Categorization and abstraction in perception

LING611 Spring 2021

Brian Dillon
Shota Momma

University of Massachusetts, Amherst
Department of Linguistics

2/3/2021

Today's goals

Our discussion today will touch on:

- Subcategorical detail in perception
- Acquisition of categories and distributional learning

Scene setting

$\Rightarrow$ This is Loki, an expert, probability-matching treat forager. He knew you do better finding treats if you know not just *where* they are, but *how likely* it is for a treat to be someplace. Tracking probabilities helps deal with uncertainty (i.e. *where are the treats??*).

Scene setting

Dealing with uncertainty is arguably a central problem in language processing. Do language users manage uncertainty about linguistic structure in the input like Loki managed uncertainty about treat locations?

- Do we track subcategorical (probabilistic) cues to linguistic categories?
- Do we exploit these cues in processing linguistic input?
- Do learners exploit ‘structured variation’ to discover categories in their experience?

Tracking activation in real-time: The visual world

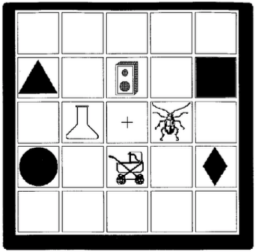
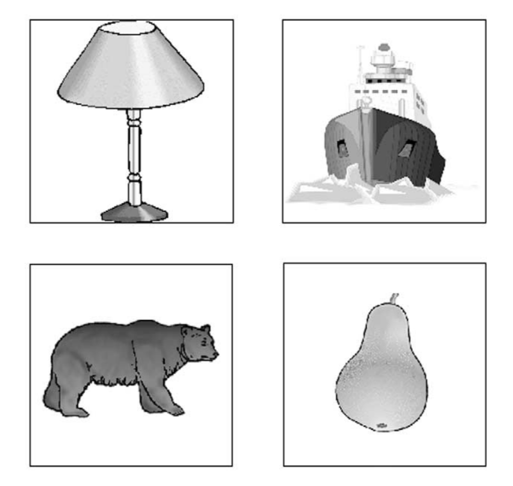


FIG. 3. An example of a stimulus display presented to participants.

Allopenna et al., 1998

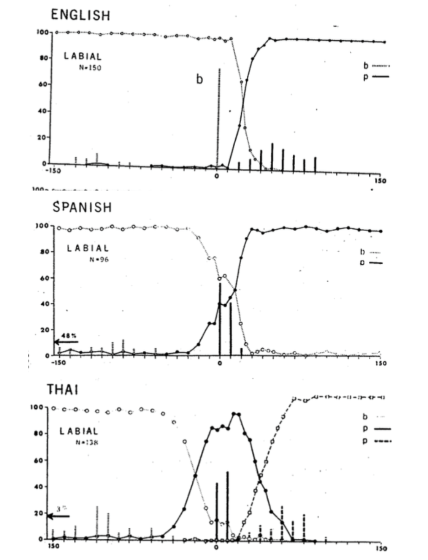
⇒ **Eye-mind hypothesis** (Just & Carpenter, 1980): Eye movements (probabilistically) reflect contents of active attention (cf Magnuson, 2019). A hypothesis about how cognitive events are related to observable measures is called a *linking hypothesis*.

McMurray et al (2002)



⇒ /bɛɹ/ vs /pɛɹ/ are (categorically) distinguished by voicing. But are listeners sensitive to subcategorical *degree* of voicing?

Distribution of VOTs by language



McMurray et al (2002)

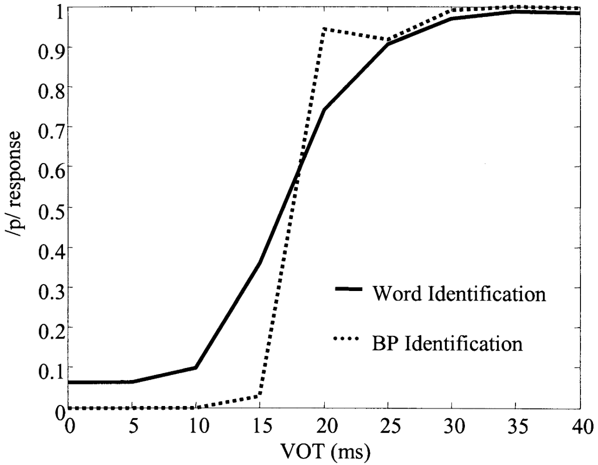


Fig. 2. Identification curves (from mouse clicks) pooled across all subjects for the word and BP identification tasks. Shown is the proportion of trials in which the p-item was selected as a function of VOT.

