Affective Computing

Week 1

Fundamentals of Affective Computing

Lecture 1

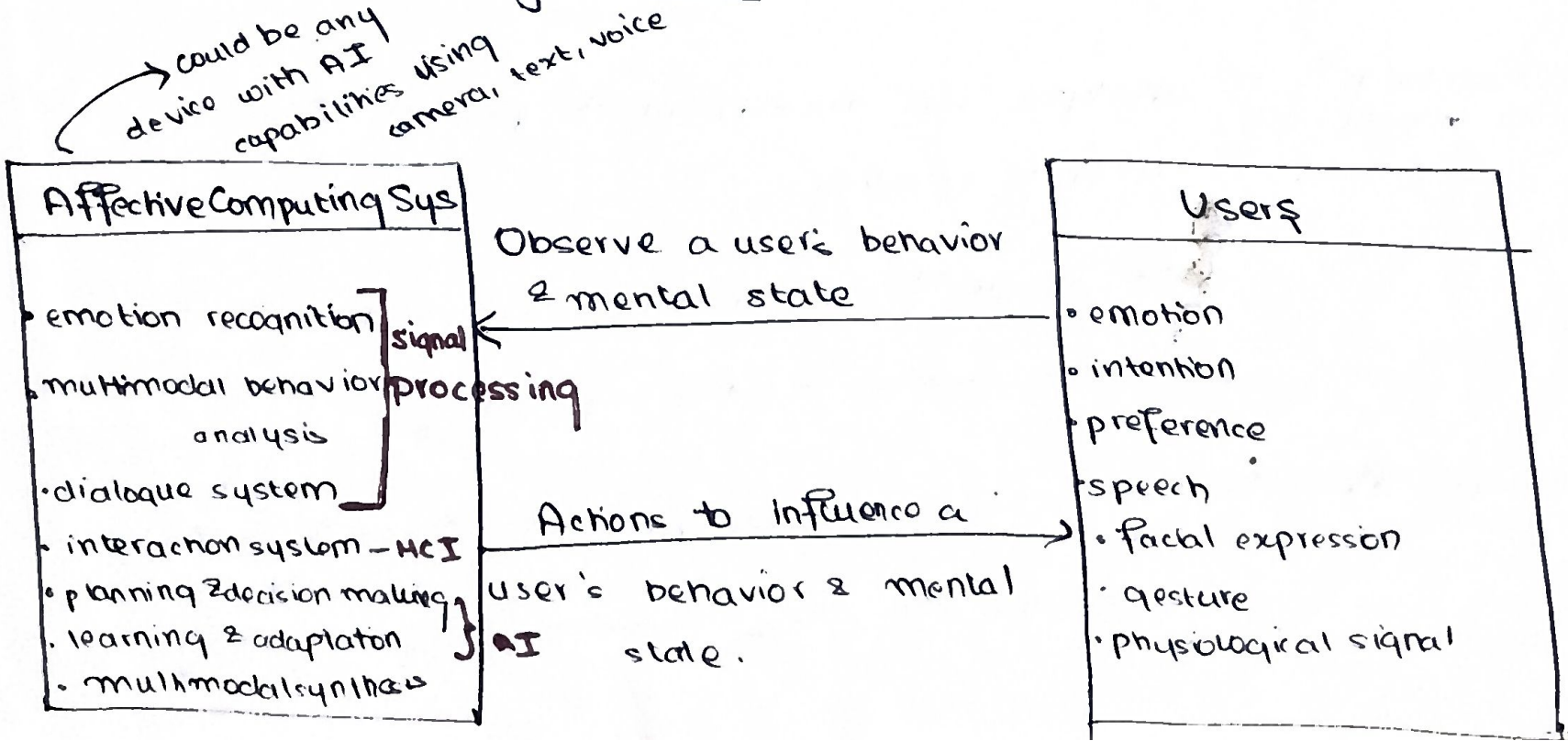
* Definition - Affective Computing - involves the creation and interaction with machine systems that sense, recognize, respond to, and influence emotions.

has 2 components:

(i) machine should understand the emotional state of the user

(ii) once the machine has this info, how should it react?, how should the info be presented via an interface to the user

* Affective Computing System



eg:- an operator is using complex machinery

- machine senses that the operator is fatigued
- affective computing system recognizes this and tells the human to take a break.

