Sound Pressure Capacity Requirements for Monitoring Immersive Audio Formats

Aki Mäkivirta, Juha Holm, Juha Urhonen, Ilkka Rissanen, Jussi Väisänen
Genelec, Iisalmi, Finland, email: aki.makivirta@genelec.com

Abstract

The statistical distribution of the sound event duration in a cinematic immersive audio track is considered as information to determine the short and long term sound pressure output capacities needed in loudspeakers and subwoofers. The statistical distribution of audio event duration in cinematic audio tracks is studied to evaluate the needed monitoring system capacity. Monitoring system capacity requirement is also related to the monitoring room size, the listening distance, and the room reverberation time. Impact of using bass management and selection bass management crossover frequency are not considered. An attempt is made to develop these factors into a design guideline enabling successful monitoring room design and selection of the acoustic and physical characteristics of the monitoring loudspeakers and subwoofers needed in the room of a certain physical size.

Introduction

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 [7], as well as commercial distribution and presentation systems such as Auro 3D [9], Dolby Atmos [5,6], 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].

Use of channels

Object based audio typically uses a *bed* in a 7.1 presentation, containing most of the audio presentation [11]. The bed presentation is usually done with the monitors on the listener ear level, set according to the ITU recommendation at equal acoustical distance and level at the listening position [12].

Object monitors or height layer monitors are usually assumed to be acoustically set to the same acoustic distance and level as the bed monitors. However, it is not untypical to find that the actual physical distance of the monitors may not be the same, and that the acoustical distance equality has been achieved with electronic delays. This produces a monitoring arrangement that performs perfectly from the sound design point of view but may require some of the monitors to work relatively harder, outputting a higher sound level, as they are located at a larger physical distances, and are operating at a higher output level after the level calibration has been completed. This should also have an influence on the monitor type choice if we know that certain monitors are physically located at a larger distance from the listening position.

Monitoring level

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 [10]. 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 [17]. 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_{mon} such that the monitoring level depends on the number of channel in the reproduction system

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

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 [10,17,18].

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 by the

standard IEC 62574 [2] proposing a multipurpose monitor layout.

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 monitoring systems to 83 dBc at a –20 dBFS input level [10] at the listening position, implying a peak level of 103 dB SPL by an individual monitor. Dolby Atmos specifies that each screen (front) monitor should have 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 [5].

Effect of reverberation

Whilst the shot-term peak SPL of a monitor is determined by the monitor design, the long-term SPL or the sound level that can be sustained by a monitor depends heavily on the room acoustics, notably the reverberation time. Larger reverberation times are linked to smaller acoustic absorption in the space, leading to higher steady state SPL.

$$T_\mathrm{m} = 0.25\,(V/V_\mathrm{o})^{\frac{1}{3}},\ V_\mathrm{o} = 100\ \mathrm{m}^2 \qquad [\mathrm{s}] \qquad (2)$$

Higher reverberation time is linked to more pronounced room resonances, and these are detrimental to accuracy of monitoring. Because of this, high quality monitoring spaces strive towards low reverberation time, and the targets are given by several recommendations. For example ITU-T BS. 1116-1 recommends a midrange reverberation time *T*m for high quality monitoring spaces. The typical reverberation time in professional monitoring rooms has the median value of 0.4 s in the mid frequencies (Figure 1).

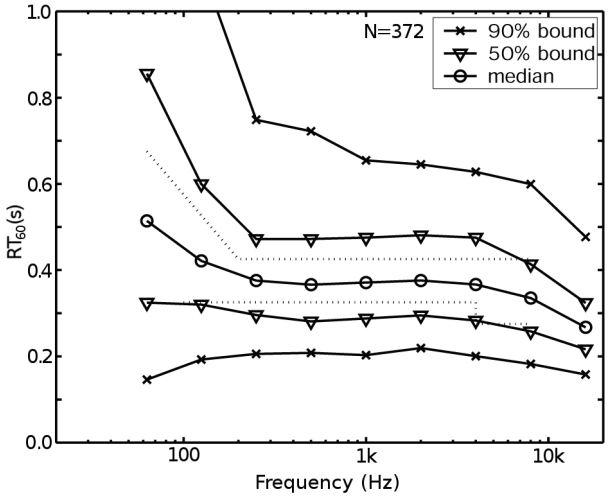

Figure 1: Percentiles of the sound reverberation time RT_{60} in professional monitoring spaces (N=372) [24]

Distance from monitor to listening position

The monitor layout for the cinema reproduction systems is not mechanically equidistant, so requirements for individual monitors vary according to monitor location. The reference listening location is typically closer to the back wall than to the front, reducing distances to the rear monitors. Auro3D places the height monitors above the respective ear-height monitors [19,20] and this leads to a larger distance from the listener to the higher-tier monitors. The voice-of-god or over the head monitor distance can vary depending on the room height and other practical factors. The relative locations of the ceiling object monitors result in varying distances to the listening location.

