies only to U.S. broadcasters and cable providers.

[2] The BS.1770-3 update in 2012 did not affect the loudness-metering algorithm.

material whose loudness falls within a floating window extending from the loudest sounds within the specified integration period to sounds that are 10 dB quieter than the loudest sounds. This is because humans tend to assess loudness based on the louder sounds in a given program.

The ATSC A/85, ITU-R BS.1770, and EBU R 128 documents are available as free downloads and their current versions can easily be located with a search engine.

CBS Loudness Meter

For many years, Orban has used the Jones & Torick loudness controller and loudness measuring technology[3] 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 the CBS algorithm and using it in its Optimod-TV 8182 in the early 1980s, Orban has continually refined and developed this technology. In the last 30+ years, 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.

Comparing the Meters

Because the ATSC recommends the BS.1770 algorithm, many broadcast and cable engineers facing the problem of controlling broadcast loudness have wondered how the CBS and BS.1770 technologies compare. An earlier version of this Orban white paper compared the CBS and BS.1770-1 (non-gated) meter. The paper you are now reading was revised in March 2012 to incorporate results from tests using the BS.1770-3 algorithm and EBU – TECH 3342 "Loudness Range" algorithm. The new measurements were performed using Version 2 of the Orban Loudness Meter[4]. This revision compares the CBS and BS.1770-3 meters.

---

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

[4] This software is available for free download at http://orban.com/meter/.

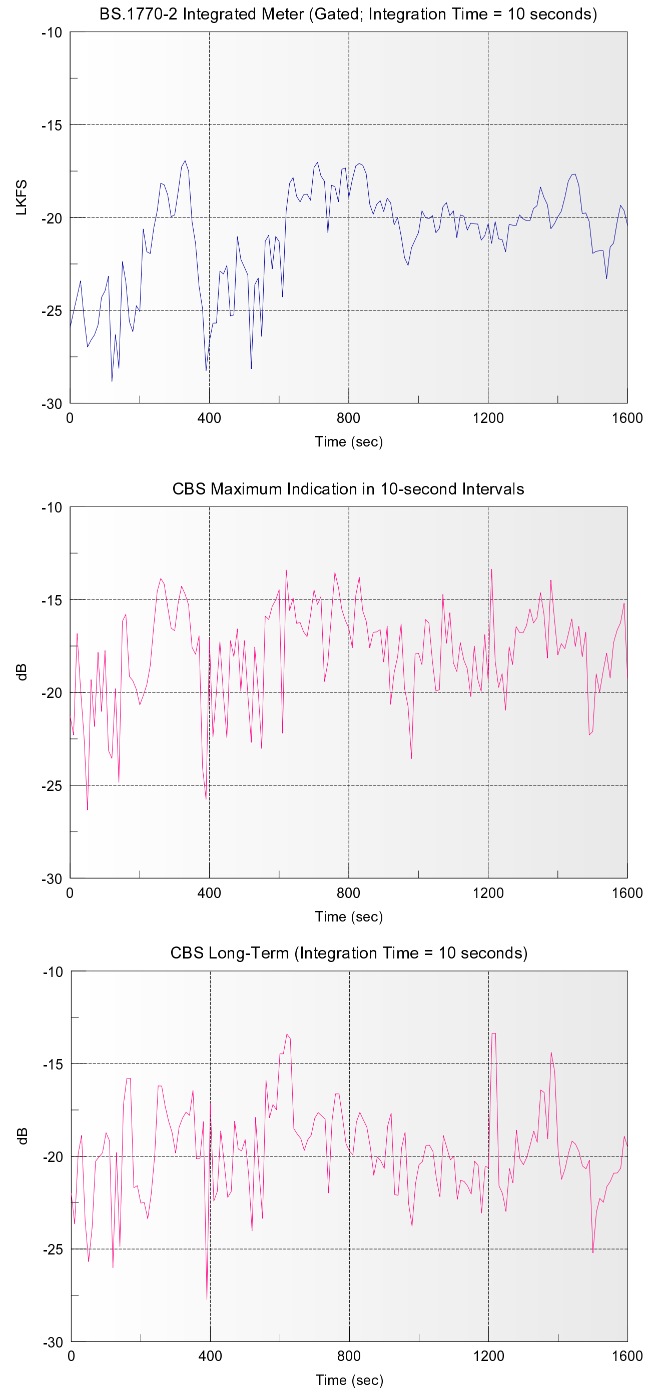

**Figure 1: Unprocessed Input—
Peak Output of the BS.1770 and CBS
Loudness Meters in each 10-second Interval as
a Function of Time**

