Head-related transfer function

A **head-related transfer function (HRTF)** is a response that characterizes how an ear receives a sound from a point in space; a pair of HRTFs for two ears can be used to synthesize a binaural sound that seems to come from a particular point in space. It is a transfer function, describing how a sound from a specific point will arrive at the ear (generally at the outer end of the auditory canal). Some consumer home entertainment products designed to reproduce surround sound from stereo (two-speaker) headphones use HRTFs. Some forms of HRTF-processing have also been included in computer software to simulate surround sound playback from loudspeakers.

HRTF filtering effect

Humans have just two ears, but can locate sounds in three dimensions – in range (distance), in direction above and below, in front and to the rear, as well as to either side. This is possible because the brain, inner ear and the external ears (pinna) work together to make inferences about location. This ability to localize sound sources may have developed in humans and ancestors as an evolutionary necessity, since the eyes can only see a fraction of the world around a viewer, and vision is hampered in darkness, while the ability to localize a sound source works in all directions, to varying accuracy,[1] regardless of the surrounding light.

Humans estimate the location of a source by taking cues derived from one ear (*monaural cues*), and by comparing cues received at both ears (*difference cues* or *binaural cues*). Among the difference cues are time differences of arrival and intensity differences. The monaural cues come from the interaction between the sound source and the human anatomy, in which the original source sound is modified before it enters the ear canal for processing by the auditory system. These modifications encode the source location, and may be captured via an impulse response which relates the source location and the ear location. This impulse response is termed the *head-related impulse response* (HRIR). Convolution of an arbitrary source sound with the HRIR converts the sound to that which would have been heard by the listener if it had been played at the source location, with the listener's ear at the receiver location. HRIRs have been used to produce virtual surround sound.[2][3]

The HRTF is the Fourier transform of HRIR. The HRTF is also sometimes known as the *anatomical transfer function* (ATF).

HRTFs for left and right ear (expressed above as HRIRs) describe the filtering of a sound source ($x(t)$) before it is perceived at the left and right ears as $x_L(t)$ and $x_R(t)$, respectively.

The HRTF can also be described as the modifications to a sound from a direction in free air to the sound as it arrives at the eardrum. These modifications include the shape of the listener's outer ear, the shape of the listener's head and body, the acoustic characteristics of the space in which the sound is played, and so on. All these characteristics will influence how (or whether) a listener can accurately tell what direction a sound is coming from.

In the AES69-2015 standard,[4] the Audio Engineering Society (AES) has defined the SOFA file format for storing spatially oriented acoustic data like head-related transfer functions (HRTFs). SOFA software libraries and files are collected at the Sofa Conventions website.[5]
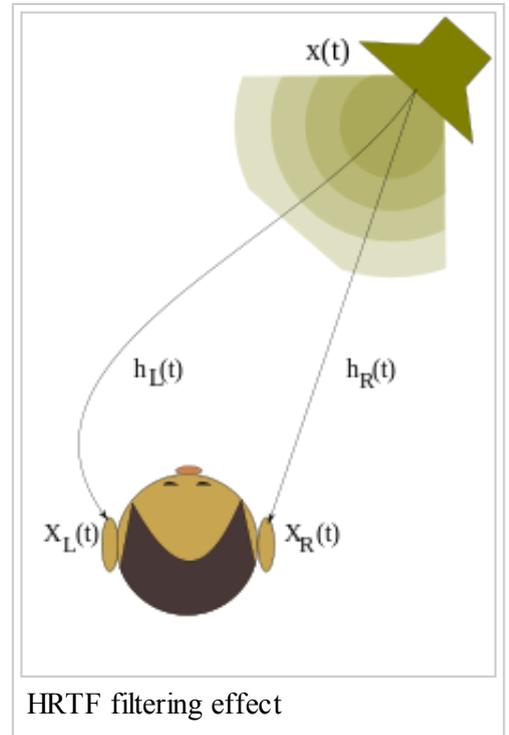
Contents

How HRTF works

The associated mechanism varies between individuals, as their head and ear shapes differ.

HRTF describes how a given sound wave input (parameterized as frequency and source location) is filtered by the diffraction and reflection properties of the head, pinna, and torso, before the sound reaches the transduction machinery of the eardrum and inner ear (see auditory system). Biologically, the source-location-specific prefiltering effects of these external structures aid in the neural determination of source location), particularly the determination of the source's elevation (see vertical sound localization).[6]

Technical derivation

Linear systems analysis defines the transfer function as the complex ratio between the output signal spectrum and the input signal spectrum as a function of frequency. Blauert (1974; cited in Blauert, 1981) initially defined the transfer function as the free-field transfer function (FFTF). Other terms include free-field to eardrum transfer function and the pressure transformation from the free-field to the eardrum. Less specific descriptions include the pinna transfer function, the outer ear transfer function, the pinna response, or directional transfer function (DTF).

