Metrics for Quantifying Loudness and Dynamics[1]

Earl Vickers

The Sound Guy, Inc., Saratoga, CA 95070
sfx@sfxmachine.com

ABSTRACT

(This material was originally intended as part of the article "The Loudness War: Background, Speculation and Recommendations" [1] but was removed for reasons of scope and to keep that article to a manageable length.) In this paper, I briefly review a variety of metrics for quantifying the loudness and dynamic spread of audio recordings. This review was motivated by the need for objective ways of measuring the effects of hypercompression as used in the loudness war.

1. LOUDNESS AND DYNAMICS DESCRIPTORS

If the loudness war is indeed a problem, we need to find measurements that correlate to the problem so we can estimate the effects of various types of processing.

It may prove impossible to measure a wide variety of effects, such as microdynamic and macrodynamic compression, clipping and aliasing, not to mention the likelihood of perceptual effects such as listening fatigue, with a single descriptor. In addition, some problems are hard to measure without access to both the 'before' and 'after' tracks. For example, blind hypercompression detection may be difficult without the original recordings because, for all we know, the audio being tested may be the original with no additional processing (e.g., double-reed instruments [2] having little dynamic range to begin with).

2. LONG-TERM LOUDNESS DESCRIPTORS

A number of long-term loudness metrics have been developed for a variety of purposes: loudness meters, loudness control, sound quality research, and estimation or prediction of the perceived loudness of an audio program. [8]

2.1.1. Single-band Loudness Models

The "equivalent sound level," *Leq*, is a long-term energy integration; it measures equivalent power over time and may be used as an estimate of sustained loudness. It can be defined as

$$Leq_w = 10\log_{10}\left(\frac{1}{T}\int_0^T\left(\frac{P_w(t)}{P_{Ref}}\right)^2 dt\right),$$

in dB, where $P_w(t)$ is the frequency-weighted sound pressure at time t , T is the time interval of interest, and the subscript w refers to the type of frequency weighting (for example, the B-weighting curve that approximates the 70 phon equal loudness contour and applies to playback at moderate levels). [8] [9] [4] [22] When measuring digital or electrical signals instead of acoustic signals, amplitude may be used instead of sound pressure.

The *Leq* metric was derived to show the potential hearing damage from exposure to sustained loud sounds with varying levels. In the article "Are Movies Too Loud?", Allen concluded that *Leq* was better than a fast, C-weighted peak level measurement for estimating subjective long-term loudness. [4]

A number of alternate frequency weightings have been suggested. Vickers proposed using a 200 Hz first-order Butterworth high-pass filter as a simplified approximation of the low end of the B-weighting curve.

---

[1] v. 1.0, September 14, 2010

This one-sided approximation was motivated by low computational cost and by the fact that musical signals typically have relatively little energy at high frequencies. The cutoff frequency was selected partly because of the filter's intended use for pre-emphasis in a single-band dynamics compressor; in addition to compensating for the hearing curve at moderate playback levels, the high-pass filter would help minimize "pumping" due to loud bass notes. [9]

Soulodre and Norcross investigated a number of objective loudness measures including the above proposal, which they referred to as *Leq(Bhp)* (for Butterworth high-pass). Other measures examined included the A-, B- and C-weighting curves, Terhardt's weighting based on the absolute threshold of hearing (referred to as *Leq(ATH)*), and a new weighting, *Leq(RLB)*, which was another revision of the low end of the B-weighting curve; some of these curves are shown in Figure 1. Their study supports the supposition that a high-frequency roll-off in the weighting function actually degrades the performance of the loudness estimation. Soulodre and Norcross found that *Leq(RLB)* was the best of the basic objective loudness measures evaluated. [22]

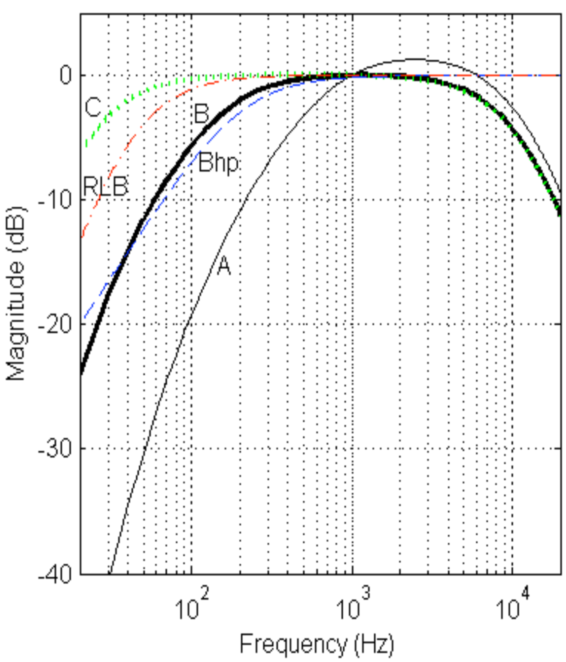

Figure 1. Weighting curves: *A* (solid line), *B* (heavy solid line), *C (*green dotted line), *Bhp* (blue dashed line), *RLB* (red dash-dotted line).

