Exploiting randomness in acoustic impulse responses to achieve headphone compensation through deconvolution

Ian J. Kelly[a)] and Francis M. Boland
Department of Electronic and Electrical Engineering, School of Engineering, Trinity College Dublin, Dublin 02, Ireland

This paper presents a method of headphone/earphone equalization based upon deconvolution of the headphone impulse response from other acoustic filters in the processing chain. The methods presented are thus applicable to areas such as spatial audio, where input signals are processed with binaural impulse responses. The extraction of low order from higher order acoustic impulse responses is justified based upon an application of the theory pertaining to the clustering of the zeros of random coefficient polynomials about the unit circle. An argument is presented supporting the applicability of this theory to acoustic impulse responses. Such an equalization method guarantees a considerable reduction in computational effort over traditional separate compensating filters. A comparison is presented of the implementation of the equalization using the homomorphic method and least squares method with and without regularization to a traditional separate least squares inverse filter. The proposed approach is shown to have a performance comparable to a separate least squares designed compensator, often giving superior performance despite a large reduction in complexity. © *2013 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4798799]

I. INTRODUCTION

Humans binaural hearing localizes sound sources by the detection of three main auditory cues: The inter-aural level difference (ILD), inter-aural time difference (ITD), and spectral coloration of the incoming audio by the pinna, head, and torso. For a given source location, left/right pairs of head-related transfer functions (HRTFs) encode all three pieces of information and can be used to create a virtual auditory space (VAS). In personal audio systems headphones are commonly used to deliver the VAS, and studies[1] have shown that spectral coloration and phase distortion, caused by the headphone transfer function (HpTF), can distort the audio reaching the listener's eardrums, damaging the subtle temporal and spectral cues contained within a HRTF. Equalization of the HpTF should compensate for this distortion, improving the listener's experience of spatialized audio.

Signals produced by convolving a source signal with the head-related impulse response (HRIR), the time domain equivalent of the HRTF, need to be reproduced at the entrance to the ear canal, which is where the HRIR has been recorded and not at the headphone/earphone drivers. Indeed, HRIRs are commonly recorded with microphones placed at the entrance to the ear canal with *blocked meatus*. The assumption is that the signal is presented to the open ear canal via headphones. Hammershøi and Møller[2] state that the sound transmission from the entrance of the ear canal to the eardrum depends only on the ear canal and the eardrum regardless of the conditions outside the ear canal's entrance.

However, the transfer function of the headphone or earphone set comprises not only the transducer itself but also the transfer function of the coupling from the transducer to the listener's ear canal.[3] The variations in the frequency responses of different sets of headphones have been shown[4,5] to be as large as the variations in the HRTFs themselves, clearly demonstrating the need for headphone equalization.

Most mobile audio device users do not use over-the-ear headphones and tend to favor in-the-ear, "*bud-type*" earphones which are inserted into the ear canal, thus blocking its open end. Here again there is a transfer function that introduces distortion and needs to be equalized for the presentation of spatialized audio. Earphone responses are more generic due to the fact that their placement is less variable; therefore a single set of earphone-compensated HRIRs can be used across many sessions. Generic compensations for both headphones and earphones have been shown by Brinkmann and Lindau[6] to promise improvements regarding maintaining spectral cues, particularly at lower frequencies.

In this paper a novel method for equalizing headphones is presented. The method is intended for use in spatial audio applications where audio processing with binaural impulse responses is performed. The equalization is performed by extracting the headphone/earphone response function from the binaural responses using deconvolution or equivalently approximate factorization.

II. PROBLEM FORMULATION

Taking the input signal to one channel of a binaural system to be $s[n]$ and the applied HRIR to be $h^\phi[n]$, one can represent the signal presented to the listeners' ear as $s_f[n]$ which is given by the convolution

[a)]Author to whom correspondence should be addressed. Electronic mail: kellyij@tcd.ie

$$s_f[n] = f[n] * (h^\phi[n] * s[n]), \tag{1}$$

where $f[n]$ represents the HpIR to be compensated for and $h^\phi[n]$ is one HRIR from a set where $\phi \in \{1, 2, ..., M\}$, with each ϕ representing a discrete point on a sphere about the listeners head. Traditionally headphone/earphone equalization attempts to design a compensator $c[n] \approx f^{-1}[n]$, such that $c[n] * f[n] \approx \delta[n]$.

However when one considers that there is a spatialization filter in the chain, $h^\phi[n]$ the HRIR, along with $f[n]$ the HpIR, the problem can be reformulated as a deconvolution or factorization. That is, $f[n]$ can be extracted, as a common factor from each HRIR, leaving a set of residuals $g^\phi[n]$. Then the HRIRs can be expressed as

$$h^\phi[n] = f[n] * g^\phi[n], \tag{2}$$

where $g^\phi[n]$ is the residual component after the HpIR is taken out as a factor and $\phi \in \{1, 2, ..., M\}$. This filter, $g^\phi[n]$, is used in place of the HRIR $h^\phi[n]$, thus removing the need for a separate inverse filter to compensate for the headphone effects.

Such a factorization approach, similar to that used by Masterson et al.[7] to simplify HRIRs, takes advantage of the fact that the HRIR and HpIR will most likely share some common, or very nearly matching zeros, as both $h^\phi[n]$ and $f[n]$, being "random-like" sequences, will have zeros clustered about and nearly uniformly around the unit circle, as is the tendency for all random polynomials.[8] Section III explores the claim that $h^\phi[n]$ and $f[n]$ can be considered as random-like sequences. In order to be able to speak in terms of polynomials, we move to the z-domain. In this case each sample in an impulse response such as a HRIR becomes the coefficient of a z^{-n} term.

III. RELEVANCE OF ROOT CLUSTERING IN ACOUSTIC SIGNALS TO FACTORIZATION

The clustering of the roots of random polynomials uniformly about the unit circle has long been an area of much interest in mathematics circles. The early celebrated work by Erdős and Turan[9] and the later work of Shepp and Vanderbei[10] proved that the zeros of infinite order random polynomials would cluster uniformly near the unit circle for a small set of probability distributions of the polynomial's coefficients, such as when the coefficients were independent normal random variables. These results have, however, been extended in recent years to include more probability distributions and to include random polynomials of finite order for which the same uniform clustering result holds. For example, Hughes and Nikeghbali require no independence restriction on their random variables.[8] They assume only that the first and last polynomial coefficients are never zero.