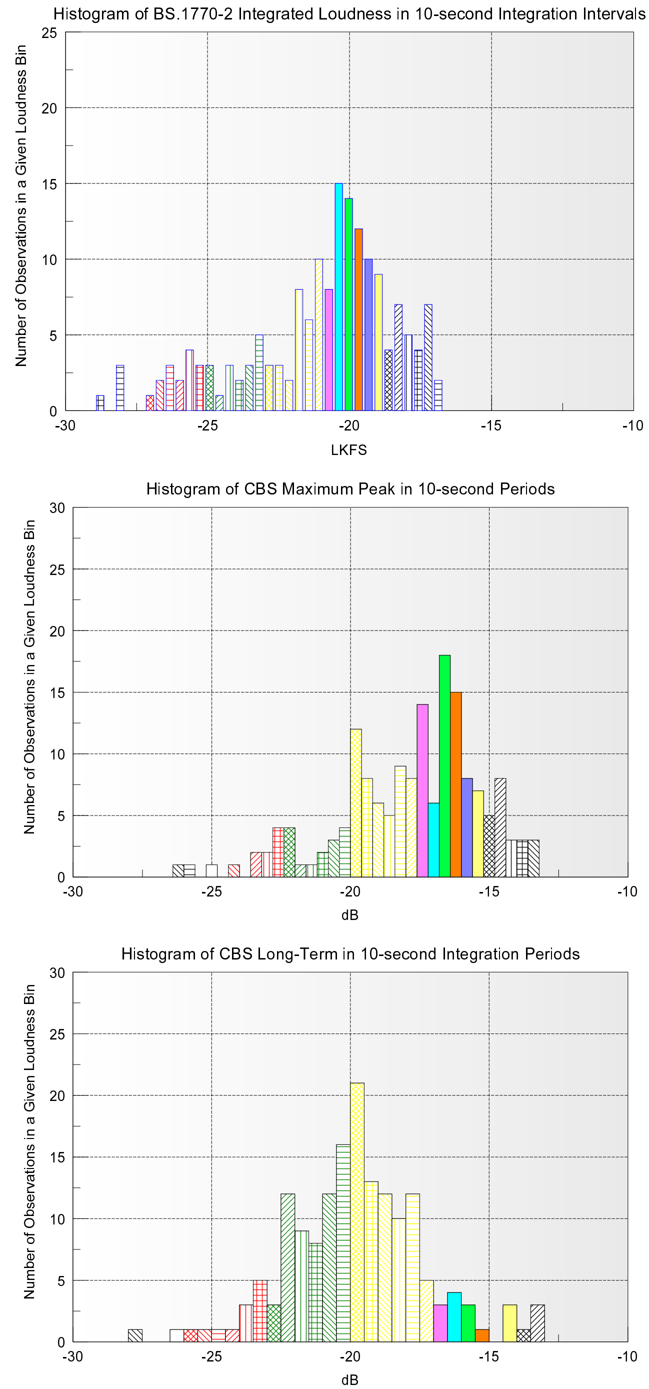

**Figure 2: Unprocessed Input—
Histograms sorting loudness measurements
into 0.25 dB bins.**

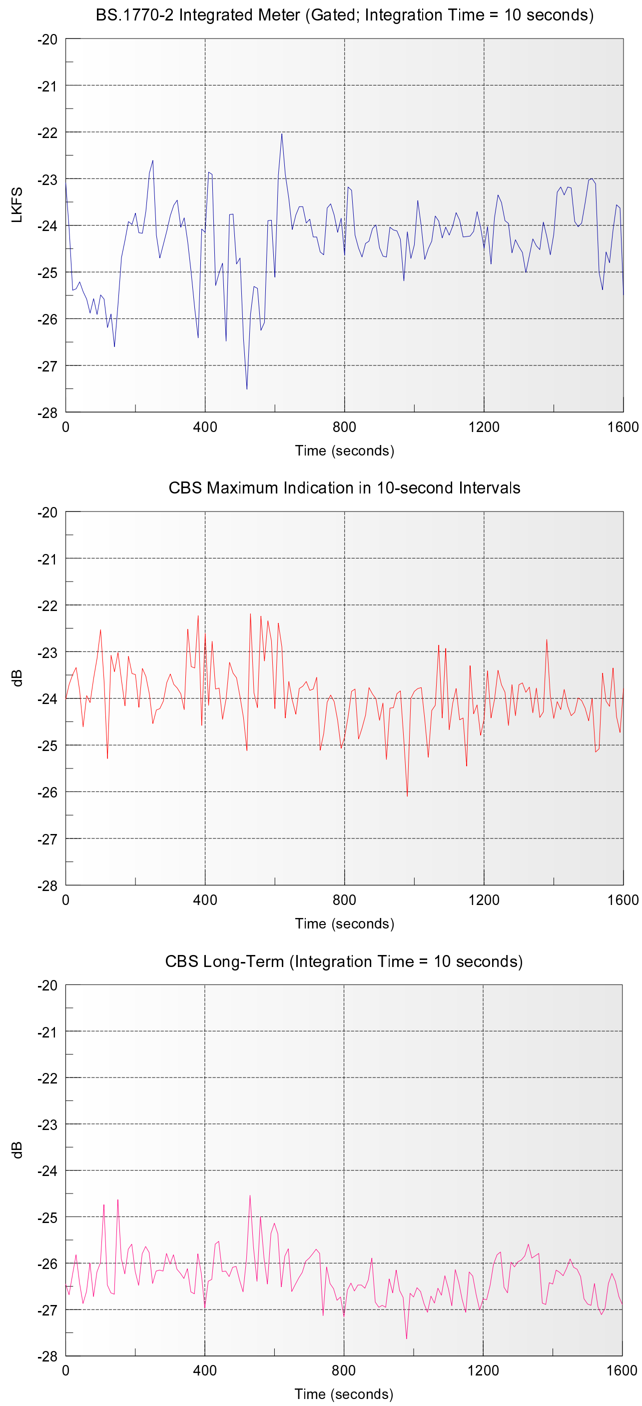

Figure 3: Loudness-Controlled Audio—Peak Output of the CBS and BS.1770 Loudness Meters in each 10-second Interval as a Function of Time

