BİL735 Speech Recognition:
Assignment 3,
Training Phoneme
3 May 2017

In this assignment we will train phoneme from the digit recordings used in before assignment.
Train Phoneme models from speech digit recordings. use the small corpus of digit recordings you recorded for before corpus.
Specifically, you will have to train models for the following phonmes. Model each phoneme using THREE emitting states: (English alphabet)

AX, AH, AY, EH, EY, F, IH, IY, K, N, OW, R, S, T, TH, UW, V, W, Z
To do so, express each of the digits from zero through nine as phoneme sequences using the following dictionary:

ONE:	W	AX	N		
TWO:	T	UW			
THREE:	TH	R	IY		
FOUR:	F	OW	R		
FIVE:	F	AY	V		
SIX:	S	IH	K	S	
SEVEN:	S	EH	V	EH	N
EIGHT:	EY	T			
NINE:	N	AY	N		
ZERO:	Z	IY	R	OW	

You will use the dictionary to represent your digit sequences as phoneme sequences. Now, using the procedure used to train digit models from continuous recordings, you can train phoneme models.

As an example, if you recorded the digit sequence 123456 as training data, and you have silences at the beginning and end of the recording, you would represent the digit sequence in the following manner to train your phoneme models:
SIL W AX N T U TH R IY F OW R F AY V S IH K S SIL

If you know you actually also paused between 3 and 4 in the recording, you'd model it as:
SIL W AX N T U TH R IY SIL F OW R F AY V S IH K S SIL

For recognition, compose digit models from the phoneme models you have trained using the dictionary above. You will NOT be recognizing phoneme sequences. You will compose word models and perform recognition of words!

Count the number of phonemes in each digit. Since we will be using isolated word recordings for initialization, we will assume that each word is bracketed by silence (has silence on either side). So, a recording of "ONE" will be counted as having 5 phonemes: SIL,W,AX,N,SIL.

Using the above initialization, train a complete set of phoneme models from your isolated word recordings. This procedure is exactly the same as training from continuous recordings, only you will be training phoneme models from isolated word recordings.

Use the final converged models as initialization for training phoneme models from continuous digit sequences.

Due: 17 May 2017.