Response to Reviewer Comments

The authors thank the reviewers for carefully reviewing the paper. We have addressed all comments (as noted below) and added references. The paper change is reasonably minor (+references, figure-scaling and clarity). The updated paper draft and in-ear demo sound files (HRTF-downmixed) are at:https://github.com/SunilGBharitkar/ICASSP2023. The link to the audio-files is provided in the paper as well. **We hope to receive an upgraded Overall Evaluation**.

Reviewer #2:
(i) Novelty: The paper extends [1, 3] by operating with impulse responses (Eq. 6) rather than the 1/3 octave magnitude responses. Recasting to time-domain enables design of optimized stimuli to simultaneously deconvolve impulse responses for (i) time-alignment between the 11-loudspeakers and (ii) de-reverberation applications, beyond magnitude-response equalization. This point was missing earlier and now we have updated the Abstract/Introduction.
(ii) Experimental Validation: The listening test stimuli were the optimized sweep, optimized MLS and optimized multitone-pink whose duration and circular shift values correspond to the values in Table I. The duration of the stimuli for 1st session (for timbre preference) was set at 5 seconds and subjects were allowed to freely switch between the three optimized stimuli during the trial. The second session did not allow the subjects to switch, forcing them to listen to the different duration stimuli to judge which stimuli they would prefer for their in-room setup. Comments were elicited for the second session. From the 11 listeners who selected logsweep, 9 preferred the timbre of multitone-pink but favored a shorter measurement time of log-sweep. Sec. IV clarifies this and includes testing interface (Fig. 4(d)).
(iii) Clarity: Improved caption for Fig. 1, parameters (GPA, ER etc) clarified, fonts and figure positioning adjusted. Fig. 4 removed to make space for additional background and clarifying content (per recommendation of Reviewer 3).
(iv) References: Added suggested and new references, including G-B. Stan *et al*. [21] (bottom of page)
(v) Detailed Assessment: Incorporated *(i)*, above, into paper for novelty. Improved Abst. and Intro (including fragmenting paragraphs). The constant T_s=1/48000 (48kHz is sampling frequency), P is duration in samples and $P=T_{stimuli}/T_s$. Table I notation is improved with optimal duration, $T_{stimuli}$, denoted in seconds. Table I shows the optimal duration and right-circular shift values M_i (relative to first channel stimuli $\mathbf{x}_1(n)$, Eq. (4)) for the remaining 10-channels. These are obtained after running Alg. 1 which employs a global hyper-parameter search technique (Bayesian optimization) over the training set of impulse responses, within the box constraints (Sec. IV) for duration and circular shift. Clarifying explanation provided in Sec. III and IV. With the explicit parameter definitions (GPA, Exploration ratio-ER, etc.), Alg. 1 can be reproduced using *bayesopt(...)* (e.g., Matlab).

Reviewer #3: *(i) Detailed Assessment*: The paper extends [1, 3] by operating on impulse responses (Eq. 6) rather than the 1/3 octave magnitude responses. Recasting to time-domain enables design of optimized stimuli to simultaneously deconvolve time-domain responses for (i) time-alignment between the 11-loudspeakers and (ii) de-reverberation. This point was missing earlier and now we have updated the Abstract/Introduction . The choice for multitone-pink stimuli is that pink noise is a standard for measuring cinemas ([19], below). Advantages of log-sweep over other stimuli is in Stan *et al.* ([21], below). We added MLS given its well-understood mathematical properties. We updated Intro section to include Huang *et al.* and Antweiler *et al.* to highlight the advantages of PSEQ for identification/tracking of MISO systems, but also indicate that the stimuli in this paper is not constrained to PSEQ. The influence of noise using the stimuli optimized for 1/3 octave magnitude response [1, 3] was already shown in [2] (Table 1 and Fig. 14) for real-world conditions. Bayesian optimization is a global hyper-parameter optimization technique (within the box constraints setup before simulation of Alg. 1 for hyper-parameters). The theory and science are documented in the context of machine-learning and improvements to the core technique [5] are ongoing for high-dimensional hyper-parameter search. Additional info/references are included in Sec. II.F. The *t-SNE* reference and a brief is provided Sec. II.E. Figure 4 has been removed per recommendation to allow space for revisions.

Reviewer #5: *Detailed Assessment:* The references are taken out of the Abstract, which is now improved.

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

[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.