

Using HRTFs for Virtual Source Positioning

Franz Zotter

May 20, 2018

Why? - Examples

To understand why we want to deal with headphone-based signal recording and rendering in general, there are some very good examples found here:

<https://www.bbc.co.uk/rd/projects/binaural-broadcasting>

The video demonstrates *binaural* recording with a dummy head (not yet explicitly using HRTF filtering). The link to the BBC production/game *The Turning Forest* might coin another essential piece of motivation for constructing artificial *binaural* worlds using HRTFs:

<https://www.youtube.com/watch?v=8YEKXUPMv18>

Another example. The award-winning pieces submitted to Europe's First 3D Audio Production Competition are also available without the big hall they were performed in: for headphones

<http://competition.ambisonics.at/>

Whenever we construct signals for virtual worlds (virtual reality or augmented reality VR/AR), use virtualized reinforcement systems and halls to present sound in, which is most of the time done for headphones, we will be using HRTFs for virtual source positioning.

Expressions and Concepts Involved

Virtual source describes the computational model of an auditory event (resulting object of our perception), whose position can be controlled within a signal processing model. If everything works ideally, the resulting auditory event location closely follows the computer-controlled virtual source position.

HRTFs abbreviate *head-related transfer functions*. When presenting a mono signal filtered by a pair of HRTFs (left and right, Fig. 1) on headphones, we aim at creating an auditory event localized in the outside world, outside of the head, outside of the headphones.

HRTFs express both time of arrival and coloration of the ear signals due to diffraction/shadowing on the human head, in the frequency domain. They are direction-dependent, come in pairs (left and right), and are called HRIRs when regarded in the time domain (head-related impulse responses).

Binaural technology refers to headphone-based presentation of outside-world signals (binaural = two-eared).

BRIRs are binaural room impulse responses. They contain early reflections and reverberation of the room that can help perceiving external sound.

For further reading, e.g., the book by Bosun Xie can be recommended [Xie13] or [Maj14] as reference to further advanced literature.

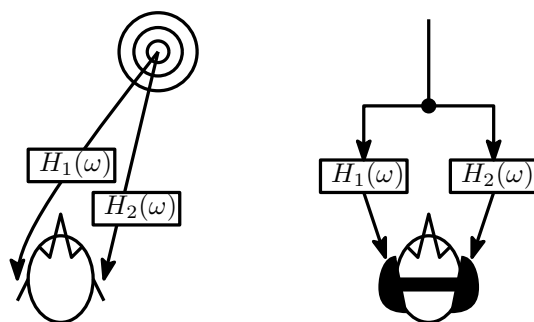


Figure 1: HRTFs in nature and as filters in binaural rendering

How do sound waves arrive at the ears?

Delayed (sphere model)

If a sound is coming from the left, it will reach the left ear earlier than the right one, vice versa.

Assuming the radius $d = 8$ cm of a spherical head, the speed of sound $c = 343 \frac{\text{m}}{\text{s}}$, and the angle φ to express the lateral angle enclosed to the left ear ($\varphi = 0$ left, $\varphi = \frac{\pi}{2}$ median plane, $\varphi = \pi$ right) and ϑ as polar angle with regard to the frontal horizontal plane ($\vartheta = 0$ horizon in front, $\frac{\pi}{2}$ above, $-\frac{\pi}{2}$ and $\frac{3\pi}{2}$ below, π horizon in the back), we could roughly model the time of arrival (TOA) to the left ear as

$$\tau_L(\varphi) = -\frac{a}{c} \cos \varphi \cos \vartheta, \quad (1)$$

assuming the two ears are just two points at $x = \pm a$, $y = 0$, $a = 0.875$ cm, picking up signals from a plane wave from φ arriving at the origin with no delay. And the right ear TOA is $\tau_R(\varphi) = \tau_L(\pi - \varphi)$ because of symmetry with regard to the median plane. More elaborated, we could consider a rigid sphere, requiring to use the arc length for the distant ear

