Under-resourced speech recognition based on the speech manifold

Reza Sahraeian[1,] Dirk Van Compernolle[1,] Febe de Wet[2]

[1]ESAT, KU Leuven, Belgium
[2]HLT, Meraka Institute, CSIR, South Africa

fReza.Sahraeian,Dirk.VanCompernolleg@esat.kuleuven.be,fdwet@csir.co.za

Abstract:

Conventional acoustic modeling involves estimating many parameters to effectively model feature distributions. The sparseness of speech and text data, however, degrades the reliability of the estimation process and makes speech recognition a challenging task. In this paper, we propose to use a nonlinear feature transformation based on the speech manifold called Intrinsic Spectral Analysis (ISA) for under-resourced speech recognition. First, the authors investigate the usefulness of ISA features in low resource scenarios for both Gaussian mixture and deep neural network (DNN) acoustic modeling. Moreover, due to the connection of ISA features to the articulatory configuration space, this feature space is potentially less language dependent than other typical spectral-based features, and therefore exploiting out-of-language data in this feature space is beneficial. They demonstrate the positive effect of ISA in the frame work of multilingual DNN systems where Flemish and Afrikaans are used as donor and under-resourced target languages respectively. The authors compare the performance of ISA with conventional features in both multilingual and under-resourced monolingual conditions.