McMurray et al (2002)

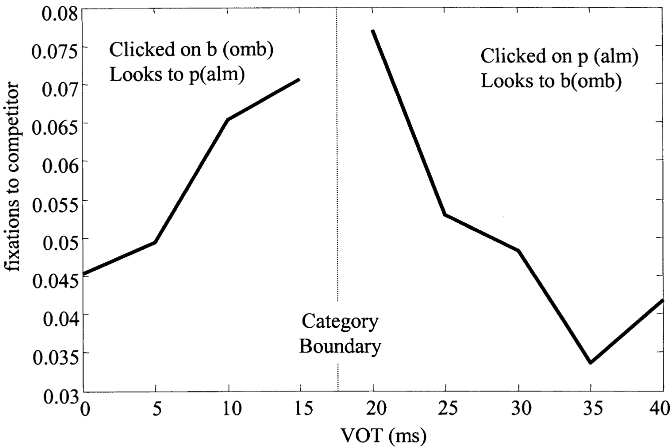


Fig. 3. Mean proportion fixation to the competitor picture as a function of VOT. The left panel displays trials in which the subject responded /b/- (and the competitor was the p-item). The right panel displays trials in which the subject responded /p/- (and the competitor was the b-item). A clear gradient effect of VOT can be seen.

McMurray et al (2002)

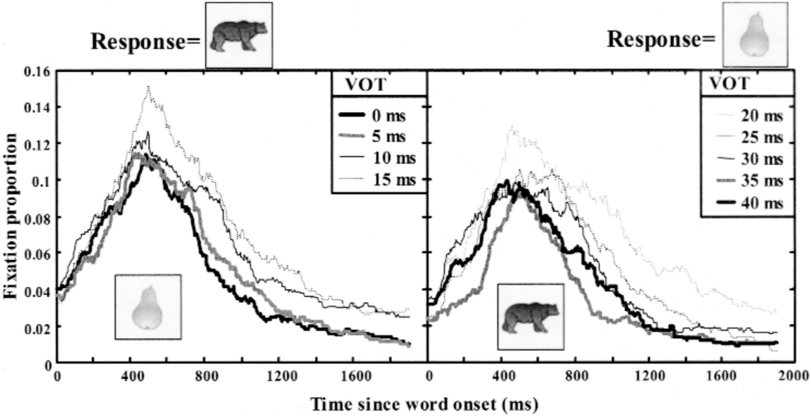


Fig. 7. Effect of VOT and time on fixations to the competitor. A clear gradient effect of VOT can be seen. Importantly, the effect of VOT is primarily on the duration of activation.

