Audio Engineering Society

Convention Paper 9183

Presented at the 137th Convention
2014 October 9–12 Los Angeles, USA

A New Approach to Impulse Response Measurements at High Sampling Rates

Joseph G. Tylka[1], Rahulram Sridhar[1], Braxton B. Boren[1], and Edgar Y. Choueiri[1]

[1] *3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ, 08544, USA*

Correspondence should be addressed to Joseph G. Tylka (josephgt@princeton.edu)

ABSTRACT

High sampling rates are required to fully characterize some acoustical systems, but capturing the system's high-frequency roll-off decreases the signal-to-noise ratio (SNR). Band-pass filtering can improve the SNR, but may create an undesirable pre-response. An iterative procedure is developed to measure impulse responses (IRs) with an improved SNR and a constrained pre-response. First, a quick measurement provides information about the system and ambient noise. A second, longer measurement is then performed, and a suitable band-pass filter is applied to the recorded signal. Experimental results show that the proposed procedure achieves an SNR of 37 dB with a peak pre-response amplitude of <0.2% of the IR peak, whereas a conventional technique achieves an SNR of 32 dB with a peak pre-response amplitude of 16%.

1. INTRODUCTION

Many tasks in both audio research and industry depend on accurate measurements of the impulse responses (IRs) of acoustical systems. For example, measuring the directivity of loudspeakers, reverberation times of concert halls, and head-related transfer functions for use in virtual audio may be accomplished through IR measurements. Due to the Nyquist-Shannon sampling theorem, acoustical systems whose frequency responses extend well beyond 20 kHz require sampling rates greater than 44.1 kHz or 48 kHz to accurately characterize the entire system response. By definition, this is the case for ultrasonic transducers, but, as we will see, even standard audio-band transducers often have non-zero responses beyond 20 kHz.

Of course, characterization of the system response above 20 kHz may seem unnecessary for typical audio applications, given the well-known inability of humans to detect sine tones above about 20 kHz. However, it has been suggested that humans are sensitive to the significant "time-smearing" of the repro-

duced analog waveform caused by the anti-aliasing and reconstruction filters at lower sampling rates [1]. Also, higher sampling rates may be necessary to adequately capture the minimum perceptible interaural time difference (ITD), a vital spatialization cue for accurate binaural audio. For humans, this minimum perceptible ITD is approximately 10 μs [2]. Such a temporal resolution corresponds to a sampling rate of 100 kHz. While we will not attempt to evaluate such claims in this work, the above arguments provide motivation to develop a reliable method of measuring system responses at high sampling rates.

A drawback to adopting high sampling rates is that the frequency response of the system under test will likely drop below the noise floor well below the Nyquist frequency. For example, this may occur when measuring low-sensitivity transducers, working in a noisy environment, or using noisy microphone pre-amplifiers. In such cases, any IR measurements may be contaminated with high-frequency noise that decreases the signal-to-noise ratio (SNR). The IR may be low-pass filtered to improve the SNR, but some filters can be expected to cause phase distortions or, as we will show, create an undesirable "pre-response" (or "pre-ringing") preceding the impulse. Therefore, in order to obtain accurate IR measurements at high sampling rates, a compromise must be made between SNR and filtering effects.

1.1. Previous Work

A well-established method for measuring IRs is to use an exponential sine sweep (ESS) [3, 4]. Advantages of this method include a high SNR and the ability to isolate and align non-linear distortion terms into distinct responses. Modified sweeps have been proposed which achieve improved SNRs based on knowledge of the ambient noise [5]. However, such modified sweep techniques require altering the time-frequency relationship of the sweep, thereby preventing the sweep from isolating distortion terms. It is worth noting that the ESS cannot isolate 100% of each distortion term from the linear response [6]. Consequently, low amplitude signals are recommended to better measure the linear response by itself (nevertheless, when higher amplitudes are required, the ESS still enables much of the distortion to be isolated and later removed).

When using an ESS, Farina recommends extracting the measured IR by convolving the recorded sweep with a time-reversed "inverse sweep" [3]. As we will show in Section 3.3, this process effectively creates a linear-phase band-pass filter (BPF) with cut-off frequencies equal to the initial and final sweep frequencies. Consequently, a high SNR can be achieved simply by restricting the sweep to the pass-band of the system under test, since any out-of-band noise will be attenuated. However, it has been pointed out that a limited-bandwidth sweep may result in a sinc function pre-response [3], so it is typically recommended to sweep up to the Nyquist frequency.

At higher sampling rates (e.g. 96 kHz), however, sweeping up to the Nyquist frequency (48 kHz) may damage the transducers due to excessive heating. This is especially true when using long-duration sweeps, as recommended in the literature [3, 4], or when driving the transducer with high voltages. Thermal models have been proposed to predict transducer heating [7, 8], though, to our knowledge, they have not been employed to suggest safe sweep durations.

1.2. Objectives and Approach

The primary objective of this work is to formulate a method to measure IRs with high SNRs but without a significant pre-response at high sampling rates. It is also our objective to make these measurements efficiently while preserving the advantages of the conventional ESS technique. In order to achieve these goals, we employ multiple sweeps which can be tailored to the system under test and the ambient noise. We also determine and apply a suitable BPF for a given system in order to achieve an improved SNR without an excessive pre-response.

Additionally, to ensure that our method is safe for the transducers, we examine the thermal effects on transducers due to sweep excitation. However, as these results are not explicitly incorporated into our proposed procedure, we discuss this consideration in the appendix.

1.3. Paper Overview

In Section 2, we give a general description of the proposed IR measurement procedure. Then, in Section 3, we describe various aspects regarding the implementation of the ESS. In Section 4, we suggest two definitions of a system's pass-band and show how the corresponding frequency bounds are computed. We then distill the previously described elements into the proposed procedure in Section 5. As

an example, we implement the procedure for a real loudspeaker system in Section 6, and present the results in Section 7. Finally, we conclude with a summary of our findings. In the appendix, we show how a thermal model may be used to derive approximate excitation duration constraints.

2. PROPOSED PROCEDURE OVERVIEW

The proposed impulse response (IR) measurement procedure begins by measuring the system response using a full-range exponential sine sweep (ESS) terminating at the Nyquist frequency. From this measurement, the approximate pass-band of the system is determined. Then, a second, limited-range sweep through only this frequency range is designed and executed. A corresponding band-pass filter (BPF) is applied to the recorded signal to yield an improved signal-to-noise ratio (SNR) compared to the initial measurement and a constrained pre-response. In the following sections, we will describe elements of this proposed procedure separately and then distill them into the complete procedure in Section 5.

3. THE EXPONENTIAL SINE SWEEP (ESS)

In this section, we define two versions of the ESS and discuss various aspects of their implementation.

