Is single microphone position enough for immersive system equalization and level calibration in production monitoring?

A. Mäkivirta[1], T. Lund[1]

[1] *Genelec Oy, Iisalmi, Finland. Email: aki.makivirta@genelec.com*

Abstract

Immersive monitoring systems contain a large number of loudspeakers, each strongly and differently influenced acoustically by the room boundaries. This creates uncertainty regarding the actual in-room frequency response, results in large acoustical differences between individual loudspeakers, and can reduce the transparency and reliability of monitoring on such system. The aim of this work is to demonstrate this influence and then study the importance of using multiple measurement microphone positions instead of one measurement microphone position when preforming in-room frequency response calibration and system alignment. An NHK 22.2 immersive monitoring setup demonstrates the variability typical of immersive layouts, and the influence individualized in-room monitoring loudspeaker equalization can bring. After this, various spatial distributions in the vicinity of the main listening location in 12 rooms are compared to the typical situation of using single microphone location at the listening location. When using equalization to remove response differences between loudspeakers, no significant differences were found between single point measurements and spatial averages for small spatial averaging displacements in rooms having low reverberation times.

1. Introduction

Loudness-based production in broadcast relies on accurate monitoring for judging balance, formats, speech intelligibility, and audio quality. Reproduction systems for monitoring immersive audio presentations [1],[2] share certain assumptions about the monitoring system setup. These include flat frequency response at the listening location, the same level and time of flight from all monitors at the listening location, and normalized monitoring level [3],[4]. Room-related acoustic effects modify frequency responses of loudspeakers. For immersive monitoring layouts, these loadings are typically unique to loudspeaker locations. Because of this, loudspeakers for critical monitoring are equalised to reduce the acoustical room influence to improve accuracy in high quality professional listening [5].

The goal in professional monitoring is to hear the audio recording content without change. The monitoring system, the loudspeaker in the room, ideally approximates an allpass system within the audible frequency range.

$$\begin{cases} \left|H(e^{j\omega})\right| = 1 \\ -\frac{\mathrm{d}\arg\{H(e^{j\omega})\}}{\mathrm{d}\omega} = \text{const.} \end{cases} \tag{1}$$

Then, the monitoring system does not change the spectral content and merely causes a constant delay at all frequencies while delivering the audio for the engineer's ears. Recommendations consequently call for flat in-room response at the listening location, constant reverberation time in the room for audible frequencies, and sufficiently short room reverberation to not interfere excessively with the direct sound radiated by the loudspeaker. Minimum-phase implementations of loudspeakers lead to group-delay variation which, however, can be compensated to fulfil Equation 1 [6]. As loudspeakers are three-dimensional radiators, recommendations for directivity characteristics are also given, e.g. [5].

The obvious method for improving the sound colour at the listening location is electronic equalization. This is not a new idea. H. Tremaine's *Audio Cyclopedia* claims that RCA's John Volkman was the first to use electronic filters for room response equalization already in the 1920s [7]. Adjustable filters for room equalization have been available in the vacuum tube era (e.g. [8]). Parametric equalizer filters were introduced in the 1970s [9],[10] and digital signal processing has made both measurements and room equalization easy and precise.

While there is a consensus that a neutral loudspeaker has a flat anechoic frequency response, some researchers have questioned the suitability of the allpass target for the in-room frequency response, suggesting a down sloping character across the audio band. However, this largely seems to be motivated by listener preference [11],[12],[13],[14],[15],[16],[17],[18] while possibly too little attention has been paid to solving the problem with the 'circle of confusion'. This refers a self-referenced system where existing recordings are used to evaluate the room sound [18],[19].

1.1. One or more measurement locations

The present work concentrates on enabling reliable measurement of the in-room frequency response at the listening location, but not on the preferred in-situ response, which we regarded a separate issue.

The sound pressure measurable in a room at a single microphone location is local. At high frequencies the typical notching due to multi-way propagation and reverberation in rooms particularly at high frequencies varies when the microphone location changes. This effect is usually reduced by applying in-frequency smoothing with a sliding variable-width averaging window to the frequency response. Part of the late reverberation may also be rejected using a time domain window on the impulse response but this leads to loss of low frequency information. Frequency smoothing reduces acoustic comb filtering and locality of measurement while extracting subjectively important low Q colorations in the midrange frequencies.

Spatial averaging of responses at multiple microphone positions can extract the common frequency response features visible in the measurement points. Spatial sampling of the room response in vicinity of the listening location has been suggested to improve the reliability of measurement and is assumed to produce a better representation of room acoustics over single point measurement, and thereby be more reliable starting point for system equalization.

Spatial sampling is done with several motivations. In the context of large seating area applications, such as cinema theatre equalization, even spatial sampling across the floor area aims to accomplish the best average equalization [5]. Similarly, spatial sampling can create information about the average coloration within an area, such as the large mixing console.

1.2. Immersive audio

Immersive reproduction systems are intended to create a spatially stable presentation of at least the frontal virtual sound images for a wide seating area [20] by having a large number of real loudspeaker sources correctly localizable at all seating position. Virtual sound images in immersive reproduction systems are created using two, three or more loudspeakers together, and localize the virtual images within the area span by these speakers, even if location error occurs for off-centre listening positions.