The mean listening distance in a stereo configuration is about 2.5 meters with the listening distances ranging from 1.2 to 4.2 meters [24]. Cinema standards require monitoring rooms emulating the film theatre and require front monitor distance of at least 5 meters, but there is also a recommendation for a room size, monitor layout and setup very similar to ITU recommendations for multichannel music productions [25].

Distance dependency of sound level

The distance to the monitor and the room reverberation time affect the sustained sound level in the monitoring room.

The sound output capacity of a monitor is usually given by the manufacturer at a certain nominal distance. The standard distance is one meter. Both the short-term peak sound level and long-term sustained maximum sound level scale with the distance.

Most monitors and subwoofers behave like point sources in terms of sound level attenuation with increasing distance to the monitor. This means that the sound level decreases by 6 dB for every doubling of the distance to the monitor until the reverberation in the room begins to dominate and the sound level reduction with distance becomes small [22].

In high quality monitoring spaces the reverberation time is relatively low. It is usually safe to assume that the square law of level reduction with distance is valid for usual listening distances of the monitors in the monitoring room.

We can follow the principles given in [22] to estimate the sound level *L*p on the acoustical axis of a monitor at the listening position when the distance to the monitor is *r*.

For this, we assume that the monitor is placed near a wall, as this is the typical installation case. The attenuation of the direct sound from the monitors as a function of the listening distance, the sound absorption in the room, and the directivity of the monitor must be considered.

The sound absorption area of the room *A* is usually best available after measuring the reverberation time RT_{60} in the room, knowing the internal volume *V* of the room.

$$A = \frac{0.161\,V}{RT_{60}} \qquad [\mathrm{m}^2] \qquad (3)$$

Monitor directivity *Q* is either given by the manufacturer or it can be estimated with measurements. Manufacturers usually give this data in the form of a directivity index *D*.

$$Q = 10^{\frac{D}{10}} \qquad (4)$$

The radiating power L_w for a monitor is not given by the manufacturer. However, the sound output level on the acoustical axis at one meter L_{p0} is usually given, and can be used to estimate the radiated acoustic power. Here we take the usual case and assume the monitors are placed at or close to a wall.

$$L_w = L_{p0} - 10\log\left(\frac{2Q}{4\pi}\right) \quad \text{[dB PWL]} \qquad (5)$$

Directivity and absorption are frequency dependent, and for precise calculations this should be considered. For estimating the achievable sound pressure, the midrange average value will be sufficient.

The sound level Lp at the listening position, on the acoustical axis of a monitor, at distance r to the monitor is expressed by

$$L_p = L_w + 10\log\left(\frac{2Q}{4\pi\, r^2}\right) + \left(\frac{4}{A}\right) \quad \text{[dB SPL]} \qquad (6)$$

Using these principles of calculation the peak short-term output level and the sustained long-term output level can be estimated based on data given by monitor loudspeaker manufacturers once the room volume and reverberation time have been modelled or measured.

Monitor output capacity

As the thermal capacity and heat conduction of a dynamic driver voice coil is not large, the long term sound output level of an active monitor is normally limited by protection systems to prevent the driver voice coils from overheating [13,14,15,16]. Because of this, even if monitor loudspeakers typically have a very large short-term sound output capacity, they have clearly lower capacity to sustain long-term sound output.

Activation of the protection lowers the highest sound level and this can also change the tonal character or fidelity of the sound output.

One fundamental aim of active monitor design is to avoid activation of the protection under all normal operating conditions. This design requires careful balancing of the short term capacity with the long term capacity. This balancing is based on assumptions about the operating conditions as well as the nature of the acoustic signals that are monitored.

The assumptions typically include (1) the sound pressure normally used in monitoring, (2) overhead sound output capacity needed for the maximums or peaks in the sound output, (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.

In the following we discuss particularly the last item in this list.

Temporal level variability in audio

Activation of the protection lowers the maximum sound level and this can also change the tonal character or fidelity of the sound output. The aim of an active monitor design is to avoid activation of protection under all normal operating conditions. In order to do this, it is useful to study the characteristics of the typical audio signal.

Audio signal value distribution has been studied as change in this value distribution over time, particularly for music recordings. A typical statistical value analysis for film sound tracks shows the highest likelihood at –20... –15 dB of full scale [26]. This coincides with the way the headroom is typically set for film sound.

