



Audio Engineering Society Convention Paper

Presented at the 129th Convention
2010 November 4–7 San Francisco, CA, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. 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.

Estimating Room Impulse Responses from Recorded Balloon Pops

Jonathan S. Abel¹, Nicholas J. Bryan¹, Patty P. Huang¹, Miriam A. Kolar¹, and Bissera V. Pentcheva²

¹*Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, Stanford, CA, 94305, USA*

²*Department of Art & Art History, Stanford University, Stanford, CA, 94305, USA*

Correspondence should be addressed to Jonathan S. Abel (abel@ccrma.stanford.edu)

ABSTRACT

Balloon pops are convenient for probing the acoustics of a space, as they generate relatively uniform radiation patterns and consistent “N-wave” waveforms. However, the N-wave spectrum contains nulls which impart an undesired comb-filter-like quality when the recorded balloon pop is convolved with audio. Here, a method for converting recorded balloon pops into full audio bandwidth impulse responses is presented. Rather than directly processing the balloon pop recording, an impulse response is synthesized according to the echo density and frequency band energies estimated in running windows over the balloon pop. Informal listening tests show good perceptual agreement between measured room impulse responses using a loudspeaker source and a swept sine technique, and those derived from recorded balloon pops.

1. INTRODUCTION

The impulse response of a room can be measured in many ways and with varying degrees of accuracy. A preferred method, for its precision and high signal-to-noise ratio, is the swept-sine technique [1, 2], which requires a suite of recording and playback equipment, including a powerful loudspeaker capable of output over a wide bandwidth. Even when using a handheld recorder to capture the

loudspeaker response, the material and logistical requirements of this technique are significant. To minimize equipment requirements, the excitation signal can be manually generated using a starter pistol or an orchestral whip; however, a starter pistol is inappropriate or prohibited in some contexts, and a whip has an undesirable radiation pattern. Both are difficult to trigger remotely, away from reflecting objects.

A logistically simple alternative to these acoustic measurement scenarios is to use a balloon pop as the excitation source and a handheld recorder as the recording device. While not as repeatable as some other methods, this system provides a number of advantages: it is physically compact and easily portable. Its power requirements are minimal, the radiation pattern of the popped balloon is relatively uniform, and the “N-wave” time response of the balloon pop is known (determined by the balloon diameter) [3].

Deriving an accurate room impulse response from a recorded balloon pop is somewhat problematic, however. While a number of important room acoustic features can be inferred from the raw response, convolving the balloon pop response with other audio material—the process used to impart the measured room characteristics on outside sound sources, for auralization purposes, for example—produces an undesirable comb-filter-like quality. This sonic distortion arises from the pattern of nulls present in the balloon pop waveform N-wave Fourier transform. The nulls are not part of the spatial impulse response being measured, but an artifact of the excitation mechanism. Because the nulls are deep and located at frequencies dependent on the balloon pop radiation direction, it is not possible to invert the filtering imposed by the N-wave, and other means must be employed to produce an impulse response suitable for use in a convolutional reverberator.

In this paper we present a method for converting recorded balloon pops into full audio bandwidth impulse responses that can be convolved with audio material to create auralizations of the acoustic effect of the measured space that do not contain artifacts of the balloon pop excitation. Rather than processing the balloon pop recording to remove its N-wave characteristics, the approach we take is to analyze the recorded balloon response and then synthesize an artifact-free impulse response by generating and filtering a pattern of echoes perceptually similar to those seen in the raw balloon response.

As an example application, we present our conversion of balloon pop impulse responses and a convolutional reverberation auralization using the estimated response as implemented for the Icons of Sound Project, an interdisciplinary exploration of the Byzantine-era perceptual experience within the

religious center Hagia Sophia. In our application, balloon pops recorded at this 1,500 year-old World Heritage Site [4] are converted to synthesized impulse responses, which are then used to create an auralization of the Cherubikon, a hymn performed there in the 6th century, in order to reconstruct a historical performance scenario. It is worth noting that in this work, the application necessitated the measurement method; Hagia Sophia is exemplary of a venue whose public context fundamentally constrains the acoustic measurement method. Other historical, archaeological and naturally formed sites similarly require a rapid measurement process with little equipment. Thus, the balloon pop/handheld recorder scenario is mandated, and a functional impulse response recovery method necessary.

This paper is organized as follows: §2 explains the dynamics of balloon pops. §3 presents methods for estimating the time evolution of the echo density over the course of the impulse response and discusses methods for synthesizing that echo density profile. §4 examines the spatial character of the balloon pop response, describing methods for estimating and synthesizing the time behavior of the inter-channel cross-correlation (synthesizing stereo pulse sequences matching the measured balloon recording cross-correlation as it evolves over time). §5 discusses the estimation of the balloon pop energy as a function of time and frequency, and methods for imprinting that energy on the different frequency bands of the synthesized echo sequence. §6 give results including synthesis of an impulse response using data from a balloon pop recording made in Memorial Church at Stanford University as it compares to an impulse response measured in the same location using a loudspeaker and swept-sine technique. Finally, balloon pops recorded in Hagia Sophia in Istanbul are converted into impulse responses having reverberation time and other acoustical parameters consistent with those reported in the literature [5, 6, 7].