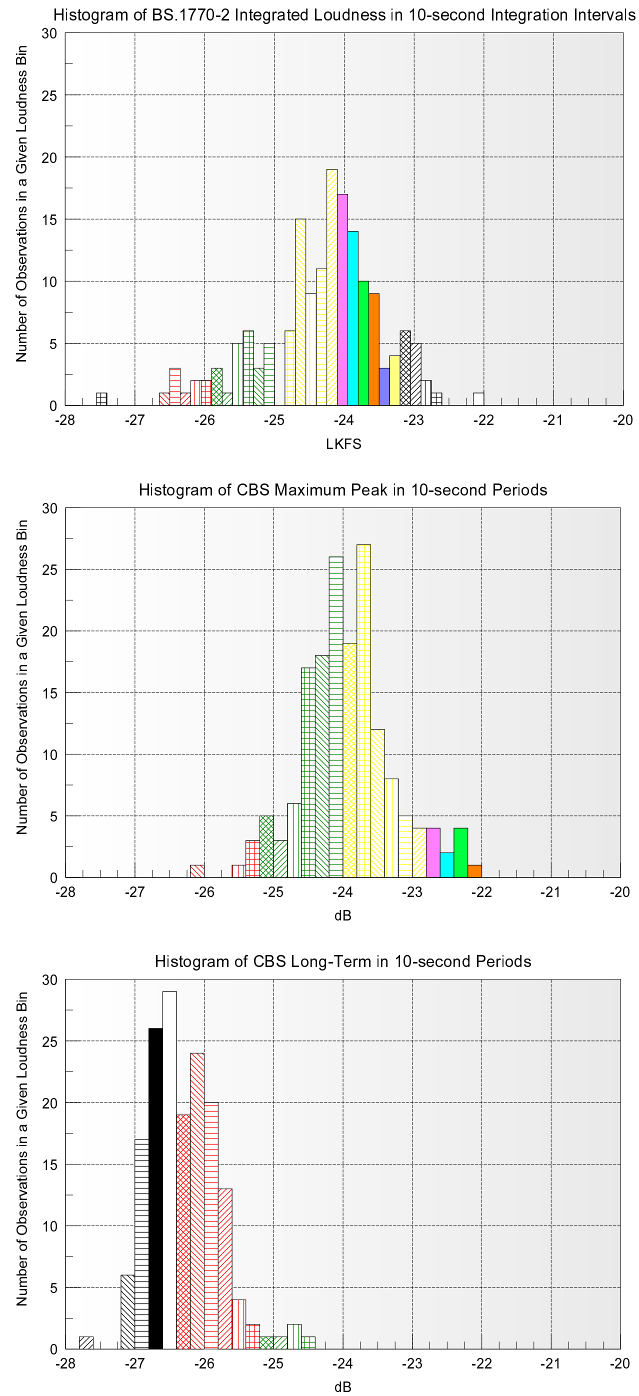

Figure 4: Loudness-Controlled Audio—Histograms sorting loudness measurements into 0.25 dB bins

A/85 and R 128 differ significantly in their philosophy and recommendations. Probably most important difference is that A/85 asserts that the loudness of a socalled “anchor element” (which is typically dialog except in programs emphasizing music, like live concert recordings) is most important, while R 128 asserts that the integrated loudness of the entire program is most important[5] and therefore, program loudness should be normalized based on an integrated BS.1770-3 measurement. The philosophy behind A/85 is similar to that of Dolby Laboratories, which for many years has asserted that dialog anchors most film and television programs and that listeners set their volume controls to make dialog comfortably intelligible[6]. (We agree more with A/85 than with R 128).

The purpose of this paper is to present, using both meters, comparative measurements of the output of Orban’s current audio processors[7] that use our latest refinement of the CBS loudness controller technology.[8]

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 meter that simultaneously computes the BS.1770-3 Integrated loudness and CBS loudness. The first 750-second segment of the program material was a daytime drama with commercial and promotional breaks, while the remainder was local news, also with commercial and promotional breaks.

The BS.1770-3 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 –3.12 dB. Data were logged every 10 seconds and included the maximum meter indication produced by both the BS.1770 and CBS meters in

---

[5] EBU – TECH 3343, “Practical guidelines for Production and Implementation in accordance with EBU R 128,” version 1 (February 2011), p. 29

[6] Riedmiller, J., Lyman, S., Robinson, C., “Intelligent program loudness measurement and control: what satisfies listeners?” AES Convention Paper 5900, 115th Convention (October 2003)

[7] Optimod-Surround 6585 and 8685, Optimod 6300 (with version 2.0 and higher software), and Optimod-PC 1101 and 1101E (with version 2.0 and higher software).

[8] For a further discussion of the CBS and BS.1770 technologies, see http://orban.com/meter/Technology.html. The ATSC A/85:2013 document also discusses the BS.1770 algorithm.

each 10-second interval. This produced 165 data points, which were imported into a scientific plotting application[9].

Orban's experimental long-term loudness measurement, based on the CBS meter and first published in 2008, was also included in the measurements and is shown in the bottommost charts. This algorithm attempts to mimic a skilled operator's mental integration of the peak swings of a meter with "VU-like" dynamics. The operator will concentrate most on the highest indications but will tend to ignore a single high peak that is atypical of the others. This algorithm can be seen to share certain characteristics with the floating gate first introduced in EBU R 128 and later adopted in BS.1770-3.

