## Echolocation

Suppose that you wished to judge the 3D position of objects around us by clapping your hands and listening for the echo. The time between hand clap and echo in principle can tell you how far you are from object. This is the method of *sonar* (sound navigation and ranging). Note that the *distance* to the object is determined by time delay between the hand clap and the arrival of the echo, namely the distance is $\frac{1}{2}v\ \tau$ where $v$ is the speed of sound and $\tau$ is the delay. The reason for the factor $\frac{1}{2}$ is that the sound has to go to the object and back again. Also note that the detailed structure of the echo could tell you the direction, for example the timing and level differences in your two years.

One issue that arises with this simple method is that the reflected sound (the echo) is much weaker than the original sound. The hand clap sends off a spherical wave of sound in all directions. For any small cone of directions, the energy remains in that cone and travels radially outward from the source. (There is also a *loss* of energy due to friction/attenuation in the air, which is greater for high frequencies). As discussed a few lectures ago, the area of each wavefront of the expanding sphere is $4\pi r^2$, and the sound energy per unit area of the sphere must fall off as $\frac{1}{r^2}$, and the sound pressure level (SPL) falls off as $\frac{1}{r}$. If the sound of the hand clap reflects off a small object in the scene, the reflection that arrives back at the source will be weak. Take a small flat surface of area $A$ which faces the source. This surface receives about $\frac{A}{4\pi r^2}$ of the energy of the hand clap. Some fraction of this energy is reflected back, though the reflection occurs in many directions: the small flat surface acts as a small sound source and the wave it reflects radiates back as a sphere. By the time the reflected wave reaches the original source (the ears of the person that clapped hands), the reflected echo energy is proportional to $\frac{A}{4\pi r^2}\frac{1}{r^2}$ which is proportional to $\frac{1}{r^4}$. Thus we see that although timing of the echo carries information about distance, it is not obvious that this echo can be reliably measured since it may be too weak.

Let's now examine echolocation in bats, and how they deal with this problem.

## Bat sonar

Bats are among the strangest of the mammals. They have horrific pointed ears and large flaring nostrils. But perhaps what makes bats most frightening and alien is that they can navigate in the dark. Humans fear the dark. Bats thrive in it.

How bats manage to live in darkness was a mystery for centuries. Bats do have eyes so obviously they can see. It was long believed that bats had much more sensitive visual systems than humans and other mammals, and that bats can see in the dark because they only need very little light and they can adapt to lower light levels than people can. Others believed that bats sensed the location of objects by "feeling" the reflection of pressure waves which are produced by the bat's beating wings. Both of these commonly held beliefs turned out to be completely wrong.

Over 200 years ago, an Italian named Spallanzani carried out experiments that showed bats use hearing to navigate in the dark. He captured a set of bats and blinded them, and verified they could navigate fine – avoiding very small obstacles and even continuing to feed on flying insects. This ruled out the hypersensitive vision hypothesis. Ruling out the wing beat theory was more difficult. Instead, he proposed another hypothesis, namely that bats rely on hearing to navigate. To test this hypothesis, he inserted small hollow cones into a blinded bat's ears. The bat could still navigate just fine. He then filled the cones with wax, blocking out most of the sound. With the

cones blocked, the bats could not navigate at all. They crashed into walls, objects, and were unable to feed on flying insects.

Spallanzani's experiments were carefully done but were largely rejected by the scientific community. The reason is that he had not explained *how* bats use hearing to navigate and locate insects. It wasn't until the late 1930's that this puzzle was solved. The key insight came from experiments done by Donald Griffin. Griffin was the first to be able to measure the cries of bats. This had been difficult to do previously because bat cries are at frequencies much higher than what people can hear (i.e. ultrasonic). There were few devices before Griffin's time for recording ultrasonic frequencies and those that did exist were expensive and so were not used to record bat sounds.

