Presented at SMPTE Film Conference, March 22, 1997

Are Movies Too Loud?

Author: Ioan Allen, Vice-President Dolby Laboratories Inc.

Introduction

Over the past few years, the subject of movies getting louder has generated increasing concern within the film production community. In addition, it has become commonplace for movie-goers to claim that movies are too loud. Many theatres now project films at an audio fader setting below the calibration setting that would match that in the dubbing theatre, presumably because of audience complaints.

The following material discusses some of the issues that can be described as soundtrack "loudness." What is a measure of "loudness" for a movie? Why do theatres turn the sound down? Have new sound formats, (Dolby SR and more recently the digital formats, Dolby SR·D, Sony SDDS and DTS) exacerbated the perceived loudness problem?

The paper describes an investigation aimed at defining perceived long-term soundtrack loudness with a meter reading, and some data is presented.

Loudness

"Loudness" has been defined many times over the years.[1,2] But the target definitions have been clearly aimed at specific issues, such as how annoying is the background noise level in a working space, how damaging can the noise exposure of a pile-driver be for eight hours, what is the noise level of a recording or transmission medium, or how to compare instantaneous loudness of different spectra. The work described in this paper discusses another "loudness" issue; have movies got "louder" over recent years, and if so, how can the change be quantified?

Those familiar with movie soundtracks would not be surprised at the subjective announcement that *The Right Stuff*, 1983 was a "loud movie". *The Right Stuff* (released in Dolby A-type 70 mm and Dolby A-type stereo optical), certainly was subjectively a "louder" movie than, say, *Shine*, a 1996 release in Dolby Digital. But *Shine* was subjectively louder than *Days of Heaven*, 1978, so perhaps a case could be made that movies have been getting both louder and quieter. Of course, this is nonsense; selected titles can be picked to prove relative "loudness" and "quietness" - potentially *Gone with the Wind* could be used to prove that films have become quieter over the last five decades!

The Current Situation

Back in the mid 1970s, Dolby Laboratories introduced a new concept to the film audio community, a calibration recommendation for monitor levels. A reference pink noise signal introduced in the record chain was used to adjust the audio monitor level to 85 dBc. All theatres equipped for playback of the new stereo optical soundtracks were set up such that an equivalent pink noise signal would generate the same 85 dBc with the playback fader set to the calibrated setting. This meant that theatres playing films at the calibrated fader setting would exactly reproduce the volume levels selected by the film director and mixers in the dubbing theatre.

For years, this system worked - Dolby Stereo (A-type encoded) films had limited headroom, and the resulting constrained dynamic range meant that there were few audience complaints of “too loud;” most theatres played films at the calibrated level.

Soundtrack format technology has been significantly enhanced since the original Dolby Stereo format (see Figure 1 and Figure 2). Dolby SR extended the headroom by 3 dB at mid-frequencies, and much more at low and high-frequencies. In recent years, the new digital formats, Dolby Digital, Sony SDDS and DTS have increased the headroom much further.

