SPEECH PERCEPTION
CHAPTERS 13

OVERVIEW OF QUESTIONS

- Can computers perceive speech as well as humans?
- Why does an unfamiliar foreign language often sound like a continuous stream of sound, with no breaks between words?
- Does each word that we hear have a unique pattern of air pressure changes associated with it?
- Are there specific areas in the brain that are responsible for perceiving speech?

STRUCTURE OF LANGUAGE

- Language is organized into a **hierarchy** of levels:
- At top of hierarchy are _____
- Sentences are composed of _____, which in turn are composed of _____.
- Words are composed of _____ and morphemes are composed of _____.

PHONEMES

- Phonemes are the smallest _____ units in language (**speech sounds**).
  - In English there are 47 (40) phonemes:
    - 13 major vowel sounds
    - 24 major consonant sounds
  - Number of phonemes in other languages varies—11 in Hawaiian and 60 in some African dialects

- **Distinctive features:**
  - ***Voicing:*** vibration of the vocal cords
  - ***Nasality:*** open the nasal passage to allow air to go through
  - ***place of articulation:*** where the air flow is restricted
  - ***voice-onset time:*** when the vibration of vocal cords begins in relation to the word's start, 0 or 60 msec in English).

ALLOPHONES

- Allophones: are _____ of phonemes that are influenced by time, frequency, and coarticulation.
- These sound the _____ to us, but created at different places in the mouth.
- These are the set of phones that sound for a particular phonetic level.
  - Keep Kool
  - Pin Spin (aspirated vs. not aspirated)
  - "Tom Burtton tried to steal a bitter plate of butter. "

Acoustic Signal: Patterns of Pressure Change

- The acoustic signal for speech is created by _____ **past the vocal cords** and into the **vocal tract**.
- The sound that is produced depends on the _____ as air is pushed through it.
- Vowels are produced by _____ and changes in the shape of the vocal tract
- These changes in shape cause changes in the _____ and produce peaks in pressure at a number of frequencies called _____ (harmonics)
- Each vowel sound has a characteristic series of **formants**. The **first formant** has the _____. The **second formant** is the next highest, and so on.

SOUND SPECTROGRAM VOWELS

- **Sound Spectrogram**: indicates the pattern of frequencies and intensities over time that make up the acoustic signal.
- **Intensity** is indicated by _____
- **Vertical Lines** are pressure oscillations caused by _____.

SOUND SPECTROGRAM: CONSONANTS

- **Consonants** are produced by a _____.
- **Rapid shifts** in **frequency** preceding or following formants are called _____ (T2 & T3) and are associated with **consonants**.

WHY IS SPEECH PERCEPTION A PROBLEM?

- **Lack of Invariance/Variability Problem:**

Reasons For Variability Of Speech Sounds

- **Coarticulation:**
- Coarticulation is an example of _____
- Speaker Characteristics: The physical properties of speech sounds, vary according to whether they have been produced by men, women, or children
- Accents
- Within Speaker Differences:
- Speech Rate: rapidly articulated conversational speech.
- Speech Rates: everyday speech is between _____words per minute (WPM).
- Maximum speech rates are limited more by the output capacity of the speaker than by the perceptual needs of the listener.

Reasons for variability of speech sounds

- People perceive speech easily in spite of the variability problems due to perceptual constancy.

Variability of the Speech Signal is a problem for Computers

- **Speaker Independent Systems:** because there is no one-to-one relationship between speech sounds and their acoustic characteristics.
  - speaker independent systems can only recognize speech of varied speakers for a _____
- **Speak Dependant Systems**: These must be **trained** to **recognize voices** (pitch variations, accents) and **pronunciations**.
  - Even after **extensive training** these systems are _____

CATEGORICAL PERCEPTION

- **Iberman, Harris, Hoffman, And Griffith (1957):** This is the experiment that discovered categorical perception.
- **Categorical Perception** of **consonants**: people are better at distinguishing between 2 sounds belonging to different phonetic categories (e.g., ba, pa) compared to 2 sounds belonging to the same category; even when the physical difference is controlled for (VOT).
- Compare consonants that only differ in voice onset time: b & p, t & d, f & v.
  - **voice onset time (VOT)**: when the _____ begins in relation to the **sound's start**: _____ in English

Methodology for Categorical Perception Experiments

- **Identification task**: present stimuli that differ from each other in **Voice Onset Time** (**10 msec steps**).
  - Stimuli are /da/ (VOT of 17ms) and /ta/ (VOT of 91ms)
- **Discrimination task**: the Ss have to indicate whether the pair of stimuli _____
- **Phonetic Boundary**: is the VOT at which the perception _____ from one phoneme to another (e.g., /da/ to /ta/).

CATEGORICAL PERCEPTION

- **Perceptual discontinuity** is demonstrated because the _____ _____ (10 msec) does not have the same perceptual impact at all locations along the continuum: /da/ on one side of the phonetic boundary and /ta/ on the other side.

- **Different languages** =

- Where you grow up determines where the _____.

