COURSE 02457

Signal Processing in Non-linear Systems:

Lecture 10

- Gaussian mixtures recap
- Speech signals
- Speech production
- Mathematical model
- Linear predictive modeling
- Cepstral coefficients
- Simple Markov models

Speech signals

- Speech signals are composed of a sequence of sounds
- These sounds and the transitions between them serve as symbolic representation of information
- The arrangement of these sounds (symbols) is governed by the rules of language
- The study of these rules and their implications in human communication is called *linguistics*
- The study of and classification of the sounds of speech is called *phonetics*

Speech production

- Speech is produced by the human vocal tract
- The vocal tract is excited either by short burst of periodic stimulus or white noise
- Voiced sounds are produce by airflow through tight vocal cords. Unvoiced sounds (or fricatives) are produced by turbulence.
- Sounds are classified broadly into phonemes: Vowels, diphthongs, semi-vowels, and consonants.
- Formants are frequencies in quasi-periodic parts of speech. Frequencies in the range: F1 (270-730 Hz), F2 (840-2290 Hz), and F3 (1690-3010 Hz).

(*After L.R. Rabiner and R.W. Schaefer: Digital Processing of Speech Signals*)

Modeling speech

- The speech signal is modeled as a linear (slowly) time-variant system excited by a high-frequency signal (periodic or random).

Linear predictive modeling

- Linear system model (IIR model)

$$x(n) = \sum_{j=1}^{p} w_j x(n-j) + \epsilon(n)$$

- System parameter are estimated from short sequences (20-30 msec)in which the signal is quasi-stationary by the Levinson-Durbin algorithm. The autocorrelation function is defined

$$R(m) = 1/N \sum_{n=m+1}^{N} x(n)x(n-m) + \epsilon(n)$$

- and the least squares estimates of the parameters satisfy

$$R(m) = \sum_{j=1}^{p} \widehat{w}(j) R(m-j)$$

Cepstral coefficients

- The cepstrum is defined at the inverse DFT of the log of the

$$
\begin{aligned}
C(m) &= IDFT\,(log(|X(k))|) \\
X(k) &= DFT(|x(m)|)
\end{aligned}
$$

- Can be used to separate a slowly varying envelope and rapidly varying excitation

Simple Markov models

- Let s^n be a sequence of symbols with K states
- Let $a_{j,j'}$ be the probability of going from j to j'.
- $a_{j,j'}$ is a stochastic matrix $\sum_{j'} a_{j,j'} = 1$
- a can be estimated by maximum likelihood.

Simple Markov models

- a can be estimated by maximum likelihood.

$$
\begin{aligned}
P(\{s_n\}|a) &= P(s_1)\prod_{j=2}^{N} P(s_j|s_{j-1}, a) \\
&= P(s_1)\prod_{<j,j'>} (a_{j,j'})^{\mathrm{n_{j,j'}}}
\end{aligned}
$$

- $n_{j,j'}$ is the occurrence of the transition.
- Condition $\sum_{j'} a_{j,j'} = 1$ leads to the solution,

$$
\widehat{a}_{j,j'} = \frac{n_{j,j'}}{\sum_{j'} n_{j,j'}}
$$