

# Speech Perception

---

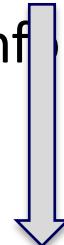
## Gesturalist Theories

## **Pam's office hours (105 Olson)**

Monday, July 1:                   3-4 pm  
Wednesday, July 3:               3-4 pm

or send me an email message ([beddor@umich.edu](mailto:beddor@umich.edu)) to arrange another time

- Keeping in mind two foundational theoretical questions
  - What is the role of (coarticulatory) variation in perception?
  - What is the nature of the information listeners recover from the variable signal?



What answer(s) have we come up with so far?

- Keeping in mind two foundational theoretical questions
  - What is the role of (coarticulatory) variation in perception?
  - What is the nature of the information listeners recover from the variable signal?
- Two Gesturalist theories
  - Motor Theory (Liberman & Mattingly 1985; Liberman & Whalen 2000)
  - Direct Realism (Fowler 1986, 1996; Best 1995)

For each:

- What are the basic claims?
- What is the main evidence – for and against (e.g., categorical perception, compensation for coarticulation)?
- What is (or isn't) controversial?

**Is phonetic variation noise or is it perceptually useful?**

Coarticulatory variation:

- is *noise* that may interfere with processing
- lawful, useful variation that *facilitates* perception

**What do listeners recover from the acoustic signal?**

Listeners recover:

- auditory information
- gestural information



***Why should we care – does it matter?***

**Is phonetic variation noise or is it perceptually useful?**

Coarticulatory variation:

- is *noise* that may interfere with processing
- lawful, useful variation that *facilitates* perception

**What do listeners recover from the acoustic signal?**

Listeners recover:

- auditory information
- gestural information

## MOTOR THEORY

From Liberman & Mattingly (1985:6):

"Speech perception is not to be explained by principles that apply to perception of sounds in general, but must rather be seen as a specialization for phonetic gestures. Incorporating a biologically based link between perception and production, this specialization prevents listeners from hearing the signal as an ordinary sound, but enables them to use the systematic, yet special, relation between signal and gesture to perceive the gesture. The relation is systematic because it results from lawful dependencies among gestures, articulator movements, vocal-tract shapes, and signal. It is special because it occurs only in speech."

## Motor Theory: three main claims

1. Listeners recover speech gestures:

Perceiving speech is perceiving vocal tract actions

2. Speech is special

"[P]honetic information is perceived in a biologically distinct system, a 'module' specialized to detect the intended gestures of the speaker that are the basis for phonetic categories." (Liberman & Mattingly 1985: 1)

3. Speech perception recruits the motor system

"... the phonetic mode, and the perception-production link it incorporates, are innately specified." (1985:24)

"... for language, perception and production are only different sides of the same coin." (1985:30)

## PERCEPTION OF GESTURES

Why *gestures*?

Lack of correspondence between acoustic cue and percept:

"... it appears that perception mirrors articulation more closely than sound"  
(Liberman et al. 1967:453 in *Psychological Review*, vol. 74)

"... acoustic patterns of synthetic speech had to be modified if an invariant phonetic perception was to be produced across different contexts" (L&M 1985:2)

