ESTIMATION OF MULTIPATH PROPAGATION DELAYS AND INTERAURAL TIME DIFFERENCES FROM 3-D HEAD SCANS

Hannes Gamper, Mark R. P. Thomas, Ivan J. Tashev

Microsoft Research, Redmond, WA 98052, USA

ABSTRACT

The estimation of acoustic propagation delays from a sound source to a listener’s ear entrances is useful for understanding and visualising the wave propagation along the surface of the head, and necessary for individualised spatial sound rendering. The interaural time difference (ITD) is of particular research interest, as it constitutes one of the main localisation cues exploited by the human auditory system. Here, an approach is proposed that employs ray tracing on a 3-D head scan to estimate and visualise the propagation delays and ITDs from a sound source to a subject’s ear entrances. Experimental results indicate that the proposed approach is computationally efficient, and performs equally well or better than optimally tuned parametric ITD models, with a mean absolute ITD estimation error of about 14 µs.

Index Terms— ITD, HRIR, HRTF, ray tracing

1. INTRODUCTION

The human auditory system is able to localise a sound source by analysing acoustic features encoded in the sound signals reaching the ear entrances. These acoustic features are introduced as a result of the sound being filtered by the listener’s head, torso, and pinnae. For a sound source in anechoic conditions, the filtering behaviour can be described in terms of a head-related impulse response (HRIR), or its frequency-domain counterpart, a head-related transfer function (HRTF). The analysis and synthesis of HRTFs is an active area of research aimed both at studying the way humans perceive spatial sound and at enabling high-fidelity spatial rendering.

HRTFs are naturally linked to the listener’s anthropometry and thus highly individual. This link is of ongoing research interest, fuelled in part by the availability of HRTF databases that include anthropometric data, for example the CIPIC database [1]. Prior work studied the mapping of anthropometric to acoustic features [2], anthropometry-based HRTF modelling and synthesis [3, 4, 5, 6], and “best match” HRTF selection [7, 8]. High-resolution 3-D scans of human subjects allow to synthesise individual HRTFs and to perform a detailed analysis of the relationship between morphology and acoustic filtering. Using finite difference time domain (FDTD) acoustic simulations of 3-D head scans, Mokhtari et al. investigated the acoustic behaviour of the pinna cavities [9]. Zolfaghari et al. studied how morphological changes of 3-D scans affect HRTFs obtained via a boundary element method (BEM) simulation [10].

Here, a computationally efficient method for estimating the propagation delays from a sound source to a subject’s ear entrances using a 3-D head scan is proposed. The spatial separation of the ear entrances and the shape of the head cause propagation delays to vary as a function of the source position. The difference between the propagation delays at the left and right ear entrances, i.e., the interaural time difference (ITD), serves as one of the main HRIR cues exploited by the human auditory system for sound source localisation [11]. Experiments by Kulkarni et al. indicate that the HRIR can be adequately modelled as a minimum-phase response plus a pure delay corresponding to the ITD [12], a result that can be applied to spatial sound rendering and HRTF interpolation [13]. For precise rendering, an accurate estimation of the listener’s ITD is crucial [14]. The ITD can be modelled as the propagation delay difference between two antipodal points on a rigid sphere. By adjusting the radius of the sphere, the model can be personalised according to the listener’s anthropometry [14]. Aaronson and Hartmann extended the spherical model to arbitrary ear angles and source distance [15]. Duda et al. proposed an ellipsoidal head model with adaptable radii and ear positions that outperformed the spherical head model in terms of the mean absolute ITD estimation error [16]. The performance of these parametric models suggests that the ITD can be modelled by estimating the direct path to the ipsilateral ear (i.e., the ear facing the source) and the shortest path from the source along the surface of the head to the contralateral ear (i.e., the ear pointing away from the source).

The estimation method proposed here estimates propagation delays and ITDs directly from a 3-D head scan, using acoustic ray tracing. 3-D head scans are becoming increasingly available due to the advent of low-cost 3-D scanning solutions. Ray tracing is a well-established tool for acoustic simulation that has recently found application in HRTF analysis [17, 18] and ITD modelling [15]. It provides an intuitive and computationally relatively efficient method of modelling acoustic wave propagation around the head. Here, ray tracing is used to estimate and visualise the propagation delays to the ears and compare estimated ITDs to acoustic measurements.
2. PROPOSED APPROACH

2.1. Propagation delay estimation

The proposed method for estimating propagation delays from a sound source position to the ear entrances relies on approximating the acoustic wave propagation via discrete rays. In the following, a subject’s 3-D head scan, centred on the interaural axis, and a sound source position, \( s \), are assumed. For the ipsilateral ear, \( e_i \), the propagation delay \( t_i \) is given as

\[
    t_i = c^{-1} \| s - e_i \|_2 = c^{-1} d_i,
\]

where \( c \) denotes the speed of sound, and \( d_i \) is the path length from the source to the ipsilateral ear entrance.