* Affective Computing in AI

① Speech — understanding speech of user

(i) automatic speaker recognition

(ii) automatic speech recognition,

sensing emotion in speech → would give the emotional state

this can be used to generate o/p based on emotion

② Facial Recognition - face detection, recognition

- understand facial emotions
- synthesize emotions on virtual agent that mirror the user's tone/ expression

③ Lip Reading - analyze lip movement to predict what the user is speaker

- tells about facial structure corresponding to a particular emotion - while speaking

④ Systems Monitoring - for a given product shown to a user, analyze the user reaction / response.

(5) Game play? - use facial exp, head poses to analyze how the user is reacting to the gameplay.

- could be used to train an individual w/ facial expression issues., could be in the form of a game.

(6) Social skills - understanding cohesion / unity in a group

- use affective computing to perceive emotions of a group

* Affect Sensing?

→ refers to a system that can recognize emotion by receiving data through signals and patterns.

→ to accomplish this task, a computer would need hardware & software

hardware ↓ sensor to capture info, high quality computing (GPUs)

software ↓

- platform to interface w/ hardware
- libraries to process the data feed coming from sensors
- ML models

→ Affect-sensing systems can be classified by modalities, each of which has a unique signature., i.e identify based on type of data / what type of sensor used.

eg. if you are using your laptop & interacting w/ software — the modality is the data input coming from the mouse & keyboard.

* Affect Sensing Modalities

① Facial activity through camera - can capture emotional state

- low cost & easy to use

- issues are: (i) privacy issues - identity, gender, age

(ii) color → use thermal cameras instead

(however, they are expensive & have lower resolution)

- can be used to analyze eyes / gaze . - gives cues, on the basis of whether someone is making eye contact or not - can assess confidence)

- can assess the pose/ gestures of an individual through a camera.

② Microphone - voice based sensing - capture what the user is speaking, analyze, predict emotion

- can also understand the scene (w/ background noise - like music)

- extract info from waveforms, analyze statistics and then use an ML model for prediction.

③ Textual - Natural Language Processing

→ in the form of documents, emails, social media, conversations (human-machine, human- human)

→ analyze tone of text

→ different models for words, sentences & documents across diff. languages

→ can also refer to processing text from transcripts from human-human interaction

→ identity can be hidden w/ this modality - unlike camera & voice.

(4) Physiological Signals

→ analyzes implicit physiological states like heart rate

→ a commonly used sensor is EDA - called electrodermal activity - studies the conductivity change in the skin due to increase in activity of sweat glands (closely related to how the person's emotional state it)

→ analyze this data w/ statistics & build an ML model

→ preserves privacy, but it is intrusive in nature. (make it more user interactive & friendly - like a smart watch).

→ This intrusiveness could be a cause for data collection bias since the user is acutely aware that the sensor is on him. (this is also in the case where a camera is an person's face).

→ usually physiological sensors are used in ensemble (a combination of sensors)

→ another commonly used sensor is EEG - electroencephalography - has electrodes, which record the electrical activity in the brain (neurons) - different neural pathways are triggered by diff. emotions

→ Here, info is collected directly from the brain - no way of bluffing (perceived emotion & actual emotion are the same).

→ however, disadvantages are : (i) not user friendly

(ii) user is aware of EEG cap's presence

(iii) if user moves, there will be noise

(iv) electrodes have to be placed in the right spots

(v) electrodes may also malfunction

(vi) not easy to take out of a laboratory setting (unlike camera & voice)

Lecture 2

* Applications of Affective Computing?

