Audio Engineering Society

Convention Paper

Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

Loudness Descriptors to Characterize Programs and Music Tracks

Esben Skovenborg[1] and Thomas Lund[2]

[1] TC Group Research, Risskov, DK-8240 Denmark
EsbenS@tcelectronic.com

[2] TC Electronic, Risskov, DK-8240 Denmark
ThomasL@tcelectronic.com

ABSTRACT

We present a set of key numbers to summarize loudness properties of an audio segment, broadcast program or music track: the *loudness descriptors*. The computation of these descriptors is based on a measurement of loudness level, such as specified by the ITU-R BS.1770. Two fundamental loudness descriptors are introduced: Center of Gravity and Consistency. These two descriptors were computed for a collection of audio segments from various sources, media and formats. This evaluation demonstrates that the descriptors can robustly characterize essential properties of the segments. We propose three different applications of the descriptors: for diagnosing potential loudness problems in ingest material; as a means for performing a quality check, after processing/editing; or for use in a delivery specification.

1. INTRODUCTION

Level control in digital audio production has traditionally been based on measuring *peak level*; often just taking the value of individual samples, which requires a minimum of processing-power. Such practice makes material with low dynamic range appear louder, and has led to the so-called *loudness war* in music production (see e.g. [1]). Furthermore, extra distortion may occur at D/A-converters and perceptual codecs in consumer equipment, as a consequence of inappropriate level control [2].

As illustrated in Figure 1, material mixed for low dynamic range reproduction generally comes out the loudest when level control is based on a peak level measurement. The challenges of level- and loudness-control in a broadcast context are further described in [3, 4, 5].

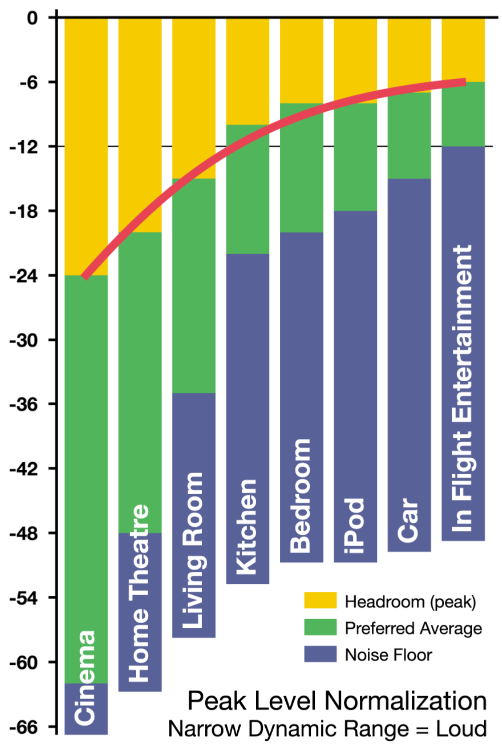

Figure 1. The bars illustrate dynamic-range properties for different playback scenarios [6]. The red line indicates average level when level control is based on peak-level measurement.

Based on loudness measurement technology, the **loudness descriptors** presented in this paper provide the means for a superior level control strategy. A descriptor summarizes an essential property of an audio segment, broadcast program, or music track. Besides giving the audience a chance to get a non-distorted audio experience, the loudness descriptors allow content of different dynamic range to be interspersed, for instance a cinema movie and a TV commercial, without more level-jumps than desired.

A single loudness descriptor is not sufficient to characterize different program genres and mixing-styles. Three different properties, relevant to the audio engineer, should be monitored and controlled independently:

a) The audio level should not exceed a certain threshold, in order to avoid distortion in the various audio media formats, processors, codecs, and converters, in the signal chain. In the digital domain, the **true peak-level** can be monitored (see e.g. [7]).

b) The overall loudness of the material should match the desired *reference* loudness of the program. Furthermore, the overall loudness should be sufficiently low, to provide headroom for the dynamics of the material (i.e., to avoid over-compression). The loudness descriptor **Center-of-Gravity** will measure this property.

c) The loudness should not make unaesthetic and annoying jumps between program segments. Furthermore, the loudness variation within a program should match the expected dynamic range capabilities of listener's playback system. The loudness descriptor **Consistency** will measure this property.