The combination of equalizing individual loudspeaker responses and careful alignment of playback levels and delays can reduce acoustic differences and can improve virtual sound image precision. Controlled in-room spectral response is the foundation of meaningful level calibration [4].

Immersive reproduction systems place loudspeakers on multiple height layers around the listener. The acoustical influence of the room differs more than for stereophonic or single-layer multichannel (surround) layouts, where it is easier to design room acoustics affecting the monitors to be similar. Increasing the number of height layers and loudspeaker channels complicates the acoustical playback system and increase acoustical differences between monitors.

1.3. Paper structure

The purpose of this work is to study the sufficiency of the single microphone position acquisition of the acoustical response at the listening position. Frequency domain smoothing (averaging in frequency) may in itself be able to increase spatial stability in a measured frequency response for features relevant to subjective experience of sound colour and for compensating the sound colour to reduce such colorations. First, a case study of the effects of electronic room calibration of an immersive NHK 22.2 system is presented. Then, research reported partially in [21] and [22] is revisited and additional analysis is provided. This study uses stereo loudspeaker setup in several monitoring rooms to study the benefit of spatial averaging for room calibration. The aim in both parts is to demonstrate the magnitude response at the nominal listening location. The study has been conducted in the context of professional monitoring spaces, typically having relatively low reverberation time and low level of the early reflections at the listening position.

Figure 1a. Method of suspending the upper layer monitors.

Figure 1b. Method of installing the lower layer monitors.

Figure 1c. Side view to the 22.2 immersive speaker layout.

2. Case study

A three-layer immersive monitoring system was set up according to the NHL 22.2 format. The installation room reverberation time in mid frequencies (f > 250 Hz) is 0.25 s on average and at low frequencies (f < 200 Hz) 0.5 – 0.8 s, increasing towards low frequencies. The room has an isolated prominent modal resonance at 44 Hz.

The loudspeaker system width is 3.22 m (from FL to FR, as well as from BL to BR). The channel name abbreviations are according to the SMPTE 2036-2 [2],[20]. The room height is 2.5 m, width 5.56 m and length 7.2 m. The listening position (measurement position) is 2.8 m from the left and right walls and 2.70 m from the front wall. The microphone height is 1.2 m. The upper layer monitors are suspended on rails attached to the ceiling (Figure 1a-1c). This places the upper layer loudspeaker front at 2.16 m height. The median distance to loudspeakers is 2.28 m, ranging from 0.96 m (TpC, room height limitation) to 3.01 m (TpFL). The time-of-flight (ToF) variation range was 4,97 ms (5,96 ms when TpC is included), see Table 1.

Table 1. Loudspeaker arrangement for the case example. Loudspeaker distance (DIST) and distance-related time-of-flight difference (DTOF) are given.

ROLE	TYPE	LAYER	DIST. (cm)	DTOF (ms)
BtFC	8330A	lower	211	2.62
BtFL	8331A	lower	257	1.28
BtFR	8331A	lower	261	1.16
BC	8250A	middle	130	4.97
BL	8240A	middle	208	2.70
BR	8340A	middle	209	2.67
FC	8330A	middle	213	2.56
FL	8330A	middle	272	0.84
FLc	8351A	middle	230	2.06
FR	8330A	middle	265	1.05
FRc	8351A	middle	230	2.06
SiL	8430A	middle	248	1.54
SiR	8430A	middle	226	2.18
TpBC	8320A	upper	139	4.71
TpBL	8330A	upper	223	2.27
TpBR	8330A	upper	214	2.53
TpFC	8330A	upper	247	1.57
TpFL	8330A	upper	301	0.00
TpFR	8330A	upper	292	0.26
TpSiL	8320A	upper	233	1.98
TpSiR	8320A	upper	214	2.53
TpC	8320A	upper	96	5.96
Sub L	7380A	sub	280	
Sub R	7370A	sub	280	