The value distribution data does not give any information about the temporal character of the audio data, i.e. for how long large levels typically last and how long are the pauses between high level peaks. This information would be very useful for understanding the performance requirements for active monitors and subwoofers.

It is important to note here that any length of audio signal is a valid audio signal, also a signal that has unlimited length in time. However, such signals are not very typical, particularly at very high output level.

The purpose of the following discussion is to review the typical temporal properties of audio in film sound tracks. The film sound tracks used as material represent several genres of films and are relatively recent multichannel productions (Table 1).

	year	type	ori.	duration h:mm:ss	size GB	no. chap.	title
A	2004	action	US	1:50:03	5.31	39	I Robot
B	2006	animation	US	1:22:43	3.99	28	Open Season
C	2009	drama	FI	1:12:03	3.47	12	Postia pappi Jaakobille
D	2006	drama	FR	1:41:40	4.90	18	Science of Sleep

Table 1: Material used in the study

The multichannel soundtracks were extracted from DVD releases of the films. The soundtracks contain a 5.1 channel sound presentation with 48 kHz sampling frequency and 24 bit sample precision (a track sample is shown in Figure 2). The data was analyzed at original, full dynamic range and original sample rate. Also the LFE channel was included in the extracted data. In total, the material includes 6.3 x 10^9 (billion) samples.

The motivation for using the DVD release sound tracks is that the word length and sample rate of this test material is the same as the word length and sample rate of most film releases with immersive 3D presentations. The test material models well the *bed* audio presentation for 3D immersive audio.

We can also assume that the audio signal levels seen in these multichannel presentations are representative also of the levels of the object channels or height channels in immersive audio presentations. This is because, once the audio content

exists in the height channels and the object channels, the relative sound level that must be presented from there to the listening location is bound to be similar.

Finally, actual 3D sound track material was not available for the study.

Source materials A and B are action films with relatively loud passages containing effects, Foley, and a significant orchestral background besides the dialogue.

Source material C represents the Nordic drama style of film making, with the dialogue as the central element, and minimal Foley and orchestral elements. This material contains a lot of very low sound level passages, but the whole dynamic range of the 24 bit data presentation has not been employed, probably due to reasons of the recording and production technology.

Source material D represents a French drama film, with very rich dialogue, Foley, and significant orchestral score.

While we wanted to include several types of film sound material for the study, it was anticipated that the action type films would present the highest audio levels and widest frequency range requirement.

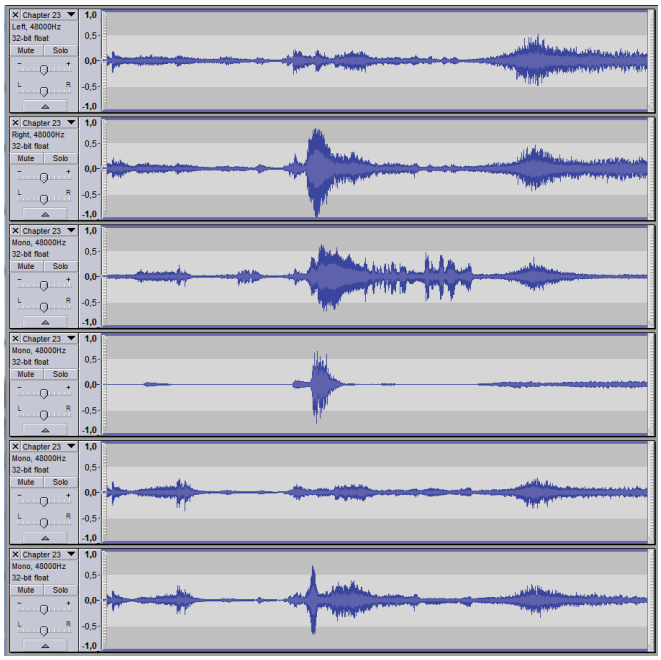

Figure 2: Example of the extracted multichannel sound track; from top to bottom: left, right and centre front channels, LFE channel, and left and right surround channels

Value distribution

The value ranges in audio tracks were studied by measuring the histograms of the signal envelopes (Figure 4), enabling precise envelope estimation also at high frequencies, fast transitions and short peaks. The envelope has been extracted as the magnitude of the analytic signal [27]. The analytic signal is a complex-valued signal consisting of the original real-valued signal and its Hilbert transform (equations 7-9).

The histogram was measured in bins having one dB width across the dynamic range of the 24-bit PCM presentation. The histograms contain the whole audio track of the film.

The two Hollywood productions (materials A and B) use high levels and hit close to the full scale more frequently than the European productions (C and D). Material C has the largest headroom, and consists dominantly of dialogue, containing very little effects and music.

