Speech Perception

Perceiving coarticulatory variation

Pam's office hours (105 Olson)

Thursday, June 27:	3-4 pm
Monday, July 1:	3-4 pm
Wednesday, July 3:	3-4 pm

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

Traditional main question in speech perception:

How do listeners interpret the input acoustic signal as linguistic forms?

Challenge in answering this question: ***phonetic variation***

Early work: Categorical perception interpreted as evidence that listeners ignored subphonemic variation

Reinforced by findings that category boundaries (and discrimination peaks and valleys) are language-specific

Now know:

- Listeners are good categorizers of variable input (e.g., “I heard [b]” or “I heard 'bear'”)
- Categorization ≠ ignore phonetic (subphonemic) detail

Now ***also*** know:

- Phonetic variation is not only tolerated but is perceptually useful (e.g., in L1 and L2 acquisition and learning new speech patterns)

Khalil Iskarous' Dynamical Systems workshop at 2015 Institute:

- There is structure that is induced only when noise / variation is present.

Coarticulatory variation is an excellent example of this — acoustic "signature" of vocal tract actions

Roadmap for today

- Reminder of 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?

- Compensation for coarticulation (at the intersection of these two questions)
  - What it tells us about perceivers
  - What is tells us about the nature of theorizing and argumentation in speech perception research

Intersection of two major theoretical issues

Is phonetic variation noise or is it perceptually useful?	**What do listeners recover from the acoustic signal?**
Coarticulatory variation: • is *noise* that may interfere with processing • lawful, useful variation that *facilitates* perception	Listeners recover: • gestural information • auditory information

Does it have to be one or the other? Why not both?

- For both questions: both answers could be correct
- Most (but not all) theories have tended to stake out one position or other
- Has advantage of pushing the envelope (strong tests of strong positions)

Compensation for coarticulation

Try this out:

You'll hear two pairs of

bed – bend

In which pair do the vowels sound more similar?

- 1st pair: vowels acoustically distinct
  both in coarticulatorily appropriate context
- 2nd pair: vowels acoustically identical
  only one (2nd) is in coarticulatorily appropriate context
- In this type of trial, phonetically naïve listeners tend to make mistakes.

Another illustration:

sibilant + [u]: anticipatory rounding for [u] lowers sibilant frequency

Mann & Repp (1981, *Perception & Psychophysics* 28):

When listeners identify members of a /s-ʃ/ continuum, they report hearing more /s/ before /u/ than before /i/

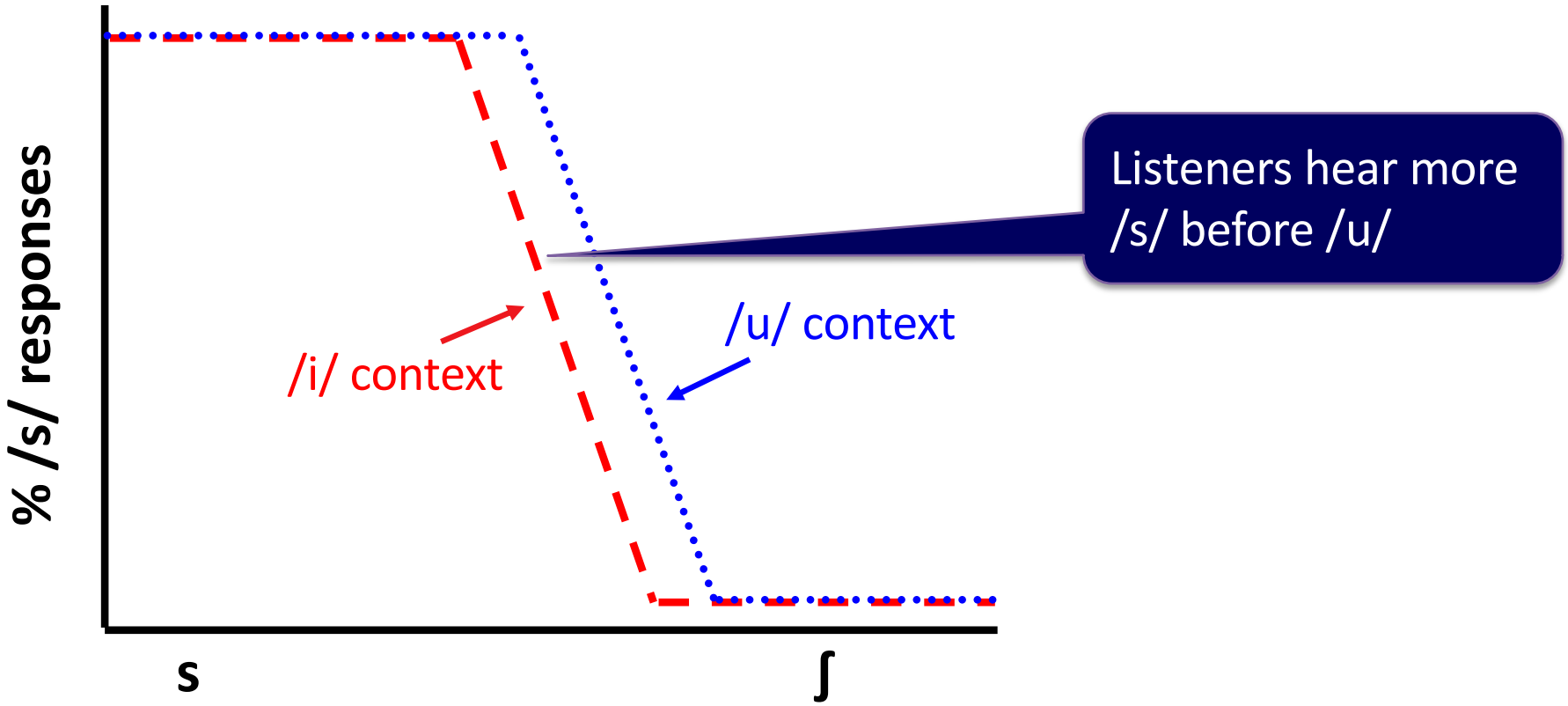
Why?

Hint: Independent of context, /s/ has higher frequency noise than /ʃ/

Perceptual Compensation for Coarticulation

Listeners perceptually "factor out" coarticulation; they attribute coarticulatory effects to their *source* rather than to the segment on which they actually occur.

A schematic figure of effects of vowel context on fricative perception:



An aside …

Kevin will show you in a couple of weeks that get comparable perceptual compensation for /s- ʃ/ differences due to talker (rather than context) variation.

