Audio Engineering Society

Convention Paper 10684

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A Comparison Between Sweep Signals for Simultaneous Deconvolution

Sunil Bharitkar[1]

[1] *DMS Audio Lab, Samsung Research America*

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

ABSTRACT

This paper compares the conventional logarithmic sweep and a perfect sequence sweep (PSEQ) for simultaneous deconvolution of multichannel loudspeaker-room impulse responses using a single time domain measurement. The optimal parameters (including sweep duration) for each sweep are determined using machine learning (ML) techniques on training data of noise-free room impulse responses. Each of these optimized sweeps is evaluated on a noise-less and noisy room-response test set to assess simultaneous deconvolution performance in the absence/presence of noise. The results show that the optimized conventional log-sweep, while shorter in duration, is better suited for noisy conditions. In contrast, the PSEQ sweep works well under conditions with little background noise. The paper compares both optimized sweeps with the shortest-length bandwise minimum-noise (BMN) sweep. Both optimized sweeps (viz., conventional and PSEQ) produce lengths shorter than the BMN sweep for simultaneous deconvolution with a single time-domain measurement. Thus, depending on the measurement time requirement and background noise condition, on the basis of this research, there is a clear consensus on which sweep would be best suited for simultaneous deconvolution.

1 Introduction

Calibrating loudspeakers in a room is essential to deliver high-quality, immersive spatial audio. The calibration involves computing and compensating for the time difference in arrival between loudspeaker signals at a listening position, level matching between loudspeakers, and frequency-dependent correction/equalization per loudspeaker by measuring the loudspeaker-room impulse responses (where the loudspeaker being the source that excites the room)[1]. The state-of-the-art (SoTA) approach involves deconvolution of the response $h_{i,j}(n)$, from a room measurement at a microphone position j, by exciting a loudspeaker i with an excitation stimulus ([5]-[18], [19], [30]-[34]), and then sequentially proceeding to the next loudspeaker. Huang [20] provides a tutorial on the identification of blind and non-blind multiple input multiple output (MIMO) systems. Antweiler *et al.* [26] present an adaptive filtering technique to perform the identification of multiple input single output (MISO) systems using perfect sequences (PSEQ) [27] that satisfy orthogonality[2], are odd symmetric in phase and have a flat magnitude spectrum [28]. Furthermore, [21] compares different measurement techniques with various stimuli as a function

[1] We interchangeably use room response and loudspeaker-room response

[2] PSEQ-sweep of length N periodically applied as an excitation signal has a periodic impulse like autocorrelation property

of noise [21]. Nakahara *et al.* [29] present a short-duration sweep optimized for single-channel deconvolution by ensuring constant normalized noise power in the frequency bands called bandwise minimum-noise sweep (BMN). Prawda *et al.* [22] presents a technique to detect measurements using logarithmic sweep which are corrupted by noise and select clean ones based on a procedure called the rule of two.

The above stimuli and techniques are based on measuring a single source/loudspeaker at a time before measuring the next source. The disadvantages of sequential measurements become apparent when the number of loudspeakers and positions increases, as the time required to deconvolve room responses from all loudspeakers increases. Furthermore, repeated measurements to increase the SNR will add to the measurement time. Newer approaches to address this limitation include the work presented by Majdak *et al.* [23] and Weinzierl *et al.* [24] based on interleaving or partial overlap sweep stimuli. Bharitkar [1], [3] present a technique for simultaneous deconvolution (i.e., exciting all loudspeakers at the same time and deconvolving responses using a single time domain measurement) using a logarithmic sweep signal. We incorporate this technique into the present study.

The advantages of logarithmic sweep over other excitation signals are provided in [21] for single channel deconvolution, whereas [25] shows the advantages of logarithmic sweep over the maximum-length sequence and multitone signal. Given the wide use and advantages of sweeps cited for room response measurements, this article compares the performance of different sweeps [35], [28], [29] (where sweeps were designed with unique properties) in the context of simultaneous deconvolution. Additionally, for consumers' measurement of room acoustics, a shorter measurement time (using a shorter excitation stimulus) is preferable, with the stimulus being robust to noise. For example, the consumer would consider a 30-second sweep undesirable, as this sweep would take at least 30$\times$ 12 loudspeakers = 6 minutes, to measure an ITU 7.1.4 home loudspeaker setup [36].