The transfer function $H(f)$ of any linear time-invariant system at frequency f is:

$$H(f) = \text{Output}(f) / \text{Input}(f)$$

One method used to obtain the HRTF from a given source location is therefore to measure the head-related impulse response (HRIR), $h(t)$, at the ear drum for the impulse $\Delta(t)$ placed at the source. The HRTF $H(f)$ is the Fourier transform of the HRIR $h(t)$.

Even when measured for a "dummy head" of idealized geometry, HRTF are complicated functions of frequency and the three spatial variables. For distances greater than 1 m from the head, however, the HRTF can be said to attenuate inversely with range. It is this far field HRTF, $H(f, \theta, \varphi)$, that has most often been measured. At closer range, the difference in level observed between the ears can grow quite large, even in the low-frequency region within which negligible level differences are observed in the far field.
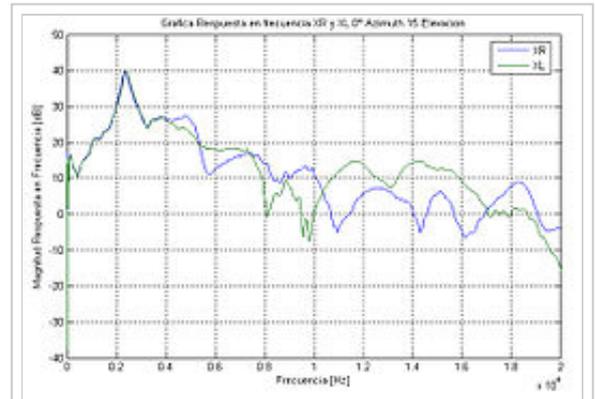
HRTFs are typically measured in an anechoic chamber to minimize the influence of early reflections and reverberation on the measured response. HRTFs are measured at small increments of θ such as 15° or 30° in the horizontal plane, with interpolation used to synthesize *HRTF*s for arbitrary positions of θ. Even with small increments, however, interpolation can lead to front-back confusion, and optimizing the interpolation procedure is an active area of research.

In order to maximize the signal-to-noise ratio (SNR) in a measured HRTF, it is important that the impulse being generated be of high volume. In practice, however, it can be difficult to generate impulses at high volumes and, if generated, they can be damaging to human ears, so it is more common for HRTFs to be directly calculated in the frequency domain using a frequency-swept sine wave or by using maximum length sequences. User fatigue is still a problem, however, highlighting the need for the ability to interpolate based on fewer measurements.

The head-related transfer function is involved in resolving the Cone of Confusion, a series of points where ITD and ILD are identical for sound sources from many locations around the "0" part of the cone. When a sound is received by the ear it can either go straight down the ear into the ear canal or it can be reflected off the pinnae of the ear, into the ear canal a fraction of a second later. The sound will contain many frequencies, so therefore many copies of this signal will go down the ear all at

A sample of frequency response of ears:

- *green curve*: left ear $X_{\mathrm{L}}(f)$
- *blue curve*: right ear $X_{\mathrm{R}}(f)$

for a sound source from upward front.

An example of how the HRTF tilt with azimuth taken from a point of reference is derived

different times depending on their frequency (according to reflection, diffraction, and their interaction with high and low frequencies and the size of the structures of the ear.) These copies overlap each other, and during this, certain signals are enhanced (where the phases of the signals match) while other copies are canceled out (where the phases of the signal do not match). Essentially, the brain is looking for frequency notches in the signal that correspond to particular known directions of sound.

If another person's ears were substituted, the individual would not immediately be able to localize sound, as the patterns of enhancement and cancellation would be different from those patterns the person's auditory system is used to. However, after some weeks, the auditory system would adapt to the new head-related transfer function.[7] The inter-subject variability in the spectra of HRTFs has been studied through cluster analyses.[8]
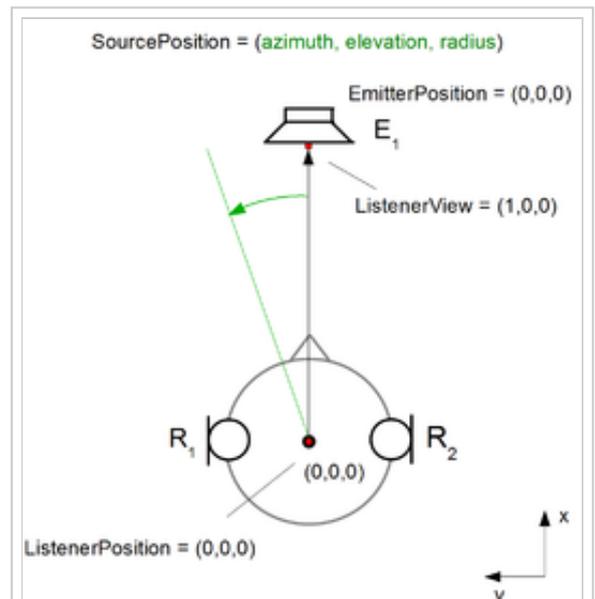
Assessing the variation through changes between the person's ear, we can limit our perspective with the degrees of freedom of the head and its relation with the spatial domain. Through this, we eliminate the tilt and other co-ordinate parameters that add complexity. For the purpose of calibration we are only concerned with the