3.1. Conventional ESS

The conventional ESS discrete-time signal is given by [3]

$$x[k] = \sin\left\{\frac{\omega_1 N}{\ln(\omega_2/\omega_1)} \cdot \left[\left(\frac{\omega_2}{\omega_1}\right)^{\frac{k}{N}} - 1\right]\right\} \qquad (1)$$

for $k \in [0, N-1]$, where $N = F_s T$ is the total number of samples of the signal, T is the sweep duration in seconds, F_s is the sampling rate in samples/second, and ω_1 and ω_2 are the initial and final frequencies of the sweep, respectively, in rad/sample. Note that ω represents the *normalized* frequency, such that the frequency in Hz is given by $\omega F_s/2\pi$.

When implementing this method, the sweep should be preceded by a brief segment of silence to ensure that the system is initially at rest. This segment of the recorded signal can be used as a sample of the ambient noise. The sweep should also be followed by a suitable segment of silence, depending on the recording environment and ultimate application, to adequately capture the desired length of the IR tail [3].

3.2. Phase-Controlled ESS

The so-called "phase-controlled" ESS requires that the phase of the sinusoid both starts and ends at an integer multiple of 2π, yielding an amplitude of zero at the start and end of the sweep [9]. This is accomplished by constraining the final frequency to be an integer number of octaves P above the initial frequency, such that $\omega_2/\omega_1 = 2^P$, and by allowing some flexibility in the sweep duration. The sweep $x_{pc}[n]$ is defined by [9]

$$x_{pc}[k] = \sin\left[\frac{\omega_1 L}{\ln(2^P)} \cdot \left(2^P\right)^{\frac{k}{N}}\right] \qquad (2)$$

for $k \in [0, N-1]$, where L is the so-called "ideal" sweep length in samples and N is the actual sweep length, equal to L rounded to the nearest integer. The ideal sweep length L is found based on an approximate sweep duration T (in seconds) such that

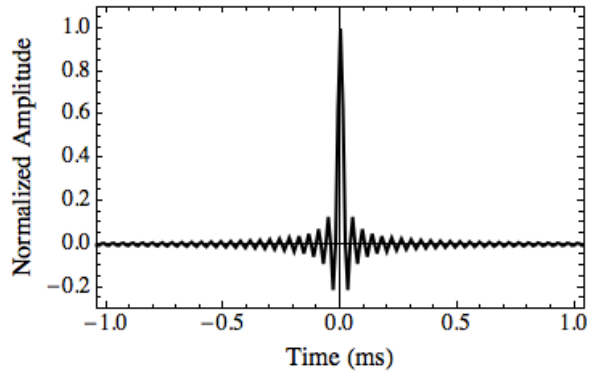
$$\frac{\omega_1 L}{\ln(2^P)} = 2\pi \cdot \text{Round}\left[\frac{\omega_1 T F_s}{2\pi \ln(2^P)}\right].$$

For a phase-controlled sweep that terminates at the Nyquist frequency, we will refer to the sweep by its nominal initial frequency $\omega_1 F_s/2\pi$ and duration T. L is then computed as shown above, and the actual ω_1 is found by rounding the nominal number of octaves.

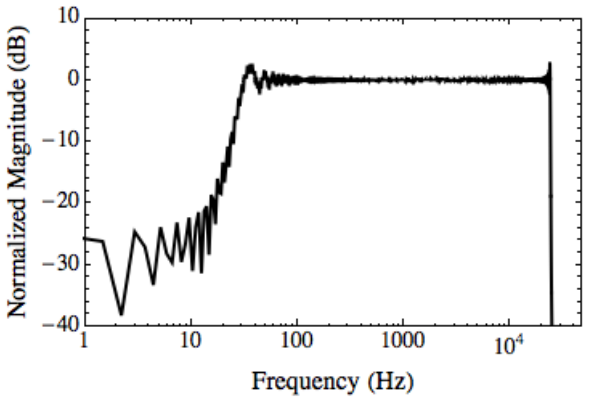
3.3. Inversion and Deconvolution

The IR of an acoustical system is obtained by deconvolving the recorded output signal by the known input signal. Deconvolution is typically performed by dividing the signals' corresponding spectra in the frequency domain via the fast Fourier transform (FFT). This is equivalent to convolving the recorded signal with the input signal's exact inverse, whose frequency spectrum is equal to the reciprocal of the input spectrum. We shall refer to this procedure as *exact deconvolution* of the recorded signal.

An alternative technique to extract the measured IR when using an ESS involves creating an "inverse sweep" by time-reversing the input sweep signal and applying an appropriate frequency-dependent amplitude envelope (+6 dB/octave) to compensate for the "pink" magnitude spectrum of the ESS [3, 9]. This signal is then convolved with the microphone signal to produce the IR. We shall refer to this procedure as *time-reversed deconvolution* of the recorded signal.



(a) Impulse response



(b) Magnitude response

Fig. 1: An example of the resulting band-pass filter due to time-reversed deconvolution for a ∼23 Hz to 24 kHz phase-controlled ESS sampled at 96 kHz.

It can be shown that time-reversed deconvolution is equivalent to performing exact deconvolution and then applying a linear-phase BPF with cut-off frequencies equal to the initial and final sweep frequencies. The resulting BPF can be obtained by convolving a given ESS with its time-reversed inverse sweep. As an example, the impulse and magnitude responses of the resulting BPF for a ∼23 Hz to 24 kHz phase-controlled ESS are shown in Fig. 1.

For both of these deconvolution techniques, linear convolution of the recorded signal with the inverse signal is recommended to prevent part of the IR from "wrapping around" via cyclic convolution [3]. The FFT can still be used, however, provided that each signal is zero-padded to at least twice its length, in which case multiplication in the frequency domain is equivalent to linear convolution in the time domain. In our method, unless otherwise noted, we perform exact deconvolution with appropriate zero-padding to extract the IR.

3.4. Loudspeaker "Pop"

One of the problems with the conventional ESS as defined in Eq. (1) is that the loudspeaker may produce an audible "pop" at the end of the sweep. This occurs when the ESS signal abruptly drops from a non-zero value to zero [3]. The pop is undesirable as it introduces energy across the entire frequency spectrum, appearing as noise in the resulting IR. Two solutions to this problem are to apply a time-domain fade-out to the end of the sweep [3] or to use a phase-controlled ESS up to the Nyquist frequency [9] to ensure that the sample values of the excitation signal converge more gradually towards zero. It is worth emphasizing that the phase-controlled sweep will prevent the pop *only* when terminated at the Nyquist frequency, since it is only under this condition that the samples leading up to the final sample converge towards zero. In our method, we use a phase-controlled ESS when sweeping to the Nyquist frequency, and otherwise apply a time-domain fade-out to a conventional ESS.

