REPORT ITU-R BS.2103

Short-term loudness metering

(Question ITU-R 2/6)

(2007)

CONTENTS

Page

1 Introduction

Radiocommunication Study Group 6 has approved a number of new Recommendations for international standardization of loudness meters. This has occurred in response to a need by many member administrations for such an instrument or set of instruments. This need has arisen because of a number of recent developments in broadcasting:

1.1 The transition to digital audio production

Digital audio production has expanded the dynamic range available to programme-makers. Where they once had a dynamic range of 60-70 dB with up to 3% distortion, they now have a minimum of 96 dB dynamic range with <0.01% distortion using uncompressed digital encoding. This has meant that headroom has been expanded from 8-9 dB up to 18-20 dB from normal recording levels. It has also meant that noise floors are now effectively inaudible under most listening conditions. This has given producers great scope to produce more dynamic material. There is a variation in domestic listening environments and reproduction systems, with some sites having both high noise level and limited dynamic range. A short-term loudness meter may be useful in programme production in order to provide information on the short-term loudness fluctuations in an audio programme.

1.2 ii) The transition to digital audio emission

As well as digital production media, broadcasters are now able to send programmes to the audience in digital form. While this generally occurs via some form of lossy codec, the dynamic range and perceived distortion can be almost as good as the original audio if care is taken in selecting codec parameters and maintaining an adequate bit rate. Some emission systems include metadata such as dialnorm, which indicates the level of typical programme dialogue, thus enabling intelligent level control at the receiver/decoder. For this control to work correctly, the metadata must be correctly chosen and encoded at the production or emission stage of the chain.

1.3 The proliferation of the number of available broadcast audio services

The advent of cable emission and digital emission, in particular internet emission, has increased the number of channels available to broadcasters to the point where they are effectively unlimited. This has led to a corresponding increase in the number of services provided by broadcasters, often with no increase in production resources and minimal increase in emission resources. More channels mean more chance of variability in levels. The increase in demand for programme material that results from increased channel capacity has led to greater reuse of existing programme material, and greater outsourcing of programme material, often with less regard for quality than for quantity.

1.4 Increasing reliance on automation and decreasing reliance on trained operators, particularly at the emission stage

The proliferation of services has meant that existing personnel are responsible for more programme material and more services. In many cases, it has also created an expansion in the need for operators and producers, some of whom are employed without as much training as in the past. While many production skills are learned on the job anyway, skilled mentors are needed to ensure that good practice is maintained. These are not always available where they are needed, so production quality sometimes suffers. One solution for this is to employ automation systems to control audio levels. While these can help, they are a simplistic solution to a problem that is often complex.

1.5 Changing practices in cinema sound production

Many of the same factors have affected soundtrack production practices in cinema. While cinema and broadcasting have significant differences in their approach to sound production, they also have much in common at a technical level. They also have common commercial interests, as most cinema productions become broadcast programmes. Loudness monitoring and control techniques developed for cinema sound are therefore of considerable interest and importance to broadcasters.

These factors have led broadcasters to look for ways to better measure and control the loudness of programmes.

While loudness meters are not a new idea, and a number have been available commercially for some time, there has been some difficulty finding a standard design that suits the needs of broadcasters.

A number of loudness meters are available as options on high-end sound level meters. These are mostly based on ISO532B [1], the Zwicker loudness assessment method. While these are undoubtedly more accurate than a VU meter for assessing loudness, they are also considerably more expensive, large, fragile and often complex to use and interpret. Broadcasters need an instrument that is small, cheap, robust, simple to use, suited to continuous monitoring and easy to interpret, even for unskilled or semi-skilled operators.

A number of proprietary loudness meters aimed at audio production are also available. While some of these may be more accurate than a VU meter for assessing loudness, no two brands are the same, with the consequence that readings taken on different brands are not consistent. This complicates the task of exchanging programme between broadcasters or between countries considerably.

There was a general consensus among members of Radiocommunication Study Group 6 that to meet these requirements, a new meter specification should be developed.

2 Existing metering – its strengths and weaknesses

Broadcasters have historically used two standardized metering systems for assessing signal levels, volume unit meters or VUs and pseudo-peak meters or PPMs. They embody two different approaches to audio level measurement. These approaches are partly products of available or affordable technology, and partly products of different priorities in signal measurement.

