

ON RESTORING PREMATURELY TRUNCATED SINE SWEEP ROOM IMPULSE RESPONSE MEASUREMENTS

Elliot Kermit Canfield-Dafilou and Jonathan S. Abel

Center for Computer Research in Music and Acoustics,
Stanford University, Stanford, CA 94305 USA
[kermit|abel]@ccrma.stanford.edu

ABSTRACT

When measuring room impulse responses using swept sinusoids, it often occurs that the sine sweep room response recording is terminated soon after either the sine sweep ends or the long-lasting low-frequency modes fully decay. In the presence of typical acoustic background noise levels, perceivable artifacts can emerge from the process of converting such a prematurely truncated sweep response into an impulse response. In particular, a low-pass noise process with a time-varying cutoff frequency will appear in the measured room impulse response, a result of the frequency-dependent time shift applied to the sweep response to form the impulse response.

Here, we detail the artifact, describe methods for restoring the impulse response measurement, and present a case study using measurements from the Berkeley Art Museum shortly before its demolition. We show that while the difficulty may be avoided using circular convolution, nonlinearities typical of loudspeakers will corrupt the room impulse response. This problem can be alleviated by stitching synthesized noise onto the end of the sweep response before converting it into an impulse response. Two noise synthesis methods are described: the first uses a filter bank to estimate the frequency-dependent measurement noise power and then filter synthesized white Gaussian noise. The second uses a linear-phase filter formed by smoothing the recorded noise across perceptual bands to filter Gaussian noise. In both cases, we demonstrate that by time-extending the recording with noise similar to the recorded background noise that we can push the problem out in time such that it no longer interferes with the measured room impulse response.

1. INTRODUCTION

In order to study room acoustics, one must measure accurate impulse responses of the space. These measurements are often challenging and time consuming to acquire. Large spaces frequently exhibit poor background noise levels, so acousticians often employ methods to improve the signal-to-noise ratio (SNR). In particular, linear and logarithmic sine sweep measurements have been shown to be highly effective [1]. Researchers have identified some of the problems and their solutions for using sine sweep measurements to study room reverberation. Specifically, [2] and [3] address issues of time-smearing, clicks/plosives, and pre/post equalization. Further, [4] discusses issues related to clock drift.

This paper addresses another problem that is often encountered when measuring impulse responses with sine sweeps in noisy environments that has not been previously discussed—what happens when the sweep response recording is stopped too early. This is common, as it is challenging to maintain quiet after the room re-

sponse has appeared to decay into the noise floor. Plosives and impulsive noises will be converted into descending sweeps and any such noises in the silence following the recorded sweep will be detrimental to the conversion process. As it turns out, one needs to record a duration of background noise equivalent to the length of the sweep following the response’s decay into the noise floor to ensure that the problem will not be encountered.

If one does not record enough silence (background noise) after the sine sweep response, the resulting impulse response will contain a time-varying low-pass filter characteristic imprinted on its noise floor. This paper addresses the cause of these artifacts and two methods for alleviating the problem in post-production. Related to this problem, [5] and [6] discuss methods for extending impulse responses through the noise floor, however there are implications for how the noise floor is measured based on these new results.

The rest of the paper is organized as follows. In §2 we review the process of converting a sine sweep measurement into an impulse response and introduce the problem associated with stopping the recording prematurely. In §3 we introduce two methods for pre-processing the response sweep by extending the recording with synthesized noise that matches the background noise present in the recording. Next §4 presents examples that demonstrate how our method for preprocessing the sweep response is desirable compared to zero padding or circular convolution. Finally, §5 presents concluding remarks.

2. CONVERTING SINE SWEEPS TO IMPULSE RESPONSES

Linear and logarithmic sine sweeps can be used to measure room impulse responses and work under the principle that the sweep can be viewed as an extremely energetic impulse smeared out in time. A linear sweep, in which the sinusoid frequency increases linearly over time, can be defined as

$$x(t) = \sin \left(\omega_0 t + \frac{\omega_1 - \omega_0}{T} \frac{t^2}{2} \right), \quad t \in [0, T], \quad (1)$$

where ω_0 is the initial frequency in radians, ω_1 the final frequency, and T the total duration. A logarithmic sweep, in which the sinusoid frequency increases exponentially over time, can be defined using the same variables as

$$x(t) = \sin \left(\frac{\omega_0 T}{\ln \left(\frac{\omega_1}{\omega_0} \right)} \left[\exp \left(\frac{\ln \left(\frac{\omega_1}{\omega_0} \right) t}{T} \right) - 1 \right] \right). \quad (2)$$

