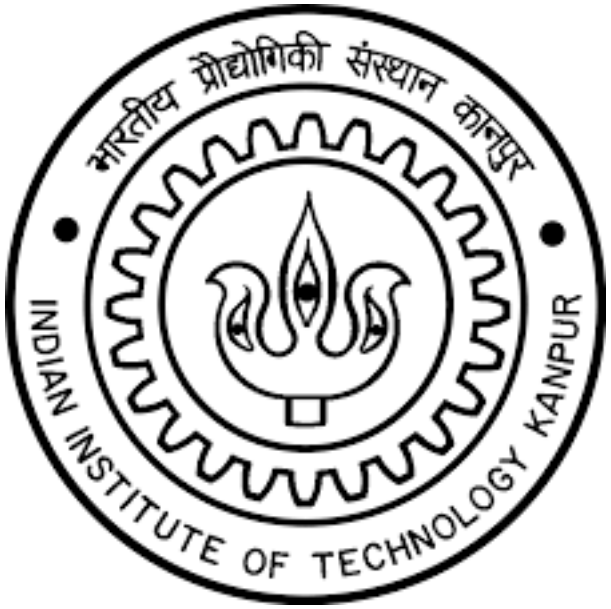

Indian Institute of Technology
Kanpur

In Collaboration
with

National Program on Technology Enhanced Learning (NPTEL)

Presents

Course Title:

Basic Cognitive Processes

By: Dr. Ark Verma,
Assistant Professor of Psychology,
Department of Humanities & Social Sciences,
IIT Kanpur

Lecture 24: Auditory Perception - II

The McGurk Effect

- Acc. to the motor theory of speech perception, understanding speech requires you to figure out which gestures created a given acoustic signal.
- the system therefore uses any sort of information that could help identify gestures.
- while acoustic stimuli offer cues to what those gestures are, other perceptual systems could possibly help out, and if they can, motor theory says that the speech perception system will take advantage of them.

- Infact, two non - auditory perceptual systems - vision & touch - have been shown to affect speech perception.

- The most famous demonstration of *multi - modal perception* is the McGurk Effect (McGurk & MacDonald, 1976).

- The McGurk effect happens when people watch a video of a person talking, but the audio portion of the tape has been altered. for e.g. the video might show a person saying /ga/ but the audio signal is of a person saying / ba/. What people actually perceive is someone saying / da/.

- If the visual information is removed (when the observing individual shuts his/her eyes), the auditory information is accurately perceived and the person hears /ba/

- The McGurk effect is incredibly robust: It happens even when people are fully warned that the auditory & visual information do not match; and it happens even if one tries to play close attention to the auditory information and ignore the visual.

- The McGurk effect happens because our speech perception system combines visual and auditory information when perceiving speech, rather than relying on auditory information alone.

- Of course the auditory information by itself is sufficient for perception to occur, but the McGurk effect shows that the visual information influences speech perception when that visual information is available.

- The McGurk effect is an example of multi -kodal perception because two sensory modalities, hearing & vision, contribute to the subjective experience of the stimulus.

- Another way to create a variant of the McGurk effect is by combing haptic information with auditory information to change the way people perceive a spoken syllable (Fowler & Dekle, 1991).

- This kind of speech perception occurs outside the laboratory from time - to - time in a specialised mode called *tadoma*.

- Hellen Keller & other hearing & vision - impaired individuals have learned to speak by using their sense of touch to feel the articulatory information in speech.

- Acc. to the motor theory, information about speech gestures should be useful regardless of the source, auditory or otherwise.

- That being the case, information about articulatory gestures that is gathered via the perceiver's sense of touch should affect speech perception.

- to test this: Carol Fowler had experimental participants feels her lips while they listened to a recording of a female speaker speaking a variety of syllables.

- Blindfolded and gloved, experimental participants heard the syllable /ga/ over a speaker while CF simultaneously mouthed the syllable /ba/.

- As a result, the experimental participant felt the articulatory gestures appropriate to one syllable but heard the acoustic signal appropriate to a different syllable.

- As in the visual version of the McGurk effect, what participants actually perceived was a compromise between the auditory signal & the haptic signal.

- So, instead of perceiving the spoken signal /ga/ or the mouthed syllable /ba/ they perceived the hybrid syllable /da/.

- the motor theory explains both versions of the McGurk effect, the visual one & the haptic one; as stemming from the same basic processes.

- The goal of the speech production system is not a spectral analysis of the auditory input; rather, it is figuring out what set of gestures created the auditory signals in the first place.

- Motor theory handles the visual & haptic effects on speech perception by arguing that both the modalities can contribute information that helps the perceiver figure out what gesture the speaker made.

- Under natural conditions, the visual, touch & auditory information will all line up perfectly, meaning that all secondary sources will be perfectly valid cues; in conditions as we saw that was not the case.

- ***Mirror Neurons:*** the motor theory has been enjoying a renaissance recently sparked off by new evidence about monkey neurons (Gallesese et al., 1996; Gentilucci & Corballis, 2006).