Broadcasters have two main goals in setting signal levels:

a) Maximizing the signal level to maximize analogue modulation at the transmitter without overloading it. This maximizes the signal-to-noise ratio, an important parameter in signal quality. This in turn maximizes the physical audience reach for a given transmitter power. (Trying to engage the audience by producing suitable programmes is an equally important but totally different problem.) Overload at the transmitter both distorts the signal, which is unpleasant but not catastrophic, but also generates interference at other radio frequencies, which has more serious consequences.

b) Matching loudness between sources within a programme, matching loudness between programmes on the same station and matching loudness between stations (in some cases of inter-station rivalry, this degenerates into an attempt to be louder than the other stations rather than equally loud).

While these goals are largely compatible, traditional level meters do not directly indicate both the loudness and the peak level of a signal, so broadcasters have opted for one or the other.

The VU meter was standardized around 1940 [2] following the initial 10-15 year phase when the broadcasting industry became established. It consists of a solid-state bridge rectifier and a high-quality moving-coil d.c. ammeter with a defined mechanical time constant and damping characteristic (Table 3). Its greatest virtue was that it was passive, requiring no expensive, unreliable valves to operate. The time constant and damping characteristic were chosen partly because of mechanical constraints in moving-coil meter design and partly for psychoacoustic reasons. The original development work indicated that, while the meter reading gave a reasonable correlation with perceived loudness, it also gave a reasonable basis for inference of signal peak level on speech (the latter probably has more to do with the characteristics of speech than of the meter). The VU meter is used extensively in many countries.

In Europe a different approach was used. Measurement of the peak signal level was considered a higher priority than measuring the loudness, so an active system was developed that sampled and held signal peaks – the PPM [7]. This did not show true signal peaks but instead averaged over a 5-10 ms window as this was considered to be the shortest audible overload duration. As with the VU meter, knowledge of the signal characteristics allowed inferences about other signal parameters to be drawn from the measured parameter, so a skilled operator could also use the PPM for loudness balancing. Loudness balancing could also, of course, be done by ear. The PPM is used extensively in Europe and has progressively replaced the VU meter in many other countries.

The advent of digital recording and digital broadcasting has brought greater dynamic range to broadcasting. This has allowed operating levels to be reduced, expanding available headroom and effectively eliminating the possibility of signal overload in all but the most extreme circumstances.

Digital broadcasting has also brought modulation systems that cannot over modulate the carrier, so even if signal overload does occur, there is no possibility of interference with other services. Both of these factors have diminished the importance of peak signal level for broadcasters.

While the VU meter is a good measure of loudness for some signals (principally those with a low crest factor) it is not always a good measure of it. The VU meter is calibrated to show the r.m.s. value of a sine wave, but it actually measures the rectified average value of a signal, not its true r.m.s. value. Loudness is closely related to the true r.m.s. value of a signal, not to its rectified average value. For signals with low crest factor, these two properties are very similar. For impulsive signals and other signals with large crest factor, the r.m.s. and rectified average value can differ substantially. This is offset to some extent by the practice of reading the peak needle swings on a VU meter, while loudness is read as the *average* of the needle swing on an r.m.s-type meter. Even so, the difference between VU reading and actual loudness can be substantial on some signals. So while a VU meter is a better measure of loudness than a PPM, it is still inaccurate on many signals. This inaccuracy is large enough to cause a substantial number of audience complaints in some circumstances, hence the need for a more accurate measure of loudness.

3 Application specific loudness meters

Loudness meters are needed for several applications, including:

– For emission, as a quality-control check.

– To confirm that loudness is within acceptable limits.

– For production, as a level meter to replace the VU meter.

In the quality control, a slow (many seconds) averaging meter is needed so that the average programme loudness can be read easily. A numerical readout is an advantage in these applications as it removes any need to interpret needle swings and perform any visual average estimation. A numerical readout also allows corrective gain to be easily determined.

In the production role, particularly if the loudness of complete programme mix is displayed by means of a single meter readout, it may be useful to have some indication of the perceived dynamic range as well as having an indicator fast enough to display individual loudnesses where these can be discriminated aurally. This gives rapid feedback to the programme maker on which to make decisions about individual levels in the mix or overall mix level. This can be particularly important in productions "on location", where for various reasons, audio monitoring may not be optimal.

