**Chapter 12**

**SPEECH PERCEPTION**

Speech perception is the most vital function of the auditory system. In human speech perception, the stimulus is the voice. The human vocal tract produces both vowel and consonant sounds, which form the bases of sounds used in human language. Formants are the frequency bands with higher amplitudes among the harmonics of a vowel sound, with each vowel sound having a specific pattern of formants. Consonants are produced by restricted airflow. Three physical features are important in determining the sound of a consonant: place of articulation, manner of articulation, and voicing. Phonemes are the basic unit of sound in human language. Linguists use the International Phonetic Alphabet, which provides a unique symbol for each phoneme. The auditory system uses several “tricks” to make fast and accurate interpretations of the speech signal. Coarticulation is the phenomenon in which one phoneme affects the acoustic properties of subsequent phonemes. Categorical perception is our perception of different acoustic stimuli as being identical phonemes until a certain point at which our perception flips to perceive a different phoneme. Voice-onset time refers to the production of certain consonants in which there is a difference between the first sound of the phoneme and the movement of the vocal cords. The McGurk effect is a phenomenon in which vision influences the sounds that a person reports hearing. Word segmentation is the ability of speakers of a language to correctly perceive boundaries between words, providing evidence for top-down processing of speech perception. Finally, the phonemic restoration effect is an illusion in which participants hear sounds that are masked by white noise but context makes the missing sounds apparent, providing further evidence for top-down processing.

Theories of speech perception can be divided into two general classes: general-mechanism theories and special-mechanism theories. General-mechanism theories argue that speech is no different than any other sound and that the same mechanisms that are used for other sounds are also used for speech. Conversely, special-mechanism theories argue that because of the importance of language to humans, special mechanisms have evolved that are specific to speech. Given that nonspeech sounds can influence perception of speech sounds, most researchers support the general-mechanism approach. Speech perception and speech production centers are closely integrated in the brain. Broca’s area is important for speech production. Wernicke’s area is important for speech comprehension. Damage to either of these areas can cause aphasia, an impairment in language production or comprehension, respectively. Finally, remarkable research has found that individual words could be decoded based only on activity recorded from the auditory cortex, though sentences were less well recreated.

Introduction

* Speech perception is the most vital function of the auditory system.
* Speech needs to be understood rapidly.
* In the cocktail party effect, a person is in a crowded room filled with many conversations taking place all at once. She may be listening to the speech of one person but her attention is distracted by the mention her name across the room. This suggests that our attentional mechanism does not screen out all perceptual inputs except the one we are focused on.

The Human Voice as Stimulus

* In human speech perception, the stimulus is a voice. Thus, in order to understand speech perception one must understand the acoustics of human voices.
* Speech sounds are made by movements of the vocal apparatus.
* Speech sounds are acoustic signals, which start as air pressure changes in the lungs that are modified by several structures in the neck and mouth to create the sounds necessary for speech.
  + Air that produces the sound passes through the **trachea** and into the **larynx.**
  + Vocal folds within the larynx open and close to change the pitch of the sound.
  + From the larynx, air passes through the **pharynx** into the mouth and nose, where teeth, tongue, lips, and the **uvula** can affect the sound signal.
  + Closing the uvula prevents sounds from going up through the nasal cavity, affecting the quality of the sound.
  + The sound signal then exits through the mouth and nose.

Vowels and Consonants

* Human vocal tracts make two categories of sounds: vowels and consonants.
* Vowels and consonants have different origins within the vocal tract.
* Vowels are produced by unrestricted airflow through the pharynx and mouth, vibrations of the vocal cords, by vibrations of the vocal cords and by changes in the position of the oral cavity, particularly the shape of the mouth.
  + As we change the shape of our mouth and vocal tract, the resonance of the sound changes, giving each vowel sound a characteristic pattern of harmonics.
  + **Formants** are the frequency bands with higher amplitudes among the harmonics of a vowel sound, with each vowel sound having a specific pattern of formants.
* Consonants are produced by restricted airflow through the pharynx and mouth.
  + Three physical features are important in determining the sound of a consonant.
    - The **place of articulation** refers to the point along the vocal tract at which the airflow is restricted, e.g., tongue, lips, teeth.
    - The **manner of articulation** refers to how the restriction occurs, e.g., whether the lips are pushed together, whether the tongue is at the front or back of the mouth, etc.
    - **Voicing** refers to whether the vocal cords are vibrating or not.