It is therefore interesting to examine the statistical distribution of the coefficients of HRIRs and HpIRs and the z-plane distribution of the zeros of the equivalent transfer functions. In Fig. 1 the root constellations of 10 headphone response functions, each 256 samples long, are plotted. Clustering about the unit circle can clearly be seen. Looking

at the distribution of the filter coefficients from 19 074 HRIRs taken from the IRCAM Listen HRIR data set which can be found in Ref. 11, it is clear that the coefficients of HRIRs such as these closely match a skewed α-stable distribution. A similar distribution was observed looking at the 112 500 HRIRs in the CIPIC data set[12] but with a far greater number of exact zero valued coefficients due to the unnecessarily long sets of leading zeros and tails. Such a distribution was also found to be typical of the coefficients of the HpIRs measured by the authors. The α-stable distribution generalizes the classical central limit theorem for sequences of random variables which may not be independent and identically distributed. Figure 2 shows a histogram of the coefficients from 19 074 HRIRs from the IRCAM Listen data set. This amounts to 9 765 888 coefficients, allowing for a good estimation of their overall distribution. Superimposed over the histogram is an approximately fitted α-stable pdf which has been scaled to match the number of entries in the histogram. The approximate distribution is parameterized via its characteristic function[13]

$$\psi(t) = \mathbf{E}\left[e^{-\gamma^\alpha |t|^\alpha (1 - j\beta \, \text{sign}(t) \tan(\pi\alpha/2)) + j\delta t}\right]. \tag{3}$$

Here the four parameters α, β, γ, and δ represent the characteristic exponent describing the tail, the skewness, the scale, and the location, respectively. The parameters used to generate Fig. 2 were $\alpha = 1.2379$, $\beta = -0.2666$, $\gamma = 0.0039$, and $\delta = -0.0013$.

One could raise the question as to whether sequences such as $h^\phi[n]$ and $f[n]$ are of sufficient length to have a uniform distribution of zeros about the unit circle. According to the celebrated Erdős-Turan result and the extension by Hughes and Nikeghbali,[8] for a polynomial defined as

$$P(z) = \sum_{k=0}^{N} a_k z^k, \tag{4}$$

where coefficients a_k are randomly distributed, the roots of the random polynomial will cluster uniformly around the unit circle if $L_N(P)$ is small compared to the polynomial order N where
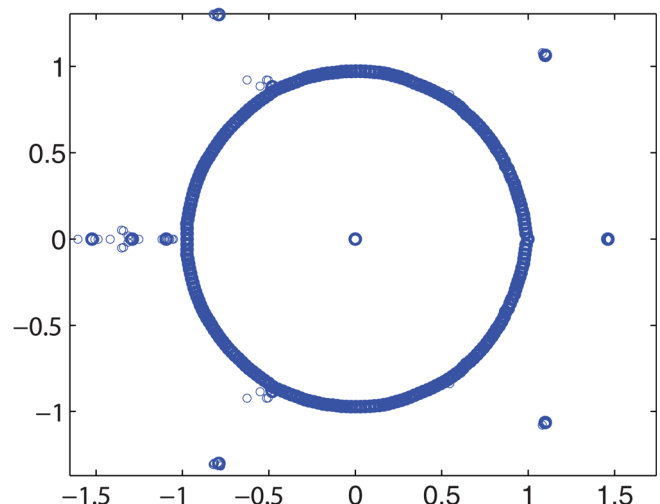

FIG. 1. (Color online) Zero distribution of ten actual headphone responses.
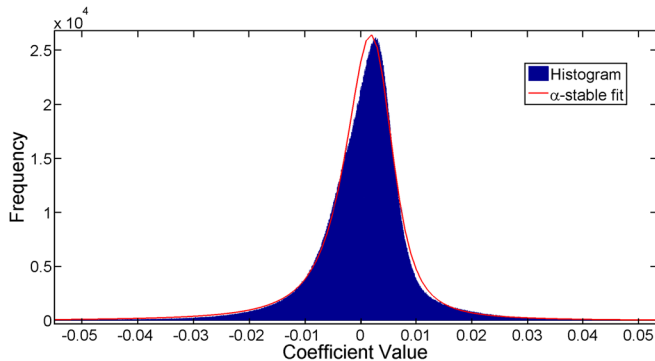
FIG. 2. (Color online) Histogram of the coefficient values for 19 074 IRCAM HRIRs. The histogram closely resembles a skewed α-stable distribution. A best fit α-stable distribution is also shown (line) with $\alpha = 1.2379$, $\beta = -0.2666$, $\gamma = 0.0039$, and $\delta = -0.0013$.

$$L_N(P) = \log \sum_{k=0}^{N} |a_k| - \frac{1}{2} \log|a_0| - \frac{1}{2} \log|a_N|, \tag{5}$$

provided a_0 and a_N are non-zero.

Taking for example HRIRs from the IRCAM data set and headphone/earphone transfer functions recorded by the authors where the impulse response lengths averaged at just 512 samples each, values of $L_N(P)$ tended to be much smaller than values for polynomial degrees N. In fact, examining the 19 074 responses from the IRCAM data set, where $N = 512$, the mean value for $L_N(P)$ was 7.2346, the median was 7.1528, and the mode was just 5.3000 with maxima and minima 11.8950 and 5.3000, respectively.

According to Hughes and Nikeghbali, the expression

$$\left(1 - \frac{v_N}{N}\right) \leq \frac{2L_N(P)}{N\rho}, \tag{6}$$

holds true for $0 < \rho < 1$. Here

$$v_N = \#\left\{z_k : 1 - \rho \leq |z_k| \leq \frac{1}{1 - \rho}\right\} \tag{7}$$

is the number of zeros lying in the annulus bounded by $1 - \rho$ and $1/(1 - \rho)$ with z_k denoting the zeros of polynomial $P(z)$. Rearranging expression (6) yields

$$v_N \geq N - 2\frac{L_N(P)}{\rho}. \tag{8}$$

This implies that for the responses examined, the number of roots falling within an annulus around the radius parameterized by $\rho = 0.2$ would always be greater than or equal to 442 for HRIRs of length $N = 512$, i.e., 86.5% of the roots.

On calculating the actual distribution of the roots of these HRIRs, however, the roots are seen to cluster even more densely. The roots of 2325 randomly chosen IRCAM HRIRs were calculated. The proportion of those with a magnitude within an annulus of $\rho = 0.2$ was then calculated; it amounted to 1 176 955 out of 1 188 075 roots or 99.06% of the roots.

