
Abstract

In this paper, we investigate the effect of speech coding on speaker and language recognition tasks. Three coders were selected to cover a wide range of quality and bit rates: GSM at 12.2 kb/s, G.729 at 8 kb/s, and G.723.1 at 5.3 kb/s. Our objective is to measure recognition performance from either the synthesized speech or directly from the coder parameters themselves. We show that using speech synthesized from the three codecs, GMM-based speaker verification and phone-based language recognition performance generally degrades with coder bit rate, i.e., from GSM to G.729 to G.723.1, relative to an uncoded baseline. In addition, speaker verification for all codecs shows a performance decrease as the degree of mismatch between training and testing conditions increases, while language recognition exhibited no decrease in performance. We also present initial results in determining the relative importance of codec system components in their direct use for recognition tasks. For the G.729 codec, it is shown that removal of the postfilter in the decoder helps speaker verification performance under the mismatched condition. On the other hand, with use of G.729 LSF-based mel-cepstra, performance decreases under all conditions, indicating the need for a residual contribution to the feature representation.