Augmented Hankel Total Least Squares Decomposition of Head-Related Transfer Functions

Kenneth John Faller II, Armando Barreto and Malek Adjouadi

Florida International University
College of Engineering and Computing
Department of Electrical and Computer Engineering

Miami, FL 33174 USA

kfall001@fiu.edu, barretoa@fiu.edu and adjouadi@fiu.edu

Abstract

There are currently 3 options for the implementation of spatial audio systems based on Head-Related Transfer Functions (HRTFs): individually measured HRTFs, “generic” HRTFs or customizable HRTFs. Individualized HRTFs require the subject to undergo lengthy measurements in an anechoic chamber under the supervision of trained personnel which limits their availability to the average potential user. To overcome this, many researchers and commercial developers have resorted to using generic HRTFs. However, it has been shown that generic HRTFs result in an increase in localization errors. The third possibility, which we are pursuing, is the customization of HRTFs in which the anatomical measurements of a potential user are utilized to determine the parameters of a structural model of the head and the pinnae. However, an initial step of decomposing the impulse responses of individualized HRTFs into a summation of Damped and Delayed Sinusoids (DDSs) is needed to reveal the parameters of the structural pinna model. Our group has already developed an exhaustive search decomposition method that seems to do this accurately when the DDSs are separated by
long latencies but not for short latencies. This paper proposes the application of the Hankel Total Least Squares (HTLS) decomposition method as a solution to the parameter extraction problem faced in customizing HRTFs when short latencies are expected between the DDSs that make up their impulse response.

0 Introduction

Head-Related Transfer Functions (HRTFs), or their impulse responses and time domain equivalents called Head-Related Impulse Responses (HRIRs), capture the localization cues implicit in binaural spatial audio. If a pair of HRTFs can be experimentally measured, a monaural sound wave can then be filtered by the HRTFs and when the resulting binaural sound waves are played to the listener they would make it appear as if the sound were emanating from the desired location in three-dimensional (3D) space, specified by azimuth ($\theta$), elevation ($\Phi$) and distance ($r$) (see Figure 1). Individualized HRTFs are usually obtained by a system similar to the one shown in Figure 2. However, a setup such as this is expensive, cumbersome and requires trained personnel for operation.

In order to avoid the need of such a measurement system used to obtain individualized HRTFs, several research groups have created “generic” HRTFs which are obtained by either measuring the HRTFs of a KEMAR mannequin (e.g., MIT’s measurements of a KEMAR Dummy-Head Microphone [1]) or creating a database of HRTFs measured on users with diverse anatomical features (e.g., the CIPIC Database [2]). These generic HRTFs have been successful in bringing binaural 3D audio to the general public, but it has been shown that they result in an increase in localization errors, especially in elevation [3]. In this context, researchers are now trying to developed methods that allow for the generation of HRTFs that do not require the inconvenient
measurements necessary for individualized HRTFs but have less localization errors than those experienced with generic HRTFs.

![Figure 1: Spherical coordinate system.](image1)

One such method was developed by our group in [4], in which a structural pinna model was proposed to generate customized HRTFs based on the anthropometric measurements of an intended user. The resulting HRTFs, on average, produced listening test results that were similar to those obtained with measured HRTFs. However, the procedures to extract the parameters needed for the model proved to be tedious and required a large amount of human operator intervention.

Our overall intent is to create an association between initial databases of measured HRIRs, encoded as sets of model parameters and dimensional measurements, to eventually elucidate a mapping from dimensional measurements to customized HRIRs.
(sets of model parameters). Once this is achieved, the mapping could be applied to the dimensional measurements obtained from a subject not originally included in the databases, to generate his/her customized HRIRs. However, this paper focuses on addressing problems found in creating the initial database of sets of model parameters from measured HRIRs. A method is proposed which addresses this problem.

1 Background

How we localize sound has been a subject of study for a long time. Over a century ago, Lord Rayleigh developed the duplex theory in which azimuthal spatial localization was achieved by two binaural cues called Interaural Time Difference (ITD) and Interaural Intensity Difference (IID) (see Figure 3) [5]. The duplex theory’s simplicity is very attractive and it is effective in explaining how humans localize azimuthally. However, it does not explain how humans are able to determine the elevation of a sound source. Furthermore, front/back reversal of the location of a sound source can occur when the IID and ITD are the only spatial auditory cues used [6].

![Figure 3: Estimate of azimuth of a sound source using ITD and IID.](image)

One way of capturing the spectral cues necessary for both elevation and azimuth localization is by measuring the directionally dependant frequency response of each ear which is represented by its HRTF. As mentioned in the introduction, measuring
individualized HRTFs requires specialized equipment that is expensive and cumbersome. To remove the need to measure every potential user using this equipment, researchers have resorted to the use of generic HRTFs. However, HRTFs are heavily dependent on the physical characteristics of the potential users, and hence they vary from user to user [7]. Accordingly, the use of generic HRTFs results in an increase in localization errors, especially in elevation [3]. In view of this, researchers have attempted to develop methods to generate HRTFs that do not require inconvenient measurements but reduce the amount of localization errors experienced when generic HRTFs are used. These include numerical computational methods, database interpolation methods, physical models and structural models. The following subsections outline these methods and models.

1.1 Numerical Computational Methods

Numerical computational methods can be used to determine the transfer function of a user’s ear for a desired location in 3D space. Numerical solutions such as the Boundary Element Method (BEM) [8-10], the Finite Element Method (FEM) [11] or ray tracing [12] are used to achieve this. Although the numerical computational methods provided useful insight into the behavior of HRTFs [13], it is known that they face three fundamental problems: 1) the difficulty of meshing the more intricate parts of the human anatomy (e.g., the pinnae), 2) the high computational cost and 3) the lack of uniformity in the physical characteristics of humans which does not allow for an exact match to a new intended user’s anthropometric measurements when the meshes are scaled [2, 7]. Although it has been shown that proper scaling reduces the errors in localization along
the median plane when compared to generic HRTFs alone [14], researchers are still seeking a simplified and computationally inexpensive method of modeling HRTFs.