Although a peak level measurement is needed to stay clear of distortion in production, distribution and reproduction; music mastering and delivery specs in broadcast should not be based on it. Standardized long-term loudness description is a superior alternative, and more suitable for the way today's consumers use a variety of audio sources.

2. DEFINITION OF DESCRIPTORS

The **Center of Gravity** (CoG) descriptor measures the overall loudness of the segment. That is, if one segment should be aligned in loudness with another using only a gain offset, that offset should be the difference between the CoG values of the two segments.

The main difference between the CoG and an integrating loudness measurement (such as the Leq-type measurement specified in BS.1770 [7]) is that the CoG employs an *adaptive gate* in order to be robust against 'silence' in the program yet making no rigid assumptions about the absolute levels of the material to be measured. By employing a gate, the measurement essentially ignores regions of the material which are too quiet to be considered part of the program – regions that would otherwise have *biased* the CoG measurement. In our development of the descriptors, we found that it is impossible to find one single fixed gate threshold that would work well with different types and genres of program material: If a fixed threshold was set to a relatively high level, it would excessively "gate out" softer parts of for example movie soundtracks; whereas if the fixed threshold was set to a lower level, the gate might *not* gate out regions consisting of (say) relatively loud background noise. Instead of trying to make this impossible compromise, we are using an adaptive gate which employs a *relative* gating threshold.

The **Consistency** descriptor measures the *variation of the loudness on a macroscopic timescale*. In other words, applying a loudness-correction processor typically *increases* the Consistency of the material; the value of the Consistency corresponds roughly to the amount of gain change that the processor might apply. In order to achieve a good compromise between precision and robustness, the measurement of Consistency is based on the *statistical distribution* of measured loudness. Thus, a short but very loud event would not affect the consistency of a longer segment significantly, and similarly the *fade-out* at the end of a music track would not decrease the measured Consistency noticeably.

As an optional supplement to the Consistency, a third loudness descriptor – **Density** – measures the variation of loudness on a *microscopic* timescale, corresponding to the amount of dynamic compression applied.

The loudness descriptors are measured in units of **LU** (Loudness Units), as recommended for loudness meters in BS.1771 [8]. In contrast to the phon- and sone-scales, used to measure loudness (level) within psycho-acoustics, the LU-scale is a dB-scale. This choice was presumably made by the ITU in order to make the LU-measurements more operational in the context of audio engineering. Hence, increasing the level of a signal by X dB will increase the corresponding loudness measurement by X LU. The **LU_{FS}** unit refers to that LU-scale which is calibrated w.r.t. absolute level, as specified in BS.1770.

Both the CoG and Consistency are negative values. The louder the program material, the closer the CoG comes to 0 LU_{FS}. Analogously, the more consistent the material is (i.e., smaller loudness range), the closer the Consistency measure comes to 0 LU, which is the maximum consistency. The less consistent the material, the more negative the Consistency descriptor. (LU is employed rather than LU_{FS} for measurements of *relative* loudness level.) Table 1 shows the theoretical and practical numerical ranges of the two loudness descriptors.

The computation of the loudness descriptors, presented here, is based entirely on the short-term loudness and long-term loudness measurements (Figure 2). The difference between the short-term and long-term loudness measurements is the analysis window length or time constants used [9]. The short-term and long-term loudness are themselves derived from a loudness measurement algorithm, such as the method specified in BS.1770 [7], which is defined for both mono, stereo, and 5.1-channel input signals; other loudness algorithms could alternatively be employed [10]. The loudness descriptors are themselves independent of the sample rate and format of the input signal.

Table 1 shows some properties of the CoG and Consistency, using the BS.1770 loudness measurement algorithm as pre-processor.

