UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering
Instructor: Mark Hasegawa-Johnson

ECE 537 Speech Processing

Problem Set 1
Fall 2009

Issued: Wed Aug. 26, 2009 **Due:** Wed Sept. 2, 2009

Reading for problem set 1: Flanagan, Allen & Hasegawa-Johnson chapters 1, 2, 3.1-2

Problem 1.1

Here are some English sentences, written in IPA. Transliterate into the Latin alphabet.

(a) wek for ðə sʌn hu skætərd ðə stɑrs bifor hɪm frʌm ðə fild ʌv naɪt draɪvz naɪt əlɔŋ wɪθ ðɛm frʌm ðə fild ʌv hɛvn

Solution: Wake! For the sun, who scattered the stars before him from the field of night, drives night along with them from the field of heaven (Omar Khayyam).

(b) lɪvɪŋ ɔn ði ɛdʒ ɪz pɛrɪləs bʌt ðə vju mor ðæn kɑmpənsets

Solution: Living on the edge is perilous, but the view more than compensates (Cassettari)

(c) bjuti ɪz truθ truθ bjuti ðæt ɪz ɔl ji no ɔn ərθ ænd ɔl ji nid tu no

Solution: 'Beauty is truth, truth beauty.' — that is all ye know on Earth, and all ye need to know. (John Keats)

Problem 1.2

Here are some English sentences, written in the Latin alphabet. Transliterate them into IPA (you need not transliterate the author's names).

(a) Remember that nothing worth learning can be taught. (Oscar Wilde)

Solution: rimɛmbər ðæt nʌθɪŋ wərθ lərnɪŋ kæn bi tɔt

(b) Eccentricity has always abounded where strength of character has abounded. (John Stuart Mill)

Solution: ɛksɛntrɪsɪti hæz alwez ʌbaʊndəd wɛr strɛŋθ ʌv kærʌktər hæz əbaʊndəd

(c) Change will soon be replaced by something new. (James Anglin)

Solution: tʃendʒ wɪl sun bi riplesd baɪ sʌmθɪŋ nu

Problem 1.3

When infants first learn to speak, some infants learn the consonants [m,b,d] and the vowels [a,i,u] before they learn any other sounds. Let's imagine a language called Infantese that contains only these six phonemes. Let's imagine that their relative frequencies are the same as in English, meaning that, for any two phonemes $(x, y) \in$[a,i,u,m,b,d], the Infantese frequencies $P_I(x)$ and $P_I(y)$ are related to the English-language frequencies $P_E(x)$ and $P_E(y)$ (given in Table 1.1 of the text) by

$$\frac{P_I(x)}{P_I(y)} = \frac{P_E(x)}{P_E(y)} \tag{1}$$

(a) Specify the probabilities $P_I(x)$ and information content $I(x)$ for all six phonemes in the Infantese language.

Solution: The probabilities of [a,i,u,m,b,d] are 0.27, 0.12, 0.09, 0.16, 0.10, and 0.25 respectively. These symbols contain 1.9, 3.02, 3.42, 2.63, 3.25, and 2.0 bits of information respectively. Notice that, for example, [d] contains exactly 2 bits of information, because its occurrence probability is exactly 2^{-2}.

(b) What is the entropy of Infantese, in bits per phoneme? How does this compare to the maximum possible entropy of a six-bit alphabet?

Solution: The entropy is the average information content, which is $\sum p(x) \log_2(1/p(x)) = 2.46$ bits. This is pretty close to the maximum possible entropy of a six-bit alphabet, which is only $log_2(6) = 2.58$ bits.

(c) Suppose that all syllables in Infantese have CV form, i.e., each syllable has exactly one consonant, and exactly one vowel. In this case, the information content per symbol drops. For example, suppose that you know that the next phoneme will be a consonant; how much *additional* information is provided by observing that it is an [m]?

Solution: The probability $p(\text{m}|\text{consonant}) = 0.31$, base-2 logarithm of which is $I(\text{m}|\text{consonant}) =$ 1.69 bits. Notice that this is about 1 bit less than the original $I(\text{m})$; knowing that the next phoneme is a consonant rather than a vowel is worth $log_2(1/p(\text{consonant})) \approx 1$ bit.

(d) Under the assumption in part (c), what is the entropy of Infantese, in bits/symbol?

Solution: The frequency of phonemes is the same as in part (b), but the average information per symbol is less; the total entropy is now only 1.47 bits/symbol.