The Orban algorithm displays the average of the peak indications of the meter over a user-determined period: 10 seconds for these measurements. The average is performed before dB conversion. All peak indications within the period are weighted equally with the following exceptions:

- If the maximum peak in the window is more than 3 dB higher than the second highest peak, it is discarded.
- All peaks more than 6 dB below the maximum (or second-to-maximum, if the maximum peak was discarded) are discarded.

Because the CBS long-term measurement discards a single peak if it is more than 3 dB higher than the second highest peak, the CBS long-term measurement tends to be biased about 3 dB lower than a measurement that shows the maximum peak indication of the CBS meter in a 10-second period. The amount of bias depends on whether or not the loudness applied to the meter's input is well controlled. This bias can be seen in the figures in the paper. Because the Orban meter allows control of the level applied to the CBS algorithm via the "CBS Gain" control, setting it 3 dB higher could better match the CBS long-term measurement to the BS.1770-3 Integrated measurement at the expense of moving the "maximum peak loudness" indication 3 dB higher.

Results

Unprocessed Audio Input: To provide a baseline for discussion of the loudness-controlled results, we measured the unprocessed audio that was applied to the Optimod 8685's input. Figure 1 and Figure 2 on page 3 show the loudness of the unprocessed audio both as a function of time and as a histogram. The histogram

[9] PSI Plot: http://www.polysoftware.com/plot.htm

sorts the meter outputs into 0.25 dB or 0.25 LK[10]-wide slices and shows the number of measurements that fit into each of these slices. The histogram thus portrays loudness consistency — when the histogram is clustered tightly within a few bins, the loudness is more consistent than it is when the histogram is spread out into a larger number of bins.

With all meters, the histogram of the unprocessed audio shows a wide spread. This is consistent with the EBU Loudness Range measurement for the entire clip, which was 16.5 LK, while the LRA for the daytime drama alone was 19.2 LK (including commercials). The BS.1770-3 Integrated loudness was –20 LKFS, integrated over the entire measurement period, although the inconsistencies between the loudness of program material and commercials are large enough to make this 30-minute measurement essentially meaningless.

In general, the loudest parts of the unprocessed audio are commercials and promos, both network and local. These are anywhere from 5 to 10 dB (or LK) louder than the rest of the program material. This inconsistency was not a problem because the station in question was using an Orban automatic loudness controller on-air, which smoothed out loudness differences before its input.

While the general shapes of the CBS and BS.1770 loudness vs. time curves are similar, there were some significant differences. For example at approximately 1250 seconds, the CBS measurement shows a sharp loudness spike that was caused by a network news report that was equalized to emphasize frequencies around 2 to 3 kHz, where the ear is most sensitive. The BS.1770-3 measurement did not indicate this as being louder than the surrounding program material although to our ears, it clearly was.

Loudness-Controlled Audio: Figure 3 and Figure 4 on page 4 show the results after automatic loudness control. (To present the data with optimum graphic resolution, we made the loudness scales of Figure 3 and Figure 4 narrower than the scales in Figure 1 and Figure 2.)

Both the loudness vs. time graphs and the histograms show the Orban 8685 controls loudness well, although the details of the meters' indications are different. Both the BS.1770 and CBS measurements indicate that most of the data points are in a ±1 dB/LK window.

The peak CBS readings fit within a ±2 dB window. The BS.1770 readings also fit within a ±2 LK window except for four short intervals, which appear as low-

[10] Unfortunately, two terms for the same loudness units have been used in different standards documents. For convenience, we will use LK and LKFS (as used in ATSC A/85); these units are the same as LU and LUFS (used in EBU R 128 and BS.1770) respectively.

probability outliers in the left side of the histogram. These intervals correspond to dialog without background music and in the author's opinion illustrate a weakness in BS.1770-3: based on our extensive listening tests, we have concluded that the meter does not effectively lock onto the A/85 "anchor element" (almost entirely dialog in the test material used to prepare this paper) and instead indicates that loudness increases when dialog level is held constant while underscoring or effects are added to the mix.[11]

