REAL-TIME VISUALISATION OF LOUDNESS ALONG DIFFERENT TIME SCALES

Esben Skovenborg

TC Group Research A/S
Sindalsvej 34, DK-8240 Risskov, Denmark
EsbenS@TCElectronic.com

Søren H. Nielsen

TC Group Research A/S
Sindalsvej 34, DK-8240 Risskov, Denmark
SHN@TCElectronic.com

ABSTRACT

We propose a set of design criteria for visualising loudness features of an audio signal, measured along different time scales. A novel real-time loudness meter, based on these criteria, is presented. The meter simultaneously shows short-term loudness, long-term loudness and peak level. The short-term loudness is displayed using a circular bar graph. The meter displays the long-term loudness by means of a circular envelope graph, organized according to an absolute time-scale – looking similar to a radar display. Typically, the loudness measured during the past hour is visible. The algorithms underlying the meter's loudness and peak level measurements take into account recent ITU-R recommendations and research into loudness modelling.

1. INTRODUCTION

Time-varying *features* of an audio signal can be visualised in different ways. Such features can be objective measures or they may represent perceptual properties of the signal. The features discussed in this paper are one of both kinds: 1) the perceptual feature *loudness*, and 2) the objective measure *true peak value*.

In the measurement of the features, the analysis of the audio signal is done over time such that the features are represented by time-varying scalars (i.e. vectors) – one for each feature. The basic time resolution of the analysis should be adapted to perceptually or technically relevant granularity. This aim might be in conflict with the possibilities of a suitable visualisation. The big challenge of visualisation is to present the desired amount of information in such a way that it is easy to comprehend – and without losing access to details.

1.1. Time scales for display

For our application, simultaneous display along three different time scales is desirable. One time scale is instantaneous value, reacting quickly to the measured feature and typically holding the indication of a possible alarm condition (e.g. overload) for a short while, to allow an operator to see it.

A more slowly moving indication is useful to assist an operator in adjusting the sound system, typically the gain. This indication should react and move with a speed similarly to an overall perception of the feature. For example, speech from a trained speaker may be considered to be of constant loudness even though short-term fluctuations occur. The display should reflect this fact.

Finally, a log or history of the fluctuations of a feature may be desirable. Such a log could, for instance, be used to verify that the loudness is aligned appropriately between different segments of a broadcast.

This paper presents a prototype of a novel real-time loudness meter, simultaneously showing short-term loudness, long-term loudness and peak level. The three metering functions have been chosen to fulfil needs in broadcasting, as well as in other production environments where a diversity of program material needs to be aligned in perceived level while also being kept within technical limits. A meter in itself does not align the levels – an operator is (ideally) present to attend to the adjustments, assisted by visual tools like meters.

In our design of the meter display, the analogue clock and radar displays were used as inspiration for the visualisation of magnitude and time dimensions.

1.2. Standardisation

Within the ITU-R (International Telecommunication Unit – Radio Communications Sector), a study group has been working on the methods of loudness and true peak level metering, and recently come up with two new recommendations: [1] and [2]. The former describes the measurement algorithms, whereas [2] describes the visual presentation of the measurements. The need for a short-term as well as a long-term loudness measure is recognized, but with only the long-term measurement method specified at present. Furthermore, methods to reliably estimate the true peak value are described.

The visualisation paradigm presented in this paper is an alternative – or supplement – to the one described in [2].

2. THE NEED FOR LOUDNESS MEASUREMENT

The audio content in the numerous formats in use, is dynamically, spectrally, and sometimes even spatially processed according to the properties of the media, format, and playback conditions, see e.g. [3, 4]. Each format requires different optimum settings of bandwidth, dynamic range etc., based on the expected listening conditions and also on the properties of the available transmission channel or storage medium such as data rate.

These different optimum settings come in addition to the dynamics processing needed to reduce undesired loudness variations. Therefore, a *loudness* meter is required as a complement to traditional *level* metering.

In many cases, a fully automatic way of setting processing parameters according to the different requirements would be desirable. This goal may not be trivial to achieve, but in all cases a monitoring function is needed: A meter which can display the relevant perceptual properties, i.e. the short-term and the long-term loudness. Furthermore, a function to monitor the measured peak level is required as an aid to avoid clipping. Such a technical measure is required, in addition to the perceptual measures (of loudness), due to the limitations of the transmission channel.