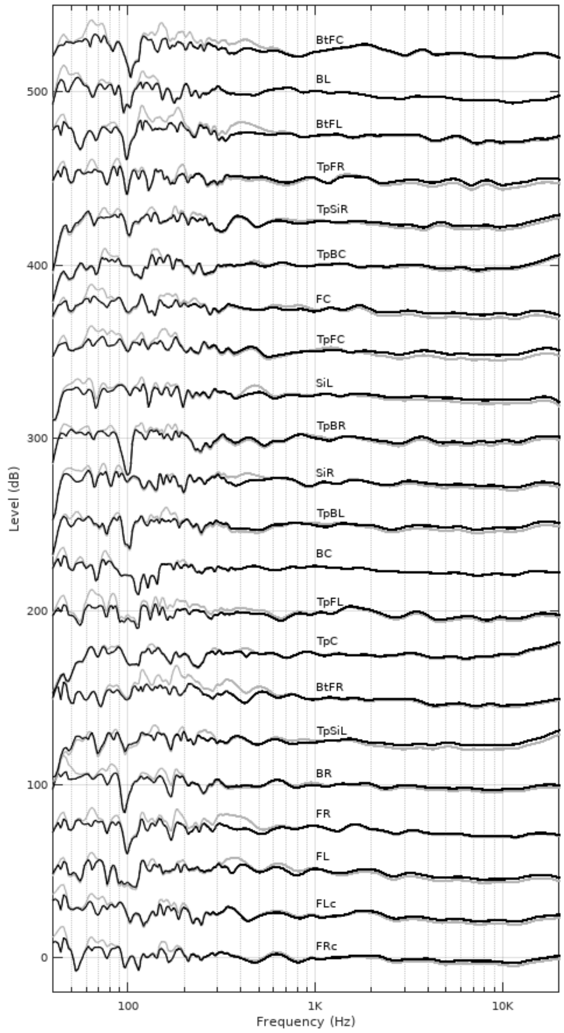

Figure 2. All response before (grey) and after equalization (black), aligned at 0.5 – 2 kHz and staggered by 25 dB to aid visibility.

Loudspeakers were measured at the listening position. In each loudspeaker, 6…10 second-order parametric notch filters and two low-frequency shelving filters were optimized for frequencies below 2 kHz, to reduce peaking above the reference level (mean level on log(f) scale in the range 0.5 – 2 kHz).

The outputs were adjusted to give the same mean output at all speakers. The times-of-flight were set the same using electronic delays in speakers.

Figure 3 presents shows the bi-smoothed loudspeaker responses (1/24 smoothing f < 200 Hz, then 1/3 octave). The 1/24 octave smoothing gives more detailed presentation of the SPL variation due to room modes at low frequencies. There, large level variation is typical due to sparse modal resonances compared to higher frequencies.

The maximum and mean levels recorded (Fig. 5, N=22) have been reduced. The audibility of mid and high Q peaks is higher than similar notches [23],[24],[25]. Equalizing the peaks reduces audible colorations and thereby renders loudspeaker responses more similar.

The equalization does not significantly change the amount of level shifting needed to bring the loudspeaker levels in alignment to the reference levels (Fig. 4).

The magnitude equalization reduces values exceeding the reference level, seen in reduction of the maximum and median response level (Fig. 5).

Looking at the pooled values exceeding the reference for all monitors (N=22), before equalization the level distribution shows a higher incidence of large deviations toward high pressure from the reference level (Fig. 6) than after the equalization (Fig. 7).

The case example demonstrates typical acoustical effects and level shifting seen in realistic, slightly acoustically suboptimal monitoring rooms. Equalization, level alignment and time-of-flight alignment can reduce acoustical differences and help create a predictable and reliable monitoring system.

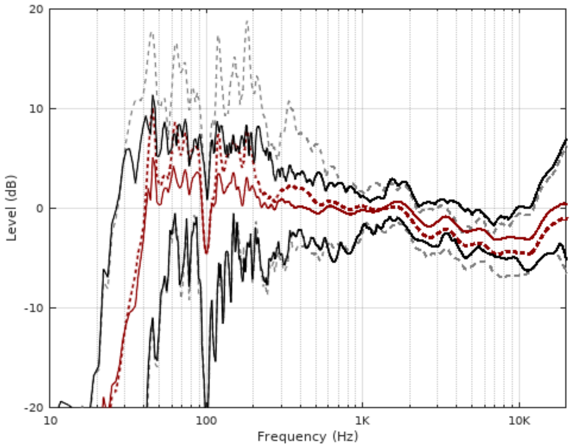

Figure 3. Minimum, maximum and mean responses after aligning the mean value calculated in log(f) in the range 0.5 – 2 kHz (N=22). The level scale is relative. The responses have been smoothed with a 1/3 octave window above 200 Hz and with a 1/24 octave window below this frequency.

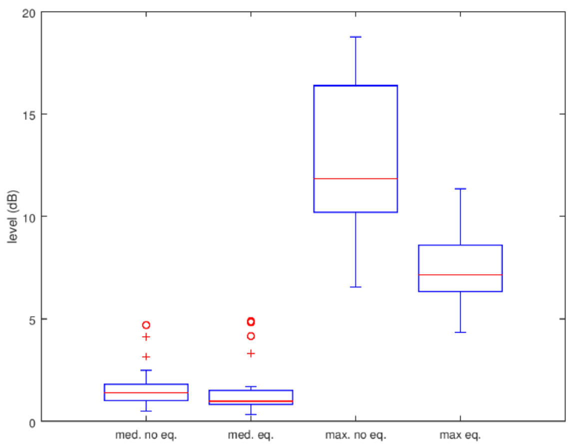

Figure 4. The median and maximum of the values exceeding the reference level, before and after calibration.