To convert sine sweep responses to impulse responses, one aperiodically convolves an equalized, time-reversed version of the

original sine sweep, \tilde{x} with the recorded response such that

$$h(t) = y(t) * \tilde{x}(t). \quad (3)$$

This works because the group delay is canceled by the time reversal, with the equalization compensating for the relative time spent in each frequency band,

$$\delta(t) = x(t) * \tilde{x}(t). \quad (4)$$

This deconvolution processing aligns the response to the original sine sweep in time effectively forming the impulse response. As a linear sweep traverses any given bandwidth in the same length of time, irrespective of the starting frequency, the equalization is a constant, independent of frequency. For a logarithmic sweep, which spends the same length of time traversing any given octave, an exponential equalization is applied to compensate for the disproportionate amount of low-frequency energy applied to the system. Naturally, the calculation is efficiently computed in the frequency domain¹ as

$$h(t) = \mathcal{F}^{-1} \left(\frac{Y(\omega)}{X(\omega)} \right). \quad (5)$$

One of the benefits to using sine sweep measurements over other integrated-impulse response measurement techniques, such as maximum length sequences, is that many nonlinearities produce only harmonics of an input sinusoid, and the deconvolution process will place the onset of unwanted harmonic component responses before the onset of the linear component of the system response. Using logarithmic sweeps, the time shift of the harmonic distortion components of the response is controlled via the length of the sweep, and the linear response is easily separated from the harmonic distortion response. While [2, 7, 8] and others have shown that useful information can be extracted from the higher order components, that is not the concern of this paper. For acquiring a linear room impulse response, one can simply window out the distortion artifacts that precede the linear response.

The fundamental problem this paper explores is caused by the desire to use acyclic convolution rather than circular convolution to convert the sine sweep response into separate linear "impulse response" and non-linear system responses. The difficulty results from the presence of additive measurement noise, Eq. (3) is actually

$$h(t) = [y(t) + n(t)] * \tilde{x}(t). \quad (6)$$

where $n(t)$ is the measurement noise (often mainly acoustic background) and is assumed to be stationary. In addition to time-aligning the frequencies of the sweep into an impulse response, the background noise is also shifted by the same transformation. Under acyclic convolution, the tail end of the converted impulse response will exhibit a frequency-dependent filter cutoff with the same trajectory as the frequency trajectory of $x(t)$. When the recording of background noise following the sweep response is sufficiently long, the filter artifact will occur after the impulse response has decayed, and can be windowed out. However, when the recording is shut off too early, linear convolution causes this frequency-dependent filter effect to overlap with the impulse response. While a recording with a high SNR may render this effect

¹In this paper, we use t as the argument to functions in the time domain and ω as the argument to functions in the frequency domain.

inaudible, in spaces with poor SNR, this unwanted effect becomes quite clear.

The naïve solution to this problem would be to use circular convolution rather than linear convolution. However, this is not ideal. Circular convolution will reconstruct the noise floor and solve the issue of the filtering effect, shifting the noise corresponding to times before the sweep to the end of the response, but any nonlinearities in the measurement will also be shifted to occur during the desired impulse response. Weak nonlinearities are all but guaranteed when measuring an acoustic space, as loudspeakers are inherently nonlinear at levels useful in impulse response measurement. The effect of circular convolution will corrupt the measurement.

Fig. 1 shows a contrived example that highlights the difference between linear and circular convolution using a linear sweep and noise signal. Under circular convolution, the noise statistics ought to stay constant (i.e., $n(t) \sim n(t) \circledast \tilde{x}(t)$). Because of zero padding, linear convolution causes the noise occupy a longer period of time. Over the course of the beginning of the processed block, the noise starts from the high frequencies, and lower frequencies enter according to the slope of the frequency trajectory of the sweep. At the end of the file, the opposite is true: the high frequencies stop before the lower frequencies. If this noise were the background noise in a space, the effect of the frequencies stopping at different times would be heard as a filter with a time-varying cutoff frequency.

