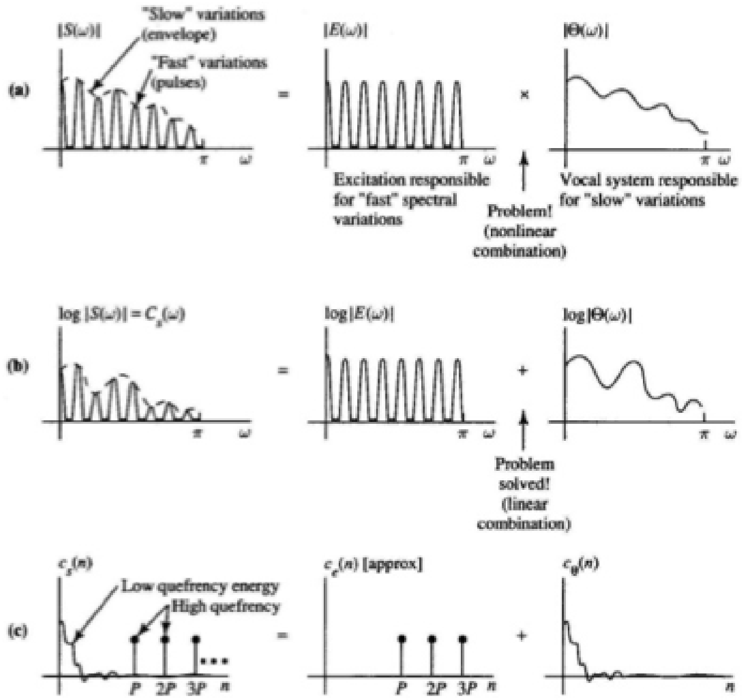


Figure 1: Procedure for feature extraction.

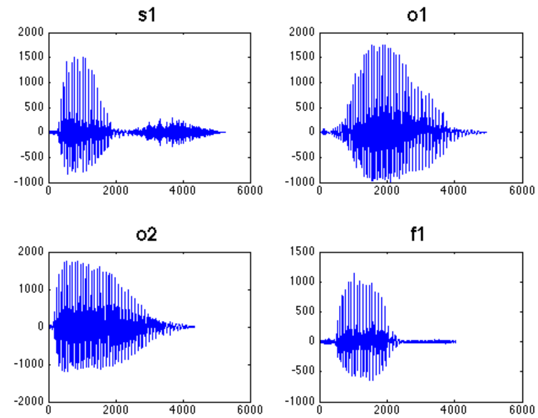
COURSE 02457

Non-Linear Signal Processing: Answers to Exercise 11

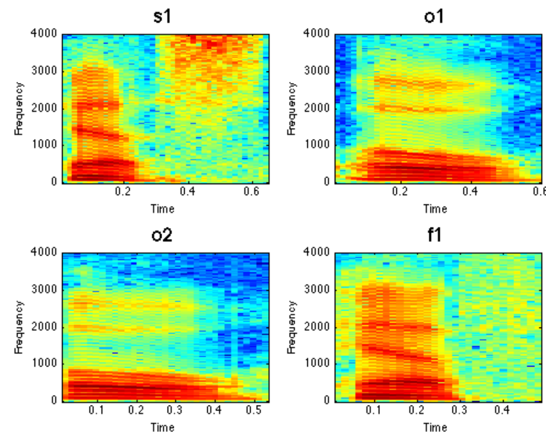
Checkpoint 11.1

In this exercise we are going to perform speech recognition. Since an audio signal contains a lot of information, it can be difficult to distinguish different sounds solely based on the temporal sound wave. In order to make it easier to distinguish the sounds, we will extract the most important pieces of information from the raw signal, and do our analysis on that. The feature extraction method we are going to use is called **cepstral liftering**, and is defined as the Inverse Fourier transform (IFT) of the logarithm of the absolute value of the power spectrum of the signal. This procedure is illustrated in Fig. 1.

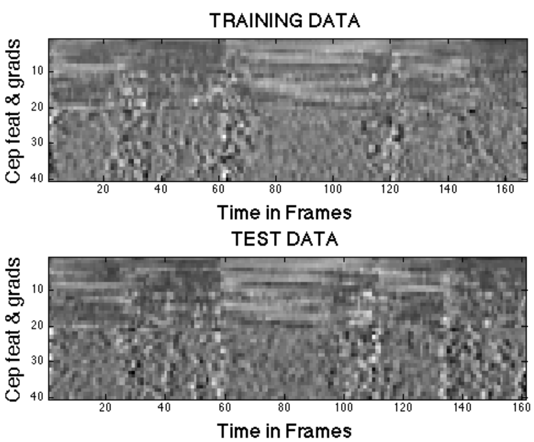
We are going to use this technique to perform feature extraction on speech signals representing the letters S, O and F.



