Continuous-direction model of the broadband time-of-arrival in the head-related transfer functions

Piotr Majdak* and Harald Ziegelwanger

*Corresponding author's address: Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, Wien, 1040, Wien, Austria, piotr@majdak.com

Head-related transfer functions (HRTFs) describe the filtering of the incoming sound by the human anatomy. They contain the so-called broadband time-of-arrivals (TOAs), which interaural differences yield the well-known interaural time differences used to estimate the lateral position of sound sources by the human auditory system. The TOAs are essential for a time-synchronous binaural rendering of multiple virtual sound sources or for interpolation of the timing information in an existing HRTF set. Estimation of the TOA is usually done separately for each spatial direction, and is thus, prone to errors and directional outliers. A method for a robust estimation of a continuous-direction TOA function from a set of listener-specific HRTFs is presented. The method relies on a geometric model of the HRTF-measurement setup represented by parameters like head position, radius, and ear position. The model parameters were fit to HRTFs of a sphere numerically calculated under various conditions. The resulting model parameters and TOA functions corresponded well to the measurement geometry and manually derived TOAs, respectively. The model was further evaluated in a setup assuming a listener placed off-axis in the HRTF-measurement setup, demonstrating the potential impact of the usually neglected aspect of listener position on the modeling results.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Head-related transfer functions (HRTFs) describe the acoustic filtering of the head, torso and pinna (Blauert, 1997; Möller et al., 1995). They provide cues like the interaural time and level differences (ITDs and ILDs) for the localization and externalization (Hartmann and Wittenberg, 1996) of virtual sound sources presented via headphones. The ITD is a relative quantity and thus sufficient for the description of an acoustic scene with a single sound source. For the description of multiple sound sources in a scene, the delay between the sound sources is important. Both, the ITD and the inter-source time difference can only be derived when the absolute delay arising from the acoustic wave propagation path from each source to a listener's ear is provided. Such a delay can be described by the time-of-arrival (TOA). For a given direction, TOA, τ, can be derived from the phase spectrum of an HRTF:

\[ \Phi(\omega) = \Phi_{\text{min}}(\omega) - j \omega \tau + \Phi_{\text{ap}}(\omega) \]  

where \( \omega \) is the circular frequency, \( \Phi \) the phase spectrum of an HRTF, \( \Phi_{\text{min}} \) the minimum phase spectrum in terms of a minimum-phase system (Oppenheim et al., 1999), \( \tau \) the time-of-arrival, and \( \Phi_{\text{ap}} \) the all-pass components of the phase spectrum.

The TOA can be approximated by the slope of a line fit to the phase spectrum (Huopaniemi and Smith, 1999; Jot et al., 1995). The TOA can also be derived from the group delay, i.e., the frequency derivative of the phase, by averaging the group delay over the frequency range of 1 to 5 kHz (Jot et al., 1995) or the frequency range of 0.5 to 2 kHz (Huopaniemi and Smith, 1999) or evaluating the group delay at a particular frequency of 0 Hz (Minnaar et al., 2000; Busson et al., 2005). The TOA can further be estimated from the impulse response of an HRTF, i.e., the head-related impulse response (HRIR). The simplest method is the amplitude threshold detection, where the TOA is the time where a pre-processed HRIR reaches a threshold for the first time (Algazi et al., 2002; Busson et al., 2005; Möller et al., 1995). In a similar way, the centroid of the HRIR can be defined as the TOA (Jeppesen and Möller, 2005; Jot et al., 1995; Minnaar et al., 2000).

An interesting TOA estimator has been proposed in (Nam et al., 2008; Wightman and Kistler, 2005) where the HRIR \( h \) and its minimum-phase version \( h_{\text{min}} \) were cross-correlated and the resulting maximum was evaluated:

\[ \tau = \text{argmax}_k \sum_{n=-\infty}^{\infty} h(n) h_{\text{min}}(n-k) \]  

with \( \tau \) as the estimated TOA and \( n \) and \( k \) as the sample indices corresponding to the discrete time.