For these reasons, a moving-indicator with a short (less than one second) averaging time is needed in production for loudness monitoring. Ideally, such a meter should show instantaneous loudness.

4 Human loudness perception

While loudness models for arbitrary sound types and for arbitrary sound levels are very complicated, ITU-R work has shown that for typical broadcast material and a limited range of loudness, a simplified model can give surprisingly good correlation with perceived average loudness over a 10-15 s interval. There is, so far, no evidence to show that this model cannot also indicate instantaneous loudness under the same constraints. In fact, it is essential to use the same model for instantaneous loudness assessment if results are to be obtained which are consistent with long-term average loudness measurements. The model is detailed in Recommendation ITU-R BS.1770.

The main concern in this Report is adapting this model to indicate instantaneous loudness.

Human sensory perception follows Stevens' Power Law [Stevens, 1957]. The general form of the law is:

$$\psi(I) = kI^{a}$$

where:

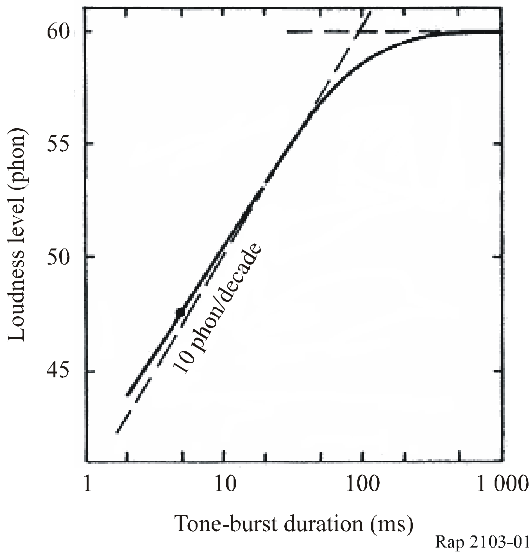
- I: magnitude of the physical stimulus
- ψ: psychophysical function capturing sensation (the subjective size of the stimulus)
- a: exponent that depends on the type of stimulation
- k: proportionality constant that depends on the type of stimulation and the units used.

This law is a consequence of the fact that biological processes, and perception in particular, can usually be modelled by first-order differential equations. These models generally apply to both the magnitude of sensation and to the onset and disappearance of the sensation. In other words, they generally apply in both amplitude and time.

In terms of temporal models of sensation, this means that sensations do not suddenly appear and disappear, but rather the sensation becomes gradually apparent when the stimulus is presented and the sensation gradually disappears after the stimulus is removed. This is illustrated by Zwicker's data on temporal loudness effects for a 2 kHz toneburst stimulus (Fig. 1) which indicate the relative loudness of a toneburst as a function of its duration. This can also be interpreted as the onset of loudness perception at the start of a continuous tone.

FIGURE 1

Zwicker's data on temporal loudness sensation



5 Review of proposed ballistic types

First order differential equations have solutions for impulse and step stimuli which are simple exponential functions.

Modelling temporal loudness perception can therefore be done with first-order analogue filters which have simple exponential rise and fall characteristics for such stimuli.

This type of filtering can be replicated in the digital domain using a first-order IIR topology or an FIR topology, although the IIR topology is far more computationally efficient. The FIR equivalent has an exponentially falling set of sample weights, corresponding to the impulse response of the first order analogue filter.

It has been suggested in discussions on loudness metering that the long-term loudness algorithm, which uses FIR filtering with rectangular weighting, could be adapted to model instantaneous loudness perception by simply shortening the time window over which it is applied. This would provide a poor model of loudness perception, as illustrated in Figs. 2 and 3.

Figure 2 shows the leading-edge step response of an FIR filter with rectangular weighting and a time window of 100 ms (the best fit for Zwicker's data), compared with a first-order IIR filter with the same time constant. The IIR filter performance follows Zwicker's data within 0.1 dB, while the rectangular FIR filter has a worst-case 2 dB error at 0.1 second. The agreement between the two is otherwise good in the early and late phases.

Figure 3 shows a similar comparison for a step response at the trailing edge. Here the disagreement is more marked. While the IIR filter falls off gradually and continuously, the rectangular FIR filter falls gradually at first and then drops precipitously at $T = 100$ ms. This FIR filter is unlikely to be a good model of perception. It gives a very discontinuous representation which does not occur in actual sensation. In other words, it is "jumpy" or "jerky".