2.1. Previous level meters: VU and PPM

Traditionally, the primary purpose of level meters has been of technical nature: They serve as an aid in fulfilling certain technical criteria, such as obtaining a good signal-to-noise ratio on an analogue medium. Here, the standardised VU- and PPM-types of level meters are discussed [5, 6, 7].

The VU (volume unit) meter [5] measures the full-wave rectified (i.e., absolute value) level with a relatively slow time constant. The response time of the meter to rising and falling levels is (ideally) identical. For judging the overall level the VU meter can be quite useful, but due to its measurement algorithm the VU meter is not suitable for loudness measurement. With the soft saturation characteristics of analogue tape recordings in mind, the VU meter has been successfully used for years to set the right recording level – often supplemented by a peak-indicating lamp, as the meter is too slow to react on short transients. The scale of the VU meter is shown in Figure 1. A mechanical instrument with a thin indicator needle is typically employed. Note the contrast in the meter, and the curved scale with approximately linear voltage scale and thus non-linear intervals on the dB scale. The overload section (above 100%) is coloured red.

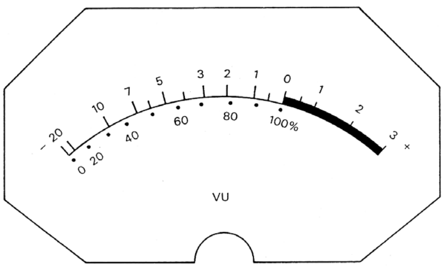

Figure 1: *VU meter scale from [5].*

For more precise control of peak levels, such as needed in radio and TV broadcast for technical and legal reasons, another type of meter was created: the peak programme meter (PPM) [6], [7]. Actually, two generations of PPMs exist: One with instantaneous response to rising levels and another with a short response time (a few milliseconds). The measurement algorithm consists of taking the peak value of the full-wave rectified signal. The decay time is chosen to be long enough that an operator may notice even brief peaks – yet not be disturbed by meter flickering. Very short peaks, which may cause problems in *digital* transmission and storage systems, are underestimated in the original PPM due to the response time, so for peak measurements in the digital domain, the peak sample value is measured [7].

Although the PPM was not designed for – and not really suitable for – loudness measurement and alignment, some rules can be made to help an operator use the PPM for that purpose anyway [8], sect. 5.2. A major disadvantage of these rules is that they require knowledge of the actual type or genre of the source material. The standards describe different appropriate display scales, their contrast, brightness, colour etc. The human factor is taken into account in the specifications for decay time and peak-hold time. Meter scales for both mechanical and opto-electronic displays have been specified. Figure 2 shows one of the scales for the mechanical display; note the linear dB-scale.

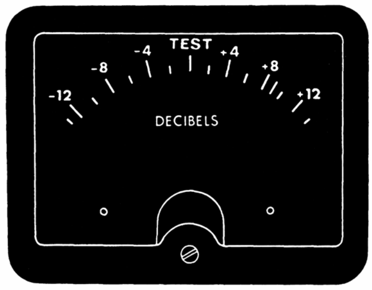

Figure 2: *PPM scale from [6].*

The opto-electronic display, as shown in Figure 3, features a non-linear correspondence between length of the bar and dB, but different from the VU meter with its linear voltage scale. Instead the bar-type of PPM takes advantage of the digital technology and adapts the scale graduation to the needs of the users, by providing a fine resolution at high levels *and* a large dynamic range. A minimum of 100 segments are specified for the bar-type instrument in order to give a smoothly changing length of the bar.

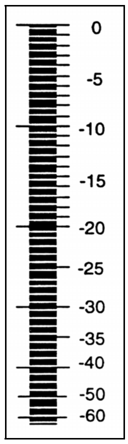

Figure 3: *Bar-type PPM scale from [7].*

The displays of traditional level meters, as described above, were determined by a mixture of technical and human factors, some of which have a scientific basis, whereas others are based on experience from their application domain. Although none of these meters are particularly suitable for measuring loudness, parts of their display properties have been applied to our presented loudness meter and associated display described in section 3.