1.2 Database Interpolation Methods

In this approach, an attempt is made to select a set of HRIRs, from a database of HRIRs, which result in the highest spatial location accuracy for a new potential listener. If a technique or method for selecting the HRIRs is established, then the appropriate set of HRIRs can be retrieved from the database and a sound signal can be convolved with the HRIRs that correspond to a desired virtual position. One example of this is the creation of a database of measured HRIRs along with the anthropometric measurements of a set of subjects (e.g., the CIPIC database [2]). Once the database is complete, a set of HRIRs are selected such that the anthropometric measurements of the new listener and a subject in the database are closely matched [15-17]. Another example is to have a small database of HRIRs and then have the listener select which of the HRIRs result in the best spatial localization (e.g., Sensaura’s Virtual Ear software [18]). Unfortunately, the diversity of the anthropometry of humans makes the selection of the “best matching” HRIRs difficult and still remains an open research problem.

1.3 Physical Models

Many of the physical models attempt to represent the human head and/or torso as spheres [19-20]. This eventually led to a simple and compact spherical-head model derived from interaural time differences (ITDs), in which the head radius was successfully associated to a weighted sum of three head dimensions (head height, width and length) [21]. The results were promising in that most of the testing subjects were able to localize accurately within 5º using binaural sounds generated with this model.
However, the head model addresses the discrimination of different azimuths, but not the differentiation of elevations.

1.4 Structural Models

Structural models, proposed by Genuit in 1983 [22], segment an HRTF into smaller components. Then an empirical relationship is found between the parameters of the components, the spatial location of a sound source and the anthropometric measurements of an intended user. Ideally, this relationship can be used to generate HRTFs for a new intended user based on anthropometric measurements. One such model was developed by Brown and Duda in [23]. The proposed structural model accounts for the effects of the head, shoulder and pinna by cascading them together to generate HRTFs (Figure 4). However, the parameters for the model were not defined from measurements of anatomical features of the individual user. Additionally, Batteau has shown that a majority of the elevation cues come from the pinnae, which may allow the simplification of the model by elimination of the shoulder sub-model and representation of the head effects through ITD and IID [24].

Figure 4: Block diagram of structural model (after Brown and Duda [23]).
1.5 **Structural Pinna Models**

According to Blauert [25], the first attempt to understand the function of the pinna was made by Schelhammer in 1684. Schelhammer hypothesized that the sound entering the ear canal was the superposition of a series of reflected components that converged in the ear canal. Batteau in 1967 revisited this idea of considering the pinna as a sound reflector [24]. He believed that the time difference between a direct sound and a sound reflected from the various structures of the pinna (see Figure 5) would change as the direction of the sound source was changed. Equations (1)-(2) show the resulting impulse response and frequency response equations, respectively. Unfortunately, Batteau’s model did not offer a method of estimating the reflection coefficients ($\rho$) and the delays ($\tau$). Furthermore, the consideration of only two reflections may be a limitation of his approach [25]. Chen in [26] increased the number of reflectors but required a very large number of parameters to achieve an acceptable fit in the time domain between a generated HRIR and an individualized HRIR.

\[
h(t) = \delta(t) + \rho_1 \delta(t - \tau_1) + \rho_2 \delta(t - \tau_2) \tag{1}
\]

\[
H(f) = 1 + \rho_1 e^{-j2\pi f \tau_1} + \rho_2 e^{-j2\pi f \tau_2} \tag{2}
\]

![Figure 5: Example of how sound waves would bounce-off of the pinna into the ear canal.](Image)
Another way of analyzing and ultimately customizing HRTFs is to investigate their behavior in the frequency domain. One recently proposed model that does this was developed by Satarzadeh [7]. Satarzadeh estimated the parameters of two band-pass filters and a comb filter from the anthropometric measurements of an intended user. The delay of the comb filter and the magnitudes of the band-pass filters were then adjusted according to the desired position of a sound source in 3D space. The resulting HRTFs displayed a good approximation to the measured HRTFs in both the frequency and time domains. However, Satarzadeh admits that his model will only work on subjects that exhibit prominent pinna reflections.

In [4] we proposed a pinna model in which the sound entering the ear canal is the summation of signals with different delays. Similar to Batteau’s model, the delays in this model are a result of waves bouncing-off of the geometrical structures of the pinna into the ear (see Figure 5). However, only the first three reflections and the first (and most prominent) resonance of the pinna are included in the model because it was shown in [27] that most of the reflections occur within 300 μs. The three reflectors are parallel to each other, but in series with the resonance, as shown in Figure 6. One of the major contributions in [4] was a method to obtain the parameters in the model of Figure 6 from individualized HRIRs. This is achieved by decomposing the individualized HRIRs into 2nd order damped sinusoids in order to reveal the parameters. The next section details the method proposed in [4] that was used to define the parameter values in Figure 6 from individualized HRIRs. Following this section is the description of our most current pinna structural model. The final two sections of this chapter detail the exhaustive search
decomposition method and the Hankel Total Least Squares (HTLS) decomposition method respectively.

![Diagram](image)

**Figure 6:** Block diagram of structural pinna model (after Gupta [4]).

# 2 Decomposition Methods

Despite the limitations of Batteau’s original structural pinna model, our research group finds its simplicity appealing as a basis for HRTF customization. The two major shortcomings with his model were the inability to estimate the value of its parameters and the small number of reflections considered. The following subsections detail the method proposed in [4], describe our most current pinna structural model, and finally outline the exhaustive search decomposition method and the Hankel Total Least Squares (HTLS) decomposition method, used to define the model parameters from measured HRIRs.