This research addresses two questions: (i) Does one of the two sweeps [35], [28] give a lower measurement time, in the absence of noise, with a small reconstruction error for a multichannel loudspeaker setup, and (ii) Does one optimized sweep outperform the other optimized sweep, in the presence of noise, at various SNRs. To address the first question, we incorporate two stochastic optimization techniques using a training dataset of room responses to design the sweeps by optimizing the sweep parameters (viz., duration and offset of sweeps relative to each other for each of the two sweeps). The latter question is answered by deconvolving the optimized sweeps with a test set of noisy room responses and measuring performance. The optimized sweeps are also compared to the BMN sweep [29], another technique used to design a sweep with a short duration based on a prior measurement of the noise present in the responses.

Section 2 summarizes the properties and construction of sweeps for simultaneous deconvolution for an ITU 7.1.4 setup [36] comprising 11 main loudspeakers. Section 3 presents the stochastic optimization algorithms that optimize sweep parameters (including the measurement time). Section 4 presents objective results (MSE) on a test set for the two optimized sweeps, the noise robustness of each optimized sweep, and the comparisons with the BMN sweep. Section 5 concludes and summarizes the paper.

2 Excitation Stimulus

2.1 Logarithmic Sine-sweep

In the case of an exponential sweep [35], assuming that ω_1 and ω_2 are the start and end frequencies (in rad/s), with a total duration of T_{log} seconds, the logarithmic sweep signal $x(t)$ is

$$x(t) = \sin\left(\frac{\omega_1 T_{\log}}{\log\frac{\omega_2}{\omega_1}}\left(e^{\frac{t}{T_{\log}}\log\left(\frac{\omega_2}{\omega_1}\right)} - 1\right)\right) \qquad (1)$$

whereas the discrete-time equivalent is $\mathbf{x}_1(n) = (x(n), x(n-1), x(n-2), \ldots, x(n-(P_{log}-1)))^T$, T represents the vector transpose and $P_{log} = T_{log}/T_s$ in samples[3].

2.2 Perfect Sequence (PSEQ)-Sweep

A perfect sequence [26], [27] $p(m)$ of duration M is a discrete sequence whose autocorrelation satisfies a Kronecker-delta impulse property. Antweiler [28] showed the advantages of a sweep constructed using the same autocorrelation property as a perfect sequence

[3] $T_s = 1/f_s = 1/48000$ (s), f_s is the sampling frequency, $T_{\text{stimuli}} = P_{log}/48000$ is duration of stimuli in seconds used in Table 1

in fast-tracking a time-varying system with robustness to nonlinear distortion. The sweep is constructed in the frequency domain with a flat magnitude spectrum and a linear group delay and is expressed as a time-stretch pulse with a stretch factor m and a length L of one period,

$$S(\omega) = \begin{cases} \exp(\frac{-(\sqrt{-1})4m\pi\omega^2}{L^2}) & 0 \le \omega \le L/2 \\ S^*(L-\omega) & L/2 < \omega < L \end{cases} \quad (2)$$

2.3 Simultaneous Deconvolution

The measurement representation, for simulations, is a linear convolution sum between the ground-truth loudspeaker-room response $\mathbf{h}_i$ and the stimuli $\mathbf{x}_i(n)$ with additive measurement noise $\mathbf{v}(n)$,

$$\mathbf{y}(n) = \sum_{i=1}^{N=11} \mathbf{x}_i(n) \circledast \mathbf{h}_i + \alpha \mathbf{v}(n) \quad (3)$$

where α controls the SNR (dB) and is expressed as a ratio of the *average* stimulus energy to the average background noise energy over the duration of the stimulus. In the absence of noise, $\alpha = 0$.

1. Log-sweep: The N-channel (with $N = 11$) log-sweep stimulus for simultaneous deconvolution is constructed by circularly shifting the log-sweep for each channel/loudspeaker stimulus relative to the stimulus applied to channel/loudspeaker 1.

