Recap…

- In Class 1, we talked a bit about the articulatory phonetics of English: what the consonant sounds are, and how we produce them.

	MANNER		VOICING	PLACE						
				Bilabial	Labiodental	Interdental	Alveolar	Palatal	Velar	Glottal
Obstruent	Stop		Voiceless	p			t		k	ʔ
			Voiced	b			d		g	
	Fricative		Voiceless		f	θ	s	ʃ		h
			Voiced		v	ð	z	ʒ		
	Affricate		Voiceless					tʃ		
			Voiced					dʒ		
Sonorant	Nasal		Voiced	m			n		ŋ	
	Liquid	Lateral	Voiced				l			
		Rhotic	Voiced					r (ɹ)		
	Glide		Voiced	w				j	(w)	

Recap…

- In Class 2, we talked about articulatory phonetics *beyond* English: what the sounds of the world's languages are, and how they're produced.

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Studying articulation

- A question that we didn't have time to address last week: how do researchers know such detailed information about how speech sounds are produced?
  - One approach: stick your fingers in your mouth.
  - Another (better) approach: study these things instrumentally.

Studying articulation

- One way to study articulatory properties of language is to use various kinds of imaging: for example, with MRI, you can observe the tongue in fascinating detail.
- https://www.youtube.com/watch?v=Nvvn-ZVdeqQ

Studying articulation

- There's also electropalatography, where a device containing many electrodes measures where the tongue makes contact with the roof of the mouth.
- https://www.youtube.com/watch?v=1w90b2xY9f8

Studying articulation

- And when these kinds of studies are unavailable (i.e. if you're doing fieldwork in a really remote area), there's always palatography…
- https://www.youtube.com/watch?v=8Sh5hFnIuS4

Class 3: Speech Perception

The Sounds of the World's Languages

HSSP Summer 2015

Three branches of phonetics

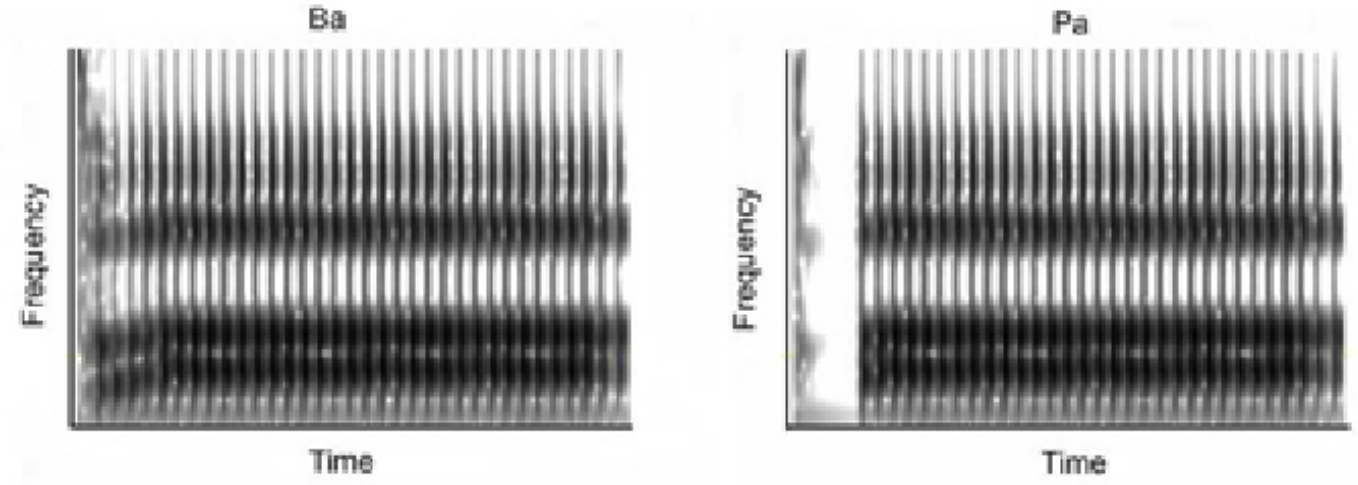
- *Articulatory* phonetics
  - The study of the production of speech sounds.
  - Aka, what we've covered so far.
- *Acoustic* phonetics
  - Study of the physical transmission of speech sounds.
  - Interesting stuff, but we won't cover it here.
- *Auditory* phonetics
  - The study of the reception and perception of speech sounds.
  - Once the sounds have reached you, what do you do with them?
  - Today we'll learn a little bit about auditory phonetics.

Remember from last week…

- There were several instances when I (and probably some of you) had a hard time telling the difference between two sounds that are distinct in another language. (Ewe fricatives.)
- This is not a personal deficit: it's a fact about how our auditory systems work.

