



Audio Engineering Society

Convention e-Brief 86

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for the contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Comparative analysis of different loudness meters based on voice detection and gating

Alessandro Travaglini¹

¹ Fox International Channels Italy, Rome, 00138, Italy
info@alessandrotravaglini.it

ABSTRACT

After decades of deep investigation, the international broadcasting community represented by technical associations and bodies has set precise standards aimed to objectively assess loudness levels of programmes. Although all standards rely on the same algorithm as described in ITU-R BS1770, there are still two possible ways to implement such metering, including Voice Detection and Gating. These two different implementations might, in some cases, provide measurements that significantly differ from each other. Furthermore, whilst the gating feature is uniquely defined in the updated version of BS1770-3, Voice Detection is not currently specified in any standard and its implementation is independently designed by manufacturers. This paper analysis this scenario by comparing the results and robustness provided by three different loudness meters based on Voice Detection. In addition, those values are compared with measurements obtained by using BS1770-3 compliant loudness meters, including tables, comments, and conclusions.

1. LOUDNESS MEASUREMENT IMPLEMENTATIONS

1.1. ITU-R BS1770 and further releases

In 2006 ITU released the first release of BS1770, the algorithm designed to assess the Loudness Level and to measure the True Peak Level of audio programmes in broadcasting. The algorithm is based on the computation of mean-square levels of the audio programme weighted according to the R2LB filtering and the gain levels described in the recommendation [1]. The audio signal is measured on all its duration and the final loudness level is produced by integrating all values gathered during the measurement. This

implies that no emphasis is put on any specific sound element, regardless their content or technical characteristic (voice, music, sound effect, intensity, pitch, background or foreground). As said, this method was not able to identify the so called “anchor sound” (typically the voice element of the audio mix) and resulted not being fully effective in case the audio programme presents variations in level such as loudness modulation between background and foreground sounds. Consequently, two approaches have been implemented in order to cope with those limitations and they include: Voice Detection and Gating. The latter is designed to compute the Programme Loudness Level by focusing the measurement on foreground sounds only. This result is obtained by discarding all values falling below a predetermined threshold. In 2012 ITU released an updated version of the its recommendation

where a gating feature was added by including a relative gating at -10LU [2].

1.2. Voice detection meters

It consists of a logic feature that automatically detects the voice content of a programme and that enables the gathering of loudness levels necessary to compute the final overall value only when those elements are detected. Consequently, only the programme parts that present voice elements will contribute in the computation of the overall loudness level representative of the whole programme. This implementation presents the positive capability to focus the measurement on the “anchor sound” that is mainly used by listeners to assess the perceived loudness level of a content, the speech. However, it also has the limitation to not being usable on content with no or very limited voice content. Furthermore, being based on non-standardized implementations and designs, the measurement obtained with voice detection meters might result less robust than others. This paper primarily aims to investigate this aspect.

The first manufacturer to implement a voice detection feature in its loudness meter was Dolby Inc. with the release of its Dialogue Intelligence technology in the Dolby LM100 Broadcast Loudness Meter in 2003. Recently, other companies have pursued the same goal and have included this option in their metering software. This research, aside Dolby LM100, includes the analysis of measurements produced by Nugen Audio VisLM-H and Waves Loudness Meter WLM.

1.3. Gating

In order to overtake the issues connected to the implementation of meters based on voice detection (lack of standardization, degree of robustness, usability on non-voice content) the gating feature was designed as previously described. The first to introduce it in a technical recommendation was EBU in R128 in 2010. Being an open standard, the gating feature rapidly gained consensus. In fact, in 2011 [3] the R128 was updated by setting the gating level at a relative threshold of -10LU, exactly how will later be amended by ITU in BS1770-3. Consequently, the current releases of both BS1770-3 and R128 describe the same method to determine the loudness level of an audio programme. The benefit of using loudness meters which implement the gating feature is represented by the possibility to

rely on standardized implementations, precisely defined in recommendations and regulations.

Despite that, gating is not able to identify and focus the measurement on the “anchor sound” that is typically used by content producers and listeners to judge the loudness level of a programme: the voice. This is the main difference in regard to Voice Detection and the test herewith described aims to highlight pros and cons of both methods.

2. TEST

The growing need of exchanging content material worldwide quickly and with the minimum possible operational impact requires solid technical references onto which basing workflows and content processing, reflecting in the necessity to apply one unique measurement implementation. Furthermore, the robustness of the measurement method could benefit from the merging of both the voice detection and the gating approaches. The uncertainty generated by the lack of standardization of the voice detection feature has so far slowed the discussion into that direction. This test aims, indeed, to provide objective analysis that could put some light into some unsolved issues.