$$\tau_L(\varphi) = \frac{r}{c} \cos \vartheta \begin{cases} -\cos \varphi, & 0 \leq \varphi \leq \frac{\pi}{2}, \\ \varphi - \frac{\pi}{2}, & \frac{\pi}{2} \leq \varphi \leq \pi, \end{cases} \quad (2)$$

instead of the direct path. The corresponding difference of the arrival times, the interaural time difference $ITD = \tau_R - \tau_L$, yields the Woodworth Schlosberg formula

$$ITD(\varphi) = \frac{a}{c} \left[\cos \varphi + \frac{\pi}{2} - \varphi \right] \cos \vartheta. \quad (3)$$

While detecting the interaural time delays (ITD) is the main mechanism of sound localization in our ear, this does not explain how we can distinguish sounds coming from the front and from the back, or how we distinguish sounds from different elevations.

Level-changed above 500 kHz (sphere model)

At frequencies above 1 kHz, there will be a level increase for the near, and a level decrease for the distant ear, because of the head reflection and shadowing. The level dependency could be roughly modeled using a high-pass filter at 1500 Hz and some compressed dependency on $\cos \varphi$

$$H_L(f) = 1.15 + 0.85 \frac{i f / 1500 \text{ Hz}}{1 + i f / 1500 \text{ Hz}} \sin \left[0.9 \frac{\pi}{2} \cos \varphi \right]. \quad (4)$$

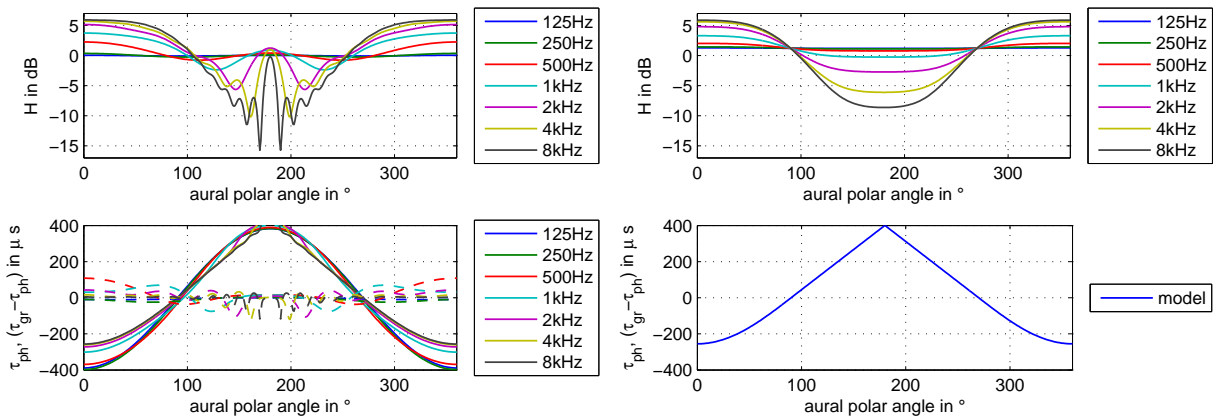


Figure 2: Magnitude response and phase delay at one ear of a physical sphere diffraction model depending on different angles (left), and above-mentioned models thereof for τ_L and H_L above (right).

Spectrally complex above 3 kHz (pinna model)

Researchers had been considering to involve later signal arrivals due to reflections on two different rings and rims of our pinna. However, our pinna is highly individual. The changes appear at frequencies above 3 kHz and are difficult to model. It can be easier to measure a dummy head's HRIRs or to measure individual HRIRs of a person.