Fig. 2 demonstrates the difference between cyclic and acyclic convolution when there are nonlinearities present. In the acyclic convolution, the nonlinear components precede the desired linear response while in the cyclical convolution they corrupt the linear response.

Ideally, a sufficiently long segment of background noise following the sine sweep is recorded so as to avoid additional processing. In real spaces, low frequencies typically take longer to decay than high frequencies. Ascending sweeps help hide issues caused by the premature ending of a recording, as the low frequency components are provided more time to decay while higher frequencies are being excited in the space. In order to guarantee that the noise roll off problem will not be encountered during the impulse response, an additional amount of background noise of length T must be captured following the decay of the room response to $x(t)$. The longer the sine sweep, the more patient one must be when recording the impulse response. (In addition, transients occurring during the time after the system response has decayed may corrupt the measured impulse response.)

3. NOISE EXTENSION TECHNIQUES

When the recording is cut off too early, our solution is to push the filter cutoff in time so that it no longer interacts with the room response. To do this, we pre-process the sweep response $y(t)$ by extending it with noise that is perceptually similar to the background noise in the physical space. We propose two methods detailed below for synthesizing this noise. In both cases, we analyze a portion of recorded room noise with no other signals present, synthesize additional noise, and splice it onto the end of the response sweep before converting it into an impulse response.

For both methods, we use a 500 ms analysis window and a 50 ms cross-fade. If the room response has decayed sufficiently, it is best to perform the analysis on samples coming from the end of the response audio file as this is where the new samples will

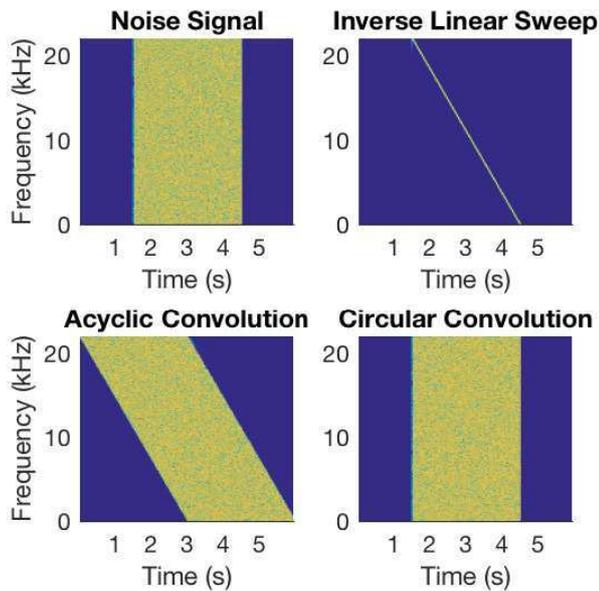


Figure 1: Spectrograms of noise, inverse (linear) sweep, acyclic convolution, and circular convolution.

be spliced. Our goal is to capture both the overall characteristic of the background noise as well as any local features necessary to make an imperceptible transition. If there is a noticeable mismatch between the recorded noise and the synthesized noise, descending-chirps will be introduced into the impulse response by the deconvolution process. If there is not enough isolated noise at the end of the file, it is possible to use a portion of background noise from somewhere else (i.e., preceding the sweep or another recording taken at the same time).

3.1. Band-pass Filtered Noise

In the first approach, we aim to match the spectrum of the background noise by analyzing the amplitude of the recorded noise in a set of frequency bands, and apply these amplitudes to synthesized white Gaussian noise to correctly color it.

We define the analysis noise $n(t)$ as 500 ms of recorded noise from the recording we intend to match. Additionally, we generate an i.i.d. Gaussian noise sequence, denoted $\gamma(t)$. In this method, we form a synthesis signal $s(t)$ such that

$$|S(\omega)| \sim |N(\omega)|. \quad (7)$$

We use a perfect reconstruction, zero-phase filter bank to split both $n(t)$ and 500 ms of $\gamma(t)$ into K composite frequency bands

$$n(t) = \sum_k n_k(t), \quad k \in [1, 2, \dots, K], \quad (8)$$

and

$$\gamma(t) = \sum_k \gamma_k(t). \quad (9)$$

Our filter bank consists of a cascade of squared 3rd-order Butterworth filters with center frequencies spaced 1/4 octave apart. The

