[SQUEAKING]

[RUSTLING]

[CLICKING]

JOSH MCDERMOTT: There was a question at the end of-- or in the middle of last lecture about the relationship of loudness to frequency, and I completely forgot to mention something that's worth knowing about, which is this thing that's shown here, which are known as equal loudness contours. And it's also sometimes called the Fletcher-Munson curves. And what this graph is showing is what the intensity of different frequencies has to be for them to sound equally loud.

So the bottom graph is our detection thresholds. That's the audiogram. That's the thing that we've seen before. And that's the thing that kind of looks like more or less like a U shape. So your detection threshold is higher for low frequencies, and then rises again at higher frequencies, and it's best here in the mids.

And then what each of these curves is showing is for higher intensities, it's like what the different frequencies would need to be set at. So everything that would fall on one of these red curves supposedly sounds equally loud to a person with normal hearing. And so you can see that as the overall intensity increases, you retain traces of the audiogram, which is to say lower frequencies and to some extent higher frequencies, they have to be higher in intensity to sound equally loud to a middle frequency. But that difference gets attenuated.

So the U is kind of deeper down here when the level is lower compared to when the level is higher. And so what this means in practice, it actually has a lot of relevance for mixing music. So what it means is that a recording is not going to sound the same when you play it at a low level compared to when you played at a high level.

And so in general, at a very low level, the mids are going to be dominant, because you can see that in order to make something equally loud, if it's a low frequency or a high frequency, you've got to increase the level of fair bit. So the mid frequencies will be fairly dominant at low levels, and then the highs and the lows become more prominent as the overall level of the recording increases. So equal loudness contours, and there are these interactions with intensity and with frequency. OK. Yeah?

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: I did. Yeah. So I just wanted to mention that. And the-- yeah, I think that's it. So let's start talking about auditory scene analysis.

So we got started on this last time. We talked about how the problem is that there are very often multiple acoustic events that are happening in the world at the same time. Your ear receives the single sound waveform, and you typically will need to estimate or comprehend something about one or more of the individual sources.

Classic example being the cocktail party problem, where you're trying to understand one person and there are other voices that are present at the same time. The problem is kind of interesting, like many perceptual problems, because it's ill posed. So you receive this one variable, which is the-- one signal, which is the waveform corresponding to the mixture of a bunch of signals. That one signal is generated by distinct source signals in the world. So it's akin to me giving you that equation and asking you to solve for x.

And we talked about how the only way that you're able to solve these ill-posed problems is by leveraging the fact that the sources that occur in the world are not random. And in fact, they're very, very far from random. And so we believe that what has happened is that over evolution and development, your brain has internalized the statistical regularities of the world. And that's what enables you to solve these ill-posed problems.

And then we talked a little bit about the idea of perception as unconscious inference, and how we can formalize that with Bayes' rule, where the task of perception can be considered that of finding the hypothesis about the world that is most probable given some observed signal O, and that can be decomposed into the prior and the likelihood. We talked a little bit about levels of analysis, and then discussed how the only way that you can solve these ill-posed problems is by having priors. The priors kind of capture your knowledge of regularities, in this case, of real world sounds, and how illusions can reveal the aspects of these priors. They reveal the constraints on sound generation internalized by the brain that make it possible to solve the problem.

And so that's where we left off. And so what we're going to do today is go through a bunch of examples of these classic illusions that illustrate principles of auditory scene analysis. So properties of sound that seem to provide constraints on how we solve the problem. And I'll note that many of these could actually be useful jumping-off points for your illusion lab. So just be thinking about these as we go through. And if you're wondering about anything, write that down as something that maybe you could investigate for the illusion lab.

So oftentimes these phenomena are discussed in the context of what are called grouping cues. So we've talked about cues as being properties of a stimulus that are informative about some aspect of the world. And so when we're talking about sound localization, we talked about localization cues. Now we're talking about auditory scene analysis, which you can think of as a problem of grouping.

So you get this sound signal here. It's got all these little bits of acoustic energy scattered across frequency and across time. We often will represent it with a picture like this, because we think that this picture kind of captures what your ear is sending to your brain. The ear, you can think of that as doing a frequency decomposition. And then there are responses at different frequencies over time.

So one way to think of the auditory scene analysis problem is as a problem of grouping the bits of energy into sources that would have a common cause in the world. And so we'll talk about grouping cues. And what is typically meant by that is a stimulus property that is informative about the organization of sound into sources. And so some standard examples would be common onset and offset, which you can see here, and we'll see some examples in a second. And then harmonicity, the fact that some sounds in the world tend to have frequencies that are harmonically related.

So this is an initial kind of classic demonstration that the brain seems to interpret abrupt amplitude changes as the onset or offset of a distinct source in the world. By contrast, gradual amplitude changes are instead interpreted as a single source that changes over time. So this is a super simple demonstration.

So what's going to happen here is we're going to have a noise signal. And then the amplitude of the noise signal will change in one of two ways. In one case, it will change gradually. So this is a graph that shows the amplitude as a function of time.

So in one case, the amplitude will start at a value, and then it'll ramp up, and then it'll ramp down. And in another case, it'll kind of shoot up very abruptly. But in both cases, essentially what happened is there's a noise signal and a little volume knob, and that volume knob is just getting manipulated in one of two ways.

But what you will see is that what you hear when you listen to this is actually quite different in the two cases. In this case, it'll really just sound like there's kind of a single thing that's kind of waxing and waning in amplitude. You might even hear it as kind of coming closer to you or further from you. Whereas in this case, you'll hear a new sound source when you get this abrupt increase in amplitude. Let's check it out.

[AUDIO PLAYBACK]

- Homophonic continuity and rise time. A noise repeatedly rises and falls in intensity. If the changes occur slowly, you hear a single sound waxing and waning in intensity. If the changes are abrupt, you hear two sounds, a soft one continuing unchanged, and an additional noise burst added to it periodically. First we hear the gradual changes, then the abrupt ones.

[PULSING WHITE NOISE]

[CLICKING WHITE NOISE]

[END PLAYBACK]

JOSH MCDERMOTT: All right. So what you probably heard in the second case is that it sounds like there's a low-intensity noise that's kind of just on continuously, and then these pulses of louder noise. Now in actuality, there's just one noise signal where the amplitude gets modulated, whereas in this case, it kind of sounds like one thing that's changing in intensity.

So these very sudden changes in intensity are interpreted by your brain as the presence of a distinct source. And this can also happen kind of at the level of individual frequency channels. So sudden onsets of new frequencies are typically interpreted as new separate events, whereas frequencies that were present before are heard as continuing. And this is often referred to as the old-plus-new heuristic.

So this is a diagram that shows how this works. So you're going to compare two stimuli. So previous graph showed amplitude versus time, whereas these are schematic spectrograms. So now we have frequency, and the gray level is kind of supposed to indicate the intensity at that frequency. And the x-axis is time.

