On Phase and Randomness in Head Related Impulse Responses

Ian J. Kelly and Francis M. Boland

***Abstract*—This paper explains why head related impulse response (HRIR) factorization algorithms based upon iterative least squares methods, such as those used by Masterson et. *al.* in [1], [2], can converge to any one of multiple equivalent solutions. The authors explain this behavior by demonstrating that it is associated with the clustering of the zeros of the equivalent z-domain Head Related Transfer Function. That is we examine the extent to which the roots of the set $H^{\phi}(z) = 0$ are clustered in a torus $\rho \leq |z| \leq \frac{\rho}{1-\rho}$ for some $0 < \rho < 1$. The paper also examines how transforming Head Related Impulse Responses to equivalent minimum phase realization forms has an influence on the results**

I. INTRODUCTION

In order to artificially spatialize a sound source for presentation with headphones, the audio must be convolved with the left and right ear Head Related Impulse Responses (HRIR's) for a given spatial position. The HRIR contains the spectral cues introduced to a signal due to the filtering effects of the pinna, head and torso, as well as the Inter-aural Time Difference (ITD) and Inter-aural Level Difference (ILD) cues introduced by the displacement of the ears. The Head Related Transfer Function (HRTF) is the Fourier transform of the HRIR. As pinna configuration and position differs from person to person, the spectral shaping is highly individual dependent. It has been proposed [1], [2] that a set of HRIRs (denoted $\mathbf{h}^{\phi}$) be simplified by factoring each filter into the convolution of a direction independent subsystem (denoted $\mathbf{f}$) which is common to the whole set and a direction dependent residual (denoted $\mathbf{g}^{\phi}$). The algorithm used in finding this common subsystem of a HRIR dataset is equivalent to finding the approximate greatest common divisor (AGCD) of the HRTF z-domain set. The task of finding the AGCD is formulated as a non-linear optimization problem over a set of M HRIRs each of length n:

$$\min_{\mathbf{f},\mathbf{g}^1,\mathbf{g}^2,\ldots,\mathbf{g}^M} \sum_{\phi=1}^{M} \left\|\mathbf{h}^{\phi} - \mathbf{f} * \mathbf{g}^{\phi}\right\| \tag{1}$$

where $\mathbf{h}^{\phi} = \left[h_0^{\phi}, h_1^{\phi}, \ldots, h_{n-1}^{\phi}\right]$, $\mathbf{g}^{\phi} = \left[g_0^{\phi}, g_1^{\phi}, \ldots, g_{j-1}^{\phi}\right]$ and $\mathbf{f} = [f_0, f_1, \ldots, f_{k-1}]$ with $m = j + k - 1$. This problem can be solved using a variant of the Gauss-Newton non-linear least squares algorithm with the exception that the usual step of linearization around the current guess is already done, as the system is bilinear in $\mathbf{f}$ and $\mathbf{g}^{\phi}$. Studies of this HRIR factorization have revealed the problem as formulated can converge to any one of multiple equivalent solutions.

II. 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 interest in mathematics circles. The celebrated work by Erdős and Turan [3] proved that the zeros of infinite order random polynomials would cluster uniformly near the unit circle for a small set of coefficient distributions. These results have, however, been extended in recent years to include more probability distributions and to include random polynomials with finite order. One such extension was made by Hughes and Nikeghbali [4]. Considering this it is therefore interesting to examine the statistical distribution of the coefficients of HRIRs and the z-plane distribution of the zeros of the equivalent transfer functions in order to determine whether zero clustering could be responsible for the multiple equivalent solutions seen in AGCD problems as described in Section I.

Looking at the distribution of the filter coefficients from 19074 HRIRs taken from the IRCAM Listen HRIR dataset which can be found at [5], it is clear that the coefficients of HRIRs such as these, closely match a skewed α-stable distribution. The α-stable distribution generalizes the classical central limit theorem for sequences of random variables which may not be i.i.d. Figure 1 shows a histogram of the coefficients from 19074 HRIRs from the IRCAM Listen dataset. This amounts to 9765888 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 [6]

$$\phi(t) = \mathbf{E}[e^{-\gamma^{\alpha}|t|^{\alpha}(1-j\beta sign(t)\tan(\frac{\pi\alpha}{2}))+j\delta t}] \tag{2}$$

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 Figure 1 were $\alpha = 1.2379$, $\beta = -0.2666$, $\gamma = 0.0039$, $\delta = -0.0013$