2.2. Loudness meter standardisation within the ITU

Measurement algorithms and display requirements for loudness and true peak level meters have recently been described in the ITU recommendations [1] and [2]. Although the algorithms specified may not be the best ones available they have now been standardised so that new meters, providing a better estimate of the perceived loudness than a VU or PPM meter, can be made. In fact, the loudness measurement specified in [1] is not really measuring loudness, but rather an estimate of the *gain offset* required to match the loudness of one sound clip to that of a reference sound.

Due to the non-linear aspects of hearing, this gain offset and the corresponding change in loudness can differ. This issue is recognised in the recommendation. For operational purposes, however, the gain offset can be quite useful, as the operator has gain adjustment tools readily available.

The loudness measurement algorithm consists of a frequency-weighted RMS value, developed with measurement time intervals in the order of 10-30 seconds, i.e. an *Leq* (equivalent level). In the case of a multi-channel input, a single loudness value is computed based on a weighted energy-sum. A measurement period of 10s of seconds is long-term rather than short-term, and not directly suitable for a real time meter. But as no reference data for continuously varying loudness matching was available, the accuracy of a short-term measurement could not be tested.

Three different displays have been specified in [2], one of which is depicted in Figure 4. Compared to the VU and PPM meters notice the relatively few segments and the linear scale in LU (Loudness Units – equivalent to dB). Furthermore, the range of the scale is rather small – which is in accordance with the primary purpose of the meter as an aid in aligning loudness, and not a general-purpose level meter.

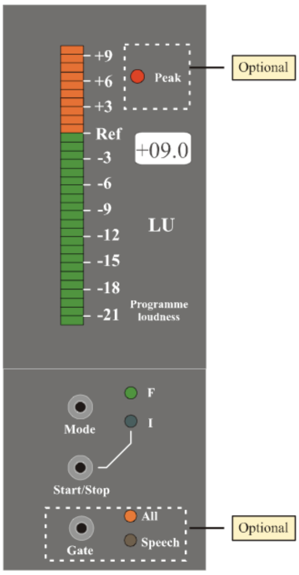

Figure 4: *Loudness meter display, according to [2].*

Having followed the ITU process of creating these two recommendations closely since 2003, we would like to make a few remarks: The contents of recommendations are not always that of the most solid science – but rather a combination of science, personal (or institutional/company) interests and political negotiations. On the positive side, [1] contains an opening for improved loudness measurement algorithms which are not as easily fooled as the simple frequency-weighted energy measure. Unfortunately, the statistics describing the results are insufficient to enable a scientifically valid comparison between the recommended method and alternative methods. Information on the statistical uncertainty, the variability of the listeners etc. are missing [9, 10]. Furthermore, the standard [1] calls for a short-term loudness measurement algorithm – which is certainly needed for real-time metering (and control).

3. A NOVEL LOUDNESS METER

In the loudness meter that we present here, we have employed successful visualisation principles of previous level meters, in combination with a new type of visualisation of long-term loudness history. The meter displays the short-term loudness, and the long-term loudness over a period of time, together with signal and overload indications. Our design of the meter's measurement algorithms and display has evolved as a mixture between science, intuition and empirical experience.

3.1. Measurement Algorithms

3.1.1. Loudness measurement

Loudness, as such, is a perceptual property of sound but can be modelled using different algorithms – and can thus be measured as an objective property of the sound.

Research into psychoacoustic models of loudness perception has been taking place for decades. Most prominently, Zwicker's loudness model has been standardised as ISO-532B [11]; however these models were developed for measuring loudness of sounds with stationary properties, such as noise and tones, and are thus unsuitable for meter applications [10]. More recently, Glasberg and Moore have presented research on modelling time-varying loudness of certain classes of signals [12, 13]. For loudness meter applications, simplified measurement algorithms have been developed, e.g. [14].

The present meter measures the loudness of the input signal by means of a simple model of loudness perception, but does not require any particular loudness model. For multi-channel input signals, a single loudness measurement is computed, combining the contribution from each channel.

The loudness meter uses the measurement unit of LU (Loudness Units). The LU is a measurement in dB, with 0 LU corresponding to a reference loudness level. The reference loudness level, and the acceptable range of fluctuation around it, might depend on the policy concerning the particular broadcast channel that the meter is monitoring.

