Frequency distribution of temporal sound pressure capacity requirement in two-way and three-way active monitors

Aki Mäkivirta
Genelec Oy, Finland, aki.makivira@genelec.com

Abstract

Active monitoring loudspeakers are optimized for very high sound pressure output using very high power amplifiers in an enclosure having a compact physical size. All drivers in such a system use thermal overload protection to limit heating in the voice coils. Optimal protection allows full lengths of the typical audio events to be reproduced before the thermal protection must activate. This work studies the temporal statistics of an audio signal to determine how the limit to the duration of the maximum power level output must be chosen for minimum audible impact. The statistical distribution of the audio event durations in cinematic 24 bit word length, 48 kHz sample rate audio track and 16 bit word length 44.1 kHz stereo audio tracks are studied for the two-way and three-way active monitoring systems to determine the temporal capacity requirement as a function of frequency. This data is analysed for each driver output channel separately in the case of a two-way monitor and a three-way monitor. Statistics are presented to describe the expected maximum sound pressure level audio event duration as a function of frequency. The mean maximum level acoustic event duration is inversely proportional to the low corner frequency of the driver channel bandwidth. Music sound tracks and film sound tracks do not differ in this respect even if the music sound track crest factor can be significantly smaller.

Keywords: sound pressure, capacity requirements, temporal distribution, maximum level, audio monitors
1 Introduction

Active monitoring loudspeakers are optimized for very high sound pressure output using high power amplifiers in an enclosure having a compact physical size. The drivers are directly connected to the drivers. The drivers in such a system require overload protection to prevent damage and limit heating in the voice coils. An optimal design for the active speaker system including the protection system allows full lengths of typical audio events to be reproduced before the protection activates.

This paper studies the temporal statistics of the events of very high level in the audio signal to determine what should be the desired limits for the duration of the maximum power level output into the drivers such that there is the minimum likelihood of an audible impact due to the activation of the protection.

This paper analyses the temporal statistics of the content of digital audio signals to determine the distribution of the durations of the signal value very close to the digital full scale. We also discuss the aspects matching the monitor loudspeaker output capacity with the acoustics of the listening space as one of the fundamental questions in setting up the monitoring system is to determine the capacity of the monitor loudspeaker once the intent of using the monitoring room and the acoustic characteristics of the room are known. The fundamental aim of this work is to create the understanding to match the monitoring loudspeaker characteristics correctly with the intent of monitoring.

2 Limitations to active monitor output level

An important benefit of an active audio monitor is that the power amplifier is directly connected to each driver in the system. This enables very large peak level to be passed from the power amplifiers into the drivers.

The thermal capacity and heat conduction of a dynamic driver voice coil are not large. The crest factor (peak power to average power ratio) of a typical music or speech signal is relatively large, about 10...20. However, the crest factor alone cannot guarantee that the voice coil cannot overheat. As continuous steady state signals can be valid audio signals the long term sound output level of an active monitor is normally limited with a protection system to prevent the driver voice coils from overheating [10,11,12,13]. Activation of the protection lowers the highest sound level and can also change the fidelity of the sound output.
A fundamental goal of the active monitor design is to prevent activation of the thermal protection under all normal operating conditions. This requires careful balancing of the short term maximal output capacity and the long term maximal output capacity using assumptions about the operating conditions, the nature of the acoustic signals to be monitored, as well as heat management capabilities in the drivers and electronic subsystems. The assumptions typically consider (1) the sound pressure typically used in monitoring, (2) headroom needed for the peak sound level above the typical monitoring level, needed for reproducing the maximums in the audio signal, (3) distance from the monitor to the listening position, (4) reverberation time of the space where monitoring takes place, and (5) typical dynamic level variability properties of the audio material.

3 Typical monitoring level

Modern immersive audio reproduction systems use a large number of loudspeakers in geometrically predetermined positions. Examples of these include the broadcasting-related systems such as ITU recommendation for immersive audio [1] and the NHK 22.2 system [6], as well as commercial distribution and presentation systems such as Auro 3D [7,16], Dolby Atmos [4,5], and DTS Neo:X [3]. ITU recommendation BS.2159-4 for broadcast applications covers both channel-based and object-based immersive presentation systems, including 10.2, 22.2, wavefield synthesis, and object-based audio presentation formats [1]. International Electrotechnical Commission IEC 62574 describes a three height layer reproduction system for consumer applications with up to 32 speaker positions including an over-the-head speaker, able to support all current multichannel formats [1,2].

Level alignment practices use 18 or 20 dB digital audio headroom. A permitted maximum signal level is usually set to −9 dBFS, but the peak level may be finally scaled to digital full scale [8]. Maximization is done as the final stage, when the print master mix is complete. During recording of the individual tracks and objects sufficient headroom is maintained so that signal fidelity can be maintained.

EBU recommends the programme loudness instead of controlling the permitted maximum (peak) level. The loudness should be regulated to -23 LUFS so that the digital audio full scale (FS) is not exceeded [14]. However, these figures reference to the digital full scale, and do not say much about the sound pressure level (SPL) used for monitoring the audio signal.