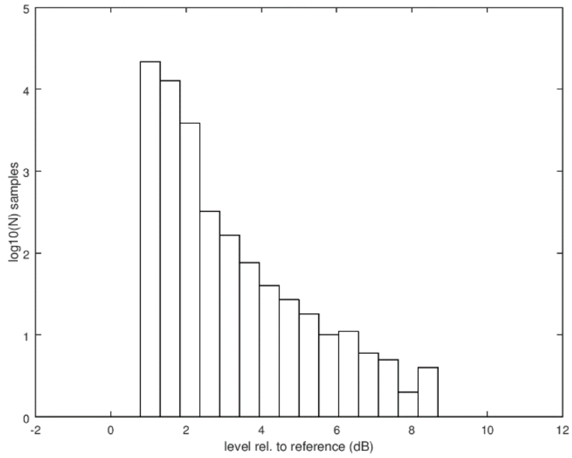

Figure 5. Histogram of the values exceeding the reference level before equalization (N=22).

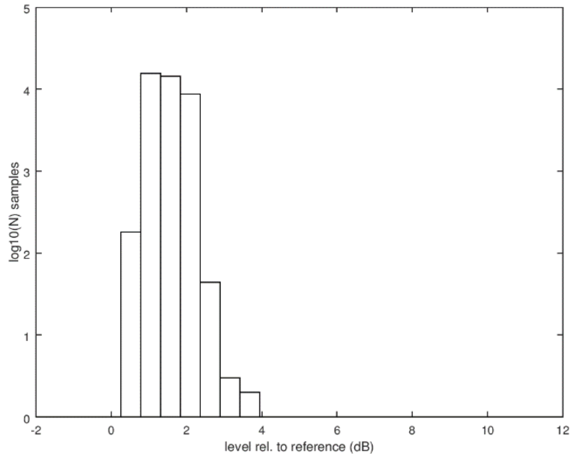

Figure 6. Histogram of the values exceeding the reference level after equalization (N=22).

3. Sufficiency of single measurement

3.1. Method

This section studies the ability of single position magnitude response measurement to render essentially the same information as spatial averaging. Methodology, choice of rooms and measurement data set used in this study has been reported in [21].

The measurement microphone has an electret capsule with omnidirectional characteristics. The main microphone position and 17 offset positions located at increasing distances are measured. The offset distances are 0.1 meters to the front, 0.1, 0.25, and 0.5 meters to the side and back, and 1.5 meters to the back, see Figure 1. Two-way loudspeakers were used (5 in woofer, 2/3 in tweeter, type Genelec 8330 or similar).

The impulse response is measured using a log sine sweep with length 256 kilo-samples (2.97 s, fs = 44.1 kHz). Two sets of 18 measurements are taken for each room. In total, 432 measurements in 24 series are taken in the study.

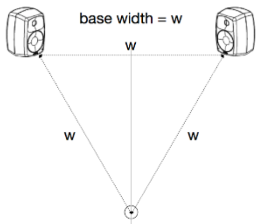

Figure 7. The principle of the loudspeaker layout follows the standard stereo pair placement. The main microphone position is marked.

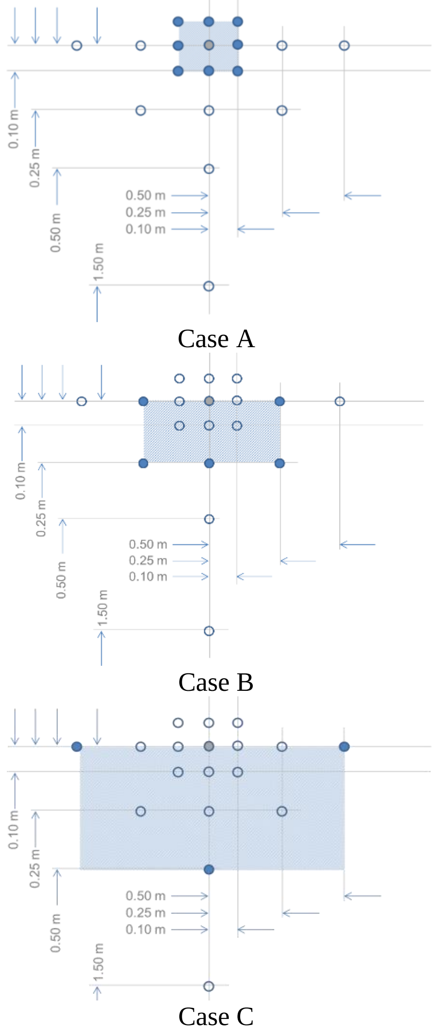

Figure 8. Definitions of the spatial average cases A to C showing the spatial averaging areas (shaded) and the numbers of measurement positions used (dark colour).

The magnitude response measurements are smoothed to standard 1/3 octave and 1/12 octave. Differences in distances are removed by level-normalizing responses. Measurements are therefore normalized in level using the mean response level across the 500 Hz to 5 kHz. The impulse responses are processed to generate (1) the main microphone position alone, (2) main position response averaged with the microphone responses at 0.1 m distance offset, (3) main response averaged with the responses at 0.25 m offset, and (4) main response averaged with the responses at 0.5 m offset.

Each spatial average response is calculated using the 1/3 octave smoothed data. All measurements in the spatial average receive the same weight. The weight of an individual measurement depends on the number of measurements included in the spatial average. After smoothing, the differences of the spatial averages to the single microphone position set at the listening location are calculated. This results in three sets of difference pressure graphs for each room, one for each spatial averaging case.