Since human perception generally follows Stevens' Power Law, and this law represents the behaviour of biological systems which can generally be modelled by first-order differential equations, and since first-order IIR filters model such equations well, and since rectangular-weighted FIR filters do not model such equations well, we should use a first-order IIR filter in any loudness meter that attempts to model instantaneous loudness perception.

FIGURE 2

Leading edge step response – comparison of two filter topologies

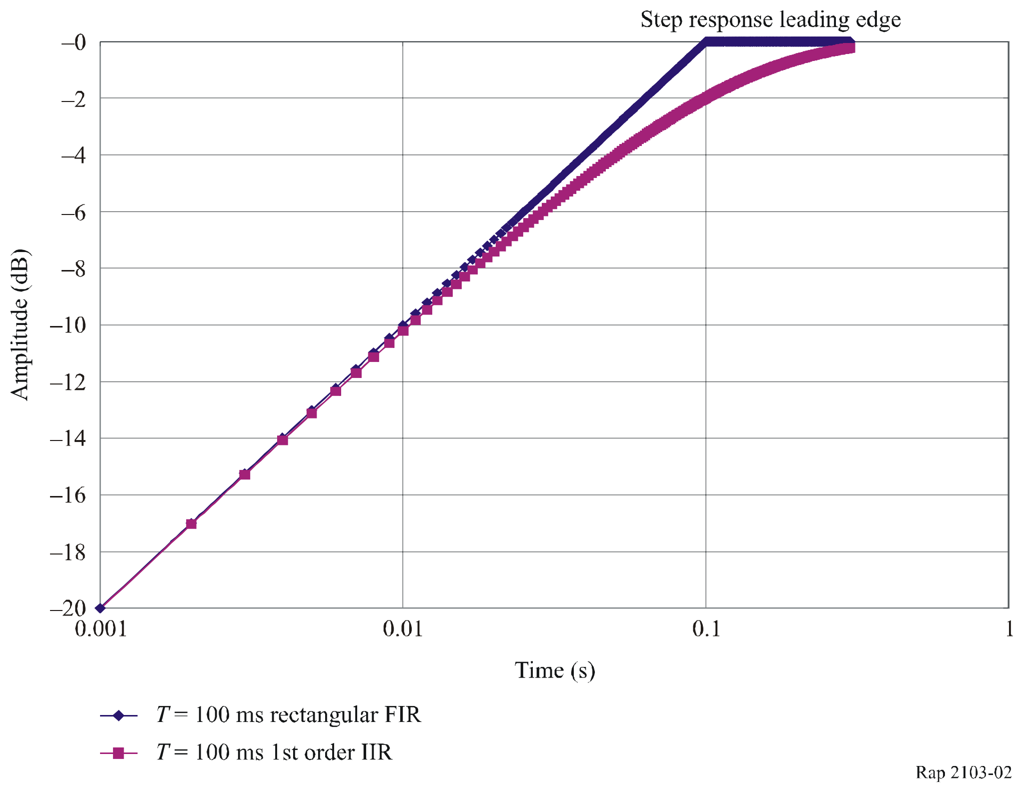
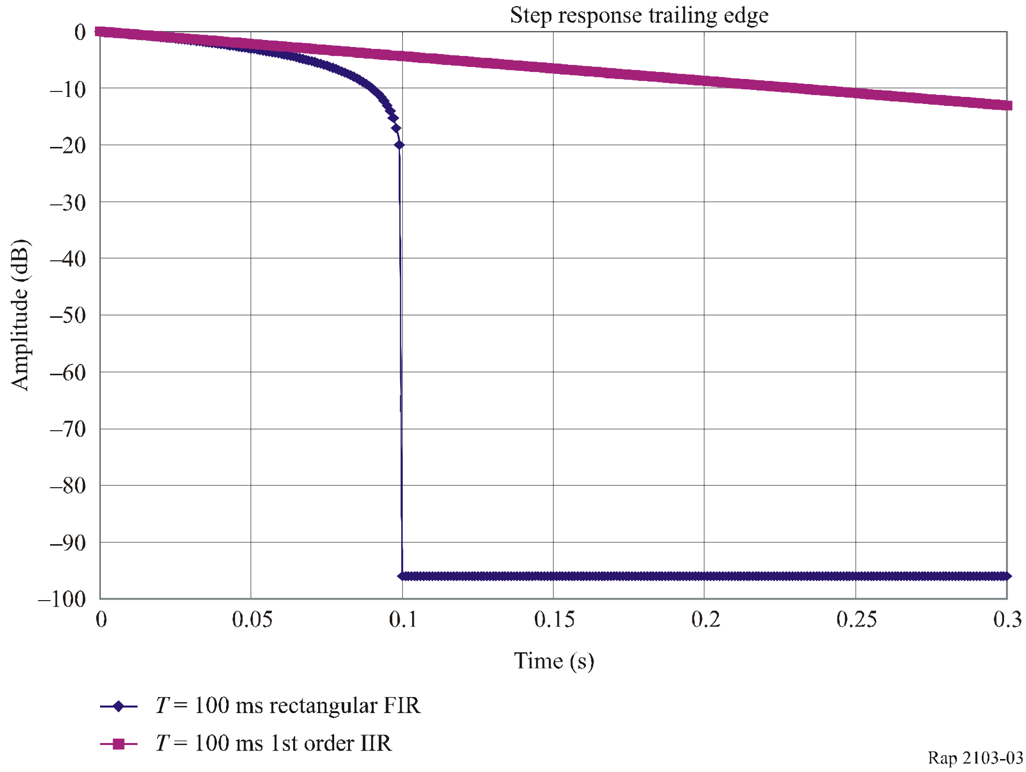