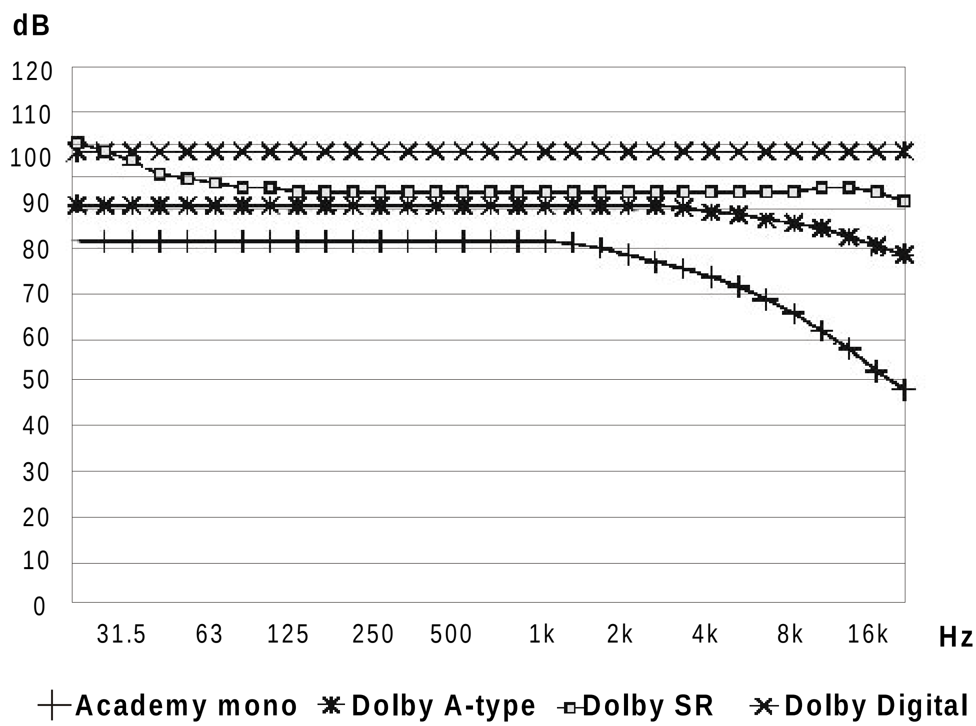

Figure 1 Peak levels of photographic soundtrack formats

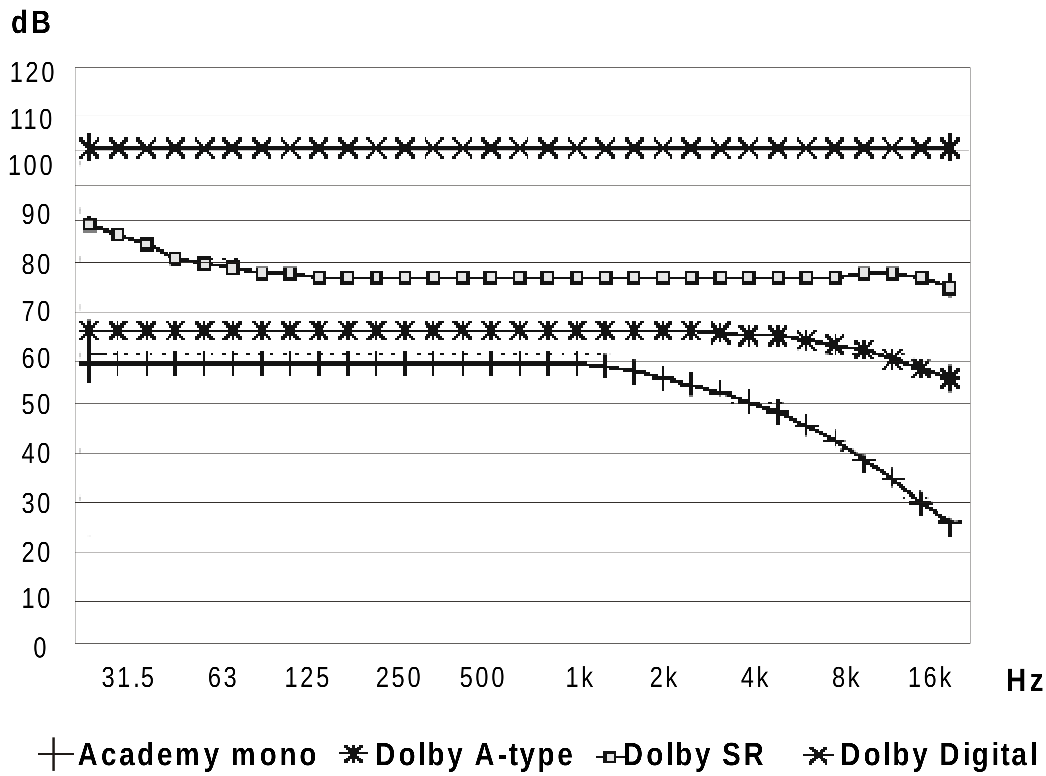

Figure 2 Dynamic range of optical soundtrack formats

Because the 85 dBc calibration technique has been maintained throughout these format changes, the extra headroom has been just that, extra "headroom". Feature films have one consistent subjective mix reference for record level, which can be termed "associative loudness." When the dubbing mixer sees an actor on the screen, and there is no "fight" with music or effects, the dialogue level in a moderate close-up will be set to be plausible for the visual. Within reason, this always holds true to within +2 or 3 dB. (This natural dialogue level does not hold true for narration, as there is no visual reference. We will meet this issue again when we analyze the level on narrated trailers.)