The mean pressure within an octave is calculated, resulting in 11 octaves from 16 Hz to 16 kHz. This results in 24 data points for each octave.

The spatial averages are compared to the single microphone position measurement. Measurements in all rooms and for all speakers are pooled. Box-and-whiskers plots are generated for each octave and for each spatial averaging distance. These describe the variability of deviation in the pressure from the single microphone position, in the octaves across rooms and loudspeaker position.

room	vol. (m3)	base width (m)	RT60 (s)	TER (ms)		LER (dB)	
				L	R	L	R
A	29.6	1.30	0.49	8.9	2.7	-11.9	-11.8
B	115.1	1.30	0.66	1.5	1.7	-10.5	-10.3
C	39.5	1.17	0.43	8.6	8.7	-10.8	-14.0
D	39.5	1.32	0.46	2.4	2.5	-13.5	-13.2
E	39.5	1.20	0.44	3.3	3.3	-15.9	-16.7
F	16.6	1.15	0.31	0.85	0.85	-10.5	-11.8
G	97.6	2.27	0.19	5.6	5.7	-10.8	-11.7
H	22.6	0.86	0.50	4.0	3.7	-16.7	-17.7
I	18.2	1.01	0.56	5.9	7.5	-17.2	-12.2
J	69.4	1.30	0.55	4.6	11.0	-14.8	-20.9
K	80	1.31	0.48	4.1	4.2	-17.2	-16.7
L	90.5	1.31	0.48	4.6	6.2	-20.9	-20.0

Table 2. Listening rooms used in the study. Reverberation time (RT60), highest early reflection level (LER) and the associated early reflection delay (TER) at the listening location for left (L) and right (R) loudspeakers.

3.2. Octave decay time in rooms

The twelve rooms share low background noise, acceptable reverberation time, and reasonably low level and acceptable direction of early reflections at the listening position (see Table 1).

The 12 rooms included in the study were evaluated for octave decay time. Three measurements were taken, at the reference location, 0.5 meters to the right and 1.5 meters to the back of the reference location.

The octave decay is compared to the recommendations [5], Fig. 9 and 10. The rooms fit recommendations for high resolution listening location in terms of the median octave decay time, but the quartile limits exceed slightly the recommended flatness for octave decay time.

Notably, the lowest octave evaluated (63 Hz) shows large variability and two rooms show large reduction in decay time at high frequencies. However, the rooms are not untypical of the acoustic in real monitoring environments [26], Fig. 11, making the data collected in these environments useful for practical evaluations of in-room acoustics and the effect of spatial sampling of responses close to the listening location.

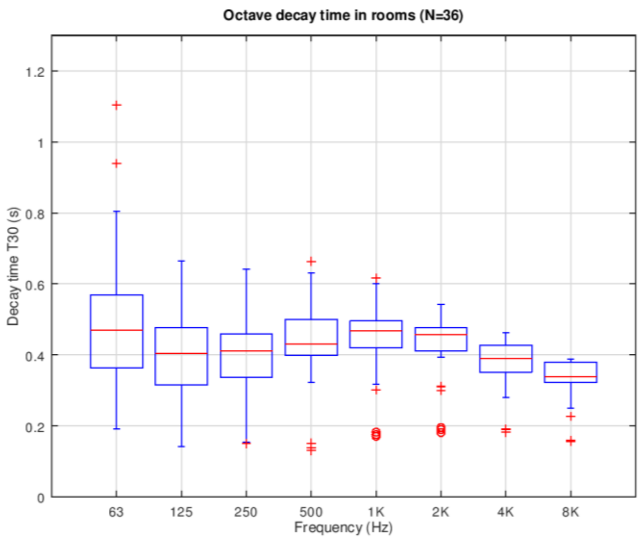

Figure 9. Distribution of the octave decay time in rooms. Three measurements were taken in each room.

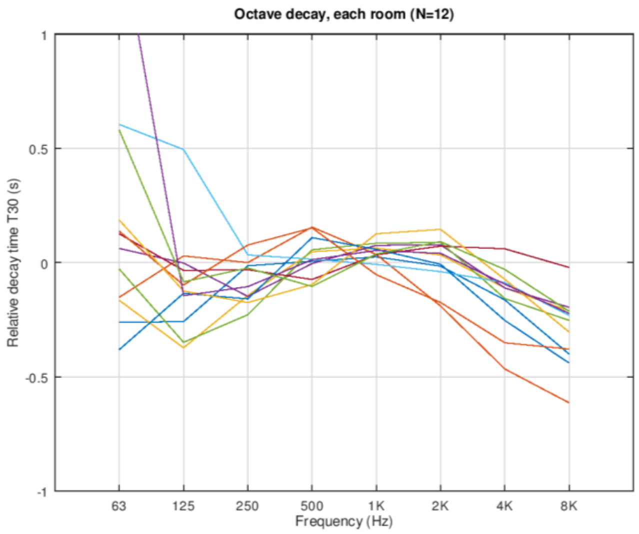

Figure 10. Deviations in octave decays in each room. Room average decay time in range [200Hz…4kHz] has been scaled to zero for each room, to enable comparison.