## 2.1 Manual Decomposition

The initial method developed by our group to decompose HRIRs and estimate the parameters of the structural pinna model, shown in Figure 6, consisted of manually windowing the HRIR and applying a signal approximation method, such as Prony, to the window [4]. The idea behind this is that if a window width could be selected such that it only contains a portion of one damped sinusoid, then Prony could be used to estimate it,
extrapolate and remove it from the overall HRIR. The remainder of this subtraction would then be shifted to the onset of the next damped sinusoid and the process would be repeated for the remaining damped sinusoids needed for the model (i.e., a total of four damped sinusoids). Performing this decomposition for each HRIR would yield the parameters (resonance transfer function, reflection coefficients and delays) to instantiate a model like the one shown in Figure 6 for the ear, the azimuth and the elevation associated with the HRIR decomposed, eventually creating a database of system parameters for multiple spatial locations and multiple individuals. Simultaneously, a database of anthropometric measurements (e.g., pinna length, pinna width, etc.) of the same users would be created. After developing large-enough databases, multiple linear regression would be used to relate the parameters to the anthropometric measurements. At this point, the anthropometric measurements of a new intended user could be “converted” to parameter values to instantiate the model at a desired location.

However, the window width selection for the decomposition method was performed manually and proved to be a tedious task that required a large amount of human interaction, which also weakened the objectivity of the process. This prompted our group to develop an automated method of extracting the desired damped sinusoids and their parameters.

2.2 Extended Structural Pinna Model

In the previous model (Figure 6) one resonance, which represents the frequency and damping component of the damped sinusoids, was initially assumed for all the paths. However, the results from the exhaustive search decomposition method (see 2.3) raised the possibility that different resonances (i.e., frequency and damping components) could
exist for each damped sinusoid. As a result of this, the model in Figure 6 was extended to allow for each path to have its own resonance (see Figure 7). Unlike the original model, which assumed a single resonance, the new model has the freedom to define a different resonance for each parallel path. Although this model adds more parameters, it allows for more flexibility in the damping and frequency components of the damped sinusoids selected to model an HRIR [28].

![Figure 7: Block diagram of structural pinna model.](image)

### 2.3 Exhaustive Search Decomposition

The idea of procedurally selecting a window width, estimating the current damped sinusoid contained within it, removing it from the current HRIR remnant, shifting to the next onset of another damped sinusoid and then repeating until all damped sinusoids from the model are obtained is used in this method as well. However, in this decomposition method an attempt was made to remove the need for human interaction during the decomposition phase. In [29], we proposed an exhaustive search decomposition method in which all possible widow widths are attempted by exploring all the branches of the associated search tree diagram. Each node represents a window width and each level represents a window. To clarify, consider the search tree diagram shown in Figure 8. This is a visualization of a search tree when two windows are considered which go from 2 to
10 samples (45.35 μs to 226.75 μs at the sampling rate of 44.1 kHz used in [29]). As can be seen, this particular example would result in 9x9=81 leaf nodes.

Figure 8: Example search tree diagram. A node in each level has been shaded to indicate a specific window succession that could result in the candidate HRIR that yields the best fit to the HRIR being decomposed.

The window widths represent the possible delay of the sinusoid segment immediately following the window (see Figure 9). The sinusoid segment, contained within each window, will be used to approximate a damped sinusoid (see Figure 10). Then every possible sequence of second-order approximations, considered at the appropriate delays (which is the sample width of the preceding window, see Figure 9), would be summed together resulting in a candidate HRIR at each leaf node of the tree diagram.

Figure 9: Example of how the window width corresponds to the following sinusoid’s delay.
Figure 10: Example of how a windowed portion of a signal (left) is used to approximate a damped sinusoid (right).

All the candidate HRIRs are temporarily stored and compared to the measured HRIR under analysis using Equations (3) and (4) to assess their individual similarity to the original measured HRIR or “fit” where $N$ is the length of the HRIR, $n$ is the current sample, $o$ is the original HRIR, $c$ is the candidate HRIR, $e$ is the error and $f$ is the percent fit. The candidate HRIR with the highest “fit” is considered to be the reconstructed HRIR that most accurately portrays the measured HRIR.

$$e(n) = o(n) - c(n)$$  

$$f = \left(1 - \frac{1}{N} \sum_{n=1}^{N} (e(n)^2) \right) \times 100\%$$  

(3)  

(4)

The HRIR decomposition cases studied in [29] using this method resulted in a 96% average fit which was very promising. However, the exhaustive search process is computationally expensive. For example, for the decomposition of the HRIR in just five windows, iterating through all possibilities would require the traversal of the associated tree diagram to $9 \times 9 \times 9 \times 9 \times 9 = 59,049$ leaf nodes. Furthermore, if more windows are added, each would increase the number of leaf nodes to be reached by a factor of 9. Additionally
the fit achieved by this method showed a tendency to be lower when two of the damped sinusoid components were separated by a small number of samples (e.g., less than 5 samples or 52.1 μs at a 96 kHz sampling rate).

To explore this potential limitation, a single damped sinusoid (x) was created and is shown in Figure 11 (top). Signal x was then tested with the exhaustive search decomposition method using Prony and Steiglitz-McBride (STMCB) function approximations (for a detailed description of these algorithms see [30]). A small window with a 3-sample width (or 31.25 μs at a 96 kHz sampling frequency) was used in an attempt to approximate the entire signal. The approximation signals resulting from the STMCB (xs) and Prony (xp) methods are also shown in Figure 11. As can be seen in the results, xs and xp fail to properly reconstruct the original signal x. This could potentially lead to inaccurate approximations of the parameters for the pinna model because it is believed that the latencies between damped sinusoids could be short for certain azimuth and elevation angles [27]. Thus, it became apparent that a method that is less computationally complex and has the ability to decompose signals when the delay between the damped sinusoids was small is necessary.