$$\begin{aligned} \mathbf{x}_1(n) &= [x(n), x(n-1), \ldots, x(n-P_{log}+1)]^T; \\ \mathbf{x}_i(n) &= [x(< n-(i-1)M >_{P_{log}}), \\ &\ldots, x(< n-(i-1)M-1 >_{P_{log}}), \\ &\ldots, x(< n-(i-1)M-P_{log}+1 >_{P_{log}})]^T; \end{aligned} \quad (4)$$

with $i = (2, \ldots, 11)$, $< m >_{P_{log}} = m$ modulo P_{log}, and $\mathbf{h}_i = [h_i(1), h_i(2), \ldots, h_i(K)]^T$ is a K-length impulse response. The loudspeaker-room responses are deconvolved using,

$$\begin{aligned} \hat{H}_j(e^{j\omega}) &= \frac{S_{\mathbf{x}_j,\mathbf{y}}(e^{j\omega})}{S_{\mathbf{x}_j,\mathbf{x}_j}(e^{j\omega})} \\ \hat{\mathbf{h}}_j &= \mathscr{F}^{-1}\{\hat{H}_j(e^{j\omega})\} \end{aligned} \quad (5)$$

where, $S_{\mathbf{a},\mathbf{b}}(e^{j\omega}) = \mathscr{F}\{\mathbf{a}(n)\}\mathscr{F}\{\mathbf{b}(n)\}^*$ and * denotes conjugation.

2. PSEQ-Sweep: The PSEQ sweep is not amenable to cyclic shift as two cyclically shifted sequences do not yield a Kronecker delta autocorrelation function. This is shown with an example in Fig. 1, where the second PSEQ sweep is cyclically shifted relative to the first. The autocorrelation sequence shows artifacts and ringing. If the second PSEQ sweep is applied at the end of the first sweep, perfect autocorrelation is guaranteed [27]. But this is equivalent to single-channel at a time (traditional) deconvolution. In this study, we determined that with the proper offset between the PSEQ sweeps, it is possible to retain orthogonality between them. Figure 2 shows an example of the sweep constructed using one of the proposed ML optimization approaches (viz., simulated annealing) for an 11 channel (loudspeaker) [4] setup. The PSEQ sweeps are offset relative to each other, and an example correlation between two sweeps is shown in the right panel.

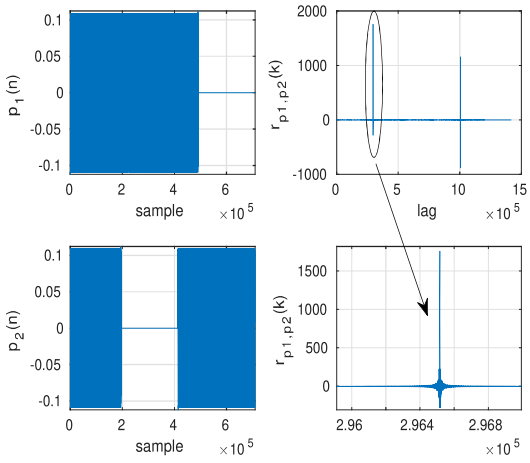


Fig. 1: PSEQ sequence (top-left), circularly shifted version (lower-left), cross-correlation result (top right and lower-right panel showing zoom-in version of panel above).

The PSEQ-sweep of length L (excluding any silence region) is denoted as $\mathbf{s}(n) \in \Re^{1\times L}$ and amount of offset Δ between PSEQ-sweeps provides the duration P_{PSEQ} of the stimuli $\mathbf{x}_j \in \Re^{P_{\text{PSEQ}}\times 1}$ using,

$$P_{\text{PSEQ}} = L + \Delta(N-1) \quad (6)$$

In the context of simultaneous deconvolution, the multichannel PSEQ-sweep stimuli for $N = 11$ loudspeaker setup is constructed from a single sweep $\mathbf{s}(n)$ obtained from (2) to ensure orthogonality between successive

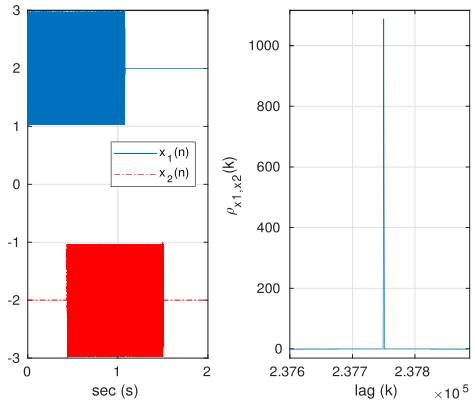


Fig. 2: Simulated annealing-based overlapping PSEQ sweeps (only two of the $N = 11$ shown in the left panel) with perfect cross-correlation between them (shown in the right panel).

sweeps, using

$$\begin{aligned} \mathbf{x}_1(n) &= [\mathbf{s}(n) \quad \mathbf{0}_{1\times(P_{\text{PSEQ}}-L)}]^T \\ \mathbf{x}_i(n) &= [\mathbf{0}_{1\times(N-12+j)\Delta} \quad \mathbf{s}(n) \quad \ldots \\ &\quad \mathbf{0}_{1\times(P_{\text{PSEQ}}-L-(N-12+j)\Delta)}]^T \end{aligned} \tag{7}$$

