Stability of the Frequency Response Estimate in Listening Rooms

A. Mäkivirta[1], T. Lund[1]

[1] *Genelec Oy, Iisalmi, Finland. Email: thomas.lund@genelec.com*

Abstract

In-room estimates of loudspeaker responses for professional use are typically taken either at one microphone location, replacing the listener with a microphone, or averaging in space, at multiple locations at and relatively close to the listening location. In this work, a number of listening rooms were measured using 18 precisely defined locations. In-space averaging in combination with frequency domain averaging can increase the stability of the frequency response estimate. However, the spatial averaging points used in taking a measurement should be chosen based on the room acoustics and on the application. Spatial averaging across a wide area may come with the risk of compromising the result at the main listening position.

1. Introduction

Professional loudspeakers intended for critical monitoring of audio are positioned and equalised to reduce the acoustical influence of the listening space [3] as room effects tend to reduce the accuracy of monitoring even in high quality professional listening rooms having controlled acoustic characteristics [1].

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 [17]. Adjustable filters make room equalization much easier and have been widely available already in the vacuum tube era (see for example [15]). More versatile parametric equalizer filters were introduced by several inventors in the early 1970s [14, 16] and finally digital signal processing has made both measurements and room equalization easy and precise while alleviating the signal quality concerns that were earlier associated with equalization.

Because the requirement in professional monitoring is to hear the audio recording content in a neutral way, without adding or removing anything, the setup ideally approximates an allpass system within the audible frequency range,

$$\left| H(e^{j\omega}) \right| = 1 \quad (1)$$

This implies a system with a flat frequency response and possibly causing a delay to the audio signal. International recommendations consequently call for the room response of the monitoring system to be flat at the listening location and the reverberation time in the listening space to be the same for all frequencies. As the loudspeaker is a 3D radiator and the listening room is a 3D medium for audio, there is also consideration about the way the loudspeaker radiates at different frequencies. Because of this, there is also a requirement for directivity characteristics in the recommendations [1].

With loudness-based production spreading globally, the spectral response of the monitoring system and its level calibration has become a cornerstone of recommendations for broadcast and streaming [11, 12].

1.1. Target curve

While there is a consensus that a neutral loudspeaker has a flat anechoic frequency response, several researchers have been commenting on the suitability of an allpass target for the in-room frequency response. Some researchers suggest a down sloping character across the full or at least a part of the audio band. This seems to be largely motivated by listener preference [4, 5, 6, 7, 8, 9] while little attention has been paid to solving the problem of the 'circle of confusion'. This refers a self-referenced system where existing recordings are used to evaluate the room sound [10, 13].

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.

1.2. One or more measurement locations

The sound pressure measurable in a room at a single microphone location is relatively local, and large variations in the pressure can be seen when the microphone location is moved. This is particularly evident at high frequencies where large and very local comb filtering effects can happen because of acoustic reflections. That effect is usually reduced by time domain windowing the impulse response estimate and by applying in-frequency smoothing with a sliding variable-width averaging window to the frequency response. These techniques are usually able to sufficiently reduce acoustic comb filtering and tend to reduce spatial locality of a measurement, thereby rendering the measurement usable for the practical purposes of evaluating room-induced sound colorations.

Spatial averaging is used in the context of cinema dubbing stage and cinema theatre equalization with even spatial sampling of the frequency response across the floor area intended for listeners [2]. Increasing the number of microphone positions can provide a more complete picture of room acoustics. Such measurements estimate the power output of the loudspeaker in the room and using that for system equalization particularly at low frequencies [4].

1.3. Scope of the work

Instead of sampling the acoustic field in the whole space, the present work aims to help minimize reliably a local effect, the room influence at one listening position. Spatial sampling of the room response in the vicinity of this listening location is one potential method of improving the reliability of the measurement of the acoustic response audible to the listener.

The purpose of this work is to study the sufficiency of the single microphone position acquisition of the acoustical response at the listening position.

The working hypotheses are that (1) spatial averaging is able to extract the common acoustic features and therefore enables focusing of the system equalization to these essential features, and (2) spatial average will be able to prevent the potentially local acoustic effects that may occur if only one microphone position is used.