2. BALLOON POP DYNAMICS

Fundamental to our study is the nature of the balloon pop acoustic disturbance. Bursting balloons generate N-wave type pressure waveforms, as put forth by Deihl and Carlson in [3]. This can be seen by considering the geometry shown in Fig. 1 (a). A spherical balloon of radius ρ bursts in the presence

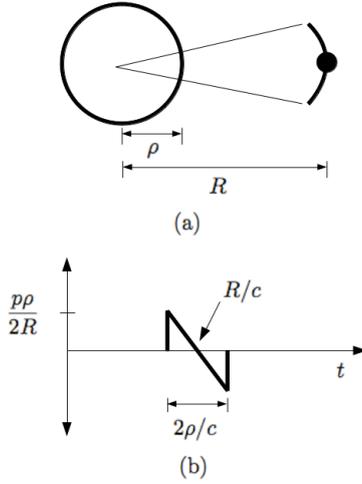


Fig. 1: Balloon Geometry and N-Wave. (a) Spherical balloon of radius ρ inflated to pressure p and microphone positioned a distance R from the balloon center. (b) An N-wave waveform $2\rho/c$ in duration radiated from the bursting balloon.

of a microphone positioned a distance R from the balloon center. Assuming the balloon skin instantly vanishes at time $t = 0$, a spherical region having pressure p remains.

Over time, half of the pressure $p/2$ radiates outward from the origin while the other half propagates inward, producing an inverting reflection from the balloon center. The pressure along an arc initially at radius $x \leq \rho$ inside the balloon will arrive at a microphone a radius R from the balloon center at time $(R - x)/c$, where c is the speed of sound in air, and with pressure $px/2R$ due to spherical spreading. Note that the pressure seen at the microphone varies linearly with time, and this trend is continued with the arrival of the portion of the pressure initially traveling inward toward the balloon center. The pressure waveform seen at the microphone is then

$$n(t) = \begin{cases} \frac{p\rho}{2R}(R - ct), & ct \in [R - \rho, R + \rho], \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

as plotted in Fig. 1 (b).

As an experimental example, a balloon pop recorded in Memorial Church is plotted on a logarithmic time

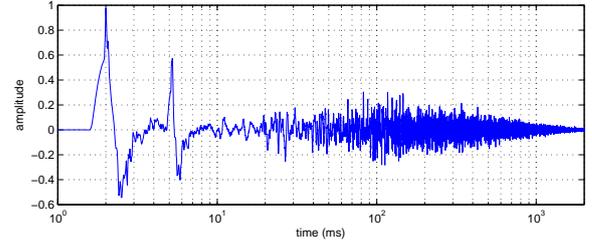


Fig. 2: Balloon Pop Recording, Memorial Church, Stanford University. The recorded waveform is plotted on a logarithmic time axis to show details of the early response. The direct path and floor reflection are visible as N-wave waveforms.

scale in Fig. 2. The recorded pressure waveform, denoted $b(t)$, reveals individual arrivals of the balloon pop response onset, giving way to a diffuse late-field response. The direct path arrival and reflection from the marble floor have the character of the idealized N-wave response, but with a prolonged onset and release likely caused by the finite time taken for the balloon to burst. Note that the leading edges of the direct path and floor reflection arrivals are somewhat peaked. This feature may be the result of the nonlinear propagation mechanism responsible for the formation of shock waves in which high-pressure waveform regions travel faster than low-pressure waveform regions [8].

Of interest in using balloon pop recordings as impulse responses is the spectrum $N(\omega)$ of the balloon pop waveform $n(t)$. The magnitude transform of the direct path arrival of the balloon pop is shown in Fig. 3. The N-wave structure of the balloon pop has spectral nulls at dc and a harmonic series of frequencies at the inverse balloon diameter (expressed in frequency units),

$$N(\omega) = \frac{pc}{2R} \frac{\nu \sin(\nu) - \cos(\nu)}{j\nu^2}, \quad (2)$$

where $\nu = \omega c/\rho$ is the frequency ω scaled by half the sound propagation time across the balloon. It is these nulls which impart an unwanted comb-filter-like character. The overall spectrum of the balloon pop response in an acoustic space does not contain this precise series of nulls, as different radiation directions from the balloon are associated with differ-

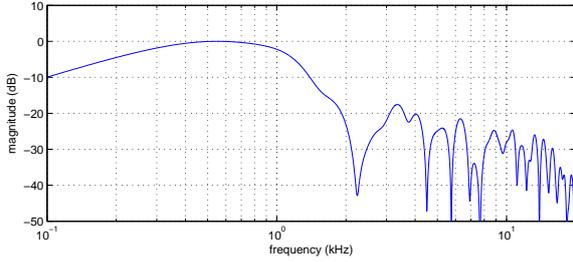


Fig. 3: Measured N-wave Transform Magnitude. The transform magnitude of the direct path arrival for the Memorial Church balloon pop response of Fig. 2.

ent diameters (except in the case of spherical balloons). Because of this, the nulls do not overlap, except at dc, and disallow any form of linear inverse filtering.

3. ROOM ECHO DENSITY

We now examine the measured balloon pop recording to produce a sequence of full audio bandwidth echoes. The echo sequence is generated in such a way as to create the same psychoacoustic impression as the original sequence of arrivals.

3.1. Balloon Pop Recording Echo Density

For a pattern of echoes of at least a modest echo density, the psychoacoustic impression created does not depend on the precise timing of the arrivals, but rather the overall echo density [9]. The approach we take here is not to identify and replace each of the N-wave arrivals present in the recorded balloon response with full bandwidth pulses, but rather to synthesize a sequence of pulses having the same evolving density as that of the measured balloon pop response $b(t)$.

The normalized echo density (NED), introduced in [10], measures the echo density of a sequence on a scale ranging from 0 to around 1 by comparing the percentage of sequence outliers to that expected for Gaussian noise. Over a sliding reverberation impulse response window, the NED $\eta(t)$ is the fraction of impulse response taps which lie outside the win-

dow standard deviation:

$$\eta(t) = \frac{1/\text{erfc}(1/\sqrt{2})}{2\Delta + 1} \sum_{\tau=t-\Delta}^{t+\Delta} \mathbf{1}\{h^2(\tau) > \sigma^2(t)\}, \quad (3)$$

where $h(t)$ is the reverberation impulse response (assumed to be zero mean), $2\Delta + 1$ is the window length in samples, $\sigma(t)$ is the window standard deviation,

$$\sigma^2(t) = \frac{1}{2\Delta + 1} \sum_{\tau=t-\Delta}^{t+\Delta} h^2(\tau), \quad (4)$$

$\mathbf{1}\{\cdot\}$ is the indicator function (returning one when its argument is true and zero otherwise), and $\text{erfc}(1/\sqrt{2}) \doteq 0.3173$ is the expected fraction of samples lying outside a standard deviation from the mean for a Gaussian distribution. In the presence of a few energetic arrivals, relatively few response samples will lie outside the window standard deviation and the NED will be small, near zero. Gaussian noise, on the other hand, produces NED values near 1. It turns out that NED is a reliable indicator of perceived echo density [9] and is used here to estimate the perceived echo density along the recorded balloon pop response $b(t)$.

