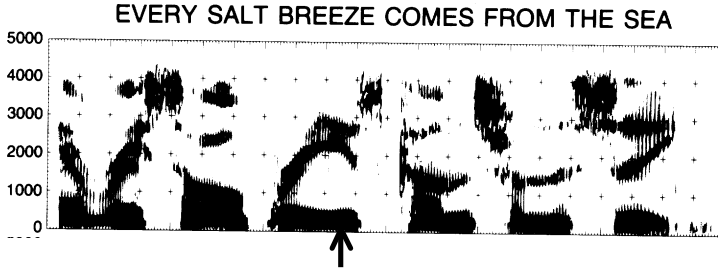
Due Wednesdayy, November 4 2009
in class or by 5:00 PM in your TA's dropbox on Rhea (zhang253).

EE 438 **Assignment No. 7** **Fall 2009**

1. Consider the spectrogram shown below for the utterance "Every salt breeze comes from the sea."



   a. Is this a wideband or narrowband spectrogram?

   b. If your answer to part a, was "wideband", sketch what a narrowband spectrogram for this same signal would look like. On the other hand, if your answer to part a, was "narrowband", sketch what a wideband spectrogram for this same signal would look like.

   c. Assuming the entire utterance lasted 2 sec,*very* roughly estimate the pitch period.

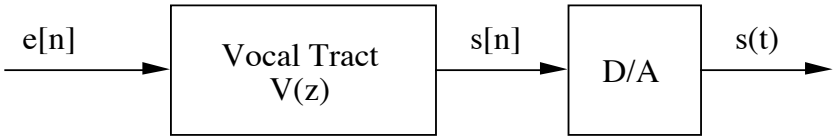
   d. Identify the formant frequencies at the time marked by the arrow.

   e. What phoneme do you think is being uttered at this point? Support your answer by comparison with the formant frequencies in the table below:

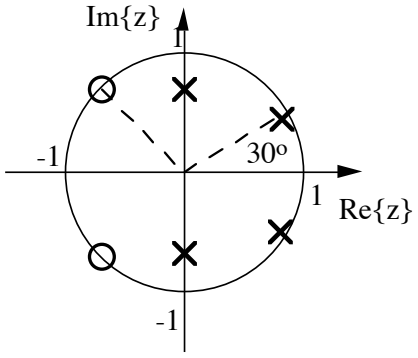
TABLE 2.2. Formant frequencies for typical vowels.

ARPABET Symbol for Vowel	IPA Symbol	Typical Word	F_1	F_2	F_3
IY	/i/	beet	270	2290	3010
IH	/ɪ/	bit	390	1990	2550
EH	/ɛ/	bet	530	1840	2480
AE	/æ/	bat	660	1720	2410
AH	/ʌ/	but	520	1190	2390
AA	/ɑ/	hot	730	1090	2440
AO	/ɔ/	bought	570	840	2410
UH	/ʊ/	foot	440	1020	2240
UW	/u/	boot	300	870	2240
ER	/ɝ/	bird	490	1350	1690

2. The digital synthesizer for voiced speech shown below operates at a 10 kHz sampling rate.



The excitation is given by $e[n] = \sum_{k=-\infty}^{\infty} \delta[n - 50k]$. The vocal tract transfer function $V(z)$ has poles and zeros at the locations shown below:



a. What is the pitch period in seconds?

b. Find the formant frequencies in Hz, and rank them according to their strength, *i.e.* how peaked the vocal tract response is at the corresponding frequency.

c. Sketch what a *wideband* spectrogram would look like for this utterance. Be sure to label the pitch and formant information appropriately.

d. Sketch what a *narrowband* spectrogram would look like for this utterance. Be sure to label the pitch and formant information appropriately.

3. Consider the STDTFT defined as

$$X(\omega, n) = \sum_{k} x[k]w[n-k]e^{-jwk}$$

where $x[n]$ is the speech signal and $w[n]$ is the window sequence. Prove the following properties

a. Linearity – if $v[n] = ax[n] + by[n]$, then $V(\omega, n) = aX(\omega, n) + bY(\omega, n)$.

b. Shifting – if $v[n] = x[n - n_0]$, then $V(\omega, n) = X(\omega, n - n_0)e^{-j\omega n_0}$.

c. Modulation – if $v[n] = x[n]e^{j\omega_0 n}$, then $V(\omega, n) = X(\omega - \omega_0, n)$.

d. Show that $X(\omega, n)$ can be put in the form

$$X(\omega, n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\theta)e^{j\theta n}X(\omega + \theta)d\theta$$

i.e. $X(\omega, n)$ is a smoothed spectral estimate of $X(\omega)$ at frequency ω.

4. Consider the signal

$$x[n] = \begin{cases} \cos(\pi n / 8), & n < 0 \\ \cos(\pi n / 3), & n \geq 0 \end{cases},$$

and assume a rectangular window

$$w[n] = \begin{cases} 1, & |n| < 25 \\ 0, & \text{else} \end{cases}$$

a. Compute the STDTFT as defined in the previous problem for the following cases:

   i. n < -25

   ii. n > 25

   iii. n = 0

b. Sketch $|X(\omega,n)|$ for all n. Be sure to label important dimensions.