Audio Engineering Society

Convention Paper 10624

Presented at the 153rd Convention
2022 October

Validation Results of Deconvolution of Room Impulse Responses from Simultaneous Excitation of Loudspeakers

Ritesh Banka[1] and Sunil Bharitkar[1]

[1] *Samsung Research America, DMS Audio, Valencia CA, USA*

Correspondence should be addressed to Sunil Bharitkar (`s.bharitkar@samsung.com`)

ABSTRACT

Traditional room-equalization involves exciting *one loudspeaker at a time* and deconvolving the loudspeaker-room response from the recording. As the number of loudspeakers and positions increase, the time required to measure loudspeaker-room responses will increase. We presented a technique to deconvolve impulse responses after *exciting all loudspeakers simultaneously* [1]. This paper presents the results of our de-convolution method compared with the traditional approach in real listening environments. We will compare the results of three different stimuli we used for testing and validating our approach: 11-channel, 7-channel, and 4-channel time-shifted log-sweeps. We measured the loudspeakers in 3 different room settings: ITU standard-based listening room, reference room, and home environment. The performance results are depicted in plots comparing the true (single-channel-at-a-time) responses with the responses obtained from the proposed approach. We also present objective metrics using log spectral distortion.

1 Introduction

This paper is an extension of the work Simultaneous Deconvolution of multiple room impulse responses [1]. Equalization of loudspeaker-room impulse responses is required to minimize the impact of audible resonances and spectral dips for creating high-quality spatial and immersive audio [1],[2],[3],[4],[5],[6]. The first step involves deconvolving the loudspeaker-room response $h_{i,j}(n)$, from an in-room measurement at a microphone position j, by exciting each loudspeaker i sequentially with an excitation signal (viz., stimuli) as shown in Fig. 1. The next step involves designing the equalization filter per loudspeaker. The traditional approach involves exciting one loudspeaker with a stimulus signal and extracting the loudspeaker-room impulse response. The stimulus is usually deterministic (log-sweep, maximum length sequences) or stochastic (white noise, pink noise, multitone). As the number of loudspeakers and positions increases, the time required to make room responses available from all loudspeakers will increase since the measurement system will measure each loudspeaker-microphone response before moving to the next loudspeaker in a multichannel setup. In [1] we described a method for deconvolving multiple impulse responses from loudspeakers with a single measurement, significantly reducing the time to equalize the speakers. We mainly emphasized 11-channel(7.1.4)

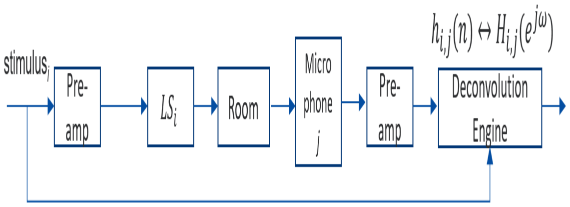


Fig. 1: Traditional approach of room response, $h_{i,j}(n)$, deconvolution for loudspeaker i and microphone position j.

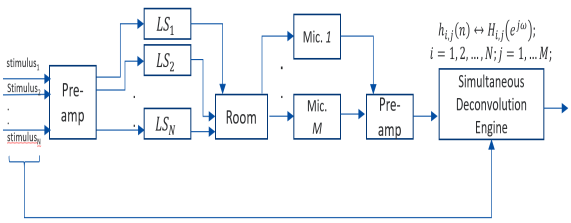


Fig. 2: Presented approach of simultaneous room response, $h_{i,j}(n)$, deconvolution for all loudspeakers *i=1,2,...,N* and microphone position *j* (multiple microphone extension being straightforward).

speakers set up to prove the working of our method. The real-world environment is far more complex than a reference room or single speaker setup. It is prevalent to have 4-channel or 7-channel speaker setups at home. Generally, the rooms with speaker systems also contain furnishing, contributing to the room response.

This paper *extends* the work [1] (based off of simulations) by validating the approach of the simultaneous deconvolution, as depicted in Fig. 2, using conventional log-sweep stimuli in real-world cases with ambient noise present. Section 2 presents the validation results of deconvolution method in all three rooms where the measurements were performed. Section 3 provides the results of objective metrics of the measurements. *The stimulus-length and circular-shift amount for each channel optimizations were done as described in [1] based on minimizing the log-spectral distortion metric*. Section 4 provides detail about the measurement system that was used to perform the measurements. Section 5 concludes/summarizes the paper.

