Chapter – 4: Speech production

- We human talk to each other most of the time and even when alone we talk to ourselves.

Vocal tract and speech production

- We human by blowing air out of our mouth, communicate complex thoughts to other member of our species.
- Speech production begins in lungs, from where air stream flows up the trachea through the glottis (voice box).
- The vocal folds consisting of a pair of membranes stretched across the opening of the glottis that can be vibrated to produce sound.
- When vocal folds are retracted, the air stream flows unimpeded for more and breath.
- The voice tract consists of oral and nasal cavities, and they serve resonating chambers for the phonation produced by the vibrations of the vocal cord.
- In English, vowels are generally produced by directing the air flow through the mouth, producing resonation in the oral cavity.
- *Consonant are produced by obstructing the flow of air through the oral cavity.
- Three factors determine consonant quality are a) place of articulation, b) manner of articulation, c) voicing (voice onset time).
- Place of articulation describes the location along the vocal tract where the obstruction occurs to produce a consonant.
- Important places of articulation include the lip, teeth, alveolar ridge, hard palate and velum.
- A consonant sound that is produced by bringing the upper and lower lips together is called a bilabial [e.g. purse lips before letter air go, *u* get the puff of "*p*" sound].
- A consonant that is produced by bringing the lower lip against the upper teeth is labia dental [e.g. repeat the syllable few, view].
- A consonant produced by protruding the tongue between the upper and lower teeth is inter dental [e.g. place blade of tongue against upper teeth and let some air come hissing "*th*" sound]. thy, thisth
- A consonant produced by pressing the tip of the tongue against the fleshy area behind the upper teeth is an alveolar consonant [e.g. new, dew, two, zoo].
- A consonant produced by pressing the blade of the tongue against the region between the alveolar ridge and the hard palate is called post alveolar [e.g. gin, chin, shin].
- Consonants produced by pressing the root of the tongue against the soft palate against the back of the mouth are referred to as velar [e.g. sag, sang, sack]. goo, cooing sound in sing

→ produced by constricting vocal folds
→ English hay ; middle sound in uh-oh, uhn-uhn is called glottal stop

Vocal tract, Major Features

(1) Nasal cavity → Resonation produces nasal consonants like m and n, nasal "twang of some vowels.

(2) Oral cavity → Most speech sounds produced inside the mouth.

(3) Alveolar ridge → Fleshy region behind upper teeth

(4) Hard Palate → Bony region along roof of mouth.

(5) Velum (Soft palate) → Fleshy region behind hard palate

Consonants: Voicing

(1) Voicing → Early or late voice onset time (VOT)

(2) Voiced → Early VOT, English boo, dew, goo

(3) Voiceless → Late VOT, English bee, tea, key

English Consonant Inventory

① Manner →

② Nasal →

③ Plosive →

④

Vowels

(1) Jaws → Continuum from high (open) to low (closed)

(2) Tongue → Continuum from front to back

(3) Lips →
- Unrounded with front vowels in English
- Rounded with back vowels in English
- Many languages use lip rounding as separate from tongue position.

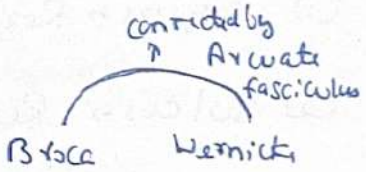
- Nasal → Produced by blocking oral cavity, releasing airflow through nasal cavity.
- Final sound in English sum, sun, sung