3.2. Echo Sequence Synthesis

In [11], an expression is presented relating the normalized echo density $\eta(t)$ to the absolute echo density (AED) $e(t)$ measured in echoes per second,

$$\eta(t) = \frac{e(t)}{e(t) + 1/\delta}, \quad (5)$$

where δ represents the typical echo duration.

To generate a sequence of echoes, the balloon pop normalized echo density $\eta_b(t)$ is estimated and converted to absolute echo density. Rearranging (5), we have

$$e(t) = \frac{\eta_b(t) c/2\rho}{1 - \eta_b(t)}. \quad (6)$$

Note that the balloon pop response echoes are N-wave doublets, not simple pulses. However, by integrating the response $b(t)$, the N-wave functions are converted into parabolic pulses as seen in the ideal case in Fig. 4,

$$\int_0^t n(\tau) d\tau = \max \left\{ 0, \frac{p\rho}{2c} \left[1 - \left(\frac{ct - R}{\rho} \right)^2 \right] \right\}, \quad (7)$$

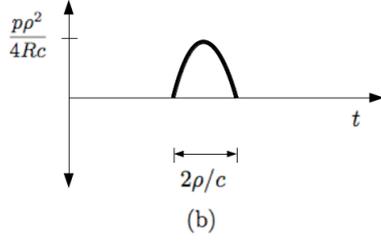


Fig. 4: Integrated N-wave waveform. The time integral of an N-wave is a parabolic pulse $2\rho/c$ in width.

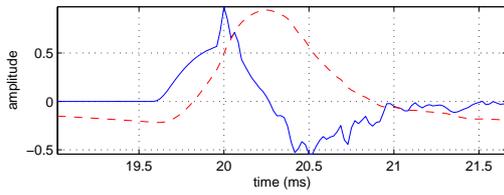


Fig. 5: Recorded Balloon Pop N-wave (solid), Integrated N-wave (dashed).

and in the measured response in Fig. 5.

Taking δ as duration of a full bandwidth pulse, and fixing $e(t)$, $\eta_b(t)$ can be used to relate the balloon response NED $\eta_b(t)$ to the “impulse response” NED, η_h , by

$$\eta_h = \frac{\eta_b}{(1 - \eta_b) \cdot 2\rho/\delta c + \eta_b} \quad (8)$$

where $2\rho/c\delta$ is the ratio of the balloon and full bandwidth pulse durations, which are on the order of 10–20 for the balloons and loudspeakers we’ve used. The relationship between the balloon and full bandwidth NEDs is shown in Fig. 7. When η_b is near zero, indicating a sparse echo sequence, so is η_h . When η_b is close to one, indicating Gaussian statistics, so is η_h . For intermediate values of η_b , η_h is smaller than η_b .

The integrated balloon response NED profile $\eta_b(t)$ is shown in Fig. 6, along with the NED profile estimated for a full bandwidth pulse sequence, denoted $\eta_h(t)$. (Note that the computed full bandwidth NED was held fixed at $\eta_h = 0.995$ after first obtaining that value.) The balloon in full bandwidth profiles take on their typical character for room re-

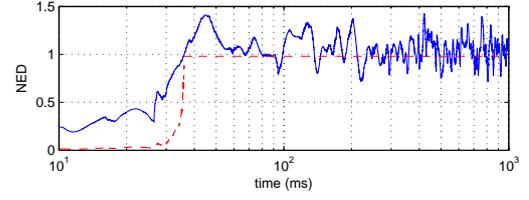


Fig. 6: Recorded, Estimated Normalized Echo Density Profiles. The NED of the integrated recorded (Memorial Church) balloon pop is shown (solid) along with the NED estimated for a 20kHz-bandwidth impulse response (dashed).

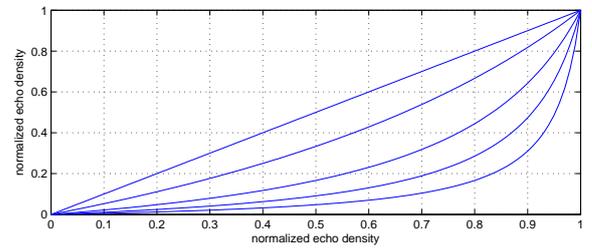


Fig. 7: Normalized Echo Density, Echo Duration Relationship. The NED of a sequence of short-duration echoes is plotted as a function of the NED of the same sequence but with longer-duration echoes. Curves for echo duration ratios of 1, 2, 5, 10, and 20 (top to bottom) are shown.

sponses, starting near zero at the room response onset, and increasing to near 1 as the room becomes mixed. In this way, the perceived echo density of a loudspeaker-generated impulse response is expected to be less than that of the balloon response until the room is fully mixed with an NED near 1.

Once the AED profile $e(t)$ is known, the process described in [10] is used to synthesize a pattern of arrivals. Pulses are generated at Poisson-distributed time intervals according to the estimated echo density profile $e(t)$. Rather than generating an echo pattern covering the entire balloon response, the first few clear arrivals (typically the direct path and floor reflection) are placed by hand. The NED of the remaining portion of the room response is used to generate the echo sequence, denoted $p(t)$. At this stage of the processing, the pulse amplitudes are drawn

from a Gaussian distribution and scaled according to the local echo density so as to give a roughly constant energy profile. Fig. 8 shows an example synthesized pattern of full bandwidth echoes $p(t)$ based on the balloon response $b(t)$ of Fig. 2.