FIGURE 3

Trailing edge step response – comparison of two filter topologies



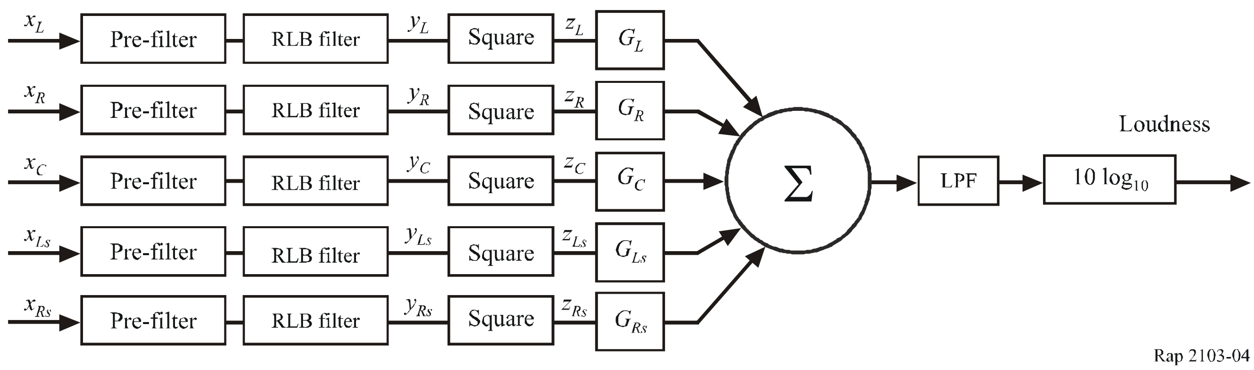
6 Implementation – theory

6.1 Algorithm topology

For consistency with the long-term loudness meter, the implementation of the short-term loudness meter followed the block diagram in Fig. 4.

FIGURE 4

Modified loudness algorithm block diagram

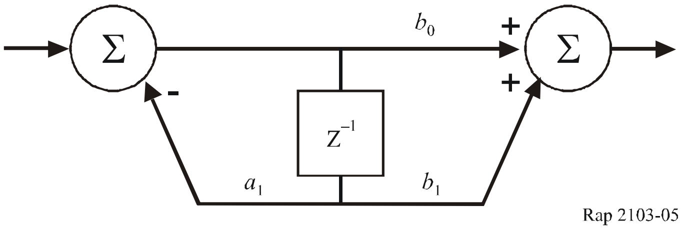


6.2 Output filter topology

The low-pass filter stage at the output was functionally equivalent to the first order IIR topology shown in Fig. 5.