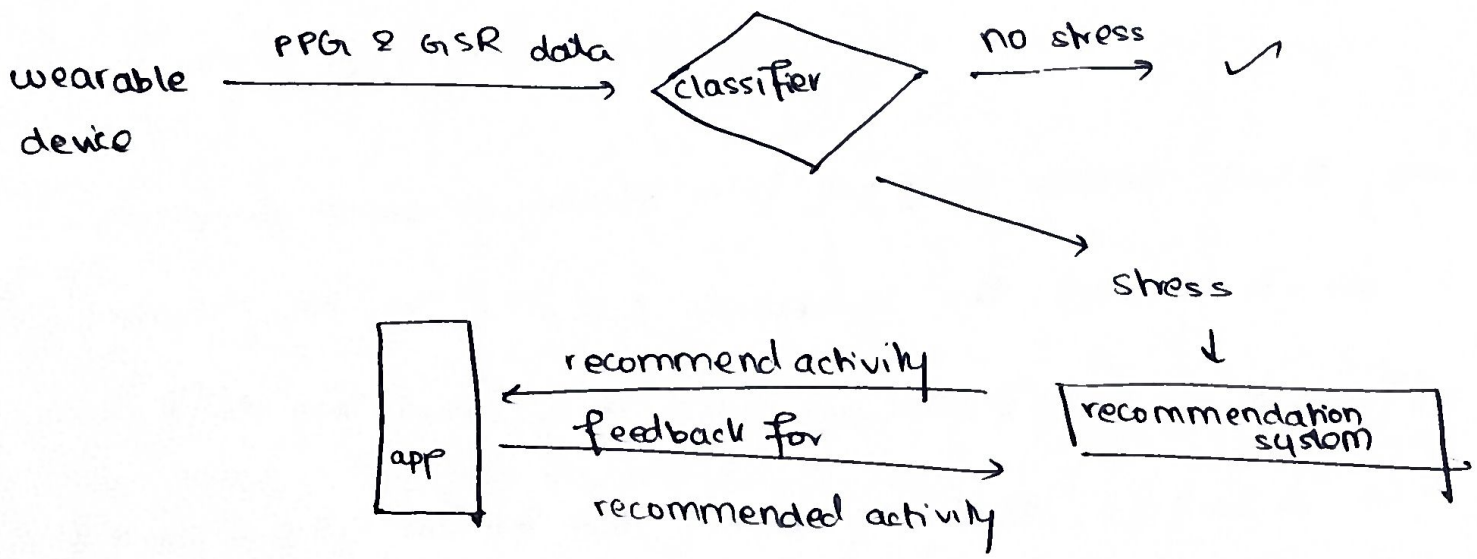
1. Detection - systems that detect the emotion of the user.

2. Expression - systems that express what a human would perceive as an emotion (eg. an avatar, robot and animated conversational agent).

3. Perception - system that actually feel an emotion

Healthcare Applications

(i) stress detection & recommendation system



(ii) For Individuals with Asperger's syndrome (AS) or high functioning autism (HFA)

→ mobile applications like SymTrend & Autism Track

(iii) For PTSD - (post traumatic stress disorder)

some applications used are:
- StartleMart
- VirtualVietnam, VirtualIraq, Virtual Afghanistan

Educational Applications

(i) EngageMe - uses skin conductance data collected from students in a classroom, along with video feeds, to help the teacher reflect on his or her classes.

(ii) Subtle Stone - a wireless, hand-held device in the form of a squeezable ball that allows students to communicate their affective & motivational experiences to teachers in real time

(iii) Automatic Tutoring - objective measurement of student engagement & learning effectiveness

(iv) Engagement Prediction - use facial expression, gaze, head pose

(v) MACH - automated conversation coach

* Emotionally Intelligent Interfaces

→ The UX design encompasses of aspects of a user's interaction with the product and the company - how easy is it for the user to achieve a goal through the UX?

→ It concerns both usability and ease of use.

→ The combination of the two leads to a better understanding of emotional experience and interaction & creates a more positive, memorable experience.

→ To check / test the ease of a user's experience, use participatory design / acceptance testing?

→ Participatory design is a process that includes the stakeholders in the early stages of design. - it leads to user-centered design

eg. the PICTIVE approach

→ Acceptance testing - during beta testing, the end-user validates the product for functionality, usability, reliability and compatability can be done with focus groups & surveys

Problems - users have to articulate exactly how they feel about the product in the form of focus groups and surveys, which can be plagued by participant bias, recall bias etc.

* Advanced UX / Neuromarketing?

→ Application of affective computing / neuroscience to marketing

→ deals with aspects like:

(i) how customers use their mind while responding to products

(ii) how this impacts their decision-making

→ for eg. can make a web-based application in tandem w/ cameras to analyse face, expression, gestures, along with EEGs.

