Audio Engineering Society

Convention Paper 8613

Presented at the 132nd Convention
2012 April 26–29 Budapest, Hungary

Defining the Listening Comfort Zone in Broadcasting through the analysis of the Maximum Loudness Levels

Alessandro Travaglini[1], Andrea Alemanno[2], and Fabrizio Lantini[3]

[1] Fox International Channels, Rome (RM), I-00100, Italy
info@alessandrotravaglini.it

[2] Department of Information, Electronic and Telecommunication (DIET) - University of Rome "La Sapienza", I-00184 Rome – Italy
andrea.alemanno@live.it

[3] Electric Light Studio, Rome (RM), I-00100, Italy
f.lantini@libero.it

ABSTRACT

Over the last few years, the broadcasting industry has finally approached the loudness issue by standardizing its measurement and recommending target loudness levels with which all programs are required to comply. If the recommendations are applied and all programs are normalized at the target level, viewers ought to experience consistent perceived loudness levels throughout transmissions. However, due to the inner loudness modulation of the programs themselves, this is not always the case. In fact, even if the overall program loudness levels perfectly match the required target level, excessive loudness modulations can still generate annoyance to viewers if the foreground sounds levels exceed the so-called "comfort zone". The fact is that we still have no clear data on which metering can provide visual/numeric feedback on the perception of "hearing annoyance." This paper investigates this issue and aims to provide objective evidence of which parameters would better represent this phenomenon. In particular, we describe an extensive subjective test performed for both the typical Stereo TV and the 5.1 home-theatre set reproductions and analyze its results in order to verify whether the Maximum Momentary Loudness Level, the Maximum Short Loudness Level and Loudness Range (LRA) values described in EBU R128 can provide robust and reliable numeric references to generate a comfortable listening experience for viewers. Furthermore, we perform a similar analysis for the loudness descriptors of the algorithm HELM and finally indicate the values of those parameters that show the most consistent and reliable figures.

1. INTRODUCTION

Human hearing is a very complex phenomenon that is not easy to reproduce artificially. Nevertheless, there are several algorithms that successfully assess the overall loudness level of broadcast content. Over the past few years, the model described in ITU-R.BS1770-2 [2] has become an international reference for organizations, professionals and manufacturers operating in the broadcasting industry. Its implementation aims to provide a very comfortable listening experience to TV viewers and radio listeners. As mentioned in several technical documents based on it (such as ITU-R.BS1864 [5], ATSC-A85 [4] and EBU R-128 [3]), normalizing all Program Loudness Levels to one specific target level allows viewers to comfortably perceive consistent volumes throughout the programing, from one content to another and regardless of genre, style or format.

However, the overall issue may be more complex than that and it might require the definition of other technical descriptors that could eventually support broadcast engineers in assessing the whole loudness characteristics of programs. In particular, there is a need to measure how foreground sounds are presented in the mix to ensure that they always stay within the comfort listening zone and that they never exceed the sound pressure tolerance possibly experienced in both the typical Stereo-TV or Home Theater 5.1 Surround Set.

The objective of this research is to verify how the following existing parameters correlate with the perception of hearing annoyance: Momentary Loudness, Short-Term Loudness, Positive Interval Loudness Level (PILL) and Loudness Range (LRA). The study also analyzes other parameters in order to provide objective evidence as to which parameter might better serve the aim of preventing annoyance. If such a parameter were agreed on, it would be a useful addition to current technical recommendations on loudness.

Our experiment involved a large number of pieces of content, which were prepared in several different versions. These tracks – mostly feature film soundtracks originally mixed for cinema presentation – were remastered in two increasingly more compressed versions. Then, we carried out a formal subjective test whereby 60 people (for a total of 70 assessments as described later) were asked to choose which version of each program they found more comfortable, both for typical home-theater 5.1 listening conditions and typical TV stereo conditions.

The results were gathered and analyzed for each parameter, as described in the Test Description paragraph section below. The whole measurement analysis was then repeated with the published algorithm HELM (High Efficiency Loudness Model) presented at the 132nd AES Convention in Budapest in 2012 and described in the corresponding AES Paper [6] in order to discover whether this method could provide a more robust loudness measurement implementation, better equipped to describe the comfort zone.

2. EBU R128

In 2010, EBU released a recommendation addressed at defining the loudness measurement methods of broadcast content. Their mathematical model is based on the algorithm described in ITU-R.BS1770-1 [1], (recently updated to ITU-R.BS1770-2 [2]). The block diagrams of the two algorithm are shown in Figure 1 and Figure 2.

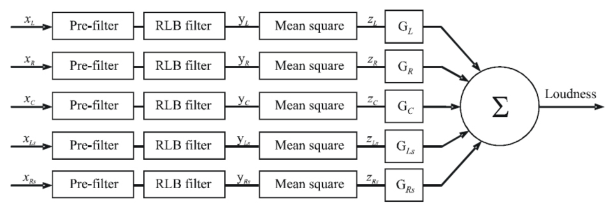

Figure 1 “Block Diagram of BS1770”

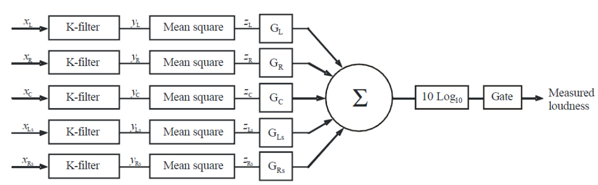

Figure 2 “Block Diagram of BS1770-2”

R128 also defines the implementation of a gating feature, set in the 2011 revision as a relative gating with a value of –10 LU, which is necessary to discard the levels of background sounds from the Program Loudness computation. As specified in EBU Tech3341 [7], the R128 recommendation describes two more meters and a descriptor meant to indicate the time-relative values and the distribution of loudness within a program. The two meters are labeled Momentary Loudness Meter (M) and Short-Term Loudness Meter (S), while the descriptor is called Loudness Range (LRA). Momentary Loudness uses a sliding rectangular time window lasting 4ms, while Short-term Loudness uses a sliding rectangular time window lasting 3 seconds. Neither measurement is gated.

2.1. LOUDNESS RANGE (LRA)

The LRA descriptor is defined in EBU Tech 3342 [8] (released in 2011) and is calculated based on the statistical distribution of measured loudness. As specified in the document, the distribution range of loudness levels is determined by estimating the difference between a low and a high percentile of the distribution. More precisely, LRA is defined as the difference between the estimates of the 10^{th} and 95^{th} percentiles of the distribution. The input to the algorithm is a vector of loudness levels, computed as per ITU-R.BS1770 [1] using a sliding analysis-window of 3 seconds for integration, overlapped with blocks of 66% (minimum 2 seconds).

LRA employs a cascaded gating method in order to focus the LRA computation on foreground sounds only and to avoid producing incorrect high values of LRA for programs containing relevant portions of background sounds (e.g. atmosphere). The two gating thresholds implemented are absolute – 70LU and relative – 20LU to the ungated integrated measurement.

EBU Tech 3342 [8] states that, thanks to the design of LRA, "a short but very loud event would not affect the Loudness Range of a longer segment" and that "similarly the fade-out at the end of a music track, for example, would not increase Loudness Range noticeably…It is noted that measurement of very short programs, where leading or trailing silence is included, or of programs consisting, for example, of isolated utterances, could result in misleadingly high values of LRA…..The lower percentile of 10%, can, for example, prevent the fade-out of a music track from dominating Loudness Range. The upper percentile of 95% ensures that a single unusually loud sound, such as a gunshot in a movie, cannot by itself be responsible for a large Loudness Range."

These statements will be commented on later in this paper.

3. OBJECTIVES OF THIS RESEARCH

EBU-R128 [3] aims to provide consistent hearing perception of loudness throughout the presentation of audio content in broadcasting. It approaches that goal by assessing program loudness levels using three different meters, described as "EBU Mode" and defined in EBU Tech 3341 [7] as Momentary (M), Short-Term (S) and Integrated (I).

In addition, as explained in EBU Tech 3343 [9], EBU "strongly encourages the use of LRA to determine if dynamic treatment of an audio signal is needed and to match the signal with the requirements of a particular transmission channel or platform". In fact, LRA is meant to indicate the maximum allowed dynamic range of the medium (including reproduction system and environment) where the programming is supposed to be reproduced.