3.5. Signal-to-Noise Ratio

It has been stated that the measured SNR can be improved either by increasing the sweep duration or by averaging multiple measurements [3, 4]. The latter technique, however, may lead to errors due to time-variant effects such as heating of the transducers or, in the case of outdoor measurements, wind. Consequently, we choose to employ a small number of longer sweeps.

4. SYSTEM PASS-BAND IDENTIFICATION

In this section, we provide two definitions of a system's pass-band and discuss how the corresponding band-limits are determined under each definition. In Section 4.1, we define the system pass-band as the widest frequency range over which the SNR spectrum is positive. Alternatively, the pass-band can be implicitly defined based on the effect of applying an ideal BPF with the same pass-band to the measured system response. This effect is expressed as a deviation from the measured IR which is often

manifested as a pre-response. In Section 4.2, we constrain the deviation due to filtering and determine the corresponding pass-band to meet this limit.

First, we define some variables. Given an input signal $x[k]$ and a system IR $h[k]$, the system output is given by

$$w[k] = (x * h)[k], \tag{3}$$

where $*$ denotes convolution. The recorded signal $y[k]$, however, includes noise $n[k]$, and is given by

$$y[k] = w[k] + n[k]. \tag{4}$$

The SNR spectrum as a function of frequency is defined as

$$SNR(\omega) \equiv 10 \log_{10} \frac{|W(\omega)|^2}{|N(\omega)|^2}, \tag{5}$$

where $W(\omega)$ is the discrete-time Fourier transform of $w[k]$, and similarly for $N(\omega)$ and $n[k]$. Also, the total SNR is defined as

$$SNR \equiv 10 \log_{10} \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} |W(\omega)|^2 \, d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |N(\omega)|^2 \, d\omega}. \tag{6}$$

The deviation spectrum caused by applying an ideal BPF to the measured system transfer function $H(\omega)$ is given by

$$D(\omega) = \begin{cases} -H(\omega) & \text{for } |\omega| \leq \omega_1 \\ 0 & \text{for } \omega_1 < |\omega| < \omega_2 \\ -H(\omega) & \text{for } \omega_2 \leq |\omega| \leq \pi \end{cases}, \tag{7}$$

where ω_1 and ω_2 are the cut-off frequencies of the BPF. Note that our definition implies that the deviation is *added* to the full system response to yield the filtered response. We define two metrics to quantify the deviation from the full system response caused by filtering. The first is the peak deviation amplitude (PDA), defined as

$$PDA \equiv \max |d[k]| . \tag{8}$$

Note that the absolute value is taken, so the term "peak" does *not* imply "positive." The second metric is the signal-to-deviation ratio (SDR), defined as

$$SDR \equiv 10 \log_{10} \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)|^2 \, d\omega}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |D(\omega)|^2 \, d\omega}. \tag{9}$$

Now we proceed to determine the frequency range over which the SNR spectrum is positive as our first definition of the system pass-band.

4.1. Optimal Signal-to-Noise Ratio

Assume that over some frequency range, the magnitude of the signal spectrum will be greater than that of the noise spectrum, while outside of that range, the opposite is true. Symbolically, that is $|W(\omega)| > |N(\omega)|$ for $\omega \in (\omega_1, \omega_2)$, and $|W(\omega)| < |N(\omega)|$ otherwise. This implies that $SNR(\omega)$ will be positive for $\omega \in (\omega_1, \omega_2)$, and negative otherwise. This may be the case, for example, if low-frequency room noise and high-frequency electronic noise levels exceed that of the system output. In general, there may be several separate frequency ranges over which the SNR is positive. However, here we consider only a single frequency range where this is the case, which is to say that we assume the measurement transducer to have only a single pass-band throughout which the SNR spectrum is positive.

Given the above assumption, we can, for example, apply an ideal BPF with cut-off frequencies ω_1 and ω_2 to the recorded signal. This yields a total SNR given by

$$SNR = 10 \log_{10} \frac{\int_{\omega_1}^{\omega_2} |W(\omega)|^2 \, d\omega}{\int_{\omega_1}^{\omega_2} |N(\omega)|^2 \, d\omega} \tag{10}$$

since the power spectra of both the signal and the noise have been set to zero outside of (ω_1, ω_2). It is clear that this total SNR will be greater than that without the BPF, as we have only eliminated regions of negative SNR. In general, it may be possible to narrow the frequency range further and improve the total SNR. However, doing so would likely create a more significant deviation due to filtering. Therefore, we define an *optimal* SNR as that given by Eq. (10), where (ω_1, ω_2) denotes the widest frequency range over which $SNR(\omega)$ is positive. We refer to this frequency range as the *optimal-SNR pass-band*.

The frequency bounds ω_1 and ω_2 are precisely the frequencies at which $|W(\omega)| = |N(\omega)|$. Therefore, at these frequencies, we can write

$$W(\omega_b) = N(\omega_b) e^{j\phi(\omega_b)}, \tag{11}$$

where ϕ denotes the relative phase between W and N, and ω_b represents either of the bound frequencies. Of course, in reality, we cannot measure $W(\omega)$ directly. Instead, we record the microphone signal $Y(\omega)$ and a noise sample to approximate $N(\omega)$. Using Eq. (11), we can rewrite $|Y(\omega_b)|$ as

$$\begin{aligned} |Y(\omega_b)| &= |W(\omega_b) + N(\omega_b)| \\ &= |N(\omega_b)| \cdot \left|1 + e^{j\phi(\omega_b)}\right|. \end{aligned} \tag{12}$$

Since ϕ is a random relative phase, we can compute the expectation value of $|Y(\omega_b)|$, which is found to be (omitting the ω_b dependence)

$$\begin{aligned} \langle |Y| \rangle &= \frac{1}{2\pi} \int_0^{2\pi} |N| \cdot \left|1 + e^{i\phi}\right| d\phi \\ &= \frac{4}{\pi} |N|. \end{aligned} \tag{13}$$

Therefore, on average, our bound frequencies will satisfy the expression

$$|Y(\omega_b)| \approx \frac{4}{\pi} |N(\omega_b)|. \tag{14}$$

Note that the above equation is equivalent to "boosting" the noise spectrum by 2.1 dB and finding the frequencies at which it intersects with the recorded microphone spectrum.

In applying our method, we use the measured microphone signal and an approximate noise spectrum to approximately determine the frequency range over which the SNR spectrum is positive. We do this by first applying 1/3 octave power-smoothing to the microphone and noise spectra, and then determining the frequencies at which Eq. (14) holds. Therefore, by applying a corresponding ideal BPF to the microphone signal, we can achieve the optimal SNR.