Figure 11: x (top) vs. xs (middle) and xp (bottom).
2.4 Hankel Total Least Squares (HTLS) Decomposition

The Hankel Total Least Squares (HTLS) decomposition method was first developed in [31] for the decomposition of Nuclear Magnetic Resonance (NMR) data. This method organizes the data into a Hankel matrix, then it decomposes this matrix using the Singular Value Decomposition (SVD) method (see [32] for a detailed description of the SVD method) and finally applies the Total Least Square (TLS) algorithm to estimate the parameters of a damped exponential model. An order of $K$ damped exponentials is modeled by their related parameters (frequency, damping factor, amplitude and phase). This method is very robust, accurate and works with both complex [31] and real signals [33].

As presented before, it is believed that HRIRs consist of damped sinusoids that are summed together when they enter the ear canal [4, 24, 27, 29]. Fortunately, damped sinusoids can be modeled by damped exponentials. Hence, the HTLS decomposition method is appropriate for the decomposition of HRIRs. However, two issues are raised when applying this method to HRIR decomposition: 1) the number of $K$ damped sinusoids that make up the HRIRs is unknown and 2) the delay parameters from the structural pinna model proposed in the previous section (see Figure 7) are not originally considered by the HTLS method. The HTLS method is a well established method for decomposition of a summation of damped sinusoids without delay. However, as will be seen in subsection 2.4.1, the HTLS method is sufficient for the decomposition of a summation of Damped and Delayed Sinusoids (DDS) with short latencies (e.g., the sinusoids are delayed less than 5 samples or 52.1 µs at a 96 kHz sampling rate).
Additionally, a method for estimating these delays is proposed to work in conjunction with the HTLS method for decomposition of HRIRs.

2.4.1 Damped Sinusoidal Detection

As mentioned in the previous subsection, the number of $K$ damped sinusoids contained in an HRIR is initially unknown. The main concern with selecting the correct order is not to under-model and potentially miss some of the damped sinusoids or over-model such that the HTLS method breaks down [31]. The worst case scenario is to under-model which causes the HTLS method to return an incomplete or empty set of the damped sinusoids. This is usually remedied by setting a high model order, such that the signal is over-modeled and the HTLS method will return the actual damped sinusoid components of the HRIR along with additional spurious damped sinusoids forced by the over-estimation of the model order. This will require the separation of the relevant damped sinusoids from the spurious ones. Once this is achieved, the sinusoids deemed spurious are discarded and the remaining sinusoids are considered the sinusoids that an HRIR is comprised of.

The separation can be achieved by searching the HTLS results in the over-modeled case for damped sinusoids that conform to the structural pinna model shown in Figure 7. The criteria for separation are as follows:

1. Eliminate the HTLS results that have frequencies above 20 kHz.
2. Eliminate the HTLS results that have amplitudes close to zero.
3. Eliminate the HTLS results that are too heavily damped.
4. Eliminate the HTLS results that do not have complex conjugate pole pairs within the unit circle.
The frequency limitation used for criterion one is based on what is believed to be the “important” frequency range (from about 1 to 14 kHz) for sound source elevation localization [7]. However, because a precise frequency range is unknown 20 kHz was selected to insure that these important frequencies are captured by the HTLS method. The second criterion is a simple empirically-defined threshold of 0.01 (normalized to the maximum value of the HRIR sequence) which must be surpassed by the absolute value of the amplitude parameter in order to be considered a relevant damped sinusoid. It was noticed that some of the spurious damped sinusoids returned from the HTLS method were heavily damped such that they resembled an impulse (which is not in agreement with the resonant nature of our model). This behavior became apparent when the damping factor was higher than 2. Hence, sinusoids with a damping factor greater than 2 are eliminated. The last criterion is from the knowledge of the relationship between the pole locations and a signal. It is known that a pair of complex-conjugate poles corresponds to a signal with oscillatory behavior [30]. Furthermore, the poles must lie within the unit circle in order for the signal to have a damping characteristic. Therefore, only signals with complex-conjugate pole pairs that lie within the unit circle are considered relevant damped sinusoids and the rest are eliminated.

In order to study the issue of assigning a proper model order $K$, several values were used for the decomposition of a synthetic signal, which was simulated as composed of four damped sinusoids. The damped sinusoids were each $N$ points in length, $n=0,\ldots,N-1$. Equation (5) was used for generation of $i$ synthetic damped sinusoids. $Fs$ is the sampling frequency, $d_i$ is the damping factor, $f_i$ is the frequency in Hertz, $\phi_i$ is the phase in radians and $a_i$ is the amplitude. The length of the signals was set to $N=256$ and the
simulation used a sampling frequency of $F_s=96$ kHz. In this representative example, it was found that the HTLS method performed best with $K=20$.

$$x_i(n) = a_i \cdot e^{-d_i \cdot n} \cdot \sin \left( 2 \cdot \pi \cdot \left( \frac{f_i}{F_s} \right) \cdot n + \varphi_i \right)$$

To illustrate the effectiveness of the method, four synthetic damped sinusoids were created using Equation (5) and the parameters shown in Table 1. The resulting parameter estimation from the HTLS method is shown in Table 2. Figure 12 shows the individual damped sinusoids ($x_1$, $x_2$, $x_3$ and $x_4$) compared to the resulting damped sinusoids from the HTLS method ($\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ and $\hat{x}_4$). Figure 13 shows the time domain plot of the summation of the damped sinusoids ($x$) versus the summation of the damped sinusoids that resulted from the HTLS method ($\hat{x}$). Figure 14 shows the frequency domain plots of $x$ (left) and $\hat{x}$ (right). As can be seen, the results from the HTLS method are very accurate in both time and frequency.

Table 1: Parameters for damped sinusoids used in simulation.

