User Acoustics with Head-Related Transfer Functions

Michele Geronazzo
Department of Architecture, Design, and Media Technology, Aalborg University, København, Denmark

Synonyms
Binaural hearing; Binaural sound; Head-related impulse response; Head-related transfer function; HRIR; HRTF; ILD; ITD; Spatial hearing; User acoustics

Definitions
Head-related impulse responses (HRIRs) or head-related transfer functions (HRTFs, in the frequency domain) describe spatial-temporal acoustic properties of a human body resulting from the interaction of user’s head, ear, and torso with the soundfield in space. The synthesis of a binaural anechoic spatial sound can be obtained by the convolution of an anechoic sound signal with left- and right-ear HRTFs chosen among a discrete set of measurements, i.e., spatial locations. Binaural and monaural localization cues contained in the HRTFs allow users to experience a 360° spatial audio experience, being fundamental component for the immersive and realistic auralization in virtual/augmented reality and games.

Introduction
Multisensory integration is all the more vital in spatial perception where auditory cues assist visual attention for seeking potential dangers around primates and mammals (Holmes and Spence 2005), apprehending location and movements of a potential prey and communicating between peers (Gridi-Papp and Narins 2008). Human hearing incorporates perception of loudness, pitch, timbre, and spatial attributes of sound. In particular, the auditory system is capable of estimating the location of sound sources in terms of direction and distance, as well as the spatial impression from the surrounding acoustic space. The position and movement of the listener inside the environment plays a key role in the perception of the characteristics of the listening space be it real or virtual.

Binaural anechoic spatial sound can be obtained by convolving an anechoic sound signal with left and right HRTFs over a discrete grid of spatial measurement points for a given listener and interpolated in order to synthesize arbitrary...
spatial locations (Xie 2013) (see Fig. 1a for a general schematic view of auralization). The minimum-phase characteristic of HRTF allows its decomposition in a pure delay, $\tau$, followed by a minimum-phase system, $H_{\text{min}}$: 

$$H(\theta, \phi, r, \omega) = H_{\text{min}}(\theta, \phi, r, \omega) \exp[-j2\pi\tau(\theta, \phi, r, \omega)],$$

(1)

where $\theta$ and $\phi$ denote source direction of arrival (DOA) in spherical coordinates and $r$ source distance from listener head. Auralization algorithms largely benefit from this decomposition in terms of stability and performances (Schissler et al. 2016). The pure delay/time-shift can be defined as the monaural time of arrival (TOA), allowing inter-aural time difference (ITD) extraction considering the difference between left and right ear TOA. With the inter-aural level/intensity difference (ILD/IID), they are the basis of duplex theory for binaural lateralization of sound (Strutt 1907). In particular, ILD can be formally described by the following equation:

$$\text{ILD}(\theta, \phi, r, f) = 20\log_{10}\left|\frac{P_L(\theta, \phi, r, f)}{P_R(\theta, \phi, r, f)}\right|,$$

(2)

where $P_{L/R}$ indicates the sound pressure in the frequency domain generated by a sound source in a given spatial position at the left/right ear. ILD is relevant for high frequencies where the wavelength is smaller than the physical dimensions of the head, shadowing sound pressure at contralateral side. Dynamic localization cues due to head movements further strengthen ITD/IID combination; accordingly, embedded IMU sensors in VR headsets become crucial for effective head-tracking and auralization.

Looking at the magnitude of $H_{\text{min}}$ (see Fig. 1b for an example), Romigh and Simpson (2014) recently identified the intraconic spectral component as the main cause of inter-individual differences where pinna acoustics becomes dominant for localization in elevation. However, recording individual HRTFs of each user requires a trade-off between resources and time, because the measurement process is heavily prone to errors, and nowadays its complexity results unmanageable for any real-world application.

In multimodal virtual environments, a common approach is to use the same generic HRTFs, e.g., recorded using a dummy head such as Neumann KU-100 (see Fig. 2 for some visual examples), for any listener in order to obtain a trade-off between average efficacy and measurement/personalization procedures taking into account the dominance of visual cues for localization. However, generic HRTFs generally introduce unnatural coloration in the frequency spectrum and degradations in localization and immersion of the listening experience. Paul (2009) provided an historical review on this topic.

**HRTF Personalization and Adaptation**

Dummy-head HRTFs can be considered an average template for user acoustics, introducing high uncertainty in localization due to differences with user anthropometry and to individual abilities in encoding generic directional information. Manipulating such HRTF template in order to personalize user acoustic contribution gives rise to different strategies: from the frequency scale factor based on anthropometric differences (Middlebrooks 1999) to ITD individualization with scaling procedures (Lindau et al. 2010), to name but a few. Moreover, one can allow users to directly manipulate a HRTF template using a series of bandpass filters to boost or attenuate specific frequency bands (Tan and Gan 1998). The post-manipulation phase usually requires user’s involvement with listening tests in order to further iterate the process until the desired level of efficacy is obtained.

Recent literature is increasingly investigating the development of personalized HRTFs for each individual user in virtual/augmented reality in order to support a listening experience which is perceptually equivalent to that with own individual HRTFs. HRTFs can be computed from basic geometry of the head, pinna, shoulders, and torso and/or accurate numerical simulations with boundary element (BEM) and finite-difference time-domain (FDTD) methods (Xie 2013; Prepelita et al. 2016).
Moreover, a structural interpretation of the acoustic contribution of head, pinna, shoulders, and torso can guide filter modeling of time and spectral (e.g., peaks and notches) variations (Geronazzo et al. 2013). HRTF features can be also defined according to principal component analysis (PCA) and subsequently personalized with self-tuning actions of weights (Hwang et al. 2010).
On the other hand, optimized selection procedures of existing nonindividual HRTFs is an alternative approach to personalization in which HRTFs are chosen by selecting the best match among several HRTF sets in a database. These approaches can benefit from the exponential increase of available HRTF data during the last 10 years, which has also provided motivations for standardization processes such as the Spatially Oriented Format for Acoustics (SOFA) (Majdak et al. 2013) (see also the official website of the SOFA project http://sofaconventions.org/ for an exhaustive list of freely available HRTF databases).

A user can select existing HRTFs in the domains of acoustic information, anthropometric features, and listeners subjective ratings (e.g., the most selected nonindividual HRTFs for a large user population). These procedures can be both automatic or guided by short listening tests aimed to assess the perceptual impact of nonindividual HRTFs for a given user. A training phase usually allows adaptation to nonindivial HRTFs thanks to automatic feedback such as correct-answer information to localization errors or audio-visual feedback adjustments. It is worthwhile to notice that there exist users able to better localize with nonindividual rather than individual HRTFs after an adaptation procedure (Klein and Werner 2016).

**Conclusions**

The need of a proper characterization of user acoustics is a key element for a realistic rendering in virtual reality scenarios, and it is even more crucial in immersive augmented reality settings where real and virtual sound sources are mixed in a natural listening experience (Ranjan and Gan 2015). Ensuring real life localization performances in VR becomes more and more important with close relations to presence, task-relevance, and effectiveness of the experience. Accordingly, users can increase their localization skills through adaptation procedures directly in VR games which act as a training tool for remapping localization cues with nonindividual HRTF (Parseihian and Katz 2012).

**Cross-References**
- Immersive Auralization Using Headphones
- Overview of Virtual Ambisonic Systems
- Sonic Interaction in Virtual Environments
- Sound Spatialization
- Spatial Perception in Virtual Environments
- Spatial Skill Training with Virtual Reality/Augmented Reality
References


