

Chapter - 3: Speech perception

Haskins laboratory

- Pattern playback machine (Liberman, 1996)

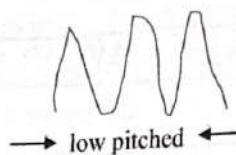
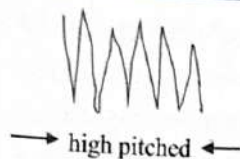
Auditory perception:

The look of sound

- Two basic properties describe a wave:

- Frequency (no. of wavelengths that pass by a given point in a given amount of time)
[cps or Hz]. *Perceived as pitch.*
- The psychological perception of sound wave frequency is called pitch.

CPS → cycle
per
second

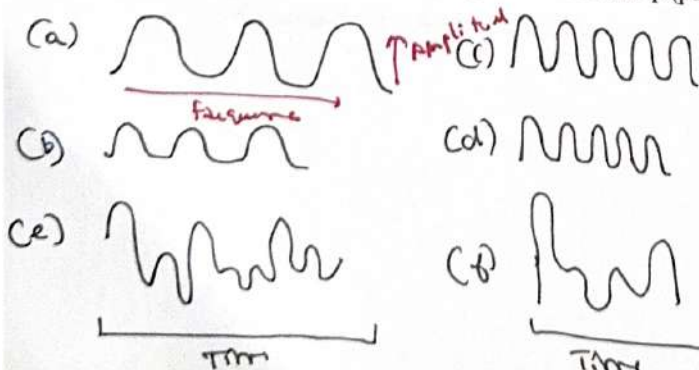


- Amplitude: amount of change that a wave undergoes during one cycle [different but highest and lowest pressure]. *Perceived as loudness*

- The psychological perception of sound wave amplitude is loudness.

These 2 properties describe a sine wave, so called because it's a wave that can be described by a trigonometric sine function.

- Vibration along the entire length of the string generates the fundamental frequency [lowest frequency produced by a vibrating object].
- Vibrations at half length, thirds length and so on generate overtones [frequencies higher than the fundamental that are also produced by a vibrating object].
- The psychological perception of sound wave complexity is known as timbre.
- A vibrating object produces a sound with a regularly repeating pattern, which known as a periodic sound.
- Two objects rubbing against each other produce a sound with no regularly repeating pattern, the periodic sound.
- Speech sounds can be broadly categorized into vowels and consonants.
- Vowels are periodic (ringing, musical character to them).
- Consonants are periodic (noisy), e.g. sh consonant are produced without function, but there is a limit [try saying sh..... and then p (puff of air)].



- Waveform (a) and (b) have same frequency, but diff amplitudes, as do (c) & (d).
- Waveforms (a) & (c) have same amplitudes but different frequency, as do waveforms (b) & (d).
- Waveform (e) is a periodic noise, while waveform (f) is one cycle of a periodic musical note played on a clarinet.

* Periodic Sound

- Regularly repeating pattern
- Produced by vibrating object
- Perceived as ringing or musical
- Vowels as periodic speech sounds

* Aperiodic Sound

- No regularly repeating pattern
- Produced by collision or friction
- Perceived as Noise
- Consonants as aperiodic speech sounds

THE SPEECH STREAM

(1) Phonation → Sound resulting from vibrations of the vocal folds as air is expelled from lungs.

(2) Prosody →

- Fluctuations in fundamental frequency during an utterance
- Conveys both linguistic & emotional information.

The auditory cortex is tucked deep inside the lateral fissure on the surface that is still considered to be part of the temporal lobe.

