Effect of microphone number and positioning on the average of frequency responses in cinema calibration

Giulio Cengarle,Toni Mateos

Dolby Laboratories, Barcelona, Av. Diagonal 177, 08018, Spain.

Abstract

When measuring the response of a loudspeaker by averaging multiple points in a room, the results typically vary according to the number of microphones employed and their positions. We present an interpretation of the average procedure which shows that averaging converges to a compromise response over the relevant listening area, at a rate inverse to the square root of the number of microphones employed. We then provide real-world examples by performing measurements in a dubbing stage and a cinema theater, and analyzing the variations of average frequency responses over a large set of different microphone number and positioning. Results confirm the predicted scaling of the deviations, and quantify their magnitude in typical rooms. The data provided help establishing the point of diminishing returns in number of microphones.

1 INTRODUCTION

The frequency response of a loudspeaker system in a room is subject to spatial variations due to, among other effects, the interaction with the room, the directivity of the speaker, and the sound absorption in the air. Different calibration techniques and workflows that take such spatial variations into account are in use in different fields. We shall focus here in the application to cinema calibration, where typically the measurements of a few microphones located in a relevant portion of the seating area are averaged and then used to design an equalization filter [1]. The question we address here is how reliable the average of a small number of microphones is in representing the actual response of the speaker across the room. If the chosen locations of microphones are not appropriate, or simply interpreted differently by different engineers, then the resulting equalization filters might also differ to a certain amount, potentially leading to different listening experiences.

Whereas the standard [1] requires at least four microphones, and some companies in the industry already promote the use of eight, there seems to be a lack of publicly available data assessing the type of error associated with the choice of different numbers of microphones or a different positioning in the room. Some studies have shown examples of the variation between single points in a room, pointing out the perils of assessing the response of a channel with the frequency response derived from a single position in the seating area [2, 3]. However, the underlying principle in taking spatial averages is the fact that the fluctuations associated to different positions, such as the effects of standing waves or some localized reflections, average out, or at least are reduced, revealing only the anomalies that are consistently present across the whole area covered by the microphones.

In this paper, we study the effect of the averaging procedure both at a theoretical and empirical level. On the one hand, we provide an interpretation of averaging as a means to converge towards a certain spatial average of the impulse response across the relevant listening area. It is shown that current workflows provide a simple Monte Carlo approximation, and that the effect of increasing the number of microphones, N_{mic}, is to effectively bound the differences that different engineers would obtain by a factor of $1/\sqrt{N_{mic}}$. The described interpretation therefore quantifies both the physical quantity that averaging provides, as well as the repeatability of the measurements.

We then verify the prediction by analyzing data obtained from a professional dubbing stage and a cinema theatre, where we measured Impulse Responses (IRs) from selected speakers to a large grid of positions in the listening area. We used this data to quantify both the spatial variation of the IRs, as well as the EQ filters that different installers would obtain when averaging over different numbers of microphones. The results are in good agreement with the theoretical prediction and help explaining experimental data obtained with similar procedures in smaller rooms [4].

The presented study might help assessing the point of diminishing returns, where the addition of more microphones does not significantly improve the results.

1.1 Outline

Section 2 discusses the theoretical interpretation of averaging and its consequences. Section 3 details the measurements, including a description of the rooms, the microphone positions, and the procedure for selecting and comparing random subsets of microphones. The results for different numbers of microphones in each room are presented in section 4. Final comments are presented in section 5.

2 THEORETICAL INTERPRETATION OF AVERAGING

Let us denote the impulse response of a given loudspeaker in a room by $IR(t, \vec{x})$, where t is time, and $\vec{x} \in R^3$ refers to points within the room. Let $S(f, \vec{x})$ be the corresponding spectrum, i.e., the square of the absolute value of the Fourier transform of IR. In what follows, we shall drop the dependency on frequency f to ease notation.

The averaging procedure consists of measuring the responses (or spectra) at a set of N_{mic} points, $\vec{x}_n$, $n = 1, ..., N_{mic}$, and then averaging:

$$\overline{S}_{N_{mic}} = \frac{1}{N_{mic}} \sum_{n=1}^{N_{mic}} S(\vec{x}_n)\,. \tag{1}$$

While different types of averages are possible, the analysis presented here applies without restrictions; without loss of generality, we have chosen to employ the energy (RMS) average, which is the preferred method in cinema calibration, and corresponds to assuming incoherent energy weighting

of the available microphones. Compared to the average of decibel values, it is less affected by large dips in the frequency response of eventual outlier positions, therefore it is less prone to large boosts in subsequent equalization filters.