Griffin was the first to record the sounds made by a bat as they navigate (in a big cage in his lab) and hunt for food. If you compute a spectrogram of these sounds, you often find a pattern such as shown in the figure below. The time and frequency scales are not shown, but here are some rough numbers. Bat cries are from 12 kHz up to 200 kHz. Most of this range is non-audible for humans. I will explain what CF and FM mean below.
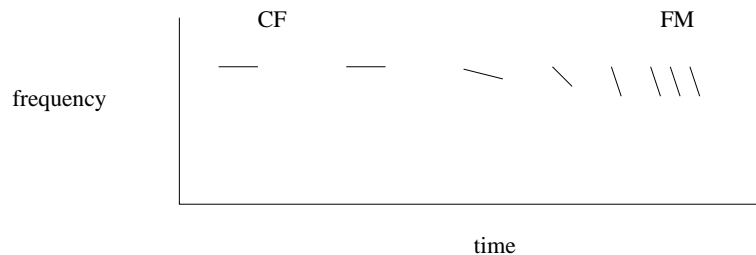


If we take 34 kHz, for example, this is 34 cycles in 1 ms which (according to $\omega\lambda = v$) corresponds to a wavelength of about 1 cm. Similarly, a high frequency such as 170 kHz corresponds to a wavelength of 2 mm. These wavelengths are behaviorally relevant for the bat. They are the size of objects such as tree branches, and edible insects such as moths which the bats eat.

There is also a large range in the durations of bat cries. A single cry can be as long as 200 ms and as short as a fraction of 1 ms. In the spectrogram above, the horizontal lines on the left have a larger time duration than the highly slanted lines on the right.

There are over 700 species of bat. Different species live in different habitats, have different shaped ears, eat different foods, and have evolved different mechanisms for using echolocation. Some bats cry through their noses, others through their mouths. Despite the variations, there are general echolocation principles that have been discovered which seem to explain much of bat echolocation behavior.

Consider again the spectrogram above. In order to explain such a spectrogram, we need to consider what is involved in hunting for food – what problems need to be solved? Suppose you are a bat flying through the air and you are hungry for flying insects. You need to solve three problems, in the following order:

1. *Detect:* is there something out there?

2. *Localize:* where is it ? (distance and direction)

3. *Recognize*: what is it? (shape, motion, material)

Roughly speaking, there are two kinds of bat cries that are used to solve these problems: constant frequency (CF) and frequency modulation (FM). These correspond to the horizontal and highly slanted lines in the spectrogram above, respectively.

## Constant frequency (CF)

The frequency composition of a CF cry does not change throughout the cry. CF cries are composed of a very small range of frequencies. In order to achieve this property, the cry must to have a long duration. Constant frequency signals are typically over 10 ms, and are often well over 100 ms. To understand why a CF cry must have a long duration, recall the properties of a time Gabor functions: if the Gabor has a small bandwidth $\Delta\omega$ i.e. a Gaussian with a small standard deviation, then the Gabor must have a large Gaussian window in time.[1] To make a CF cry, presumably the bat must use a voiced sound, but the glottal pulses can only be so fast. The high frequency presumably arises from the articulation which amplifies a small set of frequencies and attenuates most of the others.

What are the advantages and disadvantages of CF cries, with respect to solving the three problems mentioned above? First consider detection. Bats are mammals and their auditory system is like our own in that it encodes the sound using bandpass filters of varying bandwidth ("critical bands"). So, CF is good for detection because it puts alot of sound energy within a single critical band, and so the echos of a CF cry also lie within one critical band. The concentration of energy in a single critical band makes it easier to detect this signal in the presence of other sounds in the environment and noise in the auditory system in that band.

One important difference between the bat's critical bands and our own is that the bat's critical bands contain an *acoustic fovea*. Recall from our discussion of vision that the human retina packs a high percentage of photoreceptors into one small area, i.e. the direction in which we are looking. The bat's acoustic fovea devotes more cells to a particular range of positions on the basilar membrane. (An example is shown in the lecture slides.) This allows the bat very good frequency discrimination at these frequencies, as well as the ability to detect relatively quiet sounds at these frequencies. These are the frequencies near which the bat makes its CF cries. This specialization is helpful because the environment may contain sound energy at many frequencies, and the bat wishes to only hear the echos of sounds that it generates.

