Speech Perception

Theories and such

Theories of Speech Perception

- Motor
- Auditory
- Exemplar
- Some ideas from neuroscience

Alvin Liberman

As a matter of convenience, I should like at the outset to divide the consonant cues into three classes, and to make this division according to where and how the sounds are produced. I am, of course, embarrassed to introduce a discussion of acoustic cues by classifying them on an articulatory basis. However, we find here, as we so often do, that it simplifies our data quite considerably to organize them by articulatory criteria. We certainly do not mean to imply by this that there are no acoustic differences among our classes, but only that it is hard to characterize these differences very simply in acoustic terms

-Alvin Liberman, June 16, 1956, MIT conference on Speech Communication

The motor theory of speech perception

- Acoustic cues have a variable relation to speech percepts
- But a /d/, for example, is *articulated* the same way every time
- Maybe we perceive sounds by "recovering" the motoric gestures that produced them

Key Hypotheses of the MToSP

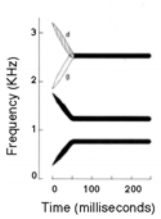
- We perceive gestures not sounds
- Speech is special — involves dedicated processing/neural systems
- Which includes the motor system, where speech motor gestures are coded in the brain

Early evidence supported the theory

- Categorical perception — seemed unique to speech at first
- Duplex perception — demonstrated the difference between "the speech mode" of perception and regular acoustic perception
- McGurk effect — showed that gestures affect perception

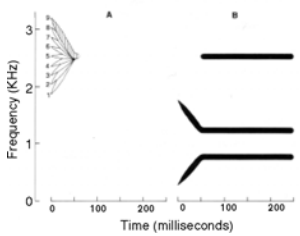
Duplex Perception

Experiments make use of stimuli in which direction of F3 transitions distinguish [da] from [ga]. Without this transition, the rest of the stimulus pattern is ambiguous between [da] and [ga].



Duplex Perception

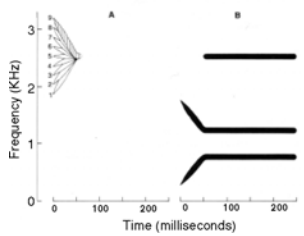
The critical formant transition (A) is presented to one ear, and everything else (the ambiguous "base", B) is presented to the other.



Duplex Perception

Simultaneous (duplex) perception:

- Listeners hear a syllable in the ear that gets the base (B), but it is not ambiguous. Its identification is determined by which of the nine F3 transitions are presented to the other ear (A).
- Listeners also hear a non-speech "chirp" in the ear that gets the isolated transition (A).



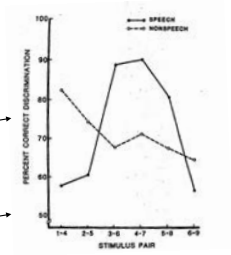
Duplex Perception

Simultaneous (duplex) perception:

- Listeners hear a syllable in the ear that gets the base (B), but it is not ambiguous. Its identification is determined by which of the nine F3 transitions are presented to the other ear (A).
- Listeners also hear a non-speech "chirp" in the ear that gets the isolated transition (A).

Categorical

Not categorical



But…

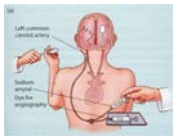
- Categorical perception is not unique to speech
- Duplex perception can be elicited by non-speech stimuli (the sound of door slamming)
- McGurk effect is probably not a motor thing, but rather a cross-sensory integration thing

And…

- There is much evidence that you don't need the motor speech system to perceive speech
  - Babies
  - Chinchillas
  - People with severe cerebral palsy
  - People with Broca's aphasia
  - Stephen Hawking (and other people with ALS)
  - People undergoing Wada procedures

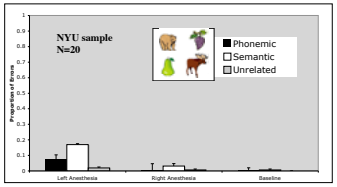
Wada procedure

- Used for pre-surgical planning in epilepsy treatment
- Assesses lateralization of language and memory function



Juhn Wada

Speech Recognition During Wada Testing



Patient is mute but can perceive speech quite well

By ~1990 MToSP was all but dead

And then mirror neurons happened

"… we understand action because the motor representation of that action is activated in our brain."

-Rizzolatti, Fogassi, and Gallese (2001) p. 661

Giacomo Rizzolatti

"…the mechanism matching action observation and execution … is very similar to that proposed by Liberman and his colleagues for speech perception…. According to this theory, the objects of speech perception are not to be found in the sounds, but in the phonetic gesture of the speaker, represented in the brain as invariant motor commands…. Considering the homology between monkey F5 and human Broca's area, one is tempted to speculate that neurons with properties similar to that of monkey 'mirror neurons', but coding phonetic gestures, should exist in human Broca's area and should represent the neurophysiological substrate for speech perception."

-Gallese, et al. 1996, p. 607

"

Liberman's intuition … that the ultimate constituents of speech are not sounds but articulatory gestures … seems to us a good way to consider speech processing in the more general context of action recognition."

-Fadiga & Craighero, 2006, p. 489

"speech comprehension is grounded in motor circuits…"

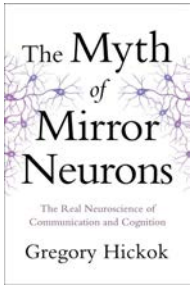
-D'Ausilio, … Fadiga et al. 2009

But… it's a zombie theory

- There is much evidence that you don't need the motor speech system to perceive speech

- Babies
- Chinchillas
- People with severe cerebral palsy
- People with Broca's aphasia
- Stephen Hawking (and other people with ALS)
- People undergoing Wada procedures

It's hard to kill but I'm trying



The Auditory Theory of Speech Perception

- Crazy idea: we perceive speech with our auditory system.