In cinema applications, the set of microphones is located more or less freely sampling the relevant listening area, subject to some constraints detailed in section 3. This allows the interpretation of (1) as a Monte Carlo [5] approximation to:

$$\overline{S}_{cont} = \frac{1}{A} \int \mathrm{d}A \; S(\vec{x}(A)) \,. \tag{2}$$

The meaning of $\overline{S}_{cont}$ is simply the continuous average of the spectrum over the relevant listening area A. Equation 2 quantifies the precise manner in which constructive and destructive points of standing waves, as well as localized reflections, combine into one single frequency-dependent function $\overline{S}_{cont}$.

On the other hand, the Monte Carlo theory quantifies the repeatability of the measurements. For fixed N_{mic}, the approximation (1) leads to different results depending on the choice of random positions $\vec{x}_n$. It can be proven [5] that

$$\overline{S}_{cont} = \overline{S}_{N_{mic}} \pm \frac{\sigma_A(S)}{\sqrt{N_{mic}}} \,. \tag{3}$$

where $\sigma_A^2(S)$ is the variance of the continuous spectrum over the listening area:

$$\sigma_A^2(S) = \frac{1}{A} \int \mathrm{d}A \; \left[S(\vec{x}) - \overline{S}_{cont} \right]^2 \,. \tag{4}$$

Equation 3 states one key beneficial effect of averaging: it enhances repeatability by reducing the effect of different choices of microphone positions. Increasing N_{mic} makes different choices lead to closer results, the improvement following an $O(1/\sqrt{N_{mic}})$ scaling.

On the other hand, Eq. 3 quantifies the intuitive fact that rooms that exhibit larger spatial variations in the spectra, i.e. larger $\sigma_A(S)$, lead to larger potential differences in the equalization filters obtained by different installers.

We shall stress the difference between the variance of the spectrum across the room, $\sigma_A^2(S)$, and the variance of a set of averages across different choices of microphone locations. The former is not reduced by using more microphones, it simply measures the degree by which $S(f, \vec{x})$ varies within the listening area in a given space. However, the use of more microphones does reduce the variation between averages, making them converge to $\overline{S}_{cont}$.

2.1 Importance sampling

Equation 2 treats all points within the listening area equally, making explicit the compromise that the averaging procedure leads to. In some applications, it might be preferable to give more importance to certain regions, like the area surrounding the mixing spot in a dubbing stage. This operation fits naturally in the context presented here.

All that is needed is defining a weight function $w(\vec{x})$, with standard properties of weight functions ($w(\vec{x}) \geq 0, \forall \vec{x}$, $\overline{w} = 1$.). The weight shall be chosen to take larger values close to the reference point, and smaller values away from it.

The weighted response to which the average procedure shall converge is now:

$$\overline{S}_{cont} = \frac{1}{A} \int dA \, w(\vec{x}) \, S(\vec{x}) \,. \tag{5}$$

The standard Monte Carlo theory states that convergence to this value can be simply ensured by distributing the microphones more densely in the regions where $w(\vec{x})$ is higher; precisely, they need to be distributed according to the probability density defined by $w(\vec{x})$. This quantifies a perfectly intuitive result that many engineers already use in their daily work.

2.2 Time-fluctuations in the IR

In the previous subsections, we neglected the fact that the spectrum is also time varying, in the sense that for fixed microphone positions, different IRs are obtained when measuring at different times, resulting from changes in environmental and electro-mechanical parameters.

We previously obtained an initial estimate of this effect by measuring the spectrum of various loudspeakers at fixed microphone positions in two different screening rooms at various times during a month. We found the variations in each frequency band to be typically within ±1.5dB, leading to a temporal variance $\sigma_t(S)$ below 0.5dB, as shown in Figure 1 for one particular loudspeaker.

This effect is also easily taken into account in the context presented here. Under the assumption that this temporal variability is statistically independent from the spatial variation discussed above, the results hold with the following minor change in Eq. 3:

$$\overline{S}_{cont} = \overline{S}_{N_{mic}} \pm \frac{\sigma_{total}(S)}{\sqrt{N_{mic}}} \,, \tag{6}$$

where

$$\sigma_{total}(S) = \sqrt{\sigma_A^2(S) + \sigma_t^2(S)} \,. \tag{7}$$

Note that the $1/\sqrt{N}$ scaling is unchanged. According to the numbers reported in this paper, the spatial variance clearly dominates, especially at low frequencies.

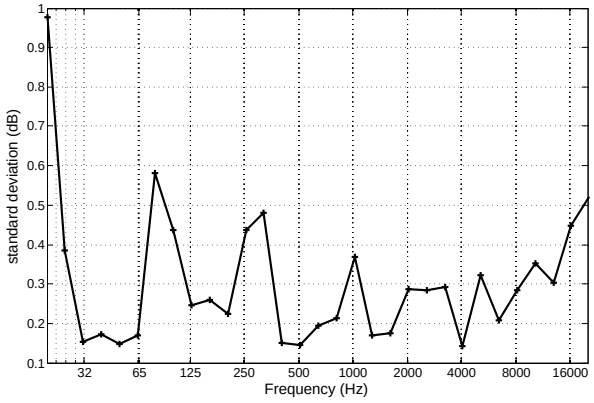

Figure 1: Temporal standard deviation $\sigma_t(S)$ per frequency band, for the left channel in the dubbing stage, obtained from eight measurements within a month, in a fixed microphone position.