The test-signals, referred to in the table, are composed as follows:

- **Sine 1 kHz @ -20**: 1 kHz sine wave, mono, -20 dBFS peak level.
- **Sine 1 kHz @ -20, -40**: 1 kHz sine wave, mono, -20 dBFS peak level, followed by another 1 kHz sine wave, mono, -40 dBFS peak level, of equal duration.

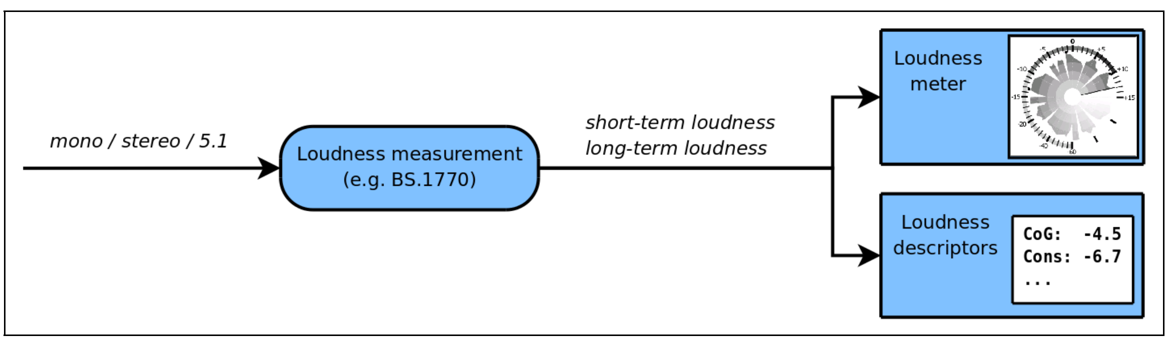

Figure 2. From audio signal, via loudness measurement, to loudness metering and/or loudness descriptors.

	CoG	*Cons*
Maximum value	12 LU_{FS}	0 LU
Minimum value	−80 LU_{FS}	−40 LU
Very high value, in practice	−5 LU_{FS}	−0.5 LU
Very low value, in practice	−30 LU_{FS}	−10 LU
Sine 1 kHz @ -20	−23.01 LU_{FS}	0 LU
Sine 1 kHz @ -20, -40	−25.98 LU_{FS}	−10 LU
+ X dB gain	+ X LU_{FS}	unchanged
Segment is repeated	unchanged	unchanged
Appending low-level signal	unchanged	unchanged

Table 1. Properties of the loudness descriptors (some values are approximate).

Vickers proposed two measures in order to *normalize* audio-files in loudness and dynamics [11]. His "long-term loudness matching level" is analogous to our CoG, while his "dynamic spread" is related to our Density; he does not present a separate measure similar to Consistency.

The MPEG-7 Audio standard specifies "*…a set of low-level Descriptors, for audio features that cut across many applications (e.g., spectral, parametric, and temporal features of a signal), and high-level Description Tools that are more specific to a set of applications*." [12]. None of those descriptors, however, deal specifically with loudness. But one of the goals of the MPEG-7 Descriptors might also apply to the loudness descriptors: to define a set of common measurement procedures that will allow the measurements to be compared and exchanged across different implementers and applications.

3. APPLICATIONS

For audio descriptors to be of value across a wide range of applications, they should not work only in a specific context, e.g. cinema, or with a specific audio content, such as spoken word. Consequently, the descriptors proposed are of a *universal* nature.

In music, post and film **production**, universal loudness descriptors help the audio engineer or journalist identifying potential reproduction and compatibility problems when a program is distributed to various platforms. The loudness meter snapshots in Figure 3 illustrate the diversity of genres. The engineer retains the artistic freedom to aim at a certain platform, and to perhaps specify a challengingly large loudness range, but the Consistency measure makes her aware how much loudness range restriction can be expected to happen during delivery.