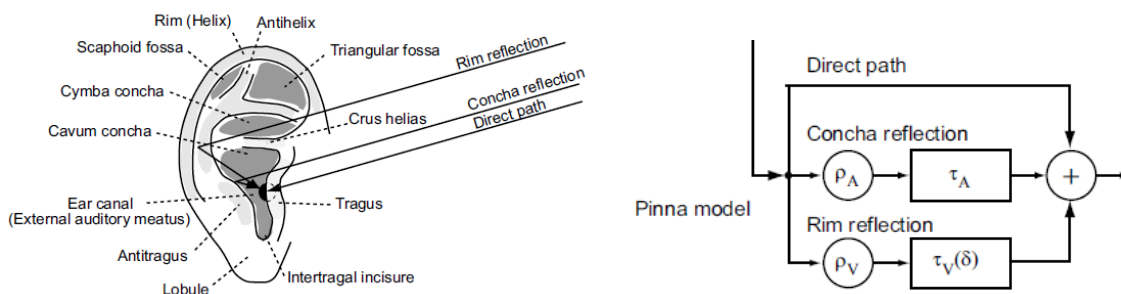


Figure 3: Model of the spectral details of the pinna according to [SAD07]

Measured HRIRs: KU 100 (SOFA file format)

Sets of left and right HRIRs of individual persons or dummy heads with dense directional sampling can be found online, e.g., here:

<http://sofaconventions.org>

For musical applications, it is often preferable to use the measurements of the Neumann dummy head KU 100, measured at the TH Köln (THK) that can be done with the most dense sampling without gaps and the least measurement noise

<http://audiogroup.web.th-koeln.de/ku100hrir.html>

for instance the one with 2° resolution in azimuth and Gauss sampling points in elevation *HRIR_FULL2DEG.sofa*, or the horizon with all 360° resolved in 1° in *HRIR_CIRC360.sofa*. Also headphone equalization filters are available for various headphones on this website *HPCF.zip*.

The left and right HRIRs of the KU100 for horizontal directions are plotted in Fig. 4, and one can readily recognize the delay pattern modeled by τ_L in the plots, and some other details. Vertically in the median plane, there is not as much difference between left and right Fig. 5, however the details of the HRIRs are observed to vary.

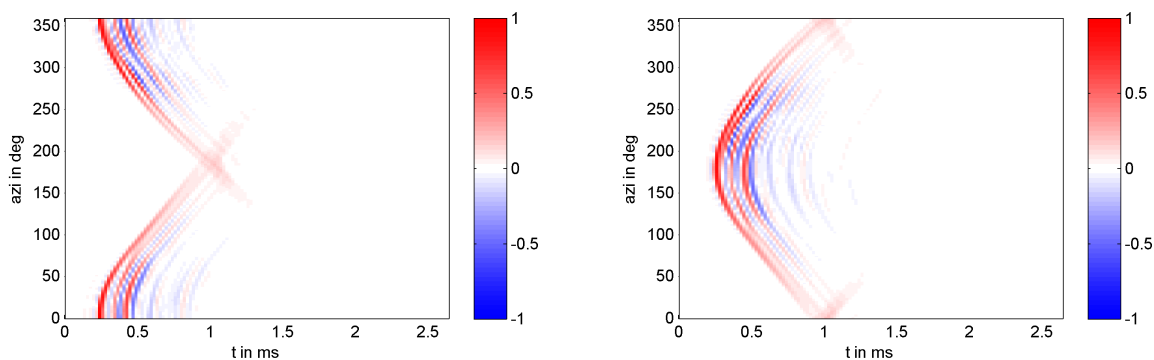


Figure 4: Left and right HRIRs (horizontal cross section) of KU100 dummy head depending on time and azimuth.

The HRTF plot (over frequency) Fig. 6 in the median plane shows distinct resonances for 8kHz at 90° elevation, one of Blauert's direction determining bands. It also contains distinct resonances for the front directions and notches for back directions above 8kHz that help to distinguish front and back.