FIGURE 5

Signal flow diagram for a 1st order filter



6.3 Time constant for instantaneous and fast response

The time constant of the filter should be $T = 100$ ms to model perceived instantaneous loudness. This will indicate the perceived dynamic range of the programme. A second time constant of $T = 400$ ms should be used, at the discretion of the operator, to indicate the average loudness with reduced meter swing.

Operational trials in Japan have shown the 400 ms time constant to be preferred over the 100 ms time constant for short-term loudness measurement and indication. It is the view of Japanese experts that a ballistic that changes faster than the commonly used VU meter is not desirable for short-term loudness measurement and indication.

6.4 Audio block decimation

Decimation by averaging blocks of samples may be used to reduce the computation load on the output filter. The decimated sample rate is arbitrary but should be significantly higher than 10 sample/s. Tables 1 and 2 give examples of filter coefficients for a decimated sample rate of 320 samples/s.

TABLE 1

Filter coefficients for first order low-pass filter, *T* = 100 ms

Sample rate	320/s	Gain	6.465674116e + 01
–	–	b_0	1
a_1	−0.9690674172	b_1	1

TABLE 2

Filter coefficients for first order low-pass filter, *T* = 400 ms

Sample rate	320/s	Gain	2.556465999e + 02
–	–	b_0	1
a_1	−0.9921767002	b_1	1

7 Implementation – Loudness meter comparison utility

7.1 Introduction

A software application called Loudness Meter Comparison Utility (LMCU V1.5) has been written by the Australian Broadcasting Corporation to allow comparison of VU, IEC Type II PPM and ITU-R loudness meter characteristics (Recommendation ITU-R BS.1770 – Algorithms to measure audio programme loudness and true-peak audio level). The software runs on a personal computer with the Microsoft Windows XP operating system and may be obtained from the downloads section of the SRG-3 Yahoo egroup (http://tech.groups.yahoo.com/group/srg3list/). The LMCU software shows a waveform display (Fig. 6) for a selected .wav file with an overlaid envelope for each of the three meters, plus a vertical bar graph display of the three meters side by side. Below each bar graph display is a statistical summary of the meter amplitude for the duration of the .wav file, showing median, 75%, 95% and 100% levels. For the loudness meter, arithmetic mean[1] and geometric mean value are also provided. The software allows a first-order low-pass output filter to be selected from a choice of no filter, fast filter (T = 100 ms) and slow filter (T = 400 ms). The raw loudness data can also be exported to an Excel spreadsheet for further analysis.

The T = 100 ms filter (Fast) gives an indication of subjective dynamic range as it corresponds to the temporal integration characteristic of the ear at 2 kHz. This can be quite fatiguing to watch for long periods however, so a second time constant of 400 ms is used to slow down the meter and reduce the size of the swings around the average value. As well as the ergonomic advantages of this approach, it is consistent with sound level metering practice in acoustics.

The software has a play control that allows the .wav file to be played through the standard audio output device. A cursor scrolls across the screen as the item is played.

