Speech Production and Comprehension

The comprehension and production of speech.
Goals

• 1. Describe the mental processes that convert ideas into sounds. (Production)

• 2. Describe the mental processes that convert sounds into ideas. (Comprehension)
• Basic processes in production:
  – Conceptualization
  – Formulation
  – Articulation

(Griffin & Ferreira, 2006)
Weaver ++ Model of Speech Production

Processor:
1. Conceptual Preparation
2. Lexical Selection
3. Morphological Encoding
4. Phonological Encoding/Syllabification
5. Phonetic Encoding
6. Articulation

Output:
- Lexical Concepts
- Lemmas
- Morphemes
- Phonological Words
- Phonological Gestural Score
- Sound Wave

(Levlt, 1989; Levelt et al., 1999; Roelofs et al., 2007)
• **Lexical concepts:** concepts for which your language has a specific word.
  – E.g., “mare” vs. “female elephant”

• **Lemmas:** a mental representation that incorporates semantic (meaning) and syntactic (combinatory) information.

• **Morphemes:** language “atoms;” smallest unit of language that carries meaning.
  – E.g., *eat vs. eats vs. eating vs. ate*
• Phonological words: a set of speech sounds produced as a single unit (≠ word).

  – E.g., “He will escort us” ➔ / ess-core-tuss/

• Phonological gestural score: a detailed map of the phonemes and syllables needed for an utterance, as well as metrical information, such as stress; used to plan specific motor movements

  – E.g. “banana” ➔ σ σ’ σ

  – vs. “Panama” ➔ σ’ σ σ
Evidence for WEAVER-style models:

1. Speech errors
2. Tip-of-the-tongue experiences
3. Picture-naming studies
I. Speech errors: Errors are not random

“Slips of the tongue can be seen as products of the productivity of language...Word errors create syntactic novelties; morphemic errors create novel words; and sound errors create novel, but phonologically legal, combinations of sounds.”

--Gary Dell, 1986
I. Speech Errors

1. Semantic Substitution Errors

e.g., “dog” for “cat”; “chair” for “table”

(Dell et al., 1997; Noteboom, 1973)
I. Speech Errors

2. Sound Exchanges

e.g., “fig beet” for “big feet”

More common after “phonological primes”

primes: fat beer, fun bed, far base

target: big feet

(Baars & Motley, 1974; Baars, Motley & MacKay, 1975)
2. Sound exchanges

- Most occur within the same phrase
- Most involve a single pair of phonemes
- Most respect the *positional constraint*

Onsets exchange with onsets; codas exchange with codas.

Because planning unit = syllable; positions within syllables are marked.

(Dell, 1986; Noteboom, 1969)
I. Speech Errors

• 3. Word exchange errors:

“My piano plays the girlfriend.”

– Obey the *category constraint*
– Reflects clause-level planning
II. Tip-of-the-tongue

• You have the correct lemma;
• You don’t have all the phonemes

(Brown, 1991; Brown & McNeill, 1966; Rubin, 1975)
II. Tip-of-the-tongue

• Data:
  – Diary methods
  – Prospecting
    • The first name of the character ‘Scrooge’
    • A small boat of the Far East, propelled by a single oar over the stern and provided with a roofing of mats.
    • A secretion of the sperm whale intestine used to make perfume.
    • A one-word name for a person who collects stamps.

(Brown, 1991; Brown & McNeill, 1966; Rubin, 1975)
• A kind of plant used to make paper; esp. in ancient times.

• A Hindu or Buddhist temple or sacred building, typically a many-tiered tower, in India and the Far East.

• The last name of the actor who plays “Fat Bastard” in the Austin Powers movies?

• The word that means mercifully killing a person who is terminally ill and in great pain?
II. Tip-of-the-tongue

• During TOT experience, people:
  – accurately predict whether they will come up with the correct word soon.
  – report the correct number of syllables.
  – accurately report the first phoneme.
  – are more accurate about the beginning and end phonemes than the middle.
  – report words that sound like the target.
  – have more TOT’s for less frequent words.
  – resolve about 40% of TOT’s within a few seconds to a few minutes.

(Nelson, 1984; Lovelace, 1987)
III. Picture-naming

• How do you find the word you need to express a concept?

• How do you activate the sounds that make up the word?
III. Picture-naming

1. People activate different concepts at about the same speed.
   - Picture-recognition speed is constant, regardless of word frequency.
   - Picture-naming depends on word frequency. Less frequent names take longer.