Problems with Low Peak-to-RMS Ratio Material

In the subjective testing to validate the BS.1770 meter, there were outliers as large as 6 dB (i.e., the meter disagreed with human subjective perception by as much as 6 dB[12].) The subjective testing to validate the CBS meter found outliers up to 3 dB, although fewer items were used in this testing. We hypothesize that the fact that the worst-case error of the BS.1770 meter was substantially larger than that of the CBS meter is caused by the BS.1770's meter's not modeling loudness summation or the loudness integration time constants of human hearing.

BS.1770-3 states:

> It should be noted that while this algorithm has been shown to be effective for use on audio programmes that are typical of broadcast content, the algorithm is not, in general, suitable for use to estimate the subjective loudness of pure tones.

We have noted that the meter tends to over-indicate the loudness of program material that had been subject to large amounts of "artistic" dynamic compression, as is often done for commercials and promotional material — in other words, the meter over-indicates the loudness of program material having an unusually low peak-to-average ratio, which, at the limit, approaches the peak-to-average ratio of a pure tone. We have encountered heated complaints by mixers[13] and producers who stated that such material, when "matched" to the loudness of the surrounding program material via the BS.1770 meter, is considerably quieter in subjective terms. In turn, this has constrained the ability of producers to specify the type of audio

---

[11] In the first published version of the paper, we observed the similar dips in the BS.1770-1 (ungated) loudness and hypothesized that they were caused by lack of gating on silence and low-level material. For this reason, we were surprised that BS.1770-2 gating made little difference in the measurements of this material.

[12] Refer to the scatter plots in Figs. 11, 12, and 13 of the ITU-R BS.1770 standard.

[13] For example: "I did a –24 [LKFS] piece for Fox that was wall to wall singing and music for two minutes. Because of the overall loudness and continued full audio signal I had to bring it down and when it aired, it was 3 db too quiet even though it matched the magic LKFS number. I have no problem using these meters or meeting specs but they are faulty." —"wheresmyfroggy," AVID board, 3-28-2011

processing they had previously used to give this material excitement and punch. We hypothesize that this problem is related to the fact that BS.1770 does not accurately indicate the loudness of pure tones.

Some studies have indicated that when people are asked to assess the loudness of a given piece of material, they state that it sounds louder when underscoring or effects are added to constant-level dialog. The EBU has used these studies to justify the position taken in R 128 that a listener's impression of total loudness is more important than dialog level[14]. In our opinion, this misses the point. A more relevant question is whether viewers would want to turn down their volume controls to make dialog quieter when underscoring and effects appear. (In other words, whether effective TV commercial loudness control requires nothing more than applying gain control to commercials such that the BS.1770-3 "short-term" loudness[15] is always limited to 0 LK.) Regarding this, Orban and Dolby Labs hold similar views. We believe that dialog is the most important element in most television audio and that listeners do not want to turn down their volume controls every time that underscoring or effects appear under the dialog. The popular Dolby LM100 Loudness Meter[16] 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 dialog-centric measurement. Moreover, the author believes it is unwise to rely on a 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. The author has experimented with "inverse short-term BS.1770 loudness control" and believes that it sounds unnatural, pumping dialog loudness up and down in a subtly inartistic way as underscoring and effects come and go.[17]

---

[14] Dash, Ian; Bassett, Mark; Cabrera, Densil, "Relative Importance of Speech and Non-Speech Components in Program Loudness Assessment," AES Convention Paper 8043, 128th AES Convention (May 2010).

[15] EBU R 128 specifies short-term loudness as a BS-1770-1 (ungated) measurement with a three-second integration time.