The present study is conducted predominantly in the context of professional reproduction spaces. Therefore we have selected rooms with acceptably low reverberation time and level of the early reflections at the listening position.

The present work does not address the topic of how the measurement will be used for room equalization or what the suitable equalization target should be.

2. Materials and methods

2.1. Room selection criteria

A variety of listening rooms were included in the study. The rooms are or have been in use for stereo mixing or editing.

Most rooms lack refined acoustic design or extensive acoustic treatment. Room selection criteria also included repeated availability, geographical location in Denmark or Finland, low background noise, acceptable reverberation time, and reasonably low level and acceptable direction of early reflections at the listening position.

All twelve rooms appointed for the study are reported, meaning that once a room had been selected, it was not excluded post-measurement. They key parameters of the rooms are summarized in Table 1.

2.2. Microphone grid

Locality of the frequency response is studied by using one main microphone position and 17 offset microphone positions (in total 18 microphone positions) located at increasing offsets from the main listener/microphone location. The offset distances are 0.1 meters to the front, 0.1, 0.25, and 0.5 meters to the side and back, as well as 1.5 meters to the back, see Figure 1.

The microphone position layout was chosen to reflect the typical professional critical listening application where the engineer is located at a defined position, typically seated at the console or audio workstation.

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 1. 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.

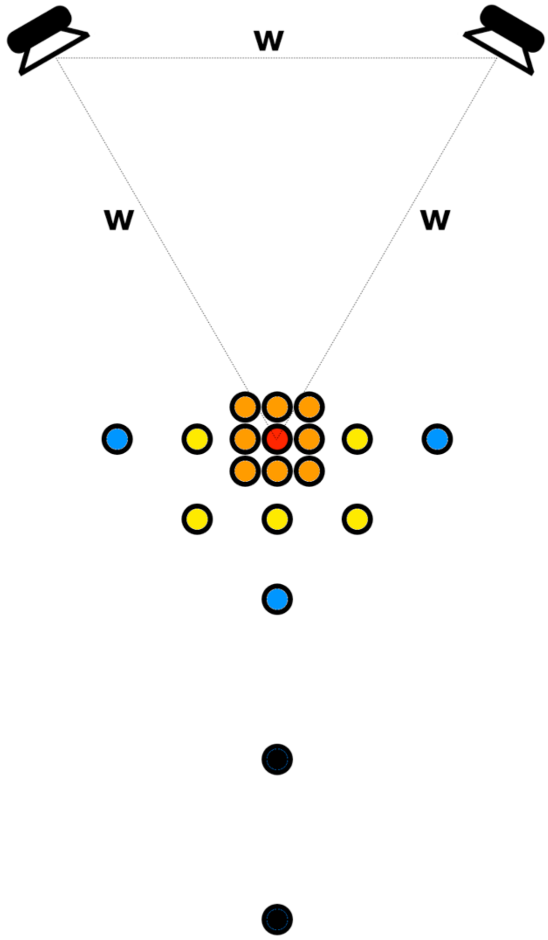

Fig 1. The full measurement grid. The main listening position (red) is one point of an equilateral triangle with the loudspeakers. Offset mic positions 0.1, 0.25 and 0.5 m plus 1.0 or 1.5 m to the back. Details in figure 2, sharing the same colour coding.

The measurements are taken at a fixed microphone height set to be the same as the listener's ear height. The microphone positioning accuracy is ensured by using a mat with the positions marked and a pointer on the microphone giving a positioning accuracy better than 2 mm for the defined measurement positions and between rooms in relation to the stereo pair. The loudspeakers were set up in a standard stereo pair configuration relative to the main microphone position (or 'listening position' or 'C').

The measurement setup in the room is generally arranged on the left-right centre axis so that the room has the best possible left-right symmetry acoustically relative to the measurement setup and the loudspeaker locations. The listening direction frequently corresponds with the longest dimension of the listening room.