The impact of the different TOA estimators can be visualized by applying the estimators to the HRTFs of an exemplary listener: 1) TOA at the maximum of the absolute HRIR (MAX; Möller et al., 1995); 2) TOA at the centroid of the HRIR (CTD; Jeppesen and Möller, 2005); 3) TOA as the average group delay of the HRTF, calculated from the negative frequency derivative of the phase, and averaged between 1 and 5 kHz (AGD; Jot et al., 1995); and 4) TOA at the maximum of the cross-correlation between the HRIR and its minimum phase version (MCM; Nam et al., 2008) according to Eq. 2. Fig. 1a shows the resulting TOAs for a sound source placed at 45\(^\circ\) to the left of the listener in the horizontal plane. It is evident that the resulting TOAs, and thus the ITDs, depend on the estimation method. Fig. 1b shows the resulting TOAs for the interaural horizontal plane. Even though the visual inspection is not always a reliable way to estimate the quality of an estimator, the directional glitches are evident and various directions might be identified as outliers. Those estimators consider each sound direction separately, and thus, can not automatically detect such outliers.

Thus, in this study, we propose a continuous-direction TOA model, which models the HRTF measurement setup in terms of the sound propagation paths around the human head as a function of direction. The model parameters are fit to an outlier-adjusted set of estimated TOAs which builds on a directional outlier detector. We evaluate the model on various sets of numerically calculated HRTFs. Further, we investigate the robustness of this model against potential translation of a listener in the HRTF measurement setup.

DIRECTIONAL OUTLIER REMOVAL

We use two coordinate systems for the directional representation of the data. In the geodesic coordinate system, the direction of a source is described by the azimuth \( \phi \) and elevation \( \theta \) angles (Majdak et al., 2007). A horizontal plane is selected by an elevation angle and the horizontal plane selected by \( \theta = 0^\circ \) is called the interaural horizontal plane.

In the horizontal-polar coordinate system (Morimoto and Aokata, 1984), the poles are placed on the interaural axis, i.e., a line parallel to the y-axis crossing the origin of the sphere, and the direction of a source is described by
the lateral $\Phi$ and the polar $\Theta$ angles. In this coordinate system, a sagittal plane is selected by a lateral angle. This coordinate system describes the human sound-localization mechanisms in a more natural way because the lateral direction is mostly associated with binaural cues and polar direction is mostly associated with spectral cues.

Let $P$ describe the set of all available spatial directions in an HRTF set of a listener. Outliers are defined and removed from $P$ in order to form the outlier-adjusted set $\hat{S}$. First, directions leading to a smooth TOA as a function of the azimuth angle are detected. All directions in $P$ are grouped to $G_0$ where the directions are in proximity $\pm \epsilon$ of a horizontal plane described by the elevation angle $\theta$. In each $G_0$, the TOA slope $\tau'_{\phi, G_0}$ is calculated:

$$\tau'_{\phi, G_0} = \frac{\tau_{\phi, G_0} - \tau_{\phi + \Delta\phi, G_0}}{\Delta\phi},$$

where $\tau_{\phi, G_0}$ is the TOA of a direction in $G_0$ and $\Delta\phi$ is the azimuth difference between two laterally consecutive directions in $G_0$. Let $R$ contain only directions resulting in smooth TOA slopes in each $G_0$ in terms of slopes smaller than the root-mean-square TOA of the corresponding $G_0$:

$$|\tau_{\phi, 0}| < \sqrt{\frac{1}{M_{\phi, 0}} \sum_{\phi, 0} (\tau'_{\phi, 0})^2} \quad \Rightarrow \quad P_{\phi, 0} \in R.$$