* Ethical Considerations of Affective Computing Applications

(i) Emotional Manipulation - Is it ethical for computers to detect, recognize, and attempt to modify certain behaviors?

(ii) Privacy - emotions are private and personal

(iii) Emotional Dependency - If users regularly uses this kind of application, codependence may occur.

Affective Computing

Week 2

Emotion Theory and Emotional Design

* Emotion Psychology

Emotion - a complex set of interactions among subjective and objective factors, mediated by neural/ hormonal systems which can

(i) give rise to affective experience such as arousal /pleasure/ displeasure

(ii) generates cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes

(iii) activate widespread physiological adjustments to the arousing conditions, and

(iv) lead to behavior that is often, but not always, expressive, goal-directed and adaptive.

Example - Imagine you see a lion in a forest:

then

(i) the affective experience is fear

(ii) the cognitive processes would evaluate the scenario, check where the lion is

(iii) physiological adjustments would be an increase in blood flow, supply of oxygen

(iv) goal directed behavior would aim to find safety for yourself.

* Emotion Generations

→ Do we feel emotions because of bodily reactions, or do the bodily reactions happen because of the emotions.

(or)

Do we run from a bear because we are afraid, or are we afraid because we run?

→ Two popular theories from a psychologists pov are:

(i) James (Common Sense Theory) - proposed that we are afraid because we run. Emotions are often accompanied by bodily responses - like a racing heart, sweaty palms, tense muscles etc.

(ii) Worcester suggested that fear is not caused directly by sense perceptions but by certain thoughts to which these perceptions may give rise.

→ The physiological responses return to the brain in the form of bodily sensations, and the unique pattern of sensory feedback gives each emotion a unique quality.

* Bidirectional Projections

→ The brain impacts on the body via visceral efferent pathways (refer to internal organs)

(neural pathways that carry signals from brain & spinal cord to the peripheral nervous system)

→ The body impacts the brain through afferent feedback.

body sends sensory information and feedback to the brain - afferent pathways carry sensory signals from the body's sensory receptors like skin, muscles and organs to the brain & spinal cord.

→ The voluntary contraction of facial muscles contributes to the emotional experience - as seen in the expt where participants who held a pencil with their lips to inhibit a smile, rate cartoons as less amusing than participants who held a pencil in their teeth to mimic a smile.

* Emotional Models

→ A distinction can be made between the perceived and felt / induced emotions.

→ Perceived emotion - is the emotion recognized in the stimuli

→ Induced emotion is the emotion that is actually experienced by the listener.

→ Representation of emotions can be:

(i) discrete or categorical models

(ii) dimensional models

A. **Categorical Model**

→ represent discrete emotions - easy to build classification models

→ use a single word to describe an affective state - based off of Darwin's evolutionary view of emotions

→ The Ekman Emotion Model describes 6 basic emotions:

(i) happiness (iii) disgust (v) anxiety
(ii) anger (iv) sadness (vi) surprise

Contempt is considered to be a 7th pan-cultural facial expression.

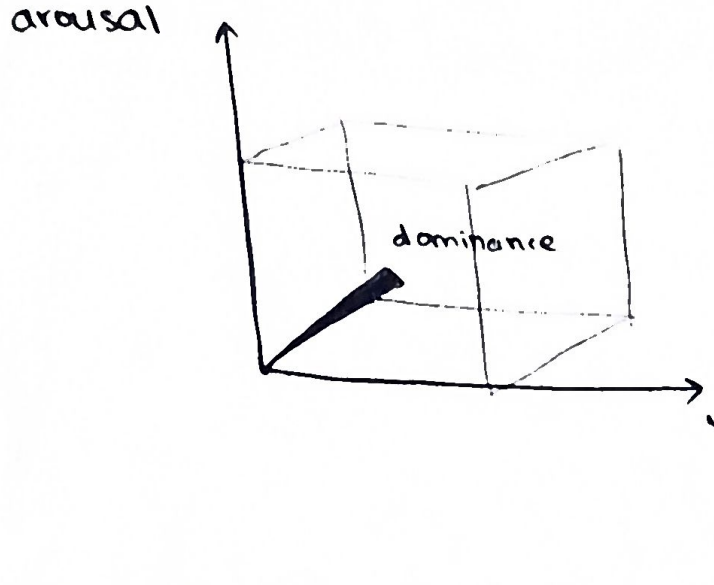
Advantages - From a computational POV, these models are easy to implement.