4.2. Constrained Deviation

We now turn to our second definition of the system pass-band based on the deviation due to filtering. In principle, the deviation can be computed exactly for any given BPF by filtering the system response and taking the difference between the filtered and unfiltered responses. However, by approximating the low-frequency and high-frequency behavior of the system response, we can efficiently approximate the PDA and SDR as functions of cut-off frequency.

We approximate the low-frequency system response behavior to zeroth order, and the high-frequency behavior to first order, and neglect the phase response. Thus the approximate deviation spectrum is a real-valued, zero-phase, piecewise-linear frequency response, given by

$$\hat{D}(\omega) = \begin{cases} -\gamma_1 & \text{for } |\omega| \le \omega_1 \\ 0 & \text{for } \omega_1 < |\omega| < \omega_2 \\ -\gamma_2 \dfrac{|\omega| - \pi}{\omega_2 - \pi} & \text{for } \omega_2 \le |\omega| \le \pi \end{cases}, \tag{15}$$

where $\gamma_1 = |H(\omega_1)|$ and $\gamma_2 = |H(\omega_2)|$. We expect that a zeroth-order low-frequency approximation will rarely underestimate the true deviation spectrum because very few acoustical system responses exhibit large peaks in amplitude as frequency decreases. Also, we expect that a first-order high-frequency approximation will rarely underestimate the true deviation spectrum because all measured responses will drop off as frequency increases towards the Nyquist frequency, either due to the actual system roll-off or the anti-aliasing filter.

It can be shown that, if the magnitude response of our approximate deviation spectrum matches closely that of the actual deviation spectrum, then the approximate deviation at $k = 0$ should be at least equal in magnitude to the actual deviation at any sample. Thus we estimate the PDA by

$$\left|\hat{d}[0]\right| = \underbrace{\gamma_1 \cdot \frac{\omega_1}{\pi}}_{\text{Low}} + \underbrace{\frac{\gamma_2}{2}\left(1 - \frac{\omega_2}{\pi}\right)}_{\text{High}}, \tag{16}$$

where "Low" and "High" indicate contributions from the low-frequency and high-frequency cut-offs, respectively. We also estimate the total deviation energy by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left|\hat{D}(\omega)\right|^2 d\omega = \underbrace{\gamma_1^2 \cdot \frac{\omega_1}{\pi}}_{\text{Low}} + \underbrace{\frac{\gamma_2^2}{3}\left(1 - \frac{\omega_2}{\pi}\right)}_{\text{High}}. \tag{17}$$

Given the total system response energy, the SDR may then be estimated using the above equation.

To validate these models, we computed the actual deviation metrics for a range of cut-off frequencies using a real system response measurement, and compared these values to those predicted by our models. For simplicity, we examined the effects of the low-frequency and high-frequency cut-offs in isolation. The results of these computations for PDA

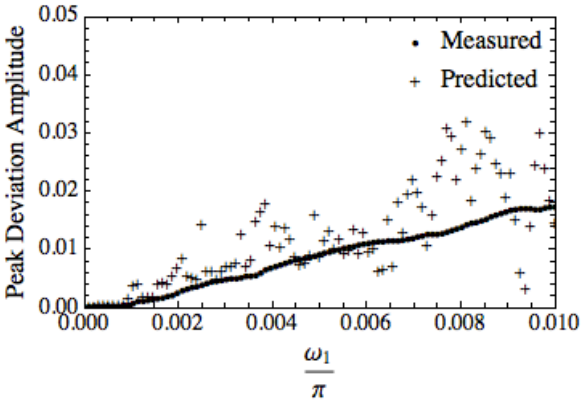


Fig. 2: Zeroth-order low-frequency model results for peak deviation amplitude.

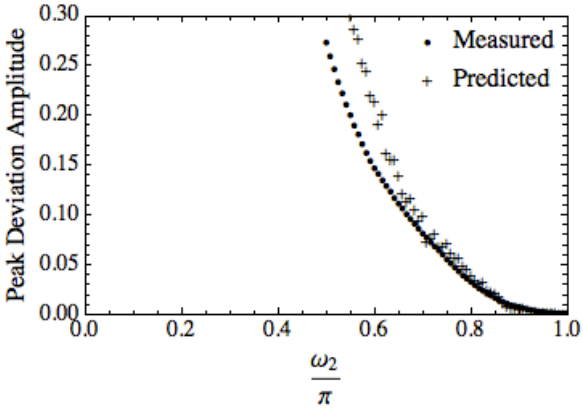


Fig. 3: First-order high-frequency model results for peak deviation amplitude.

are shown in Figs. 2 and 3. We note that, as expected, our models tend to overestimate the actual PDA. The results for SDR are not shown here, but our models were found to consistently underestimate the actual SDR, which is also to be expected.

In order to compute the pass-band for a real measurement, the user must first specify a deviation limit in terms of either PDA or SDR. Then, we determine the lower and upper frequency bounds separately such that the combined deviation will meet the specified limit. For example, the user may specify a maximum tolerable PDA equal to 10% of the measured system IR's peak amplitude before filtering. In this case, we compute each term of Eq. (16) separately using the measured magnitude response in place of γ. We then determine the outer-most frequencies at which the corresponding PDA term first exceeds 5%. We refer to such a frequency range as the *constrained-PDA pass-band*. Alternatively, the user may specify a minimum tolerable SDR and carry out a similar procedure. By defining the pass-band in this way, we ensure that applying an ideal BPF with the same pass-band to our measured system response will yield a deviation that does not exceed the user-specified limit.

We can also use this theory to quantify the resulting deviation caused by applying an ideal BPF with a different pass-band to a given measured system response. For example, when using the optimal-SNR pass-band to design the BPF, we can use Eqs. (16) and (17) to estimate the respective deviation metrics for those pass-band frequencies and then decide if such a deviation is tolerable.

5. PROPOSED PROCEDURE

The proposed IR measurement procedure consists of three phases: I) an initial measurement of the system response using a full-range excitation signal, II) determination of an appropriate pass-band given user inputs, and III) a refined measurement of the system response using an excitation signal tailored to the system under test. A flowchart illustrating this procedure is shown in Fig. 4. It is worth noting at this point that, aside from user inputs, the procedure can be completely automated. Note that the sweep durations for each measurement must be specified by the user.

5.1. Phase I: Initial Measurement

We begin by designing and executing a phase-controlled ESS that sweeps from a sufficiently low initial frequency (set by the user) up to the Nyquist frequency. The initial segment of silence that precedes the sweep allows the microphone to capture just noise, which is used to approximate the ambient noise spectrum.