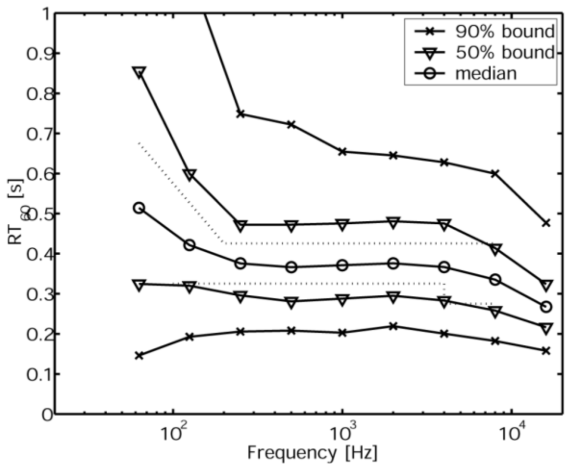

Figure 11. Distribution of decay time in professional monitoring rooms according to [26].

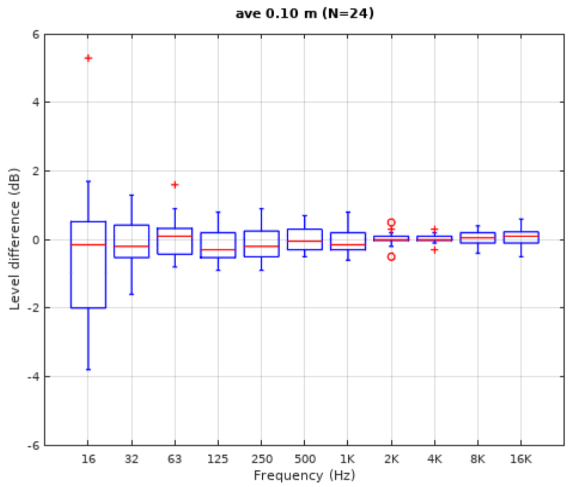

Figure 12. 0.1 m average vs. single measurement

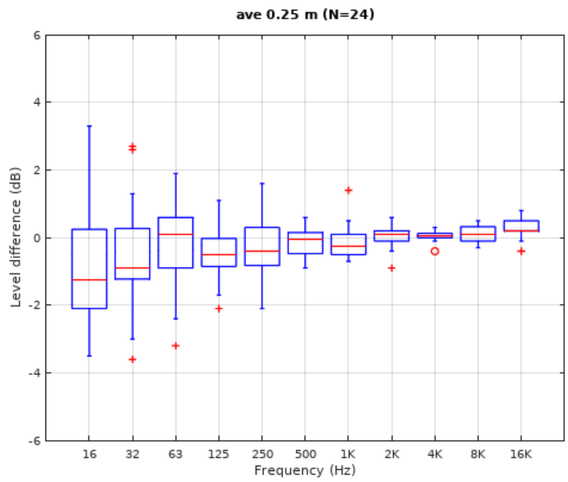

Figure 13. 0.25 m average vs. single point measurement

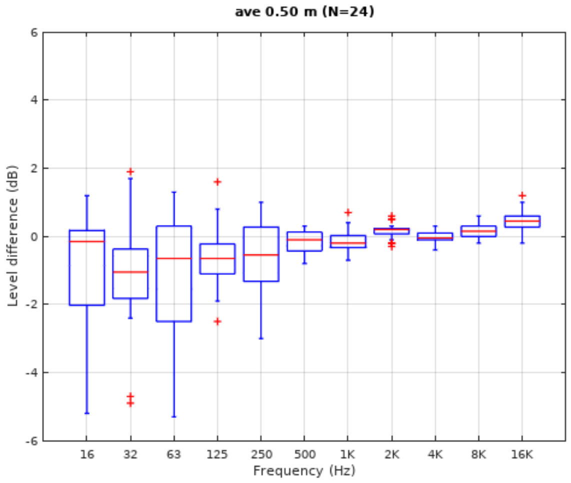

Figure 14. 0.5 m average vs. single point measurement

3.3. Frequencies f > 500 Hz

Differences to the single microphone position increase with increasing distance of the offset points.

The 0.1 m offset case (Fig. 10) shows small differences to the single measurement. The median difference remains less than 0.5 dB for all offset cases in frequencies f > 500 Hz. Quartiles, describing 50% of the deviation, are within ± 0.5 dB. Extremes (minimum, maximum in octave) remain less than ± 1 dB for f > 500 Hz. These are close to the just noticeable limit (JND) [24].

As a conclusion, using single measurement appears to result in similar equalization as using a spatial average within the spatial offset range not exceeding 0.5 m.

3.4. Frequencies f < 500 Hz

For low frequencies f < 500 Hz, acoustic interaction of the room with the loudspeaker radiation leads to larger room-dependent variability.

Spatial averages taken using smaller offset distance show smaller variability, as expected. The octave median levels across the rooms and speakers show good agreement with the single measurement position for small displacement spatial averaging (0.1 m).

For frequencies f > 63 Hz the medians in all spatial average cases remain in agreement with the single microphone measurement to ± 1 dB. The lowest frequency medians (32 and 16 Hz) show larger deviation, but even then, the deviation to single microphone measurement does not exceed 2 dB, close to the just noticeable limit (JND) [23].

