Digital Signal Processing - DSP
(PDS - Processamento Digital de Sinais)
Instituto Superior Técnico - 2º semestre 2008/2009
Jorge Salvador Marques

Spectral Analysis of Speech

Speech Recognition

Speech recognition has been a major goal during the last three decades. It aims to extract useful information from the speech signal and to use this information to convert the acoustic signal into a sequence of words. How is the information about words encoded in the speech signal? Words are made of smaller units (phonemes, diphones) and these units have a direct influence on the spetral envelope of the speech signal in short intervals (20 ms). Both the spectral envelope and its time evolution bring important information. Stationary sounds (e.g., a vowel) can be recognized from a single spectrum but the recognition of diphones and words requires the analysis of consecutive spectra computed at a frame rate of about 100Hz.

LPC Analysis

How can we compute the spectral envelope? This is usually done by modelling the speech signal $x(n)$ as the output of an AR signal i.e., we try to predict each sample from the last p samples plus a prediction error.

$$x(n) = a_1 x(n-1) + ... + a_p x(n-p) + w(n)$$

Typically $10 <= p <= 16$. The coefficients of the AR filter can be obtained by the least squares method. This leads to a set of equations $R\alpha = r$ where $R = [R_{ij}]$, $r = [r_i]$ are given by:

Assuming the excitation $w(n)$ has a flat spectrum $W(e^{jw}) = 1$, the spectral envelope of the speech signal is approximated by the amplitude spectrum of an AR filter.

Lab Tasks

1. Obtain 5 speech signals corresponding to the vowels a,e,i,o,u using the microphone available with your PC. Convert these signals to a sampling rate of 8KHz and visualize them.
2. Compute a linear predictor of order $p = 10$ for each vowel using an interval of 20 ms.
3. Visualize the prediction error $w(n)$.
4. Visualize the LPC spectra computed for each of the vowels and determine the first two formants F1, F2 by inspection. Represent them graphically by a point in the plane F1.F2.
5. Repeat the previous steps to obtain 5 new signals and represent their formants again in the plane F1, F2.
6. Discuss how these ideas can be used in vowel recognition.