Speech and speech processing II

9.59 / 24.905
April 12, 2005

Ted Gibson

Problems for Speech Perception

- Context-conditioned variation
  - One-to-many variation: Same phoneme may be superficially realized in different ways
  - Many-to-one variation: Different phonemes can have the same sound in different contexts

Summary: Problems in Speech Perception

- Problems
  - ➢ Lack of invariance, smearing (due to coarticulation)
- Solutions
  - ➢ Acoustic features
  - ➢ Categorical perception
  - ➢ Motor theory of perception
  - ➢ Context
    - Same level
      - – Phonemic context, prosodic context
    - High level
      - – Syntactic, semantic, lexical knowledge

Solutions: Categorical Perception

- For consonants, much of the difficulty of telling sounds apart is at the boundaries among sounds

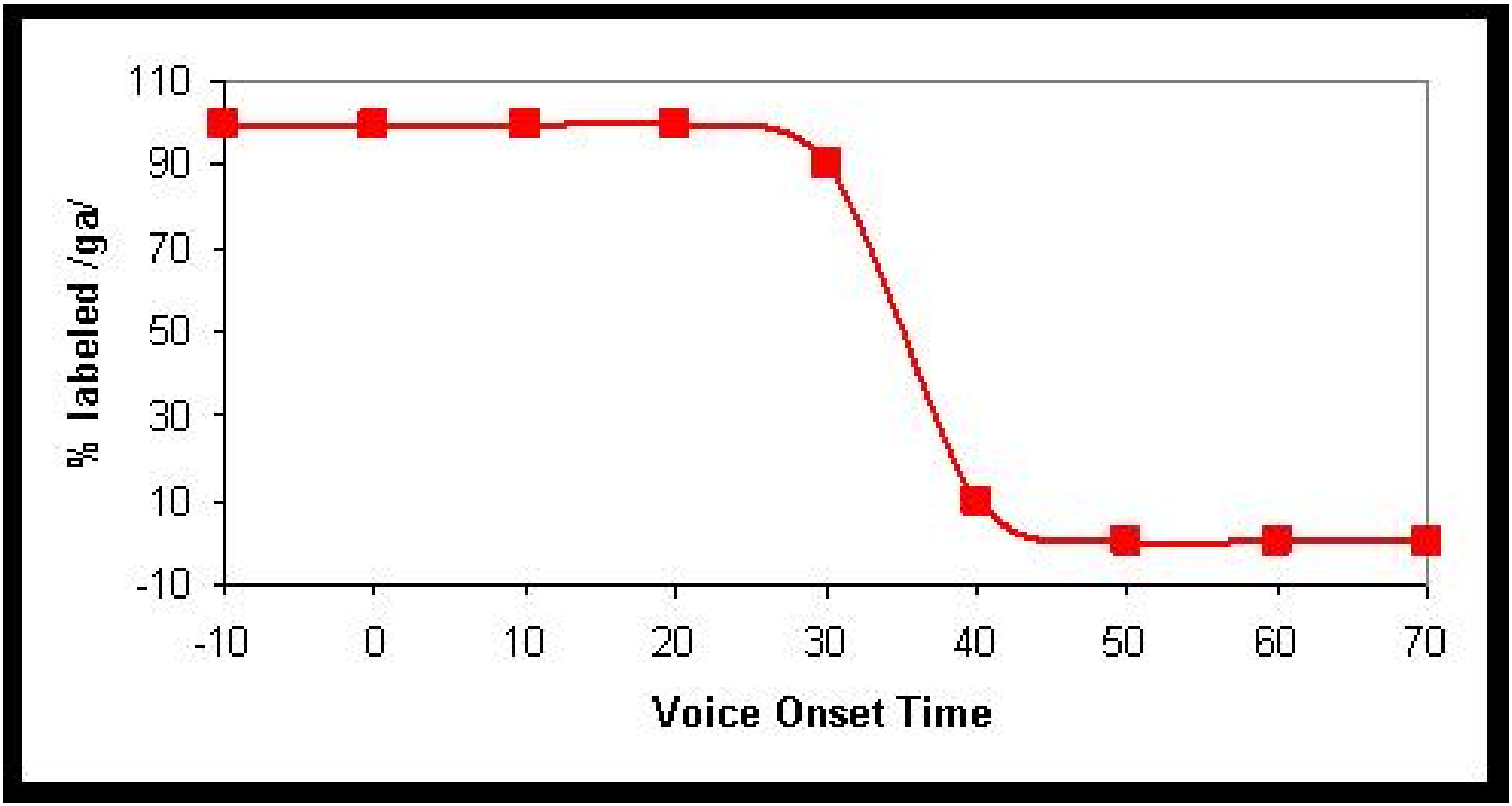
- We impose categories on physically continuous stimuli