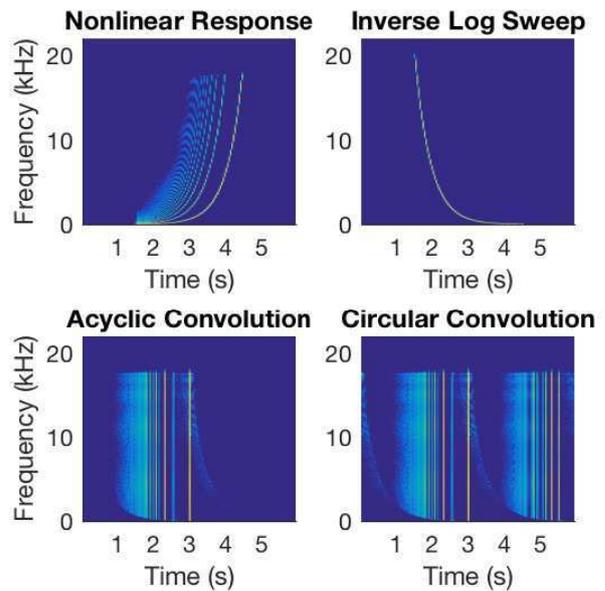


Figure 2: Non-linear sweep response, inverse (logarithmic) sweep, acyclic convolution, and circular convolution.

signals are filtered both in the forwards and backwards directions so that the phase is unaltered. The perfect reconstruction aspect of this filter bank is important because we use it both for analysis and synthesis.

Once separated into bands, we estimate the gain coefficient in each frequency band of both $n(t)$ and $\gamma(t)$ by computing the RMS level on the steady-state portion of the filtered signals. To compute the synthesis noise $s(t)$, we color a sufficiently long² amount of $\gamma(t)$ by scaling each frequency band by the ratio of measured analysis to synthesis gains and sum the result,

$$s(t) = \sum_k \left(\frac{\text{RMS}[n_k(t)]}{\text{RMS}[\gamma_k(t)]} \gamma_k(t) \right). \quad (10)$$

At this point, the steady state portion of $s(t)$ is a block of noise with the same magnitude frequency band response as the analysis signal $n(t)$. An example impulse response and magnitude spectrum resulting from this filter bank can be seen in Fig. 3. We found that 1/4 octave bands were sufficient to match the synthesis and analysis noises' magnitude spectrum. We then use a 50 ms long equal-power cross-fade between the end of $y(t)$ and the beginning of the steady state portion of $s(t)$. After this, it is safe to convert the sweep response into an impulse response as done in Eq. (5).

3.2. ERB-Smoothed Noise

Our second method synthesizes noise that is perceptually similar to the analysis noise via a filter design technique. We define $\gamma(t)$ and $n(t)$ in the same way as above, in § 3.1. We window both noise signals with a Hann window and take their Fourier transforms. In the frequency domain, we smooth both signals on a critical band

²Sufficient here depends on how prematurely the recording was halted.

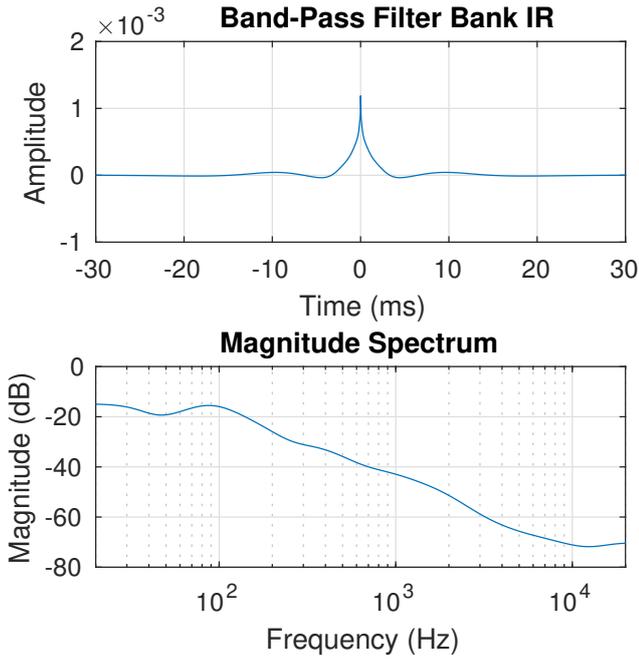


Figure 3: Band-pass filterbank impulse response and magnitude spectrum.