2.3. Loudspeakers and microphone

The loudspeakers used were mainly a two-way design with a 5 in woofer, 2/3 in tweeter and linearized phase response (type Genelec 8330 or similar).

The measurement microphone has an electret capsule with omnidirectional characteristics and has been calibrated to have a flat response when it is pointing upwards, resulting in very similar frequency response for all source directions on the same horizontal plane at a given height relative to the microphone. The effect of the height related response variation in the microphone is small and not significant for the purposes of this study.

2.4. Processing of measurements

The impulse response data acquisition is done using a log sine sweep with length 256 kilo-samples, resulting in an impulse response with length 2.97 s sampled at 44.1 kHz for each of the microphone locations.

Both the left and right loudspeakers in the stereo pair are measured, generating two sets of 18 measurements for each room. In total, 432 measurements in 24 series are taken in the study.

The magnitude response measurements are smoothed in frequency using the industry standard 1/3 octave smoothing window and 1/12 octave smoothing window.

Before averaging, differences in the measurement point distances are removed by level normalization. Monitor loudspeakers are essentially point source radiators, with the characteristic inverse square law reduction of the sound level with increasing distance to the monitor. The various measurement positions in the study cause differences in the distance which translate to level differences.

The level normalization removes the effect of geometric distance variation. Without level normalization, individual measurements being averaged receive a weight related their distance. This is not desirable as we want each measurement to have equal weight. 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 three different spatial averages in a manner that could be expected for practical applications. The cases include (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.

The 1.0 and 1.5 m mic positions offset back from the main microphone position are used for reference to see how much the response will change in the front-to-back direction.

The spatial average is compared to the single microphone position measurement statistically. The mean difference between the main position measurement and the spatial average, abbreviated MSAD, as well as the difference histogram, is calculated for each of the spatial average cases A-C.

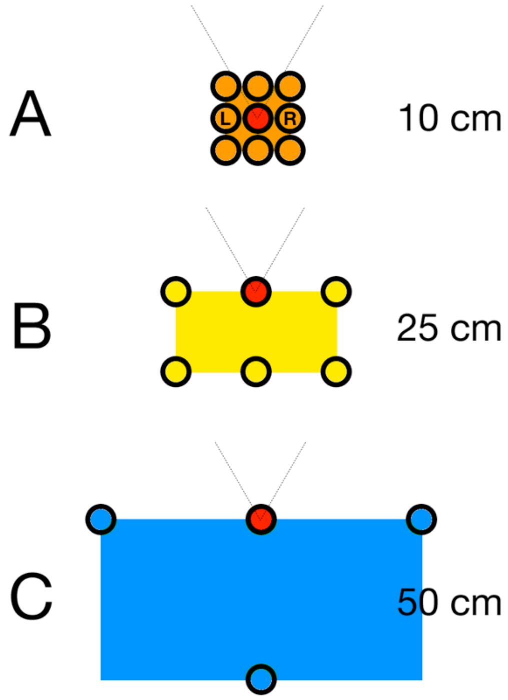

Fig 2. Definitions of the spatial average cases A to C. Main mic position is red. Spatial averaging areas (other colours) incl. number of measurement positions used.

Each spatial average response is calculated using the 1/3 octave and 1/12 octave smoothed data. All measurements receive the same weight. The weight received by an individual measurement depends on the number of measurements included in the spatial average.

Cumulative histograms of the differences between the main position measurement and the spatial averages are calculated for all rooms and measurements. The cumulative histogram is calculated using a 0.1 dB bin size.

3. Results

3.1. Effect of measurement point distance

The spatial averaging is an additional averaging process on top of the normally applied smoothing in frequency.

Both frequency smoothing and spatial averaging have the effect of reducing the extreme differences in measurements, particularly narrow band differences. The room acoustic influences to the loudspeaker frequency response are the strongest at low frequencies. This is where the largest differences are seen between a single position measurement smoothed in frequency and a spatial average calculated using several such smoothed measurements taken at and near the listening position. The largest differences are seen at the rapid turns in the frequency response, typically close to notches that remain after the in-frequency smoothing.