In-class demonstration: the /ka/ - /ga/ continuum

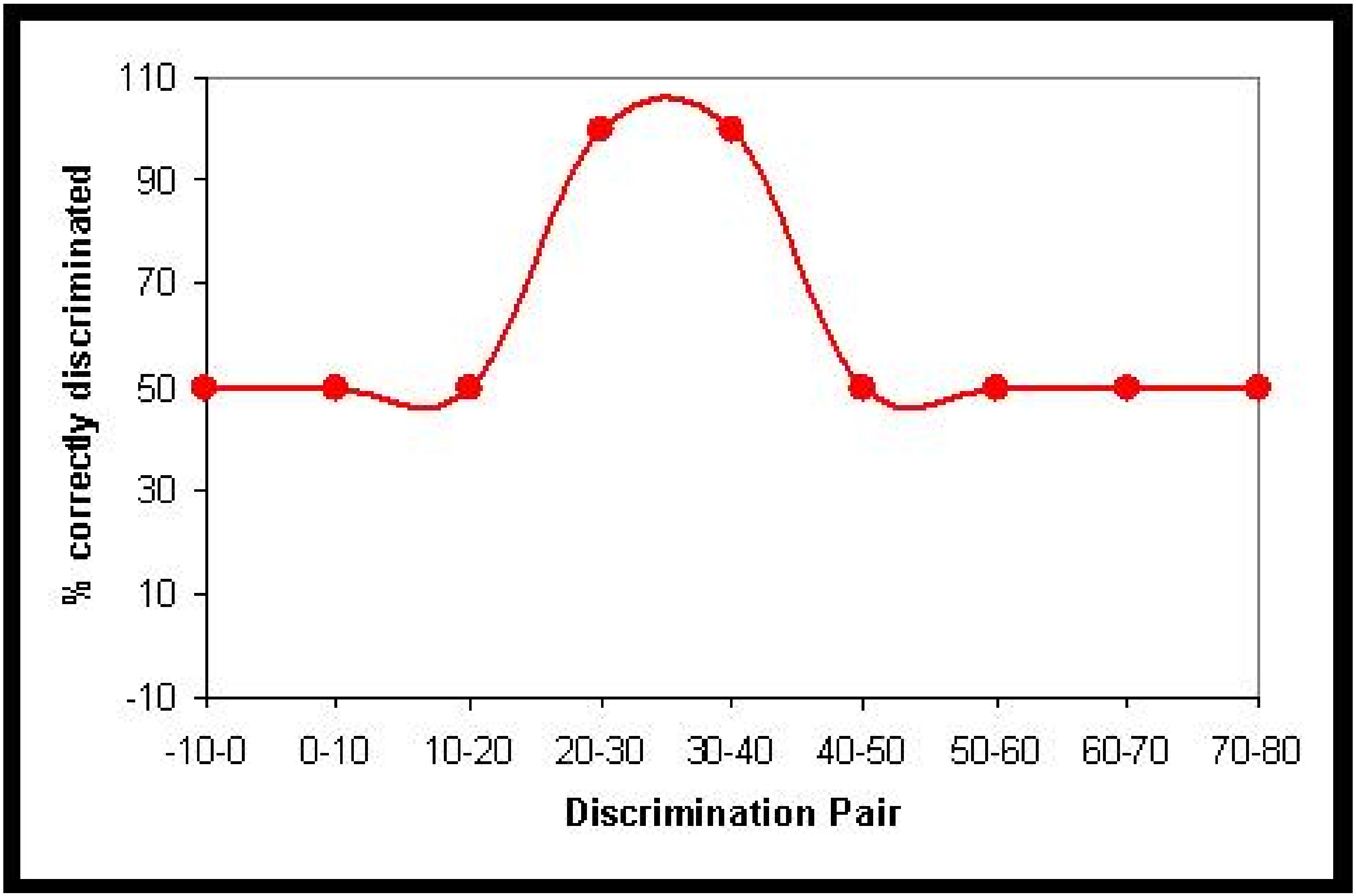
- Voicing: differences in Voice Onset Time (VOT)
- Small VOT: voiced; Large VOT: unvoiced

Graphs of frequency vs. time removed for copyright reasons.

% labeled /ga/ in /ga/-/ka/ continuum



Results of discrimination task: 10 msec intervals of VOT



- **Categorical Perception:** Can't discriminate stimuli any better than you can identify them.
  - Discriminate – tell two things apart
  - Identify – classify a sound
  - Perceptual phenomenon; Not a response strategy

What Good is Categorical Perception?

