Feature Extraction

1 Questions

1. What are requirements for cost/distance functions?
2. What are two important properties/quantities of speech signals used for the derivation of features? What is their connection to the source-filter model of speech production?
3. Which of these two properties is modeled by linear prediction coefficients?
4. What quantities are needed for the computation of linear prediction coefficients (LPC)?
5. What quantities are needed for the computation of linear prediction cepstral coefficients (LPCC)?
6. What may be the motivation for applying a logarithm in feature extraction?
7. Why can it be said that MFCCs look at speech from a different perspective compared to LPC/LPCC?
8. What is a Mel filter bank and how is it practically applied with a DFT? What properties does it offer for feature extraction (see MFCCs)?

2 Answers

1. They should be invariant to level/gain. They should also be "easy" and cheap to compute. The closer entities are to each other, the smaller should be the value of the cost funtion. It may also be advantageous if the function mirrors the human perception (does not always apply).

2. The two properties are the pitch (fundamental frequency) and the spectral envolope. Together they are the source and filter for the formation of voiced speech sounds.

3. Linear prediction coefficients (to be precise the inverse prediction-error coefficients) model the spectral envelope in the time-domain.

4. The autocorrelation function is required to fill the correlation matrix and vector in the proposed solution.

5. See LPC.

6. The logarithm is able to provide better resolution for differences in low power regions. The human loudness perception is also logarithmic.

7. LPC and LPCC are derived using a speech production model. MFCCs, however, are derived by modeling how speech/sound is perceived by the human ear.

8. A Mel filterbank performs a transformation from the frequency scale in Herz to the subjective frequency scale in Mel. It also performs a grouping of frequencies, yielding so-called mel bands as output. Both operations are executed by e.g. applying triangular filters to a DFT power spectrum. The speacing between and the width of the individual filters increase along the frequency axis. Often, the filters are normalized to the width to yield in some way the mean frequency content within a frequency range. For MFCCs the filter bank adds a subjectively motivated frequency resolution, smoothing, and data rate / dimension reduction.