While in material A all three front channels are used equally in terms of level, other productions use the centre channel more than the left and right front channels. This is possibly a sign of the dialogue placed mainly in the centre channel having a higher level in materials B, C, and D. The LFE signal seems to be usefully exploited also at low level in material A whereas the other materials use the LFE much less.

The full scale of the dynamic range is hit in material A so precisely that it must be assumed that the sound track has been digitally maximized to the full scale, also evidenced with a tiny amount clipping in the all the audio channels. The other materials do not show this. Other materials rarely show values in the last 10 dB below the full scale, with the centre channel content in the highest level. The least loud material C has a 6 dB headroom in the front left and right channels, and a 2 dB headroom in the centre channel.

Surround maximum levels remains systematically 5...10 dB lower than the front triplet level. When used, the LFE level peaks typically to the full scale. The LFE is usually played back 10 dB higher than the main channels.

Peak level distribution

The aim is to evaluate the frequency and length of peaks exceeding a set detection level L. In this way it is possible to evaluate (1) how likely a peak exceeding the detection level is to occur, (2) what is the expected length of the peak, (3) what is the expected time between repetitions of peaks.

Audio signals can be short or infinitely long but in typical audio material, high level peaks have finite length. When evaluating the signal level, time windowing has been used to extract an average of level over a certain period of time, for example in [26]. Applying a time window tends to reduce the level of short duration peaks, and particularly considering the high frequency radiator channel, accurate estimates of the signal maximum level are not obtained.

In order to avoid attenuating short peaks, we can look at the magnitude $y(n)$ of the signal $x(n)$ directly, or the envelope $h(n)$ of the signal (Figure 3).

$$y(n) = |x(n)| \tag{7}$$

$$h(n) = |x(n) + j\,\hat{x}(n)| \tag{8}$$

$$j = \sqrt{-1} \tag{9}$$

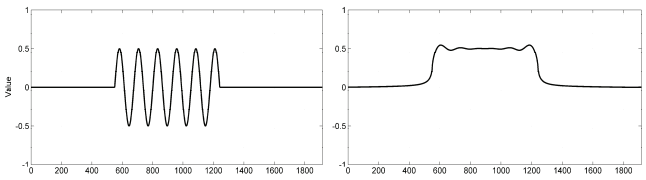

Figure 3: A sample burst signal $x(n)$ (left) and its envelope $h(n)$ (right)