The microphone positions at 0.1 m from the main microphone position are a special case in the present study because this set has microphone positions also forward of the microphone. The other spatial averaging cases (0.25 and 0.5 m offsets) have microphone positions on the line at the single microphone position or towards the back of the room.

The spatial average with 0.1 m displacement exhibits the smallest difference to the single point measurement, and differences increase with the increasing spatial average displacement.

3.2. Case example: MSAD

Each microphone position represents a full acoustic measurement. To keep the paper length manageable, typical single room data is provided only, see Figure 4. The full data set for all the rooms is summarized later.

The single position measurement and the 0.1 m off-position spatial average show the best agreement (area A). When the spatial average distance increases to 0.25 and to 0.5 meters (areas B and C) more differences are seen towards low frequency.

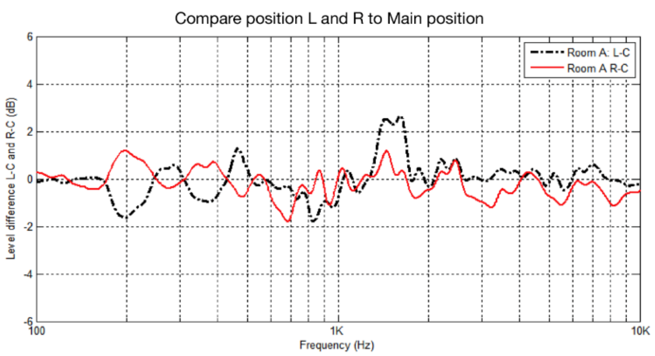

Fig 3. Example of results. Room A, left. Difference of the sound pressure at position L to the Main position (dashed black) and position R to the Main position (solid red), 100 Hz to 10 kHz. 1/3 octave smoothing has been applied.

In addition to the MSAD, responses at single microphone positions L and R (see Figure 2) were compared to the main microphone position C. Microphone locations L and R roughly model locations of the left and right ears for the case where the main microphone position C is located at the geometrical centre of where the listener's head would be.

The level difference graphs L-C and R-C in the case example (Figure 3) display the typical level differences we observed, showing symmetrical but opposite L/R excursions around the 0 dB level below 500 Hz.

3.3. Case example: cumulative histograms

A cumulative histogram is calculated for three frequency bands (30-100 Hz, 100-1000 Hz and 1-10 kHz) as this better illustrates how differences between the single position measurement and the spatial average are rather frequency dependent, see Figure 5. Above 1 kHz agreement between the single point measurement and the spatial average is generally good for all the off-position displacements tested.

The spatial average systematically deviates from the single position measurement for all cases of the spatial average. The spatial average shows a slightly smaller SPL than the single position measurement. This seems to be related to the fact that spatial average attenuates the visibility of local acoustic effects (gain maxima and minima) more than a single point measurement after smoothing in frequency. Smoothing in frequency is an averaging operation that happens in the frequency domain.

Spatial averaging multiple such measurements effectively cascades two averaging operations, so we can expect that the spatial average has a stronger smoothing tendency than the smoothing in frequency alone. In principle, a similar effect could be achieved by using a slightly wider smoothing-in-frequency range. For the example Room A, this effect of "additional smoothing" is not large, about 0.5 dB in the SPL estimation. The other room measurements do not significantly deviate from this behaviour when the room reverberation is modest.

Data is also divided into IEC octave bands (IEC61260:2014) and the mean difference statistic is calculated for each octave band for the left and right monitors separately (room A results, see Figure 6), for all monitors in the study.

The differences in the higher frequency octave bands up from 1 kHz are negligible except in the highest octave where the aiming of the monitor in combination with its increasing directivity will affect the results. Larger differences are seen for frequencies where the room exerts large acoustic influences, typically below 1 kHz.

3.4. Results of pooled data

The data of all the twelve rooms, considering left and right speakers, is pooled in the octave bands, see one example in Figure 6.

Comparing to the single point measurement, the octave band mean difference increases with increasing displacement distance of the additional microphone positions used for the spatial averaging.

