

Review of: "Solving the naming ambiguity of auditory localization mechanisms: HRTF & HRSL, and ILD, ITD, IPD"

Tim Ziemer¹

¹ Universität Bremen

Potential competing interests: No potential competing interests to declare.

This manuscript deals with the Head-Related Transfer Function (HRTF), in particular, with the ambiguous use of the term. It suggests the introduction of a new term to reduce ambiguity.

The ambiguous, and often false, use of technical terms is an important topic in science, especially when multiple disciplines are involved. In the case of the HRTF, these are, for example, acoustics, psychology, musicology, human-computer interaction and cognitive sciences. However, the manuscript neither demonstrates that an ambiguous use of the term "HRTF" is present in the literature, nor does it sufficiently argue for the suggested solution. Rather, the manuscript clearly shows the author's misconception of the HRTF. I highly recommend the author to read more textbooks, read scientific articles with more care, and study more. The manuscript is misleading and no sound, scientific work.

The introduction lacks a background section with a literature review. If the authors had consulted enough literature, he would have realized that none of the two mechanisms he describes are called HRTF. An HRTF is neither a sound source localization mechanism, nor a technological sound rendering technique.

In general, a transfer function fully describes a linear system or filter by introducing an input of "1" and measuring the output -> see https://media.springernature.com/full/springer-static/image/chp%3A10.1007%2F978-3-030-23033-3_6/MediaObjects/483725_1_En_6_Fig5_HTML.png

In acoustics, a (complex) transfer function describes a system's acoustic output for any given acoustic input. You can consider it the spectral counterpart of an impulse response in time domain. In the case of an HRTF, the system being tested is the constellation of two receivers (outputs), i.e., the sound pressure inside the two ear canals of a listener, and multiple (monopole) sound source positions (inputs), i.e., dozens to hundreds of sound source angles and distances being tested. The HRTF is just a collection or table of transfer functions with multiple source-receiver constellations. How the auditory system derives the sound source position from this output is another kettle of fish. By the way: Zhang et al. 2011 (who you cite multiple times) start their paper with "Head-related transfer functions (HRTFs), or the equivalent head related impulse responses (HRIRs) in the time domain, describe how a sound is filtered by the head, torso and pinna of a listener as it propagates from the source to the listener's ear drum in free space." You should have read the papers that you cite more carefully and reflect on what they imply.

Sound source localization mechanisms include that the auditory system derives interaural time- and amplitude differences to estimate the azimuth angle, overall loudness, attenuation of high frequencies, presence of low frequencies, and ratio of

direct sound to reverberation, to derive distance, and direction-dependent spectral peaks and notches to estimate both azimuth and altitude. Many of these pieces of information that the auditory system evaluates are contained in the HRTF. This is why a multiplication of a desired source spectrum and the HRTF-entry of a specific source position yields a very convincing listening experience, that the sound source is actually located at the desired source location outside the head, when listening over headphones. Another solution is to convolve a head-related impulse response entry with a desired sound to create this convincing illusion. A multiplication of spectra or a convolution of audio signals in time domain is a sound rendering technique, not the HRTF itself.

- "structures of torso and head" -> you forgot the pinnae, which are arguably the most important filters contributing high-frequency peaks and notches in the spectra of an HRTF. Zhang et al. 2011 (you cite them here) also explicitly say so. More importantly, your paper seems to indicate that Zhang et al. defined the HRTF as a perceptual mechanism, which they do not. You are right, however, that Risoud et al. 2018 use an unclear formulation: "Vertically, height is determined monaurally, involving just one ear: i.e., changes in incident spectral shape (reflection, diffraction and absorption) brought about by the pinna, head, shoulders and bust, known as the head-related transfer functions (HRTF)." This statement can be understood as if the HRTF was a monaural localization mechanism, but what they mean is that the reflections/deflections cause the spectral shape that can be found in the HRTF.

"The second mechanism, in contrast, is a technological sound precessing - typically..." -> This technology is referred to as "binaural audio", which is a much more ambiguous term than HRTF in my opinion.

"which is consistent and appropriate to describe the artificial processing of sound" -> A transfer function does not refer to artificial processing of sound. A transfer function "documents" or "tests" a system's behavior. You can make use of these documented results by technical means.

- Neither Zotkin et al. 2003 nor Inoue et al. 2005 state that the HRTF is technological sound precessing. They present technical ways to produce/predict individual HRTFs for better binaural audio rendering.

"ILD, ITD and IPD mechanisms share the same general scheme of operation: all are based on a perceptual processing mechanism" -> What is "operation" supposed to mean here? What operation? A difference means one thing minus the other. These terms are physical and not perceptual. They are not based on perceptual processing. The auditory system does a good job encoding and interpreting the ILD, ITD and IPD to some extent. But the underlying mechanisms are not the ILD, ITD and IPD themselves. As for the HRTF, you are the one who misunderstands these established terms. You should read some textbooks instead of focusing on conference proceedings papers.

The final section has little to do with the rest of the manuscript. It contains five sentences without an introduction, references, or discussion. It should either be removed, or included in the title, abstract, introduction and conclusion and put in the context of the HRTF.

The list of references reveals that the author has not consulted enough appropriate literature, which may be one of the courses why he misunderstands what an HRTF is. You should read textbooks in order to find definitions, or at least

elaborate descriptions of a transfer function in general, and the HRTF in particular. Even journal papers may be a reliable source to find at least a brief definition of an HRTF. But consulting papers from conference proceedings for such general topics is not a good idea.

Minor things:

- "... perceptual mechanism is based on a physical pre-processing" -> you mean physiological, referring to the deflections around the pinnae, etc.
- "-" should read "—"
- It is never a good sign when the review is longer than the reviewed manuscript.