The Auditory Theory of Speech Perception

- Evidence: Context effects on perception can be induced with non-speech sounds and it works in birds too!
- Recall al vs. ar context effect:
  - /al/ + ?
    ga? da?
  - /ar/ + ?

/al/ — tongue forward, similar to /d/ gesture
/ar/ — tongue back, similar to /g/ gesture

/al/ — tongue forward, similar to /d/ gesture
/ar/ — tongue back, similar to /g/ gesture

Due to coarticulation:
- a /g/ in an /al/ context will sound a little more /d/-like because the tongue will be moving from a forward /d/-like location
- a /d/ in an /ar/ context will sound a little more /g/-like because the tongue will be moving from a back /g/-like location

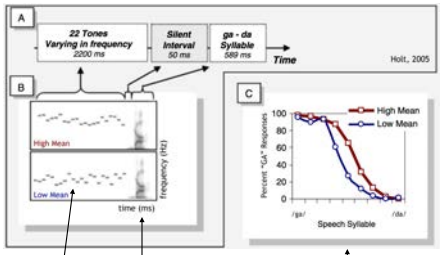
Our perception of an ambiguous [ga/da] ***compensates*** *for this coarticulation fact.*

- Maybe this coarticulation perception compensation is enabled by our motor experience in coarticulating these sounds

(Consistent with MToSP)

But no…

- It works for non-speech contexts too



context of higher (al-like) or lower (ar-like) tones

da/ga sound

category boundaries shift accordingly

And it works for quails too

Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)

Andrew J. Lotto,[a)] Keith R. Kluender, and Lori L. Holt

What's it mean?

- This means that general acoustic mechanisms can explain facts about how we perceive speech sounds, even those caused by gestural factors like coarticulation
- Ability to coarticulate is not a prerequisite for these effects

Exemplar theories

- "Standard theories" both auditory and motor typically assume that we perceive abstract phonemic categories
- Details of particular talkers or instances of hearing a sound are discarded or corrected for by the perceptual system
- Exemplar theories say we code, store, and use the variation that we hear

Exemplar theories

Attempt to account for evidence that listeners remember details relating to specific episodes

- Listeners are more accurate at recognizing that they've heard a word before if it is repeated by the same talker and at the same speaking rate (Bradlow et al., 1999)

Exemplar theories

- Each stimulus (e.g. a word) leaves a unique trace in memory
- When a new stimulus is presented, these memory traces are activated in proportion to their similarity to the stimulus
- Conscious experience of perception occurs as a result of the combined activation of these previously stored exemplars

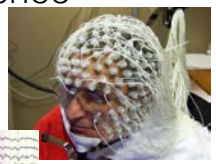
(Hidden problem)

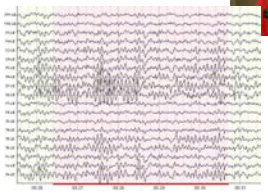
Either-Or?

- Exemplar theories and abstractionist theories can co-exist.
- Remember: listeners can categorize speech sounds as well as accurately rate similarity to a "prototype"
- Maybe a hybrid theory is the right way to think about it.

Some ideas from neuroscience

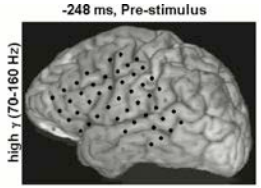
- Neural systems oscillate
- Oscillations are related to behavior





Some ideas from neuroscience

- Oscillation power varies across the brain during task performance (here, "verb generation")

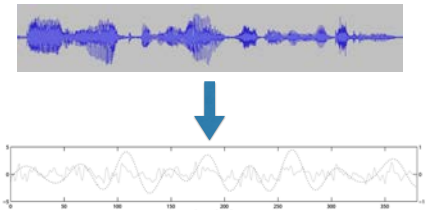


Listen to an object name, e.g., "pencil"

Say a related action name, e.g., "write"

Some ideas from neuroscience

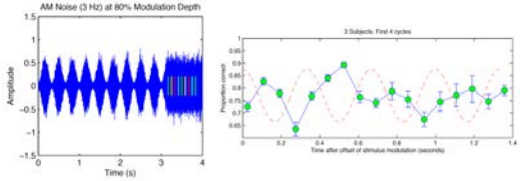
- Oscillations synchronize to speech stimuli



- And may in turn affect the perception of speech

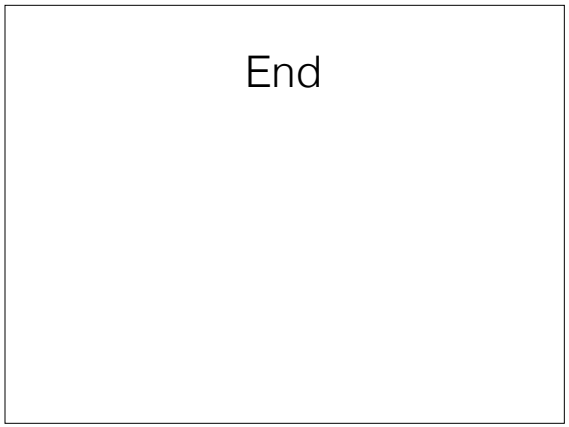
Some ideas from neuroscience

- Non-speech example: acoustic rhythmic entrainment



Some ideas from neuroscience

- Neural oscillation entrainment or phase-locking is an intense area of interest currently.
- Lot's of suggestive evidence, but still preliminary.


End