Similarly, let

$$v_{\theta\phi} = \#\{z_k : \theta \leq \arg(z_k) < \phi\} \tag{9}$$

be the number of zeros of polynomial $P(z)$ whose argument lies between θ and ϕ where $0 \leq \theta < \phi \leq 2\pi$. According to Hughes and Nikeghbali,[8]

$$\left|\frac{1}{N}v_{\theta\phi} - \frac{\phi - \theta}{2\pi}\right|^2 \leq \frac{C}{N}L_N(P), \tag{10}$$

where C is some constant. This can be reformulated as a quadratic inequality

$$v_{\theta\phi}^2 - \frac{N(\phi - \theta)}{\pi}v_{\theta\phi} + \frac{N^2(\phi^2 - 2\phi\theta + \theta^2)}{4\pi^2} - NCL_N(P) \leq 0. \tag{11}$$

It was thus found that for zeros lying between 10° and 110°, this held true for a value of $C = 5.1957 \times 10^{-9}$. In fact for the 2325 IRCAM HRIRs studied, the average number of zeros lying in each 1° segment was 1.4179. This fits very well with the ideal Erdős and Turan result which states that

$$\lim_{N\to\infty} \mathbf{E}\left[\frac{1}{N}v_{\theta\phi}\right] = \frac{\phi - \theta}{2\pi} \tag{12}$$

as

$$N\frac{\phi - \theta}{2\pi} = 1.4222, \tag{13}$$

for $N = 512$. Looking at the histogram of the root angles, one can see immediately how evenly the angles of the roots of the 2325 HRIRs studied are distributed around the unit circle. Figure 3 demonstrates such a distribution for the 2325 IRCAM HRIRs examined. The peaks at 0° and 180° are due to the real zeros of the polynomial located on the real axis.

The reason behind finding 99.06% of the roots clustered within the given annulus, which far exceeds the 86.5% calculated via Hughes and Nikeghbali's results, can be found by examining Eq. (5). It was stated that when $L_N(P)$ is small in relation to the order N of $P(z)$, then the zeros of $P(z)$, a random polynomial, will cluster uniformly about the unit circle. However this depends heavily on the magnitudes of the first and last coefficients of $P(z)$, a_0, and a_N. If either or both of these coefficients are close to zero $L_N(P)$ will grow larger. However it can be shown that despite the fact that HRIRs are likely to have very low magnitude first and last
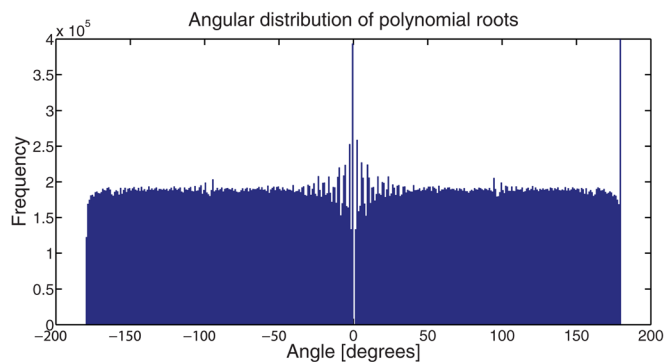

FIG. 3. (Color online) Histogram of the angular distribution of the zeros of the 2325 HRIRs from the IRCAM data set.

coefficients, their roots will cluster no less closely about the unit circle.

This is down to the reasons why HRIRs have low magnitude initial and final coefficients. HRIRs in general have some onset delay which is closely related to the ITD between left and right HRIR pairs. These onset delays are generally affected by measurement noise, transducer transfer functions, and perhaps some dc offset. This means that the delay is seen as a set of low magnitude initial coefficients as opposed to a pure delay. This accounts for a low magnitude a_0 pushing up the value of $L_N(P)$. However, the behavior of the roots is not greatly affected as the approximate delay can be seen as a convolution with something approximating a time delayed Kronecker delta. This short polynomial denoted $\widetilde{\delta}_R(z)$ of order $R \ll N$ will simply add an asymptotically negligible number of zeros to the roots of a HRIR despite increasing the value of $L_N(P)$.

This can easily be demonstrated with a HRIR with 22 samples of approximate onset delay stripped off as in Fig. 4. The lower signal in Fig. 4 was then deconvolved from the upper signal yielding a signal closely approximating a delayed Kronecker delta. The roots of this signal can be seen to form a ring outside the unit circle in Fig. 5.

A similar argument can be made for the low magnitude of the last coefficient of a HRIR, a_N. HRIRs exhibit a statistical decaying magnitude similar to an exponential decay. It was proposed by Steiglitz and Dickinson[14] that the decay in HRIR coefficients could be modeled as infinite impulse response (IIR) filtering, adding only a fixed (asymptotically negligible) number of zeros and poles to the output's z-transform. This means one can still use the white noise result on the distribution of zeros. A windowed impulse response giving an approximate finite impulse response (FIR) filter would add just a fixed set of zeros to the Argand plane. However the authors believe that such a model of the exponential decay is unsuitable. The correlation that the IIR nature of such a filter would induce on the coefficients of the response would mean that their probability distribution would be drastically changed with a removal of independence. This behavior renders Steiglitz and Dickinson's model unsuitable. Furthermore, such IIR based models do not have frequency responses which behave in a similar way as those of acoustic impulse responses. The authors instead model the decay of acoustic signals like HRIRs as a point-wise multiplication with a decaying exponential
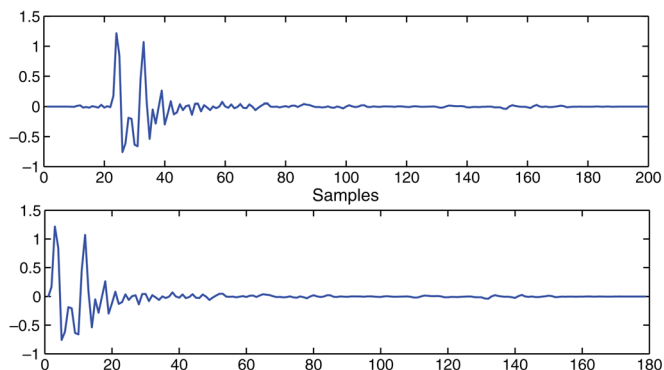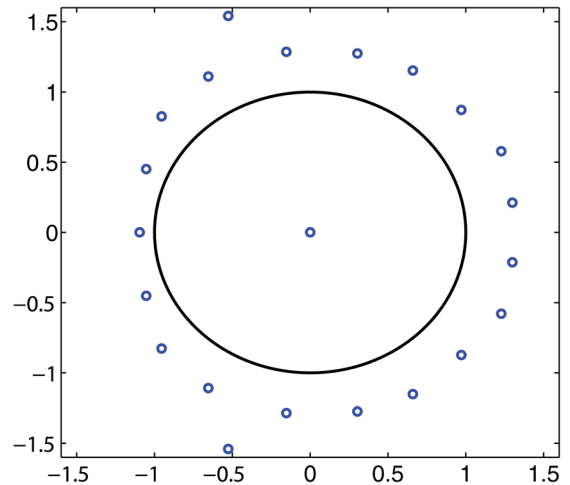