2 Validation of deconvolution method

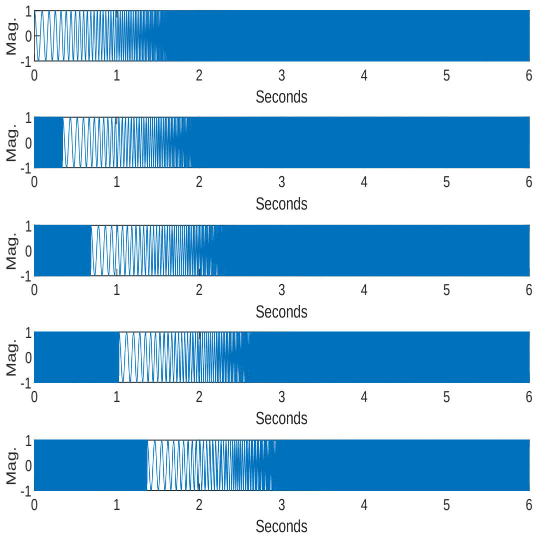
2.1 Room 1: 11-channel speaker system

The 11-channel(7.1.4) was tested in a reference room designed by Audio Lab in Valencia. Figure 3 shows the picture of the room setup with the speakers. The dimensions of the room are 3.04m x 7.01m x 5.33m. Fig 3 shows the setup of the speakers in the room.

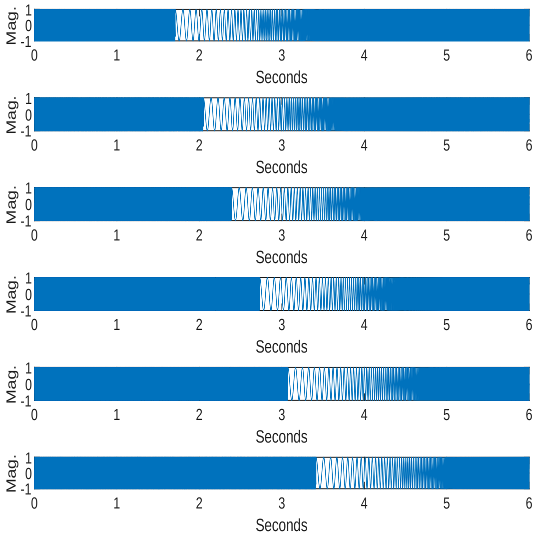
Fig. 3: Room with 11 speaker channels.

We use Genelec 8050B studio monitors as source for validation our method in this room. Since the speaker doesn't have any response below 40 Hz we only show the frequency response plots from 40-20000 Hz. The same frequency bandwidth is used for calculating the log-spectral distortion metric.

Fig 4 shows the 11-channel circularly-shifted log sweep stimulus based on the optimization of the mean log-spectral distortion [1].



(a) Channel 1 to Channel 5



(b) Channel 6 to Channel 11

Fig. 4: The 11-channel circularly-shifted log sweep stimulus.

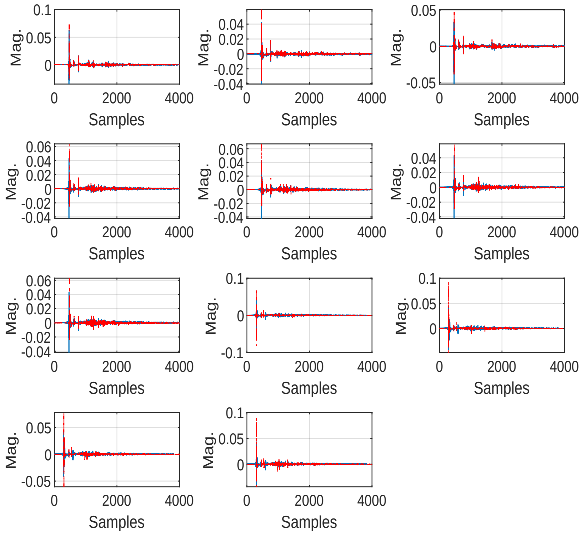


Fig. 5: Example of 11 Impulse Responses measured in the Reference Room.

