

Computational Perception

15-485/785

Assignment 2: Inference

Due: Tuesday, Feb. 12

1 Crosstalk cancellation

Simulating spatial audio is easy if you have headphones. You can just compute the signals you want sent to the left and right ears by convolving the original sound by the appropriate HRTFs, and deliver them directly into the appropriate ears. Things are trickier if all you have is a pair of stereo speakers, since in that case, both ears hear a mixture of the sounds coming out of each speaker. Still, if you know the HRTFs (that is, if the listener's head is in a known position with respect to the speakers), you can arrange to deliver the appropriate signals to each ear. This is called *crosstalk cancellation*, and we will derive it here.

Let θ_1 and θ_2 be the angles to the speakers (we're not bothering with elevation merely for notational convenience), X_1 and X_2 be the sounds (in frequency domain) being played at each speaker, and $H_L(\theta)$ and $H_R(\theta)$ be the left-ear and right-ear HRTFs for position θ .

1. (5 points) Write an expression for Y_L and Y_R , the (frequency domain) sounds heard by each ear, in terms of the X 's and H 's. You can assume this is just the sum of the contribution of each speaker.
2. (10 points) Now let Y_L and Y_R be the signals (in frequency domain) that you want to send to each ear (for spatialized audio, these would be some monaural sound convolved with some HRTFs). Solve for the X_1 and X_2 you need to play through the speakers so that the ears hear Y_L and Y_R . (If you are comfortable with matrix algebra, it is easiest to reformulate this as a matrix equation.)
3. (10 points) Unfortunately, the amplitude of the low frequency portion of all HRTFs is about equal. Why is this? This means you have a division-by-zero problem (or a noninvertible matrix) at those frequencies. How could you design a system to work around this problem?

2 Audition of the fly

Some types of flies are able to perform remarkable feats of sound localization. One particular type uses sound alone to identify singing crickets to use as a host in which to implant their larvae (nature can be quite brutal). These flies have two "ears", which are actually two small membranes, each which is just 1 mm^2 in area, and separated by only 0.5 mm. The remarkable thing about this

species is that they can use these tiny ears to localize crickets to within 2deg azimuth from several meters away.

1. (10 points) Describe the problems this system is faced with in performing sound localization. A fact you might find useful is that the cricket's song is an amplitude-modulated 5 kHz tone.
2. (10 points) Estimate the timing difference and intensity differences that would be required for the fly to localize sounds at the stated accuracy. What is the range of these acoustic cues that this fly would experience?
3. (10 points) Give some conjectures as to how the fly might accomplish this degree of sensitivity. Would it be possible knowing the constraints and limitations of our own auditory system? For those who are interested, I can provide references to this remarkable story, including some work micromechanical devices that have been designed to emulate the mechanisms used by the fly.

3 Inferring azimuth with the interaural intensity difference

As we have discussed in the class, the interaural intensity difference (IID) can be used for sound lateralization. The goal in this question is to estimate the azimuth when we (or the auditory system) observe some IID value(s). Assume that we have two IID functions $\Delta L_{f_1}(\theta)$ and $\Delta L_{f_2}(\theta)$, where f_1 and f_2 are some different frequencies of the sound, and θ is the azimuth (Figure 1). (We define these functions arbitrarily, but try to capture the general characteristics of the experimentally measured IID curves. These are defined in the MATLAB M-files and available on the course web site.)

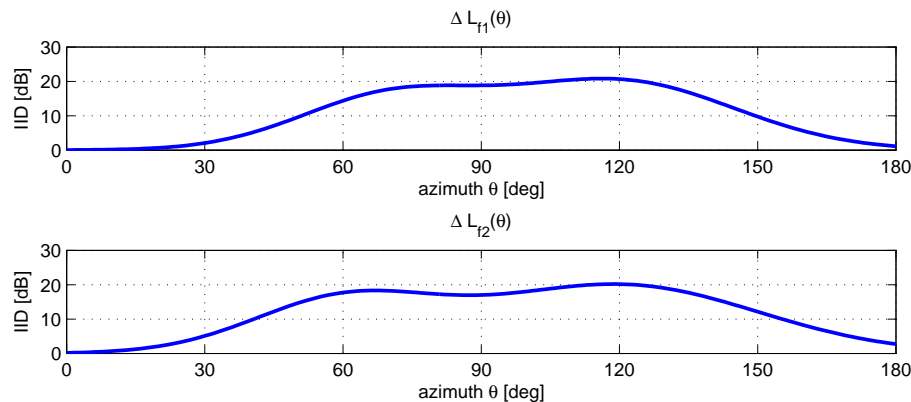


Figure 1: IID functions.

Our goal is to derive the most comprehensive expression about our certainty, or belief, about θ when we observe some IID value(s). Thus, we want to derive the *posterior* distribution:

$$p(\theta | \Delta L_{f_1}, \Delta L_{f_2}). \quad (1)$$

If the observed IID values ΔL_{f_1} and ΔL_{f_2} are noise free, then we should be able to look up the IID curves and find θ (although θ is not necessarily unique). A more realistic condition, however, is that

the observed IID values are noisy. Mathematically this can be modelled as follows:

$$\Delta L_{f_1}^* = \Delta L_{f_1} + \epsilon_{f_1}, \quad (2)$$

$$\Delta L_{f_2}^* = \Delta L_{f_2} + \epsilon_{f_2}, \quad (3)$$

where $\Delta L_{f_1}^*$ and $\Delta L_{f_2}^*$ are the noisy IID observations, and we assume $\epsilon_{f_1}, \epsilon_{f_2} \sim \mathcal{N}(0, \sigma^2)$. We need to assume (or estimate) the noise variance σ^2 ; in this question we just assume $\sigma = 2$ [dB]. Now the posterior distribution is

$$p(\theta | \Delta L_{f_1}^*, \Delta L_{f_2}^*) \quad (4)$$

(the observation is noisy; cf. eq. 1).

