
Abstract

Experiments in Gaussian-mixture-model speaker recognition from mel-cepstra, derived from mel-filter bank energies (MFBs) of the G.729 codec all-pole spectral envelope, showed significant performance loss relative to the standard mel-cepstral coefficients of G.729 synthesized (coded) speech (Eurospeech99). In this paper, we investigate two approaches to recover speaker recognition performance from G.729 parameters. The first is a parametric approach that makes explicit use of G.729 parameters, rather than deriving cepstra from MFBs of an all-pole spectrum. Specifically, the G.729 LSFs are converted to `direct' cepstral coefficients for which there exists a one-to-one correspondence with the LSFs. The G.729 residual is also considered; in particular, appending G.729 pitch as a single parameter to the direct cepstral coefficients gives further performance gain. The second nonparametric approach uses the original MFB paradigm, but adds harmonic striations to the G.729 all-pole spectral envelope. Although obtaining considerable performance gains with these methods, we have yet to match the performance of G.729 synthesized speech, motivating the need for representing additional fine structure of the G.729 residual.