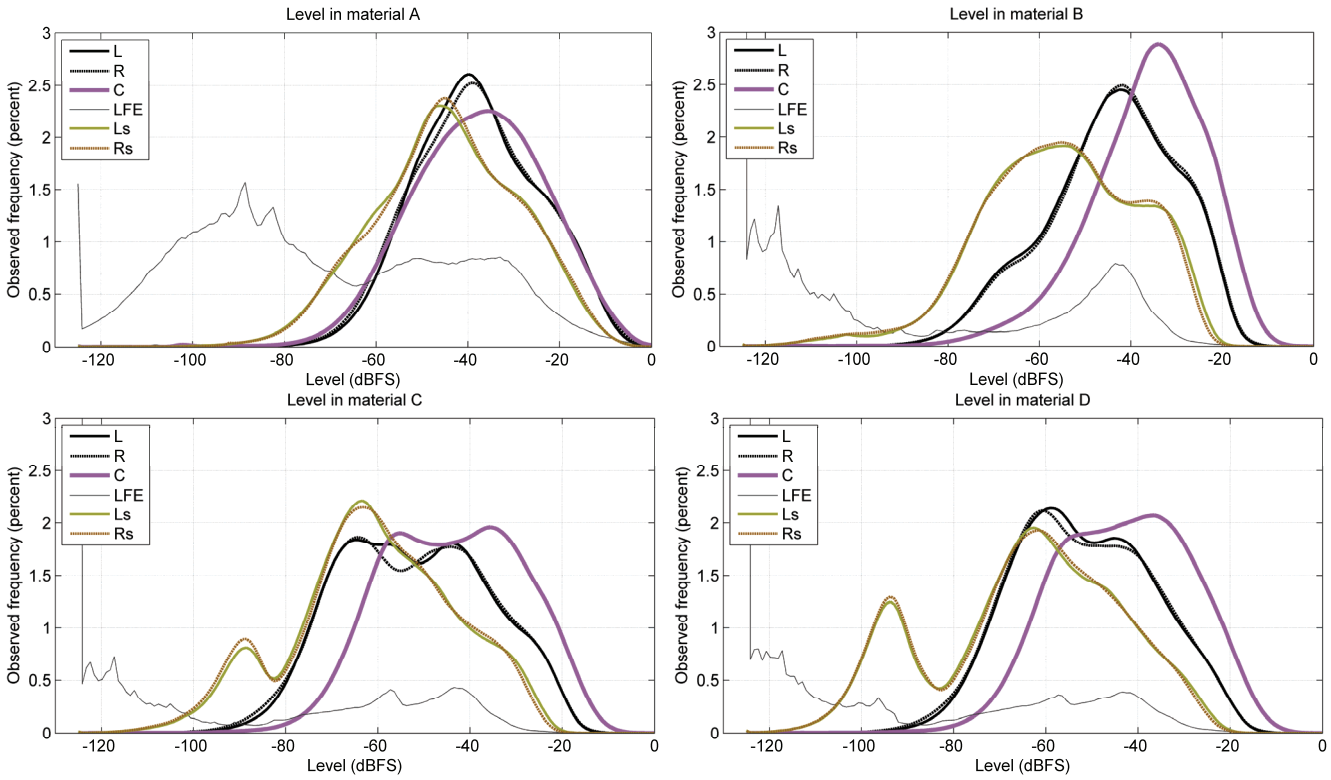

Figure 4: Value histograms for materials A (top left), B (top right), C (bottom left) and D (bottom right). Each histogram contains the whole duration of the audio track in the material. The bin spacing is 1 dB. Each histogram contains more than 10^9 samples.

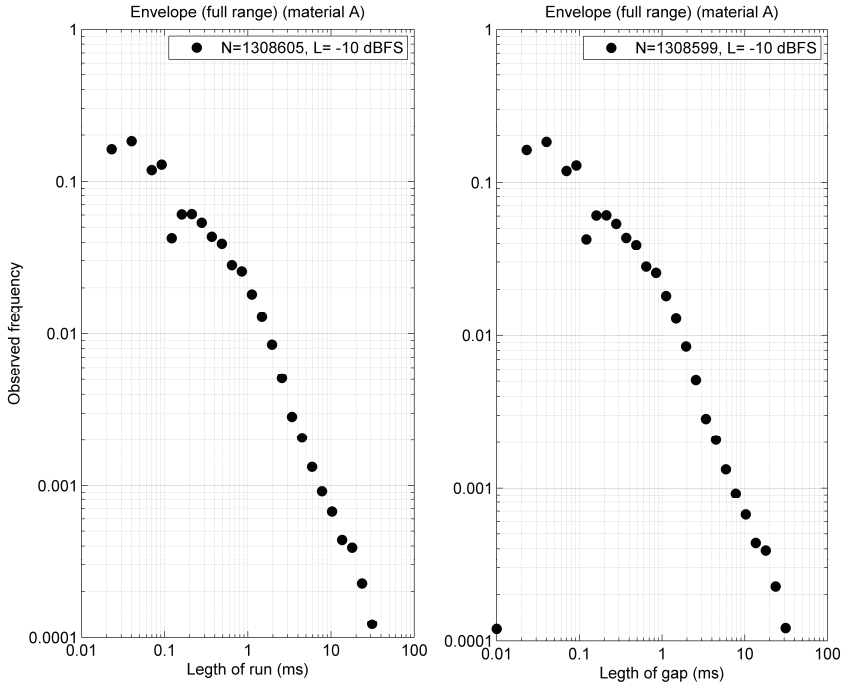

Figure 5: Lengths of peaks higher than L= –10 dBFS in the full bandwidth signal envelope. All audio channels in the material A are included. Material A has the highest likelihood of peaks values of the studied materials.