So in this case, there's a stimulus that is constructed by concatenating low-pass noise-- so it's low-pass because it contains low frequencies-- with a broadband noise. That's the B. So low-pass noise, broadband noise, low-pass noise. Whereas this one, you have high-pass noise alternated with broadband noise.

So the idea here is that the broadband noise in the two stimuli is the same. But the perceptual organization that you will hear in the stimulus is very different. So in this case, even though this was constructed by alternating low-pass and broadband noise, you're going to interpret the stimulus as a single low-pass noise that's continuous with these intermittent high-frequency pulses of noise. Whereas in this case, you'll hear a continuous high-pass noise with these intermittent low-frequency bursts of noise

So same broadband noise will get taken apart by your auditory system, or explained by your auditory system, in two different ways I'm going to let you see the spectrogram while this plays. And let's listen to it.

[AUDIO PLAYBACK]

- Demonstration 34, capturing a low band of noise from a wider band, leaving a high band as a residual. Please refer to the figure in the booklet. A short band of noise, B, is alternated with a longer band that contains only B's lower frequencies. These frequencies group with the low ones of B to form one continuous low noise, and the high components of B are heard as an intermittent high sound.

This is the residual formed when the lower frequencies of B are captured. Similarly, the high frequencies of B can be captured by a high noise, leaving a low noise as the residual. These two cases are presented twice in alternation. Remember that B is physically identical in the two cases.

[PULSING WHITE NOISE]

[CLICKING WHITE NOISE]

[PULSING WHITE NOISE]

[CLICKING WHITE NOISE]

[END PLAYBACK]

JOSH MCDERMOTT: OK. Everybody get that? All right. So what it sounds like subjectively is completely different in the two cases. And at the moment when those intermittent noises play, the stimulus is identical. And that's because your brain is interpreting that as being two different combinations of sources.

So there's kind of a related phenomena where-- so in those examples, we were just looking at these cases where you would really describe the stimulus in terms of onsets. It's like these frequencies, or the sound, kind of abruptly starts. But it's also often pretty common for the modulation to be more gradual. And we call that modulation, or amplitude modulation.

And so in particular, there's a pretty well-known effect that seems to indicate that common modulation across frequencies tends to promote grouping. And this is actually-- I like featuring this particular example, because it kind of builds on the masking experiments that we talked about a little bit earlier in the course. So in this particular phenomenon, your task is to detect a pure tone in noise, like one of those classical masking experiments. And we're going to measure your threshold for detecting the tone in the noise.

And we will, like in the classic band widening experiment of Fletcher, we're going to manipulate the bandwidth of the noise. But the catch here is that unlike the classical experiments, in this experiment, the noise can either be fully random, as in the classical experiments, or comodulated. That means multiplied by a common envelope.

So that's what is schematized here. The idea is that we've got all these different bands of noise, and they can be comodulated. And the central phenomena that we will witness is that this kind of strange thing happens, where this is supposed to be the case, where there's just a fairly narrow band of noise around the tone. You could, in principle, adjust the amplitudes of the tone and the noise so that you can't detect the tone. That's why the person is sad. It means they can't hear it.

And the phenomenon here that we will witness is that we can add these additional bands of noise at adjacent frequencies. And if they're comodulated, that will actually lower your threshold. That's why the person's happy. They can now hear the tone.

So when the noise is comodulated, the thresholds are better when noise that is far from the tone is added. So here's the result of the experiment, and then we'll listen to a demo, and you can see whether you believe it. So this is same kind of graph that we saw earlier.

So we got a threshold on the y-axis, and the masker bandwidth on the x-axis. So the graph that has the R symbol, that stands for random. And so that's the kind of classical band-widening result. What does it show?

It shows that as we increase the bandwidth of the masker, the thresholds kind go up. And then you hit the critical band, and they kind of level off. That's the thing that we looked at a few lectures ago.

So the new thing here is what happens when the noise is comodulated. And that's the M symbols. And so what happens there is actually different and pretty interesting, which is initially, the threshold goes up, but then it goes back down again. And so you end up with a pretty big difference between the thresholds that you get when the noise is fully random versus when it's modulated.

Let's see if we can hear this. So this is what happens with purely random noise. So what you're going to hear in this demo is there's a noise signal and then a sequence of tones. And I think in this demo, the tones increase in level. And so you're supposed to notice when you can hear them. And it's possible that you won't be able to hear any of them in this particular case, but let's see.

[AUDIO PLAYBACK]

[PULSING WHITE NOISE]

[END PLAYBACK]

JOSH MCDERMOTT: I heard the last two. Try it again.

[AUDIO PLAYBACK]

[PULSING WHITE NOISE]

[END PLAYBACK]

JOSH MCDERMOTT: Can you hear a tone there? Not sure? All right. That's OK. Well, let's see if you can hear it here.

[AUDIO PLAYBACK]

[LOWER FREQUENCY PULSING]

[TONE BEEPING]

[END PLAYBACK]

JOSH MCDERMOTT: Did you hear a tone at the end? OK. So that's this effect. So when that noise is comodulated, it becomes easier to detect the tone, and your threshold is lower. In this case, by about 10 dB, so it's a pretty big effect. So this is called comodulation masking release. It's a super well-known effect in the world of psychoacoustics.

And it is considered to be evidence for grouping via comodulation. So the idea being that the fact that the noise has the same envelope at all these different frequency channels helps you to separate the tone from the noise because it's helping you distinguish which parts of those signals are tone and which parts are noise, which in turn makes it easier for you to tell whether the tone is there.

Another very basic property of natural sounds that we have already talked about at some length is harmonicity. So sounds in the world are often harmonic. And the brain seems to assume that frequencies that are harmonically related belong to the same acoustic event.

And one piece of evidence for this is that when one frequency out of a bunch is not harmonically related to a bunch of others, it tends to segregate perceptually. And this is often known as harmonic mistuning. And so we'll hear a demonstration where you can listen for this for yourself.

And so what's going to happen here-- so, again, these are schematic spectrograms. So this is a complex tone that has harmonic frequencies. And all of the harmonics except the third one here are going to shift down over successive cycles of the stimulus.

And so this third harmonic is going to stop being harmonically related to all the others. And so at the beginning, this is going to sound like one tone, like one thing. But over time, what's going to happen is you'll actually start to hear two tones. So it's all just a bunch of frequencies, and they all start and stop at the same time, but the deviation from these harmonic frequency relations is going to cause you to hear two things. Let's check it out.

[AUDIO PLAYBACK]

- Demonstration 18, isolation--

[END PLAYBACK]

JOSH MCDERMOTT: In fact, I what I'm going to do, I'm going to play the intro and let you look at the picture. It seems better.

[AUDIO PLAYBACK]

- Demonstration 18, isolation of a frequency component based on mistuning. You are to listen for the third harmonic of a complex tone. First, this component is played alone as a standard. Then over a series of repetitions, it remains at a constant frequency, while the rest of the components are gradually lowered as a group in steps of 1%.

[HIGH-FREQUENCY BEEP]

JOSH MCDERMOTT: All right. Now, let's see if I can--

[MID-FREQUENCY BEEP]

Sorry, I messed it up.