(e) In a normal room, human beings understand one another with very high accuracy. Suppose that the spoken phoneme string is $X = [x(1), x(2), \ldots]$ and the heard phoneme string is $Y = [y(1), y(2), \ldots]$. With probability $(1 - \eta)$, the listener correctly hears the phoneme; with probability η, the listener fails to hear the phoneme, and must guess. If the listener must guess a consonant, she chooses a consonant at random, with probabilities proportional to $P_I(x)$; if she must guess a vowel, she chooses a vowel at random, with probabilities proportional to $P_I(x)$. What is the conditional entropy $H(y|x)$?

Solution: $p_x(x)$ is as given in part (c). The marginal distribution of y is the same as that of x, i.e., it is given by $p_x(y)$. The joint PDF is $p(x, y) = p_x(x)((1 - \eta)\delta(y - x) + \eta p_x(y))$, where

$\delta(y-x)$ is the unit pulse function. The conditional entropy is

$$\begin{aligned} H(y|x) &= -\sum_x \sum_y p(x,y) \log\left(\eta p_x(y) + (1-\eta)\delta(y-x)\right) \\ &= -\sum_x \sum_y p(x,y) \left(\log \eta + \log p_x(y) + \log(1 + \frac{(1-\eta)\delta(y-x)}{\eta p_x(y)})\right) \\ &= -\log \eta + H(y) - \sum_x \sum_y p(x,y) \log(1 + \frac{(1-\eta)\delta(y-x)}{\eta p_x(y)}) \end{aligned}$$

where $H(y) = H(x)$ because they have the same PDFs. The terms in the last summation are equal to $\log(1) = 0$ except when $x = y$ (that's what $\delta(y-x)$ means), so

$$\begin{aligned} H(y|x) &= H(y) - \log \eta - \sum_y p_x(y) \log(1 + \frac{1-\eta}{\eta p_x(y)}) \\ &= H(y) - \log \eta - \sum_y p_x(y) \left(\log(\frac{1-\eta}{\eta p_x(y)}) + \log(1 + \frac{\eta p_x(y)}{1-\eta})\right) \\ &= -\log(1-\eta) - \sum_y p_x(y) \log(1 + \frac{\eta p_x(y)}{1-\eta}) \\ &\approx \eta - \left(\frac{\eta}{1-\eta}\right) \sum_y p_x^2(y) \approx \eta G(y) \end{aligned}$$

where $G(y) \equiv 1 - \sum_y p^2(y)$ is called the "Gini impurity" of y, and behaves a lot like $H(y)$, e.g., it reaches its maximum value if y is uniformly distributed, and reaches its minimum value of $G(y) = H(y) = 0$ if y is a constant. Another interesting approximation is obtained using $H(y|x) \approx H(y, g|x)$ where $g = 1$ if the subject guesses, $g = 0$ if the subject is not guessing; the result gives

$$H(y|x) = \eta H(y)$$

(f) What is the mutual information $I(x, y)$?

Solution: $I(x, y) = H(y) - H(y|x)$. Because the listener guesses according to $P_I(x)$, $P_I(y) = P_I(x)$, therefore $H(y) = H(x) = 1.47$ bits, and $I(x, y) = 1.47(1-\eta)$.

(g) What is the channel capacity $C(\eta)$ of the channel described in part (d), as a function of η? How must the probabilities $P_I(x)$ be modified so that $I(x, y) = C(\eta)$?

Solution: Because the probability of an error is independent of x, the mutual information is the source entropy times $(1-\eta)$, and we can maximize $I(x, y)$ by maximizing $H(x)$. Maximum entropy is achieved when each phoneme occurs with frequency 1/6, and the corresponding mutual information is $C(\eta) = 1.58(1-\eta)$.

(h) Instead of limiting the alphabet to just six phonemes, suppose that it were possible to increase the size of the alphabet without bound, while maintaining a constant η. What is the highest channel capacity that you could achieve under this assumption?

Solution: If we could choose from an infinite number of equiprobable phonemes, then the source entropy would be infinite, therefore the channel capacity would also be infinite.

(i) Obviously, the channel capacity in part (h) is not possible in any real-world system, therefore there must be something wrong with the assumptions on which it is based. What is the error?

Solution: In the real world, it is never possible to achieve $\eta < 1$ for an arbitrarily large source alphabet. The error probability usually increases every time you add a new symbol to the alphabet; under certain conditions that we will discuss later in the course, this relationship can be made strictly non-decreasing.

Problem 1.4

(a) What is the formula for the speed of sound, in terms of the material properties of the fluid? Solution: The formula is $c = \sqrt{\eta P_0/\rho_0}$.

(b) Identify the variables in your answer to part (a): Names, units, and values of consonants.