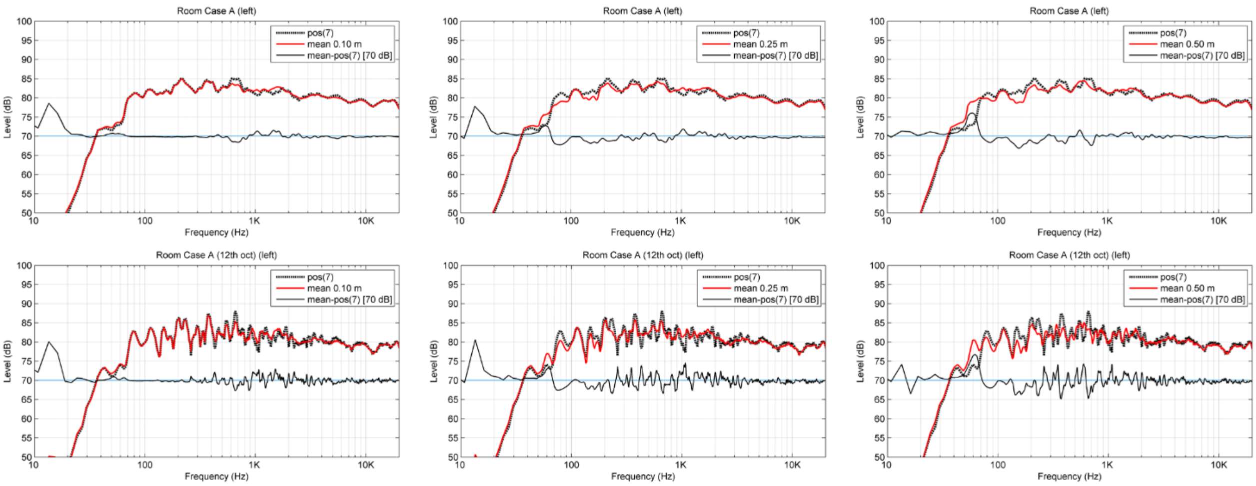

Fig 4. Example of results. Room A, left. Spatial means area A, B, C vs the response at the main position, "pos(7)". MSAD is shown with 0 dB on the blue line. 1/3 octave smoothing (upper), 1/12 octave smoothing (lower).

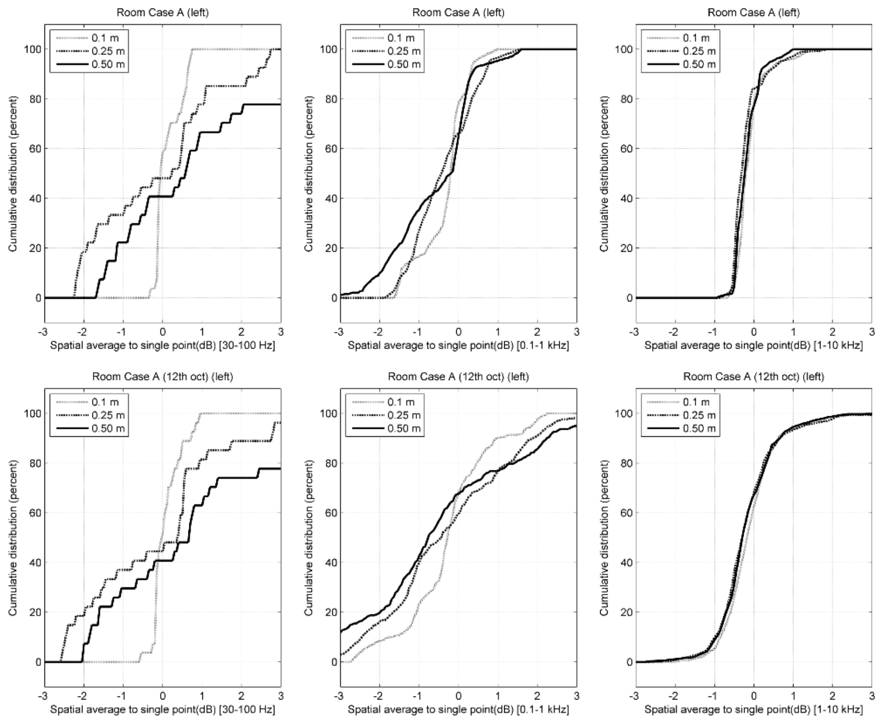

Fig 5. Example of results. Room A left. Cumulative distribution of the difference of the A, B, C average and the single position SPL. The cumulative distributions are presented in three different frequency bands (30-100 Hz, 100-1000 Hz and 1-10 kHz). 1/3 octave smoothing (upper), 1/12 octave smoothing (lower).