It helps to

- Ignore irrelevant information
- Quickly classify transient events
  - consonants versus vowels

Categorical Perception: Further Experiments

- In the context of a lexical access experiment, McMurray, Tanenhaus & Aslin (2002) found that people are sensitive to small within-category differences in VOT, close to the category borders

- Look at the "pear / bear" with VOT between 0 and 40 msec, 9 steps along the continuum, 5 msec apart

- Task: click on the appropriate item
- Dependent measures:
  - Mouse-clicks
  - Visual eye-movements

Categorical Perception: Further Experiments: McMurray et al (2002)

Diagram removed for copyright reasons.

"Look at the bear / pear"

Categorical Perception: Further Experiments: McMurray et al (2002)

Graph removed for copyright reasons.

Identification curves (from mouse clicks) pooled across all subjects for the word and BP identification tasks. Shown is the proportion of trials in which the p-item was selected as a function of VOT.

Note that the “ba/pa” (BP) identification task is more categorical than the word identification task.

Categorical Perception: Further Experiments: McMurray et al (2002)

Graph removed for copyright reasons.

Mean proportion fixation to the competitor picture as a function of VOT.
The left panel displays trials in which the subject responded /b/- (the competitor was the p-item).
The right panel displays trials in which the subject responded /p/- (the competitor was the b-item).
A clear gradient effect of VOT can be seen.

Categorical Perception: Further Experiments: McMurray et al (2002)

- Conclusion: People have access to acoustic properties of sounds when they are close to the category boundaries.
- Open questions:
  - More so in lexical access?
  - Effects of practice? Almost 2000 trials in the McMurray et al. experiment.

Motor Theory of Perception

- McGurk Effect – Visual information automatically integrated into speech percept
- Place of articulation cued by visual input
- Manner cued by ear

Solutions: Phonemic Context

- Use knowledge of how surrounding segments are articulated to interpret ambiguous segments
  - /s/ is higher frequency than /sh/
  - White noise is higher preceding /a/ than /u/
  - A sound halfway between /s/ and /sh/ is interpreted differently depending on whether it is pronounced before a /u/ or an /a/

Graph removed for copyright reasons.

Solutions: Prosodic Context

Rate Normalization

- We correct for speaking rate
  - ➢ VOT discrimination
    - Categorical boundary shifts for /ga/-/ka/ if previous syllable is pronounced faster (e.g., short /da/ versus long /da/)
  - ➢ Formant transitions
    - /ba/ vs. /wa/
      - /ba/: fast formant transitions
      - /wa/: slower formant transitions
    - If **succeeding** syllable is faster, then percept can change.

Solutions: Higher-Level Context

- Noisy perception (Miller, Heise, Lichten, 1951)

  Grammatical: *Accidents kill motorists on the highways.*

  Anomalous: *Accidents carry honey between the house.*

  Scrambled: *Around accidents country honey the shoot.*

- Shadowing – Echo speech you hear (Marslen-Wilson & Welsh, 1978)
  - Intentional mispronunciations
  - When corrected, they go completely unnoticed and do not delay shadowing
- Use syntax and semantics to perceive the input

Context can Affect Perception

- /pi/ vs. /bi/ demo: lexical knowledge affects categorical boundary

- Not just high-level percept, but perceptual discrimination is affected.

Summary: Problems in Speech Perception

- Problems
  - Lack of invariance, smearing
- Solutions
  - Acoustic features
  - Categorical perception
  - Motor theory of perception
  - Context
    - Same level
      - Phonemic context, prosodic context
    - High level
      - Syntactic, semantic, lexical knowledge

Are speech sounds learned / innate?

- Are all phonemic distinctions acoustically salient enough that they don’t need to be learned?
- Or are all phonemic distinctions learned?
- Which distinctions are acoustically salient enough that they don’t need to be learned?
- Which distinctions need to be learned?

Discriminability in Adulthood

- /r/ vs /l/
  - Japanese vs. English adults
- /t/ vs/ /T/
  - Hindi versus English

- Acquired Distinctiveness vs. Acquired Similarity

Are category boundaries learned or built-in (acoustically salient)?

- Infants: High-amplitude sucking (HAS) Eimas et al. (1971):

  More sucking responses, more interest

  Habituation: fewer sucking responses

Result: Infants have the same border as adults:
between 20 and 40 msec.

Diff. cat. Same cat. Control

Graphs removed for copyright reasons.

Perception of VOT in other species: Chinchillas (Kuhl & Miller, 1978)

Chinchillas trained on 0, 80 msec VOT

Test where they perceive a border

Graph removed for copyright reasons.

