Questions

1. (a) For a space-time sine function

$$I(X,Y,Z,t) = \sin(2\pi(k_0 X + k_1 Y + k_2 Z + \omega t))$$

that satisfies the wave equation,

$$(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2} + \frac{\partial^2}{\partial Z^2}) I(X,Y,Z,t) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2} I(X,Y,Z,t)$$

what are the constraints between the frequencies k_X, k_Y, k_Z and ω and the constant v ?

(b) The units of v are ms^{-1}. What is the three dimensional velocity $\vec{v}$ of such a sine wave?

(c) We saw with 2D image motion that by assuming image intensity of a moving point is constant over time, we could derive a "motion constraint equation". Is there an analogous motion constraint equation here in the case of a 3D sound wave ? (This involves some physical intuition.)

2. (a) Let ϕ be the angle between a distant sound source and the interaural axis, i.e. the line between the ears. This angle defines a "cone of confusion", with apex at the center of the head.

Let $\phi = 0$ point out from the left ear and $\phi = \pi$ point out from the right ear. Let the distance between the ears be D, what is the interaural time of arrival delay τ, as a function of ϕ?

(b) Compare two sounds from two nearby angles, ϕ vs. $\phi + \Delta\phi$. To discriminate the two directions, you need to discriminate the interaural arrival times. What is the difference between the delay in interaural arrival times ?

(c) Psychophysics experiments have shown that, for ϕ near $\frac{\pi}{2}$ (i.e. the "medial plane" which splits the head symmetrically left/right), we can discriminate differences in ϕ of around 2 degrees. What is the corresponding timing difference?

(d) We are poor at using timing differences to discriminate the different ϕ values near $\phi = 0$ or $\phi = \pi$. Why? (The answer has to do with geometry, not with the brain.)

3. A spectrograph and a *musical score* both plot frequency versus time. But they are not the same thing. What are the similarities and differences?

For an example of a musical score, see: `https://commons.wikimedia.org/wiki/File:Twinkle_Twinkle_Sheet_Music.png`

4. (a) Describe (or sketch) the spectrogram produced by a vibrating string of length L.

(b) How many octaves are there between the fundamental frequencies of neighboring notes C and F on a musical scale?

(c) Suppose two notes on a piano keyboard are 25 keys apart (including the black and white keys). How many octaves apart are they? What is the ratio of the fundamental frequencies of these two notes?

(d) What does it mean to sing a vowel (e.g. "a" as in the word "say") at different notes, for example, a particular C versus a slightly higher note such as the first F above that C ? What does the singer need to do physically to change sounds?

BTW, one refers to notes as pitch.

5. (a) Suppose a person makes a voiced sound with a glottal pulse train frequency of 100 Hz. What is the period of this pulse train (i.e. the time interval between pulses) ?

   (b) The fundamental frequency of the pulse train is 100 Hz. What are the frequencies of the harmonics?

   (c) When can we observe these harmonics in a spectrogram? Assume that the period of the pulse train is indeed regular and has a long enough duration (enough pulses) so that the model holds, namely the model you will derive in Assignment 3 of how a sequence of impulses is Fourier transformed to a sequence of impulses in the frequency domain.

6. Suppose you drop an empty glass bottle on the ground and it doesn't break. The bouncing bottle will make a sound that is composed of a sequence of impacts which are approximately impulses, and each impact is followed by sounds whose frequency composition depends on where the impact occured on the bottle and what are the bottle's natural modes of vibration. Sketch the spectrogram that you would expect each such impact to produce. Briefly describe the features that are due to the impact, namely the impulse itself and the following sounds (resonance frequencies of the bottle).

Answers

1. (a) Taking partial derivatives gives:

$$k_0^2 + k_1^2 + k_2^2 = \frac{\omega^2}{v^2}$$
$$\text{which implies} \quad v^2 = \frac{\omega^2}{k_0^2 + k_1^2 + k_2^2}$$
$$\text{which implies} \quad v = \frac{\omega}{\sqrt{k_0^2 + k_1^2 + k_2^2}}$$

   (b) The function is constant on planes $k_0X + k_1Y + k_2Z + \omega t = c$. The wave travels perpendicular to such planes, namely in direction (k_0, k_1, k_2). The velocity is the vector of length v in that direction. Take the unit vector in the direction (k_0, k_1, k_2), namely $\frac{1}{\sqrt{k_0^2+k_1^2+k_2^2}}(k_0, k_1, k_2)$. The velocity $\vec{v}$ is the vector of length v in that unit direction, so

$$\vec{v} = \frac{\omega}{\sqrt{k_0^2 + k_1^2 + k_2^2}} \frac{1}{\sqrt{k_0^2 + k_1^2 + k_2^2}}(k_0, k_1, k_2) = \frac{\omega}{k_0^2 + k_1^2 + k_2^2}(k_0, k_1, k_2)$$

   (c) No. There is a key difference between the image motion constraint equation and the wave equation of a sound wave. For the image motion constraint equation, the image motion can have a motion component in the direction in which the intensity is constant. Such "parallel motion" does not occur with sound waves. Rather, sound waves are longitudinal. They always travel in the direction perpendicular to the local iso-pressure surface.