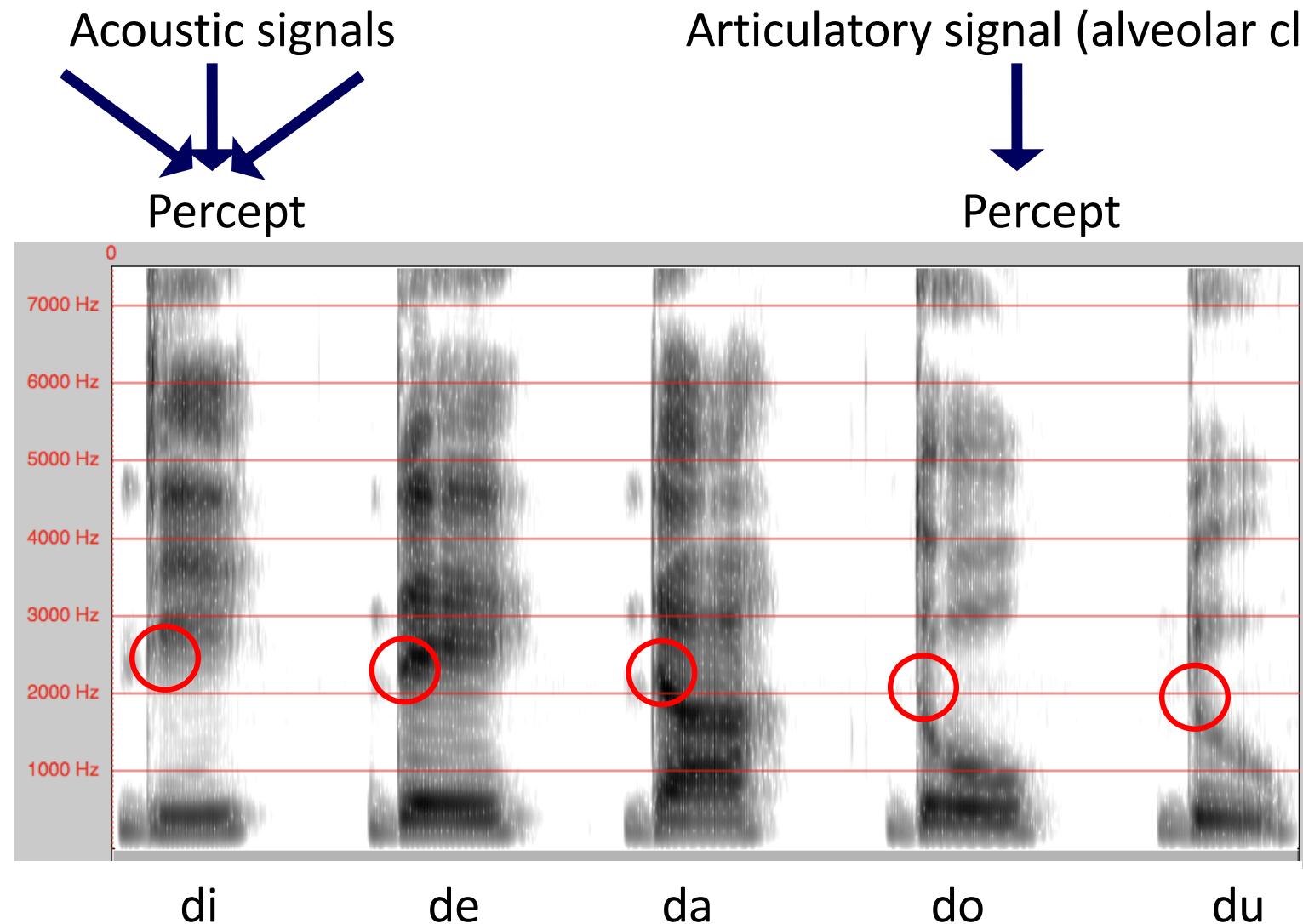
## PERCEPTION OF GESTURES

Why *gestures*?

Lack of correspondence between acoustic cue and percept:

“... it appears that perception mirrors articulation more closely than sound”  
(Liberman et al. 1967:453 in *Psychological Review*, vol. 74)

“... acoustic patterns of synthetic speech had to be modified if an invariant phonetic perception was to be produced across different contexts” (L&M 1985:2)



Not this simple!

- Acoustic invariants may lie elsewhere (e.g., Stevens 1989, *J. Phonetics* 17, 3-46)
- Articulatory signal (in no small part to coarticulation): not invariant

## PERCEPTION OF GESTURES

Why perception of *gestures*?