FIG. 5. (Color online) Roots of the approximate delay polynomial.

$$\mathcal{Z}\{e^{-\alpha n}p[n]\} = P(e^{-\alpha}z), \tag{14}$$

where $\alpha \ll 1$. This means that the expected value of the polynomial coefficients in the ensemble also have an exponential decay. This has the effect of reducing the expected value of the magnitude of the polynomials roots by a uniform factor, $e^{-\alpha}$, without altering their angular location. Thus the roots remain just as closely clustered, but about a ring just within the unit circle and not about the unit circle itself. The predominantly minimum phase nature of HRIR roots closely matches this model. The exponential decay thus results in a low magnitude final coefficient a_N, increasing the value of $L_N(P)$ without disrupting the clustering behavior of the polynomial roots.

IV. DECONVOLUTION METHODS

This section describes the use of homomorphic deconvolution and least squares factorization with and without regularization to extract HpIR from a set of HRIRs.

A. Equalization using homomorphic deconvolution

Homomorphic signal processing for deconvolution is a long established method and will be summarized.

Let $\hat{h}^\phi[n]$ denote the complex cepstrum of a HRIR, $h^\phi[n]$. It can be defined as

$$\hat{h}^\phi[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi}(\log|H^\phi(e^{j\omega})| + j\angle H^\phi(e^{j\omega}))e^{j\omega n}d\omega, \tag{15}$$

where $H^\phi(e^{j\omega})$ denotes the Fourier transform of $h^\phi[n]$.

If $z = re^{j\theta}, r > 0$ then $w = \ln r + j\theta$ is one logarithm of z; adding multiples of $2\pi j$ will give all of the others.

To unwrap the phase in Eq. (15), an additional 2π radians of phase is added on wherever the wrapped phase spectrum moves from π to $-\pi$. This is performed by simply applying a set number of samples of circular delay to $h^\phi[n]$.

Before phase unwrapping, excess phase arising from any pure delay is also removed. This excess phase is then added back into the output signal during the inverse complex cepstral procedure.

FIG. 4. (Color online) A HRIR (above) and that same HRIR with approximate onset delay removed (below).

The complex cepstrum of $f[n]$, a HpIR, denoted $\hat{f}[n]$, is calculated in the same way as that of $h^\phi[n]$.

A convolution in the time domain can be replaced by a point-wise summation in the quefrency domain. Therefore, a deconvolution should be achievable by subtracting the complex cepstrum of the HpIR from that of the HRIR. Hence to obtain the filter $g^\phi[n]$ from $h^\phi[n] \approx f[n] * g^\phi[n]$, we must simply take the inverse complex cepstrum of $\hat{h}^\phi[n] - \hat{f}[n]$.

When reinserting the excess phase after deconvolution has been performed an integer value of $-N - (|nh| - |nf|)$ can be used to maintain the differences between the HRIRs in terms of initial delay, once recombination of $g^\phi[n]$ and $f[n]$ has taken place. nh and nf denote the amount of excess phase arising from pure delay removed from $h^\phi[n]$ and $f[n]$, respectively, in terms of the number of samples of circular shift applied. N represents an additional shift, uniform across all the $g^\phi[n]$'s. The addition of such a *global delay* will not affect the ITDs which will be maintained as the difference between nh and nf. Instead it will simply induce an identical group delay in both ears. Issues with latency will not arise as long as the delay is below that detectable by humans. This extra shift ensures the $g^\phi[n]$ being recovered will be causal. For HpIR cepstral lengths of 1024 taps the authors found that introducing a delay of roughly 100 samples was more than sufficient, which at 44.1 kHz amounted to just 2.27 ms of delay. The reliable detection of audio-visual sync errors falls between 45 ms for audio leading video and 125 ms for audio lagging behind video. Thus the lag introduced here is negligible.[15]

B. Least squares deconvolution

This section describes the use of least squares factorization as a deconvolution method which could also be used to pre-deconvolve a HpIR from a set of HRIRs with a Tikhonov regularization term based upon a first difference operator.

Changing to vector-matrix notation

$$h^\phi[n] = f[n] * g^\phi[n]\mathbf{h}^\phi = \mathbf{F}\mathbf{g}^\phi. \tag{16}$$

The cost function to be minimized here is

$$\min_{\mathbf{g}^\phi} J = \|\mathbf{h}^\phi - (\mathbf{F}\mathbf{g}^\phi)\|^2, \tag{17}$$

yielding an approximate solution $\mathbf{g}^{\phi\star}$ where

$$\mathbf{g}^{\phi\star} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{h}^\phi. \tag{18}$$

However as will be shown in Sec. V, this deconvolution is poor and the filter produced $g^\phi[n]$ is often highly noisy.

Improvements can be achieved by adding an extra Tikhonov regularization term to the cost function introducing a degree of smoothness.

$$\min_{\mathbf{g}^\phi} J = \|\mathbf{h}^\phi - (\mathbf{F}\mathbf{g}^\phi)\|^2 + \lambda \parallel \phi(\mathbf{g}^\phi)\parallel^2. \tag{19}$$

Here $\phi(\bullet)$ is a measure of smoothness of $g^\phi[n]$ and λ is a weighting coefficient. An appropriate measure of smoothness is the first difference where

$$\phi(\mathbf{g}^\phi) = \nabla\mathbf{g}^\phi. \tag{20}$$

The approximate closed form solution to this problem is clearly given as

$$\mathbf{g}^{\phi\star} = ((\mathbf{F}^T\mathbf{F}) + \lambda(\nabla^T\nabla))^{-1}\mathbf{F}^T\mathbf{h}^\phi. \tag{21}$$

As one would expect with increasing λ, this extra term will penalize any high frequency components in $\mathbf{g}^{\phi\star}$. However, if λ is too low the pseudo inverse will be practically identical to the unregulated case with misbehavior in low frequency regions.