Further, directions with a similar TOA in a sagittal plane are detected. To this end, directions in $R$ are grouped to $G_\phi$ such that all directions in a group $G_\phi$ are in proximity $\pm \epsilon$ of a sagittal plane described by the lateral angle $\Phi$. In each group $G_\phi$, the average TOA $\mu_{G_\phi}$ and the TOA variance $\sigma^2_{G_\phi}$ are calculated. Further the average TOA variance $\overline{\sigma^2_{G_\phi}}$ is calculated by averaging $\sigma^2_{G_\phi}$ over all $G_\phi$. Let $S$ consist only of directions with a similar TOA in each $G_\phi$ in terms of the squared difference between TOA and the average TOA in that group being below $\overline{\sigma^2_{G_\phi}}$:

$$\left[ \tau_{\phi, 0} - \mu_{G_\phi} \right]^2 < \overline{\sigma^2_{G_\phi}} \quad \Rightarrow \quad R_{\phi, 0} \in S.$$

$S$ is considered as the outlier-adjusted set of all directions in $P$ and a TOA for a direction included in $S$ is denoted as $\hat{\tau}_S$. Fig. 2a shows the TOAs for the interaural horizontal plane calculated with the MCM method for the left-ear HRTFs of the same listener as in Fig. 1. The solid line shows all available directions (set $P$). Triangles showing down denote the outliers, i.e., directions which were removed from $P$ according to Eq. 5 with $\epsilon = 2.5^\circ$ and Eq. 4 with $\epsilon = 1^\circ$.

**CONTINUOUS-DIRECTION TOA MODEL**

The directional TOA model aims at describing the geometrical HRTF measurement setup. To this end, the propagation path between a sound source and the listener's ear is modeled, corresponding to the sound propagation path from a loudspeaker to the in-ear-canal microphone in the HRTF measurement. Listener's head is modeled as a rigid
sphere and TOA is calculated as the time required to travel along the propagation path for a given direction. The listener’s head is assumed to be placed exactly in the center of the loudspeaker array. Thus, the sphere center and the measurement center are assumed to be coincident (Fig. 2b). Note the arbitrary position of the ear on the sphere, described by $\phi_e$ and $\theta_e$. The TOA is split in a direction-independent and direction-dependent part. The direction-independent TOA $\tau$ is the propagation path between a loudspeaker and the closest point on the sphere, and is modeled as $\tau = \min |\tau_s|$. For the direction-dependent TOA $\tilde{\tau}$, and its corresponding propagation path, two cases are considered depending on the angular distance $\alpha$ between the sound source and the ear.

For ipsilateral sound sources, $\alpha \leq \frac{\pi}{2}$, the direction-dependent propagation path $s_i$ is:

$$s_i = r \left[ \sin(\theta_i) \sin(\phi_e) \cos(\theta_e) \cos(\phi_e - \phi) \right] \quad \text{for} \quad \alpha \leq \frac{\pi}{2}, \quad (6)$$

with $\phi_e$ and $\theta_e$ the position of the ear on the sphere and $r$ the sphere-radius.

For contralateral sound sources, $\alpha > \frac{\pi}{2}$, the direction-dependent propagation path additionally considers the scattering path around the sphere, and is:

$$s_i = r \left[ 1 + \cos^{-1}(\sin(\theta_i) \sin(\phi_e) \cos(\theta_e) \cos(\phi_e - \phi)) - \frac{\pi}{2} \right] \quad \text{for} \quad \alpha > \frac{\pi}{2}. \quad (7)$$

With $c$ as the speed of sound, the TOA is then

$$\tilde{\tau} = \frac{s_i}{c} + \tau. \quad (8)$$

In order to obtain the model parameters $\tilde{\tau}$, $r$, $\phi_e$, and $\theta_e$, the equations are fitted to the outlier-adjusted TOA set $S$. The fit is done by minimizing the squared error between $\tilde{\tau}$ and $\tilde{\tau}_S$.