direction level to our ears, ergo a specific degree of freedom. Some of the ways in which we can deduce an expression to calibrate the HRTF are:

1. Localization of sound in Virtual Auditory space[9]
2. HRTF Phase synthesis[10]
3. HRTF Magnitude synthesis[11]

Localization of sound in virtual auditory space[9]

A basic assumption in the creation of a virtual auditory space is that if the acoustical waveforms present at a listener's eardrums are the same under headphones as in free field, then the listener's experience should also be the same.

Typically, sounds generated from headphones appear to originate from within the head. In the virtual auditory space, the headphones should be able to "externalize" the sound. Using the HRTF, sounds can be spatially positioned using the technique described below.

Let $x_1(t)$ represent an electrical signal driving a loudspeaker and $y_1(t)$ represent the signal received by a microphone inside the listener's eardrum. Similarly, let $x_2(t)$ represent the electrical signal driving a headphone and $y_2(t)$ represent the microphone response to the signal. The goal of the virtual auditory space is to choose $x_2(t)$ such that $y_2(t) = y_1(t)$. Applying the Fourier transform to these signals, we come up with the following two equations:

$Y_1 = X_1 LFM$, and
$Y_2 = X_2 HM$,

where L is the transfer function of the loudspeaker in the free field, F is the HRTF, M is the microphone transfer function, and H is the headphone-to-eardrum transfer function. Setting $Y_1 = Y_2$, and solving for X_2 yields

$X_2 = X_1 LF/H$.

By observation, the desired transfer function is

$T = LF/H$.

Therefore, theoretically, if $x_1(t)$ is passed through this filter and the resulting $x_2(t)$ is played on the headphones, it should produce the same signal at the eardrum. Since the filter applies only to a single ear, another one must be derived for the other ear. This process is repeated for many places in the virtual environment to create an array of head-related transfer functions for each position to be recreated while ensuring that the sampling conditions are set by the Nyquist criteria.

HRTF phase synthesis[10]

There is less reliable phase estimation in the very low part of the frequency band, and in the upper frequencies the phase response is affected by the features of the pinna. Earlier studies also show that the HRTF phase response is mostly linear and that listeners are insensitive to the details of the interaural phase spectrum as long as the interaural time delay (ITD) of the combined low-frequency part of the waveform is maintained. This is the modeled phase response of the subject HRTF as a time delay, dependent on the direction and elevation.

A scaling factor is a function of the anthropometric features. For example, a training set of N subjects would consider each HRTF phase and describe a single ITD scaling factor as the average delay of the group. This computed scaling factor can estimate the time delay as function of the direction and elevation for any given individual. Converting the time delay to phase response for the left and the right ears is trivial.

The HRTF phase can be described by the ITD scaling factor. This is in turn is quantified by the anthropometric data of a given individual taken as the source of reference. For a generic case we consider β as a sparse vector

$$\beta = [\beta_1, \beta_2, \ldots, \beta_N]^T$$

that represents the subject's anthropometric features as a linear superposition of the anthropometric features from the training data ($y' = \beta^T X$), and then apply the same sparse vector directly on the scaling vector H. We can write this task as a minimization problem, for a non-negative shrinking parameter λ:

$$\beta = \underset{\beta}{\operatorname{argmin}} \left(\sum_{a=1}^{A} \left(y_a - \sum_{n=1}^{N} \beta_n X_n^2 \right) + \lambda \sum_{n=1}^{N} \beta_n \right)$$

From this, ITD scaling factor value H' is estimated as:

$$H' = \sum_{n=1}^{N} \beta_n H_n.$$

where The ITD scaling factors for all persons in the dataset are stacked in a vector $H \in \mathbf{R}^N$, so the value H^n corresponds to the scaling factor of the n-th person.