In order to choose a suitable λ the following recursive method, based upon a simple line search, is proposed.

(1) Initialize, $\lambda = 0$.
(2) The amplitude spectrum is calculated for each reconstructed HRIR, i.e., $f[n] * g^\phi[n]$. The amplitude spectrum of $h^\phi[n]$ is also calculated and the error is taken as

$$\epsilon = \sum_{i=0}^{K} \||H_i^\phi(e^{j\omega})| - |F_i(e^{j\omega})G_i^\phi(e^{j\omega})|\|_{l1}, \tag{22}$$

i.e., the sum of the L1-norms for spectral error across all of the K HRIRs. However these spectra are only calculated between 0 and $\approx$18 kHz as these frequencies are deemed to be more important to spatial perception. Thus errors in the parts of the spectrum having no effect on spatial perception are not penalized.

(3) A simple line search is performed on λ so as to minimize ϵ with respect to λ. On each iteration a value of α is added to λ. As λ approaches a suitable value, ϵ should decrease. If ϵ increases, α is set to $-\alpha/2$. The authors found that a good initial value for α was approximately 10^{-7}, although any relatively small number is clearly suitable.

(4) After several iterations λ should converge on a value that, although not optimum for any single HRIR, will be optimum across the data set, suppressing errors below $\approx$18 kHz.

Although this algorithm is heuristic, it was found, in this case, to achieve regularization factors, λ roughly equal to those found by L-curve curvature maximization,[16] with far less computational effort. These values of λ were found via a method rewarding good spectral behavior in perceptually relevant frequency bands unlike L-curve based methods which can only improve performance in a conditioning sense.

V. EXPERIMENTATION AND RESULTS

A. Experimentation

In this section the outcomes of the re-convolution of the filters generated, $g^\phi[n]$, with the HpIR, $f[n]$, are examined and compared to the original HRIRs, $h^\phi[n]$. A comparison is also made to the use of a traditional separate inverse filter, designed using the least squares method. An optimal delay is chosen to guarantee a causal inverse filter and also to maintain ITD. The separate compensation filters designed were of the same order as the HRIRs (512 samples). The novel approach to headphone compensation introduced in Sec. IV

I. J. Kelly and F. M. Boland: Headphone compensation through deconvolution

was examined in a series of investigations conducted using the G.R.A.S. Type 43AG ear and cheek simulator (G.R.A.S. sound & vibration, Holte, Denmark) in order to look at the spectral errors remaining after headphone compensation. These mean spectral errors over bands 4 barks wide were used to give an accurate estimation of the ILD performance across the frequency spectrum.

Experiments were also performed in order to examine the effect of the proposed approximate factorization based headphone compensation method on ITD. This was done using a Neumann KU100 dummy head. The headphones used were Sennheiser HD650's. These headphones are open backed and thus may introduce some acoustic crosstalk. Therefore it was deemed interesting to see if the compensation methods could improve the accuracy of the ITDs presented to the dummy's ear canals.

Finally the effects of the HpIRs/compensating filters on the preservation of the location, level, and sharpness of a distinctive set of troughs in the resulting HRTFs were examined over an elevation range of −30° to 40°, with the azimuth angle remaining at 0°. This set of troughs were shown by Shaw and Teranishi[17] to be an important spectral cue for source elevation between −45° and 45°.

The magnitude responses of the signals seen in the ear canals of the G.R.A.S. simulator were examined in 6 bands each 4 barks wide, i.e., 20 to 400 Hz, 400 to 920 Hz, 920 to 1720 Hz, 1720 to 3150 Hz, 3150 to 6400 Hz, and 6400 to 15 500 Hz. The responses examined were chosen not to exhibit a case in which the deconvolution based compensation methods perform best, but instead to show a general case, typical of the entire data set.

The reconstructed HRTFs used to measure the effects on ILD were generated by playing a set of filtered logarithmic sweeps to the G.R.A.S. ear and cheek simulator. These sweeps were filtered with:

(1) A HRIR and a separate least squares designed compensating filter $s_1[n] = c[n] * (h^\phi[n] * s[n])$;
(2) A factorized HRIR with the headphone impulse response removed from the set as a common factor via homomorphic deconvolution $s_2[n] = g_{HD}^\phi[n] * s[n]$;
(3) A factorized HRIR with the headphone impulse response removed from the set as a common factor via regularized least squares approximate factorization $s_3[n] = g_{RLS}^\phi[n] * s[n]$;
(4) A factorized HRIR with the headphone impulse response removed from the set as a common factor via un-regularized least squares approximate factorization $s_4[n] = g_{LS}^\phi[n] * s[n]$,
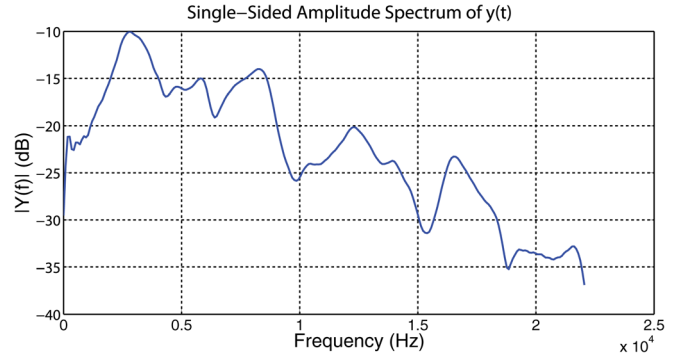

FIG. 6. (Color online) Single sided magnitude spectrum of a generic HpIRs clearly showing the need for headphone/earphone compensation. This magnitude spectrum was derived from the response of a Phillips SHE9700/00 type in-ear earphone.

where $s[n]$ denotes a 20 s logarithmic sine sweep.

This gives a clear measure of the effect of the combined HpIRs and compensating filters upon ILD across the audible spectrum.

For reference all results are compared to the case where a separate inverse filter was used to compensate for the HpTF. This filter was designed via a least squares method to be of the same length as the original HpIR $f[n]$.