**EVALUATION OF THE MODEL**

HRTFs were numerically calculated for a rigid sphere. The geometrical objects were constructed as icosphere meshes and consisted of 10,242 nodes and 20,480 triangular elements with an average edge length of 3 mm. The sound pressure in the simulated free-field was calculated with the Burton-Miller collocation boundary-element method coupled with the multi-level fast-multipole method (for more details see Kreuzer et al., 2009). Complex spectra were calculated for 100 frequencies linearly spaced between 0.2 and 20 kHz. The principle of reciprocity (Morse and Ingard, 1987) was used to decrease the computational costs in the calculation of the spectra for 1550 directions at a 9-m distance. The directions were in the elevation between -30° to 80° in steps of 5°. The azimuthal range was 360° in steps of 2.5°. Finally, the HRIRs were obtained by applying the inverse Fourier transform on the complex spectra and resampling to 48 kHz.
HRTF sets were calculated for combinations of sphere radius (77.5, 87.5, and 97.5 mm) and positions of the virtual microphones on the sphere. The parameter combinations are shown in Fig. 3 as lines. For each HRTF set, TOAs were estimated using one of the TOA estimator (MAX, CTD, AGD, MCM). Outlier-adjusted sets of directions were derived from the estimated TOAs. Continuous-direction TOA models were fit to the outlier-adjusted sets. The modeled radius overestimated the actual sphere radius, particularly by 4.26 % (MAX), 4.68 % (CTD), 4.89 % (AGD), and 2.75 % (MCM). The average radius error was 3 mm (MAX), 4 mm (CTD), 4 mm (AGD) and 2 mm (MCM). The error of the modeled ear positions was on average 1.3° and 0.6° for the azimuth and elevation angle, respectively. The details on the resulting model parameters for the MCM estimator (as the most accurate one) are shown in Fig. 3a. The comparison shows a good correspondence between the actual (lines) and model (symbols) parameters.

Further, the effect of a translation of the listener (i.e., an off-axis position) during the HRTF measurement was investigated. To this end, the model was applied to HRTFs simulated for spheres placed 0, 10, and 20 mm to the left of the center. Fig. 3b shows the parameters, in particular the lateral offset, $y_{ol}$, used for the simulations. The radius for the left and the right ears resulting from the model are also shown in Fig. 3b. With the increasing lateral offset from 0 to 20 mm, the radius errors increased from 0 to 18 mm. As a consequence, the interaural radius difference increased, showing a modeled asymmetry between the two ears. This finding suggests that a lateral offset of a listener during the measurement might cause an interaural asymmetry in the TOAs and the model parameters. Note that ITDs, being relative quantities, are not affected by the off-axis position of the listener during the HRTF-measurement system.

CONCLUSIONS

In this study, a model for a continuous-direction monaural broadband TOA was proposed. The TOA model is based on a geometric description of the HRTF measurement setup and assumes an exact placement of the listener in the measurement center. For each direction, the monaural broadband TOA from an HRTF is estimated by one of the already published estimators. Basing on the assumptions of a smooth TOA in horizontal planes and a nearly constant TOA in sagittal planes, a directional outlier detector has been introduced. The detector results in the outlier-adjusted set of directions which is used to fit the TOA model.

The TOA model was evaluated for simulated HRTFs of a rigid sphere. Generally, the evaluation showed that the model was able to accurately reconstruct the geometrical parameters used for the simulations. In particular, four TOA estimators (MAX, CTD, AGD, MCM) were tested and the MCM estimator showed best results.

The TOA model was further applied to HRTFs simulated for an off-axis-placed sphere. In the off-axis conditions, the model resulted in large interaural radius differences, which should be zero even when separately modeling the TOAs for the left and right ear. We conclude that an off-axis position of the listener might contribute to a fail of the proposed (on-axis) TOA model when applied to realistic HRTF measurements. Hence, in an accompanying study, we propose an off-axis TOA model which is able to represent measurement setups with an off-axis placement of the listener (Ziegelwanger and Majdak, submitted to JASA).
ACKNOWLEDGMENTS

This study was supported by the Austrian Science Fund (FWF, P 24124-N13). We thank Wolfgang Kreuzer for the support on HRTF simulations and Michael Mihocic for the work on the HRTF database.

REFERENCES