3 EXPERIMENTAL DETAILS

The measurement procedure based on spatial averages of frequency responses is typically applied in both cinema theaters and dubbing stages. We analyzed a professional dubbing stage and a Dolby Atmos cinema theatre, both located

in Barcelona. We gathered data that allowed the understanding of what different installers would have measured, by locating N_{mic} at different positions, all fulfilling the current requirements.

3.1 Description of the rooms and the setup

The dubbing stage is a professional venue with size 13.7m x 11.4m x 5m. The floor is sloped by a three-level stage. IRs were measured from three channels (L, Lss1, Lss2) to thirty-six microphone positions, schematically drawn in figure 2. The screen speakers are JBL 3632T and the surrounds are JBL 8340. Lss1 is a distributed channel consisting of two speakers two meters apart, while Lss2 is a single surround speaker. The microphone grid spacing was about 1m; the distance from the third row of microphones to the fourth is about 1.5m due to the presence of the mixing console.

The Atmos cinema has about 300 seats, and its floor size is 19m x 23m. The floor is sloped, with a ceiling height varying from 10m near the screen to 2.5m in the rear. IRs were measured from three channels (L, C, Lss4) to eighty microphone positions, schematically drawn in figure 2. The screen speakers are JBL 4632T and the surrounds are JBL 8350. Lss4 is a single speaker. The microphone grid spacing was about 1.8m. Each row of microphones corresponds to a row of seats, covering all the rows from number 4 (close to the screen) to number 13 (four rows from the rear wall).

Beyerdynamic M1 omni-directional measurement microphones were used, connected to an RME Micstasy preamplifier and A/D converter.

3.2 Measurement procedure

All the frequency responses were derived from time-domain IRs; the measurements were based on the exponential sweep technique [6]. The sweep we used spans the range 20Hz to 24kHz in 6s. The playback level generated an SPL around 80dB C-weighted. When playing the sweeps through the selected channels, we were not passing through the existing B-chain adjustments, which means that any equalization applied to the speakers was bypassed, ensuring that we measured the raw response of the speaker/room system. In the case of screen channels, we did run through the existing crossovers, but these had no additional equalization applied. Once the sweeps were recorded, we obtained the IRs by convolving the recordings with the inverse sweep, and trimmed them to retain a length of 1s starting from 200ms before the peak corresponding to the direct sound. Considering that the decay time in the two rooms was smaller that 0.3s, the whole decay tail was included in the trimmed IRs. The frequency response corresponding to each IR was computed by taking the absolute value of the FFT of the trimmed IR and applying moving-average smoothing with a bandwidth of $1/6^{th}$ Oct.

3.3 Microphone subsets and averaging

The guidelines regarding microphone positioning suggest that they should be located within the relevant listening area, observing the following: i) at least one microphone should be near the reference point, the position at 2/3 from the screen across the length of the room; ii) in symmetrical rooms with symmetrical channel configurations, redundancy of symmetrical measurement positions should be avoided; iii) locations should also be avoided where boundary conditions affect the response in a way that is uncommon to most of the seats, such as close to the walls, behind occlusions or along the axes of small rooms. Other than suggested to follow these generic constraints, installers locate the microphones freely within the listening area.