In our meter prototype, the *TC LARM* algorithm [10] was employed, although other algorithms could alternatively be used. The accuracy of the TC LARM algorithm was evaluated in [10], against a set of subjective reference data, using a wide selection of speech and music material. Compared to the weighted Leq measurement algorithm in [1], the accuracy of TC LARM was found to be at least as good.

The short-term and long-term loudness measurements that the meter displays, use the same underlying measurement algorithm. However, the length of the analysis window and the visualisation of the measurements, differ for the short-term and long-term loudness. The temporal properties of the short-term loudness measurement were developed with several practical criteria in mind. For example: How much does the short-term loudness drop, in the 'silent' periods that are present in normal speech? How much does the short-term loudness rise, at the snare drum beat in pop/rock music? Similarly, the long-term measurement was developed to provide a constant measurement – within plus/minus a couple of LU – for material with an overall constant perceived loudness (see also section 3.2.4). In the current implementation of our loudness meter prototype, we used the following lengths of the analysis windows: Short-term loudness: 0.5 s; Long-term loudness: 2.5 s.

3.1.2. Peak level

It is well established [15, 1] that the true peak value of a digital signal may be significantly above the magnitude of the actual samples. Especially signals that have been clipped or otherwise processed nonlinearly exhibit this property. When staying within the digital domain, and performing no subsequent processing, this will pose no problems (except from the distortion inherent in the non-linear processing). When changing domain or sample rate, however, the true – and possibly higher – peak value can appear in the new domain. That way, overload and additional audible distortion can result.

The peaks not directly represented by the samples can be estimated easily and accurately by using an interpolation technique, as used in oversampling and D-to-A conversion [1]. A short (FIR) lowpass filter near the Nyquist frequency and an interpolation factor of 4-8 will yield good estimates of the true peak value. Such a technique is employed in the presented meter.

In our loudness meter prototype, only the essential overload- and "signal present"-indications are displayed, for each input channel.

3.2. Visualisation of Measurements

3.2.1. Short-term loudness display

Figure 5 shows the display of the developed loudness meter, in a greyscale version (for better printing). In the loudness meter, a circular bar graph displays the short-term loudness of the input signal, as measured by the underlying loudness algorithm. This bar graph corresponds to the 'current' loudness that a listener would perceive. If the sound has a fairly constant loudness, the bar graph remains fairly constant (as opposed to traditional meters, that might change as a function of signal *amplitude* alone). The acceptable region of loudness is colour-coded: If the short-term loudness of a programme segment remains within the green region then the operator can easily determine that the material was neither too loud nor too soft.

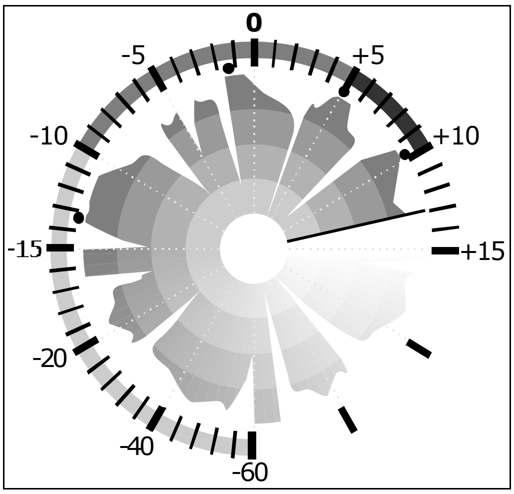

Figure 5: *A greyscale version of the developed loudness meter display.*

The range of the short-term loudness scale is larger than in the ITU meter, cf. Figure 4. The upper region from -20 to +15 LU use a linear scale, similar to the ITU recommendation. The scale tics in the upper region corresponds to the minutes on an analogue clock, and depending on the technology used, a display resolution of 0.1-0.5 LU is achievable. This resolution enables a smoothly changing display without perceivable flicker. The lower-level region of the display can be useful as a more detailed "signal present" indication, like in the PPM bar-type display (Figure 3).