Note from Figure 1 that *Leq(RLB)* appears to be slightly closer to the C- than to the B-weighting curve, which might lead one to expect that it would be more suitable for higher playback levels between 70 and 100 dB. Also note that *Leq(RLB)* is essentially identical to RMS (flat response) over most of the frequency range, with only about 1 dB of attenuation at 100 Hz and 3 dB of attenuation around 60 Hz.

Given that the (inverted) A-, B- and C-weighting curves only poorly approximate the equal-loudness contours [21], one wonders how the actual inverted equal-loudness contours might fare in listening tests.

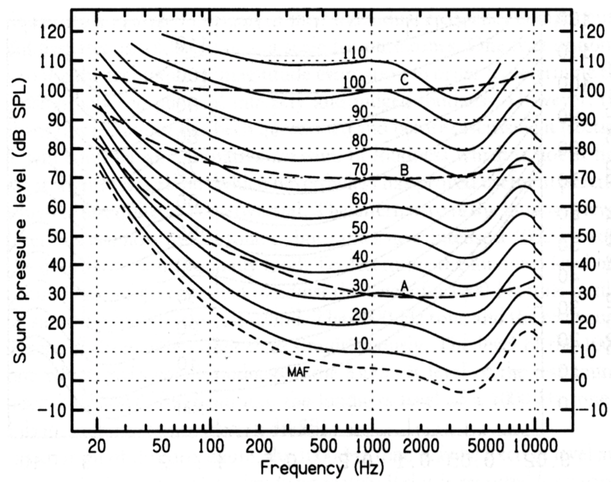

Figure 2. Equal loudness contours, with inverted A-, B- and C-weighting curves superimposed, from [21]. (Note that these contours don't necessarily apply when multiple frequencies are present simultanously.)

2.1.2. Multi-band Psychoacoustic Loudness Models

Stevens and Zwicker developed advanced multiband loudness models, which were incorporated into the ISO 532-A and 532-B standards, respectively [23]. Moore and Glasberg proposed additional improvements including a method of calculating the loudness of fluctuating sounds. [12] [24] [8]

2.1.3. Listening Tests with Speech and Music

Seefeldt et al developed a new objective loudness measure based on modifications to previous psychoacoustic models and found that it performed significantly better than other methods tested using an expanded database of audio program samples [25].

In 2004, Skovenborg and Nielsen wrote an extensive review that evaluated twelve models of long-term loudness perception, including several *Leq* models, two variants of the Zwicker method, and two new models developed by TC Electronic. The study used speech and music stimuli centered around a reference level of 70 dB SPL, as a compromise between typical TV and cinema levels. Overall, they found their new models to have the best performance, followed by *Leq(RLB)*. [8]

Despite the greater sophistication of the multi-band models, *Leq* models have done surprisingly well in listening tests. However, simple single-band descriptors such as those based on *Leq* are limited because compromises have to be made in the choice of frequency weighting, since the equal loudness contours constitute a family of curves and cannot be duplicated with a single linear filter. [9] In addition, these methods cannot model spectral loudness summation [8], and they often assume that dynamic and stationary signals with the same overall weighted energy are equally loud [15].

2.1.4. ITU-R BS.1770

In 2006-2007, *Leq(RLB)* became the basis of the ITU-R BS.1770-1 recommendation, which defined the RLB weighting curve as

$$h(z) = \frac{1 - 2z^{-1} + z^{-2}}{1 - 1.99004745483398z^{-1} + 0.99007225036621z^{-2}},$$

for a 48 kHz sampling rate; the recommendation states that coefficients for other sampling rates should be chosen to provide the same frequency response. BS.1770-1 also includes an extension for use with multi-channel signals, including a +4 dB high shelf pre-filter to account for the acoustic effects of the head. [6] A block diagram of the algorithm is shown in Figure 3. BS.1770-1 has the advantages of being simple, computationally inexpensive, widely accepted, and reasonably robust with typical audio content.

