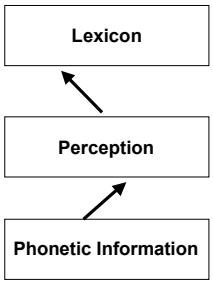
‘Perceptual warping’ and linguistic effects

Introduction to Speech Sciences: Week 9

Summary of how speech perception is challenging

In short, the phonetic cues are spread out over time, are highly variable, and can be obscured by noise

How we meet the challenges



This week: Use of linguistic structure

This week: ‘perceptual warping’ of the acoustic cues

Last week: Use of redundant multimodal phonetic cues

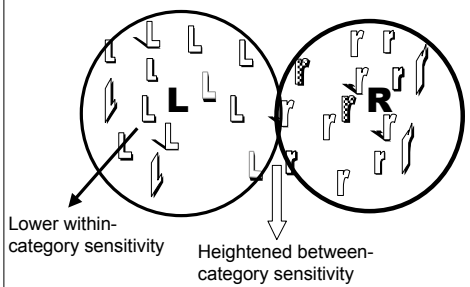
How we meet the challenge of speech perception

Part 3: We “perceptual warp” the acoustic cues

Speech has high phonetic variability

- Some of this is “good” variability
  - It can cue the differences between phonemes
  - e.g., the different burst frequencies for /b/ and /d/
- Some of this is “bad” variability
  - A lot of variability is irrelevant to phonemic categorization
  - e.g., differences in formants between adults and children
  - Some of this “bad” variability is actually useful for other things (e.g., indexical cues)
- Fortunately, our auditory system is tuned to be highly sensitive to the “good” variability and less sensitive to the “bad” variability

Notion of perceptual warping



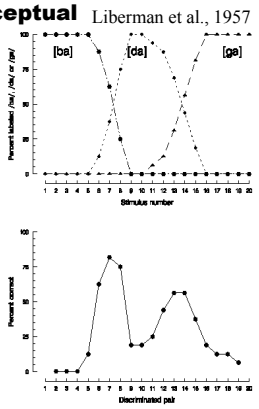
Notion of perceptual warping

- Equal physical steps perceived as unequal
- Certain differences have much greater effect on perception than others
- Tested using *discrimination experiments*
  - Play people stimuli along a continuum that have equal acoustic differences
  - Test how well they can detect the acoustic differences at all points along the continuum
  - e.g., play three stimuli and ask people which one is acoustically different

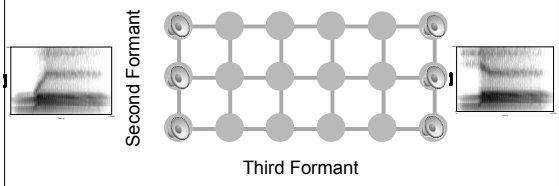
Early evidence of perceptual warping: 'Categorical perception'

Liberman et al., 1957

- *Categorical perception*: more accurate at discriminating stimuli at boundaries than within categories
  - Almost as if we perceive the stimuli in terms of their category labels
- Opposite of *continuous perception*
  - Same sensitivity for all acoustic differences



A two-dimensional view of perceptual warping: /r/ and /l/

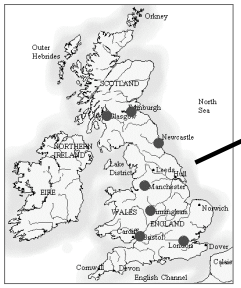


Iverson et al., 2003

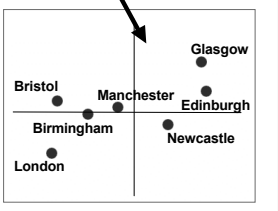
3 tasks

- Phoneme identification
- Goodness judgments
- Similarity scaling
  - Ratings for stimulus pairs
  - Analysis using *Multidimensional Scaling*

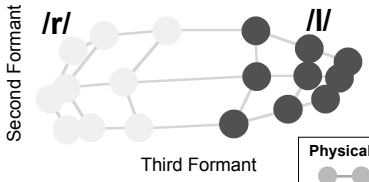
Multidimensional scaling



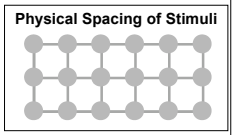
	Birmingham	Bristol	Edinburgh	...
Birmingham	0			...
Bristol	4	0		...
Edinburgh	5	7	0	...
...	...	...	...	...