[HIGH-FREQUENCY BEEP]

My technique here is questionable. Let's see.

- The group in steps of 1%

[HIGH-FREQUENCY BEEPING]

[MID-FREQUENCY BEEPING]

[DUAL TONES BEEPING]

JOSH MCDERMOTT: Raise your hand if you're hearing two tones at this point. Yeah. So almost everybody. So this works.

[DUAL TONES BEEPING LOWER IN FREQUENCY]

- Now, after two more presentations of the target component alone, the other partials of the tone are raised in steps of 1% until the target component loses its identity and once more forms part of the complex tone.

JOSH MCDERMOTT: So what this is going to show-- so now you know that there's this separate thing there. That third harmonic is going to sound like it's a second tone. But as it approaches being harmonically related to everything else, it's going to fuse back into them. And even though you know it's there, you're going to stop hearing it as a separate thing.

[HIGH FREQUENCY BEEPING]

[DUAL TONES BEEPING]

[HIGHER FREQUENCY DUAL TONES]

[END PLAYBACK]

JOSH MCDERMOTT: So at the very end there, you probably lost it. Raise your hand if you lost it at the end. Yeah, most of you. OK.

So that's harmonic mistuning. One way to measure this is to ask people whether the harmonic is mistuned or not. And the idea is that the way that you actually perform that judgment is by hearing out whether it sounds like there's a separate tone. And so people are actually pretty sensitive to this, so you can mistune the frequency of that harmonic by 2% and measure thresholds for that.

So this is just four different people. That's the threshold. This is the number of the harmonic. None of this really matters. The point is like the numbers are small right, 2%. So people are very sensitive to this.

So here's something that is related to this that is really pretty cool. And it's natural to wonder like, well, what-- so you mistune the harmonic and you hear the second thing. So that seems significant. You've changed your interpretation in terms of what the sounds are in the world.

But it's natural to wonder like the extent to which that subjective sense that there are two things there, like whether that affects other things that you do with your auditory system. And so one of the things that we often do when we listen to a sound like that is we get a sense of the pitch. Maybe because you want to sing it back or accompany it in some way.

And so it turns out that when you mistune one harmonic of a complex tone just a little bit, that will very slightly change your sense of the pitch of the tone. And so the way that this is measured is in an experiment where you perform a matching task. So on every trial of this experiment, you're going to hear a stimulus like this. It's a complex tone where one of the harmonics will be slightly mistuned.

And then following that, there will be a match tone, and that's a purely harmonic tone. And your job is to adjust the fundamental frequency of the match tone to make it sound like it has exactly the same pitch as the initial tone. Yeah?

AUDIENCE: Does the graph change depending on which [INAUDIBLE]?

JOSH MCDERMOTT: Yes. But as long as it's what we call a resolved harmonic, it looks pretty similar. So in this case, it's the third harmonic. It would look the same if it was the second or the fourth, for instance.

And that's what's shown in this graph. So this graph is showing the mean shift in the pitch match in percent of the fundamental frequency as a function of the shift in that mistuned harmonic. So what does this show? Well, it shows that as you mistune the harmonic, there's a small shift in the perceived pitch of the complex tone. It's very small. This is like a fraction of a percent, but it's measurable.

However, something pretty interesting happens, which is that once the harmonic gets shifted too much, it stops changing the pitch of the tone that it's part of. And in fact, the pitch shift kind of drops back down. So the interpretation here is that when the shift of the harmonic is very small, you don't really hear it as a separate sound, and it gets incorporated into the calculation of the pitch, and so changes the pitch.

When the shift kind of gets too big, though, you start hearing it as a separate sound, and your auditory system kind of excludes it from the computation of the pitch of the tone. So this is evidence that the estimation of the properties of sound sources is kind of influenced by these grouping phenomena. It's as though the pitch is kind of calculated on your estimates of what the sources are that explain a given stimulus. What questions you got about this? Yeah.

AUDIENCE: So does this graph change as frequency goes up based on pitch perception in general?

JOSH MCDERMOTT: It doesn't change a whole lot. I mean, I've only seen it measured over a couple octave range. So you can change the fundamental frequency of the tone, it doesn't really change very much. As I was saying in response to the earlier question, it will be affected by which harmonic you do this to.

So if you do this to an unresolved harmonic, you actually don't get this effect. So the harmonic has to be resolved. It sort of makes sense, because remember how we talked about how pitch perception really seems to be mostly dominated by these low-numbered harmonics? So that really is a big factor. Any other questions? OK.

So this effect of grouping via harmonicity is not specific to these kind of funny synthetic tones. It also works for speech. So this is a spectrogram of a speech signal. And you can see that there are these kind of horizontal stripes. So these are the harmonics that exist in voiced speech. And again, speech is usually pretty fused. So here's how it sounds.

[AUDIO PLAYBACK]

- Academic aptitude guarantees your diploma.

[END PLAYBACK]

JOSH MCDERMOTT: Good news, right? Academic aptitude guarantees your diploma. That's just a speech signal. That speech signal is composed of harmonics. And we can actually-- using modern techniques, we can take the sound apart and isolate particular harmonics. And this is the third harmonic. This is what it sounds like on its own.

[AUDIO PLAYBACK]

[HIGH-FREQUENCY WHISTLING]

[END PLAYBACK]

JOSH MCDERMOTT: Can you hear that? Sounds like a whistle, kind of. So that's what the third harmonic sounds like on its own. And it's always there, but it gets grouped with the rest of the speech signal.

So what we did here is something kind of funny. You can see how the third harmonic is now kind of shifted up relative to where it should be in the harmonic series. So it's been mistuned by a fair bit.

And so now we're going to listen to what that sounds like. And you should be able to hear-- it'll sound like there's a speech signal, but then you'll hear this whistle that's kind of on top of it, which is the third harmonic, which you're now hearing out as a separate thing because it's not harmonically related to everything else.

[AUDIO PLAYBACK]

- Academic aptitude guarantees your diploma.

[END PLAYBACK]

JOSH MCDERMOTT: Now, lest you think this is all like me hypnotizing you, let's go back to the original, and now you won't be able to hear it, I think.

[AUDIO PLAYBACK]

- Academic aptitude guarantees your diploma.

[END PLAYBACK]

JOSH MCDERMOTT: So these harmonic frequency relations are really quite powerful and play a pretty big role in the grouping of sound. But they're not omnipotent. So onset differences can cause harmonics to segregate even when they're not mistuned. So this is a classic demonstration where we're going to play you complex tones that are harmonic, and individual harmonics are going to get gated on and off. And what that's going to do is it's going to cause you to hear out that harmonic that's gated as a separate thing.

And what's going to happen is iteratively, we're going to step through the harmonics. So here, you've got like the second harmonic being gated. Next, it's going to be the third. Next, it's going to be the fourth.

And as we step through that, you'll hear each of these harmonics individually as like separate things. And the one that was gated on the previous trial will then just blend back in. So I'll let you watch this one.