Disadvantages →Difficult to model relations between the discrete states - though more complex emotions like pride, shame etc. can be derived from the 6 basic emotions, it is not clear as to what proportion to mix them in.

→ There is also inconsistency - not agreed upon as to which emotions are basic and which are not.

B. **Dimensional Model**

→ a 3D numerical vector denotes the location of an emotion within this space.



VAD/
The PAD model

P - pleasure / valence

A - arousal

D - dominance

The 3 different scales are:

(i) Pleasure - Displeasure Scale : pleasantness - denotes the positivity of an emotion

(ii) Arousal - Non Arousal Scale : denotes the intensity / energy associated with an emotion

(iii) Dominance - Submissiveness Scale : represents the controlling and dominant nature of the emotion

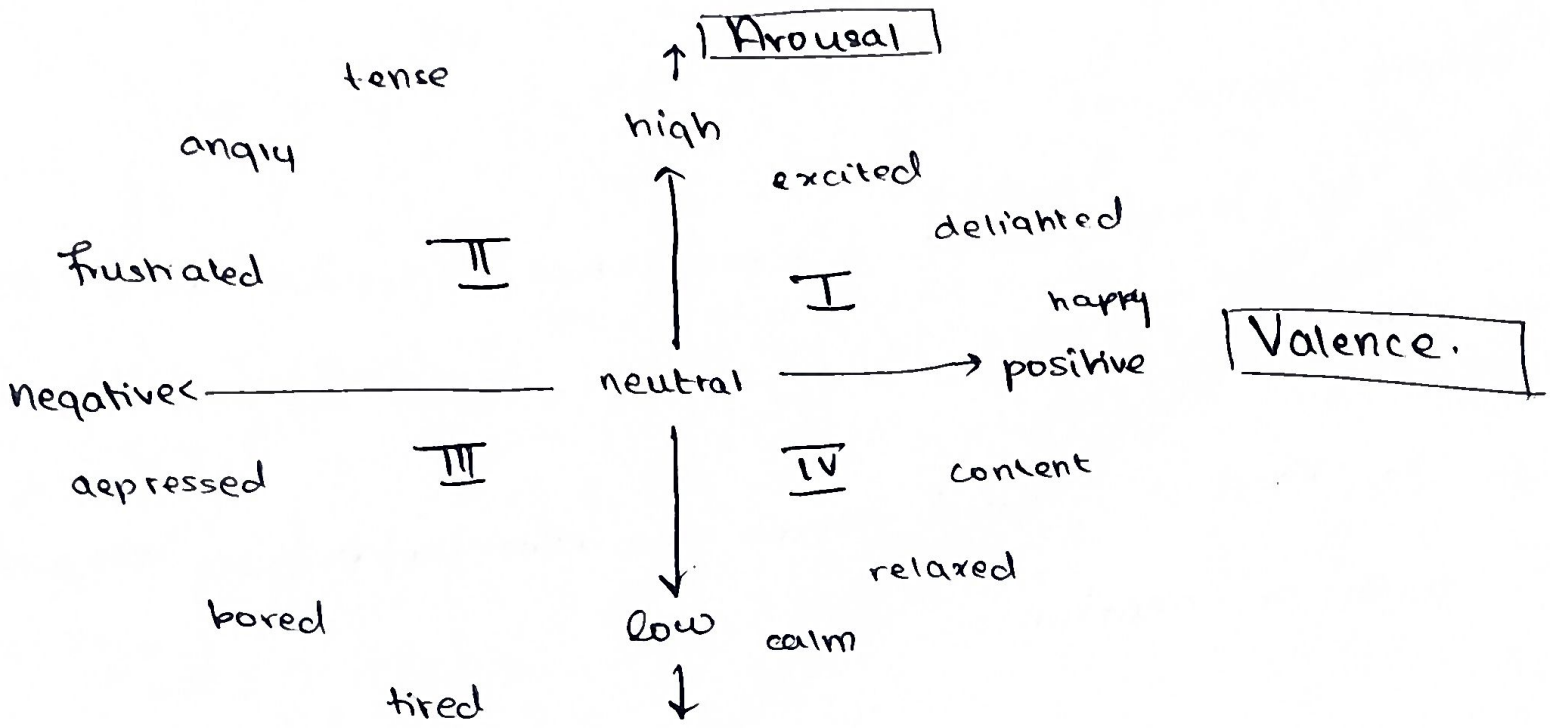
→ The PAD model is widely used in the field of emotion recognition, both in regression methods and studies that discretize dimensions into a few areas ↙ becomes categorical once again

Advantages - helps overcome the limitation of relating affective states to each other by providing a distance between them, i.e they allow for computationally interpretable relations between the emotional states.