The impulse response $\hat{\mathbf{h}}_j$ are deconvolved using cross-correlation in time between the measurement and the corresponding offset PSEQ $\mathbf{x_j}(n)$,

$$\hat{\mathbf{h}}_j = \rho_{(\mathbf{x}_j(n),\mathbf{y}(n))} \tag{8}$$

The parameters to be determined by optimizing a reconstruction error function (viz., the mean square error between the true response and estimated response over a training dataset of room responses) for each sweep are: (i) log-sweep: duration P_{log} and cyclic-shift amount M, (ii) PSEQ-sweep: duration P_{PSEQ} and offset Δ. Once optimized stimuli are obtained from room response training data, these stimuli are tested on a larger set of responses to validate performance. We have shown [2] that the traditional logarithmic sweep designed using an ML optimization technique is able to estimate room responses in the real world.

3 Sweep Parameter Optimization

The ground-truth room impulse responses $\mathbf{h}_j$ used in this article are from the MARDY [37][4] and MeshRiR [38][5] databases. The MARDY database has 72 loudspeaker-room responses, while MeshRiR has 14112 responses. Based on the combined dataset, the number of 11 channel responses available for training and testing is binomial $\binom{14112+72}{11} \approx 10^{38}$. In our optimization, depicted in Algorithm 1, we set 11 hyperparameters: (a) duration P, and (b) circular right shifts M_i $(i = 1,\ldots,10)$ for the log sweep using (4). For PSEQ-sweep, the hyperparameters were (a) duration L and (b) offset amount Δ between PSEQ-sweeps (6)-(7). The durations $T_{log} = 1/P_{log}$ and $T_{PSEQ} = 1/P_{PSEQ}$ are limited to $\{5,15\}$ (seconds), $M_i \in \{4096, 131072\}$ samples, and $\Delta \in \{19200, 22200\}$. Responses are input to the optimization process that optimizes the sweep hyperparameters by minimizing a metric $\bar{\psi}_{SD}$, where

$$\bar{\psi}_{SD} = \frac{1}{R}\sum_{k=1}^{R}\sqrt{\frac{1}{11}\sum_{j=1}^{11}\|\hat{\mathbf{h}}_j^{(k)} - \mathbf{h}_j^{(k)}\|_2^2} \tag{9}$$

is the root-mean-square error (RMSE) averaged over the training set of size $R = TR$. In the case of evaluating the performance on an independent test set of room responses, $R = TS$ where TS is the size of the test set.

3.1 Bayesian Optimization

Bayesian optimization [40] is a global hyperparameter optimization technique constrained by the limits of the hyperparameters that is best suited for optimization with 20 or fewer hyperparameters. The method builds a surrogate function for the objective and quantifies the uncertainty in that surrogate using Gaussian process regression. Furthermore, several parameters are used for initialization, including the type of acquisition function that guides the sampling of optimal hyperparameters [41]. Recent advances can be found in [42], and we have used *bayesopt*[7].

3.2 Simulated Annealing (SA)

Simulated Annealing (SA) models the physical process of heating a material and slowly lowering the temperature to decrease defects, thus minimizing the energy of

[4] https://www.commsp.ee.ic.ac.uk/ sap/
[5] https://github.com/sh01k/MeshRIR
[7] MATLAB function

the system towards the optimal solution. At each iteration of the SA algorithm, a new set of hyperparameters is randomly generated within the limits specified as box constraints, from which the distance of the trial point from the current point is calculated with a probability distribution $Pr(\nabla\bar{\psi}_{SD}) = 1/(1+\exp(\nabla\bar{\psi}_{SD}/T_{anneal,k}))$ with current temperature $T_{anneal,k}$. The algorithm accepts all new points that lower the objective function (9) but also, with a certain probability, points that increase the energy. The annealing schedule used is typically exponential in nature $T_{anneal,k} = T^0/\log(k)$ with T^0 being the initial temperature at $k = 1$. We have used *simulannealbnd*[7].