For the monitoring level EBU is recommending a scaling from an -18 dBFS pink noise input to a sound pressure level (SPL) at the listening position $L_{\text{mon}}$ such that the monitoring level depends on the number of channel in the reproduction system

$$L_{\text{mon}} = 85 - 10\log(n) \quad \text{[dBc]} \quad (1)$$

For example, the monitoring level becomes 78 dBc SPL for each monitor in a 5.1 system, 75 dBc for a 10.2 system, and 71.6 dBc for a 22.2 system [8,14,15]. This approach to setting the levels lowers the individual SPL capability requirement for each monitor when the channel count grows.
However, certain channels may need to have an output capacity higher than this, for example when the environment is intended for multiple types of productions, with varying channel count. One such approach is presented in IEC 62574 [2]. While the broadcasters recognize the capacity of a high channel count system to achieve a higher aggregate SPL, this principle is not applied by the cinema industry. SMPTE recommends aligning at the listening position to 83 dBc at a –20 dBFS input level [8], implying a peak level of 103 dB SPL by an individual monitor. Dolby Atmos specifies that each screen (front) monitor has the maximum continuous output capability of 92...105 dB SPL at the listening position, depending on the frequency range, and 99 dB SPL for the surround monitors [4].

4 Audio data value distribution and crest factor

Audio signal value distribution has been studied, particularly for music recordings. A typical statistical value analysis for film sound tracks shows the highest likelihood at –20... –15 dB of full scale [21]. This coincides with the way the headroom is typically set for film sound.

In the following discussion the typical temporal properties of audio in film sound tracks and stereo audio musical tracks are studied. The film material was selected because films are considered to contain the widest dynamic range and have been recorded with 24 bit word length, making it possible to represent a wider dynamic range. The film sound tracks used as material represent the so called action genre and is a relatively recent multichannel production (Table 1, material A). The multichannel soundtrack was extracted from a DVD release containing a 5.1 channel sound presentation at 48 kHz sampling frequency and 24 bit sample precision. Also the LFE channel was included in the extracted data, but the LFE channel content has 10 dB lower level than the main channels. The two music sound tracks (materials E and F) are 16 bit stereo recordings sampled at 44.1 kHz. The music sound tracks were selected because of high loudness and frequent occurrence of high peak levels in the sound data.

The crest factor (ratio of the peak value to the RMS value) was calculated for all materials. The film sound track crest factor is significantly larger than that of the music sound tracks (Table 1). The film sound track has about 8 dB larger crest factor than the music tracks. The film crest factor figure was calculated using Chapter no 36, containing frequency high value levels in the sound data.

<table>
<thead>
<tr>
<th>year</th>
<th>type</th>
<th>ch.</th>
<th>ori.</th>
<th>duration h:mm:ss</th>
<th>size GB</th>
<th>crest (dB)</th>
<th>title</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>film</td>
<td>5.1</td>
<td>US</td>
<td>1:50:03</td>
<td>5.31</td>
<td>19</td>
<td>I Robot</td>
</tr>
<tr>
<td>E</td>
<td>music</td>
<td>2.0</td>
<td>US</td>
<td>0:02:47</td>
<td>0.028</td>
<td>11</td>
<td>Ace of Spades (Motorhead)</td>
</tr>
<tr>
<td>F</td>
<td>music</td>
<td>2.0</td>
<td>KR</td>
<td>0:03:39</td>
<td>0.042</td>
<td>11</td>
<td>Gagnam Style (Psy)</td>
</tr>
</tbody>
</table>
5 Durations of high peak level runs

Audio signals can be short or infinitely long but in typical audio material, high level peaks have finite length. The high peak levels in the audio material are studied by evaluating the frequency and length of peaks exceeding a detection level $L$. This enables evaluation of how likely are peaks exceeding level $L$, and what are the expected lengths of the peaks. In order to avoid attenuating magnitudes of short peaks, we look at the magnitude of the signal envelope $h(n)$ (Figure 3) defined as the analytic signal [22]. The histogram bins have one dB width across the 24-bit and for the 16-bit PCM presentation dynamic ranges and contain data of the full duration of the audio track.

The durations exceeding the level $L = -10$ dBFS were studied (Figure 5). The likelihood of peaks increases with decreasing peak length. High level long duration peaks are unlikely to occur. The peak statistics change once the signal is bandpass filtered as a part of an active monitor or subwoofer crossover. The typical crossover filters are modeled as fourth-order Linkwitz-Riley frequency responses at 90 Hz for the subwoofer, 500 Hz and 3 kHz for the three-way monitor model, and at 3 kHz for a two-way loudspeaker (Figure 6). These crossover frequencies correspond with the typical values seen on the market. The peaks in the tweeter channel are typically 40 $\mu$s in length and seldom more than 1 ms long. The peak length in the
subwoofer channel is the longest, typically at 10 ms but the duration in a subwoofer can be several hundred milliseconds in length. The duration limits where 99% of peaks are shorter than the limit are given in Table 2. Material A figures are presented as they were the worst case. There was no clear difference between the film sound track and the music sound tracks. In fact the likelihood of long high peak value runs was higher in the film sound track even if the crest factor in the film material was smaller.

