ABSTRACT

To accurately render immersive auditory experience, virtual reality (VR) applications need an accurate set of head-related transfer functions (HRTFs) to serve as filters to generate spatial audio. These functions describe how individuals perceive sound originating from different source locations based on time, level, and spectral differences between sound arriving at the left and right ears. However, HRTFs are highly correlated with individuals' anthropometric features, and therefore, they are person specific. Since directly measuring HRTFs is time-consuming and inaccessible for the everyday user, prior research has focused on obtaining personalized HRTFs using subjects' anthropometric features through machine learning. This thesis proposes a machine learning approach which uses two models to estimate parts of HRTFs separately. Head-related impulse responses (HRIRs) are time-domain representations of the HRTFs. This study separated the HRIR into two key components: the time delay between the two ears and the overall HRIR shape. A regression tree is used to estimate the time delay, while an artificial neural network is used to estimate the shape of HRIR. Two listening tests were conducted to evaluate model's estimation. An ABX test assessed the perceptual differences between an actual HRIR and the model's estimated HRIR. Separately, a localization test was conducted in VR to evaluate the accuracy in the localization of auditory stimuli created with both an estimated HRIR and a generic HRIR. The listening tests revealed that the model's estimations were perceptually different from the actual HRIRs and that the generic HRIR performed just as well as the personalized HRIR. These results suggest that the models are slightly underfit for estimating individualized HRIRs even though they yield low RMSE. Nevertheless, this study demonstrated that better prediction accuracy could be achieved by estimating parts of HRIR separately.