Algorithm 1: Sweep-parameter Optimization of Stimuli (log-sweep, PSEQ)

Result: Stimuli: (P^*_{log}, M^*_i), $i = 1,\ldots,10$ or (P^*_{PSEQ}, Δ^*); minimum : $\bar{\psi}_{SD}$ (9)

Initialize: Set log-sweep box constraints $\{P_{log,low}, P_{log,up}\}$ samples and $\{M_{i,low}, M_{i,up}\}$ samples or PSEQ box constraints $\{P_{PSEQ,low}, P_{PSEQ,up}\}$ and $\{\Delta_{low}, \Delta_{up}\}$, construct base stimuli $\mathbf{x}_1(n)$ (traditional sweep use (4) or PSEQ use (8)) with initial hyperparameters within bounds of box-constraints, $TR = 500$, true dataset responses $\mathbf{h}^{(k)}_j$; $j = 1,\ldots,11; k = 1,2,\ldots,TR$;

while *time* $\leq$ *MaxTime seconds* **do**

  For each $\hat{P}_{log}$ (or $\hat{P}_{PSEQ}$) and $\hat{M}_i$ (or $\hat{\Delta}$) candidate, construct 11-channel stimuli using (4) or (7);

  Compute the convolution sum (3) using true responses and excitation stimuli with candidate hyperparameters ;

  Estimate the responses using (5)-traditional sweep or (8)-PSEQ;

  Update hyperparameters $(\hat{P}_{log}, \hat{M}_i)$ for log-sweep or $(\hat{P}_{PSEQ}, \hat{\Delta})$ using *bayesopt* or *simulannealbnd* to minimize $\bar{\psi}_{SD}$ (9);

end

$T^*_{\text{stimuli}} = P^*/48000$ (seconds);

4 Results

For *each* stimuli, the test set is of size $TS = 1000$, where each element in the test set comprises 11 responses chosen randomly from the combined dataset (with at least one response present from either of the two databases) as described in Sec. III. As shown in Table 1 (sweep comparisons on a test set of room responses without background noise), the traditional log-sweep (1) gives substantially the same duration stimuli with circular shifting using Bayesian optimization (5.24 seconds) or SA (5.65 seconds). Furthermore, SA resulted in an overlapping PSEQ whose optimal duration was also on the order of 5 seconds. However, the reconstruction error of the PSEQ test set $\bar{\psi}^{\text{test-set}}_{SD}$ calculated using (9) was lower by at least 17 dB over the traditional log sweep. In contrast, Bayesian optimization resulted in a PSEQ sweep that was substantially longer ($\approx$ 3 times) but with a significantly lower test set error of -150 dB.

Table 1: Sweep performance on a noise-free room response test set

Optim.	Param.	log-sweep	PSEQ
Bayes	$T^*_{\text{stimuli}}(sec.)$	5.24	14.78
Bayes	$\bar{\psi}^{\text{test-set}}_{SD}$	-100.4 dB	-150 dB
Simul. Ann.	$T^*_{\text{stimuli}}(sec.)$	5.65	5.39
Simul. Ann.	$\bar{\psi}^{\text{test-set}}_{SD}$	-104 dB	-121 dB

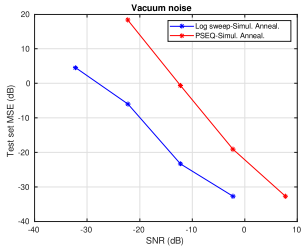
Three different types of noise $\mathbf{v}(n)$ were obtained from the Microsoft Scalable Noisy Speech Dataset (MS-SNSD)[8] and resampled to 48 kHz, where $\mathbf{v}(n)$ is vacuum, air conditioner (stationary) noise, or door closing (impulsive) interference. The lengths of each noise sequence were matched with the length of each stimulus by truncating the noise file or stitching the noise sequence until the stimulus and noise lengths were the same. Accordingly, (3) is modeled with an SNR term as

$$SNR(\alpha) = 10\log_{10}\frac{\sum_{i=1}^{N}\mathbf{x}_i^T(n)\mathbf{x}_i(n)}{\alpha\mathbf{v}(n)^T\mathbf{v}(n)}; \quad (10)$$
$$\alpha_{dB} = \{0,10,20,30\}$$

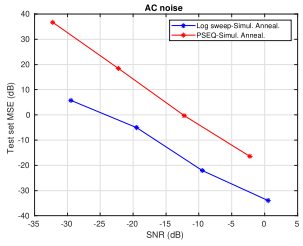
Before measuring the stimuli-dependent SNR, both 11-channel stimuli are set to a fixed loudness of -10 LKFS using the ITU-R BS. 1770-4 [43] recommendation. The mean square error (MSE), $\bar{\psi}^{\text{test-set}}_{SD}(\alpha)$, for various α_{dB} was calculated for the various stimuli. Figures 3 and 4