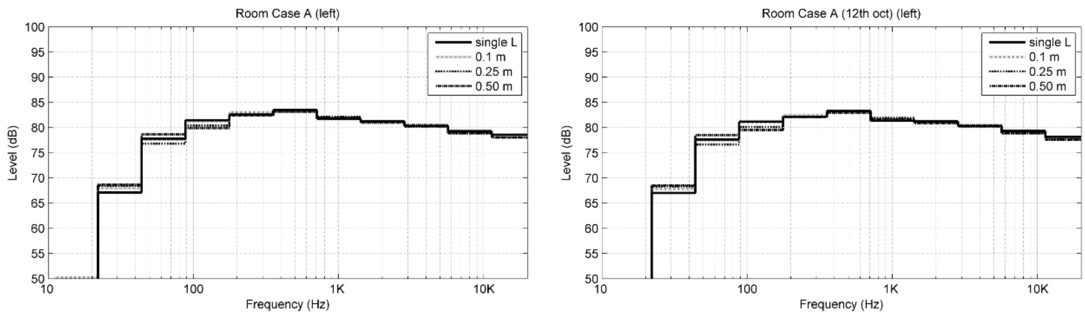

Fig 6. Example of results. Room A left. The mean difference in IEC octave bands between the spatial average SPL and the single position SPL. 1/3 octave smoothing (left), 1/12 octave smoothing (right).

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).

Large differences (minima and maxima) are typically at very low frequencies. This can also be seen in the cumulative histograms of the difference data (Figure 4). The cumulative histograms are asymmetrical. This is related to the inherently different influence of acoustic summation and cancellation in room acoustics. This happens particularly when the modal resonances in the room are strong, resulting in larger differences between the single point measurement and a spatial average.

	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 2. Pooled octave band statistics across **all rooms** for the full audio bandwidth.

	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 3. Pooled octave band statistics across **all rooms** for the octave bands 500 Hz and higher.

There is a slight trend where increasing displacement as well as increasing maximum room reverberation measured inside the octave bands 0.4-4 kHz implies larger differences between the single point measurement and the spatial average (Figure 7). This is related to the nature of the spatial average to suppress extremes in the frequency response more than applying just the in-frequency smoothing alone.

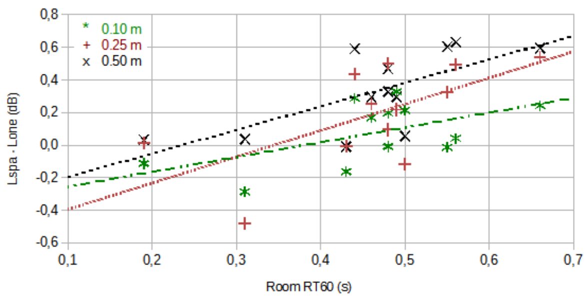

Fig 7. Mean level difference in octave bands, averaged across the audio band for the spatial offset cases 0.1, 0.25 and 0.5 m. The reverberation time RT60 is the maximum within the frequencies 0.4-4 kHz.

This effect can also be seen in the Case A room sample response plots (Figure 3). In our material the reverberation time appears to have a stronger effect on this than the spatial average displacement distance. This trend is mainly created by low frequency room effects. If only octave bands at and above 500 Hz are considered the trend disappears (Figure 8).

