Language Comprehension
Mohammed A. Aljohani
Perception of Language
Structure of Speech

• Categorization

• Complex task!

  • Context interferes with the speech signal

  • Variability of the speech signal: voice, rate, phonetic context
Prosodic Factors

Suprasegmentals

- **Stress, intonation, and rate**

- **Prosody**: “a general term that refers to the aspects of an utterance’s sound that are not specific to the words themselves.” Ferreira, 2003

- **Stress**: refers to the emphasis given to syllables in a sentence. Closely related to loudness.

- **Intonation**: refers to the use of pitch to signify different meanings; pattern of a sentence is called *intonational contour*

  - “We aim to please. You aim too, please” (Fromkin & Rodman, 1974).

- **Rate**: refers to the speed at which speech is articulated.

- E.g. Bill wants to walk but Mary wants to drive. VS Bill wants to walk to the store.
Perception of Isolated Speech Signals
Perception of Isolated Speech Signals

- Levels of speech processing:
  - auditory level
  - phonetic level
  - phonological level
Perception of Isolated Speech Signals

• Speech as a modular system

• Criteria for modularity by Fodor (1983): a cognitive system is modular if it is:
  • domain-specific
  • operates on a mandatory basis
  • fast
  • unaffected by feedback

• If speech is a modular system, then we might expect it to have a specialized neurological representation.
Perception of Isolated Speech Signals

• Speech as a modular system

• The lack of invariance problem

• There is no one-to-one correspondence between acoustic features and perceptual categorization of speech sounds.

• EX: /d/ can have different acoustic features in different syllable contexts as /da/ and /di/.

• The lack of such an invariant relationship suggests that the perception of speech segments must occur through a process that is different from and more complex than that of “ordinary” auditory perception. In other words, speech is a special mode of perception.
Perception of Isolated Speech Signals

• Speech as a modular system

• Categorical perception

• Our job: [p] or [b] not their auditory cues as frequency and intensity = speech is a special mode of perception

• [ba]: VOT = 0 ms; [pa]: VOT = 40 ms

• Vowels are perceived differently from consonants: “Stop consonants are indeed perceived differently than vowels. For while the vowel, carrier of stress, rhythm, and prosody, leaves a rapidly fading “echo,” the consonant leaves none. The initial sound of [da], for example, is difficult if not impossible to hear: the sound escapes us and we perceive the event, almost instantly, as phonetic.” (Studdert-Kennedy, 1975, p.12)
Perception of Isolated Speech Signals

• The Motor Theory of Speech Perception

  • after the discovery of categorical perception

  • Theory based on the notion that perception proceeds “by reference” to production.

  • Listeners use implicit articulatory knowledge—knowledge about how sounds are produced—as an aid in perception. To some extent, this approach is motivated by the economy of using the same mechanisms for both perception and production. But the main rationale for the motor theory is that it deals effectively with the lack of invariance discussed earlier.

  • Liberman and colleagues (1967): Sounds produced in similar ways but with varying acoustic representations are perceived in similar ways.
Perception of Isolated Speech Signals

• The Motor Theory of Speech Perception

• **Supporting studies:**

• Anecdotal evidence suggests that teaching students to produce sounds silently aids them in the identification of new sounds (Catford et al., 1991). As Catford and colleagues point out, this activity might encourage new learners to attend to subtle motor processes that would otherwise be overshadowed by auditory sensations.

• McGurk and MacDonald (1976) showed that when visual information and auditory information are in conflict, perceivers use both sources of information to arrive at a stable perception.
Perception of Isolated Speech Signals

• The Motor Theory of Speech Perception

• McGurk effect:

  McGurk (1978) demonstrated that place of articulation (especially the lips) is cued primarily by eye and that manner of articulation is cued more by ear.
Perception of Isolated Speech Signals

- The Motor Theory of Speech Perception

- Update:
  - In the revised theory, the claim is that the objects of speech are the *intended phonetic gestures* of the speaker. Phonetic gestures include such movements as rounding of the lips, raising of the jaw, and so on.
  
  By “*intended phonetic gestures,*” Liberman and Mattingly are referring to invariant motor commands sent from the brain to the structures in the vocal tract. According to this revised theory, the conversion from acoustic signal to intended phonetic gesture is done rapidly and automatically by a phonetic module.

- Support: Ojemann (1983) provides some support for the idea that the perception and production areas of the brain are closely related.


- *They speculate that infants in their first year may be sensitive to the acoustic consequences of all language gestures significant in any language and only over time narrow down to their own language.*
Perception of Continuous Speech
Perception of Continuous Speech

• Speech sounds are embedded in a context of fluent speech.

• Pollack and Pickett (1964): 1/2 of words in isolation were identified.

• Context Factors that will be discussed:
  • Prosodic factors
  • Higher-order semantic and syntactic factors
Prosodic Factors

• **Stress:** we perceive stress by a combination of acoustic cues along with our knowledge of the stress rules of the language (Lieberman, 1965).

• Bolinger and Gerstman (1957): light house keeper
Prosodic Factors

- **Rate**: Miller (1981), documented the acoustic consequences of changes in speaking rates. As we speed up, vowel duration is reduced, and the duration of the cues that signal various consonantal distinctions is also modified.