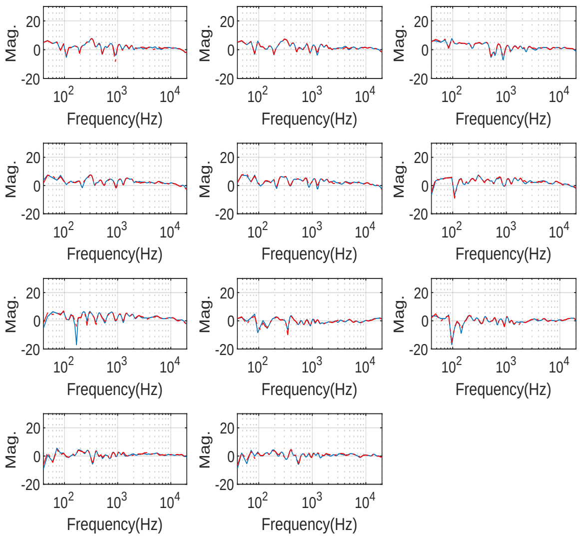


Fig. 6: Example of 11 Frequency Responses measured in the Reference Room.

To get the *true* response of each speaker, as a baseline to compare with the simultaneous deconvolution, a mono log sweep is played from each loudspeaker one at a time and measured the response at the single microphone.

2.2 Room 2: 7-channel speaker system

Room number 2 is a reverberant listening room. It's dimensions are 2.74m x 6.4m x 4.87m. Fig 7 shows the 7-channel time shifted log sweep stimulus.

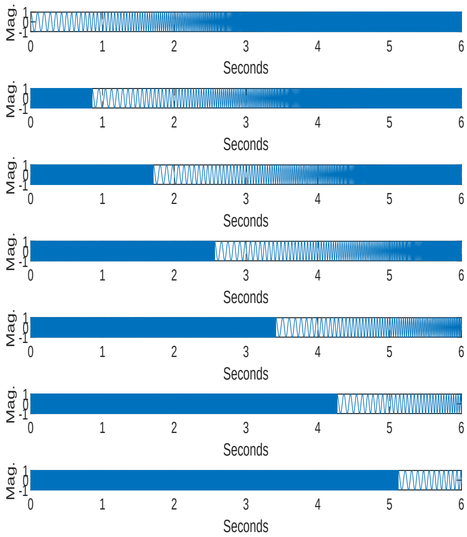


Fig. 7: Time shifted 7-channel stimulus optimized in Time Domain.

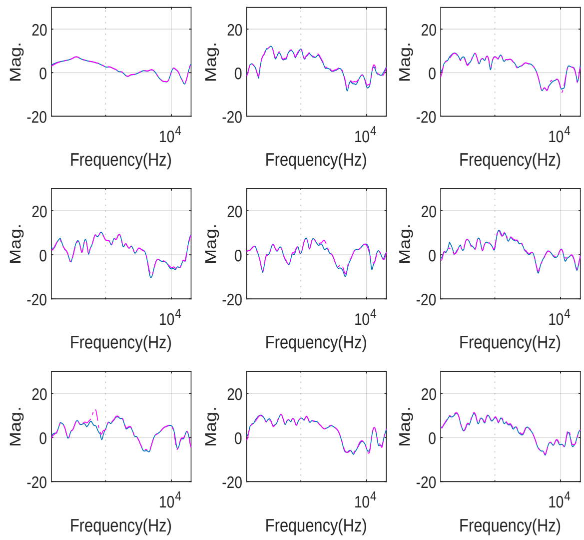


Fig. 9: Example of 9 Frequency Responses measured in the listening room.

We use custom built speakers with 4" driver as source for validation our method in this room. Since the speaker doesn't have any response below 150 Hz we only show the frequency response plots from 150-20000 Hz. Same bandwidth is used for calculating the log-spectral distortion.

