Feature Extraction for ASR: Intro

Wantee Wang

2015-03-14 15:49:44 +0800

Feature extraction is the first step for Automatic Speech Recognition(ASR), which converts the waveform speech signal to a set of feature vectors. The main goal is to make the vectors have high discrimination between phonemes.

Thus, the features should be

- perceptually meaningful, i.e., analogous to features used by human auditory system.
- invariant, i.e., robust to variations in channel, speaker and transducer.

Three main steps for feature extraction are

1. Preprocessing
2. Feature Analysis
3. Parametric Transformation

The *preprocessing* step converts the speech signal to a more suitable waveform for the following analysis, including *DC offset removal*, *pre-emphasis* and *Hamming Windowing*.

Feature Analysis is most important step, which do most of the works. Generally, it is can be divided into two main categories, *Spectral Analysis* and *Temporal Analysis*. Spectral analysis gives MFCC and PLP features, while temporal analysis produces Energy and Pitch features. MFCC involves *Cepstral Analysis* and PLP is based on *Linear Predictive Coding(LPC) Analysis*.

The final step, *Parameter transformation*, converts the features obtained by above step into signal parameters through *differentiation* and *concatenation*.

Details are in following posts:

1. Feature Extraction for ASR: Preprocessing
2. Feature Extraction for ASR: MFCC
3. Feature Extraction for ASR: PLP
4. Feature Extraction for ASR: Pitch
5. Feature Extraction for ASR: Delta