How does the bat hear the echo of its CF cries, in the presence of the cries themselves? There are two answers to this. First, the bat leaves a gap of silence between its CF cries, which allows time for the cry to propogate through space, reflect, and then return to the bat. By the time the echoed cry returns, the cry should be over. The longer the gap, the less of a forward masking[2] occurs.

Second, note that if the bat is flying forward, then the emitted cry will undergo a *Doppler shift*. The bat will be chasing the sound as it emits the sound, which will lead to an increase in the frequency of sounds that are received at the reflecting surface. These higher frequency sounds are then reflected back and the bar flies toward these sounds which results in another Doppler shift.

Suppose the bat were flying forward with velocity $v_{bat}$ and emitting sound at some frequency

---

[1]Careful: The converse is not true. I am not saying that a long duration sound always has a small bandwidth.

[2]Masking of sound B by sound A can occur even if the sounds are not played simultaneously. Foreward masking means that the mask occurs before the test. Backwards masking means that the mask occurs after the test.

$\omega_{emit}$. Once can show that the frequency of the echo observed by a forward flying bat is:

$$\omega_{observed} = \omega_{emit} \frac{v_{sound} + v_{bat}}{v_{sound} - v_{bat}}.$$

So, for example, if the bat's speed is say $\frac{1}{100}$ of the speed of sound, then we get an approximately 2% increase in the frequency from the Doppler shift. e.g. if the bat emits a cry at 100 kHz, then the shift can be about 2 kHz. This is not a lot but it is enough to put the echo into a different critical band. The idea here is that if the bat emits a sound just below the frequency of the acoustic fovea which has a very small bandwidth, then the reflected sound will fall in the fovea and masking is avoided.

Once the bat has detected an object and has a rough estimate of the depth (based on delay between cry and echo), it needs to estimate the location. For this it can use the monocular and binocular cues we have discussed in the previous lecture, namely frequency based cues arising from the HRTF, level and timing differences.

Are these cues useful for CF cries? The level differences are of little use since the bat directs most of the sound generally forward and there is presumably not much level difference between the two ears near the forward direction. The timing differences are also not very useful since the CF cries are high frequency (recall the duplex theory).

One might wonder if the *envelope* of the CF cry could be used for a timing differences. Probably not. The CF cry has a gradual envelope and so it doesn't have a well defined starting and ending point. (To understand this, think of the smooth Gaussian envelope of a Gabor function. Where does a smooth signal start?)

In order to get well defined timing signals, the bat instead needs to modulate the frequency over time using FM, as I will describe below. Before we do so, let's consider one more aspect of CF signals, namely how could be used for *recognition*. Suppose the bat would like to decide whether to pursue a flying insect, based on what kind of insect it is. Bats have tastes, just like we do. How can the bat do so? Suppose a particular species of moth beats its wings at a rate of about 40 beats per second, and so the wingbeat period is around 25 ms. When the wing is perpendicular to the direction of the bat's cry, the echo back toward the bat is maximal, and when the wing is parallel to the direction of the cry, the echo is minimal. Thus for long CF cries (100 ms), the echo contains a periodic structure – on, off, on, off, etc. Different moths have different wing beat rate, and this cue can be used for recognition!

## Frequency modulated (FM)

Frequency modulated cries are roughly of the form $\sin(\omega t)$ where $\omega$ is itself a function of $t$, for example, $\omega = \omega_0 - \beta t$. (A slightly different form of this equation was given in the lecture slides.) The idea is that the frequency near the beginning of the cry is different than near the end of the cry. The bigger is $\beta$, the faster the frequency drops over time and the steeper the slope in a spectrogram representation. Such a signal is also called a *chirp*.