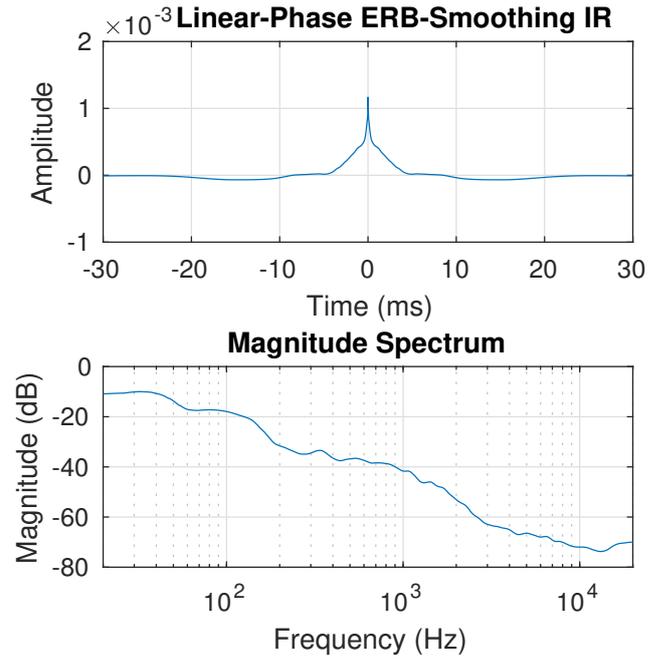


Figure 4: ERB-smoothed noise filter impulse response and magnitude spectrum.

basis by averaging the power within the DFT bins of each critical band such that

$$\hat{N}(\omega) = \sum_{\zeta=f(b(\omega-1/2))}^{f(b(\omega+1/2))} |N(\zeta)|^2 \quad (11)$$

and

$$\hat{\Gamma}(\omega) = \sum_{\zeta=f(b(\omega-1/2))}^{f(b(\omega+1/2))} |\Gamma(\zeta)|^2, \quad (12)$$

where $b(\cdot)$ defines a critical bandwidth. This results in $\hat{N}(\omega)$ and $\hat{\Gamma}(\omega)$, the critical band smoothed versions of $N(\omega)$ and $\Gamma(\omega)$. This processing reduces the complexity and detail of the noise signals' spectra but should not be perceptually audible. We then impart the spectrum of the smoothed analysis noise upon the smoothed synthesis noise in the frequency domain to obtain the transfer function

$$G(\omega) = \frac{\hat{N}(\omega)}{\hat{\Gamma}(\omega)}. \quad (13)$$

We then find a linear-phase version of this transfer function

$$G_{\text{lin}}(\omega) = |G(\omega)| e^{-j\tau\omega}. \quad (14)$$

Returning $G_{\text{lin}}(\omega)$ to the time domain as seen in Fig. 4, we filter $\gamma(t)$ with $g_{\text{lin}}(t)$ such that

$$s(t) = \gamma(t) * g_{\text{lin}}(t). \quad (15)$$

Last, we splice the synthesized noise onto $y(t)$ with a 50 ms equal power cross-fade as described above.

4. EVALUATION

To preserve the acoustics of the Berkeley Art Museum (BAM) at 2626 Bancroft Way in Berkeley, CA, acoustic measurements were taken by Jacqueline Gordon and Zackery Belanger and their team shortly before it was demolished in 2015 [9]. This space, designed by Mario Ciampi, has a very long reverberation time due to its large, concrete structure. Since this space exhibited a high noise floor, a long sine sweep was employed in order to improve the SNR. Circumstances led to the recordings being prematurely ended, and to the discovery of the problem described in this paper. Fig. 5 shows an example recorded (short) response sweep as well as versions extended with noise synthesized with our two methods. Fig. 6 shows the impulse responses computed from these sweeps. Both visually and aurally the noise extended approaches achieve better results than the unprocessed version.

