Linear Prediction

Notes on Speech and Audio Processing

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- Speech can be modeled as been produced by a periodic or noise-like source that drives an non-uniform acoustic tube.
- Based on this model of speech production, we will describe the feature set of the linear prediction coefficients. They are succinct and smooth.

Resonances and Formants

- A tube driven by periodic pulses has a set of resonant frequencies.
- Resonance corresponds to peaks in the frequency response. The resonant frequencies of vocal tract are called the formants.
- There are $4-5$ formants in the frequency range we are normally interested in, i.e., approximately one formant per kHz.

Predictive Model

- (Sec 6.8) The following transfer function represents a resonance,

$$H(z) = \frac{1}{1 - bz^{-1} - cz^{-2}}.$$

- Multiple formants can be modeled by a cascade of such resonators,

$$H(z) = \frac{1}{1 - \sum_{i=1}^{P} a_i z^{-i}},$$

$$\text{or } \ y[n] = x[n] + \sum_{i=1}^{P} a_i y[n-i].$$

Comments on Predictive Model

- The second term on the rhs $\tilde{y}[n] = \sum_{i=1}^{P} a_i y[n-i]$ is a linear predictor of $y[n]$.
- The difference between $\tilde{y}[n]$ and $y[n]$ is called the residual (or prediction error) signal.
- The energy of the error signal is minimized to compute a_i, called the linear prediction coefficients.
- The transfer function has only poles, so it is also called the all-pole model.

Getting the Coefficients

- Define the error signal

$$e[n] = y[n] - \tilde{y}[n]$$

- Let $D = \sum_{n=0}^{N-1} e^2[n]$ and we want to find the $a_i's$ that minimizes D. At the minimum value, the first-order partial derivative must vanish, and

$$\sum_{i=1}^{P} a_i \phi(j,i) = \phi(j,0), \text{ for } j = 1, \ldots, P$$

- a_i's can be solved using the Levinson-Durbin recursion.