- i.e. researchers working on macaque monkeys discovered neurons in a part of the monkey's frontal lobes that responded when a monkey performed a particular action, or when the monkey watched someone else perform that action or when the monkey heard a sound associated with that action.

- These neurons were called *mirror neurons*.

- the existence of mirror neurons in monkeys was established by the invasive single - cell recording techniques; and similar experiments in humans are not plausible; so, the existence of mirror neurons in humans remains an hypothesis rather than an established fact.

- However, the part of the brain of the macaques that have the mirror neurons (area F5) is similar to the Broca's area in the human brain.

- Neuroimaging and research involving direct recording from neurons in the Broca's area both show that it participates in speech perception (Sahin et al., 2009).

- Researchers who discovered mirror neurons propose that the mirror neurons could be the neurological mechanism that the motor theory of speech perception requires. i.e. mirror neurons in the Broca's area could fire when an individual produces a particular set of phonemes, or hear the same set of phonemes; providing the bridge between speaking & listening.

- Experiments have been conducted to non - invasively find evidence for the participation of the motor cortex in speech perception.

- the motor theory says the accessing representations of specific speech gestures underlies speech perception.

- those representations of speech gestures must be stored in the parts of the brain that control articulatory movements.

- The parts of the brain that control articulation are the motor cortex in the frontal lobes of the brain & the adjacent premotor cortex when we perceive speech.

- proponents of the mirror neurons argue that mirror neurons are the neural mechanism that establishes the link between the heard speech & the motor representation that underlie speech production.

- Now, mirror neurons have recently been fund in the monkey equivalent of the motor cortex and so, the proponents of the mirror neurons view this as evidence that the motor cortex responds to speech as supporting their view of speech perception.

- Some mirror neuron theorists argue further that mirror neurons play a role in modern humans because our speech production and perception processes evolved from an older manual gesture system (Gentilucci & Corballis, 2006).

- Evidence for mirror neurons in humans:
  - In Pulvermuller & colleagues study, participants listened to syllables that resulted from bilabial stops (/pa/, /ba/) or alveolar stops (/ta/, /da/) on listening trials.
  - On silent production trials, participants imagined themselves making those sounds.
  - Measurement of their brains activity were gathered using fMRI.

- Listening to speech caused substantial activity in the superior parts of the temporal lobes on both sides of the participant's brains, but it also caused a lot of brain activity in the motor cortex in the experimental participant's frontal lobes.

- Further, brain activity in the motor cortex depended upon what kinds of speech sounds the participants were listening tro.
  - whether the sound was a bilabial stop or alveolar stop.

- motor theory explains these results by arguing that the same brain areas that produce speech are involved in perceiving it.

- In another study, when TMS was applied to a participant's motor cortex, participants were less able to tell the difference between two similar phonemes.

- Further, when people listen to speech sounds that involve tongue movements, & have TMS applied to the parts of the motor cortex tat control the tongue; increased MEP are observed in the participants tongue muscles.

- All of these experiments show that the motor cortex generates neural activity in response to speech; consistent with motor theory of speech perception.

- ***Challenges to the Motor Theory of Speech Perception***

  - some challenges to motor theory are rooted in the strong connection it makes between perception & production.

  - infants for example, are fully capable of perceiving the differences between many speech sounds, despite the fact that they are thoroughly incapable of producing those speech sounds (Eimas et al., 1971).

  - to account for this result, we either have to conclude that infants are born with an innate set of speech - motor representations or that having a speech - motor representations is not necessary to perceive phonemes.

- additional experiments have also cast doubt on whether speech - motor representations are necessary for speech perception.
    - no one would suggest, for example that non - human animals have a supply of speech - motor presentations, especially if those animals are incapable of producing anything that sounds like human speech. Two such animals are Japanese Quail & chinchillas.
    - Once they are trained to respond to one class of speech sounds & refrain from responding to another class; they demonstrate aspects of speech perception tat resemble human performance; i.e. categorical perception & compensation for co -articulation.

- because these animals lack the human articulatory apparatus, they cannot have the speech motor - representations; but as they respond to aspects of speech very much like humans do, motor theory's claim that speech motor representations are necessary for speech production is threatened.

- further, research with aphasic patients casts further doubt on the motor theory.
    - Broca & Wernicke showed that some brain damaged patients could not produce speech but understand it & vice - versa.
    - the existence of clear dissociations between speech perception & speech production provides strong evidence that intact motor representations are not necessary for perceiving speech.

- Also, if speed perception requires access to intact motor representations, then brain damage that impair spoken language output should also impair spoken language comprehension; but this pattern does not appear much of the time.

- Another problem for either account is that there is a many -to -one mapping between gestures and phonemes.
    - i.e. the same speech sound can be produced by different articulatory gestures (MacNeilage, 1970).
    - more specifically, different people can produce the same phoneme by using different configurations of the vocal tract; because the vocal tract offers a number of locations where the air flow can be restricted & because different combinations o air - flow restrictions have the same physical effect; they wind up producing acoustic signals that are indistinguishable to the perceiver.