Sound to thought

- The organ of auditory sensation is the cochlea of inner ear. The basilar membrane extends inside the length of the cochlea, and its hair cells are sensitive to particular frequencies. *Progressive arrangements of cells sensitive to different frequencies*
- The basilar membrane exhibits tonotopic organization, with sensitivity to high frequencies at one end and low frequencies at the other end.
- The primary auditory cortex is located in the superior temporal lobe, and it has the same tonotopic arrangements as the basilar membrane.
- Neighbouring cortical regions, including Wernicke's area in the left hemisphere do higher-level processing of the auditory input.
- We can recognize objects and events just by the way they sound.
- Speech sounds are likewise auditory events and of extremely short duration, yet our ability to perceive them accurately is remarkable.
- Auditory perception unfolds along time dimension.

The speech streams

- Written text has discrete letters with each word separated through spaces from immediate neighbours.
- Speech stream doesn't consist of discrete phonemes and clear word loudness.
- Speech is continuous flow of ever changing frequency and amplitudes.
- Speech perception system infer intended phonemes and word boundaries.
- Spectrogram allows us to visualize the structure of speech stream.
- Spectrogram of speech sample consists of alteration of a periodic and periodic sounds roughly corresponding to consonants and vowels of the language.
- At bottom of spectrogram the fundamental frequency of speech is displayed.
- This is the fundamental frequency of the sound resulting from vibration of vocal folds as air is expelled from lungs (phonetics). *o-Phonation*
- Rate of vibration of vocal folds result in fluctuations of the fundamental frequency during an utterance, and this variation in pitch is called prosody.
- Prosody serves a number of linguistic function as well as conveying information about the speaker's emotional state.
- Periodic stretches of the speech stream reveal bands of high amplitude sound at certain frequencies above fundamental frequency.

Formants

- Bands of high amplitude sound at certain frequencies above fundamental frequency.
- Occur during periodic stretches of speech stream.

Formant Transition

- Modification of formant due to preceding or following consonant.

Major categories of speech sounds

Sound class	Speech Sound	Characteristic	Examples
① Periodic	Vowels	Distinguished by first two formants	I owe you a go-go
	Sonorants	Sometimes vowels, sometimes consonants	little, river
② Aperiodic	Fricatives	Constricted air stream, producing friction	Suzie has a fever
	Plosives	Air stream momentarily blocked then released	Pay Kay today

[The high energy segments are vowels, & the 'silent' segments are consonants]

eg

(a)

high

tt

(i)

high

t

(u)

high

de

- These bands are called formants and result from the fact that shape of vocal tract dampens certain harmonics and enhance others.
- The relative distance between the 1st and 2nd formants in comparison to the fundamental frequency is used to distinguish vowels. e.g. "ah" (vowel sound) – same pitch relationship.
- A periodic stretch in the speech stream may also signal a sonorant. ^{→ sometimes vowel sometimes consonant}
- This is a speech sound that usually serves as a consonant but sometimes as a vowel (l, r, n and m). e.g. the 1st l in little acts as consonant but 2nd l acts like a vowel.
- An aperiodic portion of the speech stream clearly indicates a consonant.
- The first type is fricative, which is a consonant that is produced by constricting the air stream to create friction (hissing sounds like s, sh and f).
- The second type is plosive, which is a consonant that is produced by momentarily blocking and then releasing the air stream (p, d, t "stop").

The sound of silence

- speech sounds do not occupy discrete sections of the speech stream, but rather they overlap each other in a process known as co-articulation. ^{• preceding or following consonant modifies vowel form}
- Speech sounds are perceived categorically even though they are produced differently depending on the context.
- Originally categorical perception was believed to be a unique characteristics of speech perception. It was soon understood as a general cognitive principle.
- Simply put categorical perception in one way our brain deal with the messiness of the real world.
- The speech perception system relies on context to fill in missing information from a speech stream that has been masked by ambient noise, in a process known as phonemic restoration. ^{or with contextually appropriate material by Warren 1970}

Warren and Warren (1970) modified the following sentence by splicing out the wh of wheel:

It was found that the *eel was on the axle.

* — replaced by cough, also axel was replaced by shoe. Shoe was replaced by orange.