'Perceptual map' for American English speakers

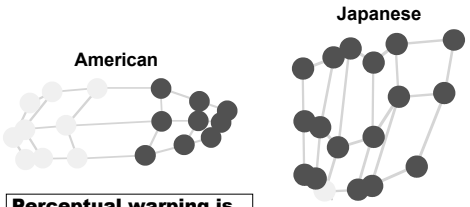


- *Stretched* at category boundaries
- *Shrunk* near 'best exemplars'
  - *Perceptual magnet effect*



Iverson et al., 2003

'Perceptual maps' for American English and Japanese speakers



Perceptual warping is critically dependent on experience (week 10)

Iverson et al., 2003

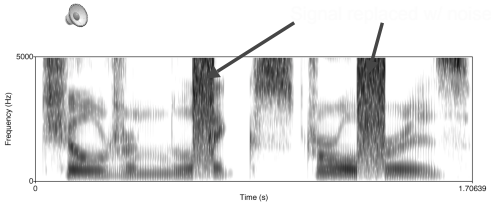
Summary of perceptual warping

- Perception is "warped" in the sense that our perceptual space is distorted compared to the acoustic/phonetic space
  - We are "tuned" so that we can better hear the important differences between phonemes, and can better ignore the unimportant variation
- Where does this warping come from?
  - Starts with basic auditory sensitivity
  - Further tuned when we learn our native language

How we meet the challenge of speech perception

Part 4: We use our knowledge of language

Example linguistic effect: Phoneme restoration



- We perceptually "restore" the missing phonetic information because of the linguistic structure

The structure of language allows us to get by with fewer cues

- Lexical information
  - Not all combinations of phonemes make real words
    - e.g., "cigarette" cannot be confused with "shigarette"
- Syntactic information
  - Rule-out words based on the structure of the sentence
- Semantic information
  - Rule-out words based on the topic of the sentence

We just need enough cues to distinguish between the words that are likely to be said. We do not need to hear all phonemes perfectly

Lexical effects: Vocabulary size

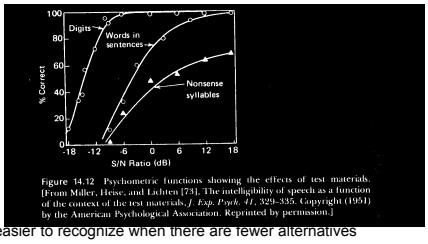


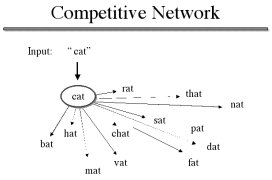
Figure 14.12 Psychometric functions showing the effects of test materials. [From Miller, Heise, and Lichten [73]. The intelligibility of speech as a function of the context of the test materials. *J. Exp. Psych. 41*, 329-335. Copyright (1951) by the American Psychological Association. Reprinted by permission.]

- Words are easier to recognize when there are fewer alternatives
  - digits > words > nonsense words
  - e.g., if we know that we are hearing digits, we won't think that "6" is "fix"
  - e.g., if we know that we are hearing nonsense words, then anything is possible (e.g., "6" could be "zicks")

Lexical effects: Word frequency of occurrence

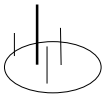
- Common words are easier to recognize than are uncommon words
  - e.g., “hat” is easier than “hack”
- All else being equal, we tend to guess that people are going to say “typical” words rather than “unusual” words

Lexical neighborhood effects (Luce & Pisoni et al.)

