Summary: Sounds and words

Last week, we talked about how our perception of sounds might be influenced by context (e.g., visual information about how a talker is moving their mouth). As it turns out, **lexical knowledge** (i.e., knowledge about which sound strings constitute real words) can also influence our understanding of speech sounds.

We see this, for instance, in the **Ganong effect**, whereby categorization of an ambiguous speech sound is influenced by lexical knowledge; if we hear a sound that's ambiguous between /g/ and /k/, we're more likely to call it /g/ if it's followed by -*ift* and /k/ if it's followed by -*iss*.

We also see this in the **phoneme restoration effect** – against a background of noise, we might report hearing speech sounds (like /s/) that aren't actually in the acoustic signal (*legi_lature*) such that the signal is consistent a word we know (*legislature*).

Lexical knowledge does not just affect processing in the moment – it can also help us learn about a talker produces their speech sounds, guiding perception in future encounters. Such **lexically guided perceptual learning** is similar to the phonetic recalibration aftereffect that we read about in the Bertelson et al. paper.

Of course, in order for words to have an influence on our interpretation of speech sounds, we need to first recognize words. And so we turn now to a discussion of spoken word recognition.

When you hear a word, you automatically also think of other words that are semantically related to it (i.e., related in meaning). It may therefore be helpful to think of the **mental lexicon** (mental dictionary) as a network-like structure, with connections between lexical items that are similar to each other in meaning. The consequence of this is that hearing one word **primes** semantically related words.

In order to understand a spoken word, a listener must map from an acoustic signal to a lexical representation and then access the appropriate meaning.

This should be a challenging process. For one thing, there are no clear gaps in the speech stream to help us segment out the words we are hearing. For another, speech unfolds over time, rendering it temporarily ambiguous; if we've just received the input "can," we may not know if someone is saying the word *can* or the first syllable of *candy*. So how do we activate the right word?

The **cohort model** of spoken word recognition suggests that that we activate all word candidates that are consistent with the input we've received so far, whittling down our set of candidates (i.e., eliminating items from the cohort) as we get input that disconfirms some possibilities. The degree to which we activate possible candidates partly depends on how frequent the words are in our mental lexicon (mental dictionary). These ideas are supported by "**visual world**" eyetracking studies as well as by cross-modal semantic **priming studies**.

While we tend to see strong competition between words that match in onset position (e.g., *candle* and *candy*), we also see some competition between rhyming words that do not share onsets (e.g., *candle* and *sandal*).

The process of lexical access can therefore be characterized as a dynamic process with competition between similar sounding words. But even once we have activated the word, it may not always be clear what it means; words like *bank* have multiple meanings, for instance. The context in which we encounter a word can help us access the right meaning by making some meanings more likely.

And so, as seems to now be a recurring theme, it seems that there is an important role for context in word recognition. Language scientists continue to debate how to think about such contextual effects. Is it the case that context directly changes our perception? Or does it only change our *decisions* about what we have perceived?