The effects on ITD were tested using a separate set of 20 s logarithmic sine sweeps which were pre-convolved with:

(1) A pair of corresponding left and right HRIRS $s_{L1}[n] = h_L^\phi[n] * s[n]$ and $s_{R1}[n] = h_R^\phi[n] * s[n]$;
(2) A pair of HRIRs and separate least squares designed compensating filters $s_{L2}[n] = c_L[n] * (h_L^\phi[n] * s[n])$ and $s_{R2}[n] = c_R[n] * (h_R^\phi[n] * s[n])$;
(3) A pair of factorized HRIRs with the headphone impulse responses removed from the set as a common factor via homomorphic deconvolution $s_{L3}[n] = g_{HD,L}^\phi[n] * s[n]$ and $s_{R3}[n] = g_{HD,R}^\phi[n] * s[n]$;
(4) A pair of factorized HRIRs with the headphone impulse responses removed from the set as a common factor via regularized least squares approximate factorization $s_{L4}[n] = g_{RLS,L}^{\phi\star}[n] * s[n]$ and $s_{R4}[n] = g_{RLS,R}^{\phi\star}[n] * s[n]$,

where $s[n]$ denotes a 20 s logarithmic sine sweep.

Corresponding responses, $d_{L1}^\phi[n]$, $d_{R1}^\phi[n]$, $d_{L2}^\phi[n]$, $d_{R2}^\phi[n]$, $d_{L3}^\phi[n]$, $d_{R3}^\phi[n]$, $d_{L4}^\phi[n]$, and $d_{R4}^\phi[n]$ were recorded for each of the above cases with HRIR/factorized HRIR pairs corresponding to 0° elevation and $\phi \in [0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°]$ azimuth. For this experiment un-regularized least squares factorized HRIRs were not used as they had previously been shown to give poor performance compared to the other compensation methods.

TABLE I. Mean absolute error for the magnitude spectra after each earphone compensation method, in different frequency bands, median value over 50 HRIRs—Phillips SHE9700/00.

Compensation type	20 to 400 Hz	400 to 920 Hz	920 to 1720 Hz	1720 to 3150 Hz	3150 to 6400 Hz	6400 to 15 500 Hz
Separate filter	0.70 dB	0.40 dB	0.17 dB	0.06 dB	0.03 dB	0.04 dB
Homomorphic	0.35 dB	0.33 dB	0.24 dB	0.15 dB	0.13 dB	0.15 dB
Regularized least-squares	0.98 dB	0.60 dB	0.73 dB	0.50 dB	0.39 dB	1.15 dB
Least-squares	1.03 dB	0.64 dB	0.77 dB	0.54 dB	0.54 dB	1.30 dB

TABLE II. Mean absolute error for the phase spectra after each earphone compensation method in different frequency bands, median value over 50 HRIRs—Phillips SHE9700/00.

Compensation type	20 to 400 Hz	400 to 920 Hz	920 to 1720 Hz	1720 to 3150 Hz	3150 to 6400 Hz	6400 to 15 500 Hz
Separate filter	0.09°	0.04°	0.02°	0.004°	0.004°	0.005°
Homomorphic	0.05°	0.03°	0.03°	0.01°	0.01°	0.02°
Regularized least-squares	0.16°	0.17°	0.12°	0.16°	0.05°	0.15°
Least-squares	0.16°	0.17°	0.12°	0.17°	0.05°	0.17°

Both ILD and ITD are only useful spatialization cues when one wishes to ascertain the azimuth angle of a sound source relative to the listener. Elevation cues are predominantly spectral in nature. One such important cue, recognized by Shaw and Teranishi,[17] was a trough in the left and right HRTFs that moved from approximately 6 to 10 kHz as the sound source moved from −45° to 45° in elevation. Tables VI–VIII demonstrate the effect of the equalized responses on the location, width, and level of this trough. The set of HRIRs used range from −30° to 40° elevation at 0° azimuth and were taken from the HRTF data set provided by Gardner and Martin.[18]

B. Results

Figure 6 shows a representative example of one of the HpTFs measured via the G.R.A.S. Type 43AG ear and cheek simulator. The need for compensation is clear as the magnitude response is far from flat and introduces a pair of troughs at approximately 9 and 16 kHz which have the potential to strongly distort the elevation cues in a given HRTF.

The values shown in Tables I, II, III and IV represent the mean absolute errors in magnitude and phase on a set of 50 HRIRs chosen from the IRCAM data set. These are presented over 6 distinct frequency bands, namely, 20 to 400 Hz, 400 to 920 Hz, 920 to 1720 Hz, 1720 to 3150 Hz, 3150 to 6400 Hz, and 6400 to 15 500 Hz. These bands are chosen so as to be 4 barks wide each. Higher frequencies than this are deemed to be of less importance with regard to spatialization. The median of these errors is presented so that results are not skewed by outliers. In the case of the separate inverse filter the error was obviously identical across all 50 HRTFs as it is dependent only on the HpTF being compensated.

The mean absolute error in the magnitude spectrum provides a good indication of the preservation of ILD in each of the frequency bands explored.

The HpIRs used were measured by the authors using a pair of Sennheiser HD650 headphones for Tables III and IV.

Meanwhile a pair of Phillips SHE9700/00 in-ear earphones were used to record the responses for Tables I and II.

As can be seen from Tables I, II, III and IV approximate-factorization/deconvolution allows for headphone/earphone compensation which, on average, maintains ILD as well as a separate inverse filter across all of the frequency bands examined when performing homomorphic deconvolution, but more efficiently as this method does not introduce any additional processing.

Figure 7 shows a comparison of the performance of the proposed headphone/earphone algorithms for all 50 HRIRs tested, in the 200 Hz to 2 kHz frequency band in terms of mean absolute error in the magnitude spectrum. Figure 8 shows the mean absolute error of the phase spectra for the same frequency band. This band was chosen as an example as its phase behavior is of importance in audio spatialization. Furthermore the behavior of each of the methods is largely similar across all the bands referred to in Tables I, II, III and IV.

The error in ITD between the left and right responses recorded using the Neumann KU100 dummy head and the Sennheiser HD650 headphones are listed in Table V below. The ITDs were calculated using the well-known inter-aural cross correlation method. The errors shown for the case where there is no compensation provided clearly demonstrate the damaging effects the presence of HpIRs can have on ITD. Each of the compensation methods explored reduce all errors to one sample or less.