Aspiration → Putt of air accompanying some plosives

Distinguishes p from b, t from d, k from g

measured by VOT

VOT → Difference b/w release of plosive (consonant) & beginning of vocal fold vibration (vowel)

Voicing onset time

It is the difference b/w release of plosive consonant & beginning of vocal chord.

* Development of Speech Perception - Long learning in Womb

(1) During 3rd trimester, fetus can hear

- Mother's voice
- Environmental sounds

(2) Change in Fetal heart rate indicates

- Fetus can discriminate stimuli

(3) At weeks 35 - 37, fetuses can recognize

- Mother's voice
- Mother's language (spoken by another woman)
- Familiar nursery rhyme.

Born Yesterday

High amplitude sucking technique:-

- (1) Measures frequency of infant's sucking on non nutritive nipple
- (2) Changes in frequency indicate discrimination of stimuli.
- (3) Used with new borns.

Newborn's Preference

→ Same

[Clearly, newborns remember what they heard in womb]

Baby Talk

Infant directed Speech

- Manner of speaking to infants, attracts their attention, helps them learn language.
- ^{Phonetic} Fundamental frequency, greater range of pitch
- Exaggerated stress contrasts and vowel length
- Nearly universal phenomenon.
- Also called **motherese** or **Caregiver Speech**

Read my lips

- Psychologists used to assume that the senses operated independently of each other.
- In recent year, researchers agree with the idea that the sense strongly interact with each other to produce our rich experience of the world.
- The McGurk effect is an artificially induced illusion in which the auditory information for one speech sound is combined with the visual information for another speech sound to produce the perception of a third speech sound.

acc 2

Development of speech perception

- Language learning begin in the womb during the 3 trimester which extend from week 28 until birth at week 38 or later.
- A foetus at 3rd trimester can hear and respond to sounds in the environment particularly its mothers voice.
- At birth a new born can already distinguish its mother's voice from other women's voices and its mother's language from other language.
- Infant directed speech is spoken with a higher fundamental frequency, a broader pitch range, and exaggerated information and stress pattern.
- These features attract the infant attention and provide cue to speech stream segmentation.
- The prosodic boot strapping hypothesis proposes that infants use intonation and stress pattern to infer phrase and word boundaries.
- There are plenty of evidence to support the notion of prosodic boot strapping. Infants are sensitive to indication of phrase boundaries in motherese.
- In ERP study 5 month old German infants were found to be sensitive to acoustic cue to phrase boundaries (pitch change, vowel duration).
- According to Metrical segmentation strategy, both infants and adults tend to segment the speech stream at the onset of stressed syllables at least in English. • Rule of thumb for segmenting speech stream
• used by infants & adults
- English has a characteristic rhythmic pattern of alternating stressed and unstressed syllables. Stressed syllables have long or complex vowel sounds, while unstressed syllables have reduced or shortened vowels.
- Morgan (1996) presented 2 syllable sequences with stress on either 1st or 2nd syllables to six and nine months infants. They found that while 6 months infants had no preference the 9 months infants preferred 1st stress syllable.

Conditioned head turn technique - Infant trained to turn head when it detects change in stimulus.

Narrowing Down

① Perceptual Narrowing

- Transition from broad to narrow perceptual categories
- At birth, infants perceive all possible speech categories.
- By the 1 year, they only perceive categories in their language.

② Distributional Learning

- Track frequency and location of various sounds in speech stream
- Aids infants in organising perceptual categories of their language.

① Approach → Motor theory

* Nativism → View that behaviour is mainly shaped by natural selection, encoded in genes.

* Language acquisition device (Chomsky, 1959)

- Specialized processing units in brain
- Guide rapid development of lang in infants.

* Module (Fodor 1983)

- Dedicated neural system evolved to perform a specific function.

* Speech is Special

- View of speech ~~perception~~ perception as distinct from General Auditory Perception.