Despite the encouragement to rely on this parameter, the document does not set out appropriate values to refer to when producing or post-producing audio programs for broadcasting. There is just the generic observation that "first experiences at broadcasting stations suggest a maximum LRA value of approximately 20 LU for highly dynamic material, such as action movies or classical music. The majority of programming will never need to fully use such a high LRA value". Furthermore, the report highlights that "EBU R128 does not include a maximum permitted LRA value, but instead strongly encourages the use of the Loudness Range parameter to evaluate the potential need for dynamic range processing according to the different criteria mentioned above". At the time of the writing, there are no other documents indicating which LRA values better suit specific media.

In terms of the use of LRA and the different meters recommended in R128 [3], Tech 3343 [9] also specifies that "to control the dynamics of a commercial in a loudness normalized world where there exists the danger of suddenly too high dynamics (overly loud 'pay-off' after a longer period of low-level signals just above the gate threshold), the parameter Loudness Range (LRA) is not suited, as the calculation is based on the short-term loudness values (3 sec. interval). Therefore, for very short items there are too few data points to derive a meaningful number for LRA….An alternative can be found in using the Maximum Momentary Loudness Level (Max ML) and/or the Maximum Short-term Loudness Level (Max SL). Especially for very short items (< 30 seconds), the Maximum Momentary Loudness Level can be used efficiently to limit loudness peaks. Experience of PLOUD members has led to a value around +8 LU (–15 LUFS) as a possible limit for Max ML". Furthermore, the authors of Tech 3343 [9] very honestly admit that "more evidence and experience is needed before this can be stated more firmly."

Indeed, our research aimed to uncover the relationship between the loudness perception of an audio stream in term of loudness, especially in terms of annoyance, and the values obtained with the loudness meters described in EBU-R128 [3]. We also investigated the correct interpretation of LRA values in order to understand how they can contribute to determining usable figures to provide consistent and comfortable sound presentations.

As recognized in the official documents, there was not enough evidence and no in-depth studies had been performed on this topic when the papers were published. In relation to the statements they made, we can safely posit the following:

- Human hearing does not behave differently according to the genre of the content or its duration for sounds above a few hundred milliseconds long. Therefore, there should be no differences in the maximum loudness levels allowed for short content compared to those allowed for long content. Basically, all programs should fulfill the same technical requirements regardless of their aesthetic or genre classification. If evidence supports a specific value for Max Loudness Level for short content, the same value should apply for programs of any other length.

- As noted in EBU Tech 3342 [8], the design of LRA means it is not able to spot offensive loud parts and therefore it cannot prevent the program from generating listener fatigue and annoyance.

- As stated in EBU Tech 3342 [8], LRA is not robust in the measurement of very short content (below 30 seconds): "It is noted that measurement of very short programmes, where leading or trailing silence is included, or of programmes consisting, for example, of isolated utterances, can result in misleadingly high values of LRA".

The goal of our study was to verify the design of the current loudness meters and descriptor recommended in R128 [3], and to identify robust figures that could reliably support the production or post-production of audio mix destined for broadcast distribution.

4. TEST DESCRIPTION

A subjective test was organized in order to spot which parameters better correlate with the sensation of loudness modulation and, more specifically, which ones better indicate the listening threshold beyond which sound levels are perceived as annoying and offensive.

First of all, we gathered 17 pieces of content, all in 5.1 surround sound at their full original dynamics as utilized for home-theater 5.1 presentation: 12 WLR (Wide Loudness Range) were feature films taken from DVD, whilst the remaining five were short trailers (of approximately 30 seconds each) produced for HDTV channels. We labeled the original one WIDE, then created two more versions of each program by applying two different grades of dynamic mono-band compression, as follows:

- MEDIUM (moderate processing):
  - attack time = 100ms
  - release time = 500ms
  - threshold = – 18dBFS
  - ratio = 5:1
  - knee = 50
- NARROW (heavy processing):
  - attack time = 100ms
  - release time = 500ms
  - threshold = – 25dBFS
  - ratio = 12:1
  - knee = 50

We also derived the corresponding stereo downmix from each multichannel 5.1 version. We ended up with six versions per program, for a total of 102 tracks, corresponding to the following groups:

- WIDE 5.1 (multichannel mix, original wide dynamics, no processing)
- MEDIUM 5.1 (multichannel mix, moderate dynamic processing)
- NARROW 5.1 (multichannel mix, heavy dynamic processing)
- WIDE 2.0 (stereo downmix, original wide dynamics, no processing)
- MEDIUM 2.0 (stereo downmix, moderate dynamic processing)
- NARROW 2.0 (stereo downmix, heavy dynamic processing)

Note that the threshold of the processor used to generate the MEDIUM versions was set just above speech level to make sure it was not engaged during dialogue or any quieter sections of the track. The processor implemented to produce the NARROW versions was more aggressive and the threshold level was set in order to have the compressor working during ordinary speech, but not during background sound.

Subsequently, all 102 tracks were measured using the three loudness meters recommended in EBU-R128 [3]: Momentary, Short-Term and Integrated. We measured the Loudness Range, and more short-term measurements were performed in order to verify the robustness of the 3-second integration time. Aside from the 400ms Momentary time window, we used the following Short-term time windows: 1 sec, 2 sec, 3 sec, 5 sec, 7 sec, 10 sec.

To have a complete picture, all measurements were repeated with a different loudness meter model. In this case, we used the HELM algorithm described in the AES Convention Paper "HELM: High Efficiency Loudness Model for Broadcasting Content" [6]. This loudness meter differs from ITU-R.BS1770-2 [2] implemented in the EBU Mode meters in many aspects, such as the frequency weighting, the gating computation and the channels gain weighting. It also includes an additional short-term measurement, the PILL (Positive Interval Loudness Level), which measures the positive interval between the 10-second integration gated (– 7) average level and the 3-second short-term level. To allow a full comparison between HELM and BS.1770-2, a PILL for the latter was measured. In this case the gating was set at – 10 as BS.1770-2 requires.

For all parameters, we tracked the Max Levels. By the end of the measurement phase, we had a database showing the following values for all 102 versions (calculated both using the BS1770-2 algorithm and using HELM): the Program Loudness Level, 6 Max Short-Term Levels, the Max-Momentary Level, the LRA and the Max-PILL.

Then, we created a short 30-second audio/video excerpt from each version, consisting of a speech part at presentation level (for approximately 20 seconds) followed by a loud part of the program containing the loudest possible part (mainly including music and effects) of that version (approximately 10 seconds long). All excerpts were aligned in order to have all speech parts play at the same perceived level. The same gain change was applied to the loud parts but the relation between them and the speech parts remained unchanged.

Consequently, each program originated six versions, each differing from the other in dynamic range: three in multichannel 5.1 surround sound and three as stereo downmix, as described above. We organized the excerpts in blocks of three excerpts, according to the program they originated from, and we placed them on an audio editor timeline.

The subjective part of the test was split in two: one for the TV-Set Stereo reproduction evaluation and the other for the 5.1 Home-Theater reproduction assessment. A total of 60 subjects took part in the tests for a total of 70 independent assessments. Subjects attended the test either alone or in groups of a maximum of three people. The majority of subjects performed only one type of test, except for 10 people who attended both the TV-Stereo and the 5.1 Home-theater one.

4.1. STEREO TV-SET TEST

48 people participated in the Stereo test. The test involved excerpts of all the downmixed program versions which were prepared in blocks of three and made into one audio track each, as described above.

The audio tracks were put on CDs of 17 tracks. Each track comprised the three excerpts representing the three versions (WIDE, MEDIUM, and NARROW) of each program.

The CDs were given to the test subjects who were asked to play them at home on their own TV sets and to select which one (1, 2 or 3 randomly corresponding to WIDE, MEDIUM, NARROW) was better perceived in terms of the loudness modulation between the speech part and the loud part. In other words, subjects were asked to discard the versions they found annoying because they were either too dynamic or too compressed. Subjects were not aware of the correspondence of the versions of the piece of content.