To estimate the propagation path length, \( d_c \), from the source to the contralateral ear, \( e_c \), ray tracing is employed. First, a set of discrete rays is cast from the source position, \( s \), towards the head. The ray directions, \( r_{k,m} \), with \( k \in \{1, ..., K\} \) and \( m \in \{1, ..., M\} \) are given as

\[
    r_{k,m} = R_{\alpha_k} R_{\beta_m} s_{\text{look}},
\]

where \( s_{\text{look}} \) is the look vector of the source (i.e., the negative source direction), and \( R_{\alpha_k} \) and \( R_{\beta_m} \) are rotation matrices that perform a rotation away from the look vector by angle \( \alpha \in \{0, ..., \alpha_{\text{max}}\} \), and about the look vector by angle \( \beta \in \{0, ..., 2\pi\} \), respectively. The resulting \( K \times M \) rays describe \( K \) cones centred on \( s_{\text{code}} \), with increasing radii, \( \alpha_k \). The upper bound, \( \alpha_{\text{max}} \), is chosen so that the largest cone fully contains the 3-D scan. Figure 1a shows the resulting ray–surface intersection points, \( p_{s,k,m} \). To estimate the shortest path from the source, \( s \), to the contralateral ear, \( e_c \), the ray–surface intersection points closest to the contralateral ear, \( p_{s,k_{\text{max}},m} \), are determined by calculating a distance metric \( w \):

\[
    w_{k,m} = \frac{p_{s,k,m}^T}{\| p_{s,k,m} \|_2} \cdot \frac{e_c}{\| e_c \|_2},
\]

where \( \cdot \) denotes the dot product (cf. Figure 1a, colour coding). The closest ray, \( k_{\text{max}} \), to the contralateral ear is found via

\[
    k_{\text{max},m} = \arg\max_k \left( w_{k,m} \right).
\]

Given a ray–surface intersection point, \( p_{s,k_{\text{max}},m} \), the shortest path along the surface of the scan to the contralateral ear, \( e_c \), is estimated as the straight path from \( p_{s,k_{\text{max}},m} \) to \( e_c \). The path is described via \( L \) node points, \( p_{c,l} \), with \( l \in \{1, ..., L\} \), along the surface of the head. The node points, \( p_{c,l} \), are determined by casting \( L \) rays from the centre of the interaural axis through the head surface. Figure 1b depicts the ray–surface intersections describing the path from \( p_{s,k_{\text{max}},m} \) to \( e_c \). As the intersection points lie on a plane, the resulting path along the surface can be fully described in two dimensions, \( x' \) and \( y' \), as shown in Figure 1c. Given a 2-D representation of the path along the surface, the wave propagation path is estimated as the line-of-flight propagation along the surface, i.e., the shortest convex path connecting \( p_{s,k_{\text{max}},m} \) and \( e_c \) that does not intersect the surface (see Figure 1c, dashed line). Assuming vectors \( x \) and \( y \) containing the \( x' \) and \( y' \) coordinates of the intersection points, \( p_{c,l} \), the node
Fig. 2. Propagation delay visualisation; a) path nodes across the surface of the head, for 1 source at 1 metre radius and -62 degrees lateral angle, 34 degrees polar angle; b-c) ITD estimations for 25 sources at 1 metre radius and polar angles ranging from -45 to 225 degrees, with a constant lateral angle of (b) -6 and (c) -39 degrees.

points of the convex path, \( p_{e, \text{convex}, j} \), with \( j \in \{1, \ldots, J \} \) and \( 2 \leq J \leq L \), are found iteratively. Set

\[
p_{e, \text{convex}, 1} = p_{e, 1} \tag{5}
\]

\[
p_{e, \text{convex}, J} = p_{e, L} \tag{6}
\]

\[
i_{\text{max}, 1} = 1. \tag{7}
\]

Then

\[
p_{e, \text{convex}, j+1} = p_{e, i_{\text{max}, j+1}}, \tag{8}
\]

where

\[
i_{\text{max}, j+1} = \arg \max_i \left( \frac{dy_i}{dx_i} \right), \tag{9}
\]

and

\[
dx_i = x_{i+1} - x_i, \tag{10}
\]

\[
dy_i = y_{i+1} - y_i. \tag{11}
\]

for \( i \in \{i_{\text{max}, j}, \ldots, L\} \). The length of the convex path, \( d_p \), can be estimated by a piece-wise linear approximation:

\[
d_p \approx \sum_{j=1}^{J-1} \sqrt{(x_{j+1} - x_j)^2 + (y_{j+1} - y_j)^2}. \tag{12}
\]

The estimation of \( d_p \) is performed for all \( M \) ray–surface intersection points, \( p_{s, k_{\text{max}}, m} \). The total propagation path length, \( d_{c, m} \), from the source to the contralateral ear can be estimated as

\[
d_{c, m} \approx d_{p, m} + \| s - p_{s, k_{\text{max}}, m} \|_2. \tag{13}
\]

