

Towards understanding the regional distribution of acoustic features from speech

THOMAS KISLER
FLORIAN SCHIEL

Institute of Phonetics and Speech Processing, LMU

The aim of this study is to investigate whether automatically extracted acoustic features can be used to predict regional distribution in large speech corpora. We are pursuing a data-driven, bottom-up approach. All speakers and utterances are treated the same, and we show that extracted features indeed show a strong regional distribution within the data. In contrast to previous studies evaluating regional distribution taking into account the speech transcriptions, gives an overview, we aim to investigate the speech signal directly. When signals are considered, at most a mid-sized set of features is examined ($\ll 100$), whereas we want to extract several hundred features. To minimize the time-consuming preprocessing of the data and to create a purely objective method, we investigate the possibility to process the data without the existence of a transcription or phonetic segmentation and labeling. Examples for the low-level acoustic features are mel-frequency cepstral coefficients, Fourier coefficients, zero-crossing rate, etc., but also infrequently applied features, such as semitone histograms. They are extracted from a corpus with recordings from speakers distributed over the German dialect areas including Austria and Switzerland. The resulting feature space leads to a high-dimensional clustering problem. Since features should be clustered in all possible subspaces, dimensionality reduction techniques like PCA are rendered inapplicable. A special focus lies on appropriate high-dimensional clustering methods, which then should reveal the underlying dialect continuum of the feature values extracted from the speech samples. After features have been clustered that show a spatial (regional) distribution, we will examine which acoustic features seem to correlate with spatial distribution and why. The resulting distribution is then mapped to known dialect areas, to unveil whether or not the automatically processed regional distributions coincide with them, and it is investigated what the implications might be for dialect research.