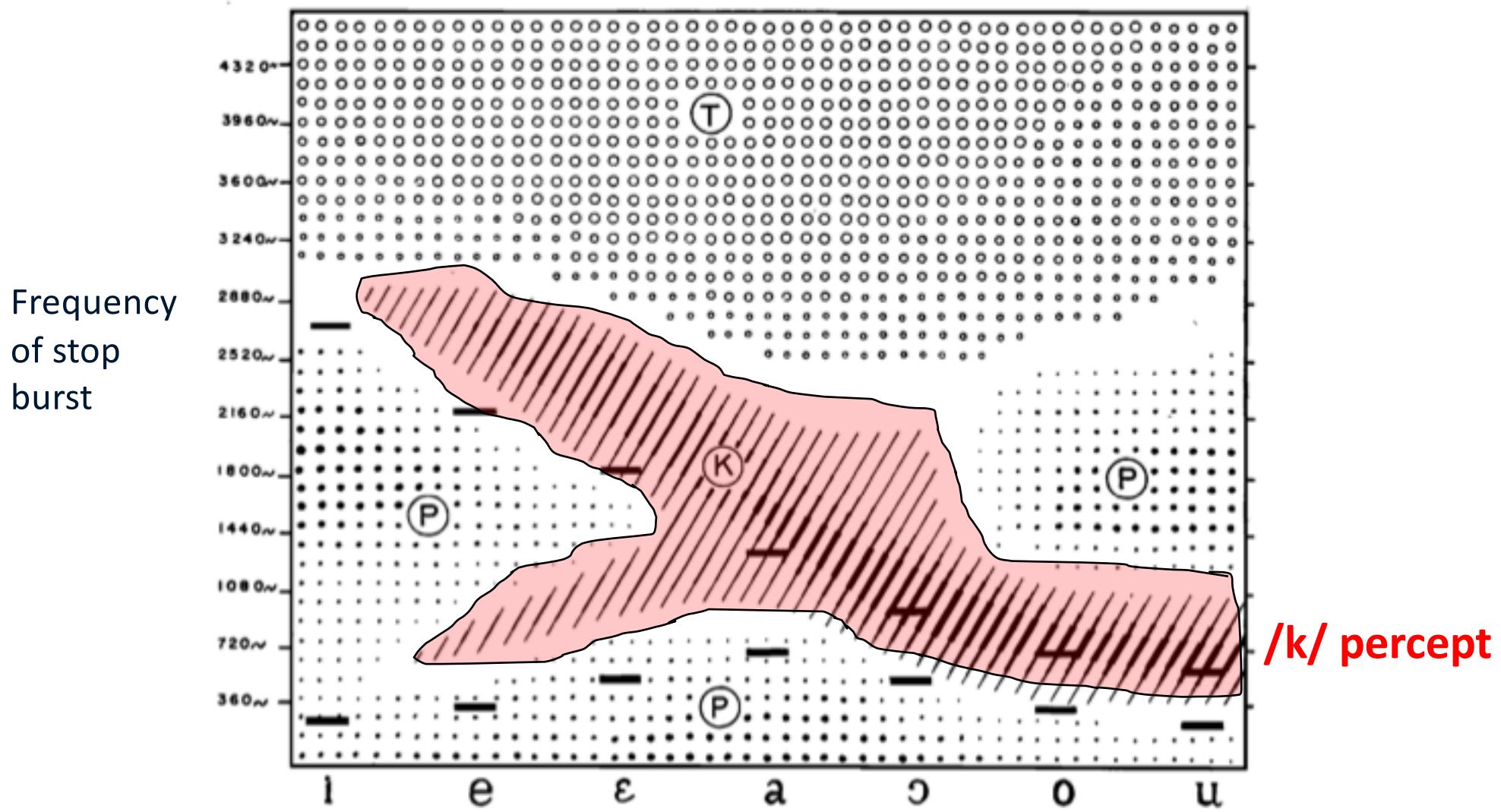
Lack of correspondence between acoustic cue and perceived phoneme:

"... it appears that perception mirrors articulation more closely than sound"  
(Liberman et al. 1967:453 in *Psychological Review*, vol. 74)

"... acoustic patterns of synthetic speech had to be modified if an invariant phonetic perception was to be produced across different contexts" (L&M 1985:2)



## PERCEPTION OF GESTURES



From Cooper et al. 1952, *J. Acoustical Soc. of America*

## PERCEPTION OF GESTURES

*Why gestures?*

- Early motivation: Lack of correspondence between acoustic cue and perceive phoneme
- As theory developed: Follows from postulation of biologically based link between perception and production

## SPEECH PERCEPTION OCCURS IN A SPECIALIZED PHONETIC MODULE

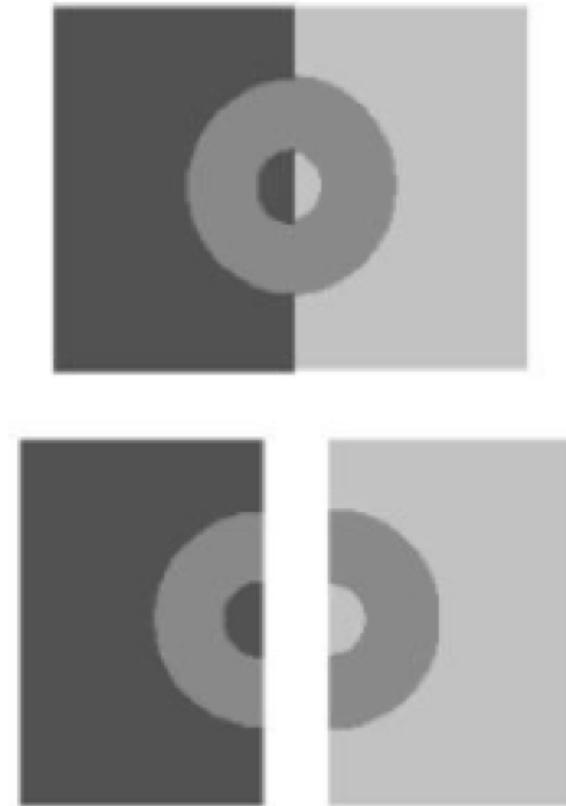
Specialized module (e.g., Fodor's *The Modularity of Mind*, 1983):

a neural system that performs specific task (e.g., sound localization, color perception, and—arguably!—speech perception)