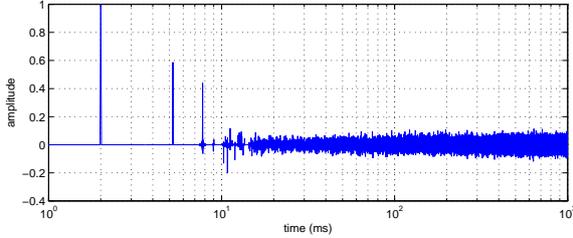


Fig. 8: Synthesized Memorial Church Echo Pattern.

4. ROOM SPATIAL CHARACTER

If the balloon pop response is recorded with stereo or binaural microphones, the perceived spatial character should be retained in the synthesized stereo impulse response. A simple method for doing so is described below, while more sophisticated and accurate methods are available in [12].

The cross-correlation between the left and right channels in the mid and high frequencies is indicative of the spatial nature of the room response [5, 15]. Small cross-correlation values near zero indicate energy independently arriving at the listener's ears, creating a sense of envelopment. By contrast, cross-correlations near 1 are produced in the presence of energy arriving coherently from a specific direction. In this work, we estimate the inter-channel cross-correlation in a running window over the recorded response and impose that correlation on synthesized stereo echo sequences.

The inter-channel cross-correlation evaluated at lag l over a running unit-sum window $w(t)$ is

$$C(t) = \frac{\sum_{\tau=t-\Delta}^{t+\Delta} w(\tau) b_1(\tau) b_2(\tau + l)}{\left[\sum_{\tau=t-\Delta}^{t+\Delta} w(\tau) b_1^2(\tau) \right]^{\frac{1}{2}} \left[\sum_{\tau=t-\Delta}^{t+\Delta} w(\tau) b_2^2(\tau) \right]^{\frac{1}{2}}}, \quad (9)$$

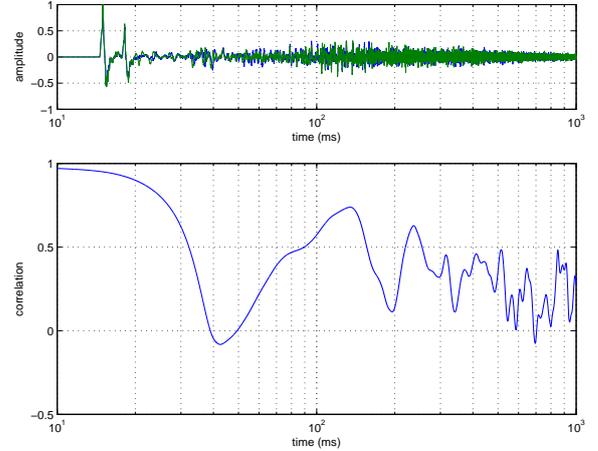


Fig. 9: Memorial Church Cross-Correlation. The time evolution of the cross-correlation computed over a 50ms running window along the stereo recording of a balloon pop in Memorial Church is shown (lower) along with the left and right channel impulse responses overlaid (upper).

Typically, the cross-correlation is evaluated at lags in the range $l \in [-1, 1]$ ms, and the maximum over that range is taken as the cross-correlation coefficient.

To synthesize a stereo echo pattern having the measured time evolution of the correlation coefficient (and correlation maximizing lag), three statistically independent sequences may be generated according to the NED profile. Two sequences are first assigned to the two channels. The third sequence is added to the first channel in proportion to the cross-correlation square root. It is then delayed, according to the maximizing correlation lag, scaled and added to the second channel. While this procedure generates the desired correlation at the desired lag, it is computationally cumbersome, and we suggest simply imprinting the measured zero-lag cross-correlation on the synthesized response.

A stereo balloon pop response was recorded in Memorial Church using an ear-width spaced pair of omnidirectional microphones aligned parallel to the front and back walls of the church. The balloon was positioned on a line between and perpendicular to the microphones.

Fig. 9 shows the stereo balloon response on a logarithmic time axis; note the direct path, initial floor reflection, and increasing arrival density. Also shown is the zero-lag cross-correlation $C(t)$. The cross-correlation at the arrival time of the balloon pop excitation is near 1.0, consistent with the nearly overlapping direct path and initial floor reflection channel responses. After this portion of the balloon response, a sizable fraction of the energy is returned from the sides of the church including from the balconies, and various wooden pew surfaces. The cross-correlation decreases accordingly, fluctuating around a value of 0.2 where it remains throughout the late-field. In many rooms, one would expect the impulse response cross-correlation to decrease to near zero in the late-field. In this example, however, a pattern of energy reflecting between the front and back surfaces of the church arrives at both microphones nearly simultaneously, producing similar waveforms and a positive correlation.

In contrast, a balloon pop recorded in Hagia Sophia, via two omnidirectional microphones (worn by the recording engineer just above the ears) has the cross-correlation profile shown in Fig. 10. While the direct path and first reflection are highly correlated, after a short time the correlation coefficient decreases to near zero where it remains.