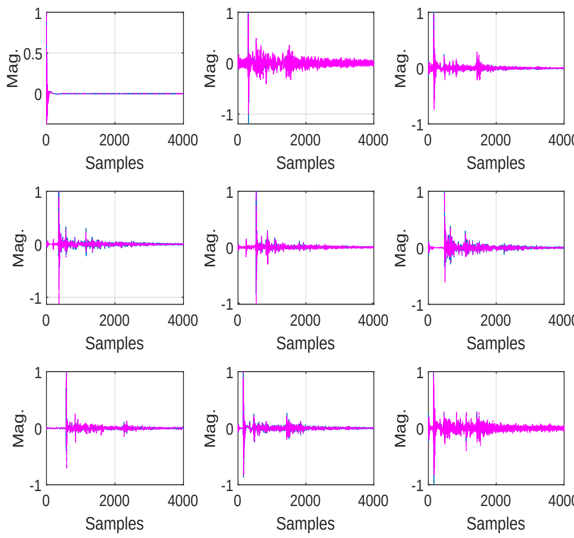


Fig. 8: Example of 9 Impulse Responses measured in the listening room.

2.3 Room 3: 4-channel speaker system

Room number 3 has the characteristics of a sound system placed in a typical home environment. It's dimensions are 3.04m x 6.55m x 4.11m. Fig 10 shows the setup of the speakers in the room.

Fig. 10: Room with 4 speaker channels placed like in a typical home environment.

We use the same speaker we used in Room 2 to validate our method and use 150-20000 Hz bandwidth to analyze the results.

Fig 11 shows the 4-channel time shifted log sweep stimulus.

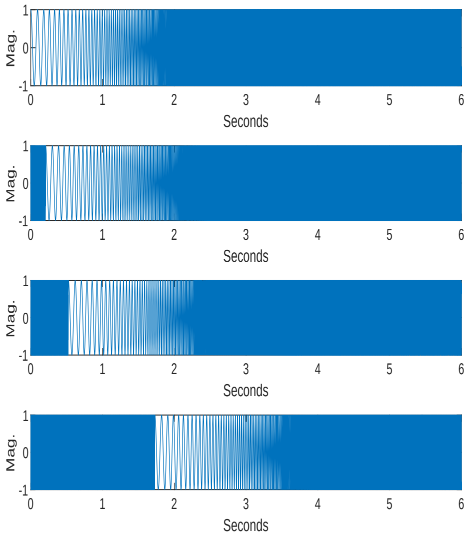


Fig. 11: Time shifted 4-channel stimulus optimized in Time Domain.

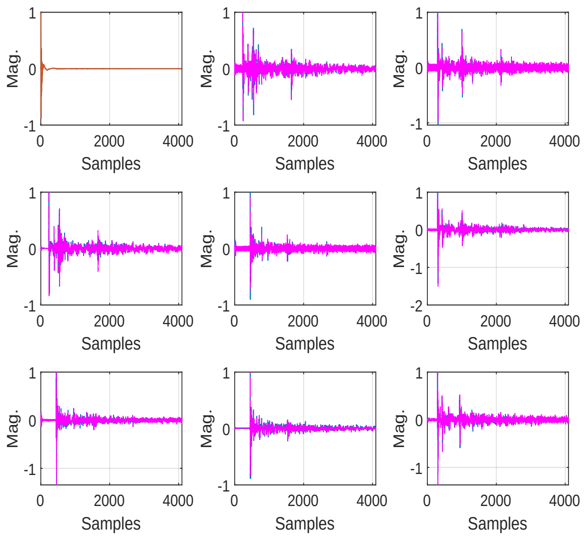


Fig. 12: Example of 9 Impulse Responses measured in Room 3.

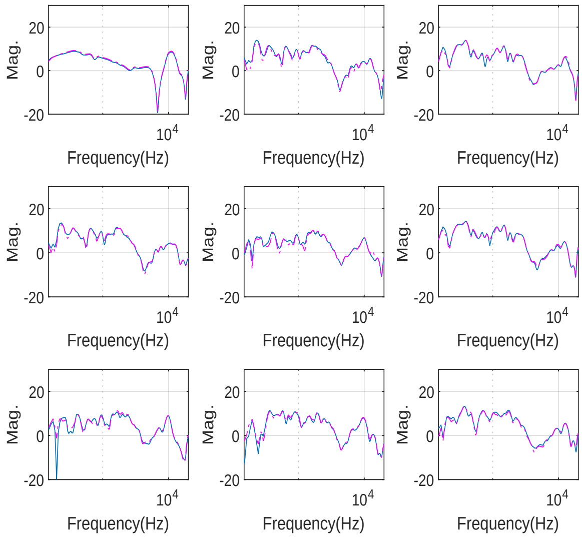


Fig. 13: Example of 9 Frequency Responses measured in Room 3.