Such modules

- have fast computations (faster than if performed by more general cognitive process)
- are informationally encapsulated (immune to other sources of information)

Example of **information encapsulation**: visual perceptual illusions



Perceptual system appears to be cut off from cognitive knowledge

McGurk effect is similar for speech perception: even though perceiver knows that visual and auditory stimuli don't match, the illusion remains.

<http://www.youtube.com/watch?v=G-IN8vWm3m0>

## SPEECH PERCEPTION OCCURS IN A SPECIALIZED PHONETIC MODULE

Specialized module (e.g., Fodor's *The Modularity of Mind*, 1983):

a neural system that performs specific task (e.g., sound localization, color perception, and (arguably) speech perception)

Such modules

- have fast computations (faster than if performed by more general cognitive process)
- are informationally encapsulated (immune to other sources of information)
- operate mandatorily (e.g., McGurk effect)
- process only domain-relevant information

Phonetic module of Motor Theory is argued to be

- responsible for (innately linked) speech perception and production
- part of broader specialization for language

### Evidence offered in support of Motor Theory

Start with two of the claims:

1. Listeners recover speech gestures
2. Speech perception is achieved through a special system (separate from ordinary audition)

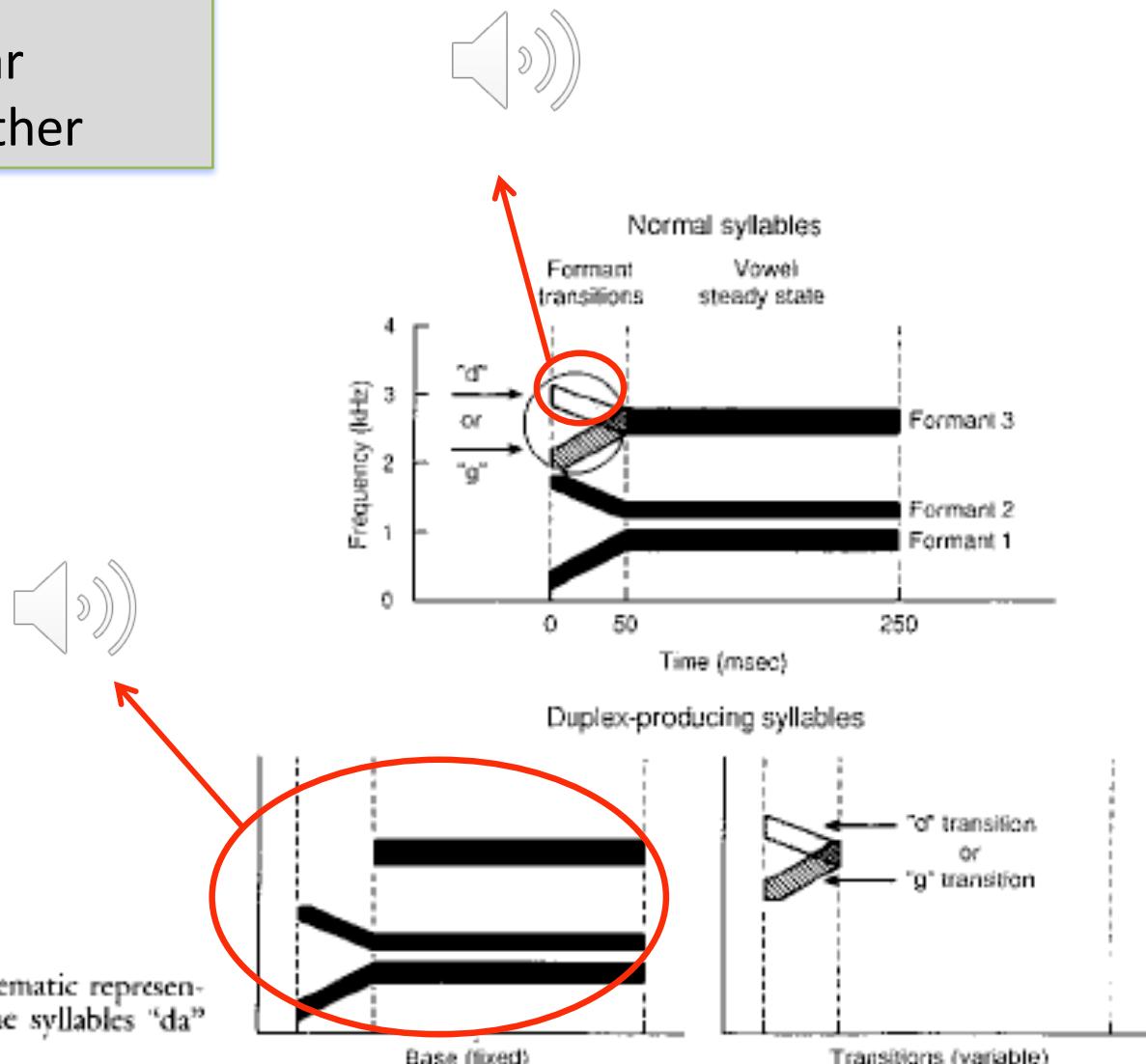
## Evidence offered in support of Motor Theory

