UCLA — Electrical Engineering Dept.
EE114: Speech and Image Processing — Midterm Exam
Wednesday, April 27, 2016

Solutions

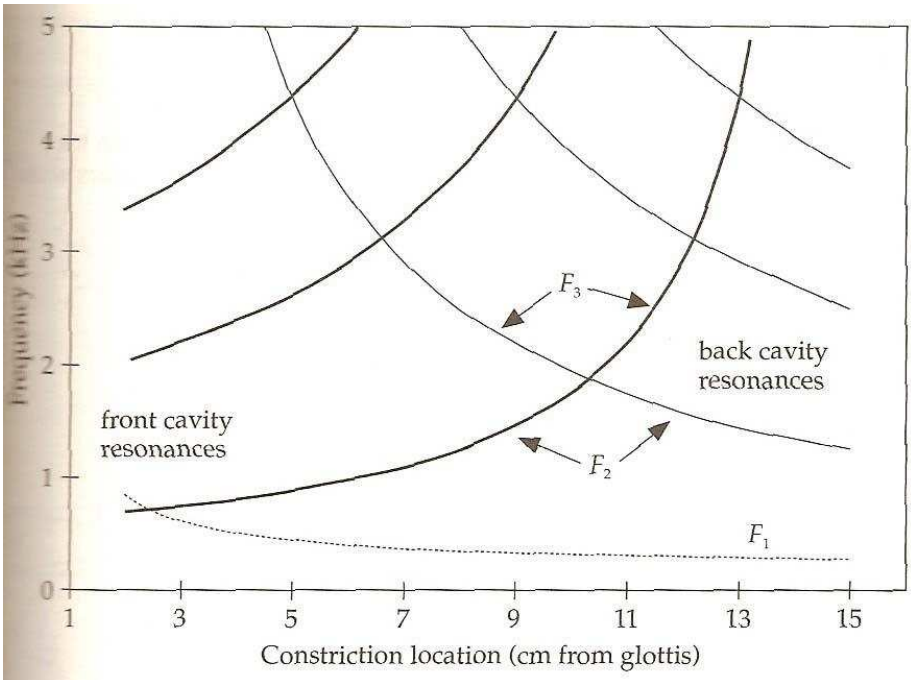


Figure 1: Nomogram for a three-tube model.

1.

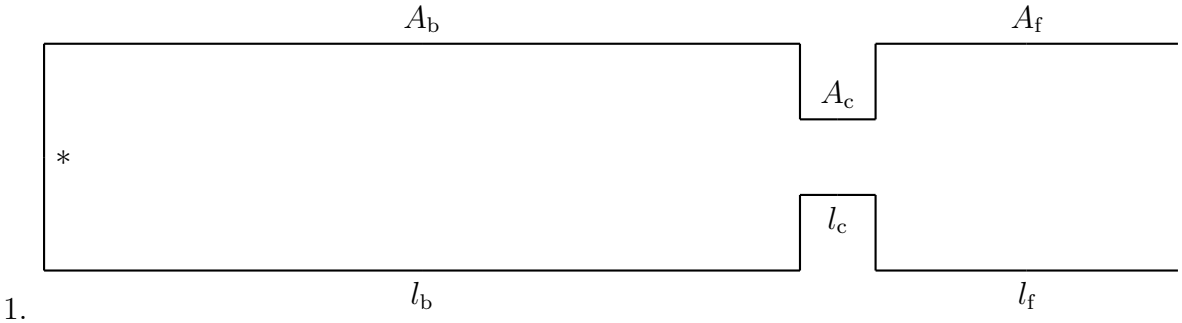


Figure 2: Three-tube model for the high vowel IY.

Fig. 1 displays the nomogram for the three-tube model of the high vowel IY of Fig. 2. The lengths of the back cavity, the constriction, and the front cavity are $l_b = 11\,\text{cm}$, $l_c = 0.5\,\text{cm}$, and $l_f = 4.5\,\text{cm}$, respectively.

(a) (4 points) Compute the first two resonant frequencies of the back cavity, $F_1^{(b)}$ and $F_2^{(b)}$, and verify that your results match the values given in the nomogram.

Solution: The back cavity can be modeled as a tube closed on both sides, i.e., a half-wavelength tube. The resonant frequencies are $F_i^{(\mathrm{b})} = i\frac{c}{2l_\mathrm{b}}$, for $i = 1, 2$, and they are given by 1590 Hz and 3180 Hz, respectively.