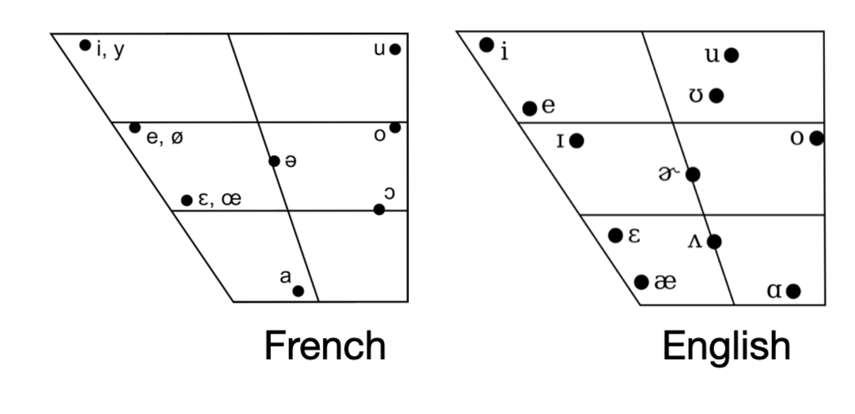
Key takeaways

- Lexical competitors activated as a gradient function of VOT: As you approach the category boundary, the activation of the competitor increases.
- Activation of the competitor was not short-lived: Competitors seem to remain active in proportion to their likelihood for upwards of a second after the ambiguous segment.
- McMurray et al's hypothesis: Subcategorical distinctions are preserved and maintained by listeners to deal with ambiguity / uncertainty. Listeners track what VOTs make a good instance of /p/ or /b/ and exploit this information 'online' during language comprehension.
- **Example:** The /d̥/ent in the fender/woods (Connine, Blasko & Hall, 1991).

Scene setting

- Do we track subcategorical (probabilistic) cues to linguistic categories?
- Do we exploit these cues in processing linguistic input?
- Do learners exploit ‘structured variation’ to discover categories in their experience?

How do learners discover phonetic / phonological categories? Consider the target of acquisition: Languages vary not just in the number and character of phonetic / phonological categories, but also the fine phonetic details of those categories:



The conditioned head turn procedure

What do pre-verbal infants know about phonetic categories? The **conditioned head turn procedure** provides one tool. One standard use: Condition infants to turn their head and look at a 'reinforcer' when there is a change in a stream of speech stimuli (Werker et al., 1997).

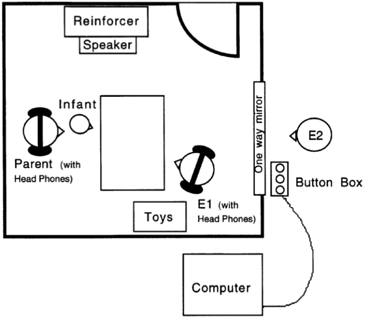
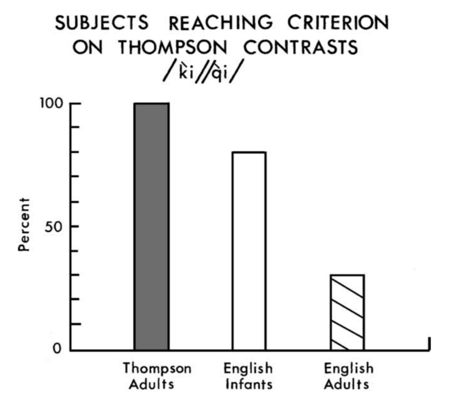


Figure 1. A schematic diagram of the Conditioned Head Turn procedure.

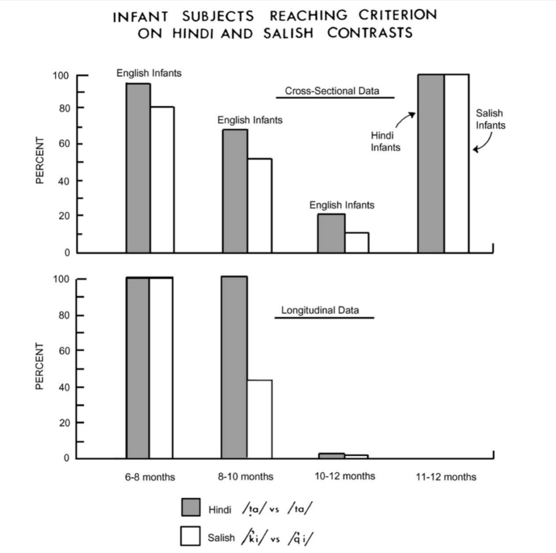
Werker & Tees (1984)

How well can adult English speakers distinguish place of articulation in a /k'/ - /q'/ contrast (e.g. Thompson Salish)? Or in a /t̪/ - /ʈ/ contrast (e.g. Hindi)?



Werker & Tees (1984)

What about infants?



Habituation paradigm

- Orient infant to visual display; play stimulus until looking time drops to pre-determined threshold (e.g. they 'get bored'.
- At habituation, continue play same stimulus on **same** trials, or change stimulus on **change** trials
- If infants detect a change, they will dishabituate and look more at the display. Increase in looking times for **change** trials compared to **same** trial baseline indicates discrimination.

