

PITCH AND TONE'S MODELING IN PARAMETRIC TRAJECTORY MODEL

Yiyan Zhang; Wenju Liu; Bo Xu; Huayun Zhang, National Laboratory of Pattern Recognition, China

It is described in this paper for the application of pitch/tone information in the parametric trajectory model. Pitch as a dynamic feature and its contours---tone as a segmental-level feature are deserved their own particular characteristics, which match case of parametric trajectory model better compared with MFCC and energy. Here we give an improved pitch extraction algorithm and especially the "total" and "parallel" integration methods to combine these information with the base model. In the experiment of Mandarin connected digit recognition, we achieve 22.87% and 33.54% error reduction respectively for them, moreover when combined with these two methods, 38.72% error reduction is obtained.

TIME-ALIGNED SVD ANALYSIS FOR SPEAKER IDENTIFICATION

Patrick Clemins; Heather Ewalt; Michael Johnson, Marquette University, United States

This paper presents a time-aligned singular value decomposition (SVD) analysis for speaker identification. SVD analysis has been used for fast spectral matching based on a global representation of an entire utterance. We incorporate temporal normalization directly into the decomposition by using a dynamic time warping (DTW) path to time-align the rows of the feature matrix prior to SVD analysis. Speaker identification results using the TI-46 database indicates that the time-aligned SVD significantly improves accuracy for most threshold choices.

TWO-PITCH ESTIMATION FOR CO-CHANNEL SPEAKERS SEPARATION

Julie Rosier; Yves Grenier, ENST, France

This paper deals with the estimation of fundamental frequency for two speakers recorded on the same channel. We estimate a set of speech models based upon sinusoidal plus autoregressive noise representations. We then detect the best model from this set using Rissanen criterion. Equivalent to a penalized log-likelihood, the criterion is also used to carry out a voicing detection. The detector compares the likelihood of a sinusoid plus noise model with the likelihood of a simple autoregressive model. Several simulations are presented to illustrate this estimation method.

PERCEPTUAL FEATURES FOR A FUZZY SPEECH-SONG CLASSIFICATION

David Gerhard, Simon Fraser University, Canada

Human speech and song seem disparate, but a range of utterances between speech and song are evident, such as poetry, chant, and rap, which have features of both singing and speaking. This work seeks to identify and characterize the perceptual features relevant for a fuzzy classification of utterances between speech and singing. The speech-ness or song-ness of an utterance depends on the speech or song features evident in that utterance. This paper presents a brief discussion of the collection and annotation of the corpus of sound clips used in this work, followed by a description of the perceptual features expected to be useful, and presentation of preliminary results for two of these features.