- Listeners integrate cross-modal information for speech gestures (McGurk effect; McGurk & McDonald 1976)
- Duplex perception

## DUPLEX PERCEPTION

Dichotic listening paradigm:

- short F3 transition in one ear
- rest of syllable ("base") in other



Schematic from Whalen & Liberman (1987, Science 237, 169-171)

## DUPLEX PERCEPTION

Dichotic listening paradigm:

- short F3 transition in one ear
- rest of syllable ("base") in other

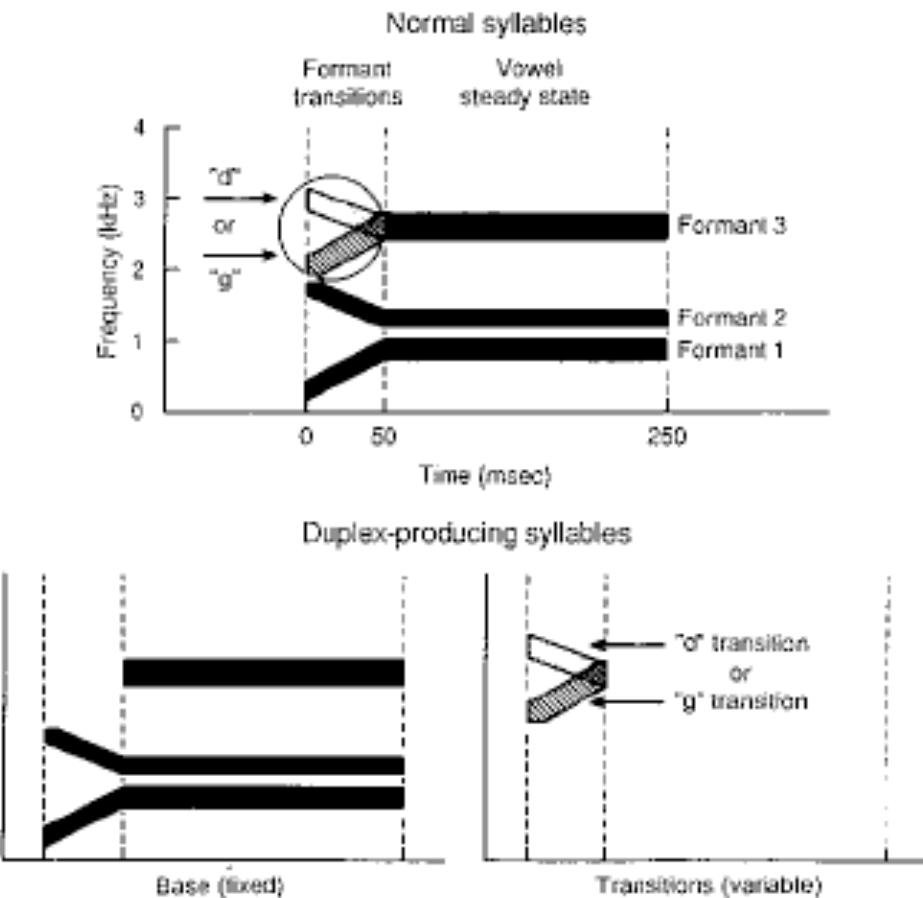
Result:

- Percepts: fused percept in ear that gets base; chirp in other ear
- Perception is duplex not triplex. Speech percept is perceived to be entirely in one ear: same signal elicits both phonetic and non-speech percept. Taken by MT as evidence of two distinct modules.

