Estimates of perceived spatial quality across the listening area

Abstract

This paper describes a computational model for the prediction of perceived spatial quality for reproduced sound at arbitrary locations in the listening area. The model is specifically designed to evaluate distortions in the spatial domain such as changes in source location, width and envelopment. Maps of perceived spatial quality across the listening area are presented from our initial results.

Precis

There are already established perceptual models that predict the sound quality impairments of speech and audio coding systems based on the timbral and temporal aspects of reproduced sound (e.g., PEAQ). The model was designed in the QESTRAL project to evaluate the effect of distortions in the spatial domain, such as changes in location and envelopment, and calculate an overall value of spatial quality. The model was calibrated by formal listening tests at multiple listener positions. It is not limited to codec evaluation, but calibrated using a range of audio processes and reproduction systems, such as downmix algorithms, loudspeaker misalignment and audio codecs. A leave-one-out cross-validation gave an R-squared value of 0.78 and root-mean-squared error of prediction of 12%. An important feature of the model is its use of binaural signals, allowing spatial quality to be predicted at multiple listening positions to create maps of spatial quality across the listening area.

In this paper, the model is used to predict measures of spatial quality for a given device under test (DUT) compared to a reference five-channel reproduction system (ITU-R BS.775-1). Probe signals, designed to stress the spatial performance of reproduced audio, are processed with and without the DUT yielding two sets of probe signals: the original reference set and an impaired set corresponding to the DUT. The DUTs tested include 64-kbps MPEG codec, 2.0 (stereo) downmix and 1.0 (mono) downmix. For each listening position, binaural signals are calculated for the two sets of probe signals by modeling the reproduction systems and acoustic listening environment as linear time-invariant systems. Hence the model calculates five metrics from the binaural signals, which are designed to evaluate various spatial aspects of the reproduced audio, including accurate rendering of localizable sound sources and listener envelopment. Differences between the reference and impaired metric values are combined in linear regression to give a single estimate of perceived spatial quality.

The listening area inside the five-channel loudspeaker configuration was sampled using a 10-cm grid. The model was used to estimate the spatial quality at each point on the grid for a number of different DUTs, and the results were plotted as maps of spatial quality. These are presented, both in terms of agreement with formal and informal listening tests and potential improvements to the model. The benefits of modeling spatial quality across the listening area are described, considering generalization of the model and mitigation of the risk of overfitting to the listening test data. Lastly, future areas for development and possible applications of the model are discussed, such as automatic system alignment, alternative rendering format and codec evaluation, and extensions to existing quality standards.