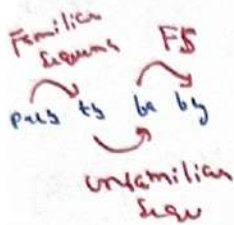
* Motor theory

- People perceive speech by inferring articulatory gestures, not analyzing speech stream.

- Transitional probabilities provide a fairly reliable cue to word boundaries that even young infants are sensitive to.

The likelihood that a particular event will occur next given the current event is known as transitional probability (TP).

- In infants the TP from *pret* to *ty* is very high and likewise from *ba* to *by* is high. however, the TP from *ty* to *ba* is low.
- Infants are born with the ability to discriminate most speech sounds, and through a process called perceptual narrowing they acquire the phonetic categories of the language they are learning during the first year of life.
- The perceptual narrowing of phonetic categories is aided by a process known as distributional learning which is the tracking of the frequency and localization of various sounds in the speech stream.



Theories of speech perception

- We consciously believe that speech stream is a sequence of speech sounds grouped into words and phrases, but in reality we know this is only an illusion.
- The observation that there is no reliable relationship between a phoneme and the acoustic signal is known as lack of invariance.
- The problem that theories of perception need to explain is how we're able to extract the original speech sounds from speech stream. → Three Approaches
- Alvin Liberman and his colleagues at Haskins laboratory built a pattern playback machine the use of which led to the developed of the motor theory of speech perception. (1) Motor theory (2) General Auditory Framework (3) Direct Realism
- Motor theory proposes that people perceive speech by inferring the movement of the vocal tract that produced those sounds, instead of analysing the speech stream into phonemes.
- In other word, we comprehend speech by imagining how we would make the sound we're hearing.
- Influenced by the theories of Chomsky and Fodor, motor theory argued that speech is special.
- Speech is processed by innate dedicated modulus that are separated from general auditory perception.
- Speech is special was based on several reasons:
 - a) Speech perception and production was viewed as unique human ability.

2nd Approach → General Auditory Framework (GAF)

Speech perception operates by same mechanism that have evolved for perceiving environmental sounds.

[Human infants can discriminate speech sounds before they can speak]

→ Non Humans animals can distinguish speech sounds:-

(i) Chinchillas categorically perceive da and ta (Kuhl & Miller, 1978)

(ii) Japanese quail perceive d in deh, dee, do as same despite coarticulation effects (Kluender et al, 1987)

→ Speech perception is so complex that we use any available cue.

Fuzzy logical model of Perception:-

- Perceptual decisions made by matching relative goodness of sensory i/p to prototypes in memory

3rd Approach → Direct Realism

- Sensory i/p sufficiently rich, allows us to completely recover object of perception.

- Hence, we have direct awareness of world, no inferences necessary.

→ Perceptual speech is based on getting not on how its acoustic is being

Rejects "Speech is special"

Motor system involved in speech perception.

Mirror Neurons:-

→ Neurons in primates that fire when perceiving or performing task.

→ Links perceptual & motor systems.

Embodied Cognition

→ Cognition rooted in body's interactions with environment.

- b) Speech perception modulus words independent of auditory perception.
- c) Speech perception was proposed via the motor system.
- d) The objects of speech perception were not the speech sounds themselves but rather the intended vocal tract gesture.

- Ample research has shown that non-human animals perceived speech sounds in much the same way as humans do, forcing motor theorists to forsake the position that speech is special.

2nd Approach -

- General auditory approaches to speech perception overcome the lack of invariance by suggesting we make use of contextual cues, including information from other senses, to make reasoned inferences about the message conveyed in the auditory signal.
- The discovery of mirror neurons, which are active both when performing and perceiving an action, has lent support to a new version of motor theory known as direct realism.

General Auditory Frame work (GAF)

- The claim that speech was special was challenged by proponent of GAF.
- Assumption that speech perception operates by means of the same mechanism that has evolved in human and other animals to perceive environmental sounds.
- GAF argue against motor theory by pointing out example of speech perception with ability to produce speech.