Another key use for audio descriptors is in program specification. Tight **delivery specifications** across genres and audio formats has been on the broadcaster's wish list for years. The more precisely external production meets the criteria of a station, the more predictable the results. At the same time, the lower the

Figure 3. Loudness meter also displaying the two loudness descriptors. The Matrix (left), WDR Nachrichten (middle), Madonna's Hung Up (right). Note the intro-jingle at the very beginning of the WDR news.

workload for correcting external programming at the station. Loudness descriptors would also be relevant in a **mastering specification** – especially when considering the ongoing loudness war.

Applied during **ingest** or inside a *station server*, level offsets may be defined, and potential loudness problems diagnosed. Is the program ready to *air* or is it not? Where correction is needed, appropriate processing leaves a window around the program's Center of Gravity unprocessed in order not to disturb the balance between foreground and background elements of a mix (game show, sports, music, news, commercial, film etc). Automatically, soft parts are brought up, and too loud parts are reduced in level. Result: A better audio quality at a lower handling effort.

In **program delivery**, a gain offset for each program is a well indicated way of avoiding inter-program level-jumps; for instance between scenes of a movie and interstitials, or between a music track from 1975 and another one from 2005. Center of Gravity is the foundation for a universal and transparent gain offset mechanism.

Post-transmission, results may be systematically and consistently logged using the same audio descriptors, thereby creating a closed loop spanning the entire range from production to delivery, and streamlining the flow of audio in a way that's based on open standards.

4. EMPIRICAL EVALUATION

The Center-of-Gravity and the Consistency were computed for a collection of audio segments. The segments were taken from different media: DVD soundtracks, music CDs, and radio- and TV-broadcasts. The collection consists of characteristic samples from different genres: present-day pop music, 80'es pop/rock, classical music, films, broadcast programs and commercials. The duration varies between 15 seconds and more than 2 hours. Note that the loudness descriptors make no assumptions about contents of the segments; they should be accurate yet robust enough to work universally.

The CoG, Consistency, genre/source, year, and duration of each segment are listed in table 2 (p. 9). The loudness descriptors expose the different strategies for level-normalizations and production that have been employed for the different genres of segments. Furthermore, the descriptors are able to quantify these differences, while being robust against 'silence', fade-outs, an atypically loud gun-shot, etc. which might otherwise have polluted the descriptor values.

All the segments analyzed here were transferred digitally from the source medium, except for the movie soundtracks from DVD. The soundtracks were recorded (digitally) from the analog outputs of a Sony NS900V DVD player as 5+1 discrete channels. In these cases, the recording gain was adjusted – per segment – to achieve a true-peak level of -1 dBFS, for the purpose of this evaluation.

Some trends can be observed within the different genres of material:

- **Contemporary pop/rock (CD, stereo):** By "contemporary" we mean that the CD was mastered within the past decade or so. This music is infamous for its (ab)use of compression, limiting and peak-normalization. The CoG of a track is frequently in the range -10 to -5 LU_{FS}, with very little loudness variation, a Consistency of -3 to -1 LU. Music which is mixed/mastered in this way suffers from the problems described in the Introduction. Note that the relation between a loudness level in LU_{FS} and the corresponding loudness in phon or sone is *not* specified by the BS.1770 standard (besides, this relation depends on the amplification of the given reproduction system). Hence we cannot say exactly how loud -5 LU_{FS} really is – only that the CoGs within this genre are 10 to 15 LU louder than those of the other genres!

- **80'es pop/rock (CD, stereo):** We use "80'es" to say that the CD was mastered within the first decade of the CD-medium's life (i.e., the 1980'es and early 90'es). Back then, some twenty years ago, the CD-mastering practice was considerably less aggressive. A common CoG is around -20 to -15 LU_{FS}. Even the well-produced – and presumably quite "radio-optimized" – *Black or White* has a CoG of no more than -13.9 LU_{FS}, which illustrates how the loudness war has developed since then.