<table>
<thead>
<tr>
<th>Signal Name</th>
<th>Frequency (kHz)</th>
<th>Phase (radians)</th>
<th>Damping</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>14</td>
<td>0</td>
<td>0.3</td>
<td>2</td>
</tr>
<tr>
<td>$x_2$</td>
<td>12</td>
<td>$\pi/7$</td>
<td>0.1</td>
<td>0.95</td>
</tr>
<tr>
<td>$x_3$</td>
<td>10</td>
<td>$\pi/6 + \pi$</td>
<td>0.2</td>
<td>0.8</td>
</tr>
<tr>
<td>$x_4$</td>
<td>4</td>
<td>$\pi/9 + \pi$</td>
<td>0.15</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Table 2: HTLS parameter estimates.

<table>
<thead>
<tr>
<th>Signal Name</th>
<th>Frequency (kHz)</th>
<th>Phase (radians)</th>
<th>Damping</th>
<th>Amplitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{x}_1$</td>
<td>14</td>
<td>0</td>
<td>0.288</td>
<td>2</td>
</tr>
<tr>
<td>$\hat{x}_2$</td>
<td>12</td>
<td>$\pi/7$</td>
<td>0.096</td>
<td>0.95</td>
</tr>
<tr>
<td>$\hat{x}_3$</td>
<td>10</td>
<td>$\pi/6 + \pi$</td>
<td>0.192</td>
<td>0.8</td>
</tr>
<tr>
<td>$\hat{x}_4$</td>
<td>4</td>
<td>$\pi/9 + \pi$</td>
<td>0.144</td>
<td>1.55</td>
</tr>
</tbody>
</table>
Figure 12: The time domain plots of the original damped sinusoids \((x_1, x_2, x_3, x_4)\) compared to the damped sinusoids that resulted from the HTLS method \((\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4)\).

To further exemplify the accuracy of this method Equations (3) and (4) were used to calculate the “fit” in the time domain where \(o\) is equal to \(x\) and \(c\) is equal to \(\hat{x}\). Furthermore, the Spectral Distortion (SD) score (Equation (6)) was used to measure the similarity between two magnitude responses, where \(|H_i|\) is the magnitude response of the \(i^{th}\) frequency of \(x\), \(|\hat{H}_i|\) is the magnitude response of the \(i^{th}\) frequency of \(\hat{x}\) and \(I\) is the number of frequencies. The lower the SD score, the more similar the magnitudes are to each other. The synthetic example presented here resulted in a fit of 99.86% and an SD score of 0.1852 dB.

\[
SD = \sqrt{\frac{1}{I} \sum_{i=1}^{I} \left(20 \log_{10} \left| \frac{H_i}{\hat{H}_i} \right| \right)^2 [dB]}
\]  

(6)
2.4.2 Exhaustive Zeroing Algorithm for time-delay estimation

It is believed that some of the individual damped sinusoids represent sound components that reflect off of the various structures of the pinna before entering the ear canal, as described in section 1.5. These reflected components are believed to follow an indirect path into the ear canal (see Figure 5) and therefore have a time delay introduced. Time delay, however, is not a parameter that the HTLS method estimates and would require an additional method to estimate. Creating a method of time delay estimation is not trivial and, to the best of our knowledge, a method does not exist that simultaneously
estimates the damped sinusoids’ parameters and time delay. Methods do exist, however, that estimate time delay but generally the choice of method depends heavily on the target application [34]. Unfortunately, none of the existing methods are equipped to handle the time delay estimation needed for the problem addressed in this paper.

The method used for time delay estimation in this paper takes advantage of an indirect property of HTLS noticed during experimentation. It was observed that the HTLS method is able to fit non-delayed damped sinusoids to damped and delayed sinusoids with short latencies. For example, the synthetic damped sinusoid $x_2$ from the previous section was delayed by 4 samples (or 41.7 µs at a 96 kHz sampling rate) to obtain $x_{t2}$ using Equation (7). Essentially this is just the expression in Equation (5) delayed by $t_i$ samples and then it is multiplied by the Heaviside function defined by $\psi(n) = 1$ for $n \geq 0$ and 0 otherwise. Once the signal was delayed, then the HTLS method was applied to $x_{t2}$ to try to obtain an estimate $\hat{x}_{t2}$. As can be seen in Figure 15, the HTLS method fitted a damped sinusoid that starts at zero ($\hat{x}_{t2}$) to the damped and delayed sinusoid ($x_{t2}$). Hence, we propose to supplement the use of HTLS for the problem at hand by zeroing out an appropriate number of initial samples (for this example that would be the first 4 samples or 41.7 µs at a 96 kHz sampling frequency) in order for the approximation from the HTLS method to be complete.

$$x_a(n) = x_i(n - t_i) \cdot \psi(n - t_i)$$

(7)

This is done by sequentially zeroing out the samples of $\hat{x}_{t2}$ using Equation (7) where $t_i$ goes from 0 to 5 (it was empirically determined that this method works well when the delay is less than 5 samples or 52.1 µs at a 96 kHz sampling frequency).
each stage of the zeroing, the two damped sinusoids ($x_{t2}$ and $\hat{x}_{t2}$) are compared using the fit equations (3) and (4) where $o$ is $x_{t2}$ and $c$ is $\hat{x}_{t2}$. The resulting fits are stored and the zeroing amount that resulted in the maximum fit is used as the effective amount of initial samples that must be zeroed. Figure 16 shows the results of this process using $x_{t2}$ and $\hat{x}_{t2}$.

As can be seen, the maximum is at four which corresponds to a delay or shift to the right of 4 samples (or 41.7 µs at a 96 kHz sampling frequency). Figure 17 shows $x_{t2}$ versus $\hat{x}_{t2}$ after the application of the zeroing algorithm. The fit achieved is 99.82% and the SD score is 0.1868 dB for this example.