Music and effects have no direct visual associative loudness. Most of us are not familiar with the real sound pressure levels of a Concorde take-off, a 50 mm howitzer, a science fiction energy transfusion machine - and the music score level is equivalently uncalibrated.

As the medium headroom capability has been extended, it has indeed been *used*. Maybe we shouldn't be surprised, but the "non-associative" loudness of effects and music has risen to fill the available space.

In addition, the style of many movies has changed. The "ride" movie is now commonplace from Hollywood, and while the film can be exciting, the phrase "cut-to-the-chase" no longer has any meaning - the chase frequently starts at the beginning of reel 1!

Perhaps the discretionary use of this increased headroom would be justified and indeed desirable on some feature films, but in practice things have gone wrong. Maybe headroom usage alone is not the problem. But here are some obvious symptoms of the situation we are in:

1. There is a growing number of audience complaints about movies being too loud. Newspaper articles are being written on the subject[3,4,5]. Even well-respected mixers are going to press discussing the problem[6].
2. Theatres are playing films way below the calibrated level. A fader setting of “5” as opposed to calibrated “7” is not uncommon, representing a level reduction of 6 dB.
3. Trailers are fighting for competitive loudness. Theatre playback levels are set by the complaints generated by the loudest element of the show. If the playback level is set to accommodate the loudest trailer, the feature will play at the same reduced level dialogue will be lowered by the same level deemed necessary for the trailer. A feature film played 6 dB low may well have serious dialogue intelligibility problems.
4. Mixers are using hearing attenuator pads to avoid the risk of hearing damage.
5. In Europe, where commercials are played before the feature, competitive loudness has led to the desire of a uniform measurement technique, and a self-disciplined constraint.

How Loud Is a Movie?

It would obviously be desirable to be able to measure how “loud” a movie soundtrack is, whether a trailer or a feature. But the conventional level meters in the recording chain are of only marginal help. A VU meter has slow time constants, and is of little use in detecting short-term peaks. The PPM meter was designed to show short-term peak levels which might clip a recording or transmission media. But neither meter system (reaction-time ballistics) will demonstrate an index of what determines the “loudness” of a film soundtrack.

Attempts have been made in Europe to restrict loudness of commercials by defining a maximum level with respect to 100% of the medium clipping level. For example, a 50% limit would mean that no peaks would be permitted to exceed a level 6 dB below the maximum excursion of a Dolby A-type stereo optical release. This constraint, though, would have the effect of restricting dynamic range, and not necessarily controlling the subjective “loudness” of the material.

While loudness has been defined for various purposes in the past, none of the existing criteria can be directly applied to “how loud is a soundtrack.” Fundamental loudness is defined by the relative loudness of frequency against level[1]. A regular VU or PPM meter takes no account of this changing ear sensitivity with frequency and level. Sound level meters commonly do, but without taking account of frequency sensitivity changes with level, are typically switchable to different “weightings”. A-weighting, for example, rolls off low-frequencies in an attempt to match the decreased ear sensitivity to low frequencies.

A further measure of how loud something is takes into account the time domain. It is well known that people attending a three-hour loud rock-and-roll concert may suffer from partial deafness for some time after the concert. But a shorter exposure to the concert (five minutes) has nothing like the same effect. So, it can be construed that the longer a loud sound lasts, the greater the apparent loudness, or perhaps in some cases, the greater the annoyance.

It is worth pointing out that a second-order effect may also be occurring. In some theatres it is possible that the increased use of headroom from Dolby A-type to Dolby SR to digital releases has not been matched by a comparable increase in power amplifier and loudspeaker capability. The resultant distortion from overloaded equipment may well exacerbate the “loudness” problems associated with recent soundtracks.

Loud sounds of short duration may cause great surprise, but little annoyance. A sudden gunshot in the middle of an otherwise quiet desert scene causes little in the way of complaints, even if very loud, but sustained gunshots can. One measure of sustained loudness is called Leq, which was derived to show potential hearing damage from exposure in the industrial environment to sustained but varying-level loud sounds. Leq can be defined as the level of a steady-state tone with the equivalent level in terms of potential hearing damage as a level time-variant signal[7]. There are several different formulae defining Leq, but all average the level of material over time. More accurately, an integral is formed of sound pressure levels, and divided by the duration of the sampling process.