To synthesize a stereo impulse response having a prescribed correlation, we start with the stereo impulse pattern generated according to the echo density above. The channels of the echo sequence are generated independently, and are expected to have a correlation near zero throughout their duration. To generate a stereo pair having the prescribed cross-correlation, each pair of independent samples is multiplied by the matrix \mathbf{M} ,

$$\mathbf{M} = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (10)$$

By setting $\theta = 0$, \mathbf{M} becomes the identity matrix, and the signals are unchanged. By setting $\theta = \pi/4$, $\mathbf{M} = \mathbf{1}\mathbf{1}^T/\sqrt{2}$ ($\mathbf{1}$ being a column of 1s), a 2-channel monophonic signal is created. A little algebra shows that to achieve a cross-correlation of $C(t)$, $\theta(t)$ should be set to

$$\theta(t) = \frac{\arcsin C(t)}{2}. \quad (11)$$

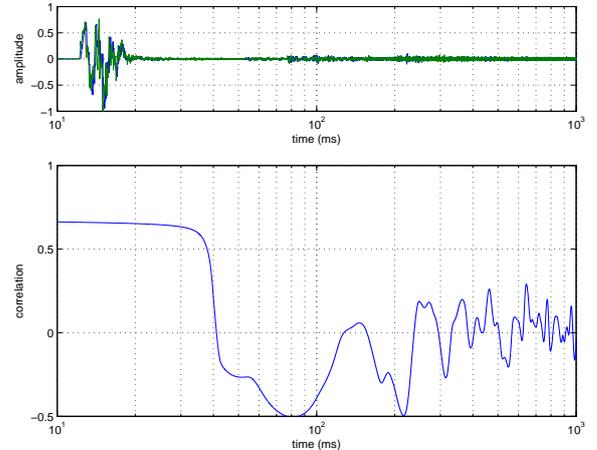


Fig. 10: Hagia Sophia Cross-Correlation. The time evolution of the correlation coefficient computed over a 50ms running window along the stereo recording of a balloon pop in Memorial Church is shown (lower) along with the left and right channel impulse responses overlaid (upper).

5. TIME-FREQUENCY RESPONSE

Starting with echo sequences having the proper echo density and cross-correlation profiles, we next imprint the spectral behavior of the recorded balloon response. This is done by splitting the balloon response into a set of third-octave frequency bands using a perfect amplitude reconstruction zero-phase filter bank. Band energies are estimated over time for the balloon response and synthesized echo sequence. The balloon response energy profiles are applied as time windows to the echo sequence bands. Finally, the windowed echo sequence bands are equalized so that the direct arrival is white and then summed.

5.1. Filter Bank

A perfect amplitude reconstruction zero-phase filter bank is used to split the balloon pop response $b(t)$ and synthesized echo sequence $p(t)$ into frequency bands $b_k(t)$ and $p_k(t)$ respectively. The filter bank is constructed via a cascade of band-splitting squared Butterworth filters as shown in Fig. 11 [13]. The band filter magnitude transfer functions are shown in Fig. 12. By using 3th-order Butterworth band splitting filters applied forward and backward

in time, zero-phase responses and transitions of 60dB/octave are achieved.

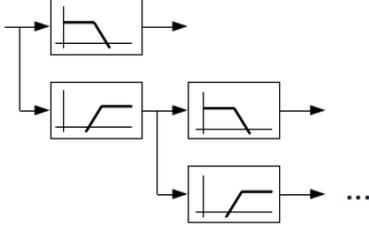


Fig. 11: Filter Bank Signal Flow Architecture.

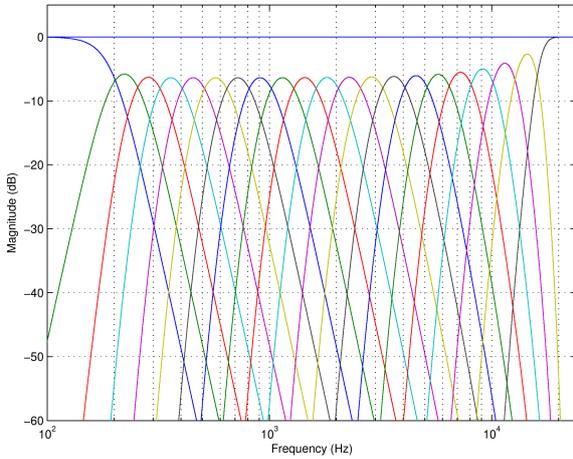


Fig. 12: Filter Bank Magnitude Response. The magnitude responses of third-octave band analysis/synthesis filters are shown.

5.2. Band Energy Profile Analysis

Once separated into bands, frequency-dependent energy profiles for the balloon response $\beta_k^2(t)$ and echo sequence $\nu_k^2(t)$ can be computed by smoothing the square band signals $b_k^2(t)$ and $p_k^2(t)$ over running windows,

$$\beta_k^2(t) = b_k(t)^2 * w(t), \quad (12)$$

$$\nu_k^2(t) = p_k(t)^2 * w(t), \quad (13)$$

where $w(t)$ is a positive smoothing window having unit sum, $\sum_t w(t) = 1$. A 10ms-long Hanning window is used so as to reveal short-duration features in the original balloon pop band energies.

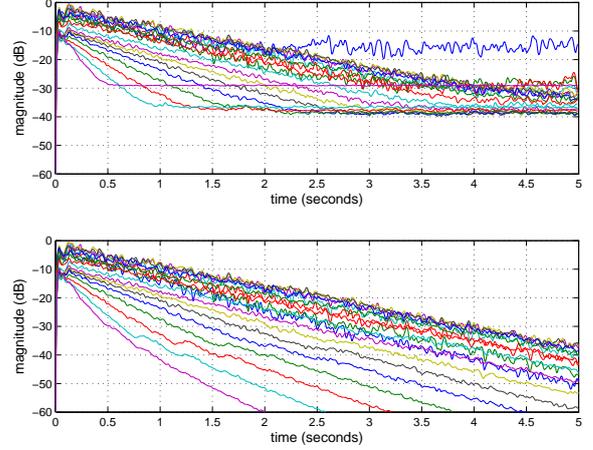


Fig. 13: Measured, Extrapolated Band Energies. The dB band energies, smoothed over 50ms Hanning windows, are shown for the balloon response recorded in Memorial Church (upper). The band energies extrapolated to below the estimated noise floor (lower).