Categorical perception

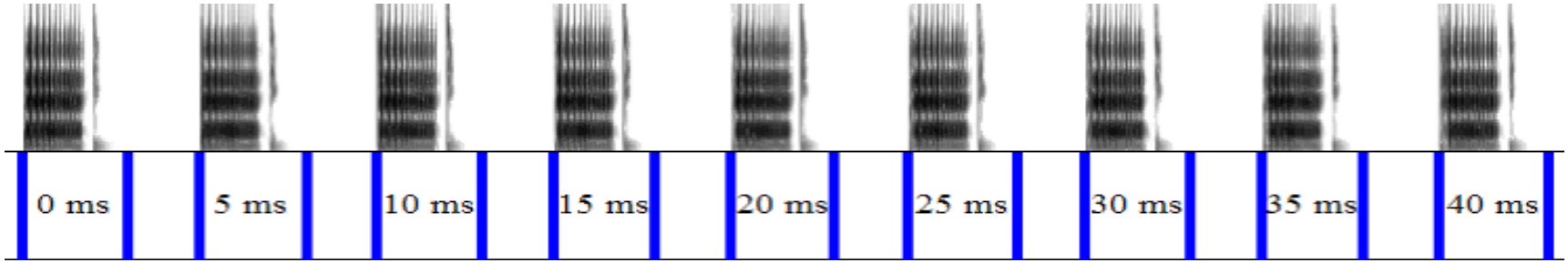
- We perceive speech in *categories*, not continuously.
- In other words, we place sounds that are similar enough into a single category and have a hard time distinguishing them.
- For example, /b/ and /p/ differ only in that /b/ is voiced and that /p/ is voiceless.



- To be more precise: in English, /p/ has a highly positive VOT. /b/ does not.

Categorical perception

- We can create a VOT *continuum*: a series of labial stops starting with a VOT of 0 ms (/b/) and ending with a VOT of 40 ms(/p/).



- Audio here: http://ucalgary.ca/pip369/mod6/speech/principles

Categorical perception

- Many of you probably began to hear a /p/ starting at a VOT of around 25 ms.
- For labial stops, that's approximately where the category boundary in English is.
- But notice something: when you were doing this task, you heard *categories*, not *sounds*.
  - Each sound in the continuum has a different VOT.
  - Yet we hear them as /p/, /b/, etc.

Categorical perception

- When we graph the results of an identification task, like the one we just did, they typically look like this:

- Point of note here: humans are *very good* at discriminating between instances of categories in their native languages.

Categorical perception

- With a VOT continuum, we can also perform discrimination tasks: is sound A different from sound B?
- Audio here: http://ucalgary.ca/pip369/mod6/speech/principles

Categorical perception

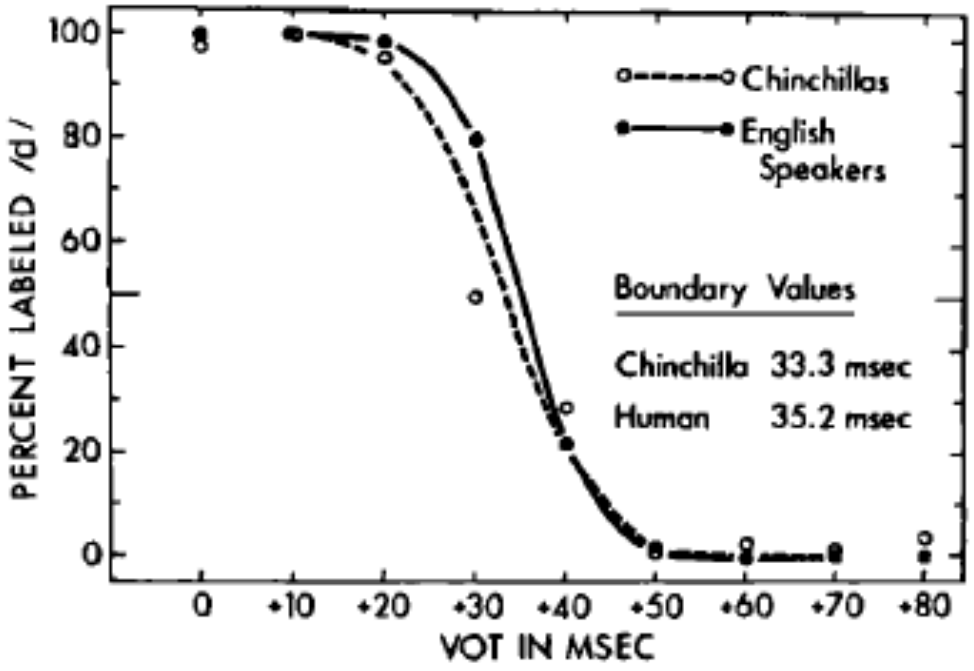
- Listeners tend to report all sounds that are on the same side of the phonetic boundary as "same", while sounds on either side of the boundary are reported as "different".
- Point of note here: humans are *very bad* at discriminating within categories in their native languages.