Both approaches for synthesizing noise produce reasonable results, and the power spectrum for the analysis noise and both varieties of synthesis noise can be seen in Fig. 7. On average, the synthesis noise tracks the nature of the room sound well, and both synthesis methods produce perceptually similar results. As it turns out, it is equally important to have a smooth cross-fade between the recorded and synthesized noise as sudden changes will create perceptual artifacts.

5. CONCLUSIONS

In this paper, we discuss how converting a sine sweep response to an impulse response requires a sufficient amount of recording time beyond what is necessary to capture the room response after it has perceptually decayed. While it is naturally best to record this in the physical space, it is not always possible like at the Berkeley Museum of Art. We propose two methods for extending the record-

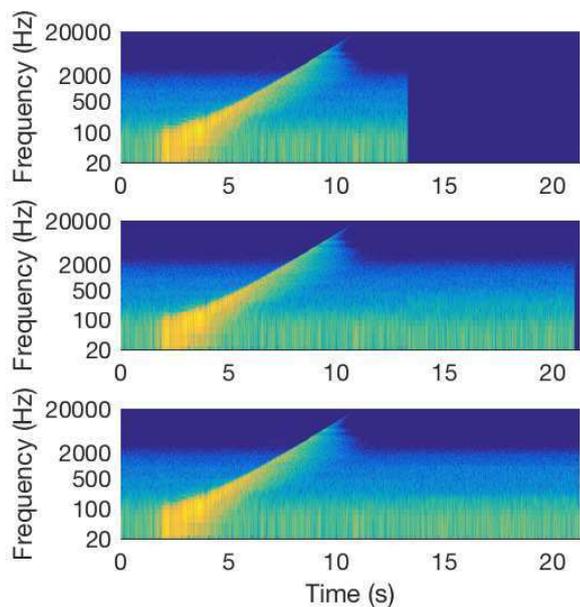


Figure 5: Spectrograms for signal sweep and sweep responses with no treatment, band-passed noise, and ERB smoothed noise.

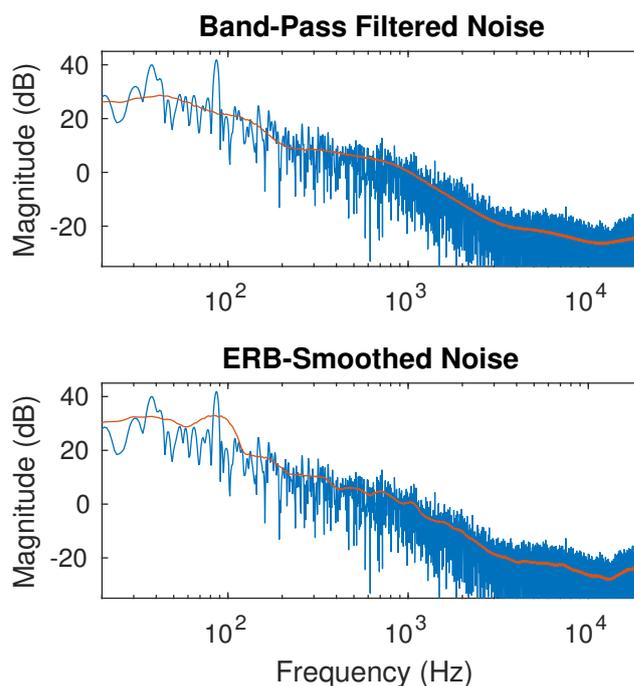


Figure 7: Spectrum of synthesized noise (red) averaged over 500 simulations compared to analyzed noise (blue) for the band-pass filter method (top) and ERB-smoothed method (bottom).

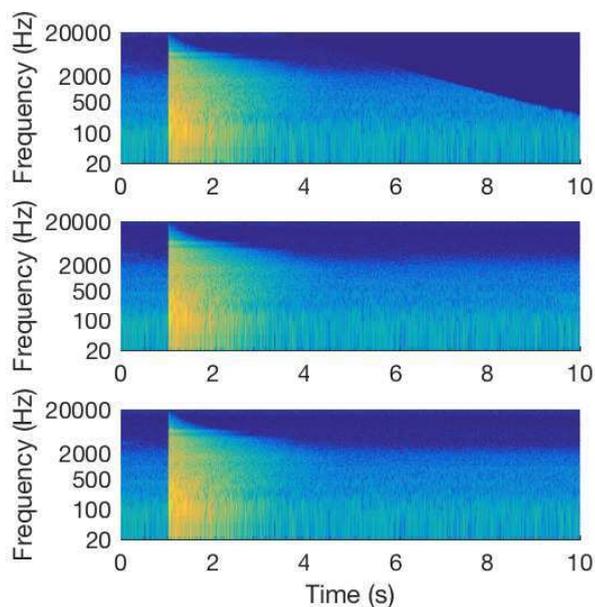


Figure 6: Spectrograms corresponding to the impulse responses calculated from the sweep responses in Fig. 5—no treatment, band-passed noise, and ERB smoothed noise.