As an example, consider the Memorial Church balloon pop of Fig. 2 (a); its spectrogram appears in Fig. 14. The balloon pop smoothed band energies $\beta_k^2(t)$ are shown in Fig. 13. The high frequencies decay relatively quickly, whereas the low frequencies persist.

Note that the band energies do not decay fully, but arrive at a frequency-dependent noise floor (around -40dB). So as to provide a more natural decay beyond the measurement noise floor, the band energies $\beta_k^2(t)$ are extrapolated using the method described in [14]. Fig. 15 and Fig. 13 (b) show the extended spectrogram and band energy profiles respectively.

5.3. Band Energy Profile Synthesis

The different filter bank bands of the echo sequence $p_k(t)$ are then imprinted with the band amplitudes by first windowing with the balloon pop amplitude profiles normalized by those of the noise sequence,

$$\gamma_k(t) = \beta_k(t) / \nu_k(t). \quad (14)$$

Then windowed band echo sequences $p_k(t)\gamma_k(t)$ are summed to form a synthesis $\hat{b}(t)$ of the balloon pop

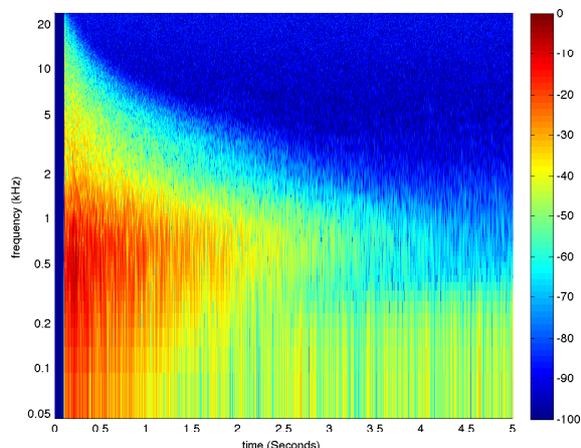


Fig. 14: Balloon Pop Spectrogram. The spectrogram above was generated from an unprocessed balloon pop recorded in Memorial Church.

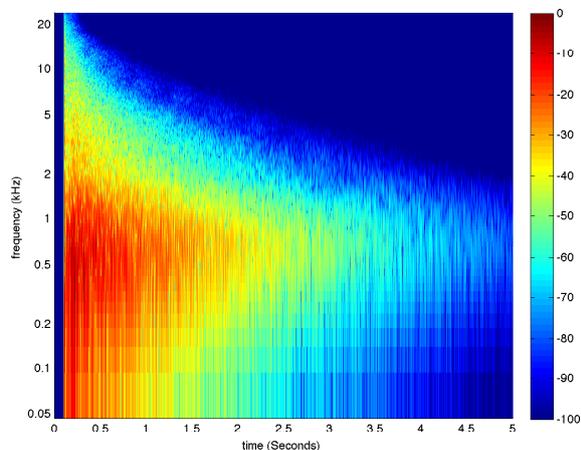


Fig. 15: Extended Balloon Pop Spectrogram. The balloon pop of Fig. 14 was extended to below its noise floor using the technique described in [14].

recording,

$$\tilde{b}(t) = \sum_k p_k(t) \gamma_k(t). \quad (15)$$

This synthesis $\tilde{b}(t)$ is expected to be perceptually similar to the original recording $b(t)$, save for the effects of having full bandwidth arrivals. To generate the final impulse response estimate $\tilde{h}(t)$, the windowed band echo sequences are equalized to whiten the direct path arrival, and summed:

$$\tilde{h}(t) = \sum_k p_k(t) \gamma_k(t) \alpha_k, \quad (16)$$

where α_k are the inverse direct path arrival band gain estimates.

6. METHOD EXAMPLES AND RESULTS

6.1. Memorial Church Measurement

Balloon pop responses were recorded in Memorial Church using a pair of ear-spaced omnidirectional microphones and balloons attached to 3-meter-long sticks held roughly 2.5 meters above the ground and approximately 6 meters from the microphone pair. The spectrogram of the recorded, extended balloon response $b(t)$ is shown in Fig. 16 (upper). A synthesized balloon response $\tilde{b}(t)$ was created; its spectrogram appears in Fig. 16 (lower). The spectrograms of the recorded and synthesized balloon pops are similar, and informal listening tests confirm their perceptual equivalence, including during the psychoacoustically important balloon pop response on-set.

A baseline room impulse response $h(t)$ was measured using a swept-sine technique and a Mackie HR824 loudspeaker positioned near the balloon burst location. The measured impulse response $h(t)$ spectrogram is shown in Fig. 17 (upper). A synthesized impulse response $\tilde{h}(t)$ was created by equalizing the balloon response estimate $\tilde{b}(t)$; its spectrogram also appears in Fig. 17 (lower). The spectrograms show similar behavior. However, since the loudspeaker energy is radiated mainly out of the front of the speaker, and the balloon pop radiates its energy rather uniformly, the direct path arrival is relatively more powerful in the case of the loudspeaker impulse response measurement. Future impulse response measurements using a dodecahedral loudspeaker source would provide a better basis for

comparison with the omnidirectional balloon pop radiation pattern.

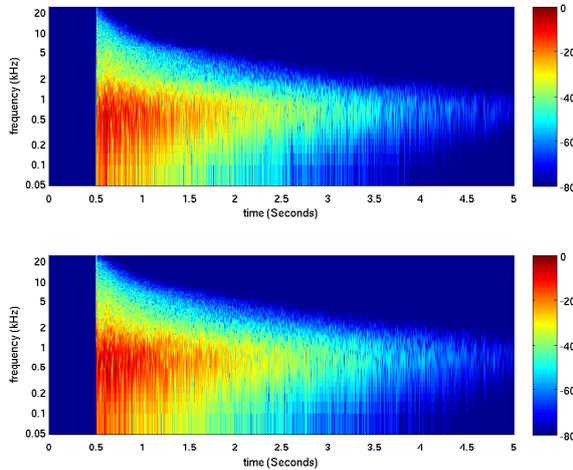


Fig. 16: Memorial Church Measured and Estimated Balloon Response Spectrograms. The extended balloon pop response (upper) and its full bandwidth estimate (lower) are shown.