Several properties of the short-term loudness display are in accord with knowledge about human visual perception, e.g., chapters 6 and 8 in [16]. One such property is *redundancy*, which increases the robustness of the readings. The short-term loudness is signalled not only by a position (of the end of the arc/curved bar) but also the size of the arc, i.e. the angle covered. Furthermore, the end of the arc changes angle according to the present loudness. As the eye is more sensitive to *angular movement* than to 'linear' movement this increases the readability of the display. Finally, *colours* are used to code relevant regions. Together, these properties also help ensuring that the short-term loudness is evident, even if a human operator was presented with several simultaneous displays, or if reading the meter from a distance.

3.2.2. Long-term loudness display

The long-term loudness is displayed in the centre of the loudness meter, by means of a *circular envelope graph*. The envelope graph is organized according to an absolute time-scale, similar to familiar analogue clocks. Thereby, the long-term loudness of the input signal during the past hour is displayed at all times. In Figure 5, the time is 9:13, hence the current long-term loudness is being displayed at the "13 minutes past the hour" position. The 'older' the long-term loudness entry is, the more it is faded into the background (white in Figure 5, black in Figure 6). Thus, the long-term loudness display appears like a "radar display" which is scanning clock-wise.

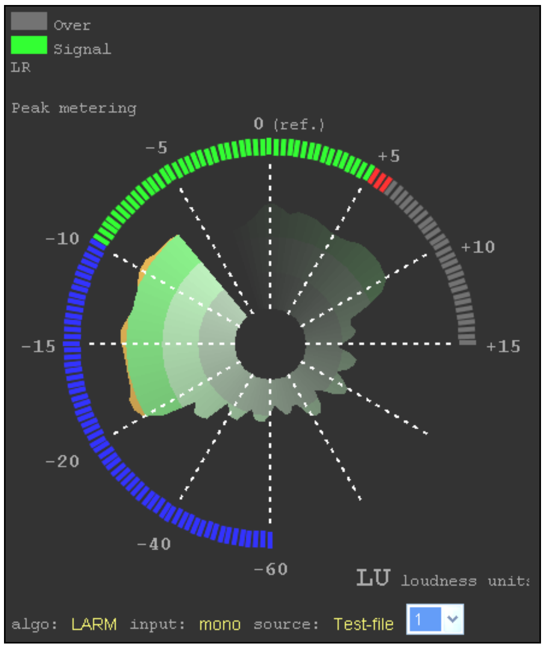

Figure 6: *A colour version of the developed loudness meter display, with a mono input signal.*

The further away from the centre the long-term loudness graph is, the louder the sound was at that time. The different colour regions correspond to “below reference loudness”, …, “above reference loudness” (Figure 6).

As an extra feature, the displayed long-term loudness could cover any period of time, from the past minute (time is zoomed in) to the past 12 hours (time is zoomed all the way out). Furthermore, a single loudness meter could record or log the *loudness history* of several simultaneous sources, which the operator could then switch between while monitoring the sources.

We have implemented the loudness meter as a virtual prototype (software). Figure 6 shows the display of the prototype.

3.2.3. *Signal level and Overload display*

The loudness meter’s signal level indicator consists of a column of LED-type components for each audio channel of the input source. The red indicator is turned on, for a short period, when signal overload is detected. Typically overload is indicated when the level is close to or above 0 dBFS, but lower levels could alternatively be used, depending on policy concerning the particular source. The green indicator is turned on when the signal level is above (say) -50 dBFS, to indicate that a signal is present on the corresponding channel – conversely to indicate a signal drop-out.

In addition, the signal overload indicator could display, using a coloured dot, any signal overloads (on any channel) that have occurred during the past hour (Figure 5). These dots are displayed along the circumference of the long-term loudness display, where their locations are used to indicate that certain events occurred at the corresponding time. Different colours could even be used to indicate (other) technical problems that happened in the past, such as signal drop out, or loss of clock sync for a digital input.

3.2.4. *Example*

Although a loudness meter is inherently an instrument to be used in real-time, for purposes of illustration the loudness measurement data can be extracted and plotted. The three graphs in Figure 7 show the short-term and long-term loudness measurements and the signal amplitude, as a function of time, using a test signal as input to the meter. This demonstrates how (much) the two loudness measures fluctuate for different types of audio material.