It is very difficult to derive the formula of eq. 4 directly. Instead, we can employ the *Bayes rule* to derive the formula. As we reviewed in the class, there are two terms required to apply the Bayes rule. One is the *likelihood* $p(\Delta L_{f_1}^*, \Delta L_{f_2}^* | \theta)$, the other is the *prior* $p(\theta)$. We can reasonably assume that the prior is a uniform distribution. In this question, let's restrict $\theta \in [0, 180]$, i.e., the sound is coming only from your left side. Regarding the likelihood, it can be expressed as

$$p(\Delta L_{f_1}^*, \Delta L_{f_2}^* | \theta) = p(\Delta L_{f_1}^* | \theta) p(\Delta L_{f_2}^* | \theta) \quad (5)$$

(you are asked to derive this). Since ΔL_{f_1} and ΔL_{f_2} are the functions of θ , eqs. 2-3 define the probability model of your observation,

$$p(\Delta L_{f_1}^* | \theta) \sim \mathcal{N}(\Delta L_{f_1}, \sigma^2), \quad (6)$$

$$p(\Delta L_{f_2}^* | \theta) \sim \mathcal{N}(\Delta L_{f_2}, \sigma^2). \quad (7)$$

(These are called *generative models* because they define how the observations are generated probabilistically.) Now we everything we need to derive 4 and are ready to evaluate it.

1. (5 points) Give a justification for the assumption made in eqn. 5. How could this be violated?
2. (5 points) Derive and plot the density of θ before observing the IID.¹
3. (5 points) Plot the posterior after observing $\Delta L_{f_1}^* = 9$. Here you don't need to normalize the posterior so that it is a proper probability density.
4. (5 points) Similarly, plot the posterior when you only observe $\Delta L_{f_2}^* = 15$.
5. (10 points) Derive and plot the posterior when you observe $\Delta L_{f_1}^* = 9$ and $\Delta L_{f_2}^* = 15$.
6. (5 points) Based on the plots above, describe how your certainty on θ changed as you observed the IID values.

¹For questions 3.2 and 3.5 you should write out the form of the posterior in addition to plotting it.

4 Inferring the interaural time delay

(20 points + up to 30 points extra credit.) In the last homework (problem 3), we calculated the interaural time delay (ITD) from a simulated stereo sound (and then calculate the sound direction). This time, let's examine this problem in a probabilistic framework. We want to estimate the time delay Δt when we observe a stereo sound $y(t) = [y_l(t), y_r(t)]$. Previously we assume and created a stereo sound from a monaural sound $x(t)$ by

$$y_l(t) = x(t), \quad (8)$$

$$y_r(t) = x(t - \Delta t). \quad (9)$$

- (5 points) Express this problem as a probabilistic model assuming a fixed noise level.
- (5 points) What is the posterior distribution that we should examine?
- (10 points) To derive the posterior, we need to define the likelihood and the prior. Define both and explain the reasoning behind your choices.
- (10 points) If the observed signals have high SNR, your estimate of Δt should be the more accurate, i.e. it should have a sharper posterior distribution. This means that louder sounds provide better cues for sound localization given a fixed background noise level. Set up two conditions of high and low SNR using the sound file `note.wav` and plot the posterior distributions in each case (again, it's not necessary to normalize). Discuss the difference in the estimation accuracy.
- (10 points) If you observe the longer sound, your estimation of Δt should also be more accurate. Again, set up (at least two) corresponding conditions using the sound file, plot the posteriors, and discuss the differences in the estimation accuracy.
- (10 points) As we discussed in the class, sound lateralization accuracy increases for louder test sounds (Fig. 2 left) and for sounds longer in duration (Fig. 2 right). Discuss to what extent your analysis could explain these measurements.

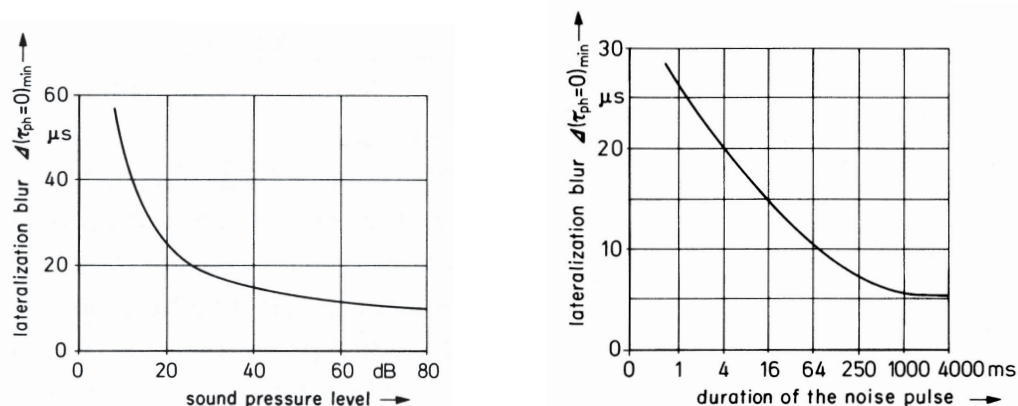


Figure 2: More signal improves localization accuracy