By replicating the test in different environments and with different reproduction systems, whilst gathering information on the typical preferred loudness modulation, we averaged the behavior of different models of TV set. Therefore the results of the TV-Set test really do represent the typical listening conditions of broadcast programming at home and therefore we consider them particularly important and reliable for their implementation in broadcasting operations.

4.2. HOME-THEATER 5.1 SURROUND SOUND TEST

22 people attended this test. It was performed in three different professional mixing rooms with the sizes of average living rooms (6x5, 4x6 and 7x10 meters). They were all equipped with full range 5.1 professional properly aligned loudspeaker sets. By performing the test in three different sound rooms, we averaged the influence of the acoustics and loudspeaker sets that could have otherwise affected the test.

Similarly to the TV-Set Stereo test, the three excerpts of each piece of content were played back to back in “blind mode,” with a few seconds’ silence between each one. At the end of each sequence of three, the subjects were asked to select which one (1, 2 or 3 randomly corresponding to WIDE, MEDIUM, NARROW) was better perceived in terms of loudness modulation between the speech part and the loud part. That is, they were asked to discard the versions they found annoying because they were either too dynamic or too compressed. Subjects were not aware of the correspondence of the versions of the piece of content.

4.3. SUBJECTS STATISTICS

The test was performed by 60 subjects in total. The gender statistics for each test are as follows:

- STEREO TV-SET: 46% Female and 54% Male
- 5.1 HOME-THEATER: 18% Female and 82% Male

The graphs distribution for age statistics are shown in Figure 3 and Figure 4.

Figure 5 shows the statistics of the TV-sets screen sizes.

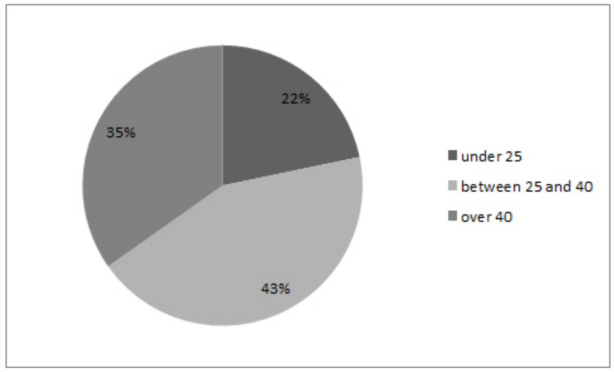

Figure 3 “Subjects’ age statistics of the STEREO-TV set Test”

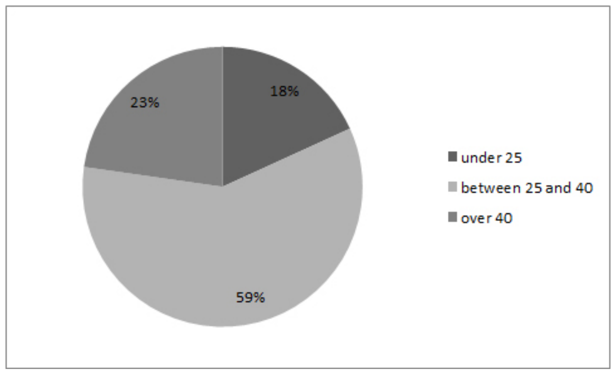

Figure 4 “Subjects’ age statistics of the 5.1 HOME-THEATER set Test”

In addition this is the statistics about the TV used by the testers:

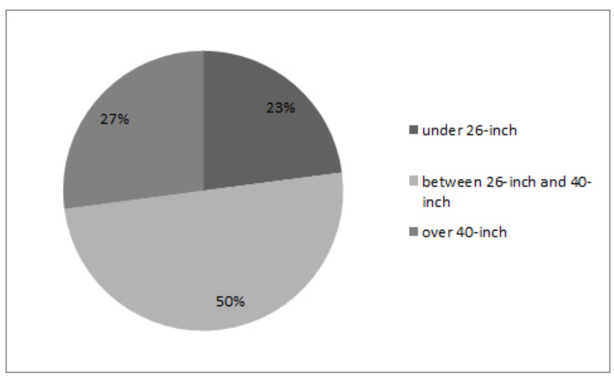

Figure 5 “Screen size’s statistics of the STEREO-TV set Test”

5. TEST RESULTS

For each of the 17 pieces of content, and for each test (5.1 Home Theater and Stereo), we looked for the most preferred version, calculating the weighted average of the subjects' selections. For that weighted average, all loudness parameters were calculated and inserted in the Table 1, Table 2, Table 3, and Table 4 that show the most comfortable version loudness values for all 17 programs. The 12 feature film programs and the 5 program trailers were analyzed in two separate groups (FEATURE FILMS, TRAILERS).

Then, we calculated the standard deviation and the average values for each parameter, and for the two groups of FEATURE FILMS and TRAILERS. The objective of the research was to identify the parameter that best indicated the upper limit within which foreground sounds should be reproduced in order to be perceived comfortably and to not generate annoyance to the listeners. To do this, we focused on analyzing the results of the FEATURE FILMS group.

We were looking for the most robust and reliable parameter of the ones previously measured: LRA, Max Momentary (MML), Max Short (MSL 1s, MSL 2s, MSL 3s, MSL 5s, MSL 7s, MSL 10s), and Max PILL (MPILL). We looked for the parameter with the lowest standard deviation as it best fulfills our requirement.

5.1. STEREO TV-SET RESULTS

The list of all measured parameters is shown in Table 5 where we can see that the results of the FEATURE FILMS group indicate that the parameter with the lowest standard deviation is the Max PILL (MPILL) measured according to HELM, which shows a value of just 0.807. This is a very low value that shows that the Max PILL is the most robust candidate for identifying the upper limit for foreground sounds of stereo programs.

The value to be used as Max PILL-HELM for stereo content is +7LU (a rounding up of +6.853).

In regard to the BS1770-2 parameters, we see that all the Max Short values and the Max Momentary ones have standard deviations figures of around 1. More precisely, the average standard deviation of all ITU-R.BS1770-2 Short-Term measurements is 1.136. The computation of the average deviation of all Short-Term measurements according to HELM gives a very slightly larger value of 1.187.

Of the Max Short values calculated with BS1770-2, the one with a 1-second integration time shows the lowest standard deviation for this algorithm, with an average value of 0.938. This makes it the best candidate for indicating the highest tolerated loudness level for all technical recommendations implementing ITU-R.BS1770-2 [2]. For the 1-second integration time, the Max Short Loudness Level is +8LU (a rounding down of +8.350). For the EBU R128, where the Short Term meter is calculated on a 3-second sliding window, the value that indicates the highest level tolerated for stereo content is +7LU (a rounding up of +6.628).

Furthermore, the results of this analysis show that the Max Momentary value that best correlates with the maximum tolerated loudness level is +9LU (a rounding down of +9.354), which is therefore close yet larger to the early informal findings developed within PLOUD which suggested +8LU.

A breakdown of the versions that subjects preferred shows that NARROW versions were chosen 49.8% of the time, the MEDIUM 36.1% and the WIDE versions only 14.1% of the time.

By contrast, if we look at the selections made for the TRAILER group we see that people tend to prefer the two most dynamic soundtracks: NARROW was chosen 24.2% of the time, MEDIUM 37.5% and WIDE 38.3% of the time. This suggests that audio over-compression (NARROW versions) is perceived as annoying by the average listener, whilst the tracks with medium or no compression are preferred.

If we now analyze the TRAILER group statistics, we see that all maximum values are lower than the ones from the FEATURE FILMS analysis. This could lead one to think that the two different content categories require different maximum levels according to the genre or duration of the program. But we do not believe this to be the case. The different maximum values resulting from the analysis are only due to the fact that, unlike the 12 FEATURE FILMS, all the 5 TRAILER soundtracks consist of a mix of music, sound effects and voice all the way through. In this case, since there is a more consistent level between the "speech" part and the following "loud" part, as the low values of LRA confirm, the integrated Program Loudness values are higher and, consequently, the expected differences of the Momentary, Short and PILL compared to the Program Loudness levels are smaller. But this does not mean that all short content such as trailers, commercials or TV channel promos require low maximum levels nor that they should be limited to smaller loudness ranges compared to other content genres. It actually may indicate that the computation of Maximum Levels is depending on the overall Programme Loudness measurement which is affected by the gating feature and by the matching between "anchor" sounds and Target Level as explained in 5.3 and 6.