**Fig. 1.** Schematic representation of the syllables "da" and "ga."



Stereo presentation:  
F3 transition to one  
ear and "base" to  
other



Schematic from Whalen & Liberman (1987, Science 237, 169-171)

## Evidence offered in support of Motor Theory

- Listeners integrate cross-modal information for speech gestures (McGurk effect; McGurk & McDonald 1976)
- Duplex perception
- Perception of sinewave speech (lacks traditionally identified acoustic cues)

Listeners perform differently on perceptual tasks depending on whether hear as speech (*Best et al. 1981, Perception & Psychophysics 29*)

## Evidence offered in support of Motor Theory

- Listeners integrate cross-modal information for speech gestures (McGurk effect; McGurk & McDonald 1976)
- Duplex perception
- Perception of sinewave speech (lacks traditionally identified acoustic cues)  
Listeners perform differently on perceptually tasks depending on whether hear as speech (*Best et al. 1981, Perception & Psychophysics 29*)
- Parsing of acoustic signal along gestural lines (compensation for coarticulation)

What about categorical perception?

## Evidence offered in support of Motor Theory

- Listeners integrate cross-modal information for speech gestures (McGurk effect; McGurk & McDonald 1976)
- Duplex perception
- Perception of sinewave speech (lacks traditionally identified acoustic cues)
- Parsing of acoustic signal along gestural lines (compensation for coarticulation)

These (and other) findings:

- Far from definitive regarding gestural perception
- Specialized system, separate from ordinary audition: very controversial

## Does speech perception recruit the motor system?

Galantucci, Fowler, & Turvey. 2006 (Psychonomic Bulletin and Review 13.3, 361-377)  
Whalen. To appear (Oxford Research Encyclopedias: Linguistics)

Motor Theory: Both production and perception systems are activated when either speaking or listening

Evidence consistent with this claim continues to accumulate

Whalen 2019:

“... the basic insight that speech perception involves active participation of neural underpinnings of the motor realization of speech has substantial support” (p. 21)

# Motor Theory: Does speech perception recruit the motor system?

Example from behavioral studies: **Bruderer et al. 2015** (Proc. Natl Ac. of Sci 112)

- 6-month olds learning English can reliably discriminate (Hindi) dental-retroflex contrast
- However, when hear these sounds while sucking on a teether that blocks tongue tip movement: dental-retroflex contrast no longer discriminated

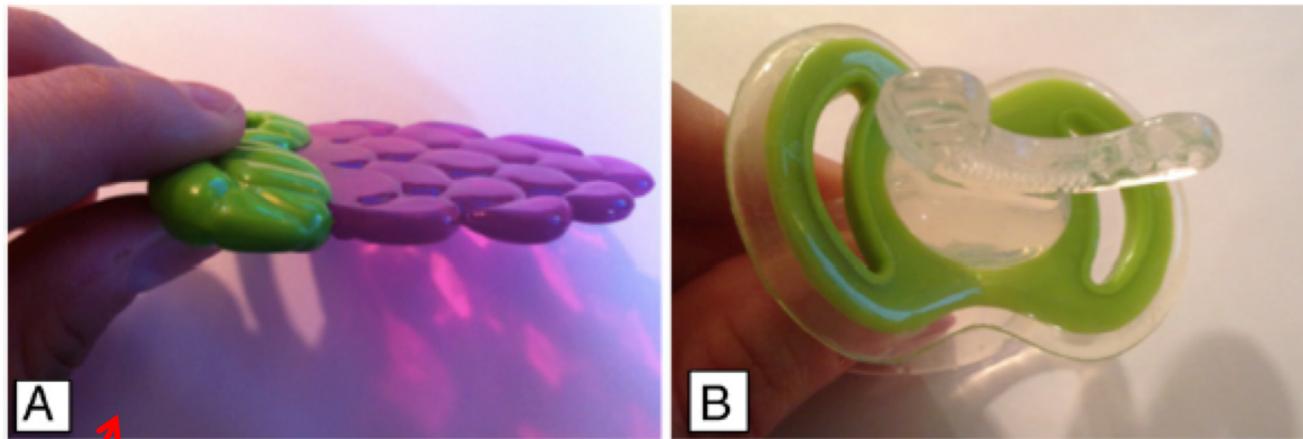


Fig. 1. Teether images. (A) Flat teether. (B) Gummy teether.