(Oldfield & Wingfield, 1965; Wingfield, 1968)
• 2. Concepts compete for selection.
  – *Picture-word interference*:
    • E.g., name the shape
  – *semantic, identity* and *phonological* conditions

(Dell et al., 1997; Garrett, 1975; Griffin & Ferreira, 2006; Cutting & Ferreira, 1999)
The Spreading Activation Model

• $\text{Weaver}^{++} = \text{feedforward}$

• Spreading Activation $= \text{interactive}$

(Dell, 1986; Dell et al., 1997)
Figure 2.3  Representation of an interactive, spreading activation model for speech production (from Dell et al., 1997, p. 805)
Evidence for Feedback

• 1. *Lexical bias effect*: errors are more likely than chance to produce real words
  
  – Phonological exchanges happen more often when the result is two real words

  Big feet ➔ fig beet ➔ big horse ➔ hig borse

• 2. Errors respect *phonotactic constraints*
These feedback loops help produce semantic substitution errors & the lexical bias effect.
Evidence for Feedback

• 3. *Mixed errors*: word substitutions sound like and have similar meaning to the target

  – E.g., *lobster* for *oyster* > *octopus* for *oyster*

Interpreted as evidence for phonological – lemma feedback during lemma selection.

(Baars et al., 1975; Dell & Reich, 1981; Levelt et al., 1991)
• Potential limitations of lemma theory.

  – Lemma representations include information about a word’s grammatical features – how the word can combine with other words.
  – If so, damage to the lemma should affect spoken and written output in the same way.
  – But, some patients make errors during spoken, but not written production; some make errors during written, but not spoken production.

  (Caramazza, 1997)
Self-monitoring and self-repair

- Self-repair happens after an error.
- Self-monitoring helps prevent overt errors (pre-output monitoring).

- When speakers make an error, they often replace the error with the correct word with no delay, or nearly no delay, between producing the error and producing the correct word.

(Postma, 2000)
• Because speech planning takes time, the plan for the correction must be undertaken as the error is being produced.

• Therefore, the error must have been detected before it was spoken.
• AND: pre-output monitors pay more attention to possibly embarrassing outputs.

  – “tool kits” is less likely to lead to a sound-exchange error than “pool kits”

  – And: galvanic skin response (GSR) higher during (correct) production of “tool kits” than “pool kits”

  – And: Error-detection is greatest when planning load is lightest; i.e., at the ends of phrases & clauses.

  (Hartsuiker & Kolk, 2001; Levelt, 1983; Motley et al., 1982; 1983; Wheeldon & Levelt, 1995; Blackmer & Mitton, 1991; Postma, 2000)
Articulation

• Articulation perturbs air flow to create different patterns of sound waves.

• *Articulatory phonology* theory:
  – Articulation consists of *contrastive gestures*
  – Each contrastive gesture creates a noticeable change in the speech signal (pattern of sound waves)
• **Articulatory phonology** theory (cont.):
  – Speech planning creates a *gestural score* that tells the articulators how to move
    • (1) Move a particular set of articulators
    • (2) Toward a location in the vocal tract where a constriction occurs
    • (3) With a specific degree of constriction
    • (4) Occurring in a characteristic dynamic manner

(Pardo & Remez, 2006, p. 217)
• **Articulatory phonology theory (cont.):**
  – Articulator movement produces *phonemes* (basic speech sounds)
  – Phonemes can be classified according to
    • *Place of articulation*
    • *Manner of articulation*
    • *Voicing*
• Speech production involves *co-articulation*:
  – The gestures for adjacent phonemes overlap in time
  – Gestures are influenced by preceding and following phonemes.

• E.g., “pool” vs. “pan”
• Properties of Sound waves.
  – Sound involves \textit{compression} and \textit{rarefaction}.
  – Each episode of compression and rarefaction constitutes a \textit{cycle}.
  – The amount of time that elapses during a cycle determines the \textit{frequency} of the sound wave.
  – The amount of energy in the sound wave determines the \textit{amplitude}.
    – \textit{Frequency (Hz)} determines pitch; \textit{amplitude (Db)} determines loudness.

• Speech consists of a complex mixture of sound waves; each component has its own frequency and amplitude.
• The pattern of acoustic energy in speech can be represented as a *sound spectrogram*.

• Simplified acoustic patterns are perceived as speech as long as they preserve information about
  
  – Formants
  – Formant transitions

(Cooper et al., 1952; Liberman et al., 1952)
Fig. 1. Showing Spectrogram (A) and Simplified Version (B) of a Spoken Phrase

Liberman et al., 1952
Motor Theory of Speech Perception

• Formants: Steady states
  – $1^{st}$ = Lowest
  – $2^{nd}$ = Next Lowest
  – $3^{rd}$ ...

• Formant Transitions: Rapid changes
• Co-articulation:
  – Speech sounds overlap in time
  – Preceding and following gestures affect target gesture.
  – “pan” vs. “pool”
• Co-articulation
  – E.g., /di/ vs. /du/

  – “What is perceived as the same phoneme is cued, in different contexts, by features that are vastly different in acoustic terms.”