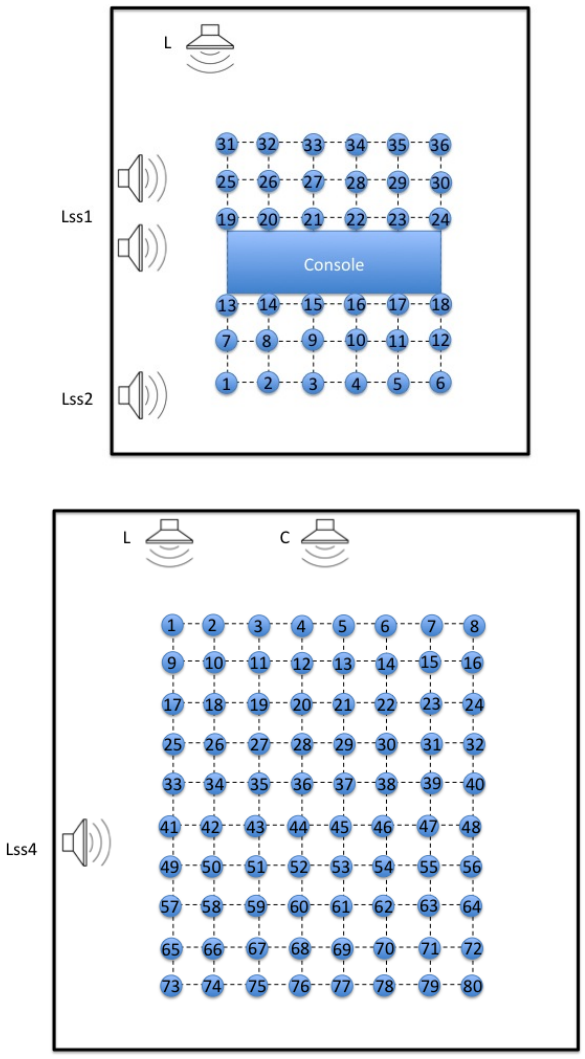

Figure 2: Measurement positions. Top: dubbing stage; bottom: cinema.

To simulate what different installers would measure, we will compare averages obtained from different subsets of N_{mic} elements of the entire set of measured positions. In order to create these subsets, we select random groups of N_{mic} out of the entire set of measurements, with the following properties:

1. The reference-point microphone is included in each subset.

2. Two subsets cannot be a permutation of the same microphones.

3. In each subset, the microphones have to be sufficiently spread.

In order to satisfy the last condition, once we have obtained a candidate subset of N_{mic}, we compute the spread Δ of the

subset as

$$\Delta = \frac{1}{N_{mic}} \sum_{n=1}^{N_{mic}} \sqrt{(x_n - \overline{X})^2 + (y_n - \overline{Y})^2}, \qquad (8)$$

where (x_n, y_n) are the coordinates of the microphones in the subset, and $(\overline{X}, \overline{Y})$ are their average. Only subsets of microphones with Δ larger than 2 grid units are accepted.

Each subset leads to a different average spectrum $\overline{S}_{N_{mic}}(f)$, as defined in Eq. 1.

4 RESULTS

4.1 Overall frequency response variation

Figures 3 and 4 show the overlaid responses of the specified channels measured in every microphone position, together with the total average and standard deviation. These responses have been normalized so that they have the same SPL level, C-weighted. Firstly, these plots indicate the range of variation that can be expected over the whole listening area: for screen speakers, this is contained within ±1.5dB at high frequencies, while at low frequencies the range exceeds 10dB.

It is worth stressing that these frequency responses are derived from the analysis of IRs in rooms where the decay of sound is so attenuated that the main contribution is the direct sound of the speaker: yet, they show the typical high-frequency roll-off described by the X-curve [7], as evidenced in figure 5, which is in this case due to the speaker response, the air absorption and the screen attenuation, rather than to measuring in a steady-state reverberant field. The high-frequency response is basically the same whether the analysis encompasses the whole IR or a smaller time window, as already shown in [8].

For individual surround speakers, such as Lss2 in the dubbing stage and Lss4 in the cinema, the high frequency curves exhibit higher deviation than screen channels (above 3dB). This is due to the fact that some microphones fall out of the coverage angle of the speakers; figure 6 shows the variation in high-frequency responses with respect to the average for an individual surround speaker and a screen channel across a line of six microphones positioned on and off axis. When distributed speakers are employed, the high frequency variation is expected to be reduced.

4.2 Frequency response variations as a function of the number of microphones

In this section we shall obtain estimations of the difference between equalization filters that different installers could potentially obtain, after averaging over N_{mic} microphones.

Following the procedure described in section 3.3, we shall first fix N_{mic}, and then obtain the average $\overline{S}_{N_{mic}}(f)$ using different subsets of all measured positions. We chose to form 20 different valid subsets, thus simulating what 20 different installers could potentially obtain.

Figures 7-8 show the maximum variation among the 20 averaged responses, $\overline{S}_{N_{mic}}(f)$, for different values of N_{mic}. Results are shown between 80Hz and 12.5kHz, since these are approximately the limits within which equalization is typically applied. In the cinema room, the first and last two rows of seats, as well as the left- and right-most lines, have not been considered, since in such room installers would probably not use those positions.