Blocks tongue  
tip movement

If articulation influences speech perception, should  
impede infants' discrimination of dental-retroflex contrast.

# Motor Theory: Does speech perception recruit the motor system?

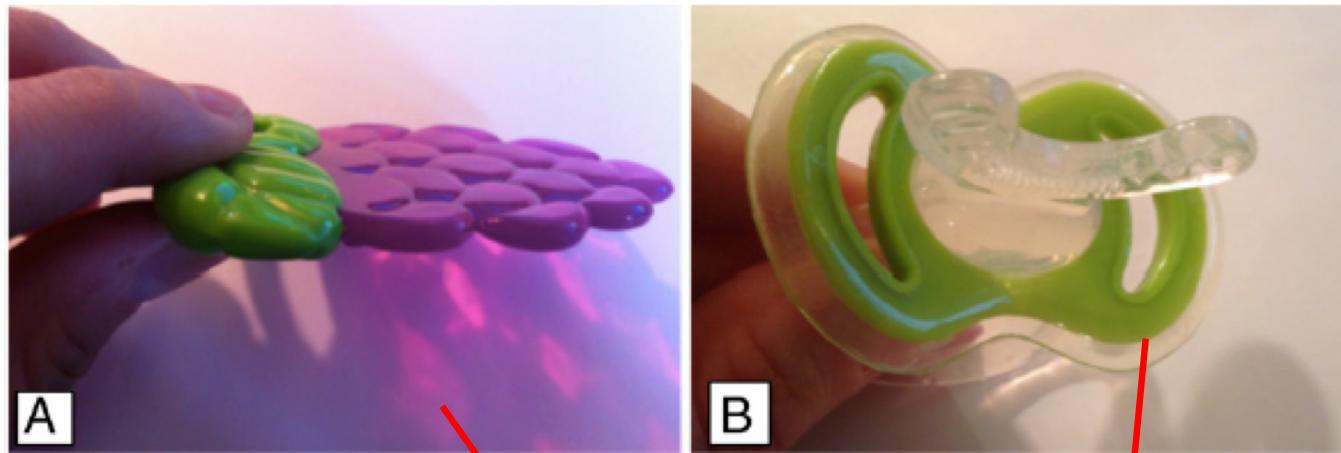


Fig. 1. Teether images. (A) Flat teether. (B) Gummy teether.

No teether

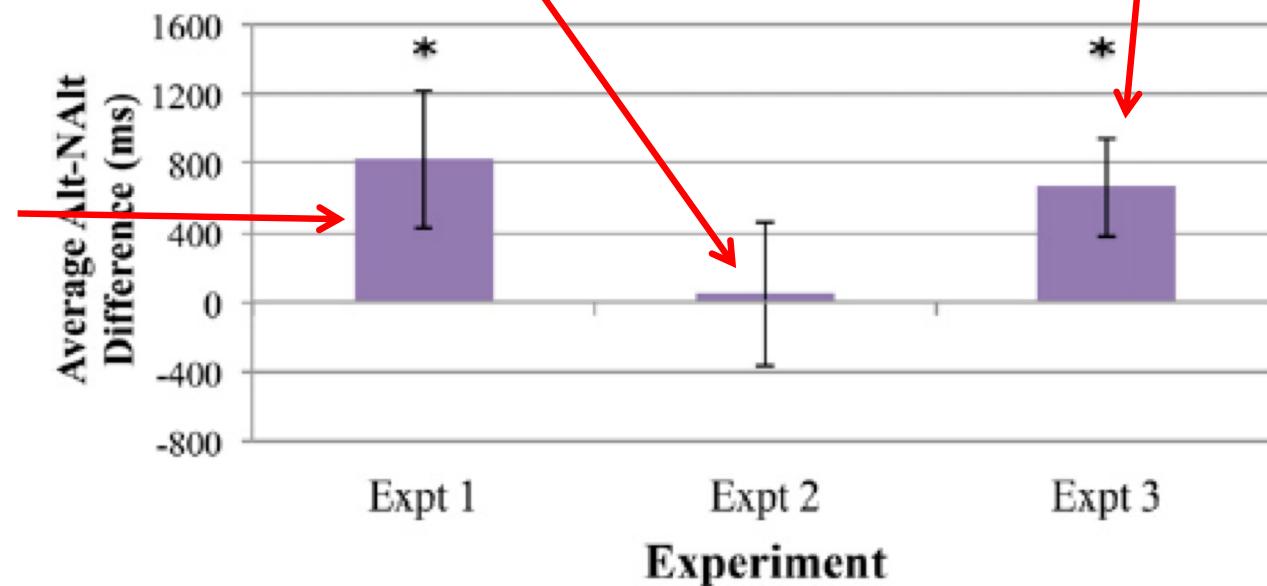


Fig. 5. Alt-NAlt difference score averages. Average difference in looking time between Alt and NAlt trials for each experiment (in ms). Scores greater than zero indicate an overall Alt > NAlt preference. Error bars denote SEM difference, and an asterisk indicates a significant difference (from zero), as reflected in the individual ANOVAs.