One could raise the question as to whether sequences such as $\mathbf{h}^{\phi}$ are of sufficient length to have uniform distribution of zeros about the unit circle. According to the celebrated Erdős - Turan result and the extension by Hughes and Nikeghbali [4], for a polynomial defined as

$$P(z) = \sum_{k=0}^{N} a_k z^k \tag{3}$$

where coefficients a_k are randomly distributed, the roots of the random polynomial will cluster uniformly around the unit

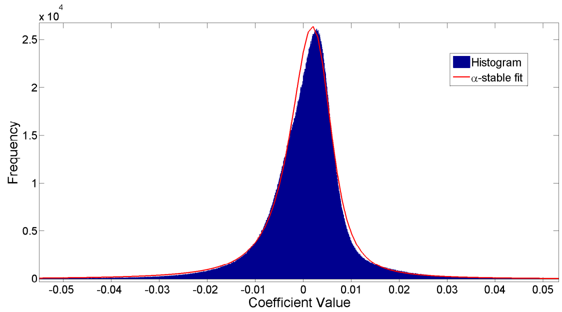

Fig. 1: Histogram of the coefficient values for 19074 IRCAM HRIRs. The histogram closely resembles a skewed α-stable distribution. A best fit α-stable distribution is shown.

circle if $L_N(P)$ is small compared to the polynomial order N where

$$L_N(P) = \log \sum_{k=0}^{N} |a_k| - \frac{1}{2}\log|a_0| - \frac{1}{2}\log|a_N| \tag{4}$$

provided a_0 and a_N are non zero.

Taking HRIRs from the IRCAM dataset 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 19074 responses from the IRCAM dataset, 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

$$(1 - \frac{v_N}{N}) \leq \frac{2L_N(P)}{N\rho} \tag{5}$$

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

$$v_N = \#\{z_k : 1-\rho \leq |z_k| \leq \frac{1}{1-\rho}\} \tag{6}$$

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 5 yields

$$v_N \geq N - 2\frac{L_N(P)}{\rho} \tag{7}$$

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 IRCAM HRIRs were calculated. The proportion of those with a magnitude within an annulus of $\rho = 0.2$ was then calculated. 1176955 out of 1188075 roots or 99.06% of the roots lay within this annulus.

Similarly letting

$$v_{\theta\phi} = \#\{z_k : \theta \leq \arg(z_k) < \phi\} \tag{8}$$

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

$$|\frac{1}{N}v_{\theta\phi} - \frac{\phi-\theta}{2\pi}|^2 \leq \frac{C}{N}L_N(P) \tag{9}$$

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

$$v_{\theta\phi}^2 - \frac{N(\phi-\theta)}{\pi}v_{\theta\phi} + \frac{N^2(\phi^2 - 2\phi\theta + \theta^2)}{4\pi^2} - NCL_N(P) \leq 0 \tag{10}$$

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

$$\lim_{N\to\infty} \mathbf{E}\left[\frac{1}{N}v_{\theta\phi}\right] = \frac{\phi-\theta}{2\pi} \tag{11}$$

as

$$N\frac{\phi-\theta}{2\pi} = 1.4222 \tag{12}$$

for $N = 512$.

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 foud by examining equation 4. 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 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 inter-aural time difference between left and right HRIR pairs. These onset delays are generally affected by measurement noise, transducer transfer functions and perhaps some d.c. 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 $R << N$ 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 figure 2. The lower signal in figure 2 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 figure 3.

A similar argument can be made for the low magnitude of the last coefficient of a HRIR, a_N. HRIRs exhibit a statistical

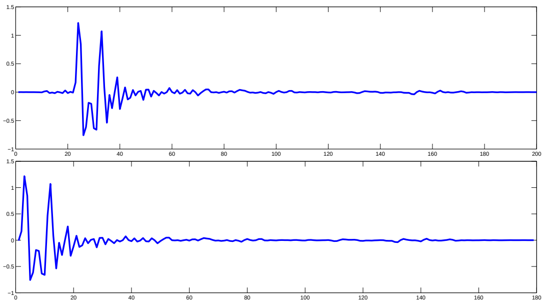

Fig. 2: This figure shows a HRIR (above) and that same HRIR with approximate onset delay removed