Figure 15: A delayed signal ($x_{t2}$) versus the results from the HTLS method ($\hat{x}_{t2}$).
The previous example shows that the zeroing algorithm works well when a single sinusoid is present. However, as mentioned in the background section, it is believed that HRIRs are comprised of multiple sinusoids. The following example is used to exemplify that the zeroing algorithm is also appropriate for cases when multiple damped and delayed sinusoids are present. The individual damped and delayed sinusoids ($x_{t1}$, $x_{t2}$, $x_{t3}$ and $x_{t4}$) are created using Equation (7). The parameters for the sinusoids used in this example are the same as the corresponding non-delayed damped sinusoids in Table 1 (e.g., the parameters for $x_t$ are used for $x_{tt}$) and the delays used are shown in Table 3.
After the delays are applied, the damped sinusoids are summed together to create \( x \). This composite signal \( x \) is then decomposed using the HTLS algorithm to obtain the parameters \( a_i, d_i, f_i \) and \( \varphi_i \) for the damped sinusoid estimates \( \hat{x}_{i1}, \hat{x}_{i2}, \hat{x}_{i3} \) and \( \hat{x}_{i4} \). At this point, the time-delay parameter \( t_i \) of each of the four components still needs to be estimated. This is achieved using the exhaustive zeroing algorithm.

The complete exhaustive zeroing algorithm is described by Equations (8)-(14) assuming that a maximum of \( p \) samples will be zeroed at the beginning of each of the \( i \) damped sinusoids obtained from HTLS. \( T(r,i) \) in Equation (8) is a \((p+1)i \times i\) matrix storing in each of its rows the amount of zeros that should be overwritten at the beginning of the damped sinusoids (i.e., each of its rows is a possible solution to the problem of estimating all the \( t_i \) delays). In Equation (10), \( \hat{x}(r,n) \) is the general form of a composite sequence, which is a candidate to be the total approximation of an original sequence \( x(n) \), being decomposed. In all cases \( \hat{x}(r,n) \) is built from the \( i \) damped sinusoids obtained from HTLS, but the pattern of zeros imposed at the beginning of each of them is determined by the row \( r \) in \( T(r,i) \), through the argument of the Heaviside function \( \beta_{ri} \). For each value of \( 1 \leq r \leq (p+1)i \), the error \( (e(r,n)) \) sequence and corresponding “fit” \( (f(r)) \) are calculated (Equations (11) and (12)) and the \( r^* \) that maximizes \( f(r) \) is chosen. The corresponding composite \( \hat{x}(r^*,n) \) is kept as the best reconstruction of \( x(n) \) and the corresponding zeroing pattern is kept as \( t_i^* \).

\[
T(r,i) = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \cdots & 0 & 1 \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
p & p & \cdots & p & p-1 \\
p & p & \cdots & p & p
\end{bmatrix}
\]  

(8)
\[ \beta_{ri} = T(r, i) \] (9)

\[ \hat{x}(r, n) = \sum_{i=1}^{j} x_i(n) \cdot \psi(n - \beta_{ri}) \] (10)

\[ e(r, n) = x(n) - \hat{x}(r, n) \] (11)

\[ f(r) = \left[ 1 - \frac{\frac{1}{N} \sum_{n=1}^{N} (e(r, n)^2)}{\frac{1}{N} \sum_{n=1}^{N} (x(n)^2)} \right] \times 100\% \] (12)

\[ r^* = \arg \max_{1 \leq r \leq (p+1)} f(r) \] (13)

\[ t_i^* = T(r^*, i) \] (14)

The results of this example are shown below. Table 4 shows the fit and SD scores for each of the damped sinusoids and the synthetic HRIRs. As can be seen the fits are very high (above 99%) and the SD scores are low (average of 0.1946 dB) which indicates an almost perfect reconstruction. Visually this is confirmed in the time domain plots (Figure 19) and magnitude responses (Figure 20). The next section describes the application of this process to real HRIRs and the evaluation of the results.

Table 3: Delay values.

<table>
<thead>
<tr>
<th>Signal Name</th>
<th>Delay (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{t1} )</td>
<td>0.00</td>
</tr>
<tr>
<td>( x_{t2} )</td>
<td>41.67</td>
</tr>
<tr>
<td>( x_{t3} )</td>
<td>20.83</td>
</tr>
<tr>
<td>( x_{t4} )</td>
<td>31.25</td>
</tr>
</tbody>
</table>

Table 4: Fit and SD score results.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Fit (%)</th>
<th>SD Score (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{t1} ) vs. ( \hat{x}_{t1} )</td>
<td>99.90</td>
<td>0.1462</td>
</tr>
<tr>
<td>( x_{t2} ) vs. ( \hat{x}_{t2} )</td>
<td>99.82</td>
<td>0.1868</td>
</tr>
<tr>
<td>( x_{t3} ) vs. ( \hat{x}_{t3} )</td>
<td>99.83</td>
<td>0.2239</td>
</tr>
<tr>
<td>( x_{t4} ) vs. ( \hat{x}_{t4} )</td>
<td>99.79</td>
<td>0.2217</td>
</tr>
<tr>
<td>( x ) vs. ( \hat{x} )</td>
<td>99.83</td>
<td>0.2848</td>
</tr>
</tbody>
</table>
Figure 18: The original damped sinusoids versus the estimated damped sinusoids.

Figure 19: Time domain plot of $x$ versus $\hat{x}$.

Figure 20: Magnitude response of $x$ versus $\hat{x}$.
3  Decomposition of HRIRs using the HTLS method

In this section the effectiveness of the HTLS decomposition method for the analysis of actual HRIRs, measured from a KEMAR mannequin and from 30 human subjects, is evaluated.

3.1 Decomposition of HRIRs Measured on a KEMAR Mannequin

The decomposition method from the previous section was first used to deconstruct HRIRs measured on a KEMAR mannequin using the AuSIM HeadZap system at a sampling frequency of 96 kHz [35]. HRIRs from azimuths of 90° and -90° and elevations from -20° to 20° at increments of 10° (see Figure 21) were measured for a total of 10 HRIRs for each ear. The application of the augmented HTLS method is targeted to the decomposition of Delayed and Damped Sinusoids (DDSs) separated by short latencies. Previous research [24, 27] indicates that HRIRs from the region covered by the azimuths and elevations chosen will likely present this challenging characteristic.