Compensation for coarticulation

Let's try this for vowel-to-vowel coarticulation:

For each of these "words", do you hear [popi] or [pepi]?

1 2 3 4 5 6 7 8 9 10

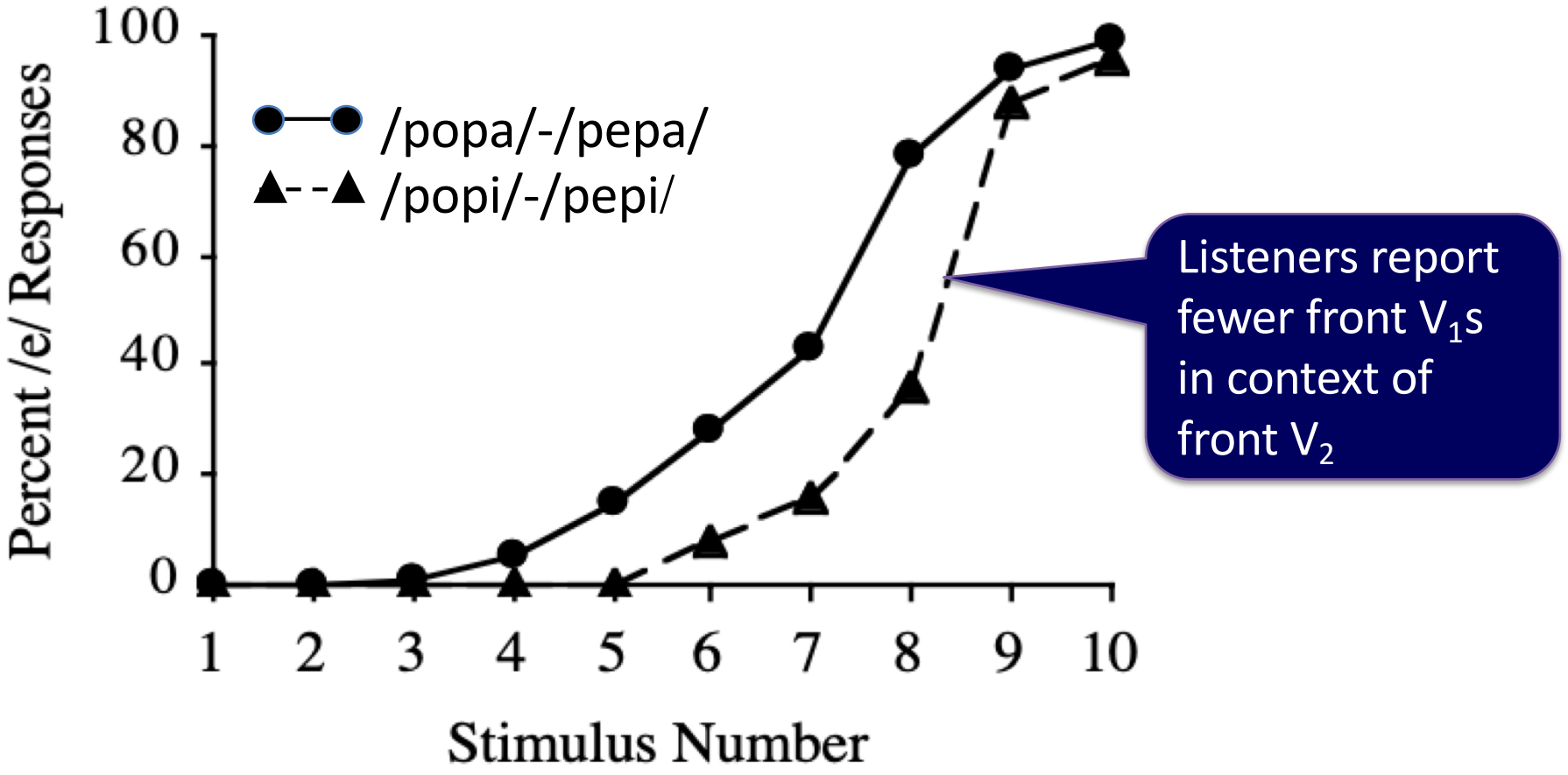
For this next set of "words", do you hear [popa] or [pepa]?

1 2 3 4 5 6 7 8 9 10

In which set did you report more [e] responses?

Why?

Results (Beddor, Harnsberger, and Lindemann, 2002, *Journal of Phonetics* 30):



From 1980-2016, several experiments were conducted on /da-ga/ continuum embedded in post-liquid contexts, /aɹ__/ and /al__/.

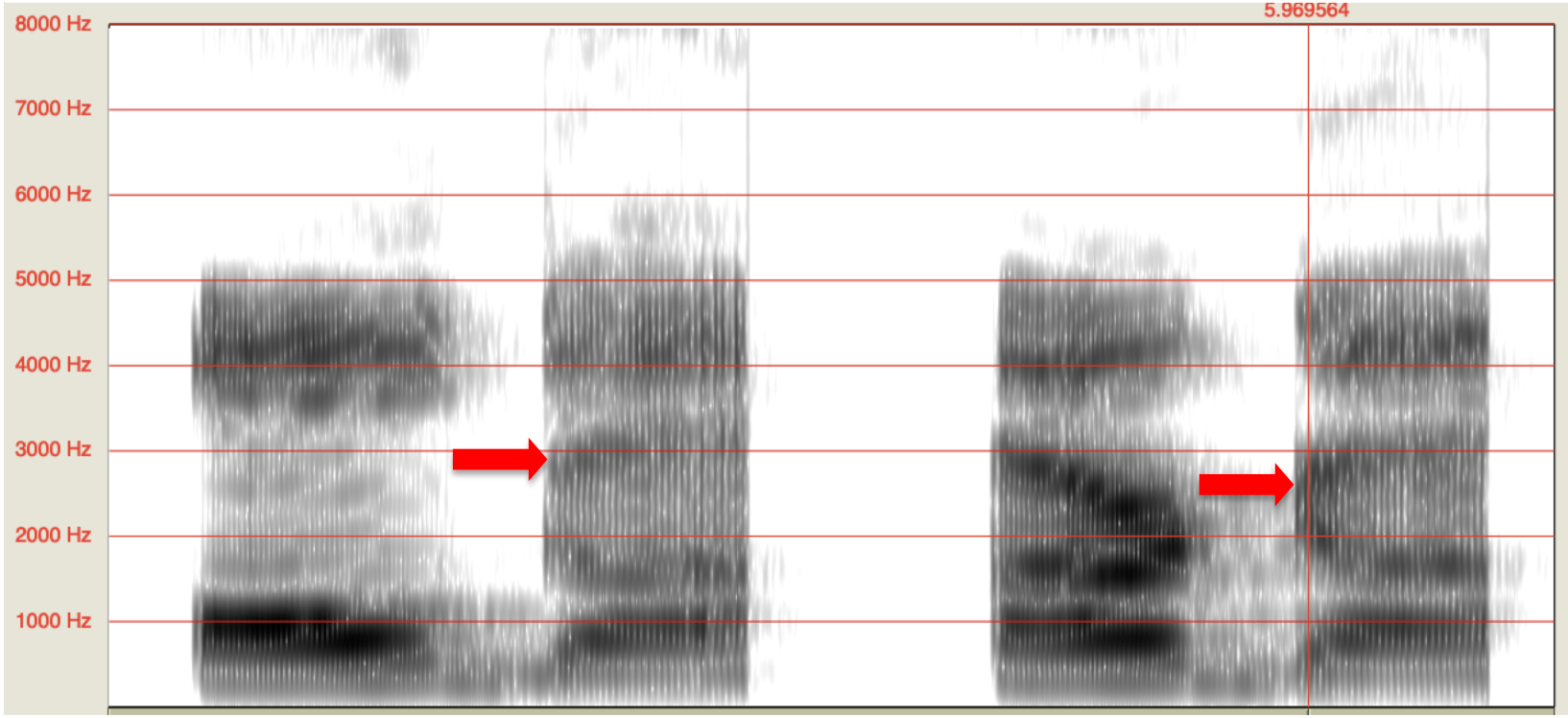
Background acoustics:

- Main acoustic difference between /ɹ/ and /l/: F3
  /ɹ/: low F3 /l/: high F3
- /g/ has lower F3 frequency than /d/ (F2/F3 "velar pinch" for /g/)

- Due to coarticulation: F3 onset of /g/ is higher after /l/ than after /ɹ/
  Acoustically, */g/ is more /d/-like after /l/*.

Due to coarticulation: F3 onset of /g/ is higher after /l/ than after /ɹ/

Acoustically, */g/ is more /d/-like after /l/*



a l g a a ɹ g a

What happens perceptually?

Perceptual task:

Members of /da-ga/ continuum (varying in F3) embedded in two contexts: /al_/ and /aɹ_/. Instructions: identify the stop.

What should happen, given that /g/ is acoustically more /d/-like after /l/?

Listeners expected to:

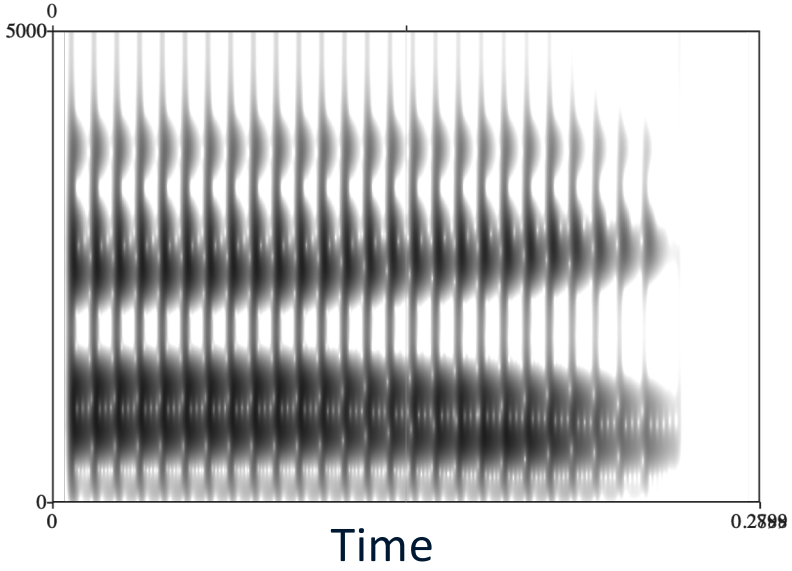
- attribute (some) F3 characteristics of transition into stop to preceding liquid (i.e., to coarticulatory source)
- hear ambiguous stops as /g/ in /al_/ context and as /d/ in /aɹ_/ context.

That is, **listeners should hear more /g/s in the /l/ context**.

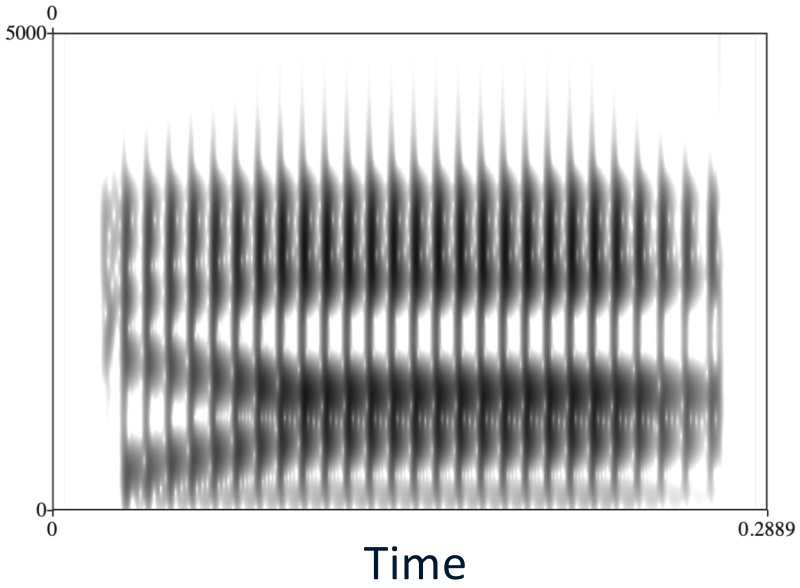
Mann (1980) *Perception & Psychophysics*

Stimulus 4 from the continuum:

[ɑl]