[AUDIO PLAYBACK]

[FLAT TONE WITH LOW-FREQUENCY BEEPING]

[FLAT TONE WITH MID-FREQUENCY BEEPING]

[FLAT TONE WITH HIGHER-FREQUENCY BEEPING]

[FLAT TONE WITH HIGHER-FREQUENCY BEEPING]

[FLAT TONE WITH HIGH-FREQUENCY BEEPING]

[END PLAYBACK]

JOSH MCDERMOTT: We could do this forever, going through the entire harmonic series. But you get the idea. And so that illustrates a whole bunch of things. Again, it's this-- similar to some of the other examples that we've seen, it's kind of cognitively impenetrable. So even though you know that those harmonics are there, and you were hearing them on the previous iteration, once they stopped being gated, they just kind of fuse back in with the rest of the tone. And that's the effect of onset.

So this is another pretty famous example where you can take a harmonic tone and cause it to sound like two things by frequency modulating part of it. So this starts out as a harmonic tone, and the even harmonics over time are given frequency modulation, and common frequency modulation. And so what happens is that at the very start of the stimulus, you hear this one thing. And then it'll kind of split into these two tones.

And so this is called the Reynolds-McAdams Oboe because it was created by two people, Reynolds and McAdams. And it's called the Reynolds-McAdams Oboe, because the oboe is an instrument that has a lot of power at the odd harmonics. And so when you start to modulate the even harmonics, and those are heard as a separate tone, kind of an octave up, what's left is the odd harmonics. And that arguably sounds a little bit like an oboe. So let's listen to that.

[AUDIO PLAYBACK]

[DUAL TONES RESONATING]

[END PLAYBACK]

JOSH MCDERMOTT: Yeah. So we're struggling here. The spectrogram of this is pretty underwhelming because there's a lot of noise right in back of me, which is kind of overlapping with some of the frequencies here. So let's just look at this example. And hopefully that's kind of what you heard. So it separates out into these two tones.

So naively, you look at this and it seems very plausible that what's going on here is that the common frequency modulation, like the changes in frequency that are shared across all those even harmonics, are what's causing them to group together. And that seems plausible, but you could also, in principle, just kind explain this by deviations from harmonicity. And the idea there is that when you're changing the frequencies of those second harmonics, they're at least instantaneously no longer harmonically related to the odd harmonics.

So in principle, you could just have something that's kind of very sensitive to the harmonic relations that would then tell you that you have two different things there. And so that's left ambiguous by this demo. And you could play around with this and try to test this. Yeah.

AUDIENCE: Can you have multiple modulation frequencies when you go to three or four?

JOSH MCDERMOTT: You should try it. I think it will-- it'll probably start to sound more like a texture, would be my prediction. But it's not that easy to pay attention to lots of things at once. And so with just two, you can do it, but if you push it up to four, I think it's going to-- you might be able to pay attention one of them at a time, but feeling like you hear all of them at the same time, I think that will be challenging, is my guess. Yeah.

But can anybody think of how you might test whether this effect might not be driven by harmonicity alone? What could we do to the tone that might give us a purer test of whether or not common frequency modulation is enough to make you hear something as separate from a bunch of other frequencies? Yeah.

AUDIENCE: Do you use inharmonic tones [INAUDIBLE]?

JOSH MCDERMOTT: Exactly. Yeah. Yeah. You should try it. Let me know what happens. Yeah. I know what happens, actually, but you should try it. Yeah. Yeah, that's exactly right. Yeah.

So the idea is that if the tones are not harmonically related, then when you comodulate a bunch of them, they're still just as inharmonic as they were before, and you might be able to tell whether the comodulation is enough to really drive the effect.

Another thing that could potentially be a powerful grouping cue, and turns out to be a powerful grouping cue, is repetition. So this is a graph that's-- it's showing you a time frequency decomposition of sound. And if you just look at this, you can notice that there's some repeating structure. You see those like blobs up at the top that seem like they're happening kind of regularly.

And lots of natural sounds kind of repeat. Not so much speech, but lots of animal vocalizations. So most nonhuman animals, they pretty much can say one thing. It means like, hey, it's me. So they have this one call, but typically, if they vocalize, they'll do it several times in a row. So repetition is actually very common in animal vocalizations.

So it seems plausible that the brain might have learned to leverage the fact that repetition is a powerful cue that you have some sound source out there in the world. So these are some experiments to try to test that. So in these experiments, we will present people with target sounds and distractor sounds. So these are modulated noises that will just sound kind of funny.

So you add these two together, and you get a mixture. And so the question is whether or not you can actually segregate that mixture and estimate the constituent sources. And so the idea behind making these kind of funny noise sources is that they don't really have any grouping cues of their own. There's not much in the way of common onset, there's no harmonicity, none of the classical stuff that we've just been talking about is present in these things.

And so the idea was to see whether repetition alone would be enough to allow you to actually segregate mixtures in this situation. And so the way that you test that is by presenting a mixture, and then presenting a probe sound. And the probe sound can either be one of the sounds in the mixture, or something different. And the idea is that if you are able to correctly segregate the mixture into the constituent sources, you should be able to solve that task of saying whether the probe was one of the sources in the mixture.

But it turns out that, for these sounds, if you just get a single mixture this is really hard. So it sounds like this.

[AUDIO PLAYBACK]

[MID-FREQUENCY STATIC]

[LOWER-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: Let me do it again.

[AUDIO PLAYBACK]

[MID-FREQUENCY STATIC]

[LOWER-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: So the question is, that sound at-- the second sound, that's the probe. Was that present in the mixture. And you don't know because for these kinds of sounds, they just really fuse together. You don't really hear them as separate things. And it turns out that probe sound actually was in the mixture.

So the sounds have some structure, but not enough to produce the segregation of a single mixture. However, if you present the target sound kind of repeatedly, and these are schematics of how this works, you're able to hear it out actually really well. So you just heard this.

So we have the target sound in red, a distractor in blue, and then a probe sound in black. And this was actually pretty hard. But let's think about what would happen here if we have the target repeated twice, each time paired with a different distractor, or three times, or five times, or 10 times.

And so this graph is plotting a measure of how well people can perform this task. The measure here is what's called the ROC area, which is related to d prime, which we talked about earlier. So the point is that chance performance here is going to be 0.5, and values above that will mean that you're able to perform the task.

And so this graph is showing that as you increase the number of mixtures, which, in this case, means hearing repetitions of the target sound, you're able to hear the target out pretty well. This is pretty easy to verify for ourselves. So I'm going to play these demos. And in all of these cases, there'll be a probe sound at the end. And the probe sound will be the same as the target sound.

And so what you need to ask yourself is whether the sound at the end resembles something that you heard in the mixture, or the sequence of mixtures. And so when you just have a single mixture, it's not going to sound very similar.

[AUDIO PLAYBACK]

[MID-FREQUENCY STATIC]

[HIGHER-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: Let's do it again.

[AUDIO PLAYBACK]

[MID-FREQUENCY STATIC]

[HIGHER-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: But with two, maybe you can hear a little something.