What are the advantages and disadvantages of FM cries? One disadvantage is that FM cries are poor for detecting objects at a distance. The cry (and hence the echo) sweeps through each critical band for a short time only. The bat has to use the signal within each critical band to detect the echo amid environmental noise, so if the bat is still far from the reflecting object then the echo will have little energy in each band. Think of an analogy in vision: consider a 2D sine grating with

noise. If the 2D sine grating covers the whole display of your monitor then it will be easier to detect in the noise than if the sine component (of the same frequency) just covers a small window say 50 x 50 pixels in the the display.

The advantage of FM comes when the bat is close to its target. FM cries are loud enough in each channel to be heard. Moreover, since the duration of the bat's cry and the echo within each critical band is short, the timing difference between the cry and echo can be computed more precisely than what we had with the CF signal where the duration of the cry was large. In this sense, FM cries and their echos carry accurate information about distance. Note that FM cries do not need to be separated by long silent gaps, as did CF cries. Provided the temporal duration within each frequency channel is short enough that the echo in that frequency channel doesn't return before the component of cry in that channel is complete, then there is no overlap in that channel. This is a big advantage with FM.

The timing differences just mentioned were between the cry and the echo and gave information about the distance. There also will be timing differences for arrival of the echo at the two ears, provided that the target is not straight ahead (in the medial plane). Note that cries are high frequencies, so the bat brain is not matching individual spikes. Rather it is measuring time differences between envelopes. Also, lots of frequency channels are activated by the FM sweep and so the auditory system can combine signals across channels to get timing differences, as humans do. The same argument can be made for level differences. So, with FM signals bats can use binaural hearing cues just as humans do.

What about recognition? Earlier I mentioned that the wing beats of a moth can be used as a cue for recognition for CF cries. For FM cries, the moth wingbeat information is useless because the duration of the cry is much shorter than the period of a wingbeats. Can FM cries be used for recognition? Yes they can! When an FM cry echos off an insect such as a moth, there is rarely a single echo, but rather multiple echos[3]: the pulse bounces off a wing and also off the body – or off the head and the wing – and these two reflecting surfaces may be at slightly different depths. To keep the analysis simple and just get the basic idea, let's look at the overlap of two echos, and we'll use a toy model where the moth impulse response function is

$$m(t) = a\delta(t) + b\delta(t - \tau).$$

where $a > 0, b > 0$. So there are two echos and the second is shifted in time by $\tau$ relative to the first. Letting $\omega$ have units cycles per second[4] and let $\tau$ be in units of seconds, we have

$$\mathbf{F}\{m(t)\} = a + b\,e^{-i2\pi\omega\tau}.$$

Observe that *constructive interference* occurs when $\omega = \frac{1}{\tau}, \frac{2}{\tau}, \frac{3}{\tau}, \ldots$ etc and *destructive interference* occurs when $2\omega\tau$ is an odd integer, so $\omega = \frac{1}{2\tau}, \frac{3}{2\tau}, \frac{5}{2\tau}, etc.$ See the Exercises for an example.

---

[3]called *glints,* like the specular reflections off water waves

[4]If you prefer, we could use the familiar units of samples $t$ and cycles $\omega$ per $T$ samples, and in that case we would write

$$\mathbf{F}\{m(t)\} = a + b\,e^{-i\frac{2\pi}{T}\omega\tau}.$$

The main idea here is that two echos separated by a small distance can produce a systematic interference pattern. If an FM cry had a roughly constant level (SPL) over a range of frequencies that it sweeps through, then the echo would *not* have constant level over frequency. In general, objects such as moths or flying beetles will have more complex echos than the simple toy model above. But whatever the echo pattern is, it will be a signature of the shape/orientation of the moth. Think of the object that is reflecting the sounds as having a transfer function, similar to what the HRTF does to a sound arriving at the head. By detecting which bands receive sound and which do not, it is possible to infer the something about the shape of the reflecting surface. Amazingly, behavioral experiments with bats have shown that bats can indeed discriminate between various spectral patterns in echos.