2.1. Test description

The test comprehended the analysis of 83 content including:

- 16 Documentaries
- 13 Movies with theatrical mixes
- 12 Movies with mixes repurposed for TV presentation
- 2 News programmes
- 3 Sport programmes
- 19 Interstitials including advertisements and promos
- 11 TV series
- 6 TV generic shows
- 1 Classical music opera

The programme languages of the content included: Dutch, English, Finnish, French, German, and Italian. The test consisted of measuring all content with some of the implementations available in the industry and that represents the possible loudness metering methods. The tools and settings that have been used are:

- Dolby Media Meter 2 using the Dialogue Intelligence feature
- Nugen Audio VisLM in Dialog LKFS mode
- Waves Broadcast Loudness Meter in ATSC A85 Dialog mode
- Nugen Audio VisLM in EBU R128 mode

2.2. Test results

The first aim of the research was to compare the measurements of the three voice detection based meters. In order to do that every programme was measured by each meter. For each programme the average of the three meters' measurements was calculated and compared to the single meter's measurement previously gathered. Then some statistics were drafted including the Median, the Standard Deviation, the Minimum and Maximum difference in between each tool results and the voice detection average. Results were arranged according to the category of the content available and that included:

- ALL programmes
- TV NATIVE programmes only
- REPURPOSED FOR TV PRESENTATION programmes only
- TV ONLY, including native and repurposed programmes
- THEATRICAL MIXES only

The following tables show the results being produced

	Dolby Dialogue Intelligence	Waves WLM	Nugen Audio VisLM
MEDIAN	-0.1	0.2	-0.1
STANDARD DEVIATION	0.4	0.8	0.4
MINIMUM	-1.8	-2.6	-2.1
MAXIMUM	1.4	3.2	1.1

Table 1 ALL programmes. Voice detection meters statistics in regard to voice detection meters average

	Dolby Dialogue Intelligence	Waves WLM	Nugen Audio VisLM
MEDIAN	0.0	0.1	-0.1
STANDARD DEVIATION	0.4	0.6	0.3
MINIMUM	-1.7	-2.6	-1.1
MAXIMUM	1.4	2.7	1.1

Table 2 TV NATIVE programmes only. Voice detection meters statistics in regard to voice detection meters average

	Dolby Dialogue Intelligence	Waves WLM	Nugen Audio VisLM
MEDIAN	-0.4	0.7	-0.4
STANDARD DEVIATION	0.4	0.8	0.4
MINIMUM	-1.8	0.3	-1.3
MAXIMUM	-0.2	3.2	0.0

Table 3 REPURPOSED FOR TV PRESENTATION programmes only. Voice detection meters statistics in regard to voice detection meters average

	Dolby Dialogue Intelligence	Waves WLM	Nugen Audio VisLM
MEDIAN	-0.1	0.2	-0.1
STANDARD DEVIATION	0.4	0.7	0.4
MINIMUM	-1.8	-2.6	-1.3
MAXIMUM	1.4	3.2	1.1

Table 4 TV ONLY, including native and repurposed programmes. Voice detection meters statistics in regard to voice detection meters average

	Dolby Dialogue Intelligence	Waves WLM	Nugen Audio VisLM
MEDIAN	-0.1	0.0	0.0
STANDARD DEVIATION	0.4	1.1	0.7
MINIMUM	-0.9	-1.0	-2.1
MAXIMUM	0.6	3.1	0.6

Table 5 THEATRICAL MIXES only. Voice detection meters statistics in regard to voice detection meters average

Furthermore, the research offered the opportunity to compare the two typical loudness measurement implementations available: voice detection vs gating. The following table shows the statistics that compare the average value of the three voice detection based meters and the gating algorithm as described in ITU-R BS1770-3.

	ALL PROGRAMME S	TV NATIVE only	REPURPOSED TV-MIX only	TV ONLY	THEATRICAL MIX only
MEDIAN	-0.8	-0.3	-4.9	-0.3	-9.0
STANDARD DEVIATION	3.4	1.2	1.5	2.0	3.6
MINIMUM	-13.6	-4.7	-6.9	-6.9	-13.6
MAXIMUM	1.4	1.4	-1.6	1.4	-1.6

Table 6 Statistics of comparison between voice detection measurement average and gating measurement

3. CONCLUSIONS

The objective results provided by this research highlight that all loudness meters that implement a voice detection feature provide values that fall within a tolerable range of robustness. This condition appears to be true for all content categories, regardless the environment or destination the programme was produced for. Furthermore, the analysis highlights that voice detection and gating might significantly differ and that it highly depends on the origin of the programme being measured. Whilst that difference is still acceptable in case of programmes natively produced for being broadcast, this difference is no longer tolerable when comparing measurements of mixes being originally crafted for theatrical presentations. In that case the test showed that the minimum difference was of 4.7 LU and up to 13.6LU, with median values of respectively -4.9LU and -9.0LU.

This leads to conclude that whilst the gating feature is ideal for assessing natively produced TV programmes it fails in determining the loudness level of the “anchor element” of programmes originally produced for theatrical presentations, in particular if they have not been repurposed for TV broadcasting.

The author concludes that it would then be highly advisable to merge the current two possible loudness implementations (voice detection and gating) into one unique standard. This way loudness measurement robustness and efficiency would benefit from both approaches and the ability of the algorithm to correlate with the real loudness level perceived by human hearing would probably improve.

4. REFERENCES

- [1] ITU-R BS1770 “Algorithms to measure audio programme loudness and true-peak audio level” (2006)
- [2] ITU-R 1770-3 “Algorithms to measure audio programme loudness and true-peak audio level” (2012)
- [3] EBU Technical Recommendation R128 “Loudness normalisation and permitted maximum level of audio signals” (2011)