- **Manner of articulation describes the degree to which the airflow is obstructed in the production of the consonant.
- The obstruction can involve a complete stop, a constriction or a diversion of the airstream.
- Plosives is a consonant produced by momentarily stopping and their releasing the air flow [*b-d-g, p-t-k*]. English:- bat, pat, dew, two, gill, kill
- A consonant that is produced by momentarily blocking the airflow and then releasing it through a light constriction is called affricate [*j*, and *ch*]. Eg gin, chin
- A consonant produced by passing an airflow through a constriction in the oral cavity is a fricative [*f*-few, *th* of thy, *z* of zoo or sue]. Eng → view, few, thy, this, zoo, sue, version, vigin
- An approximant is a consonant that is produced by diverting the airflow without constricting it [*l, r, y, w*]. approximates are turbulence to airflow and hence called semi-vowels. eg yes, way, ray, lay
- Vowels are produced by modifying the shape of the oral cavity.
- Three factors influence the shape a) the height of the jaw, b) the position of tongue and c) shape of lips.
- ***A diphthong is a vowel combination that is perceived as single phoneme.
- English has 3 (as in high, hoy, how)

  (*ah-ee oh-ee ah-oo*)

  Chinese has more (including some 3 vowels combination).

Speech area of brain



- According to the traditional Wernicke – Geschwind model,
  a) Wernicke area is responsible for speech perception.
  b) Broca's area is responsible for speech production.
  c) The arcuate fasciculus, a band of fibers extending from Wernicke to Broca's areas, connect speech perception and production.
- The Wernicke's – Geschwind model explains three common forms of aphasia (speech problem)
  a) Expressive or Broca's aphasia – a condition in which brain damage leads to a loss of speech production without a loss of speech comprehension.
  b) Receptive or wernicke's aphasia – a condition in which brain damage leads to a loss of speech comprehension and fluent but meaningless speech production.

Wernicke - Geschwind Model of Aphasia:-

Read Notes:-

Three common Aphasia:-

Read Notes

* Expressive Aphasia → Loss of speech production without loss of speech comprehension.
  - → Results from damage to Broca's areas
  - → Also called Broca's Aphasia

* Receptive Aphasia :- Loss of speech comprehension & fluent but meaningless speech production.
  - → Result from damage to Wernicke's area
  - → Also called Wernicke's aphasis

* Conduction Aphasia:- Preserved speech Reception and production but difficulty in repetition.
  - → Results from damage to arcuate fasciculus.

Cerebral Cortex - General Features

(1) Gyrus → Region of cerebral cortex that protrudes outward

(2) Sulcus → Region of cerebral cortex that is folded inward

(3) Longitudinal fissure - Deep groove separating left & right hemispheres.

(4) Lateral Sulcus → Deep fold that separates temporal lobe from frontal & parietal lobes.
  - → Also called Sylvan fissure
  - → Region inside & around lateral sulus known as Perisylvan Cortex

---

Cerebral Cortex - Functional Regions (1)

(1) Somatosensory cortex (parietal lobe) → Processes body senses, keeps track of what body parts are doing, including articulators.

(2) Primary motor cortex (frontal lobe) → Programs commands to move body, including articulators.

(3) Primary auditory cortex (temporal lobe) → Processes sensory i/p from ears, including speech. →

c) Conductive aphasia – which is a language disorder characterized by preserved speech perception and production capabilities but with a marked difficulty in repeating spoken language.

- Other areas of the cerebral cortex are also implicated in speech production, including the somato.
- Sensory cortex in the parietal lobe, the primary motor cortex in the frontal lobe, the supplementary motor area and the anterior cingulate cortex in the lateral fissure and the anterior insula in the lateral fissure. Sulcus ! Longitudinal
- Many areas of the brain responsible for walking also play a role in the production speech. - Motor loops like primary motor cortex & subcortical structures -
  a) These areas include the basal ganglia, which are involved in initiating movement.
  b) The cerebellum, which is involved in coordinating movement.
  c) Thalamus plays role in coordinating motor programs for speech production
- Dysarthria is a speech disorder that result from damage to motor area of the brain.
  a) It is characterized by the poor articulation of phonemes, irregularities in prosody, and a slow rate of speech.
- While the higher levels of language processing are mainly lateralized to the left hemisphere, neuroimaging and clinical data indicate that wide regions of both hemispheres of the brain get involved in speech production.