[16] http://www.dolby.com/professional/products/broadcast/test-and-measurement/lm100.html

[17] See Begnert, Fabian; Ekman, Håkan; Berg, Jan, "Difference between the EBU R-128 Meter Recommendation and Human Subjective Loudness Perception," AES Convention Paper 8489, 131st AES Convention, (October 2011). This paper states, "These loudness-equalized signals gave rise to a perceived maximum loudness difference of 2.8 dB." This is very close to the 3 dB number that has come up in other discussions (such as the one quoted in footnote 13

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

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 loudness contributed by the LFE channel, for good reason. Nacross and Lavoie[18] 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)[19] does not sufficiently predict how that content should be included.

An Australian study may shed light on the failure of BS.1770 when program material contains considerable energy at very low frequencies.[20] 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:

> 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

---

on page 8). While the authors of this paper consider 3 dB to be insignificant, others do not necessarily share this view, particularly advertisers who hear their expensive commercials aired 3 dB quieter than surrounding program material!

[18] 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 (October 2009)

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

[20] Cabrera, Densil; Dash, Ian; Miranda, Luis, "Multichannel Loudness Listening Test," AES Convention Paper 7451, 124th AES Convention (May 2008)

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.

Discussion and Conclusions

Several studies have shown that the loudness "comfort range" for typical television listening is +2, –5 dB[21]. 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.

In the original version of this paper, we had assumed that results using BS.1770 metering would be more consistent if that algorithm employed gating to prevent unadorned dialog from reading low compared to music and dialog with substantial background music or effects. However, this did not prove to be true with the program material we used for testing—the results from the BS.1770-1 (ungated) and BS.1770-3 (gated) measurements were similar when measuring material that had been processed by the CBS Loudness Controller. It is likely that the loudness-controlled material seldom caused the gate to act. (The CBS algorithm does not need silence gating because it is a "short-term" loudness measurement that incorporates cascaded models of the "instantaneous" and "short-term" loudness time constants of human hearing[22], which the BS.1770 algorithm does not.)

Controlling loudness to a standard such as BS.1770 says nothing about the subjective acceptability of the loudness controller's action. We have found that a simple loudness controller that uses the inverse of the BS.1770 short-term meter's output to control loudness by gain reduction can cause unnatural-sounding gain pumping of dialog when underscoring and effects appear under the dialog. More complex 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 intelligibility[23]. This is why it is important to carefully assess the audio quality

---

[21] ATSC A/85:2009 Annex E, "Loudness Ranges"

[22] For example, see Glasberg, B.R. & Moore, B.C.J. (2002) "A Model of Loudness Applicable to Time-Varying Sounds," J.AES, vol.50:5, pp.331-342, May 2002.

[23] Stone, Michael A.; Moore, Brian C. J.; Füllgrabe, Christian; Hinton, Andrew C., "Multichannel Fast-Acting Dynamic Range Compression Hinders Performance by Young,

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. Different loudness controllers do not provide equally good subjective results even if they produce identical measurements on a loudness meter.

Based on extensive experimentation with typical broadcast material, we believe that the CBS loudness meter locks onto dialog more effectively than does BS.1770, particularly when the dialog is accompanied by underscoring and/or effects. Accordingly, the CBS Loudness Controller in Orban products, which uses the CBS loudness metering algorithm as its core loudness reference, produces consistent and naturally balanced dialog levels regardless of the program material and mixing style. Unlike the BS.1770 meter, the CBS technology does not unnaturally penalize material having a low peak-to-RMS ratio, so it allows mixers and producers to freely use "artistic compression"[24] and other well-established production techniques with the knowledge that such material will be neither too loud nor too quiet when compared to the surrounding program.

Normal-Hearing Listeners in a Two-Talker Separation Task," J. AES Volume 57 Issue 7/8 pp. 532-546; July 2009

[24] It appears that the group that created R 128 may be biased against this style of production: "Again, this does NOT mean that within a programme the loudness level has to be constant, on the contrary! It also does NOT mean that individual components of a programme (for example, pre-mixes or stem-mixes, a Music & Effects version or an isolated voice-over track) have all to be at the same loudness level! Loudness variation is an artistic tool, and the concept of loudness normalisation according to R 128 actually encourages more dynamic mixing!" EBU TECH 3343, op. cit., p. 17