(b) (4 points) Compute the first two resonant frequencies of the front cavity, $F_1^{(\mathrm{f})}$ and $F_2^{(\mathrm{f})}$, and verify that your results match the values given in the nomogram.

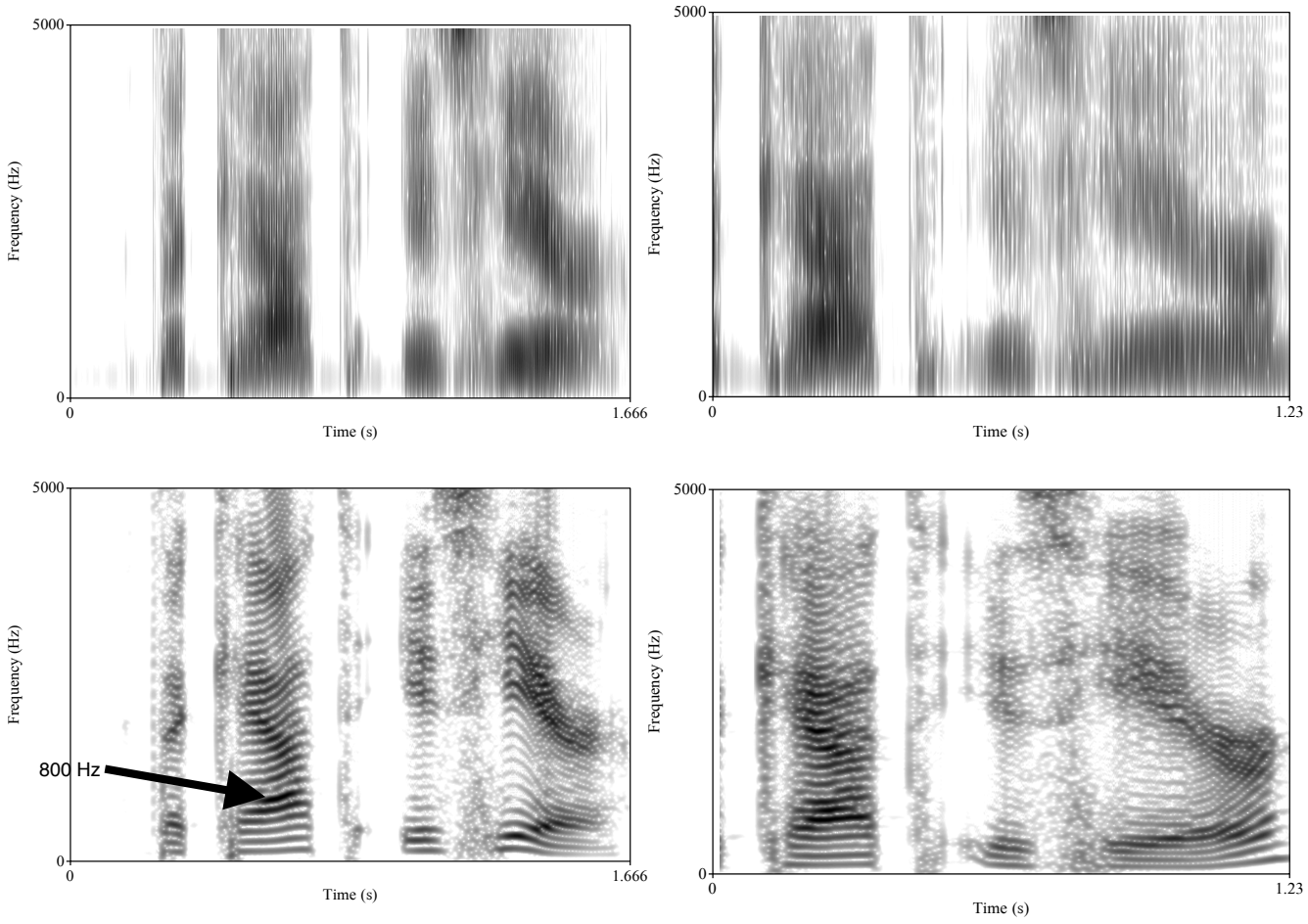
Solution: The front cavity can be modeled as a tube open on one side and closed on the other, i.e., a quarter-wavelength tube. The resonant frequencies are $F_i^{(\mathrm{f})} = (2i-1)\frac{c}{4l_\mathrm{f}}$, for $i = 1, 2$, and they are given by 1944 Hz and 5832 Hz, respectively. The higher frequency is not visible in the nomogram.

(c) (2 points) What is the approximate value of the Helmholtz frequency? (Use the nomogram for this.)