- **Classical/acoustic music (CD, stereo):** Music pieces which consists of relatively unprocessed recordings of acoustic instruments, in particular certain types of classical, are associated with quite large loudness 'dynamic' range. As an extreme example, consider Ravel's Boléro (Figure 4) which has a Consistency of -11.7 LU – this is less consistent (i.e. greater loudness range) than even the soundtrack of an action movie.

- **Movie soundtrack (DVD, 5.1):** The CoG of an entire film is generally around -26 to -21 LU_{FS}, and the Cons within -12 to -7 LU. These ranges seem to apply to a number of current, popular movies. The Consistency of around -10 LU indicates that the loudness of the soundtrack may go 10 LU *up* and 10 LU *down* from the 'central' loudness – but typically no more than that. This range presumably reflects the assumptions that the producers of the DVDs are prepared to make about the capabilities of the consumer's "home entertainment systems".

- **Broadcast program (Radio/TV broadcast, stereo):** The CoG of a program will naturally reflect the target level used by the broadcaster of the particular radio/TV channel – commonly in the range -24 to -19 LU_{FS}. Broadcast material typically exhibits a quite predictable Consistency, generally around -4 to -2 LU.

- **Commercial (TV broadcast, stereo):** Commercials tend to have a very low loudness range (i.e., a Consistency close to 0 LU), corresponding to an almost constant loudness – the level of which is measured by their CoG. The notoriously annoying transitions between commercials and programs are demonstrated by the fact that the CoG of the commercials are often 3 to 6 LU higher than that of the programs.

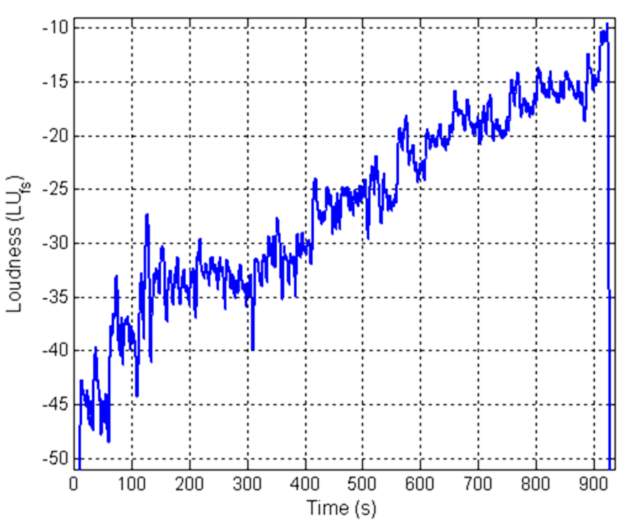

Figure 4. Ravel's Boléro, long-term loudness vs. time. A 15-minute piece of music with an impressive continuous rise in loudness spanning 37 LU.

- **Test signal (synthesized, mono):** A set of signals were synthesized for purposes of testing. They also demonstrate that even mono signals can reach levels *above* 0 LU_{FS}.

5. CONCLUSION

In times where an audio engineer isn't always available in media production, easy-to-use guidelines and tools are required to maintain audio quality across genres. The loudness descriptors presented in this paper level the playing field, and help eliminate the loudness advantage which low dynamic-range content has had in the past.

The proposed set of loudness descriptors could be extended; this paper has described Center of Gravity, and Consistency, and briefly introduced Density. Yet additional descriptors may be relevant to a given application. The scope of the ITU standards concerning loudness measurement and metering [7, 8] does not include descriptors. However, when using the descriptors in connection with exchange of material (e.g. delivery specifications) or establishing guidelines for mastering, it may be in the interest of everyone in the production chain to agree on a common, *open definition* of certain fundamental descriptors. We would be prepared to contribute to such an initiative. As simple as possible, but not simpler.

6. REFERENCES

[1] Jones, S. (2005) "The Big Squeeze -- Mastering engineers debate music's loudness wars", Internet web page: http://mixonline.com/mag/audio_big_squeeze/, Mix.

[2] Nielsen, S.H. & Lund, T. (2003) "Overload in Signal Conversion", in Proc. of the 23rd AES Intl. Conf.