5.2. Phase II: Pass-Band Selection

Using the initial measurement, the optimal-SNR pass-band is determined following the procedure described in Section 4.1. Next, an estimate of the PDA for this pass-band is computed using Eq. (16). The user is then asked to keep or reject this PDA value. If the user elects to keep this value, then

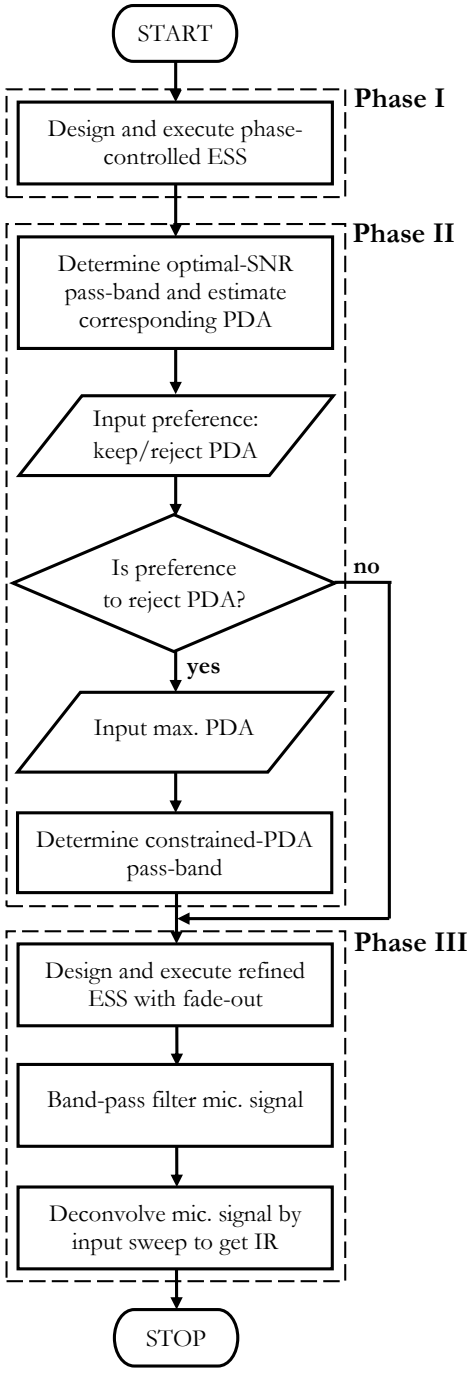


Fig. 4: Flowchart of the proposed measurement procedure. Dashed lines indicate the three phases described in the text. Note that the SDR may be used instead of the PDA, as described in Section 4.2.

the optimal-SNR pass-band frequencies will be used for the second measurement. Otherwise, the user is prompted to enter a maximum tolerable PDA value, and the constrained-PDA pass-band is determined following the procedure described in Section 4.2. The constrained-PDA pass-band frequencies are then used for the second measurement instead.

5.3. Phase III: Refined Measurement

Given the pass-band found in Phase II, a faded-out ESS over that frequency range is designed. This sweep is then executed and the appropriate BPF is applied to the microphone signal. Finally, the IR is extracted via deconvolution.

5.4. General Comments

We now provide some general comments regarding the above procedure. First of all, it may be necessary to check several sampling rates until the appropriate one is found. If the sampling rate is too low, the system output will not drop below the noise floor until just below the Nyquist frequency, if at all. In this case, the high-frequency bound that satisfies Eq. (14) will either not exist or will be approximately equal to the Nyquist frequency. This is an indication that the sampling rate should be increased. Nevertheless, this procedure may still be implemented at lower sampling rates to achieve an improved SNR and a constrained pre-response, although the high-frequency end of the system response will not be captured.

In order to carry out this procedure efficiently, the sweep duration for the initial measurement may be much shorter than that for the refined measurement. As the initial measurement is used simply to determine an approximate pass-band, the user may choose to sacrifice some accuracy in this calculation in return for a shorter overall duration. While we did not explore this trend in detail, in most cases, we do not expect this loss of accuracy to be significant.

Also, it is worth noting that this proposed procedure need not be carried out in its entirety every time. For example, when measuring a large number of IRs with the same loudspeaker/microphone pair, it may only be necessary to carry out the full procedure for the first measurement. In the case of measuring a loudspeaker's directivity with a single microphone, for example, it may be sufficient to

first run the entire procedure for the on-axis measurement, but thereafter simply use the same refined sweep and pass-band for all off-axis measurements.

Alternatively, for measurements using many microphones or loudspeakers, it may be desirable to assign the same pass-band to all measurements. For example, when measuring head-related transfer functions, a single pair of microphones typically records stimuli from a rotating array of many loudspeakers. In this case, one could perform the initial measurement for each loudspeaker/microphone pair, and determine each pass-band using the same settings (i.e. either all keep the optimal-SNR pass-band or input the same maximum PDA). Once each of these pass-bands is found, the lowest low-frequency limit and the highest high-frequency limit may be selected to define a global pass-band, to be used for all refined measurements.

The final measured SNR can be approximately computed using the band-pass filtered microphone and noise signals. In the event that this SNR is too low, additional steps such as increasing the input signal voltage or suppressing external noise sources may be taken, and the procedure re-executed iteratively until the desired SNR is achieved. Alternatively, repeating the refined measurement multiple times and averaging the resulting IRs yields an increase in SNR of 3 dB per doubling in number of measurements. Also note that the initial and refined measurements need not be performed with the same input voltage. For example, the voltage may be increased for the refined measurement to increase SNR, provided that doing so will not lead to transducer damage or excessive levels of distortion.

6. ILLUSTRATIVE EXAMPLE

In this section, we implement the proposed measurement procedure for a real system. All measurements in this example were taken at a sampling rate of 96 kHz with a Brüel & Kjaer 4189 omni-directional microphone placed 1 m away from an Ascend Acoustics CBM-170 two-way loudspeaker placed in an acoustically-damped listening room.

Our measurements were taken at an input root-mean-square (RMS) voltage level of 1.12 V, which, for a 1-kHz sine tone, produced 75 dB SPL. As indicated by the results found in the appendix, we were able to arbitrarily select our excitation durations without risk of damage to the transducers. We used a 1-second phase-controlled ESS from approximately 20 Hz up to the Nyquist frequency for our initial measurement and 5-second sweeps for all subsequent measurements, and zero-padded all input sweep signals with one second of silence before and approximately two seconds after. We applied the fade-out for the refined measurements by allowing a conventional ESS to continue an additional 1 ms beyond its designated final frequency and multiplying those additional samples by a raised-cosine window tail (equivalent to the second half of a Hanning window). We performed two refined measurements of this system: in Case 1, we elected to keep the estimated PDA value, while in Case 2, we elected to reject it. In order to demonstrate the benefits of our proposed procedure over existing methods, we also measured a 5-second conventional ESS. Results from these measurements are presented in Section 7.