One definition is:

$$Leq_m = 10\log\left(\frac{1}{T}\int_0^T \left(\frac{P_{(t)}}{P_0}\right)^2 dt\right)$$

Experimental Targets

If existing measurement techniques vary in terms of time-constant (rapid reactions to peaks, or a smoothed response), frequency weighting (which frequency ranges contribute most to loudness “annoyance”), and long term averaging, (measurement definitions such as Leq), is there some combination which can be used to result in a value which defines “how loud is a movie?”.

Measurement Procedure

It was decided to set up a variety of measurement techniques, and use them to evaluate a variety of film samples. Figure 3 shows the test set-up, which was used to measure peak levels, and alternate weightings of long-term averaged signals.

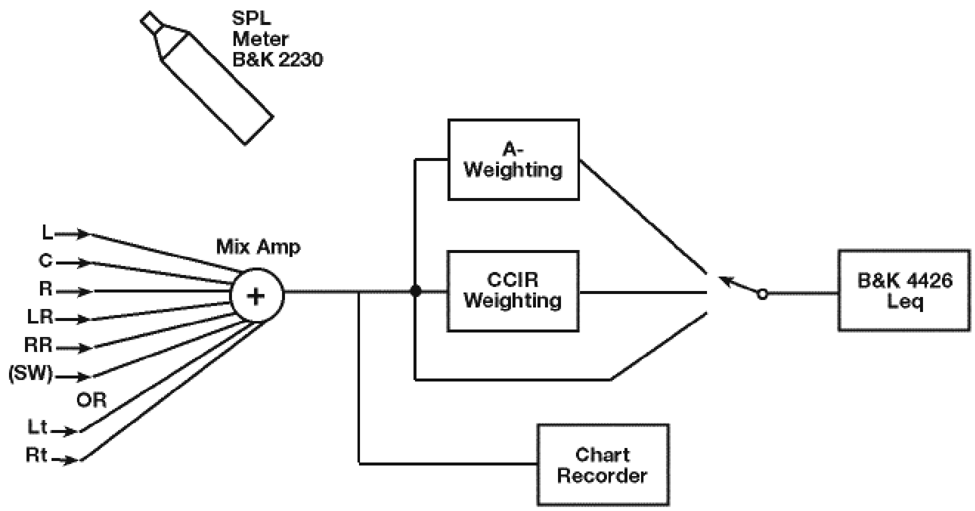

Figure 3 Test set-up

Obviously, it was impossible to measure a totally comprehensive set of material, but selected samples were available as representative of:

1. Contemporary digital trailers
2. Sections of recent digital releases
3. Maximum level Dolby SR releases
4. Maximum level Dolby A-type releases
5. Typical dialogue only recordings in various formats, including Academy mono
6. UK Commercials

Alternative measurement indices

As might be expected the samples have a different hierarchy when assessed different ways. The first comparison is between Leq unweighted and weighted, and as Leq is typically A-weighted, the low-frequency roll-off of A-weighting results in a reduction in Leq when the material has a substantial bass content. Determination of annoyance, though, does not necessarily match the A-weighting curve. The author canvassed opinion based on subjective reaction to various film samples, and concluded that a weighting curve with a heavier emphasis on the 2 – 6 kHz region would better match subjective soundtrack loudness.

A weighting curve used for low-level recording medium noise (CCIR weighting)[8] more closely matched the subjective annoyance criteria (see Figure 4); there is no technical parallel between high-level soundtrack loudness assessment and low-level recording medium noise - the CCIR curve was simply a convenient available weighting filter for the tests, and provided a better subjective match than Leq_a. An adaptation of the CCIR measurement technique, termed CCIR/ARM, offset levels by 5.6 dB for normalization purposes[9] with a 2 kHz reference point, and the same offset was used for the data presented in this paper.