Solution: The Helmholtz frequency is the lowest resonant frequency of the three-tube model. From the nomogram, the frequency is approximately 300 Hz.

(d) (6 points) What are the formants for this particular vowel? Mark their values on the nomogram.

Solution: The formants are 300 Hz, 1590 Hz, and 1944 Hz.



(a) One spectrogram.

(b) Another spectrogram.

Figure 3: Two spectrograms

2. (a) (4 points) Consider the spectrograms in Figs. 3a and 3b. Which of these is a question? Why?

Solution: Spectrogram 2 has a rising pitch at the end, and it is the question.

For the next questions, consider Fig. 3a.

(b) (4 points) Which of the spectrograms is obtained by wideband analysis and which by narrowband analysis? Explain your answer (consider both time and frequency characteristics).

Solution: The top one is obtained by wideband analysis: the spectral features are blurred and the temporal transitions are sharper.

(c) (6 points) Estimate the length in samples of the Hamming window used in the spectrogram obtained by narrowband analysis. The sampling frequency is $F_\mathrm{s} = 48\,000\,\mathrm{Hz}$.

> **Solution:** Each line is separated by 133.33 Hz. The main lobe of a Hamming window is $4F_\mathrm{s}/N_\mathrm{w} \leq 133.33$, therefore $N_\mathrm{w} \geq 1440$.

(d) (4 points) What is the pitch during the utterance of the first word? Was the speaker a man or a woman? (Remember that adult male pitch is in the range of 80 Hz to 200 Hz, and an adult female pitch is within 150 Hz to 350 Hz.)

> **Solution:** Similarly to the previous question, the first spectral line is at 133.33 Hz, which points to an adult male speaker.

(e) (4 points) In the spectrograms, see if you can identify the stops, fricatives, and voiced sounds.

3. Consider the speech segment $s(n) = [-1, 1, -2\underline{1}, 2, -1, -2, -1, 0, 1]$. Let

$$R(i) = \sum_{m=-\infty}^{\infty} x(m)x(m-i), \quad i = 0, \ldots, N_w - 1$$

and

$$\phi(i,k) = \sum_{n=0}^{p} s(n-k)s(n-i), \quad 1 \le i \le p, 0 \le k \le p.$$

Suppose $x(n) = s(n)w(n)$, where $w(n) = 1$, $n = 0, \ldots, 3$, and $w(n) = 0$ elsewhere.

(a) (6 points) Find $R(0)$, $R(1)$, and $R(2)$.

Solution: $R(0) = 1 + 4 + 1 + 4 = 10$, $R(1) = 2 - 2 + 2 = 2$, $R(2) = -1 - 4 = -5$.

(b) (6 points) Find the 2nd-order prediction coefficients, a_1 and a_2, using the autocorrelation method of linear prediction analysis.

Solution:

$$\mathbf{R} = \begin{bmatrix} 10 & 2 \\ 2 & 10 \end{bmatrix}, \quad \mathbf{R}^{-1} = \frac{1}{104}\begin{bmatrix} 10 & -2 \\ -2 & 10 \end{bmatrix} = \frac{1}{52}\begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}$$

$$\mathbf{a} = \mathbf{R}^{-1}\mathbf{r} = \frac{1}{52}\begin{bmatrix} 5 & -1 \\ -1 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \end{bmatrix} = \frac{1}{52}\begin{bmatrix} 15 \\ -27 \end{bmatrix}$$

$$\therefore a_1 = \frac{15}{52}, \quad a_2 = -\frac{27}{52}$$

(c) (4 points) Find the corresponding error, $E_{\min}$.

Solution:

$$E_{\min} = R(0) - a_1 R(1) - a_2 R(2) = 10 - \left(\frac{15}{52}\right)2 - \left(\frac{27}{52}\right)3 = \frac{409}{52}$$

(d) (6 points) Compute the expression for the poles of the corresponding vocal tract model. Are they real or complex conjugate? Write the expression (without computing it) for the frequency (in Hertz) of the formant, F_1. Let the sampling frequency be F_s Hz.

Solution: We need to compute the roots of the second-order polynomial $52z^2 - 15z + 27$, $z = (15 \pm \sqrt{225 - 4 * 27 * 52})/104 = (15 \pm \mathrm{j}\sqrt{5391})/104$. The formant can be computed

from the arctangent of the ratio of the imaginary part and the real part: $F_1 = F_\mathrm{s} \tan^{-1}\left(\frac{\sqrt{5391}}{15}\right)/(2\pi)$.