[AUDIO PLAYBACK]

[MIXED FREQUENCY STATIC]

[MID-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: Probably heard something repeat. Let's try three.

[MIXED FREQUENCY STATIC]

[MID-FREQUENCY STATIC]

[END PLAYBACK]

JOSH MCDERMOTT: There you get it pretty clearly. And with five.

[AUDIO PLAYBACK]

[MIXED FREQUENCY STATIC REPEATING]

[END PLAYBACK]

JOSH MCDERMOTT: And 10 is easy.

[AUDIO PLAYBACK]

[MIXED FREQUENCY STATIC REPEATING]

[END PLAYBACK]

JOSH MCDERMOTT: So the point is that there's this latent repetition in the signal that kind of pops out to your auditory system, even though any single one of these mixtures is kind of impossible to parse. And so this is a demonstration that kind of repetition alone is a pretty powerful grouping cue. Any questions about that?

And this is actually probably very, very important in music, where there's also sounds that kind of repeat, probably helps you hear rhythms. And my guess is it has its roots as something that co-evolved with animal vocalizations because these repeating vocalizations are very common.

So a couple lectures ago, we were talking about sound localization. Sounds often occur at different locations in the world, and it's kind of natural to wonder whether spatial location would really help you with auditory scene analysis. And there's a bunch of examples of a role for spatial cues.

This is one that's pretty famous, where spatial cues actually improve your ability to detect masked sounds. So again, we have the happy and sad people that are indicating conditions in which the tone is masked or not. And we know from the lecture on masking that you can take a tone and play it with noise, and we can adjust the levels of the tone and the noise such that the tone is masked, meaning that people can't detect it. So the person is unhappy.

Turns out, though, that if you add an identical noise-- so, OK, so this is a situation where you just play them to ear ear. So it's just going into the right ear. Turns out that you can take the identical noise and put it into the opposite ear and that will render the tone detectable. Anybody have any idea, like why would that happen in terms of what we know about sound localization?

So answer me this. In the bottom scenario, where is it going to sound like the noise is coming from? Where will the noise be situated spatially?

AUDIENCE: Probably like in front of the person.

JOSH MCDERMOTT: Yeah, like at the midline. because? We got exactly the same noise in the two ears, that means that the interaural level difference is zero. It means that the interaural time difference is zero. That's consistent with it being on the midline.

Whereas the tone is exclusively present in this case in the right ear. So you have a huge ILD, and the tone will sound like it's kind of coming from the right. So you've got very different spatial cues, and that tends to have a pretty big effect on your ability to detect the tone.

So this is known as the binaural masking-level difference. This has been studied a lot. This is one of many, many graphs. So what this graph is showing you-- this graph is like from a paper that was measuring this effect as a function of a lot of different stimulus parameters. Those effects don't really matter for the purposes of this class. I'm just giving you this as an example of the effect.

And the key comparison here is-- so these are thresholds. In this case, it's a function of bandwidth. That doesn't matter. But the key comparison here is between the circles and the diamonds and the squares.

So the circles are where the tone and the noise both have the same phase, so they're both in one ear. And then the other symbols here are when you introduce a phase difference between the tone and the noise, for instance, by putting the noise in both ears. And you can see that there's a big difference in thresholds here, up to possibly 20 dB, so it's a huge effect.

So it kind of follows, though, that if we were to add the tone back in to the other ear, it's now going to become-- it's now going to be hard to detect again, because we set the signal-to-noise ratio of the tone and the noise such that if you just play it for one year, it will be below threshold. So if you have the identical tone and the identical noise in the other ear, that would also be below threshold.

So now subject to the constraint that the tone and the noise have to be in both ears, what could you do to the tone or the noise in order to render the tone audible again, based on what you know about binaural hearing? Yes, Krista.

AUDIENCE: You would phase shift one-- like either the noise or the tone of one ear.

JOSH MCDERMOTT: Yep. Exactly. So you introduce a phase difference for one of the two signals. That will introduce a interaural time difference and the signal would be detectable again. Cool. Any questions about the binaural masking-level difference?

So one example of a case where spatial cues seems to influence auditory scene analysis. So let's keep it moving. There's another really cool spatial effect on grouping that relates to reverberation. So we were just talking about reverberation in the context of distance perception.

So reverberation refers to all the reflections that occur when you're in a natural environment. So this is like a schematic of what happens to sound when you're listening in a room. So the idea is there's a person here, there's a speaker here that's playing sound. There's this direct path from the speaker to the two ears of the person. Those are the green lines.

But then there are all these other indirect paths. So the blue lines are the paths that involve a single reflection. The red paths are the paths that involve two reflections. There's quite a lot of those. And then for the purposes of reducing chaos, the picture doesn't show you the paths that involve three reflections, and four, and five, and so forth.

And so what happens is that these alternative paths that the sound can take, well, they're longer, and so the sound arrives a little bit later. And the sound also gets filtered because any time you reflect off of a surface, the surface absorbs some of the sound energy and typically filters it a little bit. And so what happens in practice is shown here.

So we can characterize the reverberation in any setting with what's called the impulse response. So this is the sound that would be produced at a microphone at a particular position if the source was at another position, and if the sound source played an impulse that's like a click. So there's an impulse, and then this is what's recorded at the microphone in a room.

And so what happens here, well, so this is the time at which the click is played from the source, the speaker. And you can see that initially the microphone doesn't have any signal. This is the propagation time. So then you get this initial impulse in the impulse response. That's the direct sound right. That's the sound that's reaching the microphone direct from the source.

But then you can see that there are all these subsequent little things that happen in the impulse response. And so initially, these are discrete reflections that are happening from some of the shortest paths. And then after that, the reflections get sufficiently dense, so you just kind of have this long tail that slowly decays, because every time there's a reflection, the sound gets absorbed a little bit, it gets attenuated. And so eventually, this kind of fades down into something that's undetectable.

So this is how you characterize reverberation. And these impulse responses have been measured in lots of spaces. And the key thing to note here is that you always see this direct sound followed by these later reflections.

Now, one of the big challenges that this poses is for sound localization. Because the other thing that you should note here is that the sound that is traveling the direct path is coming from the direction where the sound source actually is in the world. By comparison, the sound that is reflected is actually coming from other directions, which is not where the sound actually is from. So you're essentially getting all these delayed copies of the sound coming from the wrong direction. So it's a major challenge for sound localization.

And there's this effect that's known as the precedence effect that seems to be related to this. So this was discovered by Hans Wallach, who was an early Gestalt psychologist, did lots of really interesting things back in the '40s. And the setup for measuring this is shown here.

So you have the person, they're in an anechoic chamber. Remember, that's one of those special rooms that's kind of treated so that the room itself really doesn't have many reflections. And there are two speakers, one to the left and one to the right. That's minus 45 and plus 45.

And the speakers emit clicks. But critically, there can be a time difference between the clicks. And so one of them is designated as the lagging speaker, and one is the leading speaker. That means it comes first.