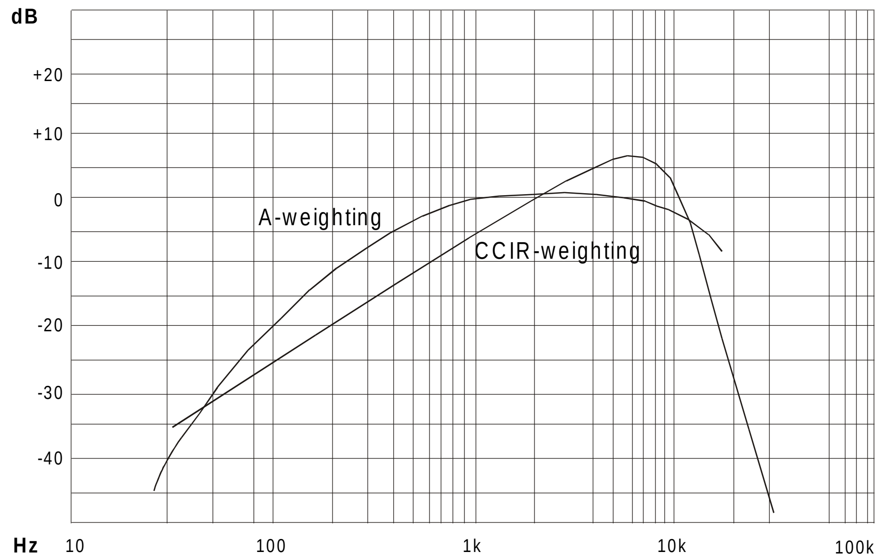

Figure 4 A-weighting vs. CCIR-weighting curve (offset by 5.6 dB)

For the purpose of this paper, the term Leq_m has been adopted for Leq with CCIR-weighting, and a 5.6 dB offset. Figure 5 shows that the *average* Leq_m number matches that of Leq_a. The higher the relative Leq_m with respect to Leq_a shows a signal content with greater emphasis in the 2 – 6 kHz region.

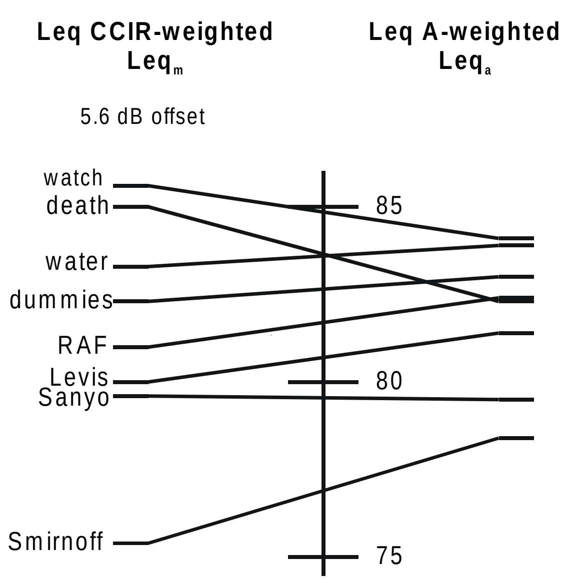

Figure 5 CCIR-weighting (Leq_m) shows influence of 2 – 6 kHz region

Analysis of the Data

Figure 6 shows the relationship between unweighted (Leq) and CCIR-weighted (Leq_m) samples. The hierarchy on the right of the figure (unweighted) is different from that on the left (weighted). The slope of the connecting lines is indicative of the amount of bass content in the program. The greater the bass content, the more positive the slope. As the slope tends to horizontal, or even negative, the less the bass content. Out of the samples shown, it will be noticed that *Indiana Jones and the Temple of Doom* is raised in the hierarchy because of the lack of bass in the five-minute sample. This matches the apparent loudness of the sample, one of the “loudest” Dolby A-type films.

