# Speech Perception
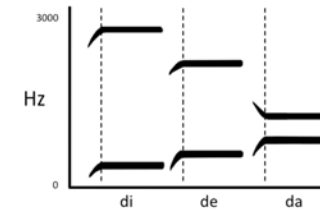
Facts and things

---

# Problems

- Lack of invariance
- Talker normalization
- Segmentation
- Speech is too fast to hear!

---

# Lack of Invariance Problem



There is no unique acoustic pattern associated with the perception of phonemes.

---

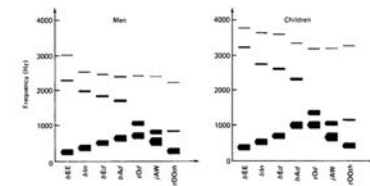# Why the lack of invariance?

- Co-articulation
- We don't articulate one sound at a time. We articulate more than one *in parallel*.
- say key, koala

---

# Lack of invariance not unique to speech



---

# Talker Normalization



Formant patterns for 7 vowels, men vs. children
Similar structure, but lots of frequency diffs

## Segmentation

Oronyms Illustrate the Problem

↓

phrases that can be segmented in more than one way

## "Mondegreens"

As a child, author Sylvia Wright heard the lyrics of The Bonny Earl of Murray (a Scottish ballad) as:

*Ye highlands and ye lowlands*
*Oh where hae you been?*
*Thou hae slay the Earl of Murray*
*And Lady Mondegreen*

It eventually transpired that Lady Mondegreen existed only in the mind of Sylvia Wright, for the actual lyrics said that they "slay the Earl of Murray and laid him on the green." And to this day Lady Mondegreen's name has been used to describe all mishearings of this type!

---

"Excuse me while I kiss this guy."
"Excuse me while I kiss the sky."
Purple Haze, Jimi Hendrix

"She's got a chicken to ride."
"She's got a ticket to ride."
Ticket to Ride, The Beatles

"You and me and Leslie."
"You and me endlessly..."
Groovin', The Rascals

"Sunday monkey won't play piano song, play piano song."
"Sont des mots qui vont tres bien ensemble; tres bien ensemble."
Michelle, The Beatles

"What a nice surprise when you're out of ice."
"What a nice surprise bring your alibis."
Hotel California, Eagles

"I'm a pool hall ace."
"My poor heart aches."
Every Step You Take, The Police

---

Manfred Mann "Blinded by the Light"

---
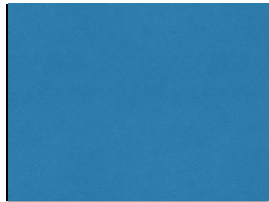
Manfred Mann "Blinded by the Light"

---

## The case of "nuther"

Slide 1:
Another
A    nuther

Slide 2:
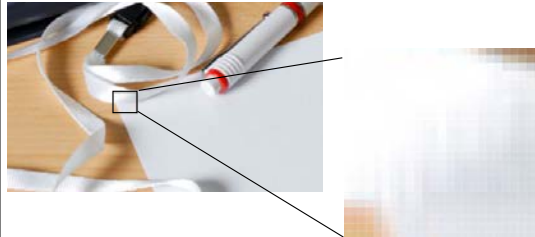Another
A    nuther
That's a whole nuther thing

Slide 3:
Segmentation problem not unique to speech

Slide 4:
Segmentation problem not unique to speech

Slide 5:
Segmentation problem not unique to speech

Slide 6:
Speech is too fast to hear! (WTF?)

1. A fast sentence: 1 phoneme every 75 msec — can you understand it?
2. A fast melody: 1 note every 75 sec — can you name that tune?
3. The same melody at normal pace: 1 note every 340 sec
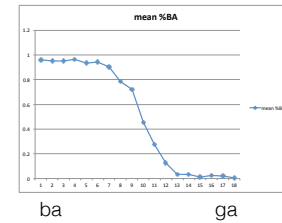4. A ~normal sentence: 1 phoneme every 110 sec <— still very fast!

## How can we hear speech that fast?

- Parallel transmission of phonemic information (co-articulation)

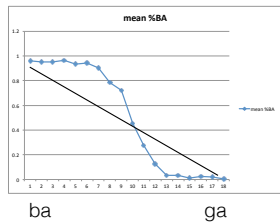- Maybe the units aren't phonemes ???