Speech

* The purpose of speech is to convey information.
* **Phonemes** are the basic unit of sound in human language.
  + When a phoneme is changed, the meaning of the utterance is changed.
  + It is important to note that phonemes refer to sounds, not letters.
  + In English and many other languages, convention in spelling does not match phonemes directly.
  + Linguists use the **International Phonetic Alphabet**, which provides a unique symbol for each phoneme.
  + According to this alphabet, there are 15 vowel sounds and 24 consonant sounds in English.

Variability in the Acoustics of Phonemes

* Speech perception requires several “tricks” that help the auditory system make fast and accurate interpretations of the speech signal.

Coarticulation

* **Coarticulation** is the phenomenon in which one phoneme affects the acoustic properties of subsequent phonemes.
  + For instance, the consonant and vowel sounds in “ba” overlap in when they are said. The “a” sound also influences how the “b” is said, and vice versa.
* However, different “b” phonemes are not heard differently.
* This is an example of auditory constancy, similar to the concepts of constancy in vision.

Categorical Perception

* **Categorical perception** refers to our perception of different acoustic stimuli as being identical phonemes up to a point at which our perception flips to perceive another phoneme.
  + Up to a point, the variation in the signal is not heard; only the phoneme is heard.
  + Then, at a certain point, the variation is too much and a different phoneme is heard.
* **Voice-onset time** refers to the production of certain consonants (called stop consonants) in which there is a difference between the first sound of the phoneme and the movement of the vocal cords.
* In a functional sense, categorical perception simplifies the category of what makes each letter and allows us to extract relevant phonemic information from individual differences in speech.

The Effect of Vision on Speech Perception and the McGurk Effect

* Until the invention of telephones, speech without vision was rare.
* Visual cues can greatly influence our perception of speech.
* The **McGurk effect** (McGurk & MacDonald. 1976) is a phenomenon in which vision influences the sounds that a person reports hearing.
  + To demonstrate, participants are shown a video of a person’s mouth saying monosyllabic sounds, such as “ba,” “da,” and “tha.”
  + However, the audio does not always match what the speaker’s mouth was actually saying.
  + The question is whether the participants hear what was dubbed or what the mouth was doing.
  + Interestingly, when participants watch the video, they perceive the sounds as being different from what they actually hear.
* Given that it is an error in perception, the McGurk effect is surprisingly robust.
* According to an fMRI study (Szycik et al., 2012), the integrative area that drives the McGurk effect seems to lie along the superior temporal sulcus.

Top-Down Processing and Speech Perception

* Speech perception is dependent on several top-down processes.
* Knowledge about language affects how we perceive speech.
  + Because we are (presumably) fluent in English, we know where the boundaries between words are, and are thus likely to hear pauses between individual words.
  + Because we may not be fluent in Russian, we do not hear the word boundaries when someone speaks Russian.
* Knowledge of specific combinations of phonemes within a language and knowledge of the context of speech provides evidence for top-down processing in speech perception.
  + For example, we know in English that “s” sounds may be followed by “p” sounds, but not by “f” sounds. If someone says “sfectacular,” we may assume that the person meant to say “specactular.”
* **Word segmentation** is the ability of speakers of a language to correctly perceive boundaries between words.