- **Japanese**: if you do a continuum between **/R/, /W/,** and **/L/** they will say that most of it is _____

  - **English** does not distinguish between **aspirated /P/ and unaspirated /P/**, although other languages such as ***Hindi*** do.

  - **Aspirated P** -- **"puff."**
  - **Unaspirated P -- "spill"**

  - The French language does not have an aspirated P.

- Categorical perception is not unique to humans: *Kuhl & Miller – chinchillas.*

Categorical Perception in Infants

- Infants are **Universal Language Perceivers** –

- **What is happening in that first year?**

MULTIPLE INPUTS USED FOR SPEECH PERCEPTION

- **MULTIPLE INPUTS:**

- Speech Perception is not a strictly bottom-up process.

- **Top-down processes:** conceptually driven, knowledge, and _____.

- McGurk Effect: McDonald & McGurk (1976)
- Phonetic Restoration Effect: Warren (1970)

McGurk Effect
McGurk & MacDonald (1976)

- Demonstrates that we _____ during speech perception.

- Why use visual information?
  - Speech sounds are often _____ so info from speakers face and lips help resolve _____.
- **fMRI measurements:** watching a person's lips make speech movements activates the _____.

CONTEXT EFFECTS
Phonemic Restoration Effect

- Warren & Warren (1970) cough study:

- Procedure: Ss listened to recorded messages in which a fraction of sound was cut out and replaced with an ordinary _____.
  - 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.

- All but one S claimed to have heard all the sounds of the message.

- We rely on syntax and _____ to replace missing sounds.

- _____: people think they hear the phoneme even though the correct sound vibrations never reach their ears.

- **Samuel (1981):** longer words increased the probability of the phonemic restoration effect.

- **Samuel (1990):** phonemic restoration effect is more likely to occur for real words than for pseudo words: prOgress vs. crOgress.

CONTEXT EFFECTS
Pollack & Pickett (1964)

- Typical speech is not clear when presented word by word (w/o context).
- Pollack & Pickett (1964) recorded conversations and then played single words cut out of these conversations to other people for identification.
- Results: single words were correctly identified _____.
- This is ambiguity at the _____.
- The perception of phonemes is _____.

SPEAKER CHARACTERISTICS

- **Indexical Characteristics**: carry information about speakers such as their age, gender, where they are from, their emotional state, and whether they are being sarcastic or serious.

Ladefoged & Broadbent (1957):

- They presented Ss w/ one of six artificially generated versions of **"Please say what word this is"** (pitch varied)
- Followed, by four artificially synthesized words: **bit, bet, but. or bat.**
- Results: the short intro sentence influenced ID of the word **"bit"**
  - when spoke in a relatively high pitched voice, the word bit was correctly identified _____ of the time.
  - when spoken in a relatively low pitched voice, the same word was incorrectly identified as "bet" _____ of the time.
- Conclusion:

SPEAKER CHARACTERISTICS
Nygaard, Sommers, & Pisoni (1993)

- Ss listened to the voices of 10 different speakers.
- Following this voice training they were given a **word intelligibility test**
- Half heard the same voices during the training and on the word intelligibility test.
- The other half heard unfamiliar voices on the word intelligibility test.
- **Results**: Those who heard familiar voices performed better on the test.
- **Conclusion**:

WHY IS SPEECH PERCEPTION A PROBLEM?

- **Segmentation Problem:**
- **Foreign Language:**

- Top-down processing, including knowledge a listener has about a language, affects perception of the incoming speech stimulus.
- Segmentation is affected by context, meaning, and our knowledge of word structure.

MEANING AND SEGMENTATION
Christmas Carols & Other Examples

THE PHYSIOLOGY OF SPEECH PERCEPTION

- **Localization of Function**:

- **Lateralization** – a particular function is processed more strongly in either the left hemisphere or right hemisphere.

  - **Left Hemisphere**:

  - **Right Hemisphere**:

- **Contralateral Conduction**: the process by which each hemisphere receives input from the _____.

- Experiments have found that speech stimuli are **more easily processed** when presented through **earphones to the right ear** than when they are presented through earphones to the left ear. Why?

- **Positron Emission Tomography (PET)**
  - **pitch stimuli** activate the _____
  - **speech stimuli** activate areas in the _____

- **Broca's aphasia** - individuals have damage in Broca's area (in frontal lobe)

  - _____

  - Example: Yes . . . ah . . . Monday ... er ... Dad and Peter H ... (his own name), and Dad ... er ... hospital ... and ah ... Wednesday ... Wednesday, none o' clock. Ah doctors ... two ... an doctors ... and er ... teeth yah. (Patient's effort to explain that he came into the hospital for dental surgery)

- **Wernicke's aphasia** - individuals have damage in Wernicke's area (in temporal lobe)

  - Speak fluently but the _____
  - They also have difficulty _____
  - Example: Well, this is ... mother is away here working her work out here to get her better, but when she's looking, the two boys looking in the other part. One their small tile into her time here. She's working another time because she's getting, too ... (Patient's description of 2 children stealing cookies while their mother's back is turned).