Lect 2

Models of speech production

- Current model of speech production is built on recent findings in neuroimaging research as well as on clinical data.
  a) Many areas of the brain in both cerebral hemispheres as well as subcortical structures are involved in speech production.
- Speech production recruits many brain areas involved in moving the limbs.
  a) The motor system is organized into feed forward and feedback control system.
  b) The feed forward system generates the overall movement plan and the feedback system provides information so that adjustments can be made.
- Jaw perturbation studies and auditory perturbation studies show that somato sensory feedback is already available during the current articulation but that auditory feedback can only influence the production of subsequence utterances.
- Forward models propose that motor system, in addition to generating a motor plan, also generates an expected sensory consequence of the motor plan. If the actual sensory

④ Primary visual cortex (Occipital lobe)
→ Process sensory i/p from eyes
→ Temps not only for reading but also for f2f speech perception

Cerebral Cortex – Functional Region (II)

① Supplementary motor region
- → Longitudinal fissure
- → Programming intentional actions.

② Anterior cingulate cortex
- → Longitudinal fissure
- → Detecting errors & monitoring conflict

③ Anterior insula
- → Dip within lateral sulcus
- → Implicated in lang processing

Minimal N/w for overt Speech Production
(Riecker et al. 2008)

Starting mechanism

→ Supplementary motor cortex initiates speech motor movements
→ Anterior cingulate cortex monitors for error.

Generation of Phonetic Plans

→ Broca's Area → anterior insula involved in preplanning of speech articulation.
→ Primary motor cortex → assembles motor plans for articulators

Lecture - 2

Stuttering

* Feedforward & Feedback control

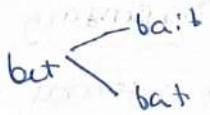
(1) Feedforward → Provides general motor plan for moving body part toward goal.

(2) Feedback control →
- Adjust forward trajectory based on real time info about likely success of movement.
- Motor system receives rapid feedback from somatosensory systems (muscles, tendons, joints)
- Feedback in speech production
  (i) Somatosensory
  (ii) Auditory.

Jaw Perturbation Technique

Tests somatosensory feedback in speech production.

Robot arms -
(1) Attached to participant's jaw
(2) Applies upward or downward force during vowel articulation
(3) Participant rapidly adjust to perturbation, produces intended acoustic signal, not gesture.



Auditory Perturbation Technique

Test auditory feedback in speech production

Technique -
(1) Participants speak into microphone, listen through headphones.
(2) Computer modifies their speech.

Participants gradually compensate for sound shift.

→ Auditory feedback much slower than somatosensory feedback.
→ Sound exits mouth & re-enters through ears (slower feedback)
→ Somatosensory info travels within nervous sys (faster feedback)
Time lag too quick to effect current production but influence subsequent productions

Auditory Suppression During Speech

Auditory suppression
- General principle of sensorimotor system
- Expected sensory effects of self-initiated action are attenuated.

Delayed auditory feedback technique
- Participants speak into microphone, listen through headphones.
- Auditory return delay by fraction of sec
- Even 50 millisec can severely disrupt speech.
- Normal auditory feedback causes no disruption because it is expected, therefore suppressed.

Forward Model

Model that explains sensory suppression.

Sensorimotor system:-
- Generates motor command
- Also generates predicted sensory consequence.

Comparison b/w prediction & i/p

(1) Match → Signals correct o/p, sensory experience dampened

(2) Mismatch → signals production error, sensory experience intensified.

(3) Input without prediction → Signals external event, sensory experience intensified.

Dual Stream Model

It is a general organising principle of sensorimotor system.

Two streams

Ventral Stream
through temporal lobe process "what" info (Object identification)

Dorsal Stream
through parietal lobe process "how information" (navigation)

Dual stream model of speech processing (Hickok & Poeppel, 2007)

Ventral Stream → Bilateral processing pathway, interprets meaning of incoming speech signal.

Dorsal stream → left-hemisphere processing pathway, links incoming speech signal with motor programs.

[Both streams are active in speech production & speech perception]

DIVA → other model

→ It is a computational model
  ↳ Comp program, simulates cognitive process
  ↳ consistent with what is currently known about human cognition.