After applying the augmented HTLS decomposition method to the 20 HRIRs used in the test, the appropriateness of the results were addressed using our “fit” measure (Equations (3) and (4)) and the SD score (Equation (6)). It was noted that the values of these performance measures depend significantly on the location of the sound source. As an example, consider Figures 22 and 23, which show the time domain fits according to azimuth and elevation for the left and right ear respectively. As can be seen, in each case, the ipsilateral ear (the ear closest to the sound source) has high fits, while lower fits (below 90%) are only observed in the contralateral ear. The SD scores were used to see if this was the case in the frequency domain as well. However, the frequency range used for calculating the SD scores was restricted to a range of 0 kHz to 20 kHz. The reason for
this, as stated in section 2.4.1, is the knowledge that frequencies that are important to elevation localization (i.e., the pinna reflections) lie below 20 kHz. In Figures 24 and 25, the higher SD scores are only observed on the contralateral sides. This could be due to the fact that contralateral HRIRs have much lower Signal to Noise Ratio (SNR) values than ipsilateral HRIRs [6]. This may not be a critical shortcoming, as related research indicates that the contralateral ear does not contribute heavily to localization at certain angles [36], similar to those tested by us. Table 5 and Table 6 below show the means and standard deviations for the fits and SD scores.

It is of interest to examine the reconstructed HRIRs that achieved high and low fits and SD scores. Figures 26 and 27 show cases where a high fit in the time domain (96.81% and 96.41%, respectively) and low SD scores (3.4539 dB and 2.2851 dB, respectively) were achieved. On the other hand, Figures 28 and 29, show the case where a low fit in the time domain (78.27% and 84.42%, respectively) and high SD scores (5.0320 dB and 7.5900 dB, respectively) were achieved. This shows that the adequacy of the decomposition obtained is reflected simultaneously in a high fit percentage and a low SD score.

Table 5: Mean fits and SD scores (for elevations -20°, -10°, 0°, 10° and 20°)

<table>
<thead>
<tr>
<th>Ear</th>
<th>Azimuth (°)</th>
<th>Mean Fit (%)</th>
<th>Mean SD Score (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>90</td>
<td>94.09</td>
<td>2.7505</td>
</tr>
<tr>
<td>Left</td>
<td>-90</td>
<td>80.26</td>
<td>4.8926</td>
</tr>
<tr>
<td>Right</td>
<td>90</td>
<td>84.91</td>
<td>5.6568</td>
</tr>
<tr>
<td>Right</td>
<td>-90</td>
<td>95.97</td>
<td>3.4141</td>
</tr>
</tbody>
</table>

Table 6: Standard Deviations for the fits and SD scores.

<table>
<thead>
<tr>
<th>Ear</th>
<th>Azimuth (°)</th>
<th>Fit Standard Deviation (%)</th>
<th>SD Score Standard Deviation (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>90</td>
<td>2.7084</td>
<td>0.4459</td>
</tr>
<tr>
<td>Left</td>
<td>-90</td>
<td>7.5917</td>
<td>0.7218</td>
</tr>
<tr>
<td>Right</td>
<td>90</td>
<td>11.4173</td>
<td>1.2373</td>
</tr>
<tr>
<td>Right</td>
<td>-90</td>
<td>0.7863</td>
<td>1.1344</td>
</tr>
</tbody>
</table>
Figure 21: Coordinate system for HRIRs used in this paper.

Figure 22: Fit plots for the left ear.
Figure 23: Fit plots for the right ear.

Figure 24: SD score plots for the left ear.
Figure 25: SD score plots for the right ear.

Figure 26: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the left ear at azimuth = 90°, elevation = 20°.
Figure 27: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the right ear at azimuth = -90°, elevation = 0°.

Figure 28: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the left ear at azimuth = -90°, elevation = -10°.
Figure 29: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the right ear at azimuth = 90°, elevation = -20°.

3.2 Decomposition of Measured HRIRs from 30 Human Subjects

In this section we present the results obtained from the HTLS decomposition of HRIRs individually measured on actual human subjects. The decomposition method from Section 2.4 was used to deconstruct the 30 HRIRs measured using the AuSIM HeadZap system at a sampling frequency of 96 kHz. Azimuths of 90° and -90° and elevations from -18° to 18° at increments of 18° were measured for a total of 6 HRIRs for each ear. As stated in the previous subsection, the application of the augmented HTLS method is targeted to the decomposition of DDSs separated by short latencies, which are expected in the HRIRs from the region covered by the azimuths and elevations selected.

Additionally, the same “fit” measure (Equations (3) and (4)) and the SD score (Equation (6)) were used to evaluate the appropriateness of these results. The performance measures seemed to return similar results for both the HRIRs measured on humans and the mannequin. That is, the values of these performance measures seem to depend significantly on the location of the sound source. As an example, consider Figures
30 and 31, which show the mean and standard deviation of the time domain fits, according to azimuth and elevation, for the left and right ear respectively. A pattern similar to the time domain fit plots obtained in the previous section (Figures 22 and 23), which are for HRIRs measured on a KEMAR mannequin, is apparent in these plots as well. Again, it can be seen that, in each case, the ipsilateral ear (the ear closest to the sound source) has high fits, while lower fits (below 90%) are only observed in the contralateral ear. The SD scores, using the restricted frequency range (see previous section), were used to see if this was the case in the frequency domain as well. In Figures 32 and 33, the higher SD scores are only observed on the contralateral sides which are consistent with the findings from the previous section.