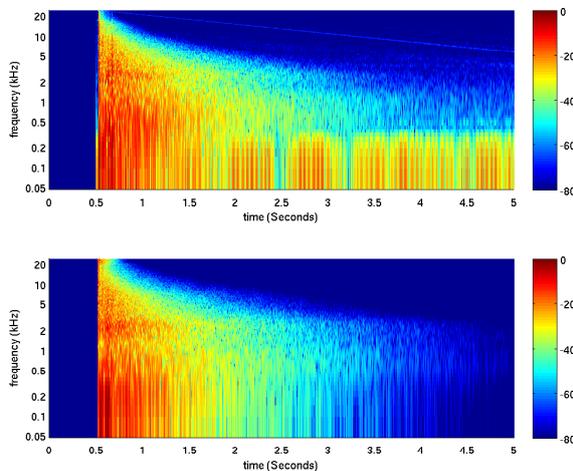


Fig. 17: Measured and Estimated Impulse Response Spectrograms. The impulse response spectrogram measured using a loudspeaker and swept sinusoid (upper) is shown along with an estimated impulse response (lower) based on a balloon pop recording from a similar location.

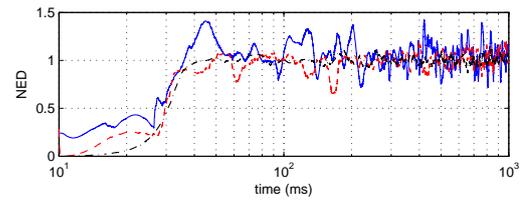


Fig. 18: Measured and Estimated NED Profiles. The NED profile of the integrated Memorial Church balloon pop response (solid) is shown along with the NED of the synthesized echo sequence designed to have the same arrival density (dashed-dot), and the NED profile of the loudspeaker-generated impulse response (dashed).

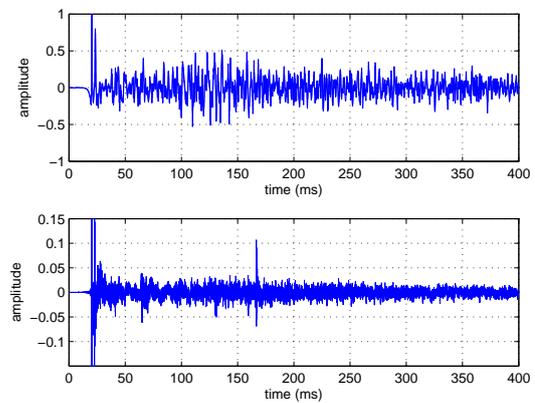


Fig. 19: Recorded Balloon Pop and Loudspeaker-Generated Impulse Response Onsets. The recorded balloon pop onset (upper) shows a number of specular reflections not present in the loudspeaker-generated impulse response (lower) Note the different amplitude scales—both impulse responses are normalized to have unit maximum amplitude.

NED profiles for the impulse response measurement $h(t)$ and balloon pop recording $b(t)$ are shown along with the NED profile of the synthesized pulse sequence $p(t)$ in Fig. 18. The NED profiles of the measured and estimated impulse responses closely track each other, and sit slightly lower than the balloon pop NED until the late-field start, as expected. It's worth examining the onsets of the measured impulse response $h(t)$ and the recorded balloon pop

$b(t)$, shown in Fig. 19. Note that while impulse response $h(t)$ contains a few specular reflections after the late-field onset in the 100–170 ms time interval, the balloon pop recording has a large number of specular reflections in the same time interval. This seems to be a result of the balloon’s uniform radiation pattern generating acoustic paths involving the church dome and floor, which are only weakly excited by the loudspeaker. Their arrivals likely affect the local spectrum, but as they appear in a high echo density background (with NED values near 1.0) their presence does not affect the perceived echo density.

6.2. Hagia Sophia Measurement

Fig. 20 shows a spectrogram of a balloon pop recorded in Hagia Sophia by one of the authors, Pentcheva, in May 2010, with omnidirectional microphones affixed to her hair, just above her ears. A balloon was popped by an assistant, roughly 3 meters in front of Pentcheva. In the spectrogram and balloon response, the direct path arrival and several early reflections are visible, as is a very long decay associated with the voluminous space of 70×80 meters and nearly 50 meters in height [4]. Stray noises made in the building were noted near 3 seconds, 6 seconds, and after 8 seconds, corrupting the recorded balloon response.

The balloon response $b(t)$ was processed using our analysis/synthesis method into an impulse response $\tilde{h}(t)$ and its spectrogram is shown in Fig. 21. As this impulse response was synthesized, neither the unwanted background noises nor the handheld recorder noise floor are present. Fig. 22 shows estimated T_{30} decay times for the partially-filled museum; these are consistent with those published by Gade et. al. [15]. Estimates of C_{80} not presented here are also consistent with that described in [15].

7. SUMMARY

Balloon pop impulse response recordings offer certain advantages over other acoustic measurement methods, and in notable contexts are the only means; however, the pattern of nulls present in the balloon pop waveform “N-wave” Fourier transform creates undesired effects when this response is convolved with other material. Our method for converting recorded balloon pops into full audio bandwidth

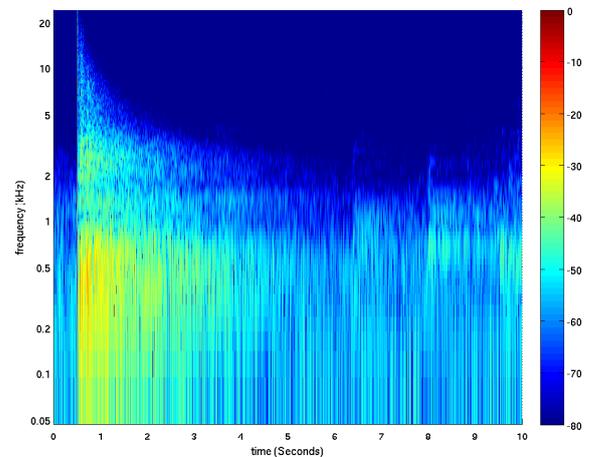


Fig. 20: Hagia Sophia Balloon Pop Spectrogram. The spectrogram of a balloon pop recorded in Hagia Sophia in the presence of background noise and sound events.