And there's two phenomena that happen in this setting. The first is what's shown in this graph. So this is a graph that plots the perceived location as a function of the delay between the clicks.

So if the delay is zero, that means the two clicks are synchronous. And what happens then is because of the symmetry, you just hear a single sound that's kind of coming from the middle. So the perceived location is zero.

But if there's a short delay of a few milliseconds, what happens is that what you hear is a single sound at the location of the first click, the leading click. So that's why it's called the precedence effect. The location that you hear is dominated by the initial bit of the sound, the sound that precedes the rest of the sound-- precedence effect

So this happens for short delays. And then there's an echo threshold, which is maybe 5 milliseconds. And at that point, once the delay gets too long, you start to hear two separate sounds at the two actual locations.

So this is a grouping effect because there's two clicks that are actually being produced. And under some conditions, you hear them as a single thing. And the location that you hear is attributed to the leading click.

And so it seems that the representation of the lagging click is actually kind of suppressed in the auditory system. And evidence for this comes from this related effect here called discrimination suppression. So here, the situation is a little more complicated.

We now have three speakers. So we still have the one that's off to the right. That's the leading speaker. But now we have two, one at minus 55, and one at minus 35. And so these are clicks that are coming second.

And so the task here is-- so what I should have told you is that on every trial, there's a click that comes from here first, and then either a click from here, or a click from here. And your job is to judge whether the second click came from minus 55 or minus 35. And so what this graph is plotting here is your performance on this task as a function of the delay between the clicks.

And so the point is that when the delay is short, you're not able to perform this task. Once the delay gets longer, you start being able to discriminate between this location and this one. So it's as though that click is kind of being suppressed by your auditory system and you lose access to the location information. Question?

AUDIENCE: If it was just undiscernible, wouldn't like the bottom be around like 50%, which is like just chance? Or when the delay is small, are they actually, like...

JOSH MCDERMOTT: Yeah, I--

AUDIENCE: [INAUDIBLE] perceive it to be--

JOSH MCDERMOTT: I think you're correct. I was thinking that to myself as I was explaining this. I think this is a mistake on the graph, because you're absolutely right that this should be at chance. And if this is a 2AFC task, then that would be 50%. So yeah, this is a schematic summary from a review paper, which may have an error in it.

But the main thing to take away from this is that you are unable to discriminate between these locations when-- under conditions in which you would experience the precedence effect, so when the delay is short. And I can go hunt down the original graph and send that to you. So that's the precedence effect.

And the precedence effect has some other kind of interesting properties. One is that it exhibits this phenomenon of what's called build-up. And that refers to the fact that you can be out in this regime here, where the delay between the clicks is above your threshold. But if the clicks are presented repeatedly in the same spatial configuration, what happens over time is the precedence effect builds up even in these settings where you initially didn't hear it. And so that's what's indicated here.

So you repeatedly present the lead and then the lag, the lead and then the lag, the lead and then the lag. And so initially, this is supposed to be indicating that you're hearing both of them. And then what happens over time is that you stop hearing the second click.

And then what you can do if you're an experimenter, is this sort of tricky thing of actually switching the order. So now the leading click is on the left, and the lagging is on the right. And you now hear both of them, and then it builds up again. And so here's a graph that shows that effect.

So this is plotting the proportion of trials on which you report hearing two sounds. And this is the presentation number. So initially, there's this particular arrangement of lead and lag, and then it switches here, and you initially shoot up to hearing both sounds. And then it kind of builds up again over the course of a few trials.

And so the kind of naive interpretation of this is that this is an adaptation that helps us localize sounds correctly in environments that have reflections. And one explanation for this build up effect is that if you're in a situation where you're hearing these two sounds that are always occurring with this very precise and consistent temporal relationship between the two of them, and they're otherwise identical, like two clicks, odds are that relationship is kind of caused by something in the world.

And a plausible candidate for that is the positioning of surfaces in the world. And so the brain is presumably interpreting that these are likely to be reflections. So it's almost as though you're kind of building up a little model of the world's reflections.

So this is kind of-- like the story that we associated with this effect is a nice story. It seems plausible. But it's often really hard to test whether this is, in fact, like an adaptation to deal with reflections. Question over there.

AUDIENCE: So if we get better at this if it's in a room where there is reflection, would the delay be less?

JOSH MCDERMOTT: Well, what this is suggesting is that, no, actually the echo threshold increases. You're getting worse. And you're getting worse at detecting the presence of the lagging sound, but better from the standpoint of correctly interpreting the world, like if, in fact, this is caused by one sound, and these are reflections. Yeah.

So the question is like, can we actually give this story that we want to tell about the function of this any teeth? And again, computational models can help us with this. So you recall from when we were talking about sound localization, we discussed how, if you take artificial neural networks and train them to localize sounds from binaural audio, they end up exhibiting some of the behavioral phenomena that we observe in human listeners.

And it turns out that they also exhibit the precedence effect. And you may recall how this model is trained in a simulated world, where we render spatially, the binaural audio that would enter the ears of a person in a room listening to sounds at different positions. And so there are reflections in this world.

So you can then go and test the model on the precedence effect. So you can render this situation where you have a click coming from here, and a click coming from here. And so this is the binaural audio. The signal for the left ear, the signal for the right ear.

And so the leading click here, the first one, is coming from the right. So you can see this small time difference between the right ear and the left ear. The lagging click, the one that comes second, is coming from the left. So you can see the left ear signal arrives a little bit before the right ear.

And then this is the precedence effect in the model. So this is the model's judged location as a function of the interclick delay. Now, this model is constrained to report a single location. So it can't tell us that it hears two things. But the location that it does ascribe to the stimulus is that of the leading click when the delay is short, and then that kind of goes away. So it looks an awful lot like the precedence effect.

So that's nice. But the other thing that we can do with a model like this is we can ask, well, how would the model behave if it had evolved and developed in a different kind of world? And in particular, because this model was trained in a simulator, we can mess with the simulator to make it deviate from the natural world in different ways.

So we can train the model in anechoic conditions. So this is a situation in which every surface absorbs all the sound that it receives. It's like if you grew up and evolved in anechoic chambers, we can get rid of background noise, make the sounds unnatural, different kinds of alternative training regimes. And so then we can optimize the model in these different kinds of worlds, and then test it behaviorally and see if it exhibits this phenomenon.

And what this graph is showing you is that in several of these alternative training conditions-- so I should say the natural one, or the normal one, is the light blue. And then in two of these other alternative training conditions, you still see the precedence effect. But look what happens in anechoic conditions. So if the model is trained in conditions without reflections, that effect goes away. So that's kind consistent with the kind of classical story that this is a mechanism that has evolved to help us cope with reflections and the challenge that they pose to sound localization.

So that's the precedence effect, and evidence that this is something that helps us deal with reflections. The bigger picture here is that this new era of machine learning-driven modeling can give us some ways to ask why our perceptual systems are the way they are, and what are the problems that they were actually designed to solve. What questions do you have about that? Yeah.