3 Results

The performance of the simultaneous deconvolution techniques measured in each three rooms can be assessed as a function of the duration of the stimuli and the amount of circular shift M. The performance metrics used are: (i) the frequency domain error between true and estimated responses, $e_i(n) = h_i(n) - \hat{h}_i(n)$ for the i-th loudspeaker, and (ii) the mean log spectral distortion metric ϕ_{SD} (dB) expressed as a ratio between the true magnitude response $|H_i(e^{j\omega})|$ and the estimated $|\hat{H}_i(e^{j\omega})|$

$$\begin{aligned} \phi_{SD,i} &= \sqrt{\frac{1}{(\omega_2-\omega_1)}\int_{\omega_1}^{\omega_2}[10\log_{10}\frac{|H_i(e^{j\omega}|)}{|\hat{H}_i(e^{j\omega})|}]^2 d\omega} \\ \phi_{SD} &= \sum_{i=1}^{11}\phi_{SD,i} \end{aligned} \quad (1)$$

where, ω_1 and ω_2 are frequency domain over which the metric is computed. In the present case $[f_1, f_2] = [0.2, 20]$ kHz are the passband cutoff frequencies of the responses.

Table 1 shows the results of log-spectral distortion for the various rooms with different noise powers. Figures 6, 9, and 13 confirms the match is extremely close between true and estimated responses for the various logsweep stimuli circularly shifted relative to each

Table 1: Mean (over all channels) log-spectral distortion and noise power in room

	Log-spectral distortion	Noise Power (dB)
Room 1	3e-13	-47
Room 2	2.68e-14	-82
Room 3	1.39e-13	-98

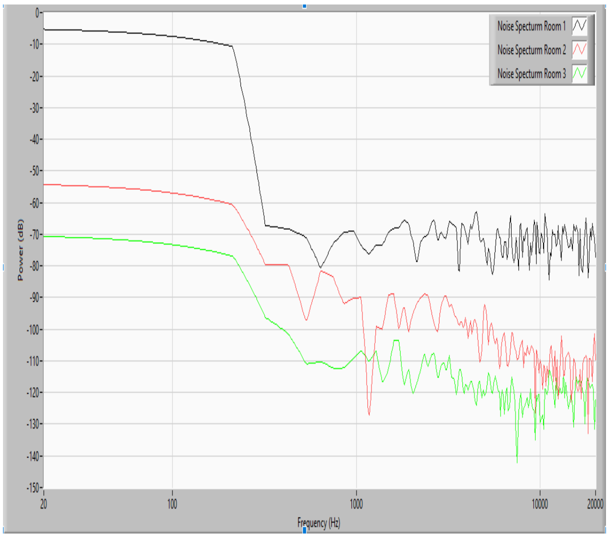


Fig. 14: Figure shows the noise power density in each room.

other using the optimizations from [1]. For most of the channels the match is very good with minor differences showing up for the estimate between 200 and 400 Hz. However, the perceptual differences after loudspeaker equalization using either the estimated simultaneously deconvolved response (in red) or the true responses (in blue), will be insignificant due to minor differences between the two spectrum.

Figure 14 shows the noise power density in each room where the measurements were taken.

A typical length of the stimulus that was used for measurement was about 6 seconds long. With the presented simultaneous deconvolution method, the time required to perform the measurements for system tuning can be reduced from minutes to only a few seconds. The autonomous nature of the method makes it viable to be used for consumer (home-theater) and professional (cinema-processors) products.