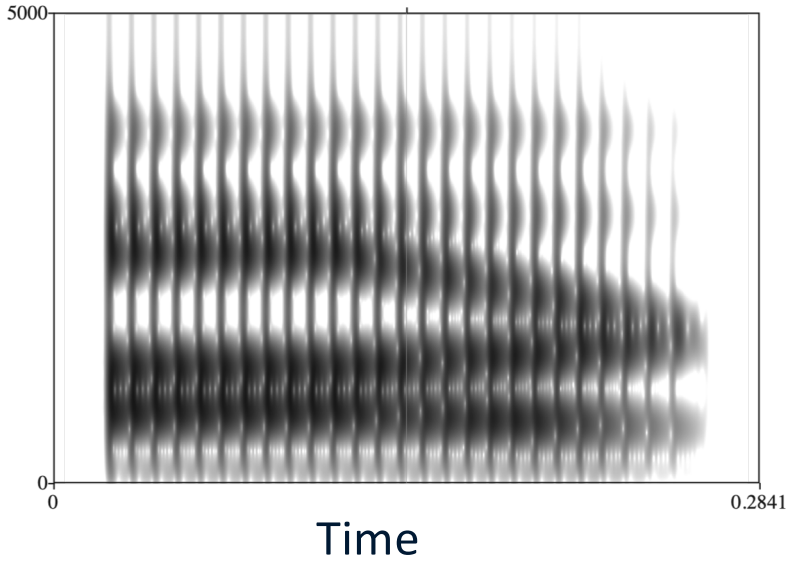


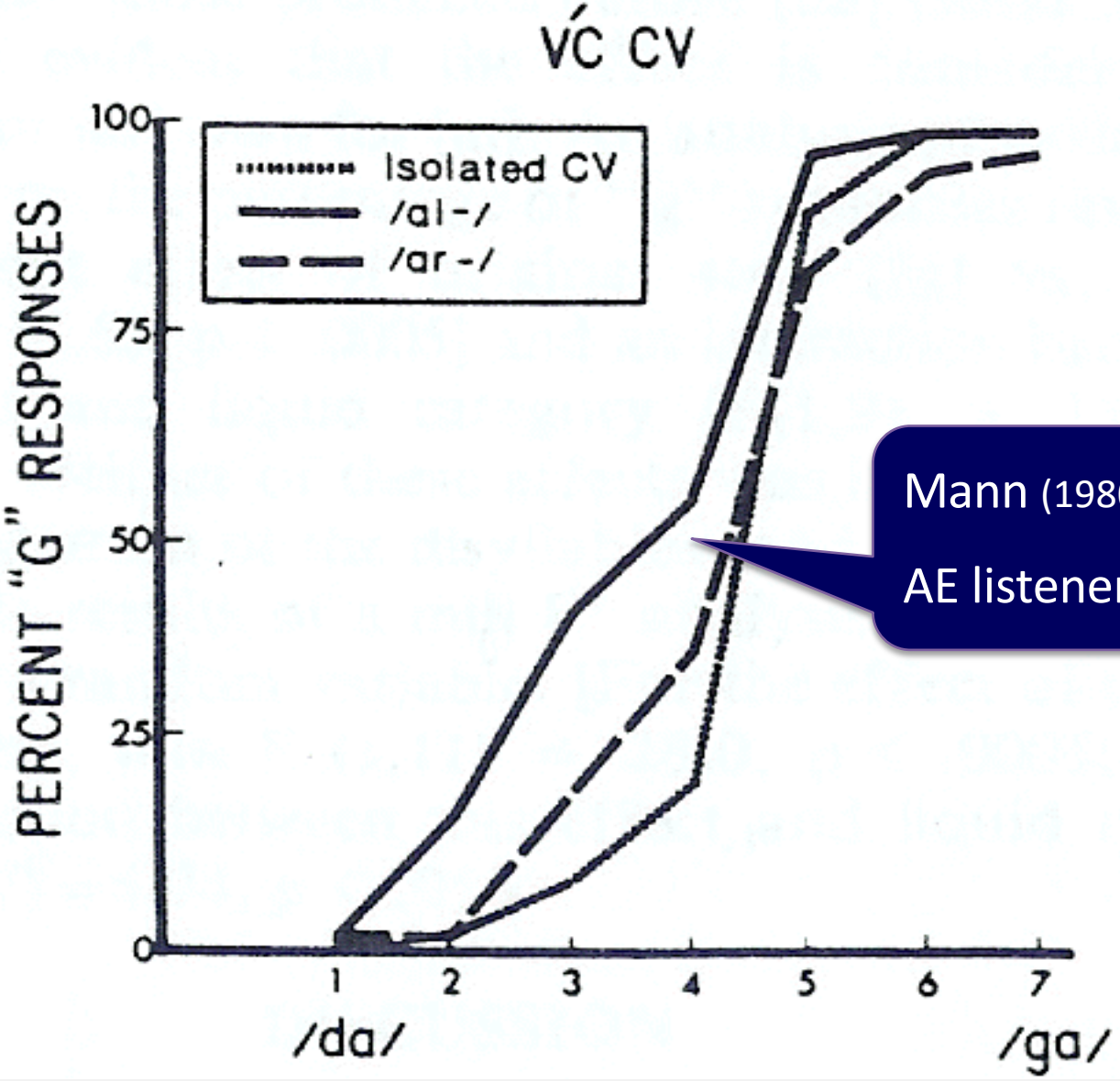
[ga]



[dɑ]

[ɑɹ]