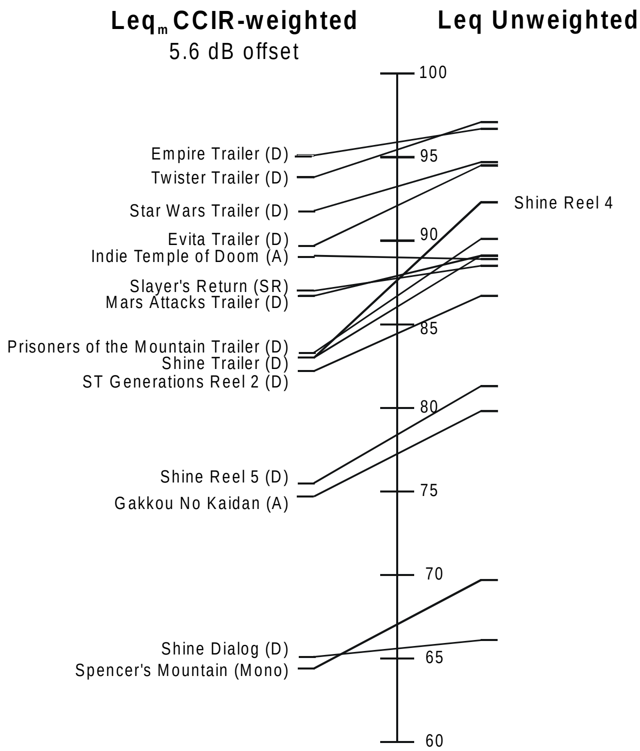

Figure 6 Comparison of weighted and un-weighted Leq

The next comparison (Figure 7) is between the maximum peak level reached in each sample, and the weighted Leq_m value. Again the slope of the line between the values is significant; in this case, the greater the slope, the greater the “dynamic range.” An unfortunate misuse of the term “dynamic range” suggests the value of the maximum peak level. Here we are using the term to show the range between average levels and the loudest peak value events on the soundtrack. Program with little change in level, whether close to peak level of the recording medium, or much lower, has little dynamic range. Program with sudden changes in level (a sudden gunshot in a quiet desert scene) has a high dynamic range.

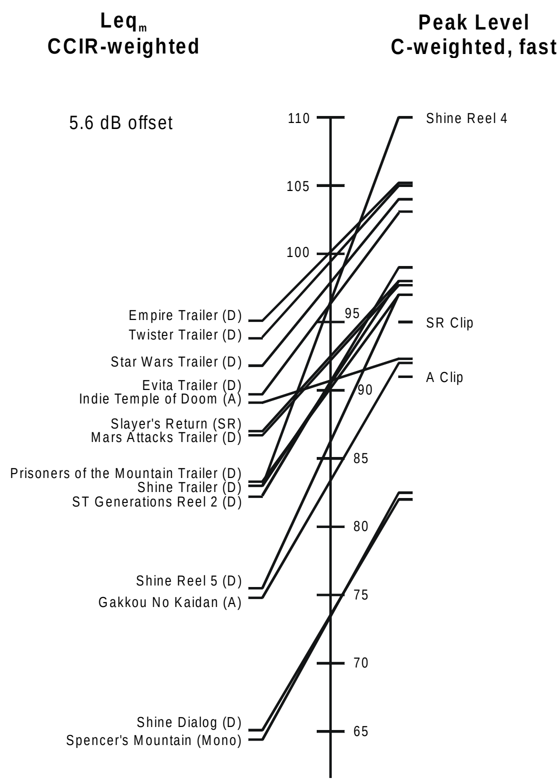

Figure 7 Comparison of Leq_m and peak levels

Several issues raised by the data are worthy of discussion. Perhaps the first that should be considered is that of dialogue level. The concept of associative loudness (discussed above) is supported by the comparison of dialogue levels of *Spencer's Mountain* (1963), an “Academy” mono film with the dialogue level of *Shine* (1996), a Dolby digital release. The two films show identical Leq levels, confirming the supposition that mixers set dialogue levels at a plausible point, matching the visual, regardless of the format of the release medium.

Next, examine Figure 7 again, and Figure 8 and compare the peak level reached in *Shine* Reel 4 with the Leq, and it can be seen that this reel has an extreme dynamic range. But as no-one has ever suggested to the author that the film was too loud, it must be assumed that the dynamic range is correct for the nature of the material, and that as might be expected the Leq is much more indicative of the subjective loudness than a peak measurement.

