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

ECE 537 Speech Processing

Problem Set 6

Fall 2009

Issued: Wed Oct. 14, 2009 **Due:** Wed Oct. 21, 2009

Reading for problem set 6: Fletcher & Munson, 1933

Problem 6.1

Maria Asymptotica, the opera singer, is well known for her ability to sing arbitrarily long notes. On one afternoon in Paris, her recording company created a CD called "Three Hours of G#" which contained a single sustained [a], three hours long, on a tone halfway between G and G#, specifically, $\Omega_0 = 2\pi 200$ radians/second. Because Maria has such a beautiful mellow voice, her three-hour recording is well represented by the sum of just ten harmonics:

$$p(t) = \sum_{k=1}^{10} |H(k\Omega_0)| \cos(k\Omega_0 t + \theta_k),$$

The vocal tract transfer function is given by

$$H(s) = \frac{G}{(s-s_1)(s-s_1^*)(s-s_2)(s-s_2)^*}$$

Her formant frequencies are

$$s_1 = j2\pi 900 - 100\pi, \quad s_2 = j2\pi 1100 - 200\pi$$

(a) Maria sings loudly, so the total intensity level at the microphone was 91dB SPL. Find the intensity level of each harmonic, $\beta_k = 20\log_{10}(|H(k\Omega_0)|/\sqrt{2}P_{ref})$, in dB SPL.

Solution: Let $H(s) = G \times T(s)$. The intensity of the sound is

$$91\text{dB} = 20\log_{10}\left(G\sqrt{|T(\Omega_0)|^2 + \ldots + |T(10\Omega_0)|^2}/\sqrt{2}P_{REF}\right)$$

where $P_{REF} = 20\mu$Pa, and $\Omega_0 = 2\pi 200$ radians/second. Solving, we find that $G = 5.8 \times 10^{13}$. The level of the kth harmonic is $\beta_k = 20\log_{10} G|H(k\Omega_0)|/\sqrt{2}P_{REF}$; these levels are $\vec{\beta} = [66, 69, 73, 84, 89, 79, 66, 58, 51, 46]$dB.

(b) Find the loudness level L_k of each harmonic, in dB SPL, using information in Fletcher & Munson's 1933 paper.

Solution: Sensation level (loudness level or SL) equals sound pressure level (SPL) between about 400Hz and 1400Hz for intensity levels between 50dB and 90dB. At 200Hz and 66dB, SL $\approx$ SPL $-$ 6dB; at 2000Hz and 46dB, SL $\approx$ SPL + 1dB. So the sensation levels are $\vec{L} = [60, 69, 73, 84, 89, 79, 66, 58, 51, 47]$dB SL.

(c) Find the loudness $G(L_k)$ of each harmonic. In order to accomplish this, you may use the tables in Fletcher & Munson's 1933 paper, or you may use the cube-root-intensity rule given in class.

Solution: Using the cube-root intensity rule, $G_{1k} = I_{ref}^{1/3} 10^{L_k/30}$, gives

$$\vec{G}_1 = 10^{-4} \times [100, 200, 271, 631, 926, 430, 158, 86, 50, 37]$$

Using Table III in the paper gives

$$\vec{G}_2 = [4350, 7440, 9850, 23100, 35000, 15800, 6240, 3820, 2350, 1780]$$

At sufficiently high levels, these differ by a scalar, $\vec{G}_2 = 40000\vec{G}_1$; at lower levels, the cube root model fails.

(d) Assume that each harmonic is masked by the harmonic just below it, in frequency, e.g., the 800Hz harmonic is masked by the 600Hz harmonic, and so on. The unmasked audibility of each harmonic is $0 \leq b_k \leq 1$. Fletcher & Munson claimed that b_k is the product of three terms:

$$b_k = g_1(f_k - f_{k-1}) g_2(L_k - L_{k-1}) b_3(f_k, \beta_k)$$

(1) Find the frequency-dependent part of the unmasked audibility, $b_1(f_k - f_{k-1})$, for each of the harmonics $2 \leq k \leq 10$.

Solution: $b_1(f_k - f_{k-1}) = (250 + f_k - f_{k-1})/1000 = 0.45$ for all components, $2 \leq k \leq 10$.

(2) Find the loudness-dependent part of the unmasked audibility, $b_2(L_k - L_{k-1})$, for each of the harmonics $2 \leq k \leq 10$.

Solution: The normalizing constant T depends on $f_k - 2f_{k-1}$, which is zero for $k = 2$, and negative for all other k. Therefore $T_2 = 32$, and $T_k = 25$ for $k > 2$.

$$\begin{aligned} b_k &= 10^{(L_k - L_{k-1})/T} \\ &= [10^{9/32}, 10^{4/25}, 10^{11/25}, 10^{5/25}, 10^{-10/25}, 10^{-13/25}, 10^{-8/25}, 10^{-7/25}, 10^{-4/25}] \\ &= [1.91, 1.45, 2.75, 1.58, 0.40, 0.30, 0.48, 0.52, 0.69] \end{aligned}$$

(3) Find the maskability correction, $b_3(f_k, \beta_k)$, for each of the harmonics $2 \leq k \leq 10$.

Solution:

$$\beta_k + 30 \log_{10} f_k - 95 = [52, 61, 76, 84, 76, 65, 59, 54, 51]$$

$$b_3(f_k, \beta_k) = Q(\beta_k + 30 \log_{10} f_k - 95) = [0.89, 0.88, 0.97, 1.13, 0.97, 0.88, 0.88, 0.88, 0.9]$$

(4) Find the total unmasked audibility b_k for each of the harmonics, $2 \leq k \leq 10$. Solution:

$$b_k = \min\left(1, b_1(f_k, f_{k-1}) b_2(L_k, L_{k-1}) b_3(\beta_k, f_k)\right)$$

$$= [0.765, 0.574, 1.00, 0.803, 0.175, 0.1189, 0.190, 0.206, 0.280]$$

Notice that the fourth harmonic, 800Hz at 84dB SPL, is completely unmasked, but the fifth harmonic, 1000Hz at 89dB SPL, is slightly masked.

(e) What is the total loudness level of Maria Asymptotica's 91dB tone? Solution: Total loudness is $G = \sum_k b_k G_k$, where $b_1 = 1$ (no other component masks the fundamental). Using the cube-root intensity rule gives $G = 0.191$ (using Table III loudnesses gives $G = 7213$). Compare this to some other ways these harmonics might have combined:

- If we added the loudnesses of all harmonics without masking, we would get $G = 0.289$.
- Suppose we played a single tone, at some frequency between 500 and 1500Hz (so that $L_k = \beta_k$), at the same intensity level of $\beta_k = 91$dB. The loudness would be $G = I_{ref}^{1/3} 10^{91/30} = 0.108$.
- Summarizing these results: spreading the power out among multiple harmonics acts to increase the effective loudness of the sound, because of the cube-root nonlinearity. Masking acts to bring the loudness back down a little bit, reducing the effectiveness of spread-spectrum audio encoding. This is not surprising, if you think about it: masking is caused because the cochlear filters add up power from neighboring sinusoids before computing the cube root nonlinearity, therefore the total loudness relationship is somewhere between $I = \sum_k I_k$ (the correct formula when all components are inside one critical band) and $G = \sum_k G_k$ (the correct formula when each component is more than a critical band away from any other).