As expected, the deviation among different choices of microphone locations is reduced by increasing the number of microphones. In both rooms, averages computed with

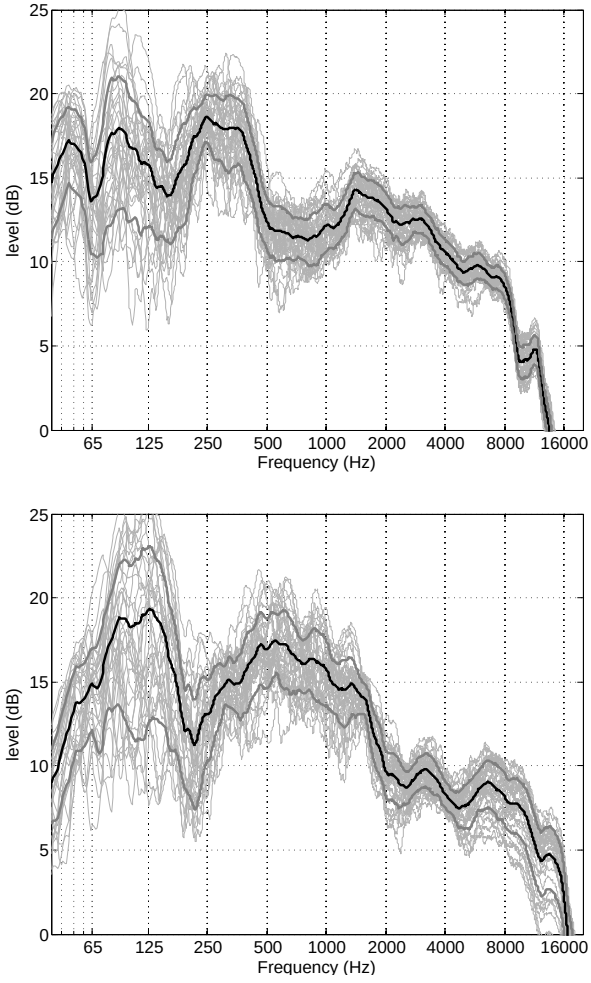

Figure 3: Overlaid responses from 36 microphone positions in the dubbing stage. The black curve is the average, and the thick grey curves are the range within the standard deviation. Top: left channel; bottom: surround speaker Lss2.

$N_{mic} = 32$ are within less than 1dB over the entire range, while for $N_{mic} = 16$ such differences increase to about 3dB in the the dubbing stage, and 2dB in the theater. For all values of N_{mic}, the differences are always larger at low frequencies, due to the higher variability of the responses in these rooms as the wavelength increases. It is worth pointing out that in the two lower octaves, $N_{mic} = 4, 8$ still lead to worst-case differences around 5dB to 7dB, which is not surprising given the range of variation shown in Figures 3 and 4.

One way to estimate the propagation from deviations in the averaged responses to deviations in the filters designed by the installers who use spatial averaging methods is as follows. Consider the simplest possible equalization strategy, by which the equalization filter $EQ(f)$ is designed to simply compensate the averaged response so that it follows a desired target curve $T(f)$:

$$EQ_{N_{mic}}(f) = T(f) - \overline{S}_{N_{mic}}(f). \qquad (9)$$

Of course, more elaborated strategies are often used, such as those that avoid the EQ filter exceeding a certain amount of

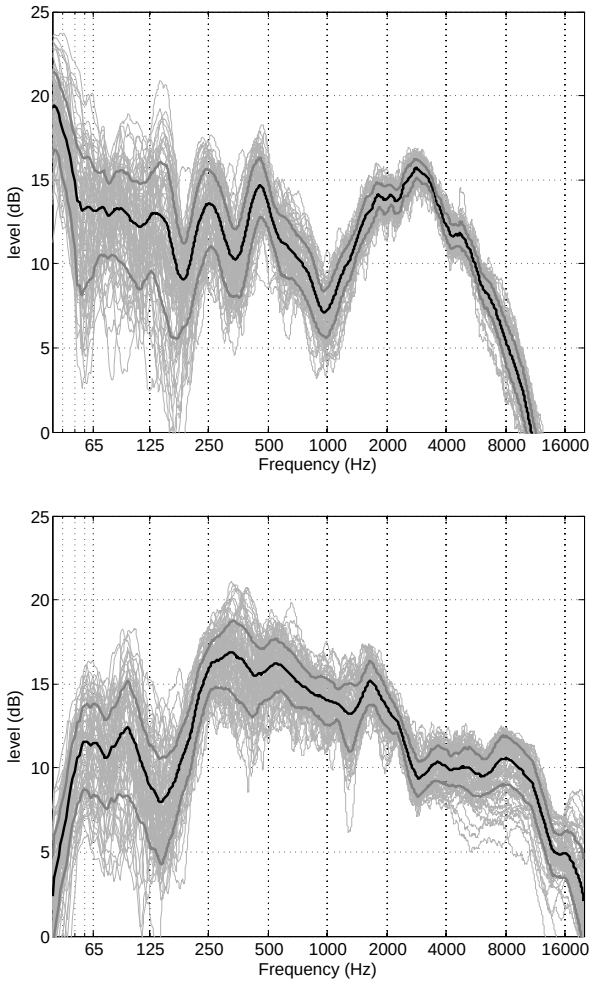

Figure 4: Overlaid responses from 80 microphone positions in the cinema. The black curve is the average, and the thick grey curves are the range within the standard deviation. Top: left channel; bottom: surround speaker Lss4.