- VOT decreases as rate increases (Summerfield, 1974).

- **Rate Normalization**: The process of taking the rate of speech into consideration when using acoustic cues during speech perception.

- **Speaker Normalization**: Listeners use the pitch of the speech signal as a cue for vocal tract size and make perceptual adjustments on this basis.
Syntactic & Semantic Factors

• Miller, Heise, and Lichten (1951): words in isolation + hissing sound VS. words in five-word sentence + hissing sound

• Result: Performance better in the sentence condition
Syntactic & Semantic Factors

• Miller & Isard, 1963:
  
  • Accidents kill motorists on the highways.
  
  • Accidents carry honey between the house.
  
  • Around accidents country honey the shoot.

• Result: Most accurate with grammatical strings, and the least accurate with ungrammatical strings.

• Top-Down processing: Proceeds from the semantic level of processing to the sensory levels.
Syntactic & Semantic Factors

• *Phonemic Restoration as evidence for top-down processing:*

• The state governors met with their respective legislatures convening in the capital city.

• **Listeners reported hearing the /s/!**

• **Listeners were asked to guess a missing sound: They were unsuccessful.**
Syntactic & Semantic Factors

- Phonemic Restoration as evidence for top-down processing:
    - It was found that the *eel was on the axle.
    - It was found that the *eel was on the shoe.
    - It was found that the *eel was on the orange.
    - It was found that the *eel was on the table.
Syntactic & Semantic Factors

- Interaction of top-down and bottom-up processing:

  - Mispronunciation Detection

  - It has been suggested that students be required to preregister.

- Cole (1973): Detection performance was better for mispronunciations at the beginning of a word compared with those later in a word, and better earlier in a sentence than later on.
Syntactic & Semantic Factors

• *Interaction of top-down and bottom-up processing:*

• Mispronunciation Detection task + *Shadowing task*

• Marslen-Wilson and Welsh (1978)

• *Results:*
  
  • *Restorations were associated with greater fluency than were exact repetitions; in particular, less pausing was observed for restorations.*
  
  • *Restorations tended to occur when the context was highly predictable, but reproductions were more likely with low levels of contextual predictability.*
  
  • *It seems that the fluent nature of the restorations suggests that semantic and syntactic constraints are naturally integrated with incoming speech during language processing.*

• So: Our immediate awareness seems to be a combination of an analysis of incoming sounds with an application of semantic and syntactic constraints.
The TRACE Model of Speech Perception

• By (1986; Elman & McClelland, 1988): it assumes that several levels of processing—distinctive features, phonemes, and words—are simultaneously active during speech perception and interact with each other.

• It assumes:

• There is a cognitive unit for each feature (for example, nasality) at the feature level, for each phoneme at the phoneme level, and for each word at the word level.

• All of these units are activated to a greater or lesser extent, as opposed to being all or none.

• When units are activated above a certain threshold, they may influence other units at the same or different levels.

• These effects may be either excitatory or inhibitory; that is, they may increase or decrease the activation of other units.

• The entire network of units is referred to as the trace, because “the pattern of activation left by a spoken input is a trace of the analysis of the input at each of the three processing levels”
the following images (depicts) are courtesy of Katharine Barden, Cambridge University
Structure of TRACE

OUTPUT: one word (or probabilities for several words)

3 layers:

*words*
- cat
- dog
- deep
- dark
- seep
- lark

*phonemes*
- b
- d
- s
- l
- o

*features*
- voice
- bilab
- alv
- cons
- voc

INPUT: featural information, one time-slice at a time
Structure of TRACE

3 layers:

words

phonemes

features

Each unit connects directly with every unit in its own layer, and in the adjacent layer(s).

Connections BETWEEN layers excite (activate) units in other layers.

Connections WITHIN a layer inhibit (lower activation of) units in same layer.
Mirror Neurons & the Motor Theory

• Doing neurones fire = other neurones fire just from watching the firing neurones!

• Problems!

• Infants, for example, are fully capable of perceiving the differences between many different speech sounds, despite the fact that they are thoroughly incapable of producing those speech sounds.

• Aphasic patients: The existence of clear dissociations between speech perception and speech production provides strong evidence that intact motor representations are not necessary for perceiving speech... If speech perception requires access to intact motor representations, then brain damage that impairs spoken language output should also impair spoken language perception, but this pattern does not appear much of the time.
The General Auditory Approach to Speech Perception

• The general auditory (GA) approach to speech perception starts with the assumption that speech perception is not special.

• “speech sounds are perceived using the same mechanisms of audition and perceptual learning that have evolved in humans ... to handle other classes of environmental sounds” (Diehl et al., 2004, p. 154).

• Studies suggest that our ability to perceive voicing is related to fundamental properties of the auditory system.

• Thus, this aspect of phonological perception could be based on a fundamental property of auditory perception, rather than the peculiarities of the gestures that go into voiced and unvoiced stop consonants.
The General Auditory Approach to Speech Perception

• The fuzzy logical model of speech perception (FLMP), one of the better known approaches within the general auditory tradition, incorporates the idea that there is a single set of “ideal” or “prototype” representations of speech sounds, as determined by their acoustic characteristics.

• A mix of Bottom-up, Top-down processing!