

EINLADUNG

Zeit: Montag, 14. März 2005, 16.30 Uhr
Ort: AH II, Ahornstraße 55 (ehem. PH)
Referent: Dipl.-Phys. Michael Pitz
Thema: Investigations on Linear Transformations for
Speaker Adaptation and Normalization

Abstract:

In current state-of-the-art speech recognition systems linear transformations are widely used to care for a potential mismatch of the training and testing data and thus enhance the recognition performance. A large number of approaches has been proposed in literature, though the connections between them have been disregarded so far. By developing a unified mathematical framework, close relationships between the particular approaches are identified and analyzed in detail. Mel frequency Cepstral coefficients (MFCC) are commonly used features for automatic speech recognition systems. The traditional way of computing MFCCs suffers from a twofold smoothing, which complicates both the MFCC computation and the system optimization. An improved approach is developed that does not use any filter bank and thus avoids the twofold smoothing. Starting from this new computation scheme for MFCC, it is proven analytically that vocal tract normalization (VTN) equals a linear transformation in the Cepstral space for arbitrary invertible warping functions. The transformation matrix for VTN is explicitly calculated exemplary for three commonly used warping functions. Based on some general characteristics of typical VTN warping functions, a common structure of the transformation matrix is derived that is almost independent of the specific functional form of the warping function. By expressing VTN as a linear transformation it is possible, for the first time, to take the Jacobian determinant of the transformation into account for any warping function. The effect of considering the Jacobian determinant on the warping factor estimation is studied in detail. Based on a close interrelationship between Maximum Likelihood Linear Regression (MLLR) and VTN, the general structure of the VTN matrix is adopted to restrict the MLLR matrix to a band structure, which significantly improves the MLLR adaptation for the case of limited available adaptation data.

Es laden ein: Die Dozenten der Informatik