

MII Notes After Midsem

LEC 1 - AM

Speech/Voice Modality : One of the most interaction modality.

Speech and Hearing :

According to **Sebock**, Speech is a subsystem of language, namely a communication using language in the acoustic channel.

According to **Stetson**, Speech is rather a set of movements made audible than a set of sounds produced by the movements.

Hierarchy of the linguistic channel (Linguistic hierarchy) :

1. Pragmatics = Deals with the meaning and intention of the user or the person uttering something. (What is the meaning the person is trying to utter ?).
2. Semantics = Meaning of a single sentence or the factual information. It is the content (factual information).
3. Syntax = The sentence structure, laws and structure of different words and parts of a sentence. How are different words arranged to form a sentence.
4. Morphology = Word forms, Different words in which the subgroups or subentities are used for building up the words.
5. Phonology = It is the sound structure, structure or rules by which sound is formed. Structure of spoken words.
6. Phonetics = Important hierarchy in speech processing. Perception of what you are saying. How spoken sounds are articulated or manifested in acoustics

and perceived at the human end. How the spoken words are received/perceived by us.

Components of Phonetics :

1. Phoneme = is a smallest segmental unit of a sound (this sound may be anything, may have meaning or not like it may be noise) employed to form a meaningful contrast between utterances. Building block of a sound. (This is just like "pixels" in image).

Another definition of phoneme : Functional class which is restricted to a given language.

Example : Analysis between two words "fit" and "bit", "kit" and "hit". These sound almost similar but are made up of different phonemes. They are differentiated by different phonemes.

2. Phone = Smallest segment or unit in phonetics and speech. Assigned to phonetics level not to phonological level. It is the subcategory of the phoneme.
3. Allophones = different phones of the same phoneme.

There are two ways to find the allophones :

- Free variants method = Let us take example of English and German, the 'R' sound in English and German. How the person utters 'R' are different (different phones) but same phoneme. The acoustics, perception, articulation of 'R' is different for English and German.
- Distributed Complementary = In german language and Greek, 'ich' and 'ah' (in german ish and hah), but in greek (as ich and ah). These two have same phoneme but uttered in different ways (different phones).

4. Morpheme = In contrast to phoneme, smallest unit actually having some meaning. Morpheme is the word stem which builds up the particular word. Morpheme is able to distinguish gender, plurals etc.

Example : Painter, paint is a thing while painter is a person who uses paint. The additional term 'er' is a small syllable which is a morpheme. 'paint' is also a morpheme. 'er' gives characteristics to the word (which is painter is a person).

5. Syllable = Not in the levels of linguistic, not in the linguistic hierarchy. This is a term in phonetics. It is a unit of articulation. An object + a person becomes a person. Syllable may be a vowel, consonant which gives the most important information in that word. Syllable helps us to identify the particular word.

“Broad umbrella is a morpheme where it can contain many syllables.”

SOURCE FILTER THEORY OF SPEECH OR VOICE :

The most important thing in speech processing is "source filter theory", this is a model which assumes voiced speech production in humans consists of two elements :

- Excitation which is happening at the vocal chords. Excitation signal is produced at vocal faults.
 - Upper part of your articulatory organ (mouth, throat, nasal cavity and so on).
-

LEC 2 - AM

Human Multimodal perception :