• Alvin Liberman, 1967
• Co-articulation: produces redundancy in the signal.
  – Individual segments provide clues about preceding and following segments.
    • E.g., “silent center vowels”
      – “bag” ➔ “b_g” ➔ perceived as “bag” (rather than “b_g”, “bog”, or “bug”)
    • E.g., “Franken-words”
      – “jo” + “b” ➔ “jog” if “jo” came from (intact) “jog”
Challenges for Speech Perception Mechanism

• Co-articulations: no 1:1 correspondence between acoustic signal and phoneme

• Variability:
  – Between speakers
  – Within speakers over time

• Possible solutions:
  – Motor Theory
  – General Acoustic Theory
Motor Theory

• Speech perception = perceiving a speaker’s intended gestures (Liberman) or speaker’s actual gestures (Fowler)

• Process:
  – Register acoustic signal
  – Determine gestures that produced the signal
  – Deduce syllables/words from gestures

  – Advantage: Relationship between gesture and phoneme closer than relationship between acoustic signal and phoneme.
Motor Theory

• Speech Perception Mechanism = Module
  – Prevents perception of “whistles” and “squeaks”
  – Leads to
    • 1. “duplex” perception
    • 2. “categorical” perception
• Duplex perception: edited speech stimuli are simultaneously perceived as intact, normal speech and “squeaks”

• Categorical perception:
  – Many signals \(\Rightarrow\) few phonemes (categories)
  – Gradual physical change but abrupt perceptual change
Figure 2.7  Simplified acoustic stimuli that are perceived as /da/ or /ga/ (from Whalen & Liberman, 1987). Researchers edited the stimuli so that a formant transition would be played to one ear, while the “base” (the rest of the signal) was played to the other ear. People perceived the stimulus as consisting of a “whistle” or a “chirp” at one ear and the complete syllable (/da/ or /ga/, depending on which formant transition was played) at the other ear.
Motor Theory

• McGurk Effect: non-acoustic information affects speech perception.
  – Because visual and haptic (touch) information can help identify gestures
  – *Tadoma*: The method used to teach Helen Keller how to understand language.
Motor Theory and Mirror Neurons

• Frontal and parietal cortex neurons in Macaque monkeys
  – Fire when monkeys perform and action
  – When monkey sees someone else perform the same action
  – And when monkey hears a sound associated with the action (e.g., peanut-crushing)
  – Area F5 ≈ Broca’s area
Motor Theory and Mirror Neurons

• “Bridge” between production and perception.
  – Perceiving speech sounds ➔ neurons fire
  – Same neurons fire during production
  – The set of neurons firing specifies the *gestural score*
Motor Theory and Mirror Neurons

• Neuroimaging and TMS experiments:

• fMRI:
  – Listening to speech activates speech motor cortex
  – Location of activity depends on how speech sounds are produced (what articulators participate in the gestures)
  – Listening to speech and imagining speaking ➔ similar patterns of neural activity
Plate 2  Patterns of neural activity in response to actual body movements (left side) and words referring to face (smile), arm (throw), and leg (walk) actions (right side). (Hauk, Johnsrude, & Pülvermüller, 2004, p. 304). Neural activity related to face movement appears in green, finger and arms movement in red, and foot and leg movement in blue.
Motor Theory and Mirror Neurons

- TMS (*transcranial magnetic stimulation*):
  - Motor TMS decreases *phoneme discrimination*
  - Increased *motor-evoked potentials* (MEPs) at tongue muscles when listening to tongue-related phonemes
Plate 1  Transcranial Magnetic Stimulation (TMS) (from the National Institute of Neurological Disorders and Stroke: http://intra.ninds.nih.gov/Research.asp?People_ID=196)
Motor Theory and Mirror Neurons

• BUT!
  – TMS ➔ leg MEPs too
  – Legs don’t talk.
  – fMRI/TMS results may reflect response-preparation or a feedback-monitoring loop.
Motor Theory

• Other problems for motor theory:
  – Infants perceive speech sounds that they cannot produce.
  – Japanese quail and chinchillas perceive speech sounds that they cannot produce.
  – Duplex and categorical perception occur in non-speech stimuli.
  – Aphasia: Provides evidence for double-dissociations between speech production and perception abilities (even in bilateral brain-damaged patients).
  – Different gestures ➔ same phoneme (e.g., bite-block vowels).
General Acoustic Theory

• Speech is not special.
• “Cue-based” approach
  – E.g., voicing contrasts: Voicing effects depend on general limits on acoustic processing (i.e., the auditory system cannot distinguish the onset of two auditory events if they start within 20 ms of each other).
General Acoustic Theory

• Fuzzy Logical Model of Speech Perception (FLMP):
  – “ideal”/”prototype” speech sounds
  – Exemplars compared to prototypes
  – Closest prototype wins
  – Experience can alter prototypes
General Acoustic Theory

• FLMP: *Bottom-up* and *Top-down* influences
  – “leam bacon” ➔ “lean bacon”
  – Because “top-down” information favors the real word over the non-word (“Ganong” effect)
General Acoustic Theory

- FLMP, *Top-down* influences: Phonemic restoration
  - “legi(*cough*)latures” → “legislatures”
  - Stronger for longer words
  - Stronger for grammatical, meaningful expressions
  - Changes depending on context
Review

• 1. Production involves processes of conceptualization, formulation, and articulation.
• 2. Models of production differ with regard to information flow.
• 3. Production errors are not random; models must account for patterns of errors.
• 4. Speech perception is harder than it seems.
• 5. Currently, speech perception models disagree about how speech is represented and what the targets of perception are (e.g., motor theory vs. FLMP).