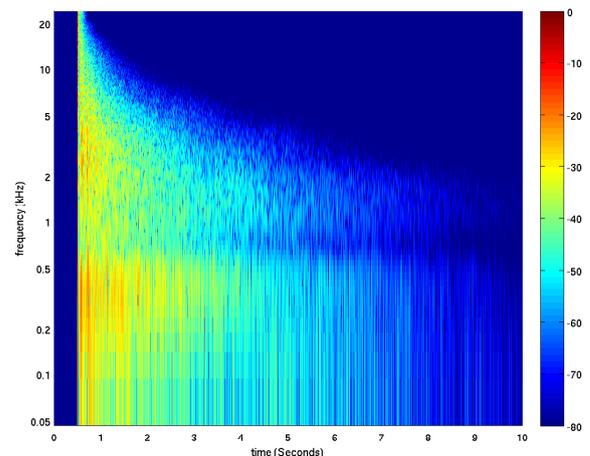


Fig. 21: Hagia Sophia Estimated Impulse Response. The spectrogram of the impulse response estimated from the balloon pop shown in Fig. 20.

impulse responses yields an accurate and clarified room impulse response that can be convolved with sample audio material to create auralizations of the acoustic effect of the measured space, free of distortions from the excitation process.

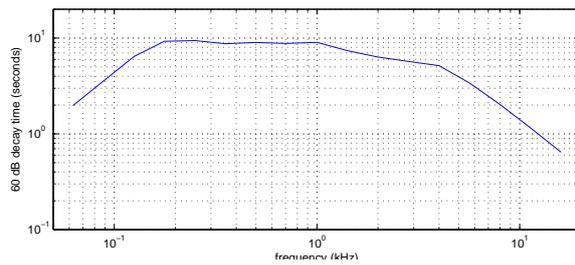


Fig. 22: Hagia Sophia (Partially Occupied) Estimated 60 dB Decay Time.

8. ACKNOWLEDGMENTS

This research was enabled by the Stanford Presidential Fund for Innovation in the Humanities, granted for “Icons of Sound: Architectural Psychoacoustics in Byzantium”, and the Stanford Institute for Creativity and the Arts (SiCa).

9. REFERENCES

- [1] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” Convention Paper 5093. Presented at the 108 Convention, Paris, February 19-22, 2000.
- [2] A. Farina, R. Ayalon. “Recording concert hall acoustics for posterity,” In the Proceedings of the 24th AES Conference on Multichannel Audio, Banff, Canada, June 26-28 2003.
- [3] D. T. Deihl and F. R. Carlson Jr. “N Waves from Bursting Balloons,” American Journal of Physics, Vol. 36, No. 5. May, 1968.
- [4] B. Pentcheva. *Icons of Sound: Hagia Sophia and the Byzantine Choros*. Chapter 2 in “The Sensual Icon: Space, Ritual, and the Senses in Byzantium,” Penn State Press, 2010.
- [5] C. A. Weitze, J. H. Rindel, C. L. Christensen, and A. C. Gade. “The Acoustical History of Hagia Sophia revived through Computer Simulation,” Rome, 2001. Retrieved from <http://www.odeon.dk/pdf/ForumAcousticum2002.pdf>, April 7, 2009.
- [6] C. A. Weitze, C. L. Christensen, J. H. Rindel. “Comparison between In-situ recordings and Auralizations for Mosques and Byzantine Churches,” May 25, 2010. Retrieved from <http://www.odeon.dk/pdf/NAM\%202002\%20paper.pdf>.
- [7] Odeon A/S. “Acoustics in ancient church, Hagia Sofia,” Retrieved from <http://www.odeon.dk/acoustics-ancient-church-hagia-sofia>, April 26, 2010.
- [8] Mark F. Hamilton, David T. Blackstock. *Non-linear Acoustics: Theory and Applications*. Academic Press, San Diego, 1998.
- [9] P. Huang, J. S. Abel, H. Terasawa, and J. Berger. “Reverberation Echo Density Psychoacoustics,” Convention Paper 7583. Presented at the 125th Convention, New York, October 2-5, 2008.
- [10] J. S. Abel and P. Huang. “A simple, robust measure of reverberation echo density,” Convention Paper 6985. Presented at the 121st Convention, San Francisco, October 5-8, 2006.
- [11] P. Huang and J. S. Abel. “Aspects of Reverberation Echo Density,” Convention Paper 7163. Presented at the 123rd Convention, New York, October 5-8, 2007.
- [12] Y. Hur, J. S. Abel, Y. Park, D. H. Youn. “Microphone Array Synthetic Reconfiguration,” Presented at the 127th Audio Engineering Society Convention, New York, October 9-12, 2010.
- [13] P. P. Vaidyanathan. *Multirate systems and filter banks*. Prentice Hall, Englewood Cliffs, 1993.
- [14] N. J. Bryan, J. S. Abel. “Methods For Extending Room Impulse Responses Beyond Their Noise Floor.” Presented at the 129th Audio Engineering Society Convention, San Francisco, November 4-7, 2010.
- [15] A. C. Gade, “Acoustics in Halls for Speech and Music,” *Springer Handbook of Acoustics*, Ed. T. D. Rossing, 301-350 (2007).