5.2. HOME-THEATER 5.1 SURROUND SOUND RESULTS

As with the analysis of Stereo content, in order to identify the parameter that best represents the upper loudness limit for multichannel programs, we compiled tables of results from the preferred FEATURE FILMS versions and looked for the parameter with the lowest standard deviation value. Once again, HELM has the lowest deviation, as shown by the 7-second integration time Short-Term measurement which has a standard deviation of just 0.739. These subjective tests indicate that the Max Short-Term (7 sec) Level measured according to HELM is just above +4LU (a rounding down of +4.171). Moreover, the top 4 most robust parameters are according to HELM and most Short-Term measurements performed with this algorithm provide very low standard deviation values: the average standard deviation of all HELM Short-Term measurements is 0.857. The average standard deviation of all Short-term measurements according to BS.1770-2 is 1.055.

Looking at Table 6 we see that the most robust parameter according to BS1770-2 is the Short-Term 10-second integration measurement, which has a standard deviation of 0.871 corresponding to a max level of +5LU (a rounding up of +4.919).

In terms of EBU R128, the maximum level for the Momentary Meter is +8LU (a rounding down of +8.427) and for the 3-second Short-Term measurement it is +6LU (a rounding down of +6.097).

In the 5.1 Home-Theater test subjects mostly preferred the MEDIUM versions (59.1 % of the time); the NARROW versions were chosen 17.8% of the time and the WIDE versions 23.1% of the time.

If we look at the selections made for the TRAILER group, we see that people tend to prefer the medium dynamic soundtracks as well: NARROW was chosen 22.7% of the time, MEDIUM 46.4% and WIDE 30.9% of the time. This suggests that a medium dynamics compression is perceived as comfortable by the average listener whilst tracks with soft compression are slightly preferred by 1 person out of 3. It also indicates that usually over-compressed content are not perceived comfortably. This statistic, although performed on a relatively small group of people, leads us to observe that the presentation of multichannel content for home entertainment should rely on the implementation of dynamics compression metadata aimed at repurposing the original wide dynamics of cinematic mixes to smaller room acoustics, letting viewers decide whether to apply further dynamic processing or not.

The same considerations presented in section 5.1 of this paper regarding the different values of maximum levels from the TRAILER group analysis are equally valid for the Home-theater 5.1 Surround Sound test: the same maximum values resulting from the FEATURE FILM test should be implemented for all other programs, regardless of their genre or duration.

5.3. LRA AND MAXIMUM LEVELS ANALYSIS

By definition, the LRA should tell the sound mixer or broadcasting operator whether the content's dynamics are adequate for transmission as they stand or whether they require audio dynamic processing. Therefore, the LRA of the preferred versions from our tests ought to have a low standard deviation. But the standard deviation of our LRA values, measured according to [3], is 3.829 for Stereo content and 4.337 for multichannel audio programs. These high figures indicate that the LRA does not seem to directly correlate with the sensation of comfortable listening and that it cannot be used to assess whether a program's dynamics is adequate for broadcast.

In fact, although the average preferred LRA value is 17 for both stereo and 5.1 surround sound (a rounding up of 16.763 and 16.922 respectively) in Table 9 and Table 10 we see that some preferred versions have low LRA values (e.g. *Transformers* 2 *LRA=9* for the Stereo and *LRA=8* for the Multichannel version) while other preferred versions have very high LRA values (*Inglorious Bastards LRA=22* for the Stereo and *The Gladiator LRA=22* for the Multichannel version). This is probably due to the fact that LRA is calculated on a statistical basis and, as described in [8], it discards 5% of the loudest values of the measured program. Indeed, this 5% most likely includes the loud foreground sounds that generate listening annoyance but that the LRA computation ignores.

LRA seems to be more useful in indicating the difference in loudness between the quasi-loudest and the quasi-softest foreground sounds of a program, calculated with a relative gating set at –20LU. As a statistical finding, we can say that amongst all preferred versions in this study, the highest LRA value was 22 for both Stereo programs and Multichannel audio programs. However, this information does not contribute to preventing listening annoyance. Some programs with a lower LRA, in fact, were perceived as clearly annoying (e.g. the stereo versions of *The Dark Knight LRA=20, Transformers 2 LRA=12*, *Wall-E LRA=16*). In another case (*Pirates of the Caribbean 3* stereo versions) we see that two program versions with the same LRA value (18) were clearly judged differently. This is because the difference between the two versions are other loudness parameters like Max Momentary and the Max Short levels which have different values for each version (see Table 9). However, also Maximum Momentary and Maximum Short-term Levels are not always providing reliable figures as explained below.

As shown in Figure 6, LRA and Max Loudness levels are not always directly proportional. The diagram represents the average preferred levels of the programs. It shows there is no relationship between the LRA values and the Max Loudness levels (either Max Momentary or Max Short): in some programs, the difference between the two values is small (*Transformers 2*) whilst in other it is large (*Tron: Legacy,* and *Inglorious bastards*), and again, any other possibility seems equally valid. This finding is also shown in Table 1, 2 3 and 4 where all values of LRA and Max Levels are indicated numerically, for both the HELM and the BS.1770-2 algorithms. However, the lack of linearity between LRA and Max Loudness Levels seems to be due to the low robustness of the latter, as you can see in Table 9 where in some cases (*The dark knight, Tron: Legacy, Star Trek*) there is no correlation between the maximum levels and the program's dynamics (see Table 9). This could be also due to other factors such as the frequency weighting or the gating, and might require further investigation. Furthermore, the fact that the LRA and the Program Loudness Level used as a reference to compute the Maximum Levels implement two different relative gating thresholds, at -20 and -10 respectively, could also explain why the two parameters are not always providing the same interpretation of the program's dynamics since the program's parts that would be discarded by the computation are not the same. Further investigation seems to be necessary in order to verify if a better correlation would be obtained by applying the same gating method to both loudness descriptors.

Finally, although the Loudness Range is inversely proportional to the ranking of the preferences (the most preferred the lower the LRA), and is indeed directly proportional to the dynamics of the program, no LRA values can be used reliably to represent the right range within which foreground sounds are perceived comfortably as there appears to be no right or wrong LRA values.

6. CONCLUSIONS

The test shows that the parameters that best indicate the upper limit for foreground sounds to be broadcast in order to avoid listening annoyance are as indicated in Table 5 and Table 6 for stereo and 5.1 surround sounds respectively (robustness is shown in decreasing order, starting with the best).

According to this study, it is not recommended to differentiating maximum values according to program genre or duration.

Moreover, we have observed that, differently to what one might expect, the maximum levels for multichannel content are very similar to the ones for stereo content. Since the figures from all tests show a very slight, if any, difference between stereo and 5.1 values, we believe that, to simplify the implementation of these parameters, cross-format values could be safely applied to any type of audio format. They are listed in order of robustness, with the most robust parameter cited first, in Table 7.

In regard to R128 [3], this study indicates that the Maximum Momentary Level should be +9LU and the Max Short-term Level should be +6LU for all content. One more LU of tolerance could be permitted in order to allow a wider loudness modulation. The test also shows that since LRA is a computation based on statistical analysis it seems to hardly provide an exact representation of the whole foreground loudness levels. However, although it does not always seem to be able to prevent hearing annoyance, the recommended LRA value should be 17. Furthermore, in some cases the Max Loudness Levels of the three program's versions used for this test have values of Max Momentary and Max Short levels that are inversely proportional to their dynamics. This could be possibly due to the loudness distribution of the program that, because of the gating, in cases of very wide loudness range programs produces that effect and lowers the correlation between the Program Loudness and the "anchor" sound level . This leads to think that the best benefit of using the Maximum Loudness descriptors is when applying real-time metering during content production. In this way, in fact, the sound mixer can verify that the "anchor" sounds are properly aligned around the Target Level and can use the foreground sounds loudness measurement in regard to that.