As can be clearly seen from Table VI the presence of a HpIR in the filter chain has the potential to significantly alter the location of notches in the signal presented at the listeners ear. This is especially true at higher frequencies where the HpTF itself has narrow notches. In terms of leaving the notch locations unaltered in frequency, the separate inverse filter shows the best performance with no alteration to notch locations relative to the original HRTFs. The homomorphic deconvolution based compensation performs next best with a mean error in notch location of just 24.5 Hz. At notch

TABLE III. Mean absolute error for the magnitude spectra after each headphone compensation method in different frequency bands, median value over 50 HRIRs—Sennheiser HD650.

Compensation type	20 to 400 Hz	400 to 920 Hz	920 to 1720 Hz	1720 to 3150 Hz	3150 to 6400 Hz	6400 to 15 500 Hz
Separate filter	1.57 dB	0.25 dB	0.11 dB	0.06 dB	0.15 dB	0.07 dB
Homomorphic	0.66 dB	0.23 dB	0.16 dB	0.12 dB	0.08 dB	0.10 dB
Regularized least-squares	1.39 dB	0.74 dB	0.85 dB	0.75 dB	1.12 dB	2.49 dB
Least-squares	1.39 dB	0.75 dB	0.86 dB	0.76 dB	1.12 dB	2.52 dB

I. J. Kelly and F. M. Boland: Headphone compensation through deconvolution

TABLE IV. Mean absolute error for the phase spectra after each headphone compensation method in different frequency bands, median value over 50 HRIRs—Sennheiser HD650.

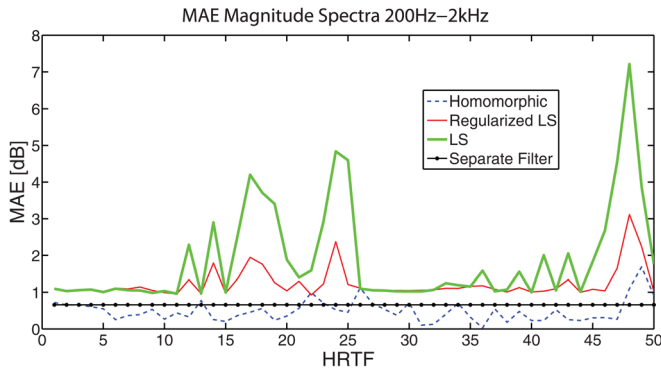| Compensation type	20 to 400 Hz	400 to 920 Hz	920 to 1720 Hz	1720 to 3150 Hz	3150 to 6400 Hz	6400 to 15 500 Hz
Separate filter	0.20°	0.02°	0.01°	0.01°	0.02°	0.01°
Homomorphic	0.08°	0.03°	0.02°	0.01°	0.01°	0.01°
Regularized least-squares	0.08°	0.05°	0.07°	0.08°	0.09°	0.42°
Least-squares	0.08°	0.05°	0.07°	0.08°	0.09°	0.43°

FIG. 7. (Color online) Plot of the mean absolute error in the magnitude spectra of 50 HRTFs after headphone/earphone compensation for the deconvolution/factorization methods described above and for a separate inverse filter in the 200 Hz to 2 kHz region of the spectrum.
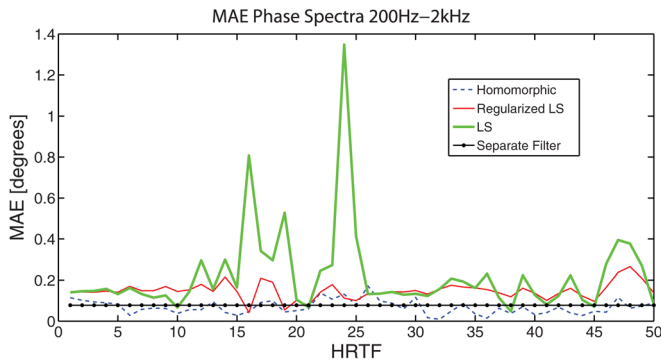

FIG. 8. (Color online) Plot of the mean absolute error in the phase spectra of 50 HRTFs after headphone/earphone compensation for the deconvolution/factorization methods described above and for a separate inverse filter in the 200 Hz to 2 kHz region of the spectrum.

TABLE V. Error in ITD. Samples at 44.1 kHz are 22.68 μs.

Azimuth angle	No compensation (μs)	Homomorphic deconvolution (μs)	Least-squares deconvolution (μs)	Separate filter (μs)
0°	22.68	−22.68	−22.68	−22.68
45°	45.35	0	0	0
90°	68.03	0	0	0
135°	22.68	−22.68	−22.68	−22.68
180°	22.68	0	0	0
225°	45.35	0	0	0
270°	45.35	0	0	0
315°	22.68	0	0	0

frequencies in the region of 6.5 to 10.8 kHz such errors are relatively small.

When looking at the 3 dB width of the same notches in Table VII it can be seen that at higher frequencies the presence of headphones can significantly alter the width of the given notches. This is due to the fact that the HpTFs contain significant troughs at these frequencies in themselves. Table VIII also demonstrates that the troughs present in the HpTFs can have a large effect on the notch magnitude at time resulting in differences of almost 8 dB. This is predominantly due to a high-pass type behavior seen in headphone responses due to leakage from the ear-cup.

VI. DISCUSSION

The arguments made in Sec. III also help to explain several issues alluded to in the approximate factorization technique of Masterson *et al.* developed to extract a common, directionally independent component from a full set of HRIRs.[19] This is formulated as a minimization problem. Masterson *et al.* found that their factorization problem appeared to have a large number of equivalent solutions. This can easily be explained based upon the clustering of the zeros of HRIRs tightly about the unit circle. Since all of the HRIRs will have a great number of shared or nearly shared roots there will be more than one approximately common component of order $M \ll N$, where N is the order of the original HRIRs. This is the same phenomenon taken advantage of for use in the deconvolution based headphone equalizations in this paper.

In the case where the HpTF $f[n]$ or the HRTF $h^\phi[n]$ have deep narrow notches, both the homomorphic

TABLE VI. The location of a set of troughs important to elevation perception.

Elevation Angle	Actual location (Hz)	No compensation (Hz)	Separate filter (Hz)	homomorphic (Hz)	Least-squares (Hz)
−30°	**6522**	6528	6519	6517	6517
−20°	**6679**	6680	6679	6677	6657
−10°	**7163**	7163	7163	7163	7152
0°	**8176**	8197	8176	8176	8179
10°	**8919**	8919	8919	8925	8908
20°	**9425**	9425	9430	9425	9458
30°	**10 125**	9910	10 125	9942	9899
40°	**10 793**	10 793	10 793	10 793	10 793
Mean Error	**0**	30.375	1	24.5	38.875

TABLE VII. The 3 dB width of a set of troughs important to elevation perception.