DIVA →
- Models both speech production & speech acquisition
- Organises functional brain regions into feedforward & feedback control system.

Contributions of DIVA
- Reinterprets Broca's area as a speech sound map.
- Explains both adult speech production & infant speech development in same framework.

Babbling

→ Infants go through predictable stages of babbling during first year
- Regardless of lang spoken
- Even with profound hearing loss.

→ Caregivers often mimic infant vocalizations.
- Provides social feedback encouraging infants to babble more
- Helps infants hone their vocalizations

→ As babbling becomes more speech-like :-
- Caregiver respond as if it were intentional speech.
- Baby's first words to develop out of babbling sequences.

Stages of babbling

Stage	Months	Characteristics
1) Phonation	0–2	Vowel-like sounds made by vibrating vocal folds
2) Cooing	2–4	Syllable-like sounds in back of vocal tract
3) Expansion	4–6	Variety of different sounds
4) Canonical Babbling	6–12	Sequences of clearly formed consonant-vowel syllables

---

Frames-then-Content Model

Explains babbling in terms of repeated jaw movements
→ Jaw oscillation + vocal fold vibration ⇒ Syllable like vocalizations
→ Certain speech sounds more basic than others, appear first in babbling
- Labial consonants :- b, m, w
- Alveolar consonants :- d, n, j
- Velar consonants :- g, ng
- Basic vowels - aa, ee, oo

→ Infant speech sound production highly variable, but centered on these basic sounds.
→ All consonant-vowel combinations occur, but certain combinations more likely.

input matches the expectation, the sensory experience is attenuated, and if there is a mismatch, it is intensified.

- The dual stream model of language processing proposes that a ventral stream from auditory cortex to the temporal lobe interprets the meaning of the incoming speech signal while a dorsal stream from auditory cortex to the frontal lobe links the incoming speech signal with speech motor programs.
- DIVA is a computational model that incorporates the latest neuroimaging data. It conceptualizes Broca's area as a speech sound map linking perception and production. It can also account for infant speech development and adult speech production with the same mechanisms.

Development of speech production

- During the first six months infants produce a variety of speech like sounds, but with the onset of canonical babbling, the child begins producing clearly perceivable consonant-vowel syllables.
- The frames- then -content model views babbling as driven by the motor system without auditory input coupling of jaw movement with phonation.
  a) As a result some consonant-vowel pairs are more likely to occur in canonical babbling.
- Care givers imitate their baby's babbling, and their social feedback helps the infant refine its production of the phonemes used in the language its learning.
  a) Caregiver's attempts to match baby's babbles with the names of object helps the infant learn its first words.
- Delays in babbling or first words can indicate an underlying disorder such as hearing loss or apraxia of speech. Infants with hearing loss will show some sign o babbling but fail to progress toward clearly articulated canonical syllable.
- Childhood apraxia of speech is a condition in which children have difficulty producing speech despite having cognitive and motor functioning in the normal range.
  a) These children often need help from a speech language pathologist.
- Some children develop expressive language more slowly than receptive language but eventually catch up with their peer.
  a) Other children have mostly normal language development except for particular speech sound errors that can persist well into elementary school or later.

Fr pronouncing child can clearly hear a distinction b/w two phonemes but use one of them while speaking

Typical Canonical Babbling Syllables

Tongue Position	Vocal Tract Configuration	Typical Syllables
Central	Mouth open, lips rounded	baa, maa, waa
Front	Tip of tongue strikes alveolar ridge	dee, nee, yee
Back	Root of tongue strikes velum	goo, ngoo

Social Aspects of Babbling

Caregivers

- Respond to babbling as if it were an invitation to conversation
- Imitate babbling within confines of own lang.

Infants use this feedback, modify babbling to sound like caregiver's lang.

Object directed vocalization

- Babbling uttered as infant approaches & manipulates novel object
- Caregivers respond with name for object the assembly the babble
- Indicates infants heightened attention, readiness to learn.