6.1. Initial Measurement

We made an initial measurement of the system using a phase-controlled ESS from ∼23 Hz to 48 kHz. Using the sampled noise spectrum and the measured microphone spectrum, we then computed the optimal-SNR pass-band.

In Case 2 only, we set the maximum tolerable PDA to be equal to 2%, allowing for 1% PDA from both the low-frequency and high-frequency cutoffs. Given this constraint, we then computed the constrained-PDA pass-band. As we will see in Section 7, our 2% PDA limit was in fact already satisfied in Case 1. However, for illustrative purposes, we nevertheless carried out the Case 2 measurement.

Additionally, we computed a *predicted deviation* for each case by actually applying the suggested BPF to the initial measured IR and taking the difference between the filtered and unfiltered IRs.

6.2. Case 1: Optimal SNR

Based on the initial system measurement, we found the optimal-SNR pass-band frequencies to be 26 Hz and 40.6 kHz, as shown in Fig. 5. Thus we designed and measured a 5-second ESS through this frequency range including the additional 1 ms raised-cosine fade-out. We then applied the appropriate BPF to the microphone signal and computed the measured SNR. Next, we deconvolved the filtered microphone

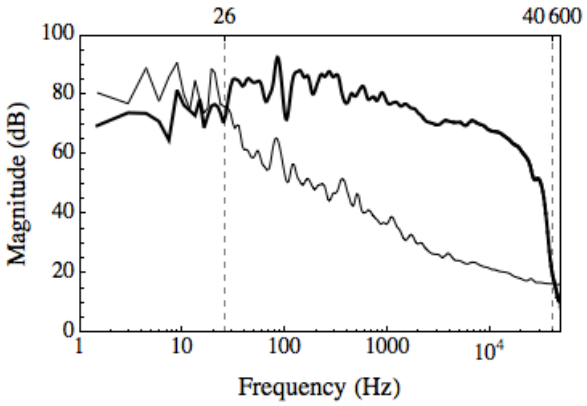


Fig. 5: Magnitude spectra used to determine the optimal-SNR pass-band. The thick curve is the microphone spectrum while the thin curve is the noise spectrum.

signal to extract the measured IR. Finally, we examined the resulting pre-response.

6.3. Case 2: Constrained PDA

Based on the initial system measurement and our specified 2% PDA constraint, we found the constrained-PDA pass-band frequencies to be 81 Hz and 36.9 kHz. Thus we designed and measured a 5-second ESS through this frequency range again including the fade-out. Just as in Case 1, we then filtered the microphone signal by the appropriate BPF, computed the measured SNR, deconvolved to extract the measured IR, and examined the resulting pre-response.

6.4. Comparison to Conventional Technique

We made a final system measurement using a conventional, 5-second ESS from 20 Hz to 24 kHz. We then extracted the system IR in two ways: 1) by exact deconvolution and 2) by time-reversed deconvolution. Next, we computed the measured SNR for each deconvolution technique. Finally, since the time-reversed deconvolution technique effectively applies a BPF from 20 Hz to 24 kHz, we also computed the measured deviation due to filtering by subtracting the IR obtained through exact deconvolution from that obtained through time-reversed deconvolution.

7. RESULTS AND DISCUSSION

We now present and discuss the results of the measurements described in the previous section. The results are summarized in Table 1.

	SNR (dB)			PDA (%)
	Measured		Pred.	
	Raw	BPF		
Initial Meas.	21	–	–	–
Case 1	24	37	33	<0.2
Case 2	26	24	25	<0.4
Conv. ESS	25	32	–	16

Table 1: Results from implementing the proposed procedure. The predicted SNR and PDA values for Cases 1 and 2 were computed by filtering the initial measurement. The raw SNR value for the conventional ESS comes from exact deconvolution whereas the BPF SNR and PDA values come from time-reversed deconvolution.

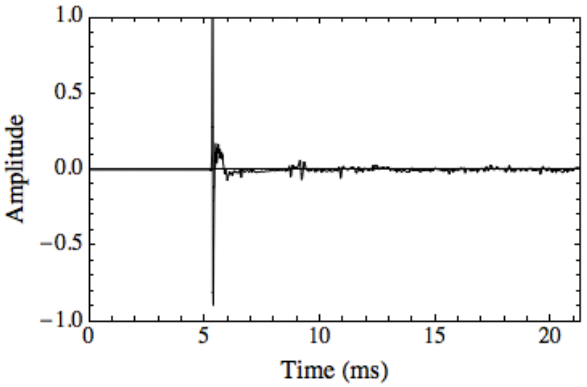
7.1. Initial Measurement

The resulting IR and corresponding magnitude response from the initial measurement are shown in Fig. 6. The IR has been truncated to 2048 samples and normalized by the peak amplitude, with the peak aligned to sample number 512. The magnitude response is obtained by taking the FFT of this normalized, truncated IR. As shown in Table 1, this measurement yielded an SNR of 21 dB.

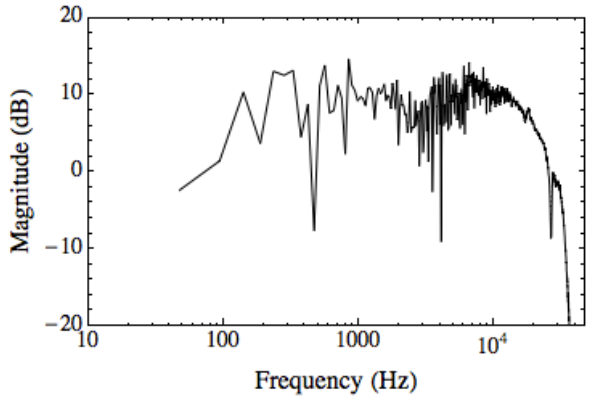
7.2. Case 1: Optimal SNR

The raw (before filtering) measured SNR for Case 1 was 24 dB. Therefore, simply by extending the sweep duration from 1 second to 5 seconds and by concentrating the sweep energy into the appropriate frequency range, we achieved an SNR improvement of 3 dB over the initial measurement. By applying the corresponding BPF, the measured SNR increased further to 37 dB; an improvement of 13 dB solely due to the BPF, and a total improvement of 16 dB over the initial measurement. Note that the predicted optimal SNR in this case was 33 dB; an anticipated improvement of 12 dB due to filtering. This agrees well with the measured improvement of 13 dB due to the BPF.