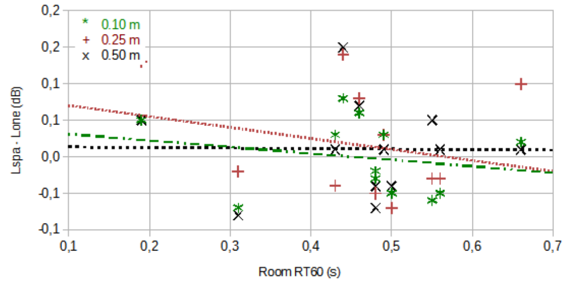

Fig 8. Mean level difference in octave bands 500 Hz and above, averaged across the audio band for the spatial offset cases 0.1, 0.25 and 0.5 m. The reverberation time RT60 is the maximum within the frequencies 0.4-4 kHz.

4. Discussion

In-frequency averaging can effectively attenuate local features at mid and high frequency. In-space averaging can effectively extract the common frequency response features visible in all the measurement points. Spatial weighting of the average combined with frequency domain averaging of the resulting estimate can increase the stability of the frequency response estimate for the features that are relevant to subjective compensation of the sound colour at the listening location.

Spatial averaging is sometimes assumed to produce a better representation of room acoustics than single point measurements, and thereby be more useful as a reliable starting point for system equalization. However, no significant difference was found between the single point measurement and the spatial average for small spatial average displacement measurements taken in professional listening rooms.

The spatially averaged responses do not deviate significantly from the single point measurement at the listening position, for small spatial averaging displacements (± 0.1 m), including the specific L and R locations at "ear distance" relative to the main position. The symmetrical L/R differences observed below 500 Hz point to the main position as the generally benign common denominator for the purposes of frequency response calibration in the listening room.

The spatial averages show less difference relative to the listening position compared to the individual off-position measurements. Partially, this may be related to the fact that in this work the measurement positions are symmetrically located relative to the single listening position used as the reference.

The spatial averaging was done using in-frequency smoothed measurements. Not smoothing before the spatial averaging does not change the outcome of the calculation as both are linear operations, and spatial averaging can be done prior to applying smoothing in frequency.

Large differences between the single point measurement and the spatial average are connected to strong modal resonances in the room. As pronounced modal resonances are related to small acoustic attenuation in the room, it is likely that high reverberation time can predict larger difference between the one point measurement and a spatial average.

5. Conclusion

Loudness-based production in broadcast and post relies on accurate monitoring for judging balance, spectrum, speech intelligibility, and audio quality.

Loudness is a *perceptual* property of an audio signal when it is reproduced acoustically and listened to. It is a complex, nonlinear function of amplitude, frequency, and time. Because of this complex nonlinear nature of perception, a controlled in-room spectral response is therefore the foundation of any meaningful level calibration, which is the requirement in all global and regional loudness standards.

ATSC A/85 devotes a comprehensive chapter to calibrated monitoring, and to dismantling the 'circle of confusion' described in section 1.1.

The use of single point equalization is confirmed [3] as a safe choice for measurement of studio monitoring rooms 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 that case.

However, in rooms with pronounced modal resonance a spatial average already with a displacement as small as 0.1 m can be useful in avoiding very local phenomena and may lead to more reliable reproduction system equalization.

Spatial averaging across a wide area may come with the risk of compromising the result at the main listening position. The spatial averaging positions should therefore be chosen based on the intention of the room equalization and the acoustic quality of the room.

6. References

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

[2] Cengarle G. & Mateos, T: *Effect of Microphone Number and Positioning on the Average of Frequency Responses in Cinema Calibration*, J. Audio Eng. Soc., vol. 63, no. 10 (2015).

[3] Mäkivirta, A. & Anet C.: *The Quality of Professional Surround Audio Reproduction: A Survey Study*, 9th Int. AES Conference: Surround Sound - Techniques, Technology, and Perception (2001).

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

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

[6] Pedersen, J.A. et al.: *Natural Timbre in Room Correction Systems*, 122nd AES Convention, (2007).Pedersen, J.A. & El-Azm, F.: *Natural Timbre in Room Correction Systems (Part II)*, 32nd International AES Conference: DSP For Loudspeakers (September 2007).

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

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

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

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

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

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

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

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

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

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

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