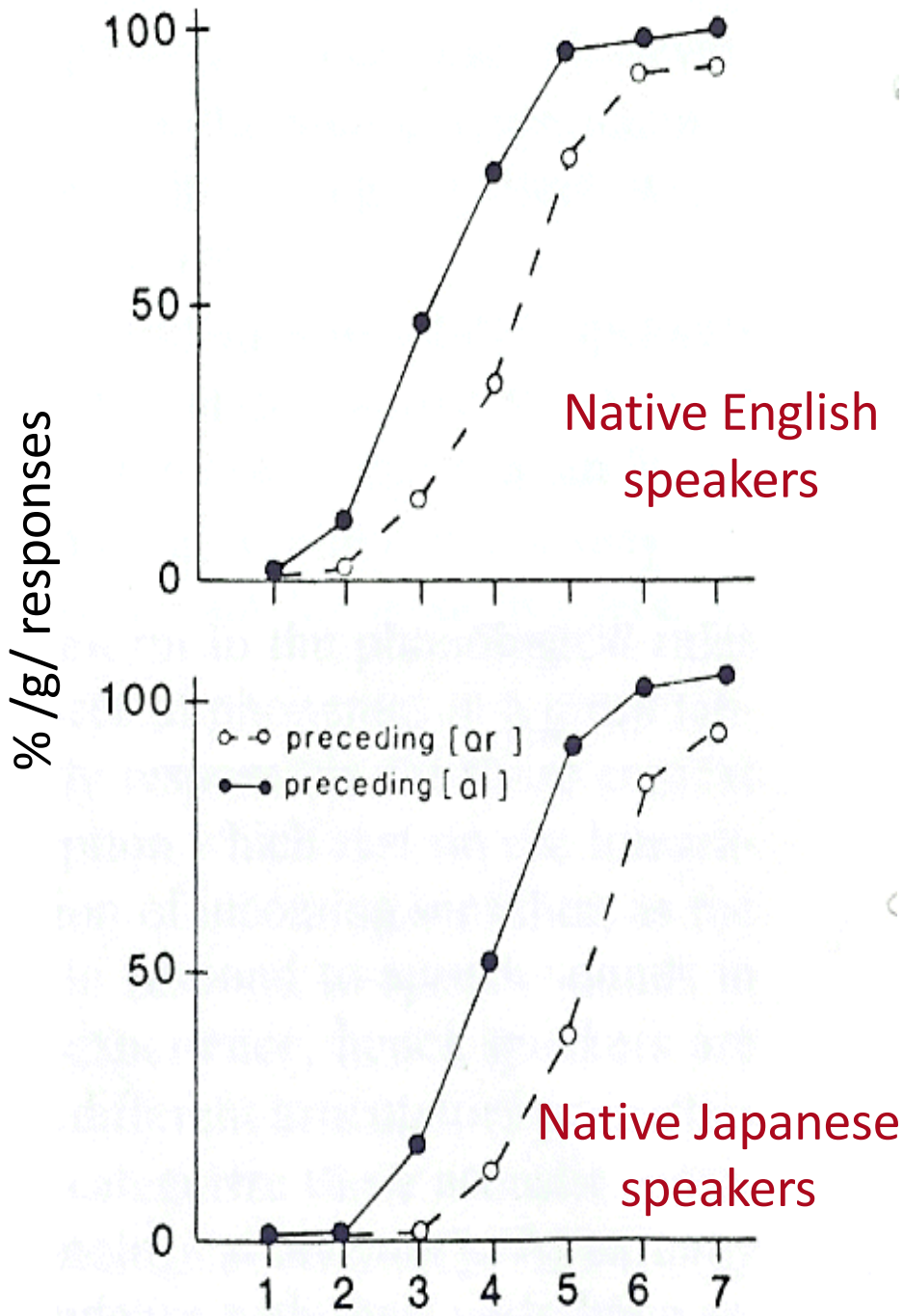




Mann (1980, *Perception & Psychophysics* 28):

AE listeners heard more /g/ after /l_/

Are these effects due to ***experience*** with the relevant coarticulatory patterns?



Not (entirely) due to experience:

Mann (1986, *Cognition* 24): Japanese-speaking listeners who don't reliably discriminate /ɹ-l/ *also* adjust for coarticulatory effects of liquid on stop

Interpreting compensation for coarticulation

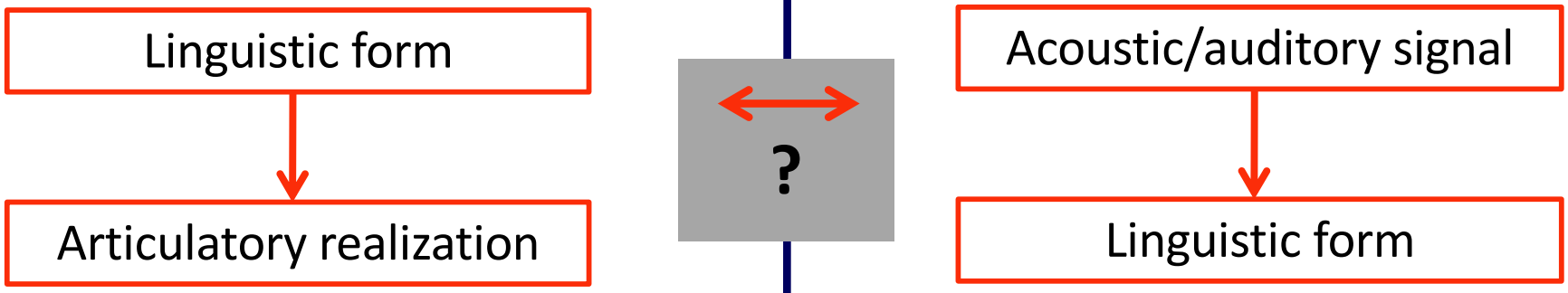
One Gesturalist Theory

Direct Realism (Fowler 1986, 1996)

- Perceivers use structure in acoustic signal as information about source of the event
- In speech, source is speaker's articulations; listeners recover vocal tract actions from acoustic signal

An aside: production-perception relation

What is the relation between speech perception and production?



Goals of theories of speech production:

- Explain how speaker's linguistic message is physiologically realized through controlled, coordinated activities of articulatory system
- Want model that predicts/derives coordinated movements that achieve linguistic goals

Goals of theories of speech perception:

- Explain how listeners map from acoustic signal to linguistic form
- Want model that predicts/derives linguistic percept from input signal (which may include cues weighted on basis of socioindexical, phonological, and lexical, etc. information)

An aside: production-perception relation

What is the relation between speech perception and production?

What are speakers controlling? What are listeners perceiving?

- For speakers, usual assumption: speakers control actions of vocal tract
- For listeners, usual assumption: object of perception is acoustic/auditory
- Fowler (2003:247, Speech production and perception, in *Handbook of Psychology, vol. 4, Experimental Psychology*, Wiley):

  "the most common type of theory of production and the most common type of theory of perception do not fit together. They have the joint members of communicative events producing actions, but perceiving acoustic structure."

An aside: production-perception relation

What is the relation between speech perception and production?

Not all theories take this approach:

- Some: domain of articulation is acoustic/auditory. Speakers control, and listeners perceive, the acoustic signal.

  E.g., rather than having articulatory goal of bilabial closure, speaker's goal is acoustic properties (burst frequency; formant transitions into flanking vowels) of sounds that result from vocal tract actions (e.g., Guenther 1995, Psychological Review 102, 594-621)

- Others: objects of speech perception are articulatory/gestural. Speakers control, and listeners perceive, vocal tract gestures.