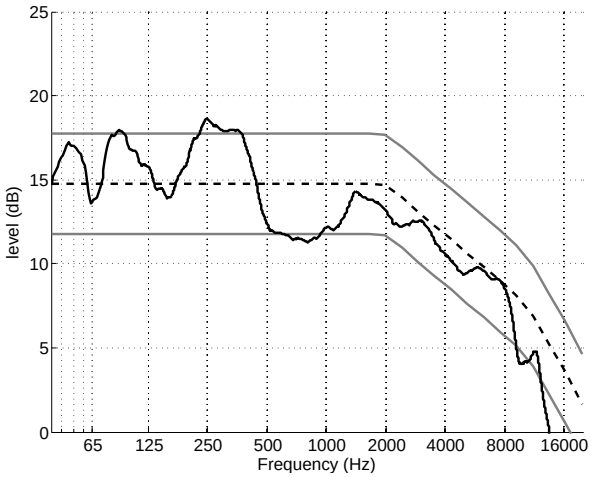

Figure 5: Average response of 36 microphones in dubbing stage, left channel, with overlaid X-Curve (dashed line) and $\pm 3dB$ tolerance range (solid gray lines).

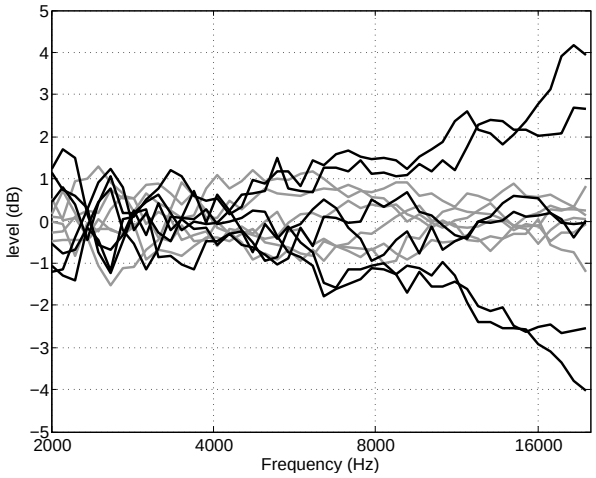

Figure 6: High-frequency deviations of speaker responses from their average across six microphones spanning a wide horizontal angle; black lines: speaker Lss2 in the dubbing stage across microphones 2, 8, 14, 20, 26, and 32; grey lines: speaker L in the dubbing stage across microphones 25, 26, 27, 28, 29 and 30).

boost, approaches that take the phase response into account, etc. The basic paradigm given by Eq. 9 should nonetheless provide a meaningful estimate of the effect under study.

Equation 9 leads to the simple property that the maximum differences $\Delta_{EQ}(f)$ across the EQ filters coincide with the maximum differences across the averaged responses:

$$\begin{aligned} \Delta_{EQ}(f) &= |\max\{EQ_{N_{mic}}(f)\} - \min\{EQ_{N_{mic}}(f)\}| \\ &= |\max\{\overline{S}_{N_{mic}}(f)\} - \min\{\overline{S}_{N_{mic}}(f)\}| \,. \end{aligned}$$

This property is independent of the choice of target response (e.g. flat, X-curve), and it leads to a re-interpretation of Figures 7-8 as showing an estimation of the maximum differences across EQ filters designed by different installers.

4.3 Deviations and repeatability

We shall here verify the effect predicted in section 2 whereby an increase in the number of microphones used in the average should reduce by a factor $1/\sqrt{N_{mic}}$ the deviations due to different microphone positioning.

A more extended analysis was done using the whole set of eighty positions in the cinema theatre. This time 40 random subsets were created for each value of $N_{mic} \in [2, 32]$. For each N_{mic} we computed the average spectra $\overline{S}_{i,N_{mic}}(f)$, and the standard deviation of the average spectra $\sigma(S_{N_{mic}}(f))$ over the 40 subsets.

Figure 9 illustrates the scaling effect clearly. Every pair of black and grey lines in the plot corresponds to one particular frequency band; black lines follow empirical values, and grey lines correspond to the exact theoretical fitting curve $\sigma_A(S)/\sqrt{N_{mic}}$.

As predicted by Eq. 3, the uncertainty in all frequencies scales in the same manner with N_{mic}, but the precise value depends on the actual spatial variation of responses across the room, which is typically larger at low frequencies.