Elevation angle	Actual width (Hz)	No compensation (Hz)	Separate filter (Hz)	Homomorphic (Hz)	Least-squares (Hz)
−30°	**177**	176	177	180	189
−20°	**339**	230	341	377	495
−10°	**92**	94	92	97	95
0°	**10**	7	10	9	12
10°	**35**	36	35	35	54
20°	**200**	436	200	213	352
30°	**460**	752	461	415	502
40°	**19**	19	18	18	17
Mean error	**0**	80.5	0.5	13.25	48.5

deconvolution and the least squares deconvolution will produce unsatisfactory filters with undesirable artifacts. This is also true of traditional separate filter compensations.

During the calculation of the complex cepstrum there is direct access to the complex coefficients of the Fourier transform of the HpIR and HRIR. Any offending notches can be *removed* from this representation by altering the corresponding Fourier coefficients such that the magnitude spectrum is a smooth function. This can be carried out by a simple interpolation of the Fourier coefficients across the notch, with care being taken not to introduce any discontinuities. Once this is done the resulting filter will equalize the HpIRs influence at all frequencies barring the notch frequency.

Clearly it is bad practice to attempt to filter out notches caused by comb filtering by introducing sharp peaks. The reason here being that the discontinuity will likely not occur in exactly the same position on every occasion and thus one would be introducing peaks which do not compensate out the offending notches. This would result in further ringing artifacts compounding the problem. Far better would be to leave the notch and simply equalize around it. The homomorphic deconvolution method offers an easily implemented means of doing just this. Such a means is not so readily

TABLE VIII. The magnitude level of a set of troughs important to elevation perception,

Elevation angle	Actual magnitude (dB)	No compensation (dB)	Separate filter (dB)	Homomorphic (dB)	Least-squares (dB)
−30°	**−19.38**	−19.99	−19.44	−19.62	−17.16
−20°	**−16.46**	−16.1	−16.35	−15.99	−15.73
−10°	**−16.95**	−15.89	−16.79	−17.15	−16.53
0°	**−34.74**	−37.48	−34.43	−33.73	−33.57
10°	**−23.12**	−22.29	−23.09	−22.91	−19.24
20°	**−16.51**	−12.12	16.39	−16.38	−15.92
30°	**−16.1**	−8.62	−16.0	−16.88	−16.23
40°	**−37.2**	−30.16	−37.1	−37.64.	−33.65
Mean Error	**0**	3.06	0.12	0.44	1.58

available when dealing with the issue via least squares based factorization.

VII. CONCLUSION

The deconvolution methods described above were shown to be a valid novel method of compensating for headphone or earphone coloration. Justification was also given for the assumption that acoustic signals can be deconvolved from one another based upon shared or nearly shared roots. The theory behind the clustering of the roots of random polynomials was also shown to be generally applicable to acoustic impulse responses. The simple homomorphic subtraction method was shown to be as good or better than a separate least squares inverse filter. Without the need for separate filters, large reductions in computational effort required can be made. A simple method to compensate despite the existence of high Q peaks was also briefly described, which can easily be included as part of the homomorphic approach.

ACKNOWLEDGMENTS

[1]T. Hohn, A. Lindau, and S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *Audio Engineering Society Convention*, Vienna, Austria, preprint no. 7032 (2007).

[2]D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am **100**(1), 408–427 (1996).

[3]K. McAnally and R. Martin, "Variability in the headphone-to-ear-canal transfer function," J. Audio Eng. Soc. **50**, 263–266 (2002).

[4]H. Møller, C. Jensen, D. Hammershøi, and M. Sørensen, "Design criteria for headphones," J. Audio Eng. Soc. **43**, 218–232 (1995).

[5]S. Adams and F. Boland, "On the distortion of binaural localization cues using headphones," in Signals and Systems Conference (ISSC 2010), IET Irish, pp. 82–87 (2010).

[6]A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," J. Audio Eng. Soc. **60**, 54–62 (2012).

[7]C. Masterson, S. Adams, G. Kearney, and F. Boland, "A method for head related impulse response simplification," in *17th European Signal Processing Conference (EUSIPCO 2009)* (EUSIPCO, Glasgow, Scotland, 2009), pp. 2569–2573.

[8]C. P. Hughes and A. Nikeghbali, "The zeros of random polynomials cluster uniformly near the unit circle," Compos. Math.**144**, 734–746 (2008).

[9]P. Erdős and P. Turan, "On the distribution of roots of polynomials," Ann. Math. **51**, 105–119 (1950).

[10]L. Shepp and R. Vanderbei, "The complex zeros of random polynomials," Trans. Am. Math. Soc. **347**, 4365–4384 (1995).

[11]I. Room Acoustics Team, "Listen hrtf database," http://recherche.ircam.fr/equipes/salles/listen/ (Last viewed 8/6/2012).

[12]V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The cipic hrtf database," in *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp. 99–102 (2001).

[13]G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman and Hall/CRC, New York, 2000), pp. 30–35.

[14]K. Steiglitz and B. Dickinson, "Phase unwrapping by factorization," IEEE Trans. Acoust., Speech, Signal Process. **30**, 984–991 (1982).

[15]ITU-R Recommendation BT.1359, "Relative timing of sound and vision for broadcasting," Technical Report, International Telecommunication Union, Geneva, Switzerland (1998).

[16]P. Hansen, T. Jensen, and G. Rodriguez, "An adaptive pruning algorithm for the discrete l-curve criterion," J. Comput. Appl. Math. **198**, 483–492 (2007).

[17]E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," J. Acoust. Soc. Am. **44**, 240–249 (1968).

[18]B. Gardner and K. Martin, "Hrft measurements of a kemar dummy-head microphone," Technical Report 280, MIT Media Lab Perceptual Computing (1994). http://sound.media.mit.edu/resources/KEMAR.html (Last viewed 3/6/2012).

[19]C. Masterson, G. Kearney, M. Gorzel, and F. Boland, "Hrir order reduction using approximate factorization," IEEE Trans. Audio, Speech, Lang. Process. **20**, 1808–1817 (2012).