We also found that the Max Positive Interval Loudness Level (Max PILL) as described in [6] seems to be able to measure the loudness of stereo content in a robust way as it shows the lowest standard deviation value for that audio format. In addition, the test shows that when assessing multichannel content, the algorithm HELM seems to provide a more correlated measurement compared with the loudness model described in [2] as it scores the four lowest standard deviation values. Other parameters, with their corresponding values, offer slightly lower degrees of robustness. Moreover, since the longest Short-term parameters (7 and 10 second integration time) might result less effective in real-time audio production, if they are removed from the analysis and only the Short-term measurements of up to 5 second integration are considered it results that the HELM algorithm still offers the top four most robust parameters adequate for measuring the upper limits for loudness modulation in broadcasting, averaging both stereo and multichannel audio formats.

The low standard deviation of HELM's values is also due to the recursive gating feature implemented in this algorithm. Because of its own design, in fact, HELM's recursive gating tends to normalize program's loud parts close to the Target Level and consequently prevents foreground sounds to be reproduced annoyingly loud. The better performance of HELM is confirmed by the fact that its recursive gating at -7 provides a better matching between the overall program level and the "anchor" level, as described in the AES Paper "HELM: High Efficiency Loudness Model for Broadcast Content" (132nd AES Convention, 2012 Travaglini et al.) [6]. This also implies that, in case HELM is used to assessing program loudness levels, sound mixers and broadcast engineers would rely on a more robust automatic "anchor" measurement and would possibly be encouraged to apply the adequate overall dynamic processing in order to maintain the proper relation between foreground sounds, background sounds and in regard to the whole program level.

For both Stereo and Multichannel audio programs in this test the least robust parameter is LRA, which has very high standard deviation values and therefore does not appear to be reliable for determining if the dynamics of a program is appropriate or not.

In conclusion we propose adding some of the parameters shown in Table 7, and possibly according to the ranking of robustness, to the international recommendations currently used worldwide in order to offer valuable support to mixing engineers.

Although this test has proved these parameters are not fully reliable, the authors still believe that by defining these values broadcasting would benefit from a more robust technical reference, which in turn would contribute to providing a more reliable and standardized approach to content creation and reprocessing, giving the professional community verified descriptors to be implemented in the technical workflows. Especially when implementing loudness models capable of competently assessing the "anchor" sounds levels, by measuring these parameters, in fact, the sound engineer will be supported in checking that the foreground sounds of the program will not exceed on average the maximum tolerated perceived level and therefore the loudness modulation will not annoy the audience.

Consequently, program exchange would be simplified, the original creative intent of content would be preserved, and both the content creators and content users' expectations would be hopefully met.

7. ACKNOWLEDGEMENTS

We would like to thank Zoë Hallington for her assistance in editing this paper. We would also like to thank Fox International Channels Italy for supporting this research and Stuart Mabey at Electric Light Studio and Alberto Colajacomo and Simone Corelli (GTCS) at Post In Europe Studio in Rome for providing the technical facilities necessary to perform the subjective tests. Finally, we thank everyone who contributed to this research by taking part in the tests.

8. REFERENCES

[1] ITU-R.BS1770-1 "Algorithms to measure audio programme loudness and true-peak audio level" (2007)

[2] ITU-R.BS1770-2 "Algorithms to measure audio programme loudness and true-peak audio level" (2011)

[3] EBU Technical Recommendation R 128 "Loudness normalisation and permitted maximum level of audio signals" (2010)

[4] ATSC Recommended Practice – Document A/85 . "Techniques for Establishing and Maintaining Audio Loudness for Digital Television" (2011)

[5] ITU-RBS1864 "Operational practices for loudness in the international exchange of digital television programmes" (2010)

[6] AES Convention Paper – Alessandro Travaglini, Andrea Alemanno, Aurelio Uncini – "HELM: High Efficiency Loudness Model for Broadcast Content" (2012)

[7] EBU Tech Doc 3341 'Loudness Metering: 'EBU Mode' metering to supplement loudness normalization in accordance with EBU R 128' (2010)

[8] EBU Tech Doc 3342 'Loudness Range: A descriptor to supplement loudness normalization in accordance with EBU R 128' (2010)

[9] EBU Tech Doc 3343 'Practical guidelines for Production and Implementation in accordance with EBU R 128' (2011)

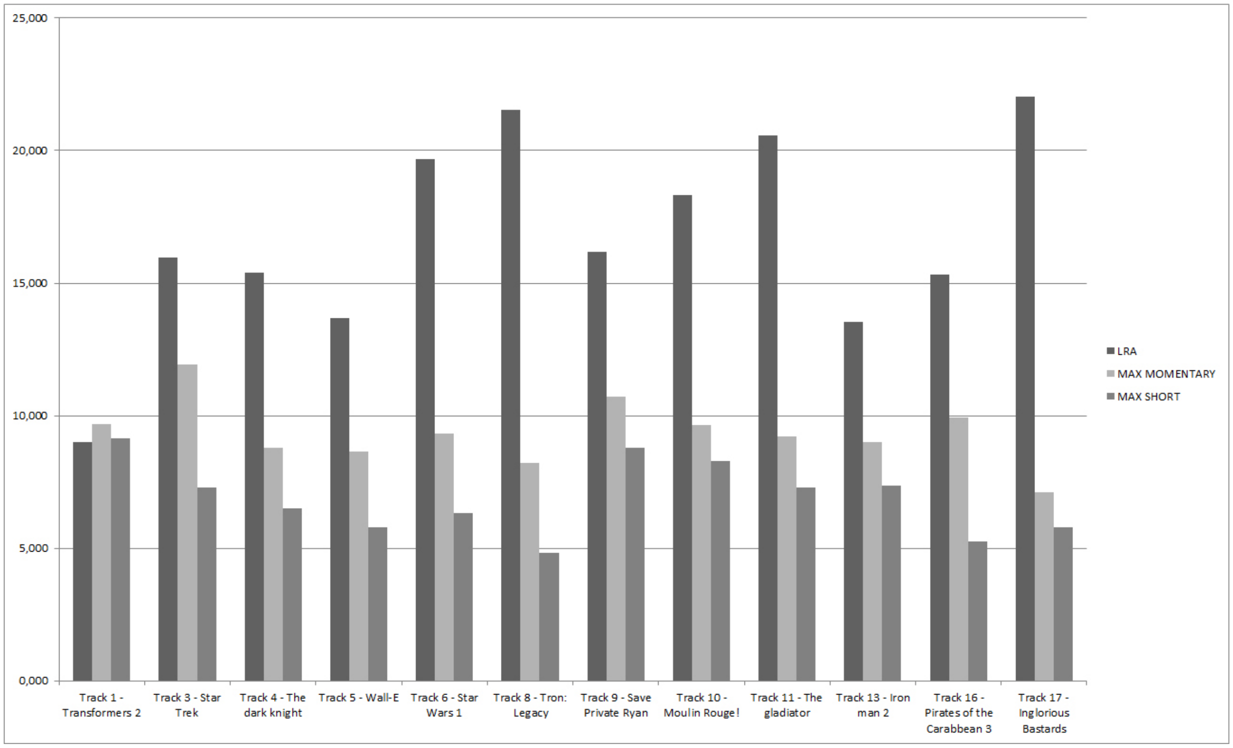

Figure 6 “LRA, MaxMomentary, MaxShort preferred versions values for ITU-R.BS.1770-2”