Solution: On the left is the speed of sound c [meters/second], and on the right are physical parameters $\eta \equiv c_p/c_v = 1.4$ [dimensionless], the barometric pressure (e.g., at sea level) $P_0 = 10^5$ [Pa], and the density (e.g., at sea level at zero degrees Celsius) $\rho_0 = 1.29$ [kg/m^3].

(c) What is the meaning of ηP_0? Solution: This combination of variables represents the adiabatic compressibility of air. The $\eta = c_p/c_v$ results from holding the temperature constant during the cycle of the wave. Heat diffusion is slow compared to the cycle at acoustic frequencies, such that the thermal energy is trapped in the air.

(d) Does P_0 on the surface of the Earth depend on temperature? Explain?

Solution: The barometric pressure depends on the weight of the air above us. While its density depends on temperature, the total weight is a constant, and is therefore independent of temperature. The barometric pressure depends on wind and weather (e.g., humidity), and on altitude (of course), but not on temperature.

(e) Does ρ_0 depend on temperature, P_0, both, or neither? Explain.

Solution: Yes, as given in class

$$\rho(T, P_0) = 1.29 \frac{273}{T} \frac{P_0}{10^5}, \tag{2}$$

where T is the temperature in degrees Kelvin.

(f) What is the form of the dependence of the speed of sound on temperature? Give the formula for $c(T)$, and explain it in words.

Solution: Since $c = \sqrt{\eta P_0/\rho_0}$ and following the state equation for a gas, $\rho_0 \propto 1/T$, we may conclude that $c(T) \propto \sqrt{T}$, where T is in degrees Kelvin. For example, between the freezing point of water (273 K) and the boiling point of water (373 K), the speed of sound changes by a factor of $\sqrt{373/273} = 1.17$.

Problem 1.5

(a) Write out the 2x2 matrix equation that describes the propagation of 1 dimensional sound waves in a tube having area A.

Solution: As shown in class the basic equations are:

$$\frac{d}{dx}\begin{bmatrix} P(x,\omega) \\ U(x,\omega) \end{bmatrix} = -\begin{bmatrix} 0 & \mathcal{Z}(x,s) \\ \mathcal{Y}(x,s) & 0 \end{bmatrix}\begin{bmatrix} P(x,\omega) \\ U(x,\omega) \end{bmatrix}. \tag{3}$$

where $\mathcal{Z} = s\rho_0/A$ and $\mathcal{Y} = sA/\eta P_0$, with A the area of the tube.

(b) Rewrite these equations as a second order equation in terms of the pressure P, and thereby find the formula for the speed of sound in terms of Z and Y: Solution: If we let P' mean the partial with respect to space, then

$$P' + ZU = 0 \tag{4}$$

and

$$U' + YP = 0. \tag{5}$$

Taking the partial wrt x of the first equation, and then using the second, gives

$$P'' + ZU' = P'' - \mathcal{Z}\mathcal{Y}P = 0. \tag{6}$$

Since the wave equation is

$$\frac{\partial^2 P}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 P}{\partial x^2} \tag{7}$$

by inspection we see that $\omega^2/c^2 = \mathcal{Z}\mathcal{Y}$, which results in the final formula for the speed of sound.

Problem 1.6

Assume that a flash bulb puts out 50 Joules when it is triggered, and it lasts for 20 μs. How much power is delivered while the flash is present, assuming the illumination is constant during the interval?

Solution: $50/20 \times 10^{-6}$ is 2.5MJ.

Problem 1.7

A bottle has a neck diameter of 1 [cm] and is l =1 cm long. It is connected to the body of the bottle "barrel" which is 5 cm in diameter and L =10 cm long. Treat the barrel as a short piece of transmission line, closed at one end, which looks like a compliance $C = V_{barrel}/\eta P_0$, and the neck which look like a mass $M = \rho_0 l/A_{neck}$. These two impedances are in series, since they both see the same volume velocity (flow).

(a) Write out the formula for the resonant frequency in terms of the dimensions of the bottle.

Solution: The formula for the Helmholtz resonator is

$$f_0 = \frac{c}{2\pi}\sqrt{A/Vl}, \tag{8}$$

where A, l are the area and length of the neck and V is the volume of the bottle.

(b) Calculate the resonant frequency in Hz for the dimensions given.

Solution: From the numbers given $A = \pi.005^2$ [m^2], $l = 0.01$ while $V = \pi 0.025^2 \times 0.1 = 1.9 \times 10^{-4}$ [m^3]. Thus $f_0 = \frac{345}{2\pi}\sqrt{0.01 \times \pi 0.005^2/0.0025} \approx 347.3$ [Hz].

(c) Blow into a blottle and measure the resonant frequency by recording the tone, and taking the FFT of the resulting waveform, and finding the frequency (Extra credit).