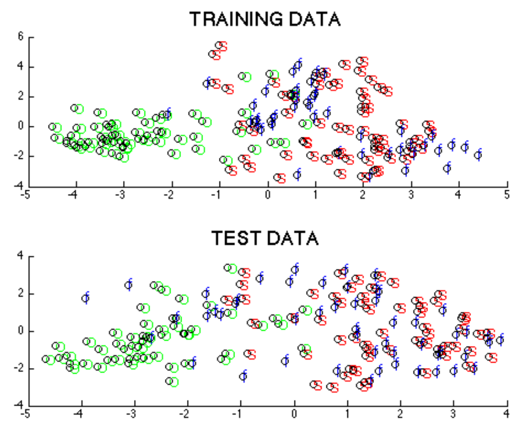
(a) Audio signal.



(b) Spectrograms (Fourier transform).



(c) Evolution of the cepstral coefficients and their gradients.



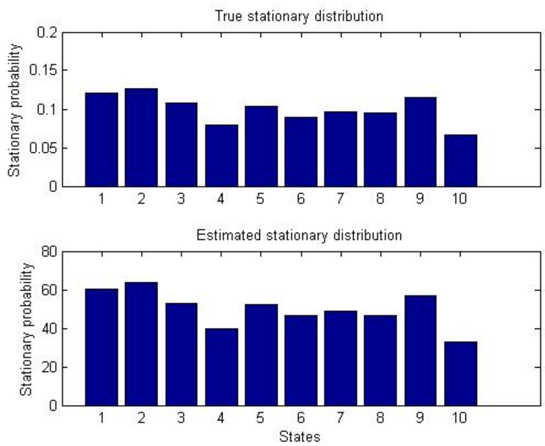
(d) Data projected on the two most principle directions in feature space.

Checkpoint 11.2

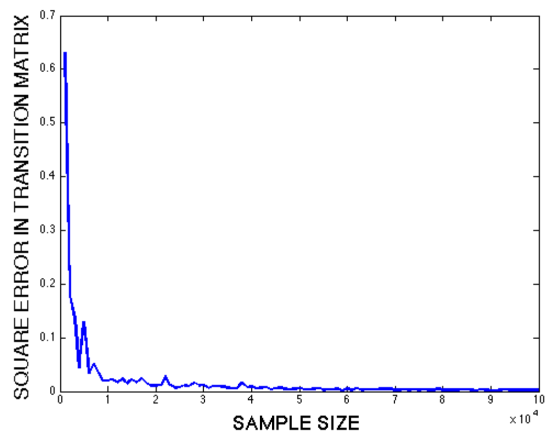
Imagine we have learned to recognize 10 different sounds/phonemes that can be used to construct sentences, we would like to estimate the transition probability for going from to another, assuming that the sequence of phonemes behaves like a **markov chain**.

To illustrate how we can estimate the transition probabilities between the phonemes, we first create a random **teacher transition matrix** A_t that creates a stationary distribution of the K different states.

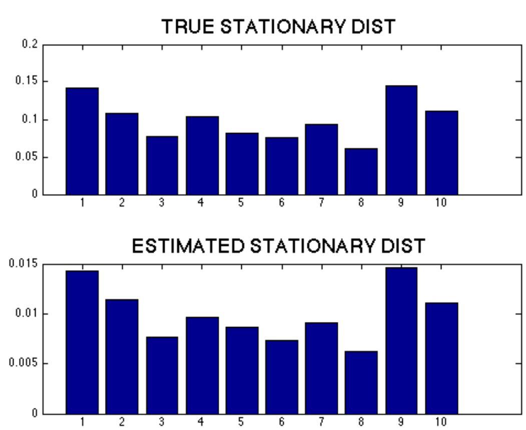
By calculating the cumulative distribution of A_t, we can make a sample h from A_t, see Fig. 2g. Based on h, we can construct a **student matrix** A_s that will act as an approximation of A_t. The matrix norm error between the student and the teacher matrix for different sizes of h is shown in Fig. 2h



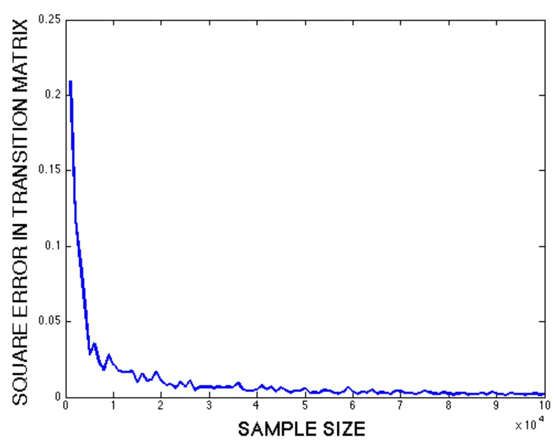
(e) Comparison between h and p for $p(y1) = 1$.



(f) Data projected on the two most principle directions in feature space for $p(y1) = 1$. The value of the error in the last iteration is 0.0019.



(g) Comparison between h and p for $p(y1) = 1/K$.



(h) Data projected on the two most principle directions in feature space for $p(y1) = 1/K$. The value of the error in the last iteration is 0.0020.