[8] https://github.com/microsoft/MS-SNSD

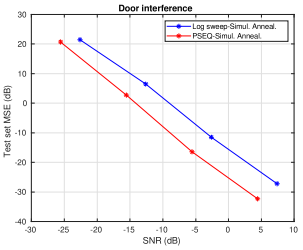
show the comparisons in the average MSE in dB computed over the 1000 test set responses with noise using each of the optimized sweeps from Algorithm 1. Panels (a), (b), and (c) show the results for vacuum noise, air conditioner noise, and door shut noise, respectively. Traditional log-sweep provides the best overall results for both ML techniques when as a function of SNR.



(a)



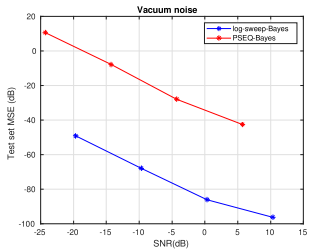
(b)



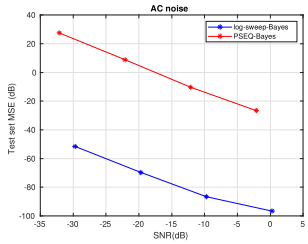
(c)

Fig. 3: $\bar{\psi}_{SD}^{\text{test-set}}(\alpha)$ using SA optimized sweeps in noise (red: PSEQ, blue: log-sweep): (a) vacuum, (b) AC, and (c) door.

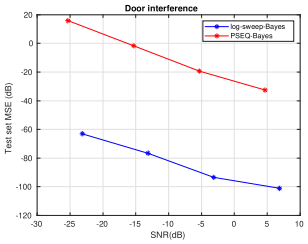
BMN sine sweep (BMN-SS) is used to find the shortest sweep in the presence of background noise [29]. If the BMN sweep was designed for this dataset using the analysis outlined in [29] (see Fig. 6 in [29]), the minimum BMN sweep length computed is $J = 16.24$



(a)



(b)

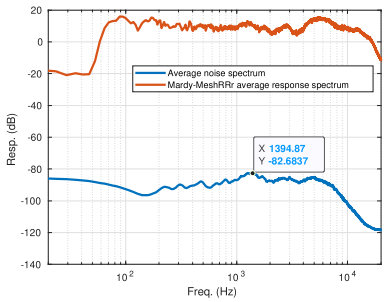


(c)

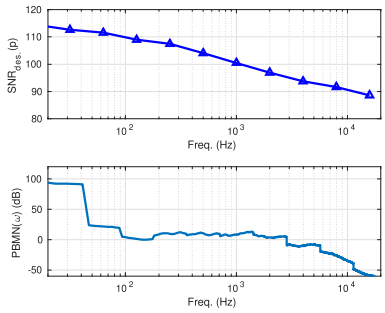
Fig. 4: $\bar{\psi}_{SD}^{\text{test-set}}(\alpha)$ using Bayesian sweeps in noise (red: PSEQ, blue: log-sweep) :(a) Vacuum, (b) AC, and (c) door.

seconds for the dataset[9]. The noise energy was calculated at the end of the impulse responses (≈ 0.85 seconds window at the end of the responses) where the responses have decayed to the noise floor. The resulting plot is shown in Fig. 5(a) of the power spectrum of the average MARDY and MESHRiR responses and noise energies. The desired noise energy was set to -83 dB (marked in Fig. 5(a) at 1395 Hz), where the BMN sweep would need to have been designed to meet the desired bandwise $\text{SNR}_{des}(p)$ for the ambient noise energy. The resulting $\text{SNR}_{des}(p)$ and power spectrum $P_{BMN}(\omega)$ is shown in Fig. 5(b). The presented

[9] In the work [29], with a desired noise floor of -45 dB, the shortest sweep length obtained was $J \approx 5.5$ seconds (cf. Sect. 5.2)



(a)



(b)

Fig. 5: Development of BMN-SS (a) orange: power spectrum of average MARDY and MeshRiR using 72 responses from each database responses, blue: noise spectrum of corresponding responses, (b) Top: desired SNR (octave-bands), bottom: $P_{BMN}(\omega)$.