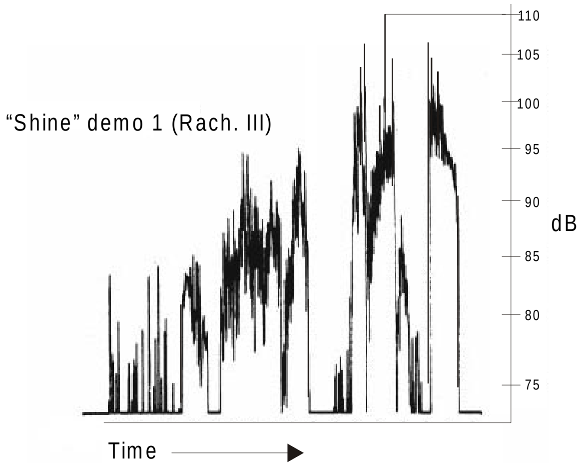

Figure 8 Peak levels in reel 4 of Shine

Now, looking at Figure 9, we can see that some of the UK commercials are pushing at the maximum available loudness of the available format. Even these Dolby A-type soundtracks show a loudness far greater than a constraint on maximum peaks would control.

But the most obvious feature of the Leq listing is the loudness of US-created digital trailers. Head and shoulders above most features, in some cases the trailers have an Leq 20 dB above the feature they advertise.

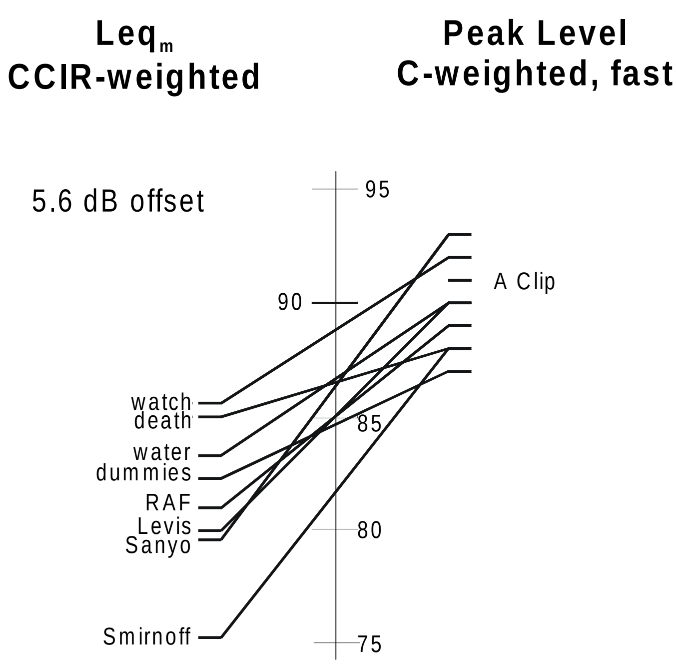

Figure 9 Comparison of Leq_m with peak levels for UK commercials

Verifications

A couple of verifications were carried out on the data – first, even though the prime measurements for the data was a direct electrical sum of the analog or digital film itself, peak levels were checked from time-to-time with a sound pressure level meter in an auditorium. In addition, the lines indicating A-type and SR on the figure shows the maximum theoretical levels of these formats. The data shows maximum recorded levels slightly in excess of these theoretical numbers, as might be expected with the slight excessive dimensional modulation consistent with contemporary practice.

Conclusions from the Data

The main conclusion to be drawn from an analysis of the data is that Leq_m provides a good measure of the subjective loudness of the film. The hierarchy of loudness shown seems a close match to the subjective hierarchy.

Are there health hazards?

This question has to be broken into several elements, but first it should be pointed out that the author is not qualified in this field, and can do no more than quote other sources.

First it seems that occasional peaks are still a substantial distance below the level which can cause instantaneous damage[10].

Sustained loudness can cause damage at much lower levels, though, and while there is no evidence to suggest that a two-hour movie once or twice a week is anywhere near damage (it is certainly way below the levels of rock concerts), there is material which suggests that *mixing* at the levels of current *trailers* could cause damage.

ISO1999/1[11] describes a method of assessing potential hearing damage from noise of varying level, absorbed over years. Here partial Leq is used to calculate the dosage. As an example, consider a mixer who works for four hours per day, mixing trailers at a level equivalent to the average of the loudest three shown in our data, *Empire Strikes Back*, *Twister* and *Star Wars*. The average Leq_m of these three items is about 94 dB. As on average Leq_m is equivalent to Leq_a to within +2 dB, we can assume an Leq_a of approximately 94 dB. Going through the procedure described in the standard:

Assuming a level of 94 Leq_a for four hours in the day (and optimistically, silence for the other twenty hours!), the standard derives a partial noise exposure index of close to 160. This leads to an equivalent continuous noise level of 92 dB(A). The standard goes on to suggest the percentage of people subjected to these noise levels suffering from hearing impairment compared with those of a non-exposed group (below 80 dB(A)) in their work environment.