Gestural theories and compensation for coarticulation

Gesturalist theories ***predict*** compensation for coarticulation: for both, listeners should "parse" acoustic signal along gestural lines

E.g., Direct Realism

- "[L]isteners use acoustic structure in speech utterances as information for the causal sources of that structure—namely, the phonetic gestures that produced the signal." (Fowler 2006, Perception & Psychophysics 68)
- Listeners compensate for coarticulation *because* they are tracking causal sources (in speech: gestures).

An alternative interpretation

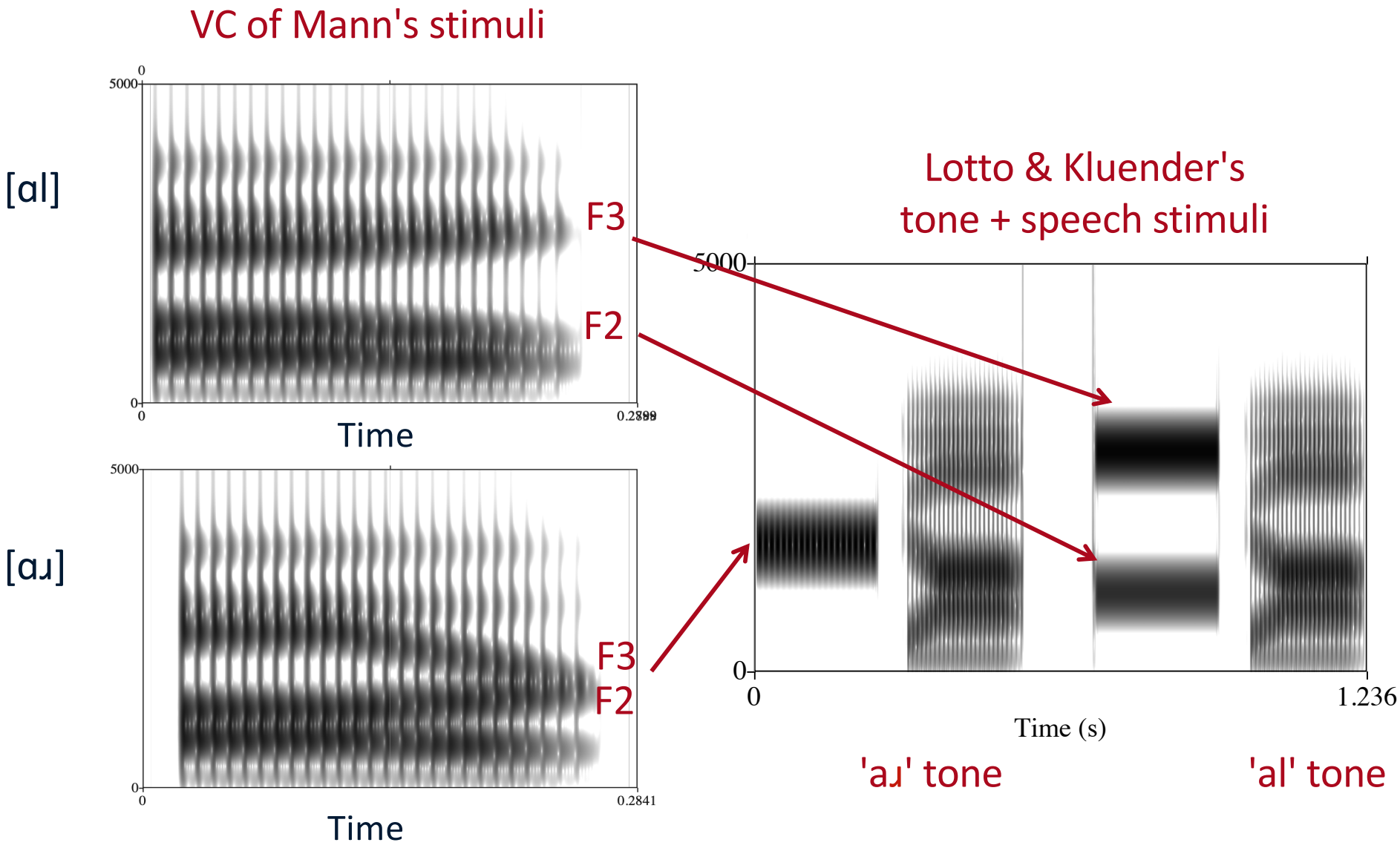
Auditory (non-gestural) theory – e.g., Lotto & Kluender (1998, *Perception & Psychophysics* 60)

- Perception is not mediated by perception of gestures, but rather relies on domain-general auditory processing and learning.
- *Apparent* compensation for coarticulation is due to auditory phenomena such as spectral contrast.

Test:

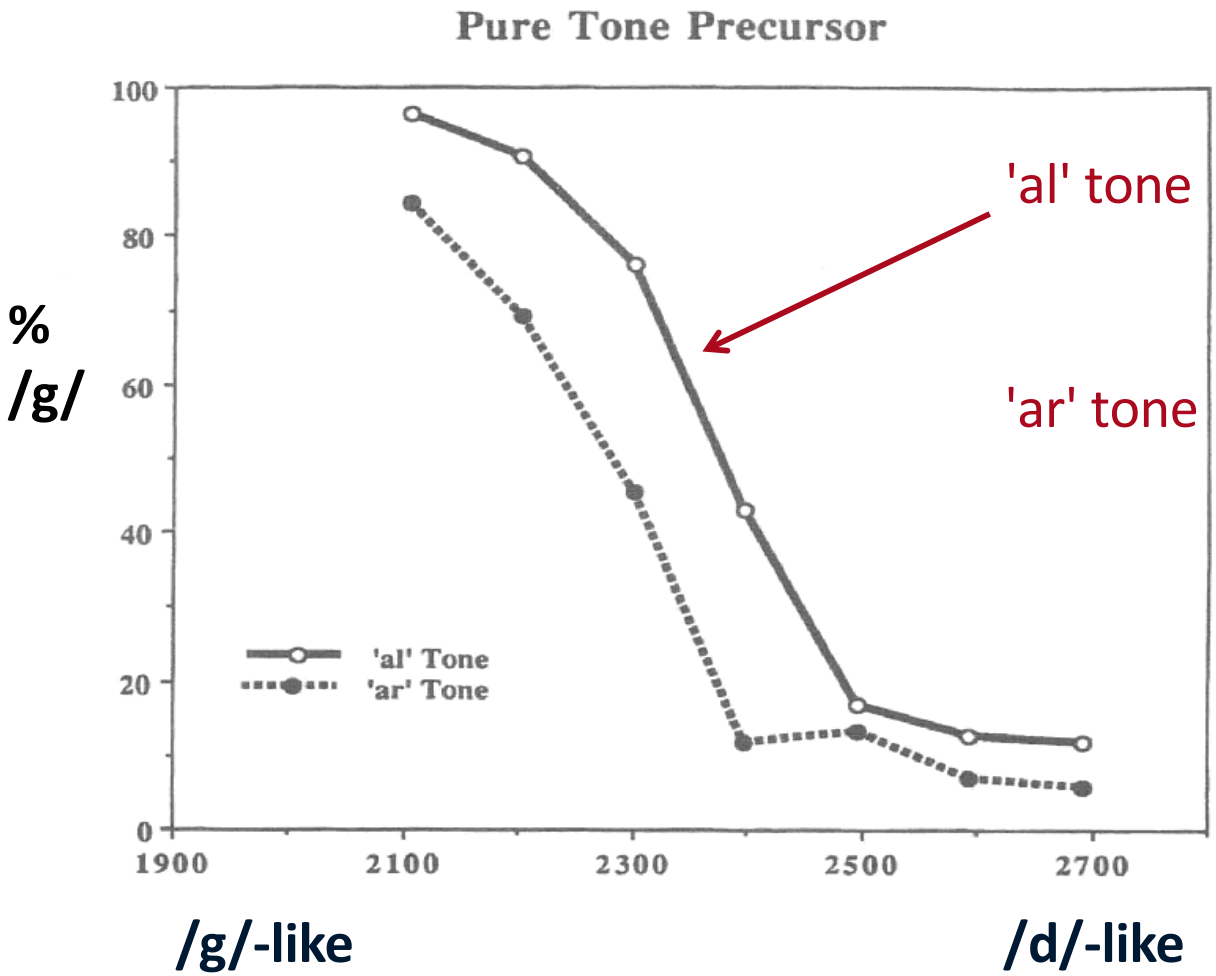
- Replace /aɹ_/ and /al_/ contexts with pure tones that model crucial acoustic properties of /ɹ/ and /l/
- *Do we still get compensation when context is not speech?*

Lotto & Kluender (1998, *Perception & Psychophysics* 60):



Lotto & Kluender

Results: as with speech, more /g/ responses after 'al' tone.

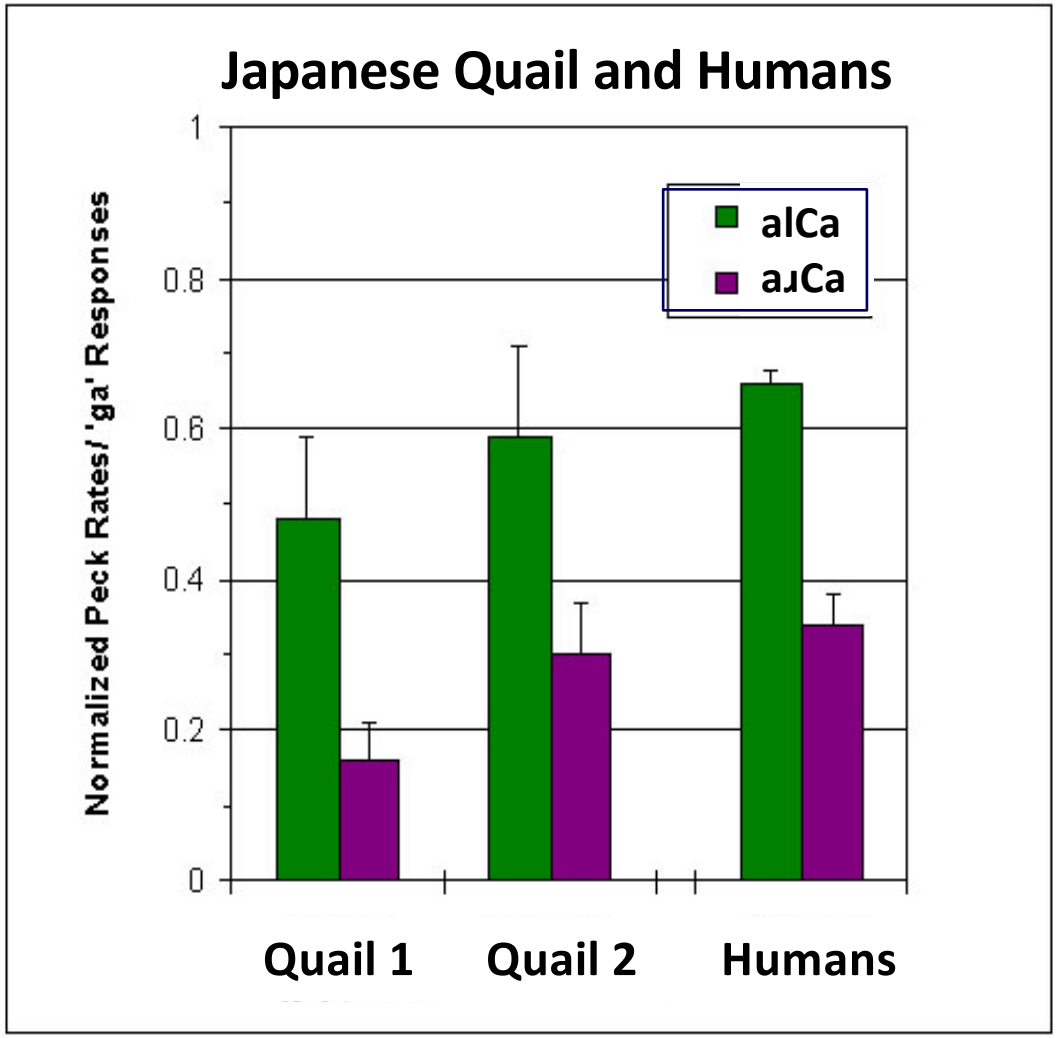


L&K's conclusion:

- Not compensation
- Auditory contrast effect: stop's F3 sounds low-frequency after high-frequency tone "F3", triggering more [g] responses

Lotto, Kluender, & Holt 1997 (*JASA* 102): Is "compensation" species-specific?