2. (a) From geometric reasoning, the sound has to travel an extra distance of $D \cos\phi$ from the left. From $D = vt$, the difference in arrival times is thus $\frac{D\cos\phi}{v}$, where $v = 343ms^{-1}$.

   (b)
$$\Delta(\frac{D\cos\phi}{v}) \approx \frac{d}{d\phi}(\frac{D\cos\phi}{v})\Delta\phi \approx -\frac{D\sin\phi}{v}\Delta\phi.$$

   (c) If $\phi = \frac{\pi}{2}$, then $\sin\phi = 1$ and so a change in ϕ causes a change in the delay between the two ears of $\frac{D}{v}\Delta\phi$. If $\Delta\phi = 2$ degrees (or $2\frac{\pi}{180}$ radians) and if the distance between the ears is about 10 cm (or 0.1 m), then the delay is about $\frac{1}{10\cdot 340} \cdot \frac{2\pi}{180} \approx 10^{-5}$ seconds, or $\frac{1}{100}$ ms. No, I didn't make a mistake. That is the sensitivity your brain has to timing differences.

   (d) If $\phi \approx 0$ or π, then $\sin\phi \approx 0$ and so a change in ϕ causes almost no change in the delay in interaural arrival times between the two ears. As such, the time delay gives poor precision about ϕ near those values. (The basic problem here is that we are in the flat part of the $\cos\phi$ function, i.e. near its peak.)

3. The frequencies axis for a musical score has increments on a log (frequency) scale, whereas spectrogram uses a linear frequency scale.

Only the fundamental is shown in a musical score. The overtones (or any other higher frequencies present) are not. These higher frequencies would be present in the spectrogram, though.

A (basic) musical score doesn't say how loud each note should be, whereas a spectrogram is grey scale.

4. (a) The spectrogram will have frequencies $\frac{c}{L}, \frac{2c}{L}, \frac{3c}{L}, \frac{4c}{L}, \ldots$. A spectrogram plots frequency as a function of time. You should have drawn a set of parallel horizontal lines, at frequencies defined by the sequence above.

   (b) There are 5 semitones from C to F, so we have a $\frac{5}{12}$ octave difference.

   (c) The notes are 25/12 octaves apart. The ratio of the fundamental frequencies of the higher to lower notes is $2^{\frac{25}{12}}$, that is, $(2^{\frac{1}{12}})^{25}$.

   (d) It means that the glottal pulse frequency is different for C versus F. To increase the frequency (”pitch”) from C to F, the singer needs to increase the frequency by increasing the tension in the vocal cords.

   Note that this has nothing to do with the fact that you are singing particular vowels “a” or “o”, etc.. The vowels are determined by the formant frequencies which are typically much higher than the glottal pulse frequency. See e.g.
   `https://en.wikipedia.org/wiki/Formant`

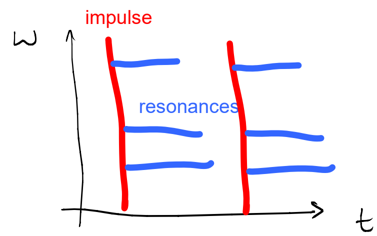
5. (a) The period is $\frac{1}{100}$ sec, or 10 ms.

   (b) The harmonics are integer multiples of the fundamentals: 200 Hz, 300 Hz, 400 Hz, etc.

   (c) No and yes. No because the articulators sometimes attenuate some frequencies so much that you wouldn't measure them. For example, the frequencies between the formants contain little 'energy' and the harmonics there would disappear.

   Yes, because glottal pulses sometimes are sometimes still visible within the frequency band covered by a formant. If you compute a spectrogram with good frequency resolution (well less than 100 Hz in the above example), then the harmonics of the glottal pulses would appear within a formant. You wouldn't know the harmonics that precisely, since the frequencies ω are still sampled in the spectrogram. But you should at least see some evidence of harmonics.

6. The impulse produces a vertical line in the spectrogram i.e. the Fourier transform of an impulse is a constant (over all frequencies). Most of these frequencies components dissipate immediately. The remaining ones (resonances) produce streaks that last over a brief time.



If the bottle bounces a few times, then this pattern will be repeated. I have shown two bounces only.