	PROGRAM	PL	LRA	MPILL	MML	MSL 1s	MSL 2s	MSL 3s	MSL 5s	MSL 7s	MSL 10s
FEATURE FILMS	**Transformers 2**	-31.476	8.825	6.641	10.171	7.765	6.619	9.008	5.430	4.989	4.527
	Star Trek	-28.246	17.502	8.213	9.488	8.303	6.848	5.967	4.875	4.181	3.779
	The dark knight	-26.120	13.187	6.246	8.486	6.886	5.624	5.109	4.580	4.290	3.539
	Wall-E	-29.257	12.858	8.276	8.745	7.778	6.815	5.973	4.929	4.274	4.019
	Star Wars 1	-26.009	18.033	6.164	5.869	4.872	4.075	3.326	2.817	2.403	2.135
	Tron: Legacy	-28.869	19.286	6.389	6.874	5.713	3.862	3.666	2.743	2.330	2.288
	Save Private Ryan	-27.819	15.487	7.284	8.914	7.797	6.280	5.999	5.705	4.801	4.127
	Moulin Rouge!	-30.394	18.740	7.441	6.559	6.198	5.924	5.633	5.228	5.096	4.889
	The gladiator	-26.076	20.248	6.401	6.673	6.452	6.027	5.500	4.931	4.461	4.274
	Iron man 2	-33.795	13.805	6.702	8.939	8.033	7.327	7.011	6.401	6.364	5.970
	Pirates of the Car. 3	-30.257	16.437	6.826	5.505	4.737	4.401	3.923	3.830	3.484	3.422
	Inglorious Bastards	-28.723	20.111	5.655	5.266	5.014	4.683	3.854	3.343	3.047	2.367
	Standard deviation	Not measured	3.486	0.807	1.681	1.319	1.179	1.164	1.153	1.171	1.133
	Average		16.210	6.853	7.624	6.629	5.707	5.139	4.568	4.143	3.778
TRAILERS	**Falling Skies**	-25.594	3.843	2.589	3.891	2.838	2.603	2.338	2.258	1.863	1.349
	AHS	-30.133	7.090	5.995	7.266	5.346	4.795	3.795	3.045	2.116	1.501
	Great migrations	-30.426	7.321	5.084	5.214	4.782	4.305	3.797	3.059	2.623	2.022
	Medici	-30.282	7.444	5.813	6.711	6.248	5.259	4.745	4.185	3.500	2.368
	Terra nova	-27.259	7.347	4.970	7.551	6.509	5.327	4.761	3.854	3.268	3.219
	Standard deviation	Not measured	1.552	1.361	1.541	1.464	1.115	0.990	0.758	0.708	0.750
	Average		6.609	4.890	6.127	5.145	4.458	3.887	3.280	2.674	2.092

Table 1 “Stereo preferred weighted levels measured according to HELM”

	PROGRAM	PL	LRA	MPILL	MML	MSL 1s	MSL 2s	MSL 3s	MSL 5s	MSL 7s	MSL 10s
FEATURE FILMS	Transformers 2	-30.743	8.991	9.213	9.687	8.367	7.037	9.142	6.996	6.405	5.675
	Star Trek	-29.052	15.954	9.687	11.918	10.021	8.318	7.305	6.869	6.393	5.817
	The dark knight	-24.570	15.378	5.638	8.781	8.038	6.999	6.498	6.130	5.677	5.237
	Wall-E	-30.077	13.671	9.395	8.657	7.614	6.543	5.795	5.076	4.798	4.574
	Star Wars 1	-26.140	19.692	8.176	9.327	8.761	7.357	6.338	5.924	5.853	5.793
	Tron: Legacy	-26.043	21.538	10.953	8.224	7.420	5.415	4.815	4.449	3.738	3.209
	Save Private Ryan	-29.179	16.182	11.606	10.707	9.671	9.209	8.780	8.153	7.214	6.052
	Moulin Rouge!	-32.413	18.318	9.259	9.660	9.110	8.687	8.286	7.672	7.372	7.074
	The gladiator	-27.625	20.572	7.248	9.225	8.451	7.866	7.304	6.674	6.242	6.032
	Iron man 2	-34.589	13.534	7.876	9.017	8.270	7.621	7.361	6.909	6.778	6.528
	Pirates of the Car. 3	-29.580	15.313	7.326	9.915	7.686	5.637	5.242	4.366	4.167	4.060
	Inglorious Bastards	-29.144	22.016	9.366	7.128	6.791	6.458	5.804	5.100	4.867	4.410
	Standard deviation	Not measured	3.829	1.656	1.210	0.938	1.159	1.202	1.235	1.173	1.111
	Average		16.763	8.812	9.354	8.350	7.262	6.628	6.193	5.792	5.372
TRAILERS	Falling Skies	-25.360	3.591	1.486	3.506	2.973	2.693	2.356	2.052	1.808	1.453
	AHS	-28.470	8.050	6.339	6.362	5.485	4.196	3.439	2.802	2.045	1.158
	Great migrations	-30.066	5.814	3.895	4.387	3.411	3.220	2.940	2.693	2.385	1.819
	Medici	-30.310	6.738	4.844	5.978	5.428	4.960	4.576	3.980	3.315	2.382
	Terra nova	-27.367	6.457	5.026	6.526	5.637	4.950	4.540	3.842	3.248	3.106
	Standard deviation	Not measured	1.636	1.807	1.335	1.285	1.022	0.980	0.818	0.690	0.776
	Average		6.130	4.318	5.352	4.587	4.004	3.570	3.074	2.560	1.984

Table 2 "Stereo preferred weighted levels measured according to ITU-R-BS1770-2"

	PROGRAM	PL	LRA	MPILL	MML	MSL 1s	MSL 2s	MSL 3s	MSL 5s	MSL 7s	MSL 10s
FEATURE FILMS	Transformers 2	-33.148	8.071	6.248	9.329	7.330	6.093	5.176	4.764	4.238	4.003
	Star Trek	-27.704	20.346	9.442	8.576	7.327	6.148	5.157	4.187	3.611	3.155
	The dark knight	-25.518	15.096	6.482	7.989	6.343	5.230	4.487	4.028	3.835	2.940
	Wall-E	-30.482	13.347	7.600	8.782	7.215	6.060	5.193	3.972	3.341	3.164
	Star Wars 1	-25.979	19.970	6.359	8.059	6.931	6.321	5.251	4.048	3.324	2.824
	Tron: Legacy	-30.375	22.059	9.212	6.734	5.071	4.278	4.129	4.256	4.154	4.198
	Save Private Ryan	-28.766	17.710	7.838	10.060	8.774	7.168	6.433	5.873	5.166	4.241
	Moulin Rouge!	-30.148	21.115	9.804	6.933	6.471	6.145	5.771	5.365	5.205	4.884
	The gladiator	-25.260	23.303	6.409	7.063	6.813	6.320	5.771	5.115	4.605	4.386
	Iron man 2	-34.729	13.601	7.752	8.449	7.131	6.387	5.907	5.283	5.302	4.891
	Pirates of the Car. 3	-31.090	19.185	7.228	5.754	4.994	4.857	4.358	6.650	3.832	3.750
	Inglorious Bastards	-30.200	21.573	6.258	5.387	5.093	4.792	4.080	3.636	3.441	2.775
	Standard deviation	Not measured	4.532	1.310	1.413	1.123	0.836	0.754	0.909	0.739	0.778
	Average		17.948	7.553	7.760	6.624	5.817	5.143	4.765	4.171	3.768
TRAILERS	Falling Skies	-25.929	4.070	2.724	3.584	2.818	2.596	2.287	2.224	1.820	1.419
	AHS	-31.399	7.159	6.450	7.656	5.788	5.209	3.895	3.148	2.209	1.368
	Great migrations	-30.519	6.778	5.713	5.236	4.723	4.252	3.699	2.923	2.478	1.857
	Medici	-31.310	7.040	5.242	5.745	5.346	4.685	4.488	4.053	3.468	2.383
	Terra nova	-28.759	7.237	5.375	7.187	6.223	5.134	4.684	3.722	3.094	3.187
	Standard deviation	Not measured	1.345	1.409	1.625	1.330	1.066	0.944	0.713	0.666	0.759
	Average		6.457	5.101	5.882	4.979	4.375	3.811	3.214	2.614	2.043

Table 3 “MCA 5.1 SURROUND SOUND preferred weighted levels measured according to HELM”