## Phenomena

- Categorical perception

- Ganong effect

- Phonetic context effect

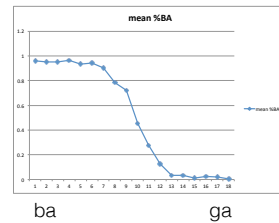- Phonemic restoration

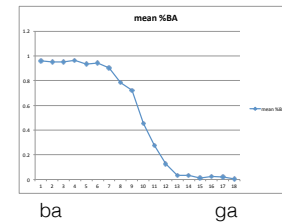- McGurk-MacDonald

## Categorical Perception



## Categorical Perception



Not a linear curve

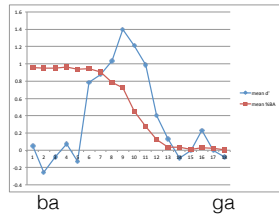## Categorical Perception



But this isn't surprising:
listeners were *asked* to categorize!
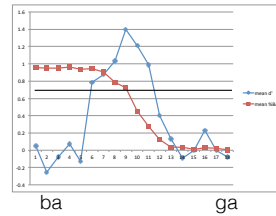
## Categorical Perception



Categorical perception is demonstrated by showing that discrimination is better across the category boundary than within it

## Categorical Perception



ba      ga

Categorical perception demonstrated by showing that discrimination is better across the category boundary than within it

## Categorical Perception



ba      ga

Not a flat function!

## Categorical Perception

BUT! If listeners are asked instead to rate how good the syllable is as an exemplar of a category, the functions look more linear (not so categorical).
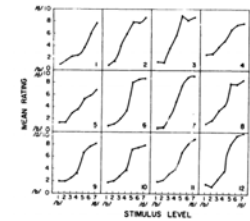
*The task is important.*



Fig. 4. Mean /ba/-/da/ ratings for each of the 12 individual subjects from the place condition.

Massaro & Cohen 1983

## Why Categorize?

- Imagine a world without categories? How would you manage *anything?*



## Why Categorize?

- Imagine a world without categories? How would you manage *anything?*



## Why Categorize?

- Categories guide our behavior by allowing us to generalize from past experience

- Same is true for speech

## A neural mechanism?



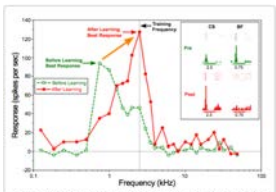Auditory cortex tunes to behaviorally relevant sound frequencies

Figure 1: Associative learning produces tuning shifts. An example of a complete shift of frequency tuning of a single cell in A1 of the guinea pig from a pre-training best frequency (BF) of 0.75 kHz to the CS frequency of 2.5 kHz after 30 trials of tone–shock pairing, during which the guinea pig developed a cardiac conditioned response. Inset shows pre- and post-training cellular responses (rasters) for the pre-training BF and the CS frequencies. Note the marked increase in neuronal discharges to the signal frequency of 2.5 kHz and the pronounced decrease in neuronal responses to the pre-training best frequency of 0.75 kHz.

## Ganong Effect

- Named for William Ganong
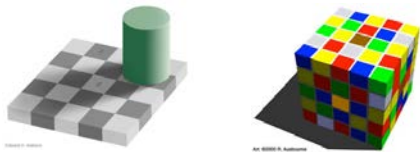- A lexical effect on speech perception
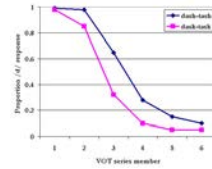


What's the last sound in these words?

## Visual "Ganongs"



## Visual "Ganongs"



## Phonetic Context Effect

- /al/ + ?

  ga? da?

- /ar/ + ?

## Visual Context Effect

## What does it mean?

- The context matters

- The brain is not just processing local bits and pieces. It is using surrounding and higher-level information to construct our perceptual experience.


## Phonemic Restoration

- "There is not a giraffe standing next to me"

- Same sentence with a cough.


## Phonemic Restoration

- "There is not a giraffe standing next to me"

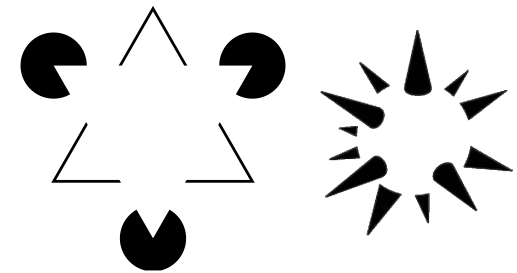- Is the sentence complete?


## Phonemic Restoration

- "There is not a giraffe standing next to me"

- Answer: no.

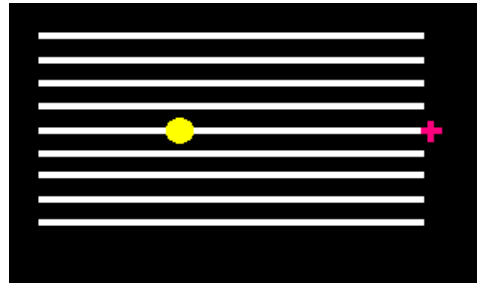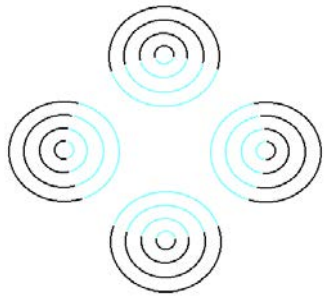- There is gap.

- What sound is missing?


## Phonemic Restoration

- "There is not a giraffe standing next to me"

- Answer: /f/ in giraffe is gone


## Visual "Restoration"

Visual "Restoration"



McGurk-MacDonald Effect



McGurk-MacDonald Effect