1. The waveform of the sentence "Irish youngsters eat fresh kippers for breakfast," is shown on Figure below. The signal was sampled at 8000 samples/ second.

(a) Label all the regions of voiced, unvoiced, and silence.

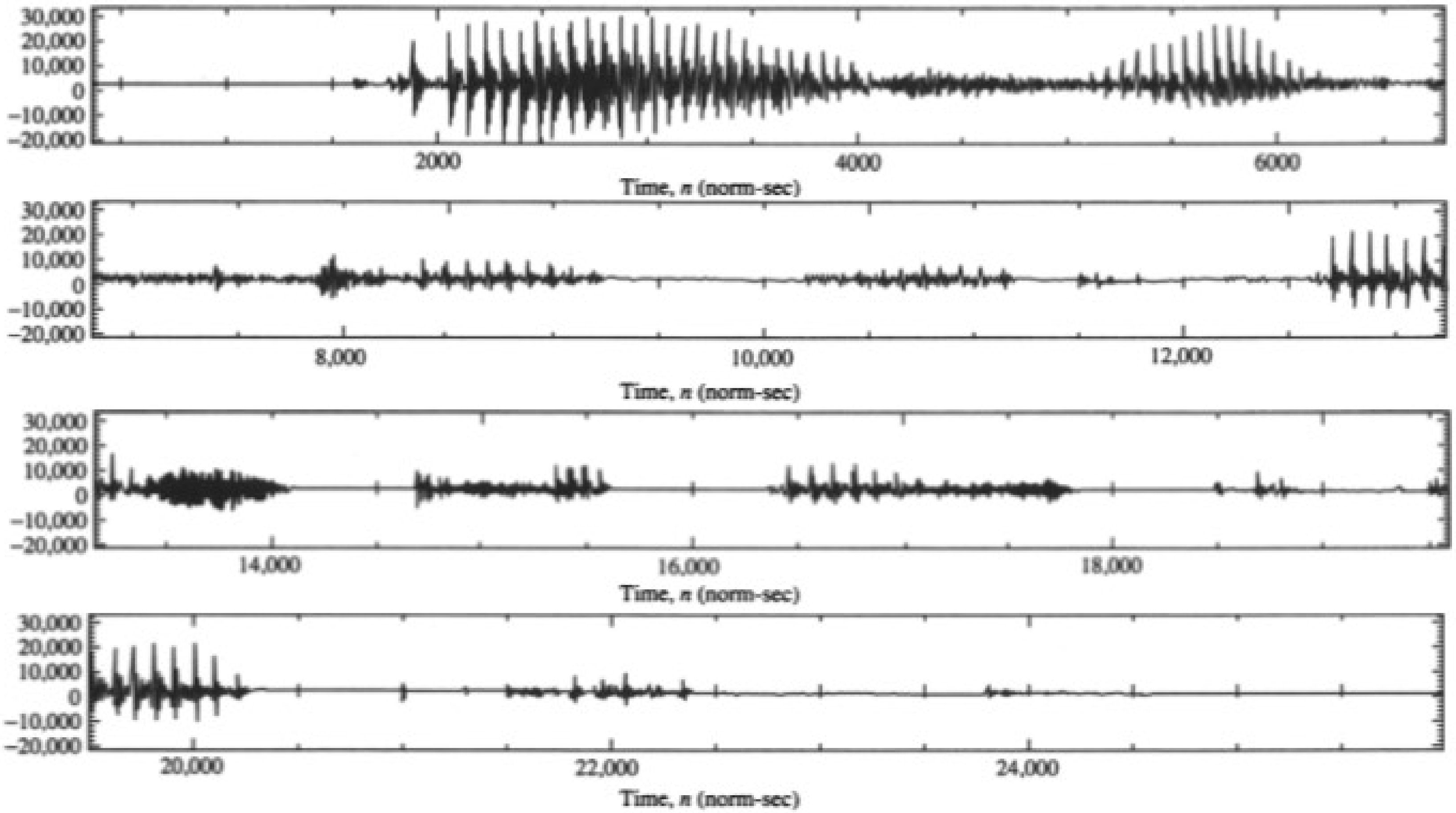
(b) For the phrase "Irish youngsters," indicate the phonemic boundaries (i.e., mark boundaries for the phonemes IYI-/r/-/I/-ISI,/y/-/A/-/G/-/s/-ltI-/R/-/z/).

(c) For each the vowel sound, estimate the average pitch period.

(d) Is the speaker an adult male, adult female, or a child?

(e) Find the range for pitch frequency, using estimated minimum and maximum pitch period values from part (c).

(f) Please pronounce the sentence and compare your signal with the given one.


30,000
20,000
10,000
0
−10,000
−20,000
2000
4000
6000
Time, n (norm-sec)
8,000
10,000
12,000
Time, n (norm-sec)
14,000
16,000
18,000
Time, n (norm-sec)
20,000
22,000
24,000
Time, n (norm-sec)

2. Prounounce /f/,/T/,/s/,/S/,/h/,/I/,/@/,/R/ and record the signal? Compare their energies. Why are fricatives always lower in energy than vowels?