Narayan et al. (2010

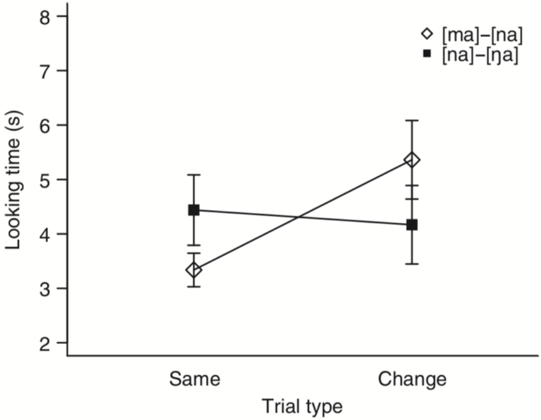
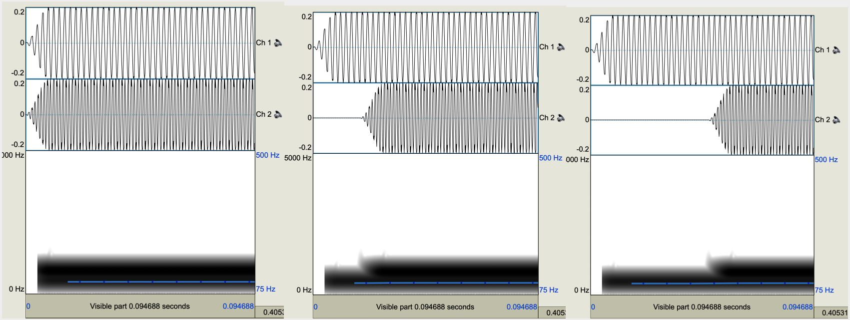


Figure 4 *English-hearing infants' (at 4–5 months) looking time to same vs. change trials for the native [ma]–[na] and non-native [na]–[ŋa] contrasts. Error bars represent standard error.*

Scene setting

Do learners exploit 'structured variation' to discover categories in their experience?
Distributional learning: A hypothesized learning procedure that leverages structured variation in the input to 'infer' the latent structure of the input. Common examples:

Challenges



⇒ Two pure tones at 440Hz and 780Hz, with 0, 20, and 50ms onset asynchrony.)

Challenges

The long-tailed (common) Chinchilla (*chinchilla lanigera*)

Challenges

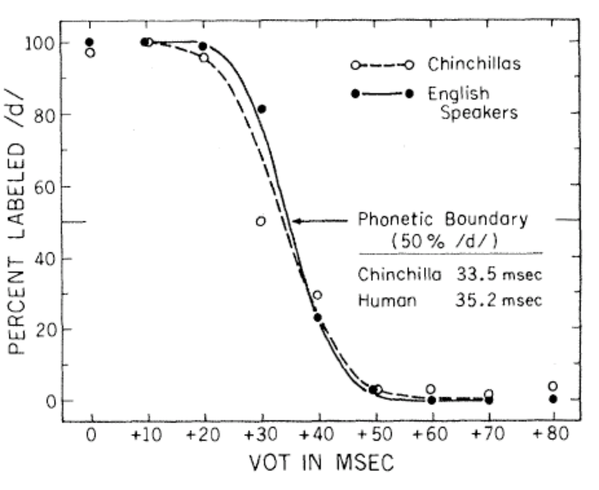


Figure: Data from Kuhl & Miller 1978

Seeking invariance: The ‘General approach‘

The ‘General approach‘ framework holds:

- The objects of speech perception are abstract linguistic units, not identified with (intended) gestures.
- Perception is mediated by general auditory mechanisms and perceptual learning (shaped but not determined by experience).
- Perceptual constancy reflect ability to combine multiple cues to categorize experience (not unique to speech).