The numbers are (not-surprisingly) related to years of exposure:

Years	0	5	10	15	20	25
% (extrapolated)	1	7	15	21	27	31

One can conclude that mixers have a real chance of hearing damage if continually working on trailers at current levels.

As far as working hours per day is concerned, Moore[12] points out that if the exposure duration is halved, the permissible equivalent sound intensity is doubled, corresponding to a 3 dB increase in level. Thus if 93 dB is permissible for 4 h, 96 dB for 1 h, 99 dB for 0.5 h, 102 dB for 15 min, 105 dB for 7.5 min, and so on.

What can be done?

The current situation is obviously unsatisfactory, for audiences, theatre operators and mixers. Now that it seems demonstrable that the Leq method can put real numbers on the loudness of a mix, we would seem to be better equipped to say “this is too loud”, or “this is OK.” It seems presumptuous were this paper to propose numbers, for this is a step that should be taken by the artistic community. But as a first step towards sanity, it would seem desirable that pressure be brought to bear on those in charge of mixing trailers and commercials, where a significant lowering of Leq would have several benefits:

1. The fader in the theatre would not be lowered specifically for the trailers and/or the commercials. This would leave the feature playing at the calibration level. We would avoid the frequent dialogue unintelligibility in features resulting from lowered fader settings.
2. There would be a substantially reduced risk of hearing damage to mixers' hearing.
3. There would also be a chance of re-establishing calibrated levels between mix rooms and playback theatres.
4. And in due course, this could lead to a more careful analysis of the levels of the feature film itself.

Metering

Existing meters for direct SPL measurements of un-weighted Leq and Leq_a include the B&K 2230 (now replaced by the 2236), and the B&K 4426. In addition, an investigation is underway to assess the possibility of a simpler analog meter, and incorporation of a finer-tuned weighting curve based on the level-offset CCIR characteristic. Such a meter would output a value for Leq_m, and would be designed for interface to a Dolby DS4 or DS10, the coding platforms for analog and digital soundtracks.

Conclusions

A variety of measurement techniques were used to evaluate a selection of film samples. It was determined that Leq_a, or better a modified method called Leq_m, provided a numerical hierarchy of the samples that closely matched subjective experience. It was shown that equalization has a major effect on ongoing loudness, and that trailers are significantly the major offenders.

References

1. Zwicker and Scharf (1965). *A model of loudness summation*. Psychol. Rev. 72, 3-26.
   Emmett (1990). Programme loudness assessment. IBC 1990. Transactions of the International Broadcasting Convention. *IEE London UK*.
2. Johnson (August 17th 1996). *Whacrashtwasthatbangcreakagain?* The Economist.
3. Stempel. (June 24th 1996) *Sound Effects and Bad Editing Often Drown Out the dialogue*. Los Angeles Times.
4. Camp. (August 12th 1996) *Who's Been Turning Up the Volume at Cinemas?* Miami Herald.
5. Thom (April 1996). *Are Movies Getting Too Loud*. Journal of the Cinema Audio Society, 3.
6. Harris (1991). *Handbook of Acoustical Measurements and Noise Control.* McGraw-Hill Publishing Company.
7. CCIR XIII Plenary Assembly Report 3982, Recommendation 468-1 and Doc X/22 presented in evidence.
8. Dolby, Robinson and Gundry (1979). *CCIR/ARM: A Practical Noise-Measurement Method.* Journal of the Audio Engineering Society Volume 27 Number 3.
9. Kryter (1985). *The Effects of Noise on Man.* Academic Press Inc.
   ISO1999-1975 (E) Acoustics - Assessment of occupational noise exposure for hearing conservation purposes.
10. Moore (1989). *An Introduction to the Psychology of Hearing.* Academic Press Inc.

Acknowledgments

The author appreciates assistance from David Schwind, Kenneth Gundry, Louis Fielder, Craig Todd, Robin Bransbury and Tim Partridge. And particular thanks to Tom Bruchs, who showed great patience while recording data.

List of Figures