ML-optimized sweeps give sweep durations lower than those estimated using the BMN-SS technique. The reason why traditional sweep performs better is that the circular shift property energizes all loudspeakers at any given time, yielding a larger SNR. In contrast, only some sweeps energize the room at any time for the PSEQ sweep due to the offset property (7) to retain orthonormality between the multiple PSEQ sweeps.

5 Summary

In this paper, both hyperparameter optimization techniques generate reproducible results for the traditional log sweep [35] (in terms of duration and $\bar{\psi}_{SD}$) but lead to different durations for the PSEQ sweep [28]. The optimized traditional sweep gives the shortest duration (i.e., 5.24 seconds using the Bayesian technique) with a negligible reconstruction error (i.e., -100 dB) without noise for simultaneous deconvolution (see Table 1). In the presence of noise, the optimized log sweep consistently outperforms the optimized PSEQ sweep over various SNRs and different noises (see Figs. 4-5). Finally, both optimized sweeps are shorter in duration than the BMN sweep for the given dataset. Thus, depending on the measurement time requirement and background noise condition, on the basis of this research, there is a clear consensus on which sweep would be best suited for simultaneous deconvolution.

6 Acknowledgements

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

References

[1] S. Bharitkar, "Deconvolution of room impulse responses from simultaneous excitation of loudspeakers," *Proc. 151st Audio Eng. Soc. Conv.,*, Online, Oct. 2021.

[2] R. Banka and S. Bharitkar, "Validation results of deconvolution of room impulse responses from simultaneous excitation of loudspeakers," *Proc. 153rd Audio Eng. Soc. Conv.,*, NY (USA), Oct. 2022.

[3] S. Bharitkar, "Bayesian Optimization for Simultaneous Deconvolution of Room Impulse Responses," *Proc. 36th IEEE Wkshp. Sig. Proc. Syst.*, Rennes (France), Nov. 2022.

[4] ITU-R BS. 2051-1, *Advanced sound system for programme production*, Int. Telecom. Union (ITU), 2018.

[5] S. Bharitkar and C. Kyriakakis, *Immersive Audio Signal Processing*, Springer-Verlag, June 2006.

[6] A. Carini, S. Cecchi, F. Piazza, I. Omiciuolo, G. Sicuranza, "Multiple position room response equalization in frequency domain," *IEEE Trans. Audio, Speech & Lang. Proc.*, 20(1), Jan. 2012, pp. 122–135.

[7] R. Mazur, F. Katzberg, A. Mertins, "Robust room equalization using sparse sound-field reconstruction," *2019 IEEE Int. Conf. Acoust., Speech & Sig. Proc.* (ICASSP 2019), Brighton (UK), April, 2019.

[8] M. Kolundžija, C. Faller, M. Vetterli,"Multichannel low-frequency room equalization using perceptually motivated constrained optimization," *2012 IEEE Int. Conf. Acoust., Speech & Sig. Proc.* (ICASSP 2012), Kyoto (Japan), 2012.

[9] M. Schneider and W. Kellermann, "Adaptive listening room equalization using a scalable filtering structure in the wave domain," *2012 IEEE Int. Conf. Acoust., Speech & Sig. Proc.* (ICASSP 2012), Kyoto (Japan), 2012.

[10] D. Menzies, P. Coleman, F. Fazi, "A room compensation method by modification of reverberant audio objects," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, Nov. 2020, pp. 239–252.

[11] J. Jungmann, R. Mazur, A. Mertins, "Joint time- and frequency-domain reshaping of room impulse responses," *2015 IEEE Int. Conf. Acoust., Speech & Sig. Proc.* (ICASSP 2015), Brisbane (Australia), 2015.

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

[13] L. Fielder, "Practical Limits for Room Equalization," *Proc. 111th AES Conv.,*, NeY (USA), 2001.

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

[15] C. Faller, "Modifying Audio Signals for Reproduction with Reduced Room Effect," *Proc. 147th AES Conv.*, NY (USA), Oct. 2019.

[16] D. Yellin and B. Friedlander, "Multichannel system identification and deconvolution: performance bounds," *IEEE Trans. Sig. Proc.* 47(5), 1999, pp. 1410–1414.

[17] T. Dobrowiecki, J. Schoukens, and P. Guillaume, "Optimized Excitation Signals for MIMO Frequency Response Function Measurements," *Proc. IEEE Instr. Meas. Tech. Conf.*,2005, pp. 1872-1877.

[18] F. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.*, 36, 1988, pp. 122–142.

[19] ST 2095-1:2015, SMPTE Standard - Calibration Reference Wideband Digital Pink Noise Signal, Nov. 30, 2015.