	PROGRAM	PL	LRA	MPILL	MML	MSL 1s	MSL 2s	MSL 3s	MSL 5s	MSL 7s	MSL 10s
FEATURE FILMS	Transformers 2	-30.399	7.820	6.649	7.734	6.508	5.721	5.157	4.905	4.629	4.097
	Star Trek	-27.750	16.442	10.728	9.403	8.860	7.907	6.965	6.420	5.904	5.405
	The dark knight	-23.963	13.856	6.102	8.434	7.070	6.156	5.680	5.168	4.903	4.126
	Wall-E	-28.771	13.928	9.803	8.226	6.625	5.528	4.794	4.049	3.602	3.616
	Star Wars 1	-26.245	18.594	5.021	8.826	8.357	7.371	6.372	5.438	4.996	4.582
	Tron: Legacy	-29.366	22.444	13.740	8.109	6.668	5.661	5.497	5.701	5.599	5.630
	Save Private Ryan	-27.576	16.196	8.771	9.614	8.965	8.144	7.645	7.059	6.140	5.207
	Moulin Rouge!	-29.985	20.544	10.041	9.496	8.704	7.976	7.471	7.114	6.931	6.544
	The gladiator	-26.319	21.402	6.969	9.423	9.030	9.781	7.748	6.905	6.440	6.089
	Iron man 2	-33.081	12.432	9.439	8.260	6.550	5.759	5.428	5.042	4.871	4.552
	Pirates of the Car. 3	-29.835	18.522	7.427	6.939	5.695	5.199	4.947	4.743	4.656	4.549
	Inglorious Bastards	-29.945	20.887	9.082	6.657	6.165	5.988	5.460	5.133	5.077	4.573
	Standard deviation	Not measured	4.337	2.380	0.984	1.245	1.195	1.092	1.007	0.922	0.871
	Average		16.922	8.648	8.427	7.433	6.648	6.097	5.640	5.312	4.914
TRAILERS	Falling Skies	-26.033	3.512	1.800	3.083	2.713	2.486	2.266	2.105	1.811	1.293
	AHS	-29.096	7.269	5.617	6.620	5.647	4.106	3.307	2.630	1.915	1.174
	Great migrations	-29.125	4.259	3.027	3.461	2.807	2.674	2.378	2.064	1.799	1.287
	Medici	-29.542	5.273	3.749	4.986	4.551	4.093	3.831	3.316	2.726	1.941
	Terra nova	-26.503	4.873	3.671	5.167	4.345	3.824	3.507	2.941	2.498	2.326
	Standard deviation	Not measured	1.414	1.384	1.426	1.247	0.793	0.698	0.539	0.432	0.504
	Average		5.037	3.573	4.663	4.013	3.436	3.058	2.611	2.150	1.604

Table 4 “MCA 5.1 SURROUND SOUND preferred weighted levels measured according to ITU-R-BS1770-2”

STEREO STD. DEV.	PARAMETER	TIME WINDOW	ALGORITHM	STEREO VALUE
0.807	Max PILL	3 sec	HELM	+6.9
0.938	Max Short Loudness Level	1 sec	ITU-R.BS1770-2	+8.4
1.111	Max Short Loudness Level	10 sec	ITU-R.BS1770-2	+5.4
1.133	Max Short Loudness Level	10 sec	HELM	+3.8
1.153	Max Short Loudness Level	5 sec	HELM	+4.6
1.159	Max Short Loudness Level	2 sec	ITU-R.BS1770-2	+7.3
1.164	Max Short Loudness Level	3 sec	HELM	+5.1
1.171	Max Short Loudness Level	7 sec	HELM	+4.1
1.173	Max Short Loudness Level	7 sec	ITU-R.BS1770-2	+5.8
1.179	Max Short Loudness Level	2 sec	HELM	+5.7
1.202	Max Short Loudness Level	3 sec	EBU-R128	+6.6
1.210	Max Momentary Loudness Level	400ms	EBU-R128	+9.4
1.235	Max Short Loudness Level	5 sec	ITU-R.BS1770-2	+6.2
1.319	Max Short Loudness Level	1 sec	HELM	+6.6
1.656	Max PILL	3 sec	ITU-R.BS1770-2	+8.8
1.681	Max Momentary Loudness Level	400ms	HELM	+7.6
3.829	LRA	program	EBU-R128	16.8

Table 5 “STEREO test results”

MCA 5.1 STD. DEV.	PARAMETER	TIME WINDOW	ALGORITHM	MCA 5.1 VALUE
0.739	Max Short Loudness Level	7 sec	HELM	+4.2
0.754	Max Short Loudness Level	3 sec	HELM	+5.1
0.778	Max Short Loudness Level	10 sec	HELM	+3.8
0.836	Max Short Loudness Level	2 sec	HELM	+5.8
0.871	Max Short Loudness Level	10 sec	ITU-R.BS1770-2	+4.9
0.909	Max Short Loudness Level	5 sec	HELM	+4.8
0.922	Max Short Loudness Level	7 sec	ITU-R.BS1770-2	+5.3
0.984	Max Momentary Loudness Level	400ms	EBU-R128	+8.4
1.007	Max Short Loudness Level	5 sec	ITU-R.BS1770-2	+5.6
1.092	Max Short Loudness Level	3 sec	EBU-R128	+6.1
1.123	Max Short Loudness Level	1 sec	HELM	+6.6
1.195	Max Short Loudness Level	2 sec	ITU-R.BS1770-2	+6.6
1.245	Max Short Loudness Level	1 sec	ITU-R.BS1770-2	+7.4
1.310	Max PILL	3 sec	HELM	+7.6
1.413	Max Momentary Loudness Level	400ms	HELM	+7.8
2.380	Max PILL	3 sec	ITU-R.BS1770-2	+8.7
4.337	LRA	program	EBU-R128	16.9

Table 6 “Home-Theater 5.1 Surround Sound test results”

STEREO STD. DEV.	MCA 5.1 STD.DEV.	AVERAGE STD. DEV.	PARAMETER	TIME WINDOW	ALGORITHM	STEREO VALUE	MCA 5.1 VALUE	CROSS-FORMAT ROUNDED AVERAGE
1.171	0.739	0.955	Max Short Level	7 sec	HELM	+4.1	+4.2	**+4**
1.133	0.778	0.955	Max Short Level	10 sec	HELM	+3.8	+3.8	**+4**
1.164	0.754	0.959	Max Short Level	3 sec	HELM	+5.1	+5.1	**+5**
1.111	0.871	0.991	Max Short Level	10 sec	BS1770-2	+5.4	+4.9	**+5**
1.179	0.836	1.007	Max Short Level	2 sec	HELM	+5.7	+5.8	**+6**
1.153	0.909	1.031	Max Short Level	5 sec	HELM	+4.6	+4.8	**+5**
1.173	0.922	1.047	Max Short Level	7 sec	BS1770-2	+5.8	+5.3	**+6**
0.807	1.310	1.058	Max PILL	3 sec	HELM	+6.9	+7.6	**+7**
0.938	1.245	1.091	Max Short Level	1 sec	BS1770-2	+8.4	+7.4	**+8**
1.210	0.984	1.097	Max Momentary Level	400ms	EBU-R128	+9.4	+8.4	**+9**
1.235	1.007	1.121	Max Short Level	5 sec	BS1770-2	+6.2	+5.6	**+6**
1.202	1.092	1.147	Max Short Level	3 sec	EBU-R128	+6.6	+6.1	**+6**
1.159	1.195	1.177	Max Short Level	2 sec	BS1770-2	+7.3	+6.6	**+7**
1.319	1.123	1.221	Max Short Level	1 sec	HELM	+6.6	+6.6	**+7**
1.681	1.413	1.547	Max Momentary Level	400ms	HELM	+7.6	+7.8	**+8**
1.656	2.380	2.030	Max PILL	3 sec	BS1770-2	+8.7	+8.8	**+9**
3.829	4.337	4.083	LRA	program	EBU-R128	16.8	16.9	**17**

Table 7 "CROSS-FORMAT test results"