Result: Chinchillas perceive the same border as humans

Chinchillas and Categorical Perception

- Longer VOT for voicelessness the farther back in the mouth that a sound is made
- Chinchillas categorize VOT just like humans
  - Motor theory is not necessary to explain VOT discrimination
- Voice-voiceless is perceptually extremely salient
  - Kikuyu adults

Evidence for innate speech discriminability

- Infants make discriminations that are not made in their native environment
- Some distinctions are made early, but are later lost
  - 3 mos: /ra/-/la/ (not produced until 3-5 years)
- Lose that ability by end of 1st year
  - /ta/ vs. /Ta/
- Head turn procedure

Some categories need to be learned

- Kikuyu distinguishes prevoiced vs. voiced consonants
  - 6mos.: infants raised in English vs. Kikuyu environment
  - If difference is large, English infants can discriminate the sounds
  - Can’t discriminate in the range of Kikuyu infants

Are phonemic categories learned or innate?

- It depends on the phoneme distinction:
  - ➢ Some seem to be innate: acoustically salient
    - /ka/ - /ga/
    - Don’t need to be maintained (Kikuyu adults)
  - ➢ Some seem to be lost
    - /ra/ - /la/, /ta/-/Ta/
    - Can be relearned with some moderate training
  - ➢ Some need to be learned: less acoustically salient.
    - prevoicing in Kikuyu (learned very early)
    - Vowels (also learned very early)
    - Probably can be trained in adults, but much harder
      - – Foreign accents

Graph removed for copyright reasons.

Auditory word comprehension

- A bottom-up left-to-right model: **Cohort theory** (Marslen-Wilson & Tyler)

  Cohort theory: A word is recognized through a left-to-right activation of phonemes wherein the first phoneme accesses all possible words beginning with that sound - the **cohort** - which narrows as incoming sounds rule out members.

  When there is a unique lexical item in the cohort: word recognition.

  /a/ : consistent with many words.

  /as/ : consistent with “ostrich”, “ostensible”, “awesome”...

  /ak/ : consistent with "awkward" ...

  /akw/ : consistent with "awkward" (any others? if not, then word recognition occurs.)

Evidence for Cohort theory: Priming studies

- “beaker” primes “glass”, a semantic associate of “beaker”
- “beaker” also primes “bug”, a semantic associate of “beetle”, which is in the initial cohort of “beaker”
- Importantly, it is difficult (impossible?) to get rhyme effects in lexical decision: "beaker" does not prime "stereo", a semantic associate of "speaker"

A top-down model which is not exclusively left-to-right: TRACE (McClelland & Elman, 1986)

- A conceptual problem with the cohort theory: The segmentation problem. Words are presented continuously, with no breaks. How do we get words out of continuous speech?

  TRACE model: spreading activation model (precursor to PDP, connectionism):

  not rigid left-to-right: includes top-down influences from what counts as a word in the mental lexicon.

  Activation from acoustic features, phonemes, words

Evidence against Cohort theory

Word-initial perception effects (Ganong, 1980):

The perception of the voiced/voiceless continuum is affected by word recognition, even in word-initial (right-context) environments.

E.g., a sound halfway between /d/ and /t/:

> is interpreted as /d/ before "ash": "dash" is a word; "tash" is not a word.
>
> is interpreted as /t/ before "ack": "dack" is not a word; "tack" is a word.
>
> is interpreted 50/50 as /t/ or /d/ before "ath": "dath" is not a word; "tath" is not a word.

Evidence against Cohort theory

Phoneme restoration effects: Samuel (1981) modification:

(a) replacement: a phoneme is replaced by white noise
(b) addition: white noise is added to a phoneme.

Subjects are asked to judge "replaced" or "added".

Results:

(1) phonemes that sound like white noise (fricatives, stops) are more susceptible to this procedure than others

(2) subjects had a harder time telling "replaced" from "added" in words than in possible non-words. E.g., in "dash" vs "dass"; or in "dash" vs. "tash"

Word position not affected: word-initial effects.

Evidence against Cohort theory

Rhyming effects: tracking eye-movements (Allopena, Magnuson & Tanenhaus, 1998).

Visual context with 4 items: target, cohort, rhyme and distractor

E.g.: target: “beaker”
cohort: “beetle”
rhyme: “speaker”
distractor: “carriage”

Instruction: “look at the beaker”

Result: get looks to target, cohort, and (crucially) also some looks to the rhyme.

The looks to the rhyme are not predicted by cohort theory. The results are evidence against a rigid left-to-right theory.

Allopena et al. (1998)

Diagram removed for copyright reasons.
“Figure 3.”

Allopena et al. (1998)

Graph removed for copyright reasons.
“Figure 4.”