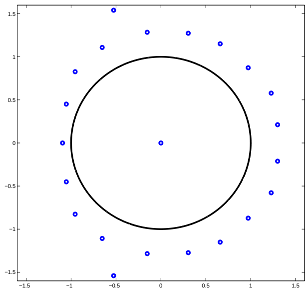

Fig. 3: Roots of the approximate delay polynomial

decaying magnitude similar to an exponential decay. It was proposed by Steiglitz and Dickinson [7] that the decay in HRIR coefficients could be modeled as 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 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 reason being that both the magnitude and phase responses of a random polynomial with an exponential decay envelope applied to its coefficients should be utterly different from those of the original random polynomial. This behavior is not implied by Steiglitz and Dickinson's model. Furthermore such a model removes the independence of the polynomials coefficients. The authors instead model the decay of acoustic signals like HRIRs as a point-wise multiplication with a decaying exponential.

$$\mathcal{Z}\left\{e^{-\alpha n}p[n]\right\} = P(e^{-\alpha}z) \tag{13}$$

where $\alpha << 1$ This has the effect of reducing 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.

Due to the fact that the roots of HRIRs cluster nearly uniformly about the unit circle, it is to be expected that a full set of HRIRs will share or very nearly share many roots. When Masterson *et al.* used an iterative least squares algorithm to factor out a common component $\mathbf{f}$ of order K from a set of HRIRs $\mathbf{h}^{\phi}$ of order N [2], [1], $K << N$. It was found that for different initial conditions, different $\mathbf{f}$'s could be extracted. This is due to the fact that it is likely that each of the $\mathbf{h}^{\phi}$ shares or very nearly shares many of the $K-1$ roots. This allows the iterative least squares algorithm to converge on differing solutions given a different initial $\mathbf{f}$.

III. Effects of using minimum phase HRIRs

An FIR system $X(z)$ of length L is said to be minimum phase when all of it's $L-1$ zeros lie within the unit circle on the Argand plane. Such a system will undergo a net phase change of zero between $\omega = 0$ and $\omega = \pi$. HRIRs are in general mixed phase systems, with zeros lying both inside and outside the unit circle. Therefore when one transforms a HRIR, $\mathbf{h}^{\phi}$, to its minimum phase equivalent, changes are made to the distribution of the polynomial's zeros. With a true minimum phase approximation of $\mathbf{h}^{\phi}$, denoted $\mathbf{h}^{\phi}_{min}$, excess phase zeros should be reflected in the unit circle into its interior at the reciprocals of their original positions.

A major contribution to the excess phase of a HRIR $\mathbf{h}^{\phi}$ is from the inter-aural time difference (ITD) related onset delay, associated with both the left and right HRIRs. This delay has the effect of forming roots within a ring, well outside of the unit circle, with the number of roots being directly proportional to the number of samples of delay before the direct.

This can be shown by reintroducing such annular zeros back into $\mathbf{h}^{\phi}_{min}$ and noting that the effect on the impulse response and phase response is roughly the same as that of introducing a pure delay identical to that which was removed when a minimum phase version of the HRIR was calculated.

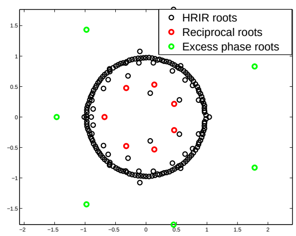

Fig. 4: Roots of a CIPIC HRIR. The roots highlighted in green in the outer ring can be shown to correspond to the response's onset delay. The reciprocals of these roots are shown highlighted in red

Figure 4 shows the zero distribution of a HRIR from the CIPIC dataset [8]. The excess phase zeros corresponding to the onset delay have been highlighted in green. Other excess phase zeros are un-highlighted. The reciprocals of these roots also highlighted in red.

By removing just these roots from this outer ring via a least squares factorization and convolving the result with a polynomial whose own roots are the reciprocals of these roots, we can show that that this ring of zeros is indeed due purely to the onset delay.

$$\hat{\mathbf{h}}^{\phi} = ((\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{h}^{\phi}) * \bar{\mathbf{r}} \quad (14)$$

Here $\mathbf{R}$ is a Toeplitz convolution matrix formed from a vector $\mathbf{r}$ which represents the polynomial whose roots are the annular excess phase roots of $\mathbf{h}^{\phi}$. $\bar{\mathbf{r}}$ is a time reversed version of $\mathbf{r}$ and thus its roots are the reciprocals of $\mathbf{r}$'s. $\hat{\mathbf{h}}^{\phi}$ will have the same magnitude response as $\mathbf{h}^{\phi}$ but with a phase response that differs only due to the removal of initial delay.