The equivalent propagation delay, \( t_{c, m} \), is given as

\[
t_{c, m} = c^{-1} d_{c, m}. \tag{14}
\]

### 2.2. ITD estimation

The ITD, \( \tau \), can be expressed in terms of the propagation delays, \( t_l \) and \( t_r \), from the source position to the left and right ear, respectively, as

\[
\tau = t_l - t_r. \tag{15}
\]

Using the propagation delay estimation method proposed here, the ITD can be estimated as

\[
\hat{\tau} = t_i - t_{c, m_{\text{min}}} \tag{16}
\]

for a source to the left of the listener's median plane, and

\[
\hat{\tau} = t_{c, m_{\text{min}}} - t_i \tag{17}
\]

for a source to the right. The contralateral path delay, \( t_{c, m_{\text{min}}} \), is approximated as the minimum estimated delay, \( t_{c, m} \):

\[
m_{\text{min}} = \arg \min_m (t_{c, m}). \tag{18}
\]

### 3. EXPERIMENTAL EVALUATION

The performance of the proposed delay and ITD estimation approach was evaluated using a 3-D scan of the B&K 4128C head-and-torso simulator (HATS). The scan, consisting of over 300,000 triangles, was obtained using a Flexscan3D optical scanning setup. To improve the performance of the ray tracing algorithm, the triangle mesh was organised in an octree structure [19]. Ray–surface intersections were obtained using a fast ray–triangle intersection algorithm [20].

In the following, source directions are given in interaural coordinates, i.e., in terms of the lateral angle off the median plane and the polar angle describing the rotation about the interaural axis [21]. Figure 2a shows the estimated path lengths for 50 rays from a fixed source position to the contralateral
ear. As can be seen, the path lengths vary as a function of the angle, $\beta$, about the source look direction (cf. Figure 1a). Figure 2b and 2c illustrate the dependence of the ITD on the polar angle. At both -6 and -39 degrees lateral angle, the estimated ITD increases for sources above and behind the head, and appears to be minimal for sources in the front, where the shortest path to the contralateral ear passes along the subject’s face. These observations are in line with findings from prior work [16]. As seen in Figure 2b and 2c, the shortest paths from the tested source positions tend to lie on a straight line from the source to the contralateral ear.

To evaluate the accuracy of the ITD estimation algorithm, the ITDs of the HATS were estimated from the 3-D scan using the proposed method. As a ground-truth reference, ITDs were obtained from blocked-meatus HRIR measurements of the HATS for 400 source positions, using the hardware setup described by Bilinski et al. [4]. A parametric spherical head model [16] served as a baseline algorithm:

$$\hat{\tau} = c^{-1} r_{\text{opt}} (\varphi + \sin \varphi),$$

where $r_{\text{opt}}$ denotes the optimised sphere radius, and $\varphi$ the lateral angle of the source. The optimal radius, $r_{\text{opt}}$, was obtained by minimising the mean-squared estimation error (MSE) of the spherical model:

$$r_{\text{opt}} = \arg \min_r \left( \sum_n (\tau_n - \hat{\tau}_n(\varphi_n, \theta_n, r))^2 \right),$$

where $\tau_n$ and $\hat{\tau}_n$ are the measured and estimated ITD for the $n$-th source position, respectively.

As shown in Figure 3, the estimation accuracy of the proposed ray-tracing approach compares favourably to the baseline algorithm. Using the proposed algorithm, the total computation time for estimating the ITDs of 400 source positions, with $K = 25$, $M = 50$, and $L = 25$, was about 4 minutes on a PC with a 3.07 GHz quadcore processor with 12 GB RAM. Table 1 summarises the ITD estimation results. The performance of the baseline algorithm is in line with the results of prior work. The proposed ray-tracing method achieves a root-mean-squared error (RMSE) of about 19$\mu$s and a mean absolute error (MAE) of about 14$\mu$s, which constitutes an improvement of about 30% over the baseline algorithm. This improvement is partly due to the fact that the spherical model does not account for the dependence of the ITD on the polar angle. To put the errors into perspective, the human auditory system is able to resolve time delay differences as small as 10$\mu$s [11].

### 4. CONCLUSIONS

A method for estimating sound propagation delays from a source to a listener’s ear entrances was proposed. The estimation is performed using ray tracing on a 3-D head scan. The ITD is estimated as the difference between the direct path to the ipsilateral ear and the shortest path along the surface of the head to the contralateral ear. Experimental results indicate that the proposed method proves useful for analysing and visualising the multipath propagation delays from a source to the contralateral ear. Compared to an acoustic measurement of a head-and-torso simulator, the proposed method exhibited an absolute ITD estimation error of about 14$\mu$s, averaged over 400 source positions. The error rates achieved with the proposed method are equal or better than optimally tuned parametric models from the literature. Unlike the ground-truth optimised parametric models, the proposed method does not require prior knowledge of the measured ITDs. Future work includes the validation against 3-D head scans obtained from different subjects, and a study on the effect of the scan resolution and/or quality on the algorithm’s performance.
5. REFERENCES