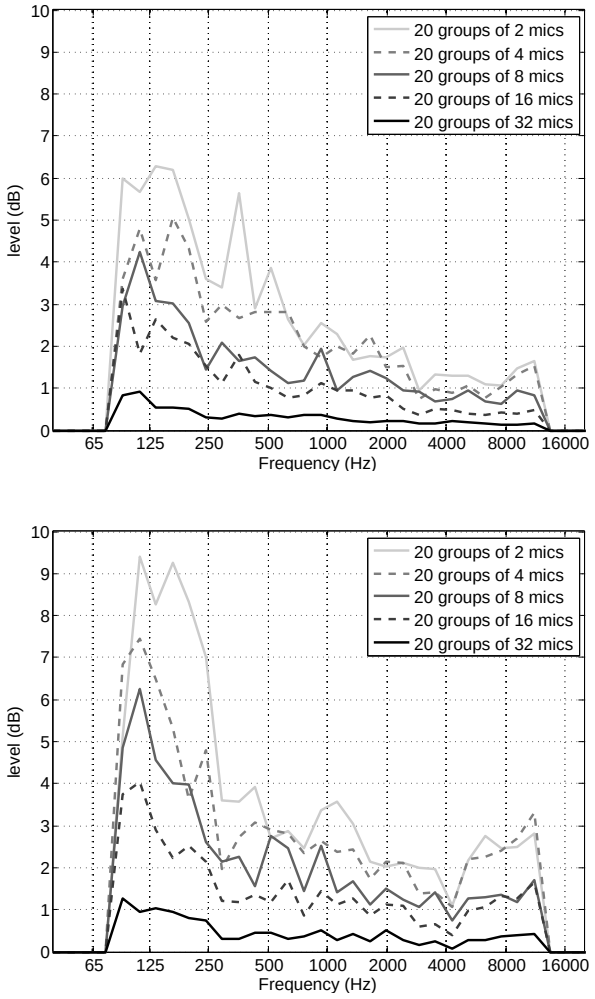

Figure 7: Maximum differences in averaged responses across twenty different subsets of 2, 4, 8, 16 and 32 microphones in the dubbing stage. Top: screen left channel; bottom: surround speaker Lss2.

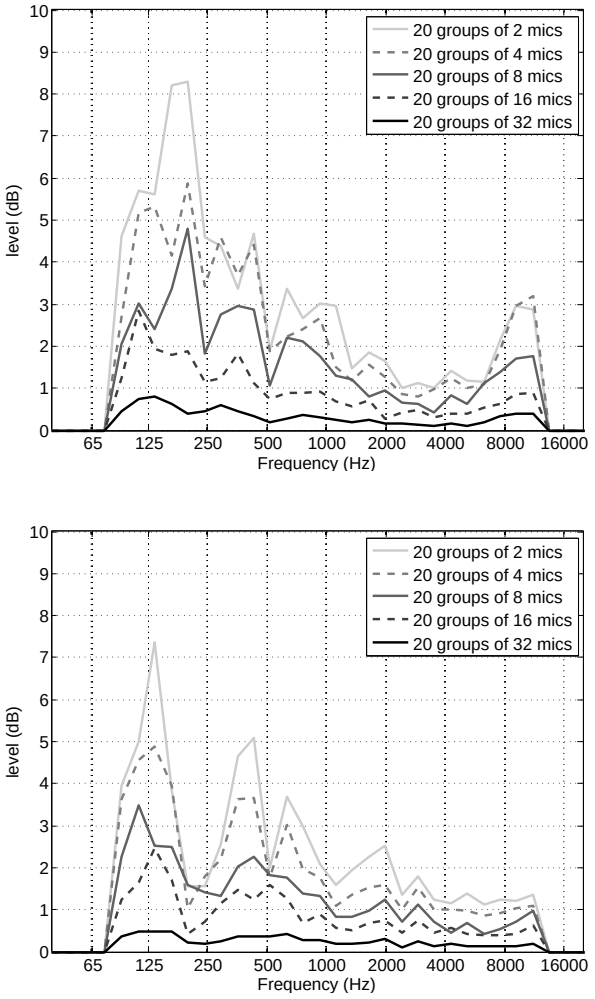

Figure 8: Maximum differences in averaged responses across twenty different subsets of 2, 4, 8, 16 and 32 microphones in the cinema. Top: left channel; bottom: surround speaker Lss2.

5 CONCLUSIONS

The analysis presented here might help understanding better the procedure of response averaging over multiple microphones. The interpretation of the method as a Monte Carlo approximation to a certain integral allowed the precise definition of i) the precise magnitude the averaging method converges to, and ii) the rate of convergence as a function of the number of microphones used.

The variability due to locating the available microphones differently is shown to decrease as $\sigma_A(S)/\sqrt{N_{mic}}$ both theoretically and empirically. This scaling behavior is valid at each frequency. The variance of the spectrum across the listening area, $\sigma_A^2(S)$, does not depend on N_{mic}; it captures a footprint specific to the loudspeaker-room system. As expected, this intrinsic variability depends on frequency, as shown in the analysis of the two rooms considered, with larger effects at low frequencies.