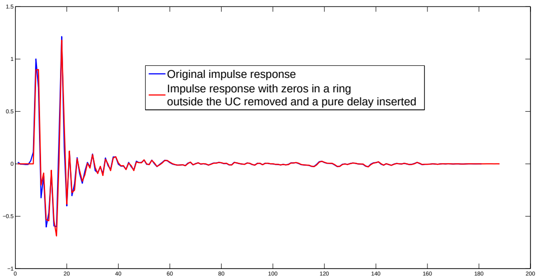

Fig. 5: Response of the original HRIR $\mathbf{h}^{\phi}$ with the shifted response of $\hat{\mathbf{h}}^{\phi}$ superimposed

Figure 5 shows the original impulse response $\mathbf{h}^{\phi}$ with $\hat{\mathbf{h}}^{\phi}$ superimposed over it shifted by seven samples. Seven samples of shift correspond directly to the seven roots of $\mathbf{h}^{\phi}$ highlighted in figure 4. Figure 5 clearly shows that beyond the initial delay both $\mathbf{h}^{\phi}$ and $\hat{\mathbf{h}}^{\phi}$ are nearly identical.

HRIRs however are not always simply minimum phase systems uniformly shifted in time. As can be seen from figure 4, there are also a small number of excess phase roots not belonging to this *'outer ring'* and not simply due to onset delay. Such roots are far less likely to be common across a HRIR set and thus could lead to errors in factorization problems.

It should be noted that not all algorithms designed for calculating minimum phase versions of signals will actually produce output signals who's roots include the reciprocals of the excess phase zeros of the original function. Many algorithms designed to calculate minimum phase versions of a signal use what is known as the Kolmogoroff spectral factorization method, also referred to as the Hilbert transform method within the signal processing community [9]. However common implementations of this method such as the *rceps()* function found in MATLAB do not exactly calculate a minimum phase version of the given response. The principal disadvantage of the Kolmogoroff method is that a summation around the unit circle is always different to an integration about the unit circle thus introducing errors. A description of this method can be found in [9].

IV. Other applications for root clustering in acoustic responses

Root clustering in acoustic responses can also be of benefit with regards to deconvolution/equalization. The clustering can for example be exploited when deconvolving a headphone response from a head-related impulse response (HRIR) or a binaural room impulse response (BRIR) in order to compensate for the headphone's impulse response. This can be done without increasing the order of the HRIR/BRIR beyond that of the original and without the need for a separate compensating filter.

In a soon to be published paper the authors demonstrate how such a compensation performed using homomorphic deconvolution can outperform a separate least squares compensator of the same order as the original response. This can be achieved without introducing any extra computational costs.

V. Conclusion

It was shown that a set of approximately delayed exponentially decaying polynomials such as a HRIR have their roots densely clustered. This effect can explain why head related impulse response (HRIR) factorization algorithms have a tendency to converge to any one of multiple equivalent solutions given different initial conditions. The effect of transforming Head Related Impulse Responses to equivalent minimum phase realization forms on the root locations was also examined and the majority of excess phase zeros were shown to be due to onset delay and thus were not of consequence to the factorization algorithm.

Acknowledgment

The authors would like to acknowledge support for this research from Trinity College Dublin, Google Inc. and Science Foundation Ireland Project(08/IN.1/I2112) - Content Aware Media Processing

References

[1] C. Masterson, G. Kearney, M. Gorzel, and F. Boland, "Hrir order reduction using approximate factorization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 6, pp. 1808 –1817, aug. 2012.

[2] C. Masterson, G. Kearney, and F. Boland, "Hrir factorisation: A regularised approach," in *18th European Signal Processing Conference (EUSIPCO 2010)*, Aalborg, Denmark, August 2010.

[3] P. Erdos and P. Turan, "On the distribution of roots of polynomials," *The Annals of Mathematics*, vol. 51, no. 1, pp. 105–119, Jan 1950.

[4] C. P. Hughes and A. Nikeghbali, "The zeros of random polynomials cluster uniformly near the unit circle." *Compositio Mathematica*, vol. 144, pp. 734–746, March 2008.

[5] I. Room Acoustics Team, "Listen hrtf database," http://recherche.ircam.fr/equipes/salles/listen/, accessed: 08/06/12.

[6] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic models with infinite variance*. Chapman and Hall/CRC, 2000.

[7] K. Steiglitz and B. Dickinson, "Phase unwrapping by factorization," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 30, no. 6, pp. 984–991, dec 1982.

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

[9] J. Claerbout, *Fundamentals of Geophysical Data Processing*. McGraw-Hill, 1976.