Mobile phone embedded digit-recognition

Christophe Levy, Georges Linares, Jean-Francois Bonastre

Abstract

Speech recognition applications are known to require a significant amount of memory. However, the targeted context of this work - mobile phone embedded speech recognition system - only authorizes less than 100kB of memory. In order to fit the memory resource, a global codebook of Gaussians is learned to derive state-dependent probability density functions. This strategy aims at storing only the transformation function parameters for each state. In this paper, two upper limits (concerning the acoustic model size) are set to 50kB and 100kB.

The proposed approaches are evaluated on the French corpus VODIS (digit recognition - recorded into car with or without fan/opened window/radio - with a very low Signal/Noise Ratio). This preliminary study allows to build systems fitting the memory constraint with a DER (Digit Error Rate) around 10.9% (for model less than 100kB) which represents a DER absolute increase less than 1% compared to an HMM-based baseline system respecting the same memory constraint. Despite this increase, performance of both approaches remains comparable since the DER is still in the confident interval.