Table 2: Lengths and relative occurrences of peaks in driver-specific signals for material A

<table>
<thead>
<tr>
<th>crossover channel</th>
<th>passband (Hz)</th>
<th>run length 1% limit (ms)</th>
<th>occurrence (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tweeter</td>
<td>&gt; 3 k</td>
<td>0.4</td>
<td>7.7</td>
</tr>
<tr>
<td>midrange</td>
<td>0.5 k - 3 k</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>two-way woofer</td>
<td>&lt; 3 k</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>three-way woofer</td>
<td>&lt; 0.5 k</td>
<td>40</td>
<td>16.4</td>
</tr>
<tr>
<td>subwoofer</td>
<td>&lt; 90</td>
<td>300</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Figure 3: Two-way monitor case.
6 Sound level dependency on monitor properties

An active monitor loudspeakers has a large short-term sound output capacity but lower sustained long-term sound output capability mainly because the thermal capacity and heat conduction from the dynamic driver voice coils is limited and the long term output level must be limited to prevent driver voice coil overheating [10,11,12,13].

Protection lowers the sound level. An active monitor must be selected for an application so that protection activation can be avoided under all normal operating conditions. This selection is making assumptions about the operating conditions. The parameters to consider include the (1) sound pressure used in monitoring, (2) overhead capacity needed to reproduce the peaks in the sound output, (3) distance from the monitor to the listening position, (4) reverberation time of the monitoring space, and (5) statistics of the level variability in the typical audio material.

The typical installation case is that the monitor is placed near an acoustically hard wall. The reverberation time in high quality monitoring spaces is relatively low.

The distance to the monitor and the acoustical characteristics of the monitor and room (reverberation time, directivity) influence the sustained sound level in the monitoring room. Most monitors and subwoofers are small radiators in relation to the radiated wavelength, and therefore exhibit sound level reduction of 6 dB for doubling of the distance until the room reverberation slows down the sound level reduction with distance.
Following the principles in [18] we can estimate the sound level $L_p$ on the acoustical axis when the listening position distance is $r$. The sound absorption area $A$ for the internal volume $V$ of the room is obtained once the reverberation time has been measured.

$$A = \frac{0.161 V}{RT_{60}} \text{ [m}^2\text{]} \quad (2)$$

Monitor directivity $Q$ is either given by the manufacturer or it can be estimated with measurements. Some directivity data is also available from public sources [23,24]. Manufacturers may give this data in the form of a directivity index $D$.

$$Q = 10^{10^D} \quad (3)$$

The sound output level on the acoustical axis and referred to the one meter standard distance $L_{p0}$ is usually given by the manufacturers. The radiated acoustic power $L_w$ can be estimated assuming the usual use case with the monitors placed close to an acoustically hard wall.

$$L_w = L_{p0} - 10\log\left(\frac{2Q}{4\pi}\right) \quad \text{[dB PWL]} \quad (4)$$

Loudspeaker directivity and room absorption are typically frequency dependent. For precise calculations this should be considered. The midrange average values can be sufficient for estimating the typically achievable sound level $L_p$ at the listening position, on the acoustical axis of a monitor, at distance $r$ to the monitor can be expressed, expressed as

$$L_p = L_w + 10\log\left(\frac{2Q}{4\pi r^2}\right) + \left(\frac{4}{A}\right) \quad \text{[dB SPL]} \quad (5)$$

**Figure 5:** The effects of reverberation and source directivity on SPL at the listening location.
The main determinants of the maximum SPL at the listening location are the maximum sound pressure $L_{p0}$ generated by the loudspeaker and the listening distance. A typical listening distance is about 2.5 m and ranges from 1.2 to 4.2 m [24]. The room reverberation has a minor effect when the reverberation time is small [17,20]. This is normal in professional monitoring spaces. The directivity effects mainly the early reflection level and other room related modifications of the imaging and sound colour at the listening position mainly.

7 Conclusions

The peak SPL and sustained long-term SPL output capacities of a monitor loudspeaker are the decisive factors in low reverberation level rooms such as typical audio monitoring rooms. The other main determinant for the applicability of a monitor is the distance from the monitor to the listening location. With modern multichannel loudspeaker layouts [16,20] the distances should be carefully studied as the physical layout of the speakers in the room usually determines the actual distances, and with electronic compensation of the time-of-flight for audio it is not necessary to put all the speakers at equal physical distance [9]. This can make the peak SPL demand for the individual monitors different.

Comparing the lengths of the signal runs at very high output level, in this study defined as level higher than -10 dBFS, the film sound track and music recordings did not differ significantly in terms of the likelihood of high level runs.

The crest factor of the film sound track was higher than that of music recordings. This is well known and music recordings have sustained a long trend of reducing crest factor due to the so called "loudness war" [19] where the recording engineer attempts to maximize the loudness in a recording with various signal compression tools.

References