A test signal was made up of two audio segments, each of 15 seconds duration, representing characteristic signals: speech and pop music (Table 1). Whereas the speech signal is a fairly ‘dry’ recording, the pop track has undoubtedly been processed with dynamics compression and other mastering techniques. Each of the segments were ‘level-normalised’ individually, i.e. scaled to peak at 0 dBFS.

In test signal	CD	Track	Start time	End time
0-15 s	EBU SQAM: Sound Quality Assessment Material	#49: Speech - Female, English	0:00	0:15
15-30 s	Madonna: Confessions On A Dancefloor	#1: Hung Up	0:45	1:00

Table 1: *The contents of the test signal, displayed in* Figure 7.

The *short-term loudness* graph reveals that the speech segment consists of two spoken sentences, but also that the speaker achieves a fairly constant loudness, with variations within +/− 5 LU. Even though both segments were peak-normalised, the *long-term loudness* graph shows that the pop segment is consistently louder that the speech segment. In fact, the pop segment manages to stay 6-8 LU above the reference loudness level, with virtually no variation.

Imagine these two segments had been broadcast on the radio, with no further processing, while the loudness meter was ‘listening’ to the signal. The meter’s long-term loudness history could then suggest that the pop segment should have been attenuated by say 5 dB, in order to spare the listeners for a noticeable increase in loudness. Lund describes the loudness meter’s application in the context of broadcast for Digital TV and other media [17].

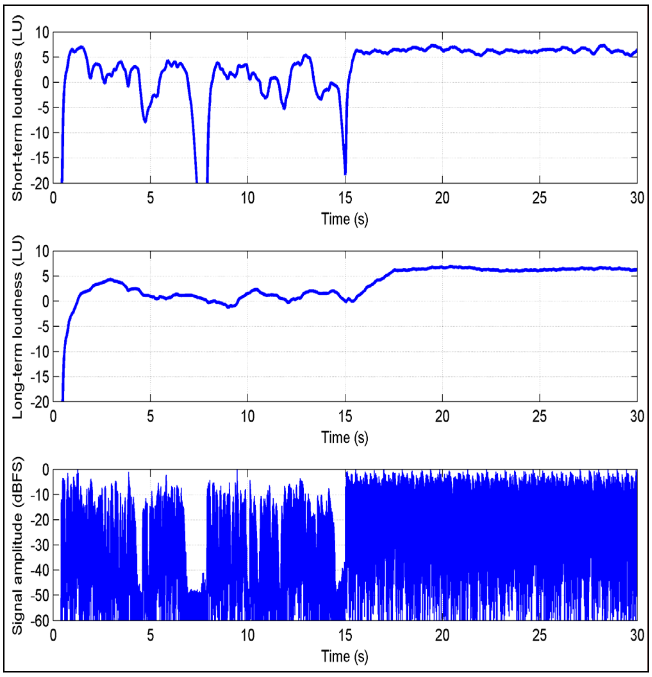

Figure 7: *The short-term loudness, long-term loudness and amplitude of a test signal consisting of 2 segments. First 15 sec: Female speech; Last 15 sec: Madonna - Hung Up.*

4. TOWARDS EVALUATION

The effectiveness of any loudness meter will depend on both the graphical appearance and dynamic behaviour of its display, as well as on its underlying measurement algorithms. All of these factors must be taken into account, when assessing the meter's overall quality and usability.

Formal evaluation of a visualisation system, such as the one described in this paper, is challenging: First of all, one or more metrics must be defined by which the display should be evaluated. The correspondence between the sound heard and the picture seen is one aspect to be evaluated. Another metric could characterise the speed of reading the meter reliably. A very high-contrast and flashing meter could cause eye fatigue (even though such a display might be immediately more readable). The usefulness of having several types of loudness measurements available at one glance may be hard to measure directly, but has to do with the compactness of the display, which again determines where the display can fit in the application, and how many independent sources can be displayed on individual meters, in a given workspace.

So far our loudness meter has only been verified informally – we have received positive response from potential users in different application areas. No systematic experimental assessment of the proposed meter has been attempted yet.

4.1. An example of an evaluation method

A method for designing and evaluating a short-term loudness meter has been proposed in [18]. The purpose of that study was to address the need expressed by the ITU in [1]. At least two challenges are described in [18]: First, a way of creating a continuously varying measure of the perceived loudness must be found – i.e., a set of reference data for evaluating the measurement by the meter. Second, these time-varying reference data must be compared to the meter display. As a result of the evaluation, the technical parameters of the meter (or its measurement algorithm) can be set to appropriate values.

