Speech Recognition Spring 2026 Project 1

Due date: Jan 13th 2026 12pm Beijing time

Problem 1

Write a program to capture speech data. It must include the following:

• Record multiple instances of digits

o Zero, One, Two etc.

o 16Khz sampling, mono channel, 16 bit PCM

For reference, you can consult PyAudio's codebase.

https://people.csail.mit.edu/hubert/pyaudio/docs/

Problem2

• Write a function for computing MFCC from audio

o Compute log spectra and cepstra

▪ Use 40 Mel spectral filters. They must cover the frequencies between 133.33Hz and 6855.4976Hz (you may use a different setting if you choose).

▪ No. of features = 13 for cepstra (use first 13 DCT coefficients)

o Visualize the original and the reconstructed spectrogram

▪ Note similarity in different instances of the same word

o Modify number of filters to 30 and 25 (over the same frequency range).

▪ Patterns will remain, but be more blurry

o Perform delta and double delta operations as well as the mean variance normalization

Some suggestions

You may utilize the provided framework and implement the missing functions.

- Dan Ellis has nice matlab code on his website.

  https://www.ee.columbia.edu/~dpwe/resources/matlab/rastamat/

- **How to visualize the spectrogram represented by cepstra**

The Mel-log spectrum can be directly visualized as a matrix.

However, the cepstrum is a dimensionality-reduced and transformed version of the log spectrum. It is not visually meaningful. The truncated cepstrum can be converted back to a log spectrum by zero padding it to 64 or 128 points and computing an inverse DCT (if you used a DCT to derive cepstra from log spectra). The IDCT-derived logspectrum is what the cepstrum really represents.

You need to visualize the IDCT-derived logspectrum offline and shown in the report. We have provided an example audio file in the asset folder, along with visualizations of both the Mel-log spectrum and the IDCT-derived log spectrum for your reference. You can also compare your results with those generated by the Librosa toolkit. Additionally, you should create visualizations for the digit recordings you have made.

Your submission should include the source code, the recorded files, the saved features, the spectrogram images, and a word or pdf report.