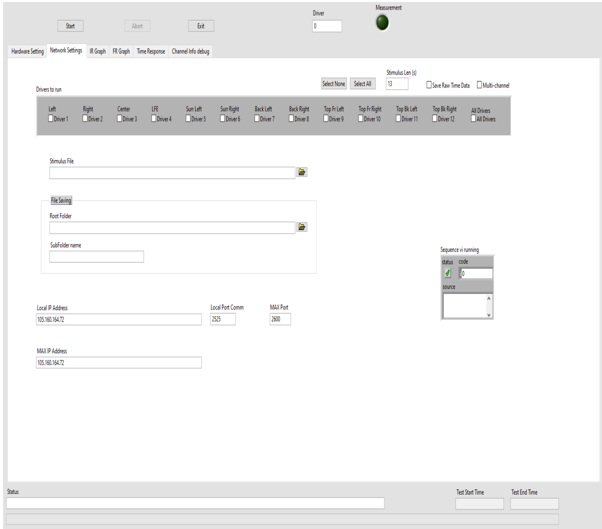


Fig. 15: Graphical Interface of the tool used to perform the measurements and process the data.

4 Measurement System

The measurements were conducted with National Instruments hardware and we developed our own measurement tool in-house in LabVIEW [7] and MAX [8] to perform the measurements and process the data. For single microphone measurements, we used NI USB-4431 to acquire the microphone signal and BSS BLU-160 to playback the multichannel signal. For multiple microphone measurements, we used NI-PXI chassis with PXIe-4497 card which can capture up to 16 input channels simultaneously. Since the programs developed on LabVIEW and MAX were running on separate clocks, we added white noise at the beginning of stimuli to time-align the signals. For single microphone measurements, we used the Genelec 8050B studio monitors as source. For multiple microphone measurements, we used custom made speakers with 4" driver having a microphone at the front of the diaphragm.

Figure 15 shows the front panel of the graphical interface written in LabVIEW that was created to perform the measurements and analysis.

Figure 16 shows the block diagram of the complete measurement chain.

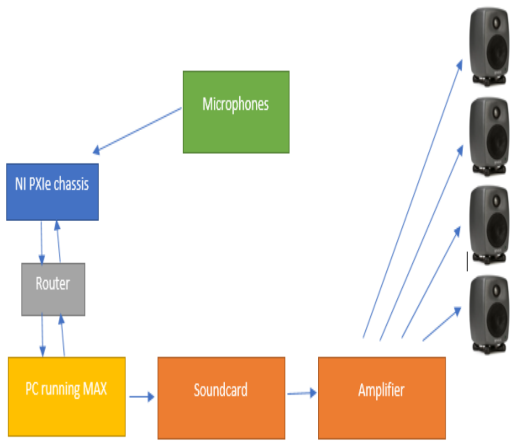


Fig. 16: Block Diagram of the measurement system.

5 Conclusions and Future Directions

The presented simultaneous deconvolution technique, emphasizing three different loudspeaker systems: 11-channel, 7-channel, and 4-channel works well for widely popular log-sweep stimuli by proper circular-shifting of the stimuli and cross-correlation techniques. The technique also works well, whether the measurement is done with a single microphone or multiple microphones in various real-world noise conditions. In future work, we will explore whether the approach can be replicated in different home environments with longer reverberation time and lower signal-to-noise ratio.

6 Acknowledgement

Samsung Electronics and Samsung Research America supported this work. The authors would like to thank the entire staff of Samsung's US Audio Lab, who helped with all aspects of this research, offered insightful suggestions, and contributed to this work.

References

[1] Bharitkar, S. and Kyriakakis, C., *Immersive Audio Signal Processing*, 2006.

[2] Cecchi, S., Carini, A., and Spors, S., "Room Response Equalization—A Review," *Applied Sciences*, 16(8), 2018.

[3] Toole, F. and Olive, S., "The Modification of Timbre by Resonances: Perception and Measurement," *Journal of Audio Engineering Society*, (36), pp. 122–142, 1988.

[4] Bharitkar, S., Robinson, C., and Poulain, A., "Equalization of Spectral Dips Using Detection Thresholds," *Proc. 140th AES Conv., Paris, France*, 2016.

[5] Fielder, L., "Practical Limits for Room Equalization," *Proc. 111th AES Conv., New York, NY, USA*, 2001.

[6] Faller, C., "Modifying Audio Signals for Reproduction with Reduced Room Effect," *Proc. 147th AES Conv., New York, NY, USA*, 2019.

[7] "National Instruments LabVIEW," `https://www.ni.com/en-us/shop/labview.html`, 2022, Accessed: 2022-07-14.

[8] "Cycling 74 MAX," `https://cycling74.com/products/max`, 2022, Accessed: 2022-07-14.