Categorical perception

- Why do we have categorical perception?
  - Allows us to "ignore" irrelevant variations in the speech signal. It doesn't matter if a /p/ has a VOT of 35 or 40 ms, it's still a /p/.
  - The difference between a /b/ and a /p/, however is very important!
- But sort of interestingly, categorical perception is not speech-specific… nor is it even human-specific.

Categorical perception

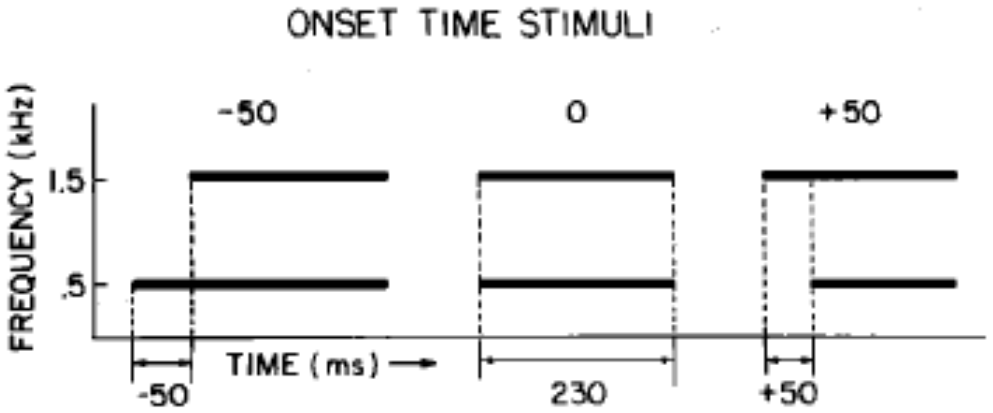
- Chinchillas also have categorical perception of VOT, and somewhat surprisingly, the phonetic boundary resembles English speakers'.