The predicted deviation due to filtering is shown in Fig. 7 (a), and the corresponding PDA was <0.2%. The actual deviation caused by this BPF is most notably seen as a pre-response, shown in Fig. 7 (b) as

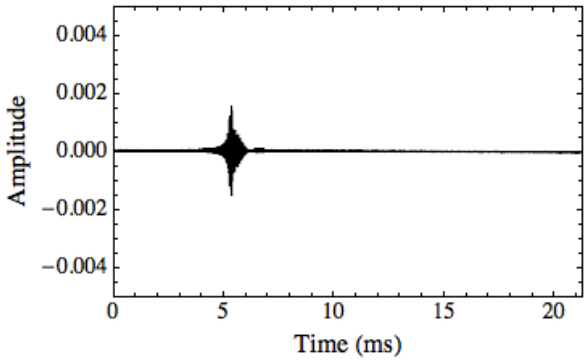


(a) Impulse response

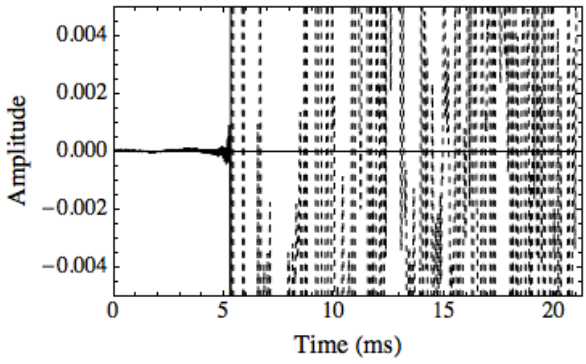


(b) Magnitude response

Fig. 6: Sample system responses measured using the 1-second phase-controlled ESS.



(a) Predicted deviation from initial measurement



(b) Actual pre-response

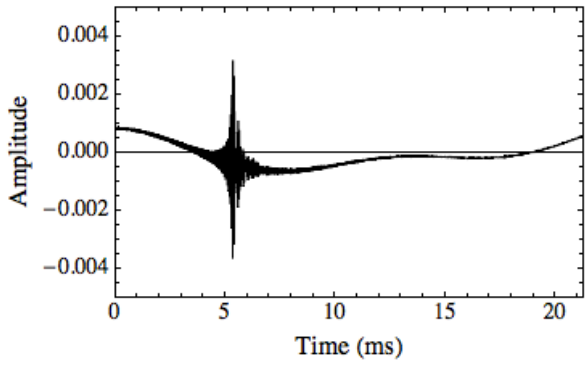
Fig. 7: Case 1 deviation and pre-response. The amplitude has been normalized by the peak of the unfiltered IR. In (b), the solid curve depicts the pre-response up to the peak, while the dashed curve depicts the rest of the IR.

the solid curve. We see very good agreement between the measured pre-response and that of the predicted deviation, shown in Fig. 7 (a).

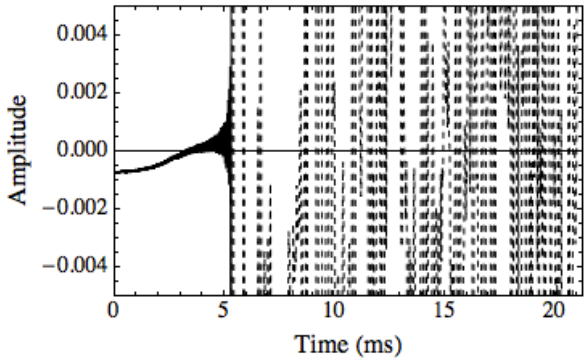
7.3. Case 2: Constrained PDA

The raw measured SNR for Case 2 was 26 dB. Thus, again we see that by concentrating more energy into the system pass-band, we achieved an SNR improvement of 5 dB over the initial measurement. In contrast to Case 1, applying the BPF in this case resulted in a reduction of the measured SNR to 24 dB; a loss of 2 dB. This is surprising, since the predicted SNR in this case was 25 dB; an anticipated improvement of 4 dB due to filtering.

The predicted deviation due to this BPF is shown in Fig. 8 (a), and the corresponding PDA was <0.4%. Note that maximum tolerable PDA limit of 2% has not been reached due to the fact that our PDA models tend to overestimate the actual PDA. The actual deviation caused by this BPF is shown as a pre-response in Fig. 8 (b). Again we see very good agreement between the measured pre-response and that of the predicted deviation, shown in Fig. 8 (a), taking into account the sign convention in our definition of the deviation, given in Eq. (7). We also note the prominent low-frequency components of the deviation for this measurement compared to Fig. 7 due to the higher low-frequency cutoff.



(a) Predicted deviation from initial measurement



(b) Actual pre-response

Fig. 8: Case 2 deviation and pre-response. The amplitude has been normalized by the peak of the unfiltered IR. In (b), the solid curve depicts the pre-response up to the peak, while the dashed curve depicts the rest of the IR.

Although, compared to Case 1, the SNR in this case is much lower and the PDA is actually higher, we note that this measurement accomplished what was specified, which was to limit the PDA to, at most, 2%. Generally, however, the user would likely specify a PDA value that is lower than that obtained with the optimal-SNR pass-band.

7.4. Comparison to Conventional Technique

Exact deconvolution of the conventional ESS yielded a measured SNR of 25 dB. In comparison, time-reversed deconvolution yielded a measured SNR of 32 dB; an improvement of 7 dB due to filtering and a total improvement of 11 dB compared to the proposed procedure's initial measurement. Of course, had different sweep frequencies been chosen for the conventional measurement, we may have obtained a larger SNR. However, unlike in our proposed procedure, we lacked knowledge of the system response for this measurement, so we could not know which sweep frequencies would yield an improved SNR.

The measured deviation is shown in Fig. 9 (a), and has a PDA of 16%. We see that this signal is noisy since the effective BPF is also filtering out the high-frequency noise which has been amplified by exact deconvolution. The pre-response from time-reversed deconvolution is shown in Fig. 9 (b).

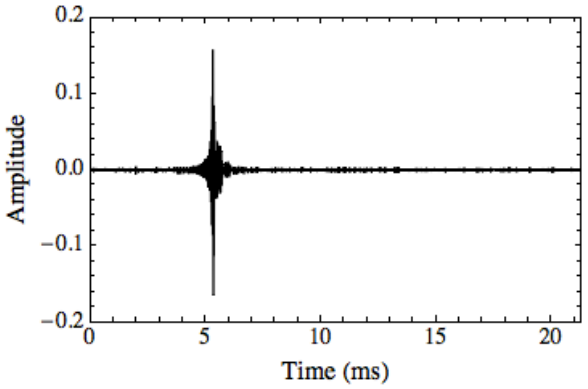
From this measurement we see the consequences of using time-reversed deconvolution. The effective BPF yields a large SNR improvement since it is filtering out a significant amount of noise, but also results in a large pre-response since the cut-off frequencies were not matched to the system under test.