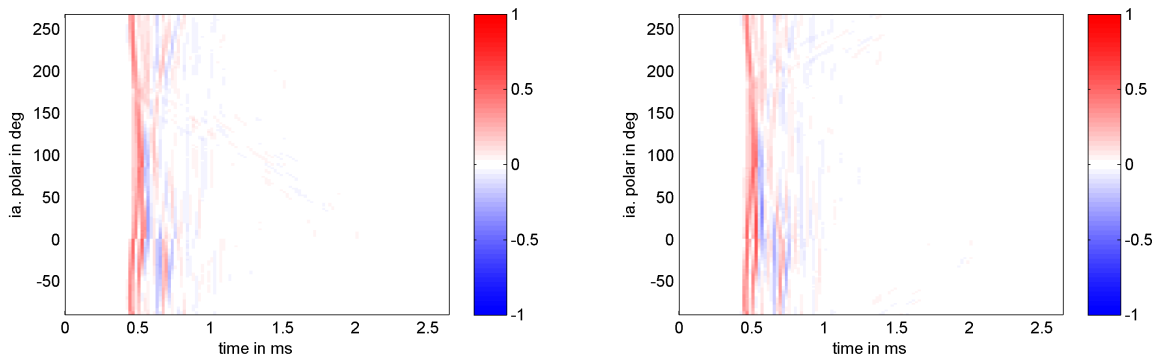


Figure 5: Left and right HRIRs (median plane vertical cross section) of KU100 dummy in time and elevation.

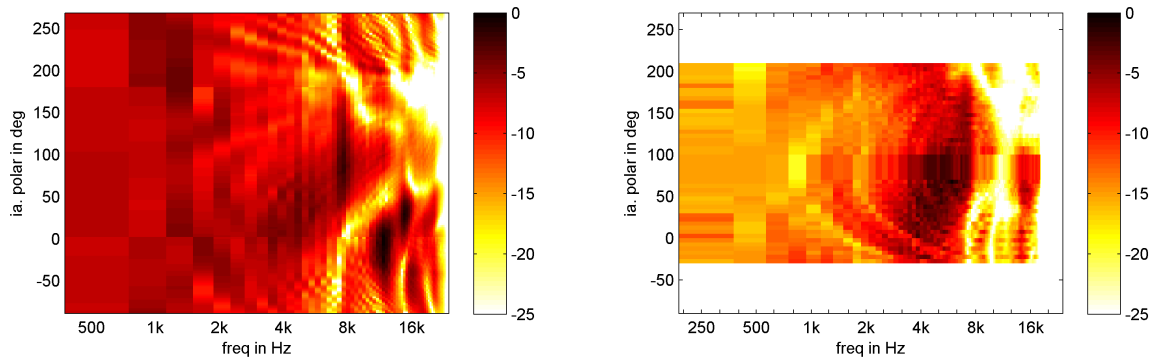


Figure 6: Right HRTF (median plane vertical cross section) of KU100 dummy in frequency and elevation, F.Zotter's left HRTF (measured by ARI in Vienna).

Implementations in Real-Time Signal Processing

Pd example

Reaper Session: psypan, BinAural, AMBEO Orbit, ambix_encoder_o5+BinauralDecoder, RoomEncoder+BinauralDecoder
 Read open source code by Tomasz Woźniak (BinAural), Daniel Rudrich (BinauralDecoder) to get prepared for own action(?)...

How to interpolate between HRIRs?

-switching a highly-resolved set ($\leq 5^\circ$) Hm... comb filtering

Tune-In 3D

IEM BinauralDecoder / TAC

Why does it not externalize well?

Room/anechoic

indiv/noniniv

(Begault, Völk, Werner, Hassager)

Room Divergence

(Werner,Sinker,...)

Advanced things

Bernschütz's HpTF equalizers for KU100

individual HRTF measurement (Majdak,Masiero,Brinkmann)

hefio individual HpTF equalization (Hippaka, PhD Thesis)
HRIR individualization/models with MRT or photometry (Jin, 3dsoundlabs,...)

References

- [Maj14] P. Majdak, “Head-related transfer functions and their role in the localization of sound sources in human listeners,” Ph.D. dissertation, 2014.
- [SAD07] P. Satarzadeh, V. R. Algazi, and O. Duda, “Physical and filter pinna models based on anthropometry,” in *prepr. 122nd AES Conv.*, Vienna, 2007.
- [Xie13] B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. J. Ross Publishing Inc., 2013.