- this means that there is no single gesture for syllable like /ga/.
- Studies involving the production bite - block vowels also show that very different gestures can lead to the same or nearly the same set of phonemes.
- The motor theory can account for this set of findings in one of two ways:
    - either by proposing that more than one speech - motor representation goes with a given phoneme or that there is a single set “prototype” of speech -motor representations & that an acoustic analysis of speech signals determines which of these ideal gesture most closely matched the acoustic input.
- Both, violate the spirit of the theory!

Other Theories of Speech Perception

- **The General Auditory Approach to Speech Perception**
  - starts with the assumption that speech perception is not special (Diehl & Kluender, 1989; Pardo & Remez, 2006); instead "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).

- Researchers in this tradition look for consistent patterns in the acoustic signal for speech that appear whenever particular speech properties are present.

- further, they seek to explain commonalities in the way different people and even different species react to aspects of speech.

    - for e.g. some studies have looked at the way people and animals respond to *voicing contrasts* (the difference between unvoiced consonants like /p/ and voiced consonants like /b/).

    - these studies have suggested that our ability to perceive voicing is related to the fundamental properties of the auditory system.

- i.e. we can tell whether two sounds occurred simultaneously if they begin more than 20ms apart.
    - if two sounds are presented within 20 ms of each other, we will perceive them as being simultaneous in time. if one starts 20ms before than the other, we perceive as occurring in a sequence, one before the other.
    - the voicing boundary for people & quail sits right at the same point.
    - if vocal fold vibration starts within 20ms of the burst, we perceive the phoneme as voiced; but if there is more than a 20ms gap between the burst & the vocal fold vibration, we perceived an unvoiced stop.
- 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 the voiced & unvoiced consonants.

- the general auditory approach does not offer an explanation of the full range of human (or animal) speech perception abilities.

- it's chief advantage lies in its ability to explain common characteristics of human & non - human speech perception & production.

- since the theory is not committed to gestures as the fundamental unit of phonological representations, it is not vulnerable to many of the criticisms leveled at the motor theory.

- ***The Fuzzy Logic 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 (Massaro & Chen, 2008).

- Acc. to the FLMP, speech perception reflects the outcomes of two kinds of processes:

  - *bottom - up & top - down*: the bottom up processes are those mental operations that analyse the acoustic properties of a given speech stimulus. these processes activate a set of potentially matching phonological representations

- stores representations of phonemes are activated to the degree that they are similar to acoustic properties in the speech stimulus; more similar phonemes attain higher degrees of activation, less similar phonemes attain lower degrees of activation.

- top - down processes are this mental operations that use information in the long - term memory to try & select the best possible candidate from among the set of candidates activated by the bottom up processes.

    - this may be specially important if the incoming information is ambiguous or degraded. for e.g. when the /n/ phoneme precedes the /b/ sound (as in *lean bacon*), often times coarticulation makes the /n/ phoneme comes out sounding more like /m/.

- So, when someone listens to *lean bacon*, bottom - up processes will activated both the prototype /n/ & /m/ phoneme, because the actual part of the signal will be intermediate between the two types.

- Acc. to the FLMP, our knowledge the *lean bacon* is a likely representation in English should cause us to favour the /n/ interpretation.

- However, if the /n/ sound were in a non - word, such as *pleat bacon*, a listener would be more likely to favour the /m/ interpretation, because the opening sound would not receive any support from top - down processes. This tendency to perceive the ambiguos speech stimuli as real words if possible is known as the *Ganong Effect,* after William Ganong (1980).

- FLMP, also offers a mechanism that can produce *phonemic restoration effects* (Sivonen et al., 2006).
    - phonemic restoration happen when speech stimuli are edited to create gaps. for example. remember the legi(cough)lators experiment.
    - these phonemic restoration effects are stronger for longer than shorter words and they are stronger for sentences that are grammatical and make sense than sentences that are ungrammatical & don't make sense.
    - further, the specific phoneme that is restored can depend on the meaning of the sentence that the edited word appears in.

- for e.g. if you hear *The Wagon lost its (cough)eel,* you will most likely hear the phoneme /w/ in place of the cough. But if you hear *The circus has a trained (cough)eel,* you will more likely hear the phoneme /s/.

- Research involving ERPs show that the nervous system does register the presence of the cough noise very soon after it appears in the stimulus (about 200ms).

- All of these suggest that a variety of possible sources of top - down information affects the way the acoustic signal is perceived.

- Further they suggest that the perception of speech involves analysing the signal itself as well as biasing the results of this analyses based on how well different candidate representations fit in with other aspects of the message.

- These other aspects could include whether the phonological representations results in a real word or not, whether the semantic interpretations of the sentence makes sense or how intact the top - down information is.

To Sum Up

References

- Traxler, M. J. Introduction to Psycholinguistics: Understanding Language Science. *Wiley – Blackwell*.