Rather than tracking the time-varying loudness itself, the task of the subjects in [18] was to continuously adjust a gain control to keep the loudness constant, that is, a continuously varying gain correction factor was registered. A couple of difficulties using this method were found: 1) When subjects adjusted the gain they tended to overshoot a bit. This must be taken into account when analysing the data. 2) The subjects tended to drift in their loudness reference. This means that their gain correction factor for *identical* sound segments changed over time.

To evaluate variations of a short-term loudness meter, the output of the ITU loudness measurement algorithm using different lengths of the analysis time window was plotted against the subjective gain adjustment data. The evaluation consisted of a visual inspection of these plots, and based on that a time window of 3 seconds was chosen as optimum for a “short-time loudness” measurement.

4.2. A proposed evaluation method

A loudness meter with its underlying measurement algorithms and display methods contains many parameters – more than could easily be adjusted in a traditional adjustment-type of experiment. Furthermore, the task of evaluating the complete meter in an experiment would require a considerable amount of time, as the inclusion of a signal history depends on listening to several sound segments within a single session.

One way to overcome the difficulties of performing a multi-parameter adjustment experiment would be to present a number of different *complete* meters, with pre-set variations of the underlying algorithms and their parameters, and maybe even display types. The task of the test subjects would in that case be to *rate* the different meters according to specific criteria (as the metrics mentioned above), as well as the subjective overall impression.

5. CONCLUSION

We have proposed some design criteria for visualising loudness features of an audio signal, measured along different time scales. We then presented a novel loudness meter, simultaneously showing three time-varying features of an audio signal: short-term loudness, long-term loudness history and a overload indicator. Our meter displays the short-term loudness using a circular bar graph. The long-term loudness is displayed by means of a circular envelope graph, organized according to an absolute time-scale – looking similar to a radar display. The presented real-time loudness meter thus provides one complete solution to the requirements for an effective loudness meter. The algorithms underlying the meter prototype's loudness and peak-level measurement take into account recent ITU-R recommendations and research into loudness modelling. Finally, different aspects of evaluating a loudness meter were discussed.

6. REFERENCES

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

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

[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] IEC (1990) "IEC 268-17. Sound system equipment - Part 17: Standard volume indicators", International Electrotechnical Commission.

[6] IEC (1991) "IEC 268-10. Sound system equipment - Part10: Peak programme level meters", International Electrotechnical Commission.

[7] IEC (1995) "IEC 268-18. Sound system equipment - Part 18: Peak programme level meters - Digital audio peak level meter", International Electrotechnical Commission.

[8] Dickreiter, M. (1987) "Handbuch der Tonstudiotechnik, Band 1" (5. Auflage. ed.), München: K. G. Saur.

[9] Skovenborg, E., Quesnel, R. & Nielsen, S.H. (2004) "Loudness Assessment of Music and Speech", in Proc. of the AES 116th Convention, Berlin.

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

[12] Glasberg, B.R. & Moore, B.C.J. (2002) "A Model of Loudness Applicable to Time-Varying Sounds", Journal of the Audio Engineering Society, vol.50:5, pp.331-342.

[13] Moore, B.C.J., Glasberg, B.R. & Stone, M.A. (2003) "Why Are Commercials so Loud? -- Perception and Modeling of the Loudness of Amplitude-Compressed Speech", Journal of the Audio Engineering Society, vol.51:12, pp.1123-1132.

[14] Jones, B.L. & Torick, E.L. (1982) "A New Loudness Indicator for Use in Broadcasting", in Proc. of the 71st AES Convention, Montreux.

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

[16] Goldstein, E.B. (1989) "Sensation and Perception" (3rd. ed.), Belmont: Wadsworth Publishing Company.

[17] Lund, T. (2006) "Control of Loudness in Digital TV", in Proc. of the NAB-2006 Convention.

[18] Soulodre, G. & Lavoie, M.C. (2006) "Development and Evaluation of Short-Term Loudness Meters", in Proc of the 121st AES Conv.