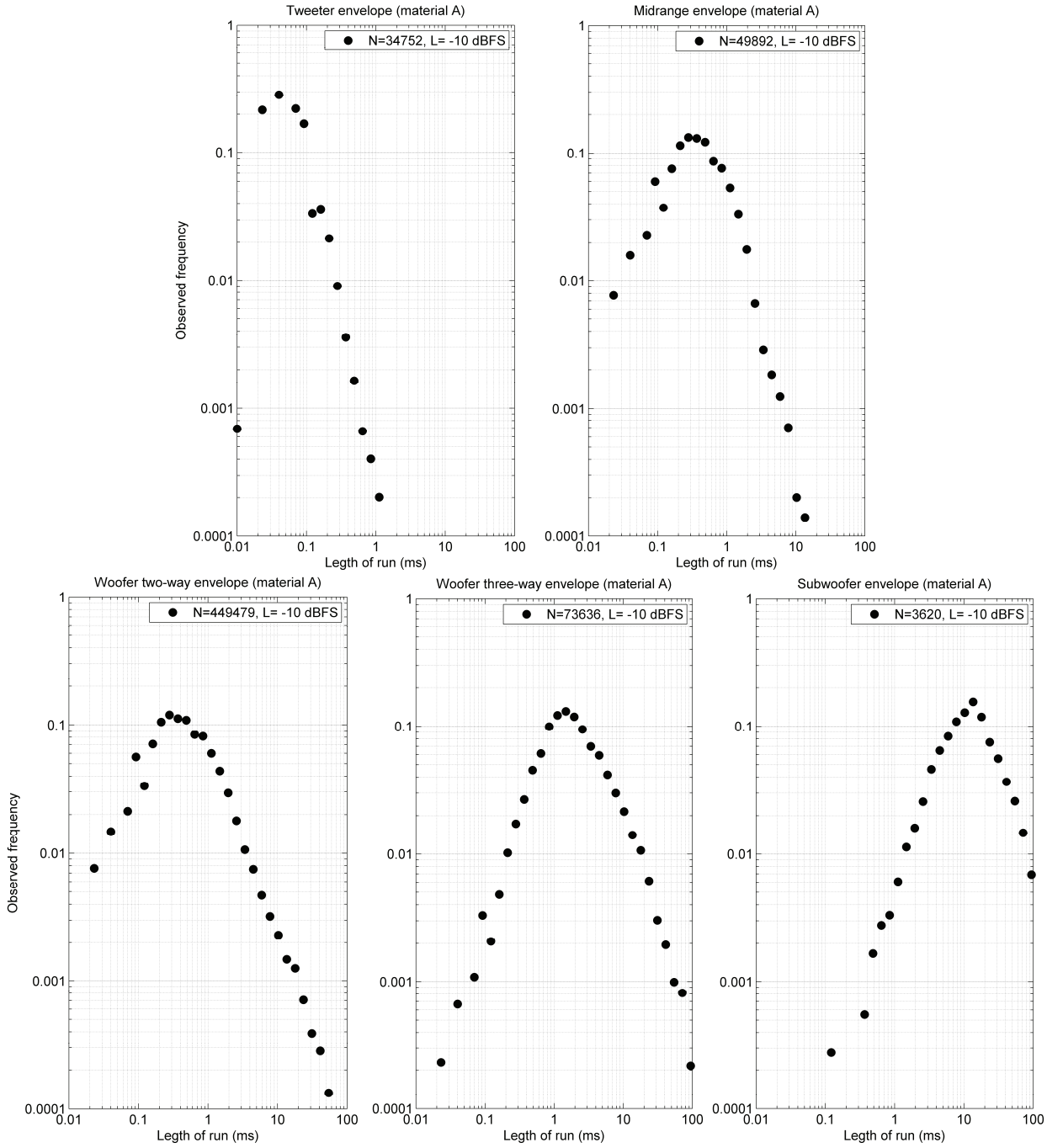

Figure 6: The observed frequency versus the length of peaks higher than L= –10 dBFS, in envelopes of driver-specific bandwidth signals. All audio channels in the material A are included. Material A has the highest likelihood of peaks values of the studied materials. Bandwidth limitation is explained in the text.

The duration of the output peak levels was studied for peaks exceeding the level $L = -10$ dBFS in the full bandwidth envelope of the signal (Figure 5). The data is presented as histograms. The likelihood of peaks increases with decreasing peak length. The gap between peaks follows similar statistic. Looking at the likelihoods of peaks and the gaps between peaks, the expectation is that a 99% of the peaks have a length less than 4 ms. Peaks longer than 10 ms are infrequent in our material but do occur.

The immediate conclusion is that high level long duration peaks are unlikely to occur. While a continuous signal is a perfectly normal audio signal, it is not very likely to occur at a very high level, close to the maximum output level.

The peak statistics changes after the signal has been filtered with bandpass filters being a part of an active monitor or subwoofer crossover filtering.

The crossover filters were modelled with fourth-order Linkwitz-Riley frequency responses. The crossover cut-off frequencies were modelled at 90 Hz for the subwoofer, 500 Hz and 3 kHz for the three-way monitor model, and 3 kHz for the two-way monitor model. These frequencies represent well the typical monitoring applications seen on the market.