Kuhl and Miller, 1978

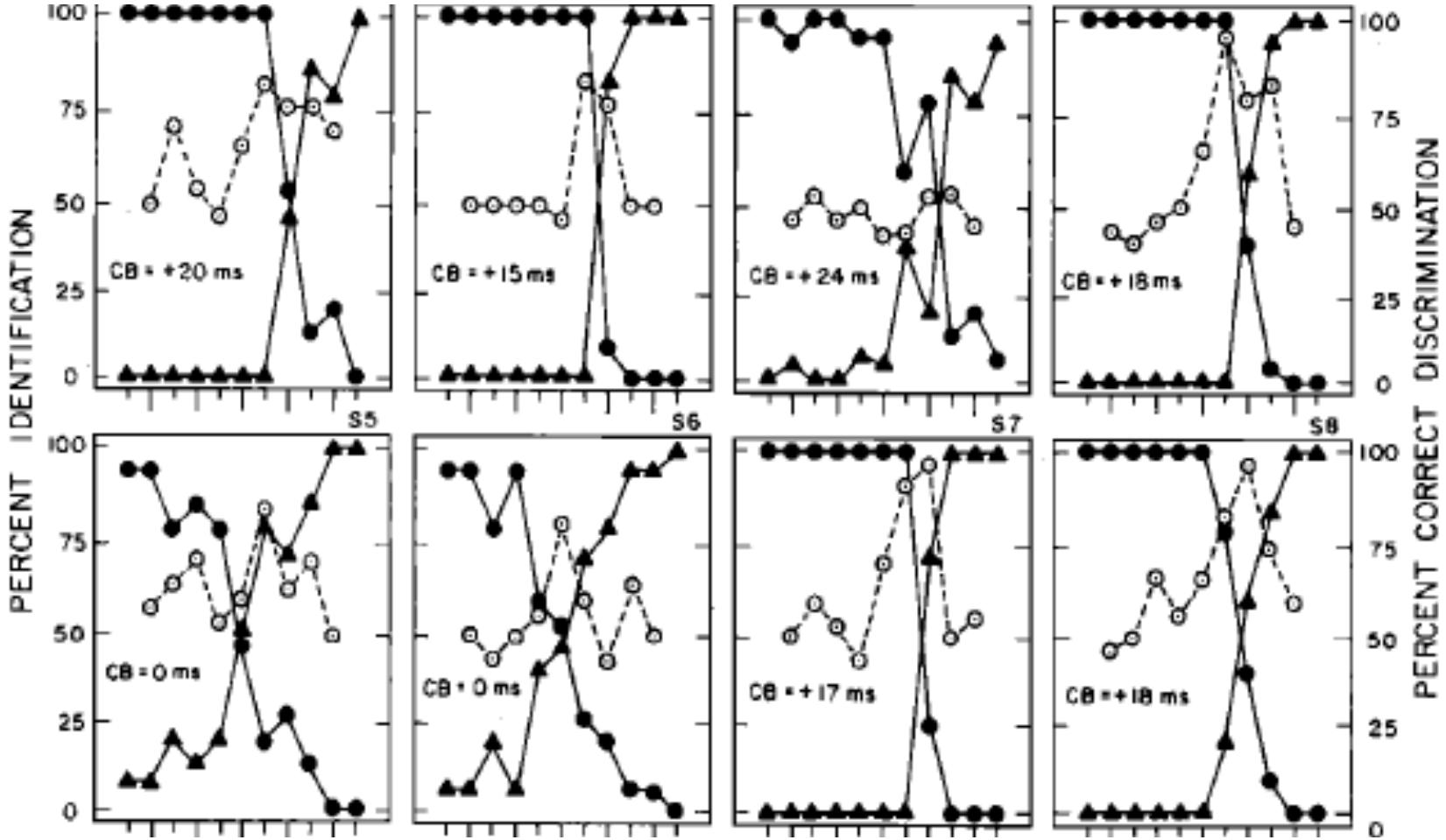
Categorical perception

- We can also elicit the categorical perception effect with non-speech stimuli.
- Pisoni (1977) asked listeners to judge whether two tones came on at the same time or not, varying the onsets of the two tones.



Pisoni, 1977

Categorical perception



Pisoni, 1977

Conclusions: categorical perception

- When we perceive speech sounds, we tend to hear *speech categories*, and not the small acoustic variations across individual instances of the same category.
- People's sensitivity at category boundaries reflects general auditory sensitivities, but also depends on memory and other processes.
- Speech and hearing evolved together: speech takes advantage of our pre-existing auditory sensitivities.

credit where it's due: much of this presentation was adapted from Lynne Werner, U Washington.

Interlude

- We've learned from categorical perception that what we *hear* is not what we *perceive*.
  - What do we hear? What you saw in the spectrograms: a bunch of formant frequencies and noise.
  - What do we perceive? Categories, more or less.
- But it's not even this simple. There's a good amount of evidence that "speech perception" involves a complex interplay of a number of factors: what we hear, what we see, what we know about the speaker, what we know about ourselves…
- We'll discuss some of this other factors next.

What do we perceive?

- What does the listener recover from the acoustic signal? And how is that information present in the input signal?
- There's no universally accepted answer, but there are a number of different theories.
  - *Gestural* theories: the objects of speech perception are articulatory/gestural. Speakers control, and listeners perceive, vocal tract gestures.
  - *Auditory* theories: domain of articulation is acoustic/ auditory. Speakers control, and listeners perceive, the acoustic signal.

What do we perceive?

- Some evidence that neither (and both) of these theories are correct comes from the McGurk effect:
- https://www.youtube.com/watch?v=G-lN8vWm3m0
- The effect is automatic and mandatory: visual speech information, when present, *cannot be ignored*.
- So when we perceive speech sounds, we are synthesizing at least two types of information:
  - *Auditory* information: what we hear.
  - *Gestural* information: what we see.

What do we perceive?

- There are a lot of questions about what it means to perceive a gesture.
- There are also questions about what it means to perceive a sound.
  - **Question**: which acoustic elements are essential for the perception of speech?
  - Rubin et al.'s **answer**: none of them.
  - Sine wave speech discards all of the acoustic elements of the speech signal, keeping only the changing pattern of vocal resonances. Yet to some speakers, the percept of speech is still there…
  - http://www.haskins.yale.edu/featured/sws/swssentences/sentences.html#

Beyond gestures and sounds

- When we perceive speech sounds, we also take into account a lot of other information:
  - Coarticulatory context:
    - A /k/ before an /i/ sounds different than a /k/ before an /a/.
    - Yet we perceive them both as /k/s.
  - Knowledge of the speaker's native language (you can adapt to foreign accents pretty well).
  - Socio-indexical knowledge (e.g. dialect, age, ethnicity, occupation).
  - Lexical knowledge: the *Ganong Effect*.
    - Ganong (1980): speakers, when presented with a sound that is ambiguous between "duke" and "tuke", for example, speakers will respond that they heard "duke."
    - We are biased to perceive real words.

Wrap-up

- Speech perception: transformation of an acoustic signal from a speaker to an intended communicative message heard by a listener.
- Cool point: for the listener, speech perception is *effortless* (well, for the most part, anyway).
- For the researcher, however, it is a complex and multifaceted problem.