	STEREO			HOME THEATER 5.1 SURROUND SOUND		
PROGRAM	**NARROW**	**MEDIUM**	**WIDE**	**NARROW**	**MEDIUM**	**WIDE**
The dark knight	25	20	3	7	10	5
Inglorious Bastards	36	8	4	4	16	2
Iron man 2	18	18	12	1	14	7
Pirates of the Car. 3	26	17	5	1	14	7
Save Private Ryan	16	23	9	3	9	10
Star Trek	23	12	13	5	10	7
Star wars 1	24	20	4	3	14	5
Transformers 2	27	13	8	6	11	5
Tron: Legacy	24	19	5	4	14	4
Wall-E	15	25	8	2	15	5
The Gladiator	21	21	6	3	17	2
Moulin Rouge!	32	12	4	8	12	2
AHS	12	20	16	1	11	10
Falling skies	12	27	9	6	13	3
Great migrations	11	12	25	4	10	8
Medici	9	22	17	4	10	8
Terra nova	14	9	25	10	7	5
TOTAL	**345**	**298**	**173**	**72**	**207**	**95**
	42.3%	**36.5%**	**21.2%**	**19.3%**	**55.3%**	**25.4%**

Table 8 “Total number of preferences selected by subjects for both STEREO and Home Theater 5.1 Surround Sound Tests”

PROGRAM	VERSION	PREFERENCES	PL	LRA	MML	MSL 3s
The dark knight	NARROW	25	-26.449	13.029	9.428	6.782
	MEDIUM	20	-22.957	17.557	8.364	6.476
	WIDE	3	-19.663	20.427	6.163	4.273
Inglorious bastards	NARROW	36	-29.062	21.518	7.023	5.775
	MEDIUM	8	-28.865	23.293	6.895	5.832
	WIDE	4	-30.439	23.945	8.546	6.017
Iron man 2	NARROW	18	-34.346	12.126	8.474	7.247
	MEDIUM	18	-34.401	13.889	8.780	7.343
	WIDE	12	-35.235	15.111	10.186	7.557
Pirates of the Carabbean 3	NARROW	26	-29.445	13.005	8.831	4.543
	MEDIUM	17	-29.504	18.035	10.835	5.868
	WIDE	5	-30.542	18.068	12.425	6.748
Save Private Ryan	NARROW	16	-31.155	12.004	9.516	7.768
	MEDIUM	23	-28.974	16.986	11.321	9.465
	WIDE	9	-26.187	21.555	11.258	8.828
Star Trek	NARROW	23	-31.107	11.175	12.253	6.860
	MEDIUM	12	-29.098	17.691	12.443	7.595
	WIDE	13	-25.374	22.808	10.841	7.826
Star wars 1	NARROW	24	-27.445	16.326	9.608	6.813
	MEDIUM	20	-25.050	22.598	9.037	5.789
	WIDE	4	-23.762	25.356	9.085	6.234
Transformers 2	NARROW	27	-30.817	7.614	7.971	4.701
	MEDIUM	13	-30.696	10.108	10.801	6.827
	WIDE	8	-30.572	11.823	13.665	9.113
Tron: Legacy	NARROW	24	-27.481	20.384	8.453	4.943
	MEDIUM	19	-24.926	22.685	8.029	4.635
	WIDE	5	-23.386	22.717	7.869	4.892
Wall-E	NARROW	15	-30.496	11.196	7.111	5.051
	MEDIUM	25	-30.004	14.341	8.912	5.960
	WIDE	8	-29.517	16.215	10.759	6.679
The Gladiator	NARROW	21	-29.631	15.581	9.054	7.007
	MEDIUM	21	-26.657	23.688	9.350	7.455
	WIDE	6	-23.997	27.131	9.386	7.820
Moulin Rouge!	NARROW	32	-33.262	15.686	8.699	7.793
	MEDIUM	12	-31.355	22.647	10.935	9.030
	WIDE	4	-28.800	26.387	13.517	10.001
AHS	NARROW	12	-28.820	4.283	3.856	1.996
	MEDIUM	20	-28.360	7.758	6.280	3.344
	WIDE	16	-28.346	11.242	8.345	4.641
Falling skies	NARROW	12	-25.791	2.444	2.413	1.421
	MEDIUM	27	-25.036	3.652	3.556	2.452
	WIDE	9	-25.756	4.939	4.813	3.317
Great migrations	NARROW	11	-29.102	4.693	3.021	2.405
	MEDIUM	12	-29.403	6.268	4.480	3.124
	WIDE	25	-30.809	6.090	4.944	3.088
Medici	NARROW	9	-29.385	4.678	4.387	3.358
	MEDIUM	22	-29.909	7.033	5.718	4.604
	WIDE	17	-31.319	7.448	7.157	5.186
Terra nova	NARROW	14	-26.416	5.797	4.702	3.335
	MEDIUM	9	-26.435	6.606	6.401	4.579
	WIDE	25	-28.234	6.772	7.592	5.201

Table 9 "Detailed Stereo Program Loudness Values of all versions measured according to BS1770-2"

PROGRAM	VERSION	PREFERENCES	PL	LRA	MML	MSL 3s
The dark knight	NARROW	25	-27.224	9.288	7.373	5.026
	MEDIUM	20	-24.038	14.194	8.539	5.860
	WIDE	3	-19.250	19.575	9.710	6.237
Inglorious bastards	NARROW	36	-29.688	19.090	6.131	5.155
	MEDIUM	8	-29.820	21.186	6.608	5.519
	WIDE	4	-31.460	22.088	8.101	5.601
Iron man 2	NARROW	18	-32.777	9.856	6.180	5.040
	MEDIUM	18	-32.836	12.175	7.108	5.101
	WIDE	12	-33.615	13.314	10.862	6.138
Pirates of the Carabbean 3	NARROW	26	-29.264	12.916	4.575	3.206
	MEDIUM	17	-29.434	18.220	6.561	4.628
	WIDE	5	-30.719	19.928	8.033	5.833
Save Private Ryan	NARROW	16	-30.510	9.364	6.301	4.679
	MEDIUM	23	-28.854	14.207	8.804	6.983
	WIDE	9	-25.547	20.035	11.338	9.130
Star Trek	NARROW	23	-30.048	9.997	7.579	5.447
	MEDIUM	12	-28.517	16.003	9.368	7.033
	WIDE	13	-25.012	21.673	10.756	7.951
Star wars 1	NARROW	24	-28.073	12.295	6.083	4.543
	MEDIUM	20	-26.322	18.633	8.262	5.922
	WIDE	4	-24.935	22.266	12.050	8.731
Transformers 2	NARROW	27	-30.461	5.529	5.575	3.627
	MEDIUM	13	-30.342	7.958	7.169	4.829
	WIDE	8	-30.449	10.265	11.569	7.714
Tron: Legacy	NARROW	24	-29.136	16.169	5.318	3.203
	MEDIUM	19	-30.716	24.015	8.856	6.253
	WIDE	5	-24.871	23.222	8.287	5.144
Wall-E	NARROW	15	-29.585	9.893	5.643	3.917
	MEDIUM	25	-28.858	13.893	7.435	4.719
	WIDE	8	-28.185	15.646	11.633	5.369
The Gladiator	NARROW	21	-28.882	14.394	8.632	7.070
	MEDIUM	21	-26.187	22.083	9.469	7.800
	WIDE	6	-23.590	26.124	10.216	8.325
Moulin Rouge!	NARROW	32	-31.545	15.836	7.828	6.660
	MEDIUM	12	-29.455	22.711	9.961	7.754
	WIDE	4	-26.926	26.380	13.377	9.019
AHS	NARROW	12	-29.197	2.535	3.107	1.373
	MEDIUM	20	-28.904	5.824	5.690	2.655
	WIDE	16	-29.299	9.331	7.995	4.218
Falling skies	NARROW	12	-26.475	2.157	1.827	1.427
	MEDIUM	27	-25.719	3.731	3.296	2.406
	WIDE	9	-26.510	5.273	4.673	3.339
Great migrations	NARROW	11	-28.494	3.266	2.477	1.931
	MEDIUM	12	-28.695	4.612	3.541	2.548
	WIDE	25	-29.978	4.315	3.854	2.388
Medici	NARROW	9	-28.726	3.651	3.727	2.685
	MEDIUM	22	-29.153	5.472	4.837	3.894
	WIDE	17	-30.437	5.834	5.801	4.326
Terra nova	NARROW	14	-26.194	4.592	4.237	2.768
	MEDIUM	9	-26.170	4.954	5.480	3.916
	WIDE	25	-27.587	5.320	6.589	4.412

Table 10 “Detailed MCA 5.1 Program Loudness Values of all versions measured according to BS1770-2”