![Plot of the mean and standard deviation (dashed) of the fit for the left ear.](image)

Figure 30: Plot of the mean and standard deviation (dashed) of the fit for the left ear.
Figure 31: Plot of the mean and standard deviation (dashed) of the fit for the right ear.

Figure 32: Plot of the mean and standard deviation (dashed) of the SD scores for the left ear.
Figure 33: Plot of the mean and standard deviation (dashed) of the SD scores for the right ear.

In addition to showing the aggregated results for the whole set of 30 subjects tested, Figures 34 and 35 display the results obtained for a single human subject, considered “typical”. These results also confirm the general trends observed in the collective results.

Figure 34: Plot of the fit for subject RS.
4 Conclusion

We have presented a new method, created to address the challenging problem of decomposing signals into damped sinusoid components with onsets separated by relatively small delays. Other available methods might be applicable to the separation of Damped and Delay Sinusoids (DDSs) when the delays between them are much longer than the sampling interval used for their digitization. Similarly, there are existing methods that tackle the problem of decomposing a signal into damped sinusoids with a common onset. However, the scenario proposed by our structural pinna model, involving multiple damped sinusoids with onsets delayed by short latencies was not appropriately addressed by previous methods. We have, therefore, augmented the Hankel Total Least Squares (HTLS) method with an exhaustive zeroing process to make it possible to also estimate the short latencies expected in the decomposition of Head-Related Impulse Responses (HRIRs), particularly those for cases that correspond to a sound source placed at locations that neighbor the inter-aural axis.
The augmented HTLS decomposition method was first tested with a synthetic combination of damped and delayed sinusoids. The method was capable to estimate the parameters of the synthetic components, including the onset times, with very good accuracy. The augmented HTLS method was then used for the decomposition of real HRIRs and the results were evaluated by assessing the similarity of a reconstructed signal, formed by the estimated components, and the original signal, in both time and frequency. The results showed that the method is better suited for the decomposition of ipsilateral-ear HRIRs, which naturally have a larger amplitude and better SNR than contralateral-ear HRIRs. Nonetheless, time and frequency visualization of the reconstructed HRIRs for contralateral ears (e.g., Figures 28 and 29) show that the few components identified in the contralateral signals still reflect correctly the key temporal and spectral features of the HRIRs analyzed. Additionally, other researchers have proposed that the level of significance of the contralateral ear in source localization may be low.

It seems that the augmented HTLS decomposition method presented here will allow us to perform the decomposition of HRIRs of certain azimuth and elevation ranges (around the inter-aural axis) which have been a challenge for other methods. This, in turn, will enable us to develop a database of structural model parameters, matched to a database of anthropometric subject features, from which we will explore the systematic relationships between these two sets of data. Knowledge of these relationships may provide a means to instantiate structural model parameters for a new user from his/her anthropometric measurements.
5 Acknowledgments

This work was sponsored by NSF grants CNS-0520811, HRD-0833093 and CNS-0426125.

6 References


http://findarticles.com/p/articles/mi_m0EIN/is_2000_Nov_14/ai_66917801/


Figure and Caption List

Figure 1: Spherical coordinate system.

Figure 2: HRTF measurement system.

Figure 3: Estimate of azimuth of a sound source using ITD and IID.

Figure 4: Block diagram of structural model (after Brown and Duda [23]).

Figure 5: Example of how sound waves would bounce-off of the pinna into the ear canal.

Figure 6: Block diagram of structural pinna model (after Gupta [4]).

Figure 7: Block diagram of structural pinna model.

Figure 8: Example search tree diagram. A node in each level has been shaded to indicate a specific window succession that could result in the candidate HRIR that yields the best fit to the HRIR being decomposed.

Figure 9: Example of how the window width corresponds to the following sinusoid’s delay.

Figure 10: Example of how a windowed portion of a signal (left) is used to approximate a damped sinusoid (right).

Figure 11: $x$ (top) vs. $x_s$ (middle) and $x_p$ (bottom).

Figure 12: The time domain plots of the original damped sinusoids ($x_1$, $x_2$, $x_3$ and $x_4$) compared to the damped sinusoids that resulted from the HTLS method ($\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ and $\hat{x}_4$).

Figure 13: The time domain plot of $x$ and $\hat{x}$.

Figure 14: The frequency domain plots of $x$ (left) and $\hat{x}$ (right).

Figure 15: A delayed signal ($x_{t_2}$) versus the results from the HTLS method ($\hat{x}_{t_2}$).

Figure 16: Resulting fits from example.
Figure 17: $x_{i2}$ versus $\hat{x}_{i2}$ with initial samples zeroed.

Figure 18: The original damped sinusoids versus the estimated damped sinusoids.

Figure 19: Time domain plot of $x$ versus $\hat{x}$.

Figure 20: Magnitude response of $x$ versus $\hat{x}$.

Figure 21: Coordinate system for HRIRs used in this paper.

Figure 22: Fit plots for the left ear.

Figure 23: Fit plots for the right ear.

Figure 24: SD score plots for the left ear.

Figure 25: SD score plots for the right ear.

Figure 26: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the left ear at azimuth = 90°, elevation =20°.

Figure 27: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the right ear at azimuth = -90°, elevation =0°.

Figure 28: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the left ear at azimuth = -90°, elevation =-10°.

Figure 29: Time domain (left) and magnitude response (right) plots of original (solid) and reconstructed (dashed) HRIRs for the right ear at azimuth = 90°, elevation =-20°.

Figure 30: Plot of the mean and standard deviation (dashed) of the fit for the left ear.

Figure 31: Plot of the mean and standard deviation (dashed) of the fit for the right ear.

Figure 32: Plot of the mean and standard deviation (dashed) of the SD scores for the left ear.

Figure 33: Plot of the mean and standard deviation (dashed) of the SD scores for the right ear.
Figure 34: Plot of the fit for subject RS.

Figure 35: Plot of the SD scores for subject RS.