To briefly summarize, the bat uses the CF signals to detect and recognize the object producing the echo. If there is an object and it is worth pursuing (wing beat frequency corresponds to edible moth species), then it gradually switches to an FM cry. The FM cry is a shorter duration signal and the length of the signal within each band is much shorter. This provides better timing information which allows accurate distance estimation, as well as direction estimation: more channels are active which allows binaural and monaural spectral cues to be used for localization. It also allows spectral cues to be used for recognition, since the object's size and shape determine the constructive and destructive interference in the reflected echos.

## How dolphins and porpoises use echorecognition

Dolphins and porpoises and other marine mammals also use echolocation and echorecognition. These animals are very sociable with people and so they can be trained to perform many behavioral tasks. A key difference between dolphin and bat sounds is that porpoises do not use FM cries, but rather they use "clicks". They are called clicks because that is what they sound like to a human listener. The center frequency is often in the 60-150 kHz range and there are 2-3 cycles within the envelope's half height which corresponds to about half an octave. For example, for a 120 kHz center frequency, you typically have about a 60 kHz bandwidth. (We can't hear these clicks, but porpoises also emit clicks at much lower frequencies that do fall within the human range.[5])

Dolphins can echolocate objects and also use echos to recognize them. They can distinguish the shapes and materials of the objects that produce the echos. This allows them to distinguish different types of fish, for example, some of which are easier to catch or to digest.

Let's just look at one aspect of this, namely the constructive and destructive interference patterns in the echo. The idea is similar to what we saw with the bat, but the click is different so new issues arise. Suppose a target fish is aligned so its body axis is perpendicular to the line between the dolphin and the fish. When dolphin click is reflected off a fish, the sound reflects off both the front surface as well as the back surface – the sound passes through the fish and back out. (There are multiple reflections within the fish, but let's keep things simple and just consider two echos, as we did above i.e. two echos using a formula such as written above.) The reason that the sound passes through the fish is that the fish is made mostly of water. (Discussion of 'impedence' omitted.)

The speed of sound in water is about 1500 m/s which is about four times faster than in air. Suppose we take center frequency of 75 kHz. Verify using the $v = \omega\lambda$ formula that the wavelength is $\lambda = 2cm$. Consider possible thicknesses of the fish and whether these produce constructive or

---

[5]https://www.dolphins-for-kids.com/dolphin-clicks-and-whistles

destructive interference for a given wavelength. Note that the echo that passes through the fish must pass through it twice (go and come back). Thus, if the fish thickness is 1/4 wavelength of the sound of some frequency, then twice the fish thickness (aller-retour) will be 1/2 a wavelength which will give destructive interference, whereas if the fish thickness is 1/2 wavelength then we get constructive, etc

fish thickness

| | | |
|---|---|---|
| .25 | $*\lambda$ | destructive |
| .5 | $*\lambda$ | constructive |
| .75 | $*\lambda$ | destructive |
| 1 | $*\lambda$ | constructive |
| 1.25 | $*\lambda$ | destructive |
| 1.5 | $*\lambda$ | constructive |
| 1.75 | $*\lambda$ | destructive |
| etc | | |

Keep in mind that the above is just a toy model, meant to illustrate some basic properties of echose. Hopefully this is enough for you to appreciate how the reflected sound might be quite different than the emitted click.

Like all hearing mammals, the dolphin processes the reflected sound by using bandpass auditory filters. The frequencies are far too great for the details of the sound wave to be captured by the timing of spikes. Rather, auditory nerve cells that encode the echos hear will respond to those frequencies that are present and will compare those that are present with those that are absent. As with bats, this can be used not just to locate objects but also to recognize them. [6]

---

[6]https://books.google.ca/books/about/The_Sonar_of_Dolphins.html?id=Q3MIsrPDA5EC&redir_esc=y