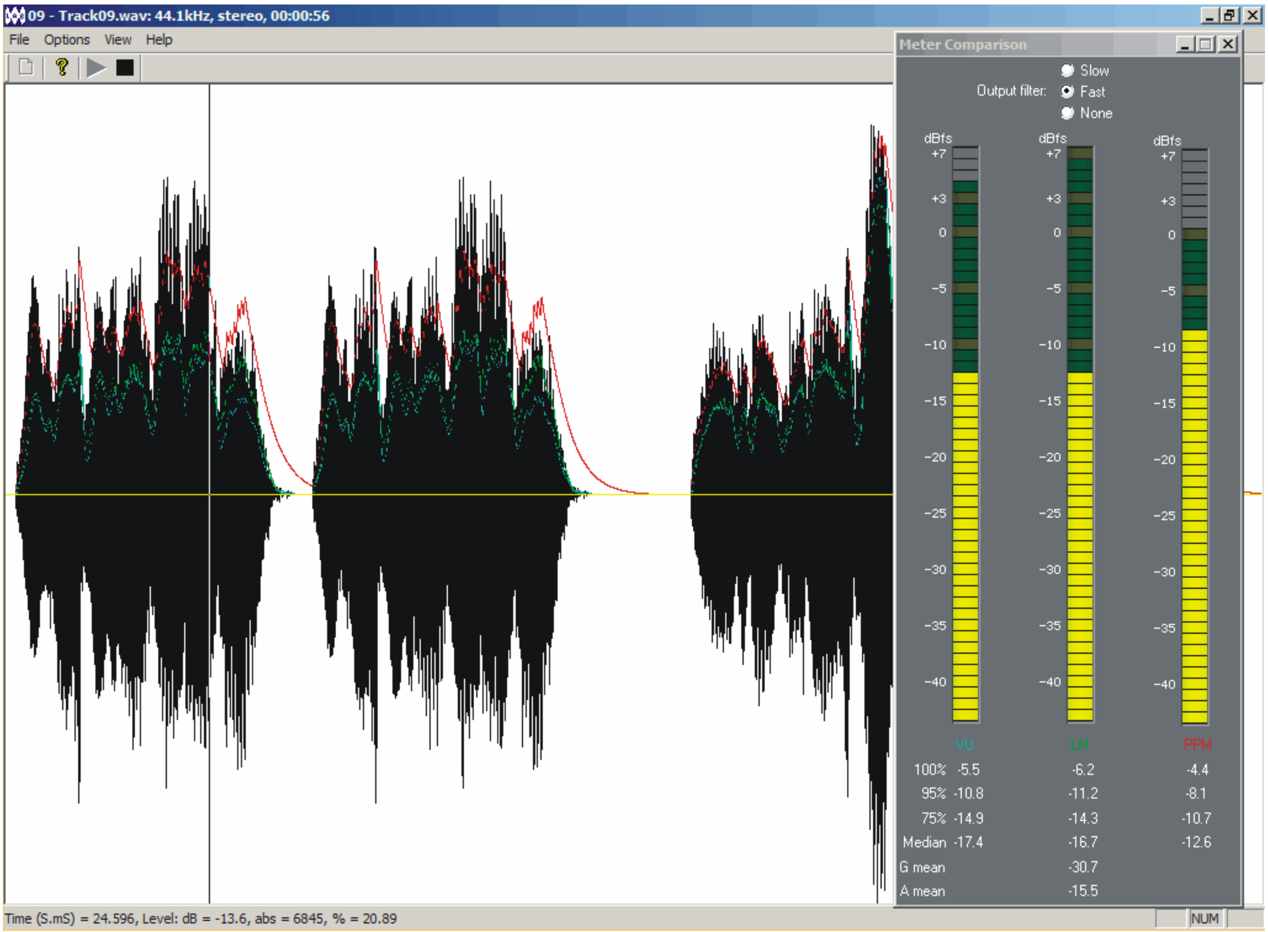
7.2 Calibration

The levels are calibrated in dB, with the VU meter calibrated so that a sine wave peaking at clipping level reads −3 dB. The loudness meter is also calibrated to read −3 dB for a sine wave peaking at clipping level at 1 kHz. The meter time constants are shown in Table 3.

[1] The arithmetic mean loudness is the measure specified in Recommendation.

FIGURE 6

Screen shot of LMCU software (showing track 9 from EBU SQAM test CD)



Rap 2103-06

TABLE 3

Meter time constants

Meter	LP Filter order	Rise time	Fall time
PPM (IEC Type II)	1	−2 ± 0.5 dB from peak for 10 ms toneburst at 5 kHz	2.8 ± 0.3 s for 24 dB fall
VU	2	99% of final value in 300 ms, overshoot 1%-1.5%	Symmetrical
Loudness, no filter	–	instantaneous	instantaneous
Loudness, fast filter	1	First order low-pass filter, T = 100 ms	Symmetrical
Loudness, slow filter	1	First order low-pass filter, T = 400 ms	Symmetrical

7.3 Decimation

Audio sample decimation (sample rate reduction) was incorporated in the LMCU software. The decimation algorithm in the LMCU uses a simple arithmetic average for each block of samples. A $\sin(x)/x$ weighting window could be used instead if desired. The block size was set to be significantly less than 100 ms, the time constant of the fast output filter. The block size chosen, 3.125 ms (150 samples at 48 kHz) was to some extent arbitrary and shorter blocks could also be used without interfering with the output filter characteristic.

Decimation was incorporated for three reasons:

– to reduce the computation load on the output filter;

– to simplify graphic display of the meter envelope;

– to simplify animation of the bar graph display.