[3] Emmett, J. (2003) "Audio levels - in the new world of digital systems", EBU Technical Review, vol.2003:January.

[4] Moerman, J.P. (2004) "Loudness in TV Sound", in Proc. of the AES 116th Conv.

[5] Lund, T. (2007) "Specifying Audio for HD", in Proc. of the NAB-2007 Convention.

[6] Lund, T. (2007) "Audio Delivery Specification", in Proc. of the AES 123rd Conv.

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

[8] ITU-R (2006) "Rec. ITU-R BS.1771, Requirements for loudness and true-peak indicating meters", International Telecommunications Union.

[9] Skovenborg, E. & Nielsen, S.H. (2007) "Real-Time Visualisation of Loudness Along Different Time Scales", in Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France.

[10] Skovenborg, E. & Nielsen, S.H. (2004) "Evaluation of Different Loudness Models with Music and Speech Material", in Proc. of the AES 117th Convention, San Francisco.

[11] Vickers, E. (2001) "Automatic Long-term Loudness and Dynamics Matching", in Proc. of the 111th AES Convention, New York.

[12] MPEG (2003) "MPEG-7 Overview (version 9)", N5525, ISO/IEC JTC1/SC29/WG11.

Sound segment	**Genre/source**	**Year**	**Duration [h:m:s]**	**CoG [LU_{FS}]**	**Cons [LU]**
Donald Fagan, I.G.Y	80'es pop/rock	1982	6:04	**-19.4**	**-1.8**
Jennifer Warnes, Bird on a Wire	80'es pop/rock	1987	4:40	**-22.4**	**-3.0**
Lou Reed, Dirty Boulevard	80'es pop/rock	1989	3:28	**-18.3**	**-4.1**
Eagles, Hotel California	80'es pop/rock	1990	6:33	**-20.6**	**-6.4**
Ry Cooder, Don't Mess Up a Good Thing	80'es pop/rock	1990	4:02	**-24.2**	**-1.7**
Bonnie Raitt, Come To Me	80'es pop/rock	1991	4:20	**-16.7**	**-2.6**
Michael Jackson, Black or White (incl. Intro)	80'es pop/rock	1991	4:36	**-13.9**	**-5.7**
U2, Mysterious Ways	80'es pop/rock	1991	4:04	**-12.3**	**-2.5**
Lyle Lovett, Family Reserve	80'es pop/rock	1992	3:56	**-19.9**	**-2.9**
Pink Floyd, Shine On You Crazy Diamond	80'es pop/rock	2000	13:20	**-15.9**	**-7.8**
Friends, Episode 9 (TV show)	broadcast	2003	22:00	**-23.4**	**-3.6**
WDR2, Male News, processed	broadcast	2003	5:54	**-18.4**	**-1.9**
WDR2, Male News, unprocessed	broadcast	2003	5:54	**-21.2**	**-3.0**
Klovn, Biggie Black (TV sitcom)	broadcast	2005	25:20	**-21.8**	**-2.4**
Orange County, Episode 2/13 (TV show)	broadcast	2005	41:30	**-23.8**	**-4.1**
Orange County, Episode 2/16 (TV show)	broadcast	2005	41:30	**-23.3**	**-3.8**
Klovn, Kennedy (TV sitcom)	broadcast	2006	23:10	**-22.2**	**-2.6**
BBC News Count In	broadcast	2008	1:35	**-16.6**	**-2.6**
China Olympics, Cycling, DTV	broadcast	2008	34:30	**-24.7**	**-3.6**
China Olympics, Handball, DTV	broadcast	2008	15:57	**-22.3**	**-3.8**
China Olympics, Rowing, DTV	broadcast	2008	9:16	**-24.3**	**-2.6**
Count Basie, Bruce Swedien orig. stereo recording	classical/acoustic	1960	3:55	**-19.7**	**-4.3**
Vivaldi, La Primavera, Allegro, Wiener Phil.	classical/acoustic	1984	3:19	**-20.8**	**-7.8**
Bach, Violin Concerto BWV1041, Andante, Agostini	classical/acoustic	1990	6:16	**-25.8**	**-4.3**
Ravel, Bolero, London Symphony Orchestra	classical/acoustic	1993	15:30	**-20.5**	**-11.7**
Schubert, The Trout, Scherzo, Forellenquintett	classical/acoustic	1994	4:20	**-19.7**	**-6.7**
Beethoven, Cello Sonata Op 17, Rondo, Kliegel & Tichmann	classical/acoustic	2002	4:59	**-18.7**	**-5.9**
Bach, Toccata BVV565 for marimba, Makoto Nakura, 5.1	classical/acoustic	2007	8:50	**-17.6**	**-3.6**
Mediamarkt	commercial	2001	0:20	**-17.4**	**-0.9**
D2	commercial	2002	0:20	**-16.5**	**-0.6**
Noodle	commercial	2002	0:30	**-16.4**	**-0.7**
Head and Shoulders	commercial	2006	0:30	**-20.5**	**-1.7**
Colgate, DTV	commercial	2008	0:30	**-19.1**	**-0.6**
Danish Crown promo, DTV	commercial	2008	0:15	**-17.5**	**-1.0**
Garnier, DTV	commercial	2008	0:20	**-18.5**	**-0.6**
Lisa Nilsson, Himlen runt hörnet	contemp. pop	1999	5:02	**-13.3**	**-3.3**
Tosca, Boss on the Boat	contemp. pop	2000	6:03	**-10.4**	**-2.6**
Anastacia, Don't Stop	contemp. pop	2002	4:21	**-6.0**	**-2.4**
Macy J. Blige, Family Affair	contemp. pop	2002	4:02	**-8.2**	**-1.0**