Quartiles range covering 50% of data in each octave remains within ± 2 dB while extremes (minimum, maximum) are seen deviating by as much as 6 dB from the single point measurement at frequencies f < 125 Hz and this happens mainly toward smaller pressure, indicating that this may be related to notches in the frequency response at low frequencies. Notches at low frequencies are known to have poor audibility [23].

	spatial average offset		
	0.1 m	0.25 m	0.5 m
mean	0,074	0,190	0,327
std	0,67	0,89	1,02
min	-3,50	-2,65	-3,00
max	4,30	3,30	4,80

Table 3. Pooled octave band statistics across all rooms for the full audio bandwidth [21].

	spatial average offset		
	0.1 m	0.25 m	0.5 m
mean	0,0	0,016	0,011
std	0,16	0,27	0,24
min	-0,45	-0,80	-0,35
max	0,40	0,60	0,70

Table 4. Pooled octave band statistics across all rooms for the octave bands 500 Hz and higher [21].

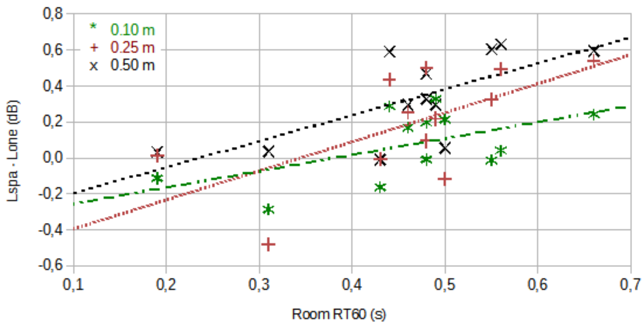

Figure 15. Mean level difference in octave bands, averaged across the audio band for the spatial offset cases 0.1, 0.25 and 0.5 m. The decay time is the maximum within the frequencies 0.4-4 kHz [21].

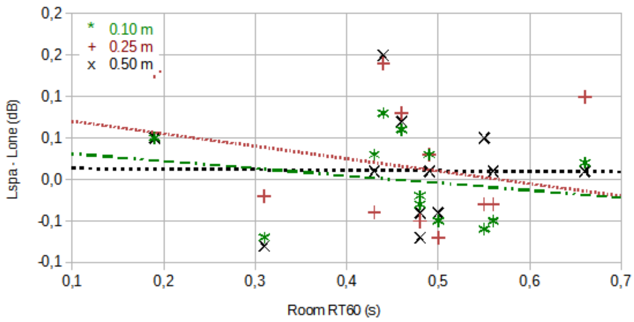

Figure 16. Mean level difference in octave bands 500 Hz and above, for the spatial offset cases 0.1, 0.25 and 0.5 m. The decay time is the maximum within the frequencies 0.4-4 kHz [21].

3.5. Significance of spatial averaging

Spatially averaged responses do not deviate significantly from the single point measurement at the listening position for small spatial averaging displacements (± 0.5 m) except for very low frequencies. The largest differences between the single point octave mean levels and the spatial average octave mean levels occur at low frequencies, in octave bands with frequency lower than 500 Hz. The differences are small at 500 Hz and higher frequency octave bands (tables 2 and 3).

When full bandwidth data is considered, a slight trend exists where increasing distance from the reference position and increasing reverberation measured inside the octave bands 0.4-4 kHz is linked to larger difference between the single point measurement and spatial average (Fig. 7) but this trend is mainly created by low frequency room effects and if octave bands at and above 500 Hz are considered, the trend disappears (Figure 8). The reverberation time appears to have a stronger effect on this than the spatial average distance.

4. Discussion

For immersive reproduction systems, controlled in-room spectral response is the foundation for meaningful loudness calibration and accurate monitoring [2],[3],[4]. For example, the ATSC recommended practice A/85 [1] devotes an entire chapter to monitoring system calibration to dismantling what they call the 'circle of confusion' for audio monitoring. Reproduction system calibration is the key to achieving this.

We have studied the ability of single position magnitude response measurement to render essentially the same information as spatial averaging. The design of the present research is suitable for modelling typical audio engineering workflow, where decisions are taken primarily in seated position. The spatial average collection microphone position layout in our work consequently reflects the single engineer situation. Fixed microphone height at the listener's ear height was used, therefore the present work cannot reveal sound colour stability in vertical orientation although similar findings can be expected also for vertical direction.

In-frequency smoothing is the typical tool used for in-room frequency response measurements and it can attenuate very local features at mid and high frequencies, to extract the subjectively significant wider bandwidth features in the system response. Excessive smoothing easily leads to loss of precision at very low frequencies.

Spatial averaging can extract the common spectral features visible in all the spatial measurement points, rejecting local differences. Spatial averaging is typically used for very wide seating area applications, such as cinema auditoria, to remove the shared spectral coloration. Nevertheless, spatial averaging has also been offered as a more reliable starting point for system equalization over single-point measurement even for single-person working environments.