The effect of decimation on the mean loudness value is negligible. As it may be a useful tool for loudness meter builders, it is suggested that it should be an optional part of the instantaneous loudness meter specification.

7.4 Gating

For instantaneous loudness indication, gating should not be used. Gating will give a false indication of background noise on wide-range meters during silences.

8 Assessment of performance

It has been suggested that a figure-of-merit is required to assess meter ballistics for instantaneous loudness indication. It has also been suggested that such a figure-of-merit could be derived by operational tests which assess how "accurately" loudness is set using a given meter ballistic. Such a scheme however is very one-dimensional. The purpose of an instantaneous loudness meter is not merely to set long-term average loudness. The purpose is to give visual feedback to an operator on instantaneous loudness so that he/she may make production decisions on the levels of that material. Long-term average loudness is only one parameter of several that may be important in the final product. Others include dynamic range and relative loudness in a mix.

If a figure-of-merit is needed to assess the value of a ballistic, it would be better to use Zwicker's temporal loudness data as a reference and then use r.m.s. error from this reference as the figure-of-merit. On this basis, the first-order IIR filter is ideal, with an error of virtually zero.

9 Conclusions

Recent developments in broadcasting have created a need for a new type of level meter which can indicate perceived loudness more accurately than existing level meters. Existing metering systems were developed to indicate either peak level or a combination of peak level and an approximation of perceived loudness. The importance of peak level however has diminished with the introduction of digital audio production and emission methods with their expanded headroom and immunity to over-modulation.

While it is sufficient to know a single-number long-term average loudness for quality control in emission, it is also necessary to have short-term loudness information in production work. This indicates that several types of loudness meter may be needed.

It is important however that these various types of loudness meter give consistent readings. They therefore should use the same loudness measurement algorithm, as specified in Recommendation

ITU-R BS.1770. The difference between the types of meter is then reduced to the display format and the temporal characteristic, or ballistic.

While loudness perception is a complex phenomenon, for the purposes of broadcasters it has been found that a relatively simple model, described in Recommendation ITU-R BS.1770, can indicate perceived loudness with surprisingly good accuracy for typical programme material over a limited loudness range. The model conforms with Stevens' Power Law [Stevens, 1957].

Stevens' Power Law is a consequence of the fact that biological systems, and perception in particular, can be modelled by first-order differential equations. These apply in the time domain as well as in the amplitude domain.

In the time domain, first-order differential equations can be precisely modelled by first-order analogue filters. Such filters are therefore ideal models for temporal loudness perception.

A first-order analogue filter can be replicated in the digital domain by a first-order IIR filter or by an FIR filter of arbitrary order with exponentially-decreasing filter weights. The IIR filter is far more computationally efficient however and is therefore preferred.

It has been suggested in discussions on loudness metering that a shortened version of the rectangular-weighted FIR filter used for the long-term loudness algorithm could be used for instantaneous loudness indication. The step response for this filter has been compared with the step response of the first-order IIR filter, both optimized to model Zwicker's data on temporal loudness integration. The IIR filter and the rectangular FIR filter have similar attack characteristics, although with less error in the IIR case. The decay characteristics however differ greatly. The IIR filter has smooth decay. The rectangular FIR filter has an uneven decay, initially smooth, then falling away precipitously. The IIR filter is greatly preferable in this regard for short time constants (100-400 ms). For a very slow indication, an integration window may be a suitable solution.

Implementation details have been given for the IIR filter, using a decimated sample rate. The decimation has no significant effect on the reading as long as the decimated sample rate is well above 10 samples per second.

The IIR filter has been implemented in demonstration software called Loudness Meter Comparison Utility V1.5. The software is described briefly. Readers are encouraged to download and try it for themselves to verify the perceptual accuracy of the filter.

It has been suggested that proposed instantaneous loudness meter ballistics be compared on the basis of setting a long-term loudness level. For a short time-constant (100-400 ms) meter, this is not considered a good measure of merit as it ignores other short-term aspects of performance which are equally important. A better reference for an instantaneous loudness ballistic is conformance with Zwicker's temporal loudness integration data. On this basis the IIR filter, with a 100 ms time constant is virtually ideal.

References

STEVENS, S.S. [1957] On the psychophysical law. *Psychological Rev.* 64(3):153–181. A précis is available at http://en.wikipedia.org/wiki/Stevens%27_power_law