## Motor Theory: Does speech perception recruit the motor system?

Example from electrocorticography: Glanz et al. 2018 (Nature: Scientific Reports  
doi: 10.1038/s41598-018-26801-x)

- Placed electrodes on brains of preoperative epilepsy patients, to record cortical activity in real-life communicative interactions
- Provides temporal and spatial precision
- Mouth motor region: activated during both speech production and perception (but not during non-speech noise perception)

## Direct Realism (Fowler 1986, 1996; Best 1995)

- Direct Realism is embedded in J. Gibson's more general theory of perception
  - Perceivers gain direct information from world around them; perceptual systems have function of perceiving real-word causes of structure in media (light, air, etc.)
  - All perception involves direct recovery of distal source of event being perceived
- Speech perception is not different from other types of auditory perception or, more generally, from other types of perception

## Motor Theory

- Listeners perceive (intended) gestures
- Speech is perceived in a specialized speech module
- Speech perception recruits the motor system

## Direct Realism

- Listeners perceive (actual) gestures
- Speech perception is not special
- No proposed role of motor system in perception

## Motor Theory

- Listeners perceive (intended) gestures
- Speech is perceived in a specialized speech module
- Speech perception recruits the motor system

uncoarticulated

## Direct Realism

- Listeners perceive (actual) gestures
- Speech perception is not special
- No proposed role of motor system in perception

vocal tract gestures have invariant properties

## Motor Theory

- Listeners perceive (intended) gestures
- Speech is perceived in a specialized speech module
- Speech perception recruits the motor system

## Direct Realism

- Listeners perceive (actual) gestures
- Speech perception is not special
- No proposed role of motor system in perception

For both: perception of gestures achieves parity between forms of speaking and listening

But source of parity is different:

- Motor theory: perceiving and producing part of same system
- Direct realism: parity due to perceivers being realists (perceive source of event)

## Motor Theory

- Listeners perceive (intended) gestures
- Speech is perceived in a specialized speech module
- Speech perception recruits the motor system

## Direct Realism

- Listeners perceive (actual) gestures
- Speech perception is not special
- No proposed role of motor system in perception

### Direct Realism and Articulatory Phonology

- Gestures are units of production, units of perception, and phonological forms
- Articulatory Phonology (specifically, task dynamics model):  
Gestures directly structure the acoustic signal
- Direct Realism:  
Acoustic structure has sufficient information to specify gestures

## Evidence offered in support of Direct Realism

Recovery of gestures: evidence overlaps with that for Motor Theory

- McGurk effect: listeners perceive gestures, which are specified not only acoustically, but also optically and haptically (Fowler & Dekle 1991, J. Exp. Psych. Hum. Perc. Perf. 17, 816-828)
- Shadowing response times (Fowler et al. 2003, J. Memory & Lang. 49, 396-413)
  - Reaction times: typically, simple (constant response) < choice (variable)
  - Speech shadowing RTs: simple, choice responses nearly identical
- Parsing the signal along gestural lines
  - Compensation for coarticulation (both carryover and anticipatory)
    - *But:* tones, quail (Lotto, Kluender et al. findings)
  - Companion finding: listeners closely track, and use, coarticulation

## Motor Theory

- Listeners perceive (intended) gestures
- Speech is perceived in a specialized speech module
- Speech perception recruits the motor system

especially controversial

## Direct Realism

- Listeners perceive (actual) gestures
- Speech perception is not special
- No proposed role of motor system in perception.

Can acoustics-to-articulation “inversion problem” be solved?

## Next time: Perceivers as producers of contextual variation

- Individuals differ in how they produce and perceive coarticulatory variation
- Is there a link between the two – and, as always, why might we care?
- My answer to why we care
  - Important to theories of perception
  - Important to theories of the role of listeners in sound change
- Reading: test of this link (Beddor et al. 2018)