8. SUMMARY

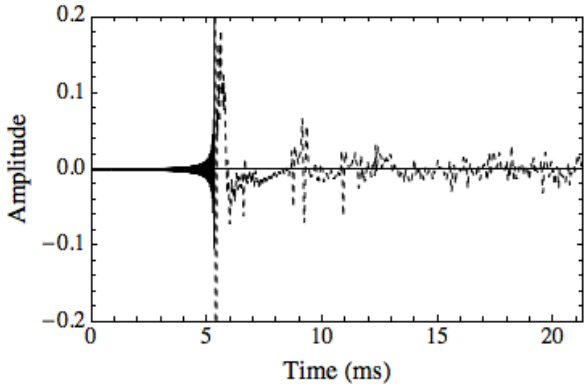
In this paper, we addressed challenges that arise when measuring acoustical impulse responses (IRs) using an exponential sine sweep (ESS) at high sampling rates. First, we discussed various aspects pertaining to the implementation of the ESS. We then proposed two definitions of a system's pass-band, and suggested applying a linear-phase band-pass filter (BPF) to the measured response in order to improve the signal-to-noise ratio (SNR) while constraining any pre-response. We combined these ideas to develop an iterative procedure to measure IRs efficiently and with optimal results in terms of SNR and pre-response. Additionally, we examined the thermal behavior of a typical moving-coil loudspeaker when excited by an ESS to gauge safe operating conditions (see the appendix).

The proposed procedure begins with a quick measurement of a system's response, from which an appropriate pass-band is determined. A second, longer sweep is then designed to cover this pass-band. When implemented in conjunction with a corresponding BPF, this refined measurement achieves a compromise between SNR and pre-response, depending on user input.

We demonstrated the performance of the proposed procedure by implementing it to measure the IR of



(a) Measured deviation



(b) Actual pre-response

Fig. 9: Conventional measurement deviation and pre-response. The amplitude has been normalized by the peak of the IR obtained through exact deconvolution. In (b), the solid curve depicts the pre-response up to the peak, while the dashed curve depicts the rest of the IR.

a real loudspeaker system and comparing its performance to a widely-accepted conventional IR measurement technique. In this experiment, the proposed procedure achieved an SNR of 37 dB with a peak pre-response amplitude of <0.2% of the IR peak. The conventional technique, however, achieved an SNR of 32 dB with a peak pre-response amplitude of 16%. This shows that the proposed procedure achieves superior results compared to the conventional technique in terms of measured SNR and peak pre-response amplitude.

ACKNOWLEDGEMENTS

This research was conducted under a contract from the Sony Corporation of America.

APPENDIX: THERMAL CONSIDERATIONS

When sweeping up to very high frequencies, damage to the loudspeaker due to heating may be a concern. At high voltage levels, the sweep may need to be limited to a safe duration. In order to estimate this duration limit, we modeled the thermal behavior of a typical tweeter's voice-coil. We examined only the tweeter since it heats up much more quickly than a woofer and is notoriously more prone to thermal failure. We used the well-known direct-radiator electromechanical equivalent circuit to estimate the power being converted to heat, which we then input into the thermal model proposed by Zuccatti [7] to predict the voice-coil temperature as a function of time. Voice-coil windings have rated maximum safe operating temperatures, typically at least 155°C [7]. Therefore, we set the sweep duration limit τ as the time at which the voice-coil temperature reaches this upper limit.

Table 2 shows thermal simulation results for a 29-mm dome tweeter[1] using a second-order 2.5-kHz crossover. From these results we see that the thermal limit is not reached until the SPL exceeds approximately 94 dB at 1 m. Therefore, we expect that a sweep up to 48 kHz (or even higher with higher sampling rates) at an RMS voltage level of 2.83 V or less should not cause thermal damage to moving-coil loudspeakers such as the one used in this model.

While these results lend us insight into the approximate thermal behavior of a tweeter when excited by an ESS, we note that they cannot be blindly applied in all cases. For example, when using transducers other than moving-coil loudspeakers, clearly this model will not apply. Moreover, for some transducers, such as electrostatic loudspeakers, electrical arcing, rather than heating, may be the primary limiting factor for the sweep duration. Also, the direct-radiator model is well-known to only accurately predict impedance at low frequencies, and therefore cannot be expected to yield an accurate estimate of the power being converted to heat at the

[1] Electromechanical parameters are from the Scan-Speak D2905/990000 datasheet and the thermal parameters are from Chapman [8]

SPL (dB)	V_{RMS} (V)	τ (s)
88	2.00	-
91	2.83	-
94	4.00	638
97	5.64	133
100	7.97	2.02

Table 2: Thermal simulation results for a 29-mm tweeter for various nominal sound-pressure levels at 1 meter and corresponding RMS voltage levels. The duration limit τ is the time at which the simulated voice-coil temperature reaches 155°C.

high frequencies covered by our sweeps. Nonetheless, we consider these results to be useful guidelines for selecting appropriate sweep durations for similar moving-coil loudspeakers. As for other transducers, we present this work as an example to show how suitable sweep limitations may be determined.

9. REFERENCES

[1] P. G. Craven. "Antialias Filters and System Transient Response at High Sample Rates". *J. Audio Eng. Soc.*, 52(3):216–242, 2004.

[2] A. W. Mills. "On the Minimum Audible Angle". *J. Acoust. Soc. Am.*, 30(4):237–246, 1958.

[3] A. Farina. "Advancements in Impulse Response Measurements by Sine Sweeps". Presented at the AES 122nd Convention, May 2007.

[4] S. Müller and P. Massarani. "Transfer-Function Measurement with Sweeps". *J. Audio Eng. Soc.*, 49(6):443–471, 2001.

[5] H. Ochiai and Y. Kaneda. "A Recursive Adaptive Method of Impulse Response Measurement with Constant SNR over Target Frequency Band". *J. Audio Eng. Soc.*, 61(9):647–655, 2013.

[6] A. Torras-Rosell and F. Jacobsen. "A New Interpretation of Distortion Artifacts in Sweep Measurements". *J. Audio Eng. Soc.*, 59(5):283–289, 2011.

[7] C. Zuccatti. "Thermal Parameters and Power Ratings of Loudspeakers". *J. Audio Eng. Soc.*, 38(1/2):34–39, 1990.

[8] P. J. Chapman. "Thermal Simulation of Loudspeakers". Presented at the AES 104th Convention, May 1998.

[9] K. Vetter and S. di Rosario. "ExpoChirpToolbox: a Pure Data implementation of ESS impulse response measurement". Presented at the 4th Pure Data Convention, 2011.