AUDIENCE: How does that even model sound waves? Like when it comes to [INAUDIBLE]. I mean, all the presentations we've seen [INAUDIBLE] single lines. So what's the intuition for how do you manage that in a 3D space?

JOSH MCDERMOTT: Yeah, I mean, so this is probably a respect in which-- I mean, so there's a simulator that kind of does it. And it treats them as though they are traveling along lines, and that that's an approximation. It's not fully accurate. All this stuff is suggesting that it's probably good enough for our purposes, but it's not perfect. And those are things that will get better with time.

So here's a quick summary of these auditory grouping cues that we've discussed. We talked about common onset, comodulation, harmonicity, we talked about how there's this ambiguous role for frequency modulation, talked about repetition, and then some role for spatial cues. And all of the things that we've been discussing so far really relate to how the auditory system assigns energy that's kind of received in of local neighborhood in time, different sources in the world.

But there are also a whole bunch of phenomena that relate to the grouping of sound energy over longer periods of time. And this is often referred to as streaming. And so I'm going to give you a few examples of that.

So this is the most classic streaming phenomenon. This was really introduced back in the 1970s. And so back then, really people were very limited in terms of the kinds of sounds that you could actually do experiments on, and so there were lots of experiments that were done on beeps and boops and noises. That was pretty much it.

But they nonetheless, discovered some fairly interesting effects. And so in this particular stimulus, you will be presented with alternating high and low tones. So a high tone, a low tone, a high tone, a high tone, a low tone, a high tone.

This sequence of tones can be perceived as part of a single process, kind think of that as a single melody, in which case there's a galloping rhythm. 1, 2, 3; 1, 2, 3; 1, 2, 3; or as two separate streams, one with high-frequency tones and one with low-frequency tones.

So in general, if the difference in the frequency gets too great, or if the time between the tones gets too short, people will hear two separate streams. Now, you might think that this distinction between one and two streams is kind of like a weird and subjective distinction. It's like, well, obviously like this could be one thing or it could be two things.

But it just turns out that, in practice, when you listen to this, there's a pretty clear, subjective sense that it either sounds like one thing or two. And it's very consistent, and it works for lots of people. So let's have a listen.

[AUDIO PLAYBACK]

- Demonstration three, loss of rhythmic information as a result of stream segregation. You will hear a cycle of two high tones and one low tone in a galloping pattern. Listen to what happens to the experience of the gallop as we speed up the sequence by shortening the tones. First, the tones are far apart in frequency.

[GALLOPING HIGH AND LOW TONES INCREASING IN SPEED]

[END PLAYBACK]

JOSH MCDERMOTT: And I should have said-- so we got the tones far apart. Initially, they're probably going to sound like they're two distinct things. What probably will happen is as they speed up at the very end-- or sorry, initially, because they're so slow, you will hear them kind of potentially as a single melody. But what will happen as they speed up is it's going to split. And you'll start to hear this high beep-beep-beep and a low beep-beep-beep

[AUDIO PLAYBACK]

[HIGH AND LOW TONES BEEPING WITH INCREASING SPEED]

- Next the frequency separation is small.

[END PLAYBACK]

JOSH MCDERMOTT: And so now the frequency separation is small, and so there's going to be a tendency for you to hear this as one stream. So you're going to still get this effect of speeding up, but you'll probably hear it as one thing until the very, very end. So see what happens.

[AUDIO PLAYBACK]

[HIGH AND LOW TONES BEEPING WITH INCREASING SPEED]

[END PLAYBACK]

JOSH MCDERMOTT: So at the very end, they split for me. I don't know what happened to you. So this seems like a very subjective phenomenon, but it turns out that there's this fairly interesting objective consequence to this difference in subjective state, and that's what's stated here.

So when a sequence is parsed by your auditory system into two streams, it becomes very difficult for you to judge temporal relations between the tones or the elements of the two streams. I mean, it's a very interesting and important phenomenon because it suggests that there is some fundamental change in the way that the sound is represented.

So here's the essential idea. So you're going to-- there's my cursor. So you're going to hear these high and low tones. And in one case, the frequency separation will be small, and so you're going to be a lot more likely to hear this as one thing. In another case, it'll get big, so you're going to tend to hear it as two streams.

And what we're going to be looking at is the extent to which you can tell whether the low tone is kind of centered right between the two high tones or offset a little bit. And the phenomenon that you'll probably notice is that when the frequency separation is small, so that you tend to hear this as one stream, it's pretty easy to tell the difference between those two rhythms.

Whereas when the frequency separation gets big, that distinction kind of gets lost because you just hear it as two streams. So you still have that same physical time offset between the two tones, but you're not going to really hear it very clearly. And actually, I'm not going to show you the spectrogram for this because you could use it to cheat. So let's listen.

[AUDIO PLAYBACK]

- Demonstration 13. The effects of stream segregation on the judgment of timing. We start with regular repetitions of a single pitch and occasionally add a lower tone to create a galloping rhythm. Sometimes the low tone is halfway in time between the high ones, and sometimes closer to the second one.

In each sequence, you are to judge whether the low tone is exactly halfway between the high ones. First, we present two examples in which the low tone is close in frequency to the high one. Next, we present two examples in which the low tone is far in frequency from the high one.

[GALLOPING RHYTHM OF LOW AND HIGH TONES]

JOSH MCDERMOTT: OK, could you hear the difference between those two rhythms? Yeah like one-- it's a little bit subtle, but one, just- - the timing sounds different. So now we're going to do the same thing, but with this bigger frequency separation.

[HIGH-FREQUENCY RHYTHM WITH LOW TONE SEPARATION]

[END PLAYBACK]

JOSH MCDERMOTT: So the argument is that in this setting here, where we have this big frequency separation, the subjective difference between the timing is very subtle. And objectively, you can measure people's ability to discriminate and show that they're worse. So that's just an important fact, and an objective consequence to the streaming phenomenon.

So there's this kind of-- one way to think about the problem that we've been talking about, where you get some auditory input like this. It's just energy that's spread across frequency and over time. And you could think of the scene analysis problem as possibly initially requiring you to determine the grouping of the observed sound elements.

And so what's shown here is mixtures of the target talker with a bunch of additional talkers. And here, the pixels in the cochleagram representation have been colored red if they correspond to distractor sources. So this is energy from the target talker, and then this is from some other talker. So you might imagine that you need to be able to figure out which bits of energy belong to one source and which belong to another source.

But there's this other thing that happens. You can see that a lot of the pixels are colored green. And so those are places where the target talker had a fair bit of energy, but where all the other talkers actually have more energy. And so that's a situation where the target talker is going to be masked. The other sources will make it difficult to actually detect the content of the target source at those frequencies at those moments in time.

And as we've discussed, because of the fact that sounds add in the air, masking is ubiquitous in the world. Things are always masking other things. And it seems plausible that you might want to be able to infer the content of sources in these places where they're masked. And in fact, there's a fair bit of evidence that this happens all the time.