The Phonemic Restoration Effect

* In the **phonemic restoration effect**, top-down processing of what one expects to hear overrides input from the cochlea.
  + To demonstrate, an experimenter uses a computer to delete or mask a particular sound in a sentence in which the context clearly indicates what the missing sound should be.
  + The experimenter then asks the participant what he/she just heard.
  + For example, the sentence might be “American viewers flocked to the opening \*\*ight of the movie *The Hobbit*,” with the “n” in “night” replaced by white noise.
  + Listeners report hearing the word “night” complete with the “n” sound.
* If a person is engaging in a secondary task, the illusion of hearing the missing sound becomes stronger, indicating that the phonemic restoration effect occurs at ta preattention stage in speech perception processing.
* According to a magnetoencephalography study (Sunami et al., 2013), areas within the auditory cortex in the temporal lobe are involved in the phonemic restoration effect.

Theories of Speech Perception

* Theories of speech perception can be divided into two general classes.
  + **General-mechanism theories** argue that speech is no different than any other sound and that the same mechanisms that are used for other sounds are also used for speech.
    - These theories start with the premise that speech perception occurs through the same neurocognitive processes as other forms of auditory perception.
    - What makes speech unique is its importance, which is learned rather than based on an innate neural mechanism
    - One general-mechanism theory has the view that speech perception involves to-down mechanisms of categorization.
      * When we hear speech, we immediately classify it as such.
  + **Special-mechanism theories** argue that because of the importance of language to humans, special mechanisms have evolved that are specific to speech.
    - These theories start with the premise that there is a unique neurocognitive system for speech perception.
    - For instance, the **motor theory of speech perception** suggests that we have a special mechanism that allows us to detect speech as unique and then relate the sounds to the presumed speech movements of the speaker.
* Most researchers think of speech perception as a form of auditory perception, in general.
  + This comes from studies that suggest that nonspeech sounds can influence perception of speech sounds.
  + However, things become complicated when we try to understand how speech signals, such as phonemes, are extracted from complex and variable stimuli.

The Development of Phoneme Perception

* During the first 6 months of life, an infant can distinguish among all of the phonemes used in a language and can discern subtle differences in sound among equivalent phonemes across languages.
* In fact, 6-month old infants can distinguish sounds that adults cannot distinguish, such as the different “t” sounds in the Hindi language.
* **Perceptual narrowing** refers to the developmental process whereby regularly experienced phonemes are homed in on, as well as the simultaneous diminishing of the ability to discriminate unfamiliar phonemes.
  + As infants get older, they focus attention on relevant stimuli rather than to all stimuli.

Speech Perception and the Brain

* There is close integration of the perception and comprehension networks for language and the brain regions responsible for producing speech.
* **Broca’s area** is an important area in the production of speech, located in the left frontal lobe.
* **Wernicke’s area** is an important area in the understanding of speech, located in the left temporal lobe.
* **Aphasia** is an impairment in language production or comprehension arising from neurological damage.
  + Caused by damage to Broca’s area, **Broca’s aphasia** is characterized by nonfluent speech.
    - Broca’s aphasics have a halted speech pattern and difficulty speaking sentences.
    - Speech perception and language comprehension are largely not affected.
  + Caused by damage to Wernicke’s area, **Wernicke’s aphasia** is characterized by deficits in language comprehension.
    - Wernicke’s aphasics have largely fluent speech but it may not make sense to listeners, because the patients themselves cannot understand what they are saying.
* A **voice area** in the superior temporal sulcus becomes more active in response to human voices than it does to nonspeech sounds.
* Remarkable research by Pasley et al. (2012) found that individual words could be decoded based only on activity recorded from the auditory cortex, though sentences were less well recreated.

# In Depth: Hearing Loss and Speech Perception

* Individuals with hearing impairments choose to wear hearing aids in order to help them understand speech.
* However, when hearing aids compensate for loss of sensitivity, they may introduce distortion by amplifying irrelevant aspects of the signal.
* The combination of the hearing aid and the impaired cochlear results in a number of ways in which speech perception can be affected adversely.
* New hearing aids with faster and more efficient processing are needed.