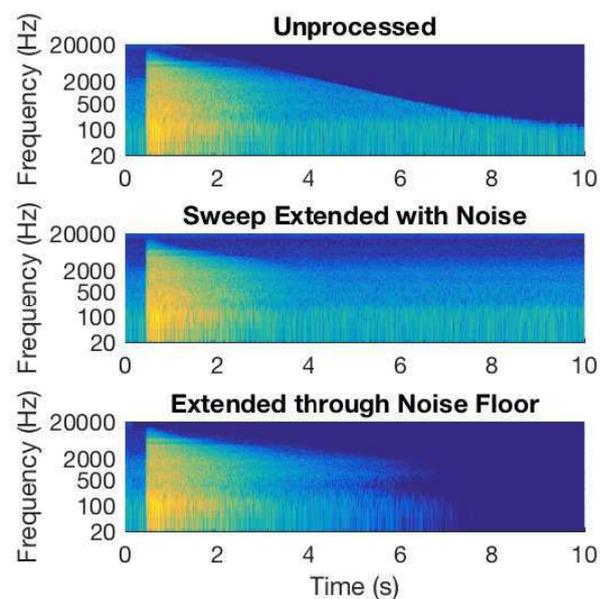


Figure 8: BAM impulse response without processing, extending the sweep with noise, and extending impulse response through the noise floor.

ing after the fact that both depend on analyzing recorded background noise and using this information to color Gaussian noise. One technique measures the frequency-dependent amplitude levels with a bank of band-pass filters while the other involves filtering Gaussian noise with a perceptual-based filter. Both methods work well and eliminate the undesired artifact from the resulting impulse response.

Naturally, an impulse response with a large noise floor is not ideal. In such cases, the techniques described above can be used to prepare an impulse response measurement for further processing, such as described in [6], to extend the measurement through the noise floor. The result of such processing applied to another of the BAM measurements appears in Fig. 8. Note that the extended impulse response shows none of the time-varying low-pass filtering artifacts present in the original measured impulse response.

6. ACKNOWLEDGMENTS

The authors would like to thank Jacqueline Gordon for bringing the BAM acoustics measurement project [9] to our attention, and for setting us onto this work.

7. REFERENCES

- [1] Angelo Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Audio Engineering Society Convention 108*, Feb 2000.
- [2] Swen Müller and Paulo Massarani, “Transfer-function measurement with sweeps,” *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–71, 2001.
- [3] Angelo Farina, “Advancements in impulse response measurements by sine sweeps,” in *Audio Engineering Society Convention 122*, May 2007.
- [4] Jonathan S. Abel, Nicholas J. Bryan, and Miriam A. Kolar, “Impulse Response Measurements in the Presence of Clock Drift,” in *Audio Engineering Society Convention 129*, Nov 2010.
- [5] Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel, “Analysis and synthesis of room reverberation based on a statistical time-frequency model,” in *Audio Engineering Society Convention 103*, Sep 1997.
- [6] Nicholas J. Bryan and Jonathan S. Abel, “Methods for extending room impulse responses beyond their noise floor,” in *Audio Engineering Society Convention 129*, Nov 2010.
- [7] Antonin Novak, Laurent Simon, Frantisek Kadlec, and Pierrick Lotton, “Nonlinear system identification using exponential swept-sine signal,” *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 8, pp. 2220–9, 2010.
- [8] Jonathan S. Abel and David P. Berners, “A technique for nonlinear system measurement,” in *Audio Engineering Society Convention 121*, Oct 2006.
- [9] Jacqueline Gordon and Zackery Belanger, “Acoustic deconstruction of 2626 Bancroft Way, kickstarter project 632980650,” 2015.