Sound segment	**Genre/source**	**Year**	**Duration [h:m:s]**	**CoG [LU_{FS}]**	**Cons [LU]**
Kelly Klarkson, Since U Been Gone	contemp. pop	2004	3:09	**-5.8**	**-3.2**
Eric B&Rakim, I Know You Got Soul	contemp. pop	2005	4:45	**-11.0**	**-1.2**
Madonna, Hung Up	contemp. pop	2005	3:21	**-7.0**	**-1.5**
Mariah Carey, It's Like That	contemp. pop	2005	3:33	**-7.7**	**-1.4**
Pussycat Dolls, Don't Cha-02	contemp. pop	2005	4:32	**-8.2**	**-1.7**
Jack Johnson, Losing Keys	contemp. pop	2008	4:48	**-16.4**	**-9.6**
Shining (Digital version, 2001)	movie	1980	1:54:20	**-28.2**	**-8.2**
Jean de Florette	movie	1986	2:01:05	**-26.2**	**-9.0**
Star Wars, Episode 1	movie	1999	2:07:20	**-25.3**	**-7.0**
The Matrix	movie	1999	2:15:50	**-21.8**	**-11.8**
Norah Jones Live in New Orleans (concert)	movie	2002	1:06:30	**-19.3**	**-4.9**
Pirates of The Caribbean, The Curse of the Black Pearl	movie	2003	2:10:30	**-24.3**	**-8.1**
Oliver Twist (Polanski version)	movie	2005	2:09:40	**-24.9**	**-7.3**
Sine FSD, one channel, 997 Hz	test signal	n/a	0:28	**-3.0**	**-0.1**
Sine FSD, one channel, fs/4	test signal	n/a	0:27	**+3.4**	**-0.1**
Sine FSD, one channel, fs/6	test signal	n/a	0:29	**+1.6**	**-0.1**
Sine FSD, one channel, fs/8	test signal	n/a	0:28	**+1.0**	**-0.1**

Table 2. The two loudness descriptors measured for characteristic audio segments of different types and genres. Year refers to the particular release or broadcast that was measured (CDs may be re-mastered and released several times).