Usage of Spectral Distortion for Objective Evaluation of Personalized HRTF in the Median Plane

Fabián C. Tommasini

Facultad de Matemática, Astronomía y Física (FaMAF), Universidad Nacional de Córdoba, Argentina.
Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina.

Oscar A. Ramos

Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina.
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

Mercedes X. Hüg and Fernando Bermejo

Centro de Investigación y Transferencia en Acústica (CINTRA), Unidad Asociada del CONICET, Universidad Tecnológica Nacional, Facultad Regional Córdoba, Argentina.

Measuring the head-related transfer functions (HRTFs) for each subject is a complex process. Therefore, it is necessary to develop procedures that allow the estimation of personalized HRTFs. It is common to estimate the weights of the principal component analysis (PCA) of a group of subjects based on some anthropometric parameters using multivariable regression modelling. Moreover, to objectively evaluate the goodness of fit between the original HRTFs and the personalized ones, the spectral distortion (SD) is usually used too. However, its suitability in the median plane, in which the spectral profiles are crucial to localize a sound source, has not yet been demonstrated. This paper analyses the validity of the SD as a measure of the quality of the HRTF personalization in the median plane, from the localization point of view. The HRTFs were modelled from the weights estimated by multiple linear regression and artificial neural networks (ANNs). The SD was used to compare the HRTFs measured with those estimated. Likewise, the level of fitting accuracy of characteristic resonance and notches in the median plane was also compared. Despite the fact that the SD scores of ANNs are lower than those of the multiple linear regression and are similar to those reported by other studies, the errors obtained from analysing both central frequencies and levels for resonance and notches could be discriminated.

1. INTRODUCTION

The purpose of the acoustic virtual reality (AVR) is to recreate the hearing experience that a person would undergo in a real environment, thus provoking a feeling of immersion in that environment. The principle that supports acoustic simulations states that acoustically equivalent stimuli produce equivalent sensations.[1] That is to say, if the biologically correct signals are applied to a listener's eardrums by means of headphones, it will be possible to stimulate the listener's feeling of immersion in the modelled environment.[2]

In an AVR system, the sound source, the room, and the listener must be modelled. Basically, the sound source is specified by the directivity characteristics and the frequency response; the room by its impulse response between a sound source and a receiver; and the listener by the head-related impulse responses (HRIRs) in the time domain, or the head-related transfer functions (HRTFs) in the frequency domain.

Due to the separation between both ears, the sound waves travel different paths, causing an interaural time difference (ITD), and, according to the position of the sound source, one of the ears may remain hidden by the head, thus also originating an interaural level difference (ILD). These two phenomena, together with the frequency spectra, are cues that humans use to localize a sound source in space.[1]

Each HRTF contains all the transformations produced in a sound wave before reaching a listener's eardrums when interacting with the head, pinna, torso, and shoulders. It is different for each ear and varies systematically with the location of the sound source. It is known that a HRTF not only depends on this, but also on the subject's anatomical characteristics: head and pinna size, and shoulders and torso width, among others.

If the HRTFs used correspond to the listener, the source is perceived as compact, external, and well defined in a position in space. By contrast, if they belong to another individual, the source is perceived as diffuse, located in the interior of the head, and the front-back confusion increases.[1,3] This means that it is essential to measure a subject's own HRTFs to experience a genuine perception of space. However, these measurements are complex and expensive and require special equipment. Therefore, it is necessary to develop procedures that allow estimating personalized HRTFs by means of simpler and less expensive approaches.

Different studies have addressed the problem of personalizing the HRTFs in different ways. A review of the methods can be consulted in Xu et al.,[4] in which seven potential methods for HRTF personalization were identified. One of these methods obtains the relation between some individual's anthropo-