HRTF magnitude synthesis[11]

We solve the above minimization problem using Least Absolute Shrinkage and Selection Operator (LASSO). We assume that the HRTFs are represented by the same relation as the anthropometric features. Therefore, once we learn the sparse vector β from the anthropometric features, we directly apply it to the HRTF tensor data and the subject's HRTF values H' given by:

$$H'_{d,k} = \sum_{n=1}^{N} \beta_n H_{n,d,k}$$

where The HRTFs for each subject are described by a tensor of size $D \times K$, where D is the number of HRTF directions and K is the number of frequency bins. All $H_{n,d,k}$ corresponds to all the HRTFs of the training set are stacked in a new tensor $H \in \mathbf{R}^{N \times D \times K}$, so the value $H_{n,d,k}$ corresponds to the k-th frequency bin for dth HRTF direction of the n-th person. Also $H'_{d,k}$ corresponds to kth frequency for every d-th HRTF direction of the synthesized HRTF.

Recording technology

Recordings processed via an HRTF, such as in a computer gaming environment (see A3D, EAX and OpenAL), which approximates the HRTF of the listener, can be heard through stereo headphones or speakers and interpreted as if they comprise sounds coming from all directions, rather than just two points either side of the

head. The perceived accuracy of the result depends on how closely the HRTF data set matches the characteristics of one's own ears.

See also

- 3D sound reconstruction
- A3D
- Binaural recording
- Dummy head recording
- Environmental audio extensions
- OpenAL
- Sound Retrieval System
- Sound localization
- Soundbar
- Sensaura
- Transfer function

References

1. Daniel Starch (1908). *Perimetry of the localization of sound*. State University of Iowa. p. 35 ff.
2. Begault, D.R. (1994) 3D sound for virtual reality and multimedia. AP Professional.
3. So, R.H.Y., Leung, N.M., Braasch, J. and Leung, K.L.(2006) A low cost, Non-individualized surround sound system based upon head-related transfer functions. An Ergonomics study and prototype development. Applied Ergonomics, 37, pp. 695–707.
4. "AES Standard » AES69-2015: AES standard for file exchange - Spatial acoustic data file format". *www.aes.org*. Retrieved 2016-12-30.
5. "Sofa Conventions Website". Acoustics Research Institute, a research institute of the Austrian Academy of Sciences.
6. Blauert, J. (1997) Spatial hearing: the psychophysics of human sound localization. MIT Press.
7. Hofman, Paul M.; Van Riswick, JG; Van Opstal, AJ (September 1998). "Relearning sound localization with new ears" (PDF). *Nature Neuroscience*. **1** (5): 417–421. doi:10.1038/1633. PMID 10196533.
8. So, R.H.Y., Ngan, B., Horner, A., Leung, K.L., Braasch, J. and Blauert, J. (2010) Toward orthogonal non-individualized head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study. Ergonomics, 53(6), pp.767-781.
9. Carlile,S (1996). "Virtual Auditory Space and Applications". *Austin, TX, Springer.*
10. Tashev, Ivan (2014). "HRTF PHASE SYNTHESIS VIA SPARSE REPRESENTATION OF ANTHROPOMETRIC FEATURES". *Information Technology and Applications Workshop,San Diego, CA, USA, Conference paper*: 1–5. doi:10.1109/ITA.2014.6804239.
11. Bilinski,Piotr; Ahrens, Jens; Thomas, Mark R.P; Tashev, Ivan; Platt,John C (2014). "HRTF MAGNITUDE SYNTHESIS VIA SPARSE REPRESENTATION OF ANTHROPOMETRIC FEATURES". *IEEE ICASSP, Florence, Italy*: 4468–4472. doi:10.1109/ICASSP.2014.6854447.

External links

- Spatial Sound Tutorial (http://interface.idav.ucdavis.edu/sound/tutorial/)
- CIPIC HRTF Database (http://interface.cipic.ucdavis.edu/sound/hrtf.html)
- Listen HRTF Database (http://recherche.ircam.fr/equipes/salles/listen/index.html)
- AIR Database (HRTF database in reverberant environments) (http://www.ind.rwth-aachen.de/air/)
- Full Sphere HRIR/HRTF Database of the Neumann KU100 (http://www.audiogroup.web.fh-koeln.de)
- MIT Database (one dataset) (http://sound.media.mit.edu/resources/KEMAR.html)
- ARI (Acoustics Research Institute) Database (90+ datasets) (http://www.kfs.oeaw.ac.at/index.php?optio

n=com_content&view=article&id=608:ari-hrtf-database&catid=158:resources-items&Itemid=606&lang=en)

Retrieved from "https://en.wikipedia.org/w/index.php?title=Head-related_transfer_function&oldid=758243903"

Categories: Acoustics | Electrical circuits | Signal processing | Control theory

- This page was last modified on 4 January 2017, at 08:33.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.