3. (a) In the speech files available, you will find a file that contains samples of the vowel /oo/, sampled at an 16-kHz sampling rate. Extract 512 samples of this speech sound and plot the time waveform. Next, apply an FFT to the segment and plot the magnitude spectrum.

(i) What features characterize the time waveform?

(ii) From the magnitude spectrum, estimate the fundamental frequency. Is this a male or female speaker?

(iii) From the same magnitude spectrum, estimate the formant locations. How do your measurements compare to typical values for the vowel?

(b) In the same set of computer files, you will find some vowel files whose phonemes are not labeled. For one or more of these files, repeat the steps of part (a), and determine which vowel the file(s) likely represent(s).

4. The glottal shaping filter is often modeled by a two-pole system whose impulse response is given by

$$g(n) = [\alpha^n - \beta^n]u(n), \qquad \beta < \alpha < 1, \quad \alpha \approx 1.$$

(a) Plot several impulse responses for various values of α and β. Make sure that your pulse is practically "closed" at time n= 64 in every case.

(b) Based on 128-point DFTs, plot the magnitude and phase responses for two typical examples in part (a). As wediscussed in the chapter, more realistic glottal pulse shapes can be obtained using the Rosenberg pulse

$$g(n) = \begin{cases} \frac{1}{2}[1 - \cos(\pi n/P)], & 0 \le n \le P \\ \cos([\pi(n-P)]/[2(K-P)]), & P \le n \le K. \\ 0, & \text{otherwise} \end{cases}$$

(c) Plot several impulse responses for various values of P and K. Make sure that your pulse is "closed" at time n=64 in every case.

(d) Based on 128-point DFT, plot the magnitude and phase responses for two typical examples in part (c).

(e) Compare the time-domain and frequency-domain properties of the two glottal pulse models. What realistic time-domain feature of the Rosenberg pulse is apparently unachievable with the two-pole model? How is this feature manifest in the phase spectrum?

5. (a) Perform the following operations with a vowel utterance of your choice:

(i) Compute the short-term autocorrelation for a Hamming window of length N=512 for $\eta = 0,1,\ldots,256$.

(ii) Compute the N =512-point magnitude spectrum of the waveform based on a Hamming window and an stDFT. (Note: The stDFT and conventional DFT are equivalent here because only the magnitude spectrum is required.)

(iii) Repeat steps (i) and (ii) after center clipping the waveform according to:

$$C[s(n)] = \begin{cases} s(n) - C^+, & s(n) > C^+ \\ 0, & C^- \le s(n) \le C^+. \\ s(n) - C^-, & s(n) < C^- \end{cases} \qquad (4.57)$$

(b) Comment on the changes in both the autocorrelation and the spectrum. What do these changes indicate about the effects of the clipping operation on the waveform?

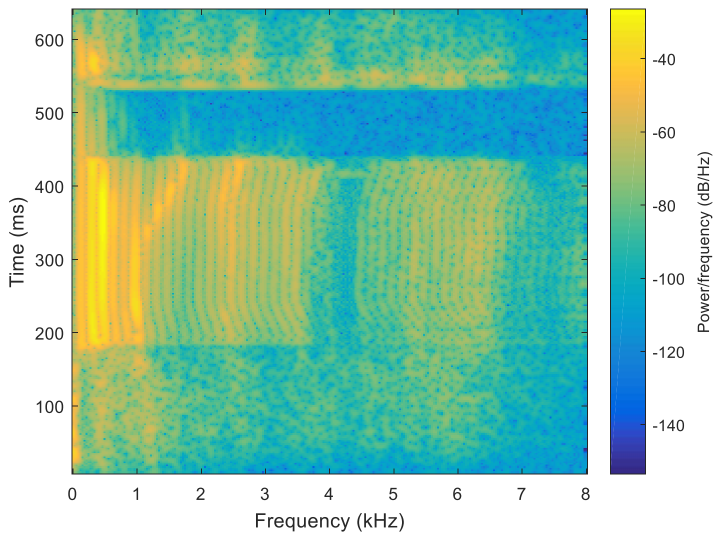
(c) Estimate the pitch using the two autocorrelation results. Which result would provide better performance in an automated procedure?

6. Write simple programs for the computation of the short-term AMDF and the general short-term autocorrelation estimator $r(71; m)$ given in (4.35).

$$\begin{aligned} r_s(\eta; m) &= \frac{1}{N} \sum_{n=-\infty}^{\infty} \mathcal{J}(\eta)\{s(n)w(m-n)\} \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} s(n)w(m-n)s(n-|\eta|)w(m-n+|\eta|). \end{aligned} \qquad (4.35)$$

Explore the use of different window types and lengths in estimating the pitch every 1000 samples for the /i/ glissando. (Note: The speaker of the glissando (pitch sweep) is an adult male and the sampling rate of the data is 16kHz. This should help reduce the number of ηs considered. This is important to reduce the amount of computation to a reasonable level.)

7. (a) Using the overlapped windows, write a program to compute stDTFT, and plot the result as this:



(b) Apply your program on data of Question 3, and compare results with the "spectrogram" function of MATLAB.

(c) Do the question #3 and compare some signals of /oo/ according to the spectrogram.