Often, the third dimension is omitted - this gives the PA/VA model - called the circumplex model

c.) Circumplex Model

→ a 2D model, with valance and arousal



* Advantages of D in the PAD Model

→ With the circumplex model, we cannot differentiate emotions that are overlapping. However, the dominance factor in the PAD model makes them distinguishable

→ Consider the following example:

Fear	Anger
Valence - negative	valence - negative
Arousal - low/high	arousal - low/high

both are same in the VA domain

Dominance - submissive Dominance - Control

D - factor is able to differentiate both emotions

* Problems in traditional Affective Computing

→ Human emotions not only include the emergency emotion stimulated by intense instantaneous stimuli, but also the process emotion stimulated by the accumulation of continued weak stimuli over a period of time ⇒ studies would need to be more resource-intensive.

→ Emergency emotions are quick and low precision resources. The computational complexity of traditional precision-oriented affective computing is too high to handle emergency emotions

→ Human emotions are personalized and the emotions of different human individuals excited by the same stimulus can be different. (individual variability)

* Specificity of Emotions

→ Although basic emotions are characterized by specific facial expressions, a single set of facial actions can become different emotional expressions in different contexts. (eg. just looking at eyes is not indicative of a specific emotion - anger & fear may both result in wide-eyed faces)

A. Fear

→ characterized by raised eyebrows & drawn together, wide open eyes, tense lower eyelids and stretched lips

→ Associated with activation in the frontoparietal brain regions and a broad pattern of sympathetic activation, including reduced heart rate variability (HRV).

→ Associated with more skin conductance responses and larger electromyographic (facial muscles) corrugator activity than anger.

B. **Anger**

→ Lowered eyebrows drawn together, tense lower eyelids and pressed lips.

→ left pre-frontal cortex (PFC) is activated.

→ no change in HRV

→ anger may elicit either an anger-mirroring or a reciprocating fear response.

c. **Disgust**

→ raised upper lips, wrinkled nose bridge & raised cheeks

→ There is a differential skin conductance response based on the type of emotion.

(i) if core-disgust inducing stimuli (dirty toilets) - unchanged or decreased skin conductance

(ii) if body-boundary violating stimuli (mutilation scenes) - increased skin conductance.

D. Sadness

→ raised inner eyebrows, lowered lip corners

→ increased blood flow in ventral regions

→ The responses for the 2 types of sadness are different:

① Crying-related sadness

(i) increased heart rate

(ii) no change in HRV

(iii) increased skin conductance

② Non-crying sadness

(i) reduction in heart rate

(ii) reduced HRV

(iii) reduced skin conductance

(iv) increased respiration

E. Happiness

→ tensed lower eyelids, raised cheeks and raised lip corners

→ The intensity of smiling in photographs has been found to predict longevity

(i) no smiles - 72.9 yrs

(ii) partial smiles - 75 yrs

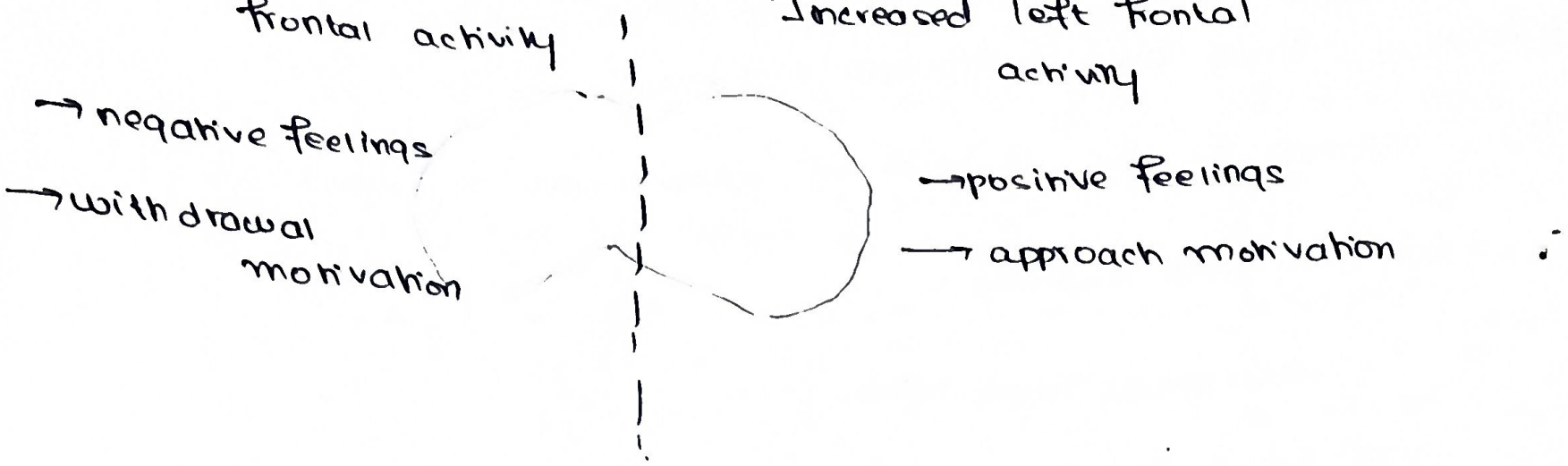
(iii) duchenne smiles - 79.9 yrs

→ has activation in medial prefrontal and temporo parietal cortices

→ left PFC also activated in +ve affects.

→ Happiness ⇒ decreased HRV, amusement, joy ⇒ increased HRV

* Emotion and Brain Asymmetry?



→ Higher engagement of the left - relative to the right frontal brain is related to positive feelings and higher engagement.

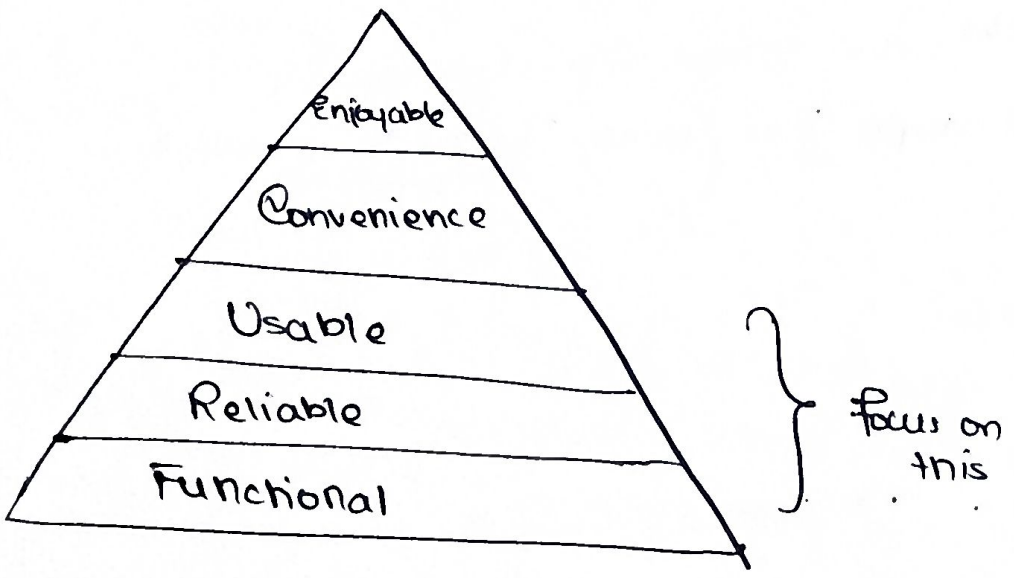
→ This is with the exception of anger - which has a left brain bias

*Emotional Design

→ creation of designs that evoke emotions that result in positive user experiences.

→ designers primarily focus on user's needs in their interaction, but also need to focus on user's emotional responses

UX pyramid