2.1.5. Crest Factor, and "Loudness Crest Factor"

Crest factor, or peak-to-average ratio, is a signal's peak amplitude divided by its RMS value:

$$CF = \frac{|x|_{peak}}{x_{rms}},$$

typically expressed in dB. The inverse of the crest factor, then, is the RMS average of a signal after peak-normalization. Processes such as multiband compression and reverberation tend to lower the crest factor.

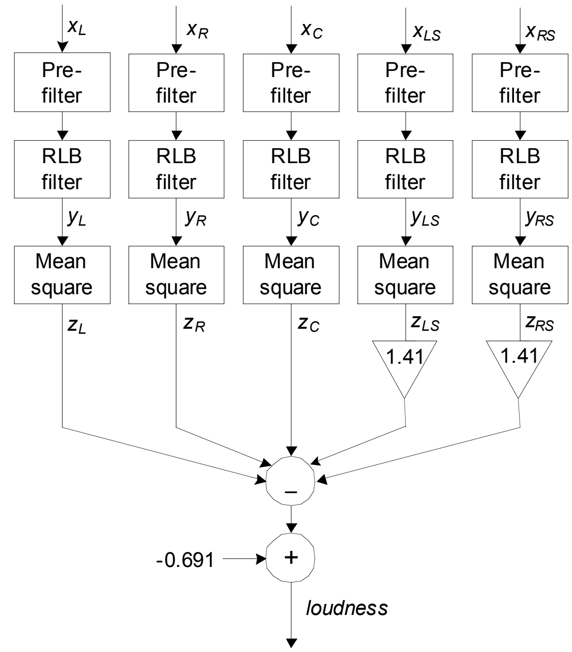

Figure 3. Block diagram of BS.1770 loudness meter. (The -0.691 value is intended to compensate for the total gain of the filters at 1 kHz.)

A perceptual metric related to crest factor, which might be termed "loudness crest factor," could be defined as a signal's peak amplitude divided by its BS.1770 loudness (converted to linear units):

$$LCF = \frac{|x|_{peak}}{10^{0.05\,BS.1770}}.$$

Alternatively, the inverse of this metric would estimate the loudness of a peak-normalized signal.

2.1.6. Replay Gain

Replay Gain, proposed in 2001 and currently supported in various products, computes a gain that adjusts the music's loudness to match that of pink noise at -20 dBFS when reproduced at 83 dB SPL (slow, C-weighted). The intent is to embed the desired Replay Gain value within a music track as metadata, using tagging standards such as ID3v2 [26]. The Replay Gain procedure applies frequency weighting, computes an RMS value over 50 ms blocks and uses a histogram to select the loudness exceeded by only 5% of the blocks in a track. [7]

2.1.7. Dealing with Silence and Fluctuating Intensities

With averaging methods, periods of relative silence can cause an undesirable reduction of the estimated loudness. More generally, little is known about how people form subjective long-term loudness estimates of audio with fluctuating intensities: e.g., according to the average of the non-silent portions; according to the level of the louder portions; or as a weighted average of the short-term loudness values, with louder frames weighted more heavily.

As mentioned in [9], we speak of the "loudness" of a sound, not its "softness" or "quietness"; this suggests the possibility that we may tend to weight louder portions more heavily when judging the long-term loudness of signals with fluctuating intensities. Zwicker and Fastl stated that in many cases, the long-term loudness of a dynamically changing sound is well characterized by its N_5 loudness; i.e., the loudness value that is exceeded only 5% of the time [13]. A study by Zhang and Zeng found that, for two stimuli with the same RMS level, the one with greater temporal fluctuations was sometimes perceived as significantly louder [14]. Johnston conducted a small test that also suggested that dynamic signals are perceived as louder than stationary signals having the same overall energy [15].

The loudness of compressed speech

The opposite conclusion might be drawn from a study by Moore et al, which showed that amplitude-compressed speech was perceived as louder than dynamic, uncompressed speech at the same RMS level. This latter effect may be due to the fact that, while compressors reduce the amplitude of the major peaks, they spend more time amplifying the low-level portions. [3] Still, this study raises new questions about how we perceive loudness – why would the amplification of low-level portions have more influence on our perception of loudness than attenuation of the loudest portions, and more effect than the overall RMS level?

One possibility is that the ear may pick up on fluctuations in low-level background noise and interpret those as straining or shouting [16]. A more likely possibility is that the ear may interpret non-linear distortion (caused by the compressor's rapid gain changes) as straining and therefore louder. It might be interesting to study the effects of slow (macrodynamic) compression and fast (microdynamic) compression separately.