As expected, the frequency bandwidth affects the duration distribution of the peaks (Figure 6). While the peaks in the tweeter channel are typically 40 μs in length and seldom more than 1 ms long, the peak length in the subwoofer channel is about 10 ms and can be several hundred milliseconds in length. If we take the 1% occurrence level as the criterion, meaning that 99% of peaks are shorter in duration than a limit, then these limits are given in Table 2. The figures given do not imply that runs longer than the limits could not occurs or that one cannot produce a sound track that would have a higher incidence of long runs, but the limits given do give an idea of the typical situation we can expect.

crossover channel model	passband (Hz)	run length 1% limit (ms)	relative occurrence (percent)
tweeter	> 3 k	0.4	7.7
midrange	0.5 k - 3 k	5	11
two-way woofer	< 3 k	15	100
three-way woofer	< 0.5 k	40	16.4
subwoofer	< 90	300	0.8

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

Discussion

While it is not a great concern for theatrical presentations where the listener's distance to the loudspeakers is large, the physical size of the loudspeaker becomes a serious consideration for the compact monitoring rooms. The room designer must consider the loudspeaker's influence to the acoustic performance of the room particularly for immersive audio reproduction systems where a large number of loudspeakers can create acoustically significant reflecting area in the room. While avoiding the acoustic reflection and working in smaller-sized rooms imply using physically small loudspeakers, high maximum sound pressure requirement combined with the requirement that the frequency bandwidth extends to low frequencies leads to selection of physically large monitoring products. Bass management can offer at least a partial solution to this, particularly in small monitoring rooms, if the use of bass management is acceptable.

Certain cinema standards present requirements on the maximum sound level capacity of the audio reproduction system. An example of these is the Dolby Atmos, specifying the SPL at the reference listening location.

In this paper, an attempt has been made to describe the methods to consider the variables in monitoring room design when selecting the acoustic performance of the monitoring loudspeakers and subwoofers. Calculation methods to estimate the sound pressure level at the listening location have been given for a case where the physical size and acoustical characteristics of a room are known or have been estimated using simulations. These have been described for the typical case where the monitors are located close to one acoustically hard surface, typically a wall or ceiling. The method to estimate the sound level at the listening location has been provided as a collection of formulas. The use of the formulas requires some data describing the characteristics of a particular monitor. This data can typically be obtained from the manufacturers or with measurements.

The material studied demonstrates that the surround channels are operating at about 5...10 dB lower maximum level than the front monitor triplet. Most likely this also applies to the height output channels in immersive audio presentation systems, too, due to the tendency to concentrate a lot of the action in the front at the screen in cinema mixes.

The realistic duration at a high output level was studied using the multichannel sound track material on DVD records. It can be assumed that the characteristics of this material can be extended realistically also to 3D immersive audio systems because the method of building the 3D presentation uses heavily the standard multichannel audio presentation on the ear height of the listener.

The statistical distribution of the high level sound event duration in a full bandwidth cinematic immersive audio track envelope is less than 4 ms when all audio channels are considered collectively. In a multiway monitor and subwoofer the high sound level event duration depends on the frequency band, with the highest occurrence of high levels in the two-way woofer channel. The benefits of a three-way design over a two-way design were evident as a lower occurrence of high levels in each driver channel.

The high level audio event duration was estimated using modelled Linkwitz-Riley crossover filter bands at the typical passbands for active monitors and subwoofers. After crossover filtering the envelope peak level durations change to reflect the passband of the crossover filter. The statistical limits found in this work do not mean that longer duration maximum output levels could not occur, as any duration of the audio signal at any given frequency is completely normal and can be produced at will. The statistics merely describe the typical case in real material, and therefore can be taken as an indication of what typically will be required of a monitoring loudspeaker or subwoofer.

When evaluating the maximum SPL capacity of the reproduction system it would make sense to use test signals that resemble the typical signals that are reproduced. Attempts towards this direction have already been made, for example the Jiffy signal by Dolby and the use of non-continuous pulsed pink noise [5]. Signals with realistic characteristics should be preferred over using continuous technical test signals, such as sinusoids or uninterrupted pink or white noise, as the continuous signals will produce misleading estimation of the systems real-life performance and are likely to lead to severely over-specified and expensive system implementations not necessarily delivering better performance in real life.

References

[1] International Telecommunication Union (2015) Multichannel sound technology in home and broadcast applications. Report ITU-R BS.2159-7. http://www.itu.int/pub/R-REP-BS.2159-7-2015.

[2] International Electrotechnical Commission. IEC 62574: 2011, "Audio, video and multimedia systems – General

channel assignment of multichannel audio". https://webstore.iec.ch/publication/7218.

[3] DTS Inc. DTS Technology for Home Theatre, DTS Neo:X and Neural Surround. http://www.dts.com/professionals/sound-technologies/audio-processing/dts-3d-audio.aspx.

[4] Jürgen Herre, Johannes Hilpert, Achim Kuntz, Jan Plogsties (2014) MPEG-H Audio, The New Standard for Universal Spatial/3D Audio Coding. *Proc. AES 137 Convention, Los Angeles.*

[5] Dolby Inc. (2015) Dolby Atmos Specifications, Issue 3. http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-specifications.pdf.