Tuning perception: Lisker & Abramson (1964)

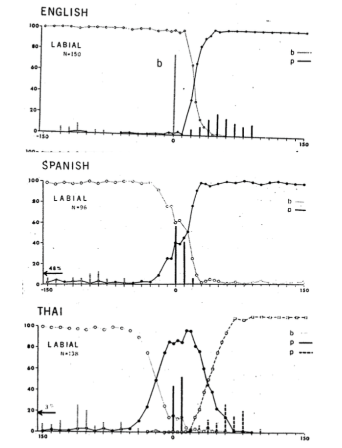


Figure: ⇒ Possible VOT boundaries align with discontinuites in perception of 'TOT' boundaries. VOT contrasts may 'maintain' some boundaries, and elide others.

Exploiting audition in perception: Coarticulation

- Stop closures /ga/, /da/ are fronted after more anterior /al/, backed after more posterior /aɹ/: Coarticulation.
- In isolation, the onset frequency of the third formant (F3) is one cue to place of articulation: /d/ has higher F3 than /g/.
- Coarticulation alters this acoustic cue: fronting with /al/ raises onset of F3, making /ga/ more similar to /da/. Conversely for the backing with /aɹ/.
- As a consequence the formant cues to the stop in /alga/ and /arda/ can be difficult to distinguish on their own.
- How does this impact perception?

Exploiting audition in perception: Coarticulation

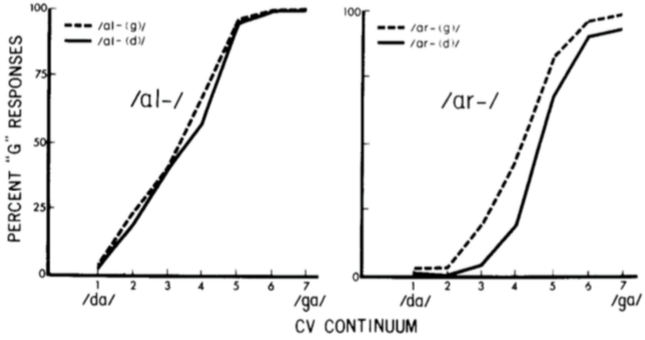


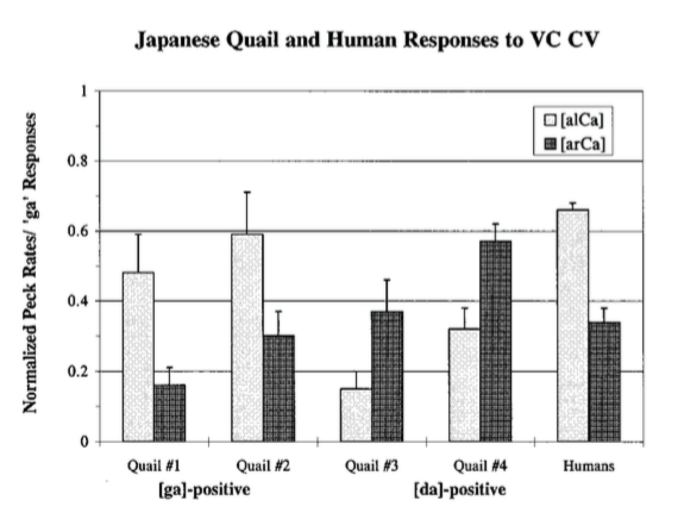
Figure 3. Percentage of "g" responses given to CV stimuli as a function of whether the preceding liquid had originally been produced before [d] or [g].

⇒ Fewer /g/ responses after /ar/: Compensation for coarticulation. From Mann (1980).

Exploiting audition in perception: Coarticulation

The Japanese Quail (*coturnix japonica*)

Exploiting audition in perception: Coarticulation



⇒ Japanese quails show similar compensation for coarticulation.
From Lotto et al. (1997).

Exploiting audition in perception: Coarticulation

- A (GA) spectral contrast account: following high frequency Fs, lower frequency F3 perceptions are reported and vice versa. Contrast heightens perception of a difference.
- Spectral contrast effects can be cued by pure tones preceding /da/-/ga/ syllables (Kluender, 1998)
- A general perceptual function seems well suited to accounting for coarticulation effects in this instance.

Stepping back

- Speech perception involves recovery of relevant phonetic categories.
- This is shaped by general perceptual properties of the auditory system, in humans and non-humans.
- Despite this alignment, linguistic experience appears to shape or exploit these natural boundaries. More on this next time.