So this is a classic effect that's often known as the continuity effect. And it consists of a stimulus where you have tones that are separated by a gap, and, in some cases, the gap will be filled by noise. And the phenomenon is that if the noise is sufficiently high in intensity and contains frequencies close to those of the tone, you will subjectively hear the tone as continuing during the noise, even though the tone is not physically present there.

And the idea is that the brain is inferring that the tone's absence during the noise is best explained by masking, and so it gets filled in. So the idea is that here's a stimulus-- again, this is sort of spectrogram. You got your tone. You got your noise. And there's two possible explanations for the stimulus-- at least two.

One is that you have two tones that just kind of happen to be sandwiching this noise. The other is that the noise is continuous-- sorry, the tone is continuous, and that the reason that you don't actually have physical evidence for the tone here is because it gets masked by the noise. And most of the time, this is what you hear if the noise is sufficiently high in intensity so as to plausibly mask the tone. let's check this out.

[AUDIO PLAYBACK]

- Pulsation threshold. You will hear a 2,000 hertz tone alternating with a band of noise centered around 2,000 hertz. The tone intensity decreases one decibel after every four tone presentations. Notice when the tone begins to appear continuous.

JOSH MCDERMOTT: So just to make sure everybody gets what's happening, you're going to hear this tone-- beep, beep, beep, beep, beep-- alternating with noise. Initially, the tone is going to be pretty high in intensity. It's going to gradually decrease. It will reach a point at which the tone is low enough in level that it could plausibly be masked by the noise. And at that point, you're going to start to hear this thing as continuous, even though the tone is always going to be pulsing.

[HIGH-FREQUENCY TONE PULSES]

[TONE PULSE CONTINUOUS]

[END PLAYBACK]

JOSH MCDERMOTT: At the end, were you hearing that beep? Yeah. OK. So that's what's known as the continuity effect.

And it happens for tones. It also happens for speech. And speech is often referred to as phonemic restoration. And this is a famous picture that is thought of as a visual analog of this, where what we're going to be comparing here, it's just like going to be like the tone, where we can have tones separated by gaps that can be filled with noise or not.

And in this case, we've got these letters that are occluded by of splatter. And if you just get the fragments of the letters on their own, it's kind of hard to interpret. But when you have visual evidence that they're occluded, in this case, it's a lot easier to see that they're Bs.

So the speech demo is kind of the analog of that. So we're going to have a speech signal. Little bits of it will be deleted, replaced by silence. And then in some cases, those silences will be filled by noise. And we're going to ask whether it's easier to hear the speech when the noise is present.

[AUDIO PLAYBACK]

- Demonstration 31. The picket fence effect with speech. We play a sentence in which half of the sound has been eliminated by taking out every other 1/6 second segment and replacing it with a silence. Then the same sentence is played with loud noise bursts replacing the silences. Does this make the speech sound more complete? Finally, you will hear the intact sentence.

- (CHOPPILY) Auditory scene analysis involves a grouping of sounds. The principle of similarity is very important.

(WITH STATIC PULSES) Auditory scene analysis involves a grouping of sounds. The principle of similarity is very important.

(NORMAL) Auditory scene analysis involves the grouping of sounds. The principle of similarity is very important.

- Again, half the speech is deleted, but this time, every other quarter-second segment has been deleted. Again, judge the apparent completeness of the sentence.

- (CHOPPILY WITH LONGER DELETIONS) Auditory scene analysis involves grouping of sounds. The principle of similarity is very important.

(WITH LONGER STATIC PULSES) Auditory scene analysis involves grouping of sounds. The principle of similarity is very important.

[END PLAYBACK]

JOSH MCDERMOTT: So did you feel like you were hearing the speech when the noise was is present? Yeah. So that's phonemic restoration. And did you have the sense that it worked better the second time? Yeah.

So I think there's some optimal stimulus parameters for this that are partially a function of how much context there is to be able to fill in the stuff in the noise, and some evidence that there's actually linguistic constraints on what you fill in. It's an interesting effect. Phonemic restoration.

The last thing I'll leave you with is an example where this happens with sound textures. So textures are sounds that result from large numbers of similar acoustic events. So things like rain or swarms of insects, fire, wind, things like that. And there seems to be this pretty powerful inference of texture when it's masked. And so I'll play this example.

So what you're going to hear is two seconds of texture followed by two seconds of masking noise. And I'm actually going to turn the volume down here. Just don't want anybody to lose their hearing. And you will probably have the sense-- and the texture is the sound of applause. You probably have the sense that it continues during the noise.

[AUDIO PLAYBACK]

[APPLAUSE]

[STATIC]

[APPLAUSE]

[END PLAYBACK]

JOSH MCDERMOTT: So it's kind of interesting because the noise is a full two seconds long, and you feel like you hear the thing throughout the whole thing, unless you think it's the power of suggestion. This is a version where there's a small gap between the end of the texture and the start of the noise, which makes it a lot less plausible that the texture continues, and that tends to kill the effect.

[AUDIO PLAYBACK]

[APPLAUSE]

[STATIC]

[APPLAUSE]

[END PLAYBACK]

JOSH MCDERMOTT: And so the big idea here is that these processes of filling in, we think are going on all the time, like behind the scenes without you knowing it. So you probably notice there's this kind of annoying humming sound that kind of comes from the cabinet in back of me. But that is continually masked intermittently by my voice.

So when I talk, there will be periods of time where that noise is fully masked. Yet it probably sounds like the noise is totally continuous. And of course, physically it is, but the fact that it sounds continuous is because your brain is just kind of filling it in subjectively. So that you're kind of hearing your estimate of what's actually out there in the world, and that estimate is intended to compensate, in part, for the fact that things are masked all the time.

Similarly, I can talk, and I can clap my hands like this. So my clapping is going to be masking my speech and it sounds just continuous to you. So this stuff goes on all the time, kind of behind the scenes. Yeah.

AUDIENCE: Would this still work with like a texture that doesn't sound like the noise? If I really think about it, the applause does sound a bit like static noise. But if you replace the applause with popping bubbles, for instance, would this effect still hold?

JOSH MCDERMOTT: Yeah, so there's a whole paper that explores the dependence on stuff like that. And in general, the effect is strongest when the sounds are stationary, which is to say kind of stable over time. So something that has these intermittent popping bubbles gets a lot harder to hear in this way. And in particular-- so if you do this exact same-- set up this exact same thing, but with, say, speech or music instead of the texture, you don't hear the speech or music kind of continuing over two seconds.

So if you shorten the noise to, say, like 200 milliseconds, you'll get the continuity effect with speech, just like we heard. You get the continuity effect with music. But it doesn't last for that long. So the situations in which you subjectively hear this thing lasting for two seconds are those where the sound is stationary. And one reason for that is potentially because those are sounds that are, A, very predictable, but, B, also very likely to continue on very long timescales in the world.

So textures are typically generated by lots of similar events. Probably makes it unlikely that they stop abruptly. And so you probably have a pretty strong prior that they just go on for long periods of time would be my guess, whereas for things like speech or music, you're not going to have that same prior. That's all for today.