[6] Dolby Inc. (2015) Dolby Atmos Next Generation Audio for Cinema, Issue 3. Accessed 2015-07-05 at http://www.dolby.com/us/en/technologies/dolby-atmos/dolby-atmos-next-generation-audio-for-cinema-white-paper.pdf.

[7] Kimio Hamasaki (2011) 22.2 Multichannel Audio Format Standardization Activity. *Broadcast Technology* No.45, Summer 2011. 14-19.

[8] K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando (2008) A 22.2 Multichannel Sound System for Ultra-High-Definition TV (UHDTV). *SMPTE Motion Imaging J.*, Vol. 117, No. 3, 40-49.

[9] Bert Van Daele, Wilfred Van Baelen (2012) Productions in Auro-3D, professional workflow and costs. White paper. http://www.auro-technologies.com/wp-content/uploads/documents/Professional-Workflow-White-Paper-v0-6-20120228.pdf.

[10] Francis Rumsey, David Griesinger, Tomlinson Holman, Mick Sawaguchi, Gerhard Steinke, Günther Theile, Toshio Wakatuki (2001) AES TC-MBAT Information Document: Multichannel Surround Sound Systems and Operations. http://www.aes.org/technical/documents/AESTD1001.pdf.

[11] Dolby Inc. (2013) Authoring for Dolby Atmos Cinema Sound Manual, Issue 1. http://www.dolby.com/uploadedFiles/Assets/US/Doc/Professional/Authoring_for_Dolby_Atmos_Cinema_Sound_Manual(1).pdf.

[12] International Telecommunication Union (2012) Multichannel stereophonic sound system with and without accompanying picture. https://www.itu.int/rec/R-REC-BS.775/en.

[13] Aki V. Mäkivirta (2008) Loudspeaker design and performance evaluation. In David Havelock, Sonoko Kuwano, Michael Vorländer (ed.) Handbook of Signal Processing in Acoustics. Springer. 649 – 667.

[14] Klippel GmbH (2015) Thermal Parameter Measurement, Application Note 18. http://www.klippel.de/fileadmin/_migrated/content_uploads/AN_18_Measurement_of_Linear_Thermal_Parameters.pdf.

[15] Moo-Yeon Lee, Hyung-Jin Kim (2014) Numerical Investigation on the Temperature Characteristics of the Voice Coil for a Woofer Using Thermal Equivalent Heat Conduction Models. *Entropy*, 16, 4121-4131.

[16] Clifford A. Henricksen (1987) Heat-Transfer Mechanisms in Loudspeakers: Analysis, Measurement, and Desig. *JAES*, 35 (10) 778-791

[17] European Broadcasting Union (2011) Practical guidelines for Production and Implementation in accordance with EBU R 128 EBU – TECH 3343. Geneva. https://tech.ebu.ch/docs/tech/tech3343.pdf.

[18] European Broadcasting Union (1998) Listening conditions for the assessment of sound programme material, EBU Tech Doc 3276-E and its supplement 1 (2004).

[19] Wilfried Van Baelen, Tom Bert, Brian Claypool, Tim Sinnaeve (2015) Auro-3D, A new dimension in cinema sound. http://www.barco.com/projection_systems/downloads/Auro-3D_v3.pdf

[20] Günther Theile, Helmut Wittek (2011) Principles in Surround Recordings with Height. Proc. 130 Conv. AES.

[21] International Telecommunication Union (2015) Methods for the subjective assessment of small impairments in audio systems, BS.1116-3. https://www.itu.int/rec/R-REC-BS.1116/en.

[22] Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, James V. Sanders (2000) Fundamentals of Acoustics, 3rd ed. John Wiley, New York.

[23] Esben Skovenborg (2012) Loudness Range (LRA) – Design and Evaluation. Proc. 132 Conv. AES.

[24] Aki V. Mäkivirta and Christophe Anet (2001) A Survey Study Of In-Situ Stereo And Multi-Channel Monitoring Conditions. Proc. 111 Conv. AES.

[25] Dolby Inc. (2003) Dolby 5.1-Channel Music Production Guidelines, Issue 2. http://www.beussery.com/pdf/beussery.dolby5.1.pdf.

[26] Miomir Mijić, Draško Mašović, Milan Petrović, Dragana Šumarac-Pavlović (2009) Statistical Properties of Music Signals. Proc. AES 126 Conv.

[27] Julius O. Smith (2007) Introduction to Digital Filters with Audio Applications. W3K Publishing.