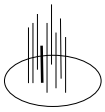


- We activate “neighborhoods” of phonetically similar words during speech perception
  - e.g., words that differ by one phoneme
- We make a guess about which word in the neighborhood matched the input
  - We use lexical frequency in our guess

Lexical neighborhood effects (Luce & Pisoni et al.)

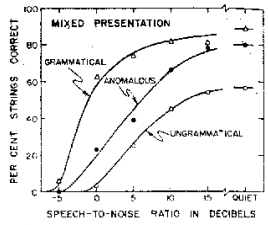


- ***Easy words*** are high frequency and have few neighbors. They can be recognized with less phonetic information. e.g., “cigarette” or “orange”



- ***Hard words*** are low frequency and have many higher-frequency neighbors. They need a lot of phonetic information to be recognized. e.g., “hack” or “vat”

Effects of grammatical structure



Grammatical sentences
Gadgets simplify work around the house.

Semantically anomalous sentences
Gadgets kill passengers from the eyes

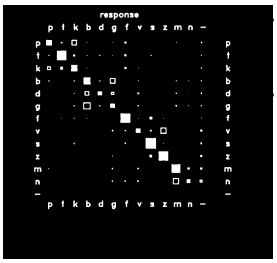
Ungrammatical sentences
Gadgets accidents country honey the shoot

- The grammar and semantics of sentences give us more information about what words are likely to be right
- Word recognition is kind of a guessing game where we pick the most likely word based on the phonetic information we perceive and the words that would make most sense linguistically

Implications for the design of speech audiometry materials

- All words and sentences are not equally easy to recognize.
  - Need to compare people using standardized tests
  - Bad idea to test people repeatedly on the same test, because they learn the materials and learn to guess
- Different tests use context to different degrees, so it is important to know what level you want to test.
  - Use of phonetic information?
  - Use of semantic context?
  - “Real-world” speech recognition skills?

VCV test



Excellent for measuring what phonetic information we perceive.

Does not measure any kind of higher-level processing (e.g., lexical, semantic, or syntactic effects).

SPIN test

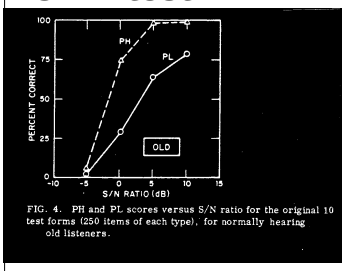


FIG. 4. PH and PL scores versus S/N ratio for the original 10 test forms (250 items of each type), for normally hearing old listeners.

HP The watchdog gave a warning growl.
HP She made the bed with clean sheets.

LP The old man discussed the dive.
LP Bob heard Paul call about the strips.

Tells us how well people can make use of semantic information, but tells us less about the use of phonetic information

"Global" tests

- ***Text comprehension***: presentation of paragraph level material followed by a set of open or closed questions
- ***Connected Discourse Tracking***: how many words in a passage can you 'transmit' to a listener per minute?
- ***Sentence verification task***: reaction time for "true/false" responses to sentences such as "Mud is dirty" and "rockets move slowly"

Gives us a better indication of how well someone is functioning in the real world, but does not allow us to analyze where people are making errors

Levels of assessment

ANALYTIC ---------->GLOBAL

Far from "normal communication"	Close to "normal communication"
Word level	Sentence/paragraph level
Provide reliable information about the use of acoustic information	Cannot reliably be used to evaluate the use of acoustic information

Summary of linguistic effects

•We use our knowledge of language to help guess what was said
  •e.g., lexical neighborhoods, lexical frequency, semantic and grammatical probabilities

•The guesses "constrain" the amount of phonetic information that we need to perceive
  •It helps us solve the challenge of speech perception because we do not need to hear every cue perfectly

•We can use different clinical tests to examine speech perception at global or analytic levels

Today's Lab: Neighborhood Activation Model

Test yourself on lexically "hard" and "easy" words mixed with noise.