[20] Y. Huang, J. Benesty. and J. Chen, "Identification of acoustic MIMO systems: Challenges and opportunities," *Sig. Proc.*, 86, 2006, pp. 1278–1295.

[21] G-B. Stan, J-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.*, 50(4), April 2002, pp. 249–262.

[22] K. Prawda, S. J. Schlecht, and V. Valimaki, "Robust selection of clean swept-sine measurements in non-stationary noise," *J. Acoust. Soc. Amer.*, 151(3), March 2002, pp. 2117–2126.

[23] P. Majdak, P. Balazs, and B. Laback, "Multiple Exponential Sweep Method for Fast Measurement of Head-related Transfer Functions," *J. Audio Eng. Soc.*, 55(7/8), July/Aug. 2007, pp. 623–637.

[24] S. Weinzierl, A. Giese, and A. Lindau, "Generalized multiple sweep measurement," *Proc. 126th AES Conv.*, Munich (Germany), May 2009.

[25] S. Bharitkar, E. Souza-Blanes, and P. Brunet "Excitation Stimuli For Simultaneous Deconvolution of Room Responses," *Proc. 155th AES Conv.*, NY (USA), Oct. 2023.

[26] C. Antweiler, A. Telle, and P. Vary, "NLMS-type system identification of MISO systems with shifted perfect sequences," *Proc. IWAENC*, Seattle (USA), 2008.

[27] C. Antweiler, S. Kuhl, B. Sauert, and P. Vary, "System identification with perfect sequence excitation-efficient NLMS vs. inverse cyclic convolution," *Proc. 11th ITG Conf. Speech Comm.*, Erlangen (Germany), 2014.

[28] C. Antweiler, A. Telle, P. Vary, and G. Enzner, "Perfect-sweep NLMS for time-variant acoustic system identification," *Proc. Int. Conf. Acoust.*

Speech Sig. Proc. (ICASSP), Kyoto (Japan), 2012.

[29] Y. Nakahara, Y. Iiyama, Y. Ikeda, and Y. Kaneda, "Shortest impulse response measurement signal that realizes constant normalized noise power in all frequency bands," *J. Audio Eng. Soc.*, 70(1/2), Jan. 2022, pp. 24–35.

[30] I. F'ellejero, M. Zivanovic, I. Pmoabarren, A. Carlosena, "Application of multitone signals in room acoustics measurements," *Proc. IEEE Instr. & Meas. Tech. Conf.*, Budapest (Hungary), May 2001.

[31] J. Schoukens, R. Pintelon, "Identification of linear systems," *Pergamon Press*, 1991.

[32] E. Czerwinski, A. Voishvillo, S. Alexandrov, and A. Terekhov, "Multitone testing of sound system components–some results and conclusions, Pat 1: History and theory," *J. Audio Eng. Soc.*, 49(11), Nov. 2001, pp. 1011–1048.

[33] A. Potchinkov, "Low-crest-factor multitone test signals for audio testing," *J. Audio Eng. Soc.*, 50(9), Sept. 2002, pp. 681–694.

[34] J. Vanderkooy, "Aspects of MLS Measuring Systems," *J. Audio Eng. Soc.*, 42(4), April 1994, pp. 503–516.

[35] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," *Proc. 108th Audio Eng. Soc. (AES) Conv.*, Paris (France), Feb. 2000.

[36] ITU-R BS. 2051-3, *Advanced sound system for programme production* , Int. Telecom. Union (ITU), 2022.

[37] J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," *Proc. IWAENC*, Paris (France), Sept. 2006.

[38] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnstrom, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," *Proc. IEEE-WASPAA*, Mohonk, NY (USA), 2021.

[39] L. van der Maaten, and G. Hinton, "Visualizing Data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, 2008, pp. 2579–2605.

[40] J. Snoek, H. Larochelle, and R. Adams, "Practical Bayesian optimization of machine learning algorithms," *Proc. Neural Inf. Proc. Syst. (NIPS)*, 2012.

[41] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv:1807.02811v1*[stat.ML], July, 2018.

[42] J. R. Doppa, V. Aglietti, and J. Gardner, "Advances in Bayesian Optimization," *Neural Inf. Proc. Syst. (NIPS) Wkshp.*, Dec. 2022.

[43] ITU-R BS. 1770-4, *Algorithms to measure audio programme loudness and true-peak audio level* , Int. Telecom. Union (ITU), 2015.