Magnitude responses measured have been smoothed to *de facto* standard 1/3 octave smoothing. This may limit the accuracy low frequency for narrow band resonances, audible particularly when associated with long decay time. This was demonstrated by the case example reported in this paper, where significantly less smoothing was applied to low frequencies, offering better visibility to modal resonance effects in the room.

Octave-band averaging of pressure can also affect extremes but is suitable for modelling the wideband sound colour variation as a function of frequency.

When in-frequency smoothing is applied mostly non-significant differences were found in our study between the single-point measurement and spatial averages taken with distances less than 0.5 meters for listening rooms similar to professional rooms in decay characteristics. Large differences between a single point measurement and a spatial average are connected to strong modal resonances in the room, typically happening at low frequencies where the natural room absorption is not sufficient to reduce the decay time. However, at low frequencies the audibility of such medium-to-high Q resonances is much reduced comparing to mid frequencies. Studio monitoring rooms typically have relatively short reverberation times, reducing local variation in pressure and further decreasing significance of spatial averaging.

5. Conclusion

The use of single point equalization appears to be a safe choice for measurement of studio monitoring rooms typically having relatively low reverberation times and well controlled room modes. The use of spatial average is not likely to significantly change or improve the outcome of equalization in the case of equalizing an audio reproduction system for a limited listening area essentially intended for one person.

6. References

[1] International Telecommunication Union: "Advanced sound systems for programme production", ITU-R BS.2051-1 (2017).

[2] International Telecommunication Union: *Multi-channel sound technology in home and broadcasting applications*, Report ITU-R BS.2159-6 (2014).

[3] Advanced Television System Committee (ATSC): *ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television (A/85:2013)*, ATSC A/85

[4] European Broadcasting Union (EBU): *R128 Loudness Normalisation And Permitted Maximum Level of Audio Signals*, (2014).

[5] International Telecommunication Union (ITU): *Recommendation ITU-R BS.1116-3*, (2015).

[6] Mäkivirta A., Liski J., Välimäki V.: *Effect of Delay Equalization on Loudspeaker Responses.* AES 144 Convention (2018).

[7] Tremaine, H.: *Audio Cyclopedia*, 2nd. ed., H.W. Sams, Indianapolis, (1973).

[8] Villard, J.R.: *Adjustable frequency selective apparatus*, US Patent 2672529 (1954).

[9] Flickinger, D.: *Amplifier system utilizing regenerative and degenerative feedback to shape the frequency response*, US Patent 3752928 (1973).

[10] Massenburg, G.: *Parametric Equalization*, AES 42th Convention (1972).

[11] Pedersen, J.: *Loudspeaker-Room Adaptation for a specific Listening Position using Information about the Complete Sound Field*, AES 121 Convention (2006).

[12] Olive, S. et al.: *The Subjective and Objective Evaluation of Room Correction Products*, AES 127 Convention (2009).

[13] Pedersen, J. et al.: *Natural Timbre in Room Correction Systems*, 122nd AES Convention, (2007).

[14] Pedersen, J., El-Azm, F.: *Natural Timbre in Room Correction Systems (Part II)*, 32nd International AES Conference: DSP For Loudspeakers (September 2007).

[15] Toole, F.: *Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1*, J. Audio Eng. Soc., vol. 34, 4 pp. 227-235, (1986).

[16] Toole, F.: *Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 2*, J. Audio Eng. Soc., vol. 34, 5 pp. 323-348, (1988).

[17] Toole, F.: *Sound Reproduction*, Focal Press (2008).

[18] Toole, F.: *The Measurement and Calibration of Sound Reproduction Systems*, J. Audio Eng. Soc., Vol. 63, No. 7/8, pp. 512-541 (2015).

[19] Toole F.: The Acoustics and Psychoacoustics of Loudspeakers and Rooms – The Stereo Past and The Multichannel Future, AES 109 Convention (2000).

[20] Society of Motion Picture and Television Engineering, *Ultra High Definition Television, audio characteristics and audio channel mapping for programme production*, SMPTE 2036-2-2008 (2008).

[21] Mäkivirta A, Lund T.: *Spatial stability of the Frequency Response Estimate and the Benefit of Spatial Averaging,* 141 AES Conv. (2016).

[22] Mäkivirta A., Lund T.: *Efficacy of Using Spatial Averaging in Improving the Listening Area of Immersive Audio Monitoring in Programme Production*, AES Int. Conf. Spatial Reproduction (2018)

[23] Olive S., Schuck P., Ryan J., Sally S., Bonneville M.: *The Detection Threshold of Resonances at Low Frequencies,* J. Audio Eng. Soc. 45 (3) pp. 116 – 128 (1997).

[24] Moore B., Oldfield S., Dooley G.: *Detection and discrimination of spectral peaks and notches at 1 and 8 kHz*, J. Acoust. Soc. America, 85 (2) 820 - 836 (1989).

[25] Bahritkar S., Robinson C., Poulain A.: *Equalization of Spectral Dips Using Detection Thresholds,* 140 AES Convention (2016).

[26] Mäkivirta A., Anet C.: *A Survey Study of In-Situ Stereo and Multichannel Monitoring Conditions*, AES 111 Convention (2001).