The empirical results have highlighted that using just one or two microphones leads to excessive variations, and therefore, lack of repeatability. Eight microphones allowed to restrict the maximum deviation within 4dB in the analyzed cinema and 6dB in the smaller dubbing stage. The rate of convergence indicates that there is little gain in using more than eight positions, especially considering the extra effort that it would require.

Our results evidence that care should be taken when a particular set of measurements shows high variation between microphones, as this indicates a large chance of disagreement between two different installers, if a strict approach of adjusting the average to a target curve is pursued. In such cases, it is necessary to keep in mind that the correction might just be an aberration due to a poor sampling of the listening area; moving the microphones might even result in an average requiring less equalization.

Although our experimental data are limited to one room per type, the responses of the speakers in these rooms and their overall building quality are representative of the state of the art, so we expect our results to apply in the majority of other rooms with similar size and acoustical properties, in particular regarding the maximum amount of variation between different subsets of microphones.

One thing we would like to point out is that the analy-

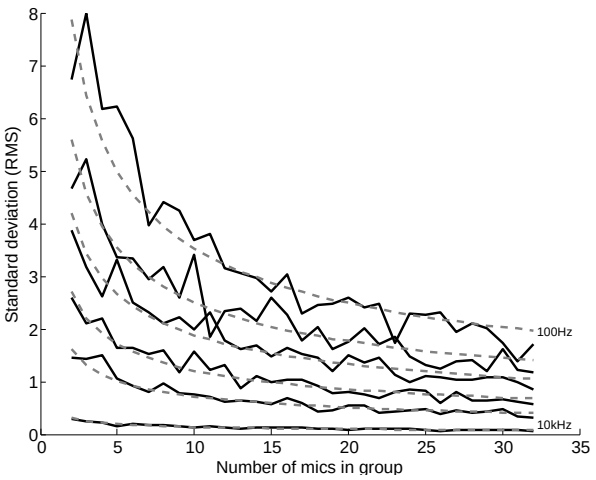

Figure 9: Variance of the averaged spectra due to different microphone positioning, as a function of the number of microphones used. Black lines follow empirical values; grey dashed lines follow the theoretical $1/\sqrt{N_{mic}}$ prediction. Each pair of black-grey lines corresponds to one fixed frequency band. From top to bottom: 100Hz, 400Hz, 500Hz, 2kHz, 4kHz and 10kHz.

sis presented so far did inevitably ignore the human aspect behind the measurements, namely the interpretation of the measured curves and the assessment of the perceived sound, since the focus was solely on the physical variation of sound within a selected area. For example, when computing filters given an average response, no question was posed on whether the outcoming correction was sensible or perceptually desirable at all. Ultimately, listening to the sound of the speaker in the room and switching the EQ on and off will often remove any doubt about the necessity and amount of the correction.

6 ACKNOWLEDGEMENTS

The authors would like to thank the following people: Louis Fielder, Charles Robinson, Mark Davis and Sunil Bharitkar for their valuable input in the discussions related to this topic; Xavi Pitarch for helping with the measurements; two anonymous reviewers for their comments and suggestions.

References

[1] SMPTE 202-2010: Motion-Pictures - Dubbing Theaters, Review Rooms and Indoor Theaters - B-Chain Electroacoustic Response.

[2] P. Newell, G. Leembruggen, K. Holland, J. Newell, S. Torres-Guijarro, D. Gilfillan, D. Santos-Dominguez, S. Castro, "Does 1/3 Octave Equalisation Improve the Sound in a Typical Cinema?", Proceedings of the Institute of Acoustics, Vol. 33, Pt 6, Reproduced Sound 27th conference, Brighton, UK (Nov. 2011).

[3] P. Newell, K. Holland, J. Newell and B. Neskov, "New Proposals for the Calibration of Sound in Cinema Rooms", presented at the AES 130th convention, London, UK, 2011, May 13–16.

[4] J. A. Pedersen, 'Sampling the energy in a 3D Sound Field", presented at the AES 123rd convention, New York, USA, 2007, October 5-8.

[5] S. Weinzierl, "Introduction to Monte Carlo methods", lectures at NIKHEF, the National Institute for Nuclear Physics and High Energy Physics in Amsterdam, Holland, 2000, http://arxiv.org/pdf/hep-ph/0006269.pdf.

[6] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique," presented at the AES 108th convention, Paris, France, 2000 February 19–22.

[7] I. Allen, "The X-Curve: its Origins and History", SMPTE Motion Imaging Journal, July/August 2006.

[8] L. Fielder, 'Frequency Response versus Time-of-Arrival for Typical Cinemas", SMPTE Annual Technical Conference & Exhibition, Hollywood, USA, 2012.