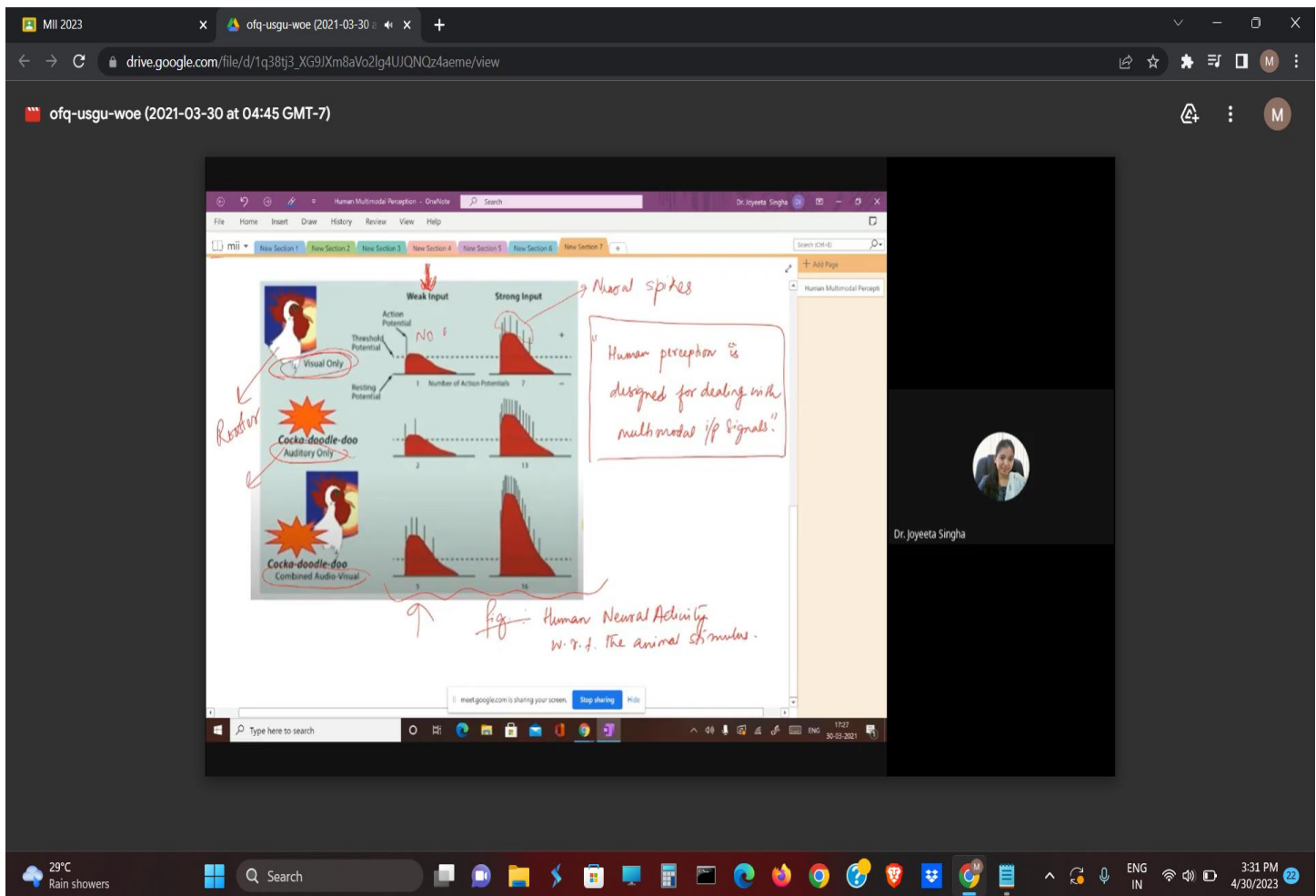
Human generally rely in multimodal input. We require multiple modalities.

Difficult to find information from noisy image, if we know what we are looking for, it becomes easy for us to find that thing.

Topics :

- 1) Processing multiple signals
- 2) Multimodal dual task
- 3) Effects of discongruency in multimodal signals
- 4) relevance to human-computer interaction

Statement = "Human perception is designed for dealing with multimodal I/P signals" - this statement needs to be explained with the help of a figure :-



About the figure above : This figure is example of human neural activity as response to animal stimuli.

- Only visual provides "no neural spikes" in weak signal.
- Auditory only - few neural spikes with auditory weak signals.
- Audio + video = neural spikes in weak signal is very much more.

Therefore, Human perception is much more "higher/better" in case of multimodal input than in the individual modality.

Super additivity = property for multimodal systems. Multimodal input stimulates the human stronger than the sum of the single individual inputs.

As in the figure, 1 + 2 combined has a lower neural response than 3, which is a multimodal input.

ofq-usgu-woe (2021-03-30 at 04:45 GMT-7)

Visual Only
Auditory Only
Combined Audio-Visual

Weak Input
Strong Input

Neural spikes (NS)

Human perception is designed for dealing with multimodal i/p signals.

1 + 2 < 3

Property: Super-additivity

Multimodal system

Multimodal i/p stimulates the human stronger than the sum of individual i/p's

Fig: Human Neural Activity w.r.t. the animal stimulus.

Dr. Joyeeta Singha

29°C Rain showers

Search

3:36 PM 4/30/2023

Statement - “Using multimodal system, Output will improve (Jacko et al., 2005).”

Advantages of multimodal system :

- The reaction time will decrease for user.
- Increased salience and prominence (Ex: Degree of attention grabbed).
- Decreased workload (as multiple modalities will be searching for region of interest).
- More robustness
- More naturalness/authenticity (Ex: 2 factor authentication in mobiles).
- Shorter task completion time because multiple modalities work.
- Improved Usability and user experience if multiple modalities are there
- May be used by certain target groups (such as older adults, people with certain deficits, such as visually impaired, auditory impaired).
- Can be used to design new interface paradigms.

Example of multimodal system : Technische Universitet, Berlin

Spatialaudio.net/ssr = developed a multimodal system in certain virtual environment. 3D audio synthesis system + 3D virtual environment (with visual information), they combined audio and video to form a multimodal system.

Statement - "Welch et al 1986 said - Each modality is unique."

Example :

- Visual system = used for detailed information (special information), useful for gathering complex information and special information in HCI.
 - Auditory system = good for general information (temporal information), and those information not in view. Useful for warning, grabbing attention.
-