- Trained Japanese quail to peck differentially to clear cases of /da/ and /ga/
- Tested quail on ambiguous /da-ga/ stimuli in /aɹ_/ and /al_/ contexts
- Results: similar to those of humans



Same basic conclusion:

- Not compensation
- Auditory contrast effect

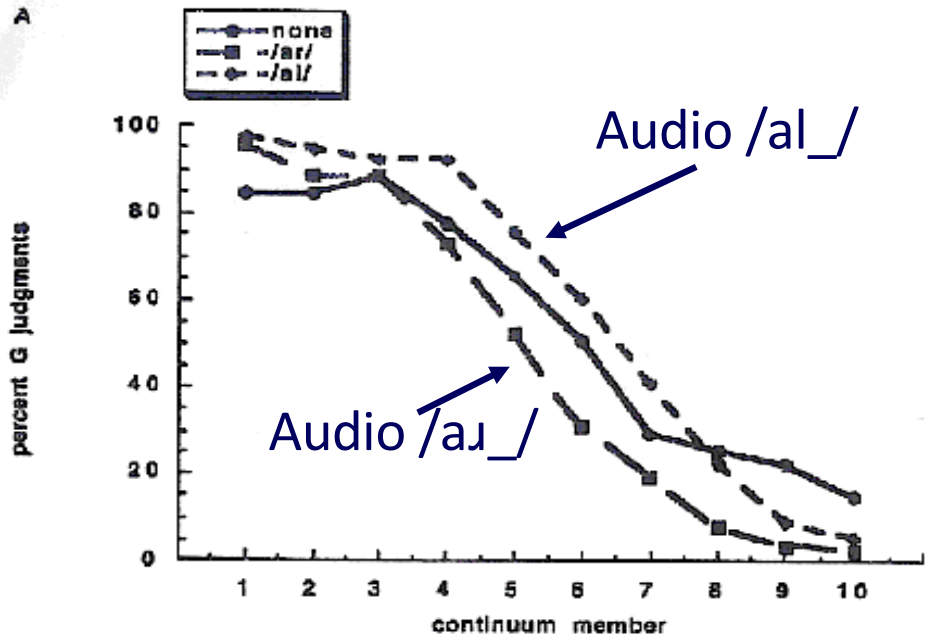
ald/ga-ard/ga: non-auditory context

Fowler, Brown, & Mann (2000, *Journal of Exp. Psych.: Human Perc. & Perf.* 26):

In response to work by Lotto, Kluender & colleagues:

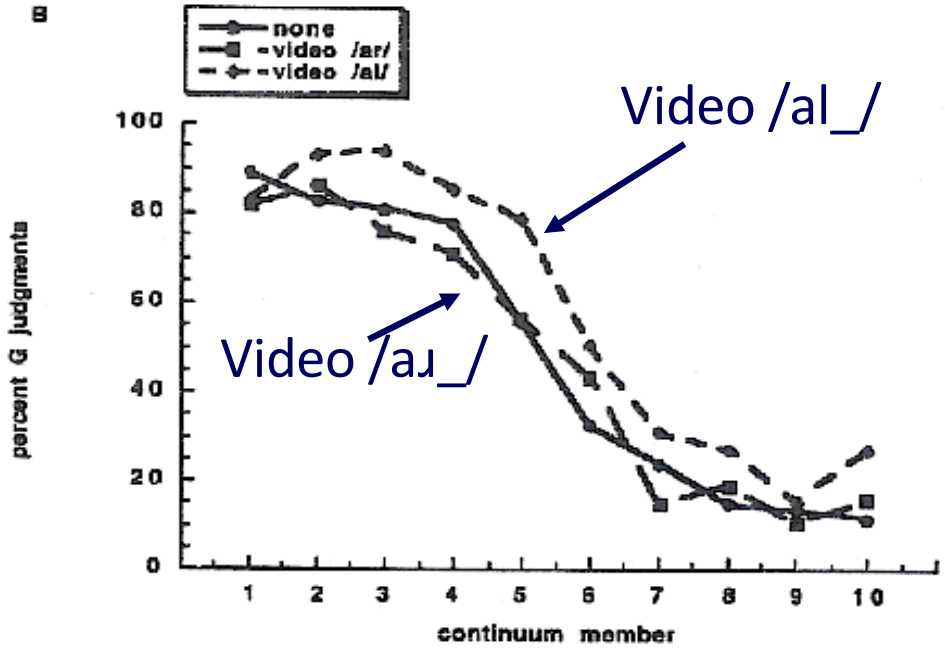
- If perceptual "compensation" is purely auditory phenomenon, then effect shouldn't hold when information for liquid is visual and not acoustic
- Variant of McGurk paradigm:
  - Visual: clear /ɹ/ or /l/
  - Auditory: constant, ambiguous [al/ɹ] precursor + /da-ga/ stimuli
- If auditory phenomenon: visual /ɹ/ or /l/ shouldn't influence /da-ga/ responses (because acoustic information for liquid is constant)
- If compensation for coarticulation: listeners should integrate visual + auditory info; should report more /ga/ after /l/ (even though acoustic cues for liquid remain constant)

ald/ga-ard/ga: non-auditory context



Fowler, Brown, & Mann (2000)

Replication condition: replicated original Mann (1980) findings with audio versions of their stimuli.



Main result:

- Listeners compensated for ***visual*** coarticulatory effects
- Since precursor audio was constant, *compensation could not be auditory effect*

Coarticulatory compensation or not?

- Lively, continuing exchange in literature (let me know if you would like more references)
- Gets at heart of dynamics of perception of coarticulation: what are listeners attending to as acoustic information evolves?
- Illustrative of theoretical and experimental exchanges in speech perception
  - Reasonably good agreement about relevant data
  - For ald-ga, aɹd-ga:
    - L2 learners
    - Infants
    - Non-speech (tones)
    - Non-human animals (quail)
    - Visual signal

Coarticulatory compensation or not?

- Lively, continuing exchange in literature (let me know if you would like more references)
- Gets at heart of dynamics of perception of coarticulation: what are listeners attending to as acoustic information evolves?
- Illustrative of theoretical and experimental exchanges in speech perception

So … is it coarticulatory compensation or not?

Next time: Gesturalist theories of speech perception
Motor Theory
Direct Realism

- What does it mean to say listeners perceive gestures?
- Why do we care—indeed, do we care—about what listeners recover from the acoustic signal?
- Two different views of perceiving gestures
  - Motor Theory: speech is perceived in a special 'module' that recruits the motor system

    Whalen's paper: reviews and updates MT claims, including based on neurological evidence
  - Direct Realism: perceptual systems perceive causes of structure (in air, light, etc.); in speech, vocal tract gestures structure the acoustic signal
- Is speech perception different from other kinds of auditory processing?