CS498 Fall 2011 - Problem Set 4 - Synchronizing Signals

In this problem set you will make use of dynamic programming in order two align two spoken utterances. Get this archive that contains two soundfiles of the same sentence spoken slow and fast. Your goal is to find the optimal warping to get these two to line up.

Features

As describes in class you will have to find a feature representation if you hope to find an alignment easily. Load the two sounds and for each of them compute the log of the magnitude of their spectrograms. Use an FFT size of 1025 samples, an overlap of 512 samples, and employ a Hann window (MATLAB hints: `spectrogram()`, `hann()`).

Plot the two feature sets and make sure that you use the same axes for both plots. You should be able to see that one of the spectrograms is much narrower than the other implying that it is the faster sentence. Each column of the spectrograms is the spectral feature for that time point. Let's say you will have M columns in the first sound and N for the second one.

Distances

In order to perform the time warping you need to create a matrix of pairwise distances between all the spectra you have. Create an M by N matrix $\mathbf{D}$, where the $(i, j)^{th}$ element is the Euclidean distance between the i^{th} time feature for the first sound and the j^{th} time feature for the second sound. Plot the distance matrix.

Cost matrix

You will now have to use the distance matrix to find the cost of possible paths. Remember that each time you advance to a new point in your path your cost will be the existing cost plus the distance value associated with that jump. Use the constraint that each sound can either stay at the same time point or advance by one. Your cost matrix should be an M by N matrix $\mathbf{C}$, such that the $(i, j)^{th}$ element is the overall optimal cost that a path ending in (i, j) would have.

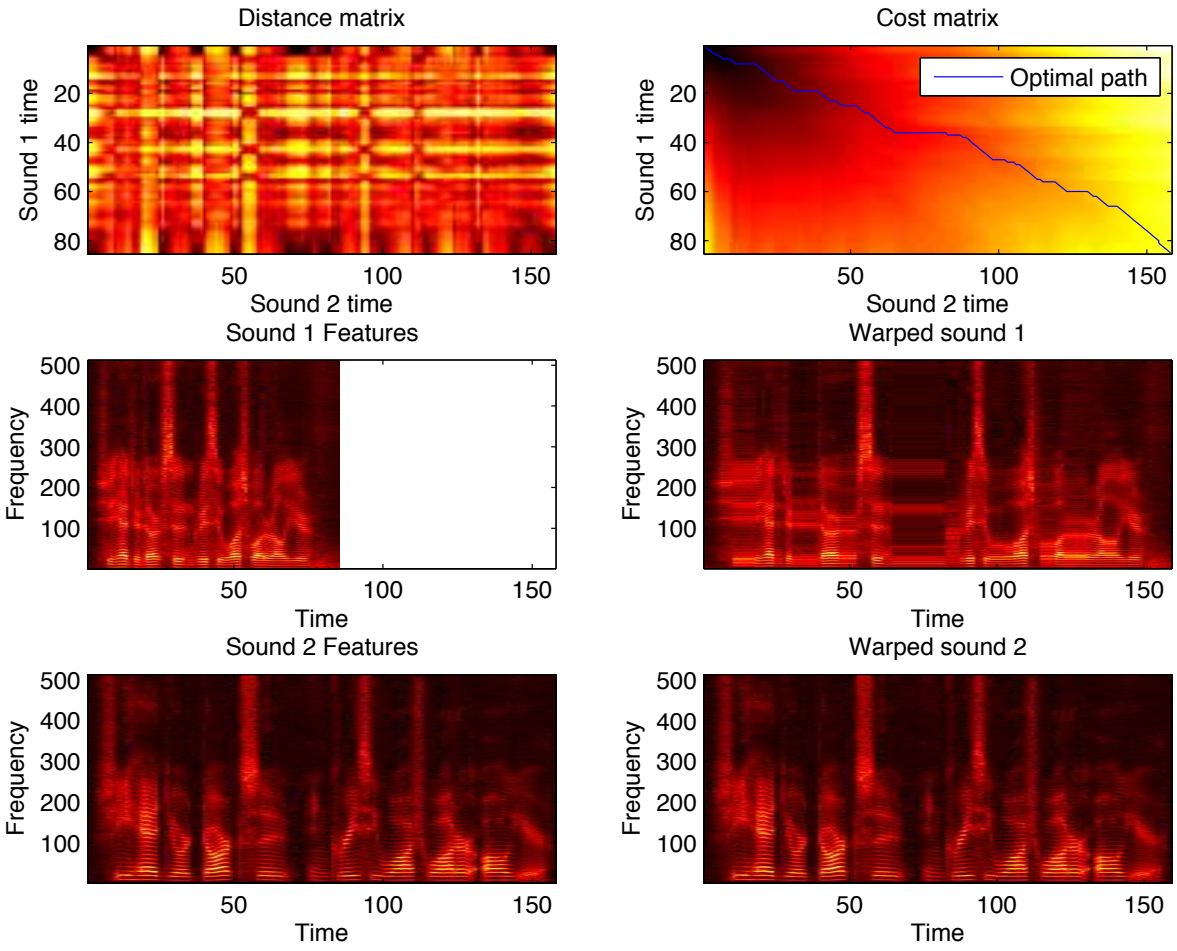
Optimal path

Once you compute the cost matrix you should be able to backtrack and find the overall optimal path. That path will tell you how to warp the two feature representations so that they line up. Do so and plot them, you should see that the warped features line up in time and that both sequences are the same length.

Extra credit

If you are ambitious you can try to convert the warped features to a time series and submit with your homework the resulting sounds. They should be in sync if you play them simultaneously. In order to do so you need to implement an inverse spectrogram function.

This problem set is due December 9th. An example of the plots we expect is shown in the following graph.


Distance matrix
Sound 1 time
20
40
60
80
50
100
150
Sound 2 time
Cost matrix
Optimal path
Sound 1 time
20
40
60
80
50
100
150
Sound 2 time
Sound 1 Features
Frequency
500
400
300
200
100
50
100
150
Time
Warped sound 1
Frequency
500
400
300
200
100
50
100
150
Time
Sound 2 Features
Frequency
500
400
300
200
100
50
100
150
Time
Warped sound 2
Frequency
500
400
300
200
100
50
100
150
Time