Is this result specific to speech? For example, are there similarities between compressed speech and shouting? The brain has a great deal of *a priori* knowledge about how speech normally sounds. Perhaps we overlook the compressed peak levels and judge the loudness of speech largely on how well we can hear the relatively quiet consonants – we may decide that if the articulation is quite audible, the person must be talking loudly. It would be interesting to see if the results of this study hold for music or other sounds, such as nature sounds.

Another interesting and useful test would be to determine whether compression increases the perceived loudness of speech and music even after normalization to the same BS.1770 loudness value. This test would be similar to the Moore et al study [3], but it would use RLB weighting instead of simple RMS and would be conducted using music as well as speech. If compressed music is significantly louder than uncompressed music, even after BS.1770 normalization, this might slightly reduce the effectiveness of loudness normalization in removing incentives for hypercompression.

Strategies for estimating the loudness of dynamic signals

Skovenborg and Nielsen stated that, especially with methods such as *Leq*, the measurement can benefit from using a silence detector, to avoid giving silent segments excessive influence over the loudness estimate [8]. Other ways of dealing with varying intensities include statistical methods such as the N_5 loudness [13], and the hastily named "long-term loudness matching level," *LLML*, which averages short-term loudness measurements and optionally weights louder frames more heavily [9].

Each of these strategies for estimating the perception of time-varying loudness relies on setting seemingly arbitrary constants: gating threshold and block size (for silence detectors), the loudness percentile threshold (for methods like N_5), and the "emphasis parameter" that determines how heavily to weight louder frames (for *LLML*) [8] [13] [9]. It is likely that a full comparative evaluation of these strategies has never been performed; such evaluation is especially difficult because of the need to calibrate the constants for the best fit with human perception, as well as the dependence on the type of audio material being tested.

Silence gating and R128

Grimm et al found that silence gating was useful for wide loudness range programs such as movies with relatively quiet segments; they suggested adding a gate to the ITU-R BS.1770-1 standard [17].

The European Broadcasting Union's PLOUD group is currently finalizing recommendation R128, which encourages the abandonment of peak normalization in favor of loudness normalization using a Programme Loudness descriptor, based on BS.1770 with the addition of a gating function. R128 offers two additional audio descriptors: Loudness Range (see Section 0) and True Peak Level. [18][19]

3. DYNAMICS DESCRIPTORS

Dynamics descriptors are used to measure the variation of loudness. The most familiar such descriptor, dynamic range, refers to the ratio between the smallest and largest amplitude values; for example, a 16-bit CD has a dynamic range of about 96 dB.

The term "dynamic spread" was proposed as a generalization of dynamic range. Range is one measure of the spread of a data set and is defined by only two numbers (the largest and smallest measurements). Under this definition, a massively clipped song ending in a reverb tail that decays to one bit might be said to have full dynamic range – range actually tells us little of interest about such a signal. A statistical measure of dynamic spread, such as one based on the standard deviation or mean absolute deviation, might be more useful and robust. [9]

The conventional definition of dynamic range looks at signals one sample at a time and has no notion of perceptual loudness or of different time-scales or frame sizes. Skovenborg and Lund proposed Loudness Range (LR, based on their earlier Consistency measure [10]), as a descriptor for measuring the variation of loudness on a macroscopic time-scale; Loudness Range uses a statistical method similar to Interquartile Range (IQR), along with an adaptive measurement gate [11]. Another descriptor, "Density," measures the variation of loudness on a microscopic time-scale [10]. Skovenborg and Lund regarded Density as similar to dynamic spread [10], though dynamic spread may respond either to macrodynamic or microdynamic variations, depending on the frame size selected [9].

The Pleasurize Music Foundation provides a Tischmeyer Technology "Dynamic Range Meter" that measures the average cumulative difference between peak and RMS levels over the duration of a song or album, taking into account only the highest 20% of the levels. The resulting value, DR, is presented as a whole number in dB for ease of comprehension. [5] This measurement does not include frequency weighting. In addition, it appears to be a variant of crest factor, rather than dynamic range per se [27]. While the term "dynamic range" has the advantage of widespread familiarity, and while crest factor may in fact be more useful than true dynamic range as a measure of loudness variation, this use of the term "dynamic range" is potentially confusing.

4. ACKNOWLEDGMENTS

The author wishes to thank Martin Walsh, Rob Toulson, Jon Boley, and William Campbell for their helpful comments and suggestions.

5. REFERENCES

[1] Earl Vickers, "The Loudness War: Background, Speculation and Recommendations," presented at the AES 129th Convention, San Francisco, CA, 2010 November 4-7.

[2] Melville Clark, and David Luce, "Intensities of Orchestral Instrument Scales Played at Prescribed Dynamic Markings," *J. of the Audio Engineering Society*, Vol. 13, No. 2, 1965 April.

[3] Brian C. J. Moore, Brian R. Glasberg, and Michael A. Stone, "Why Are Commercials So Loud? – Perception and Modeling of the Loudness of Amplitude-Compressed Speech," *J. Audio Eng. Soc.*, Vol. 51, No. 12, 2003 December.

[4] Ioan Allen, "Are Movies Too Loud?", presented at SMPTE Film Conference, March 22, 1997, http://www.dolby.com/uploadedFiles/zz-_Shared_Assets/English_PDFs/Professional/54_Moviestooloud.pdf, accessed June 26, 2010.

[5] Pleasurize Music, http://www.pleasurizemusic.com, accessed 2010 June 5.

[6] ITU-R, "Rec. ITU-R BS.1770, Algorithms to measure audio programme loudness and true-peak audio level," International Telecommunications Union, 2006.

[7] http://www.replaygain.org, accessed 2010 June 5.

[8] Esben Skovenborg and S.H. Nielsen, "Evaluation of Different Loudness Models with Music and Speech Material," presented at the AES 117th Convention, San Francisco, 2004 October 28-31 .

[9] Earl Vickers, "Automatic Long-term Loudness and Dynamics Matching," presented at the AES 111th Convention, New York, 2001.

[10] Esben Skovenborg and Thomas Lund, "Loudness Descriptors to Characterize Programs and Music Tracks," presented at the AES 125th Convention, San Francisco, CA, 2008 October 2-5.

[11] Esben Skovenborg and Thomas Lund, "Loudness Descriptors to Characterize Wide Loudness-Range Material," presented at the AES 127th Convention, New York, NY, 2009 October 9-12.

[12] B. C. J. Moore and B. R. Glasberg, "A Revision of Zwicker's Loudness Model," *Acoustica – Acta Acoustica*, vol. 82, 1996, pp. 335–345.

[13] Zwicker, E., Fastl, H., *Psychoacoustics: Facts and Models*, Springer-Verlag, 2nd Edition, 1999.

[14] C. Zhang and F-G. Zeng, "Loudness of Dynamic Stimuli in Acoustic and Electrical Hearing," *J. Acoust. Soc. of Am.*, vol. 40, pp. 71-78, 1966.

[15] James D. (jj) Johnston, "Testing Loudness Models – Real vs. Artificial Content," presented at the AES 125th Convention, San Francisco, CA, 2008 October 2-5.

[16] Earl Vickers, personal communication (dream), 2010 September 14.

[17] Eelco Grimm, Esben Skovenborg, and Gerhard Spikofski, "Determining an Optimal Gated Loudness Measurement for TV Sound Normalization," presented at the AES 128th Convention, London, UK, 2010 May 22-25.

[18] "Loudness normalisation and permitted maximum level of audio signals," EBU Technical Recommendation R128 (draft), European Broadcasting Union, http://tech.ebu.ch/loudness, 2010.

[19] "Practical Guidelines for Production and Implementation according to EBU Technical Recommendation R128," EBU Technical Document 3343 (draft), European Broadcasting Union, http://tech.ebu.ch/loudness, 2010.

[20] Ole Kirkeby, "A Balanced Stereo Widening Network for Headphones," AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, 2002 June.

[21] Brian C. J. Moore, *An Introduction to the Psychology of Hearing*, Fourth Ed., London: Academic Press, 1997.

[22] Gilbert Soulodre and Scott Norcross, "Objective Measures of Loudness," presented at the AES 115th Convention, New York, NY, 2003 October 10-13.

[23] ISO, "Acoustics. Method for calculating loudness level. International Standard ISO 532 (1.ed.)", International Organisation for Standardisation, 1975.

[24] B. R. Glasberg and B. C. J. Moore, "A Model of Loudness Applicable to Time-Varying Sounds," *J. Aud. Eng. Soc.*, Vol. 50, No. 5, 2002.

[25] Alan Seefeldt, Brett Crockett, and Michael Smithers, "A New Objective Measure of Perceived Loudness," presented at the AES 117th Convention, San Francisco, CA, 2004 October 28-31.

[26] ID3, http://www.id3.org/, accessed 2010 July 12.

[27] William Campbell, personal communication, 2010 August 27.

[28] Jon Boley, Michael Lester, and Christopher Danner, "Measuring Dynamics: Comparing and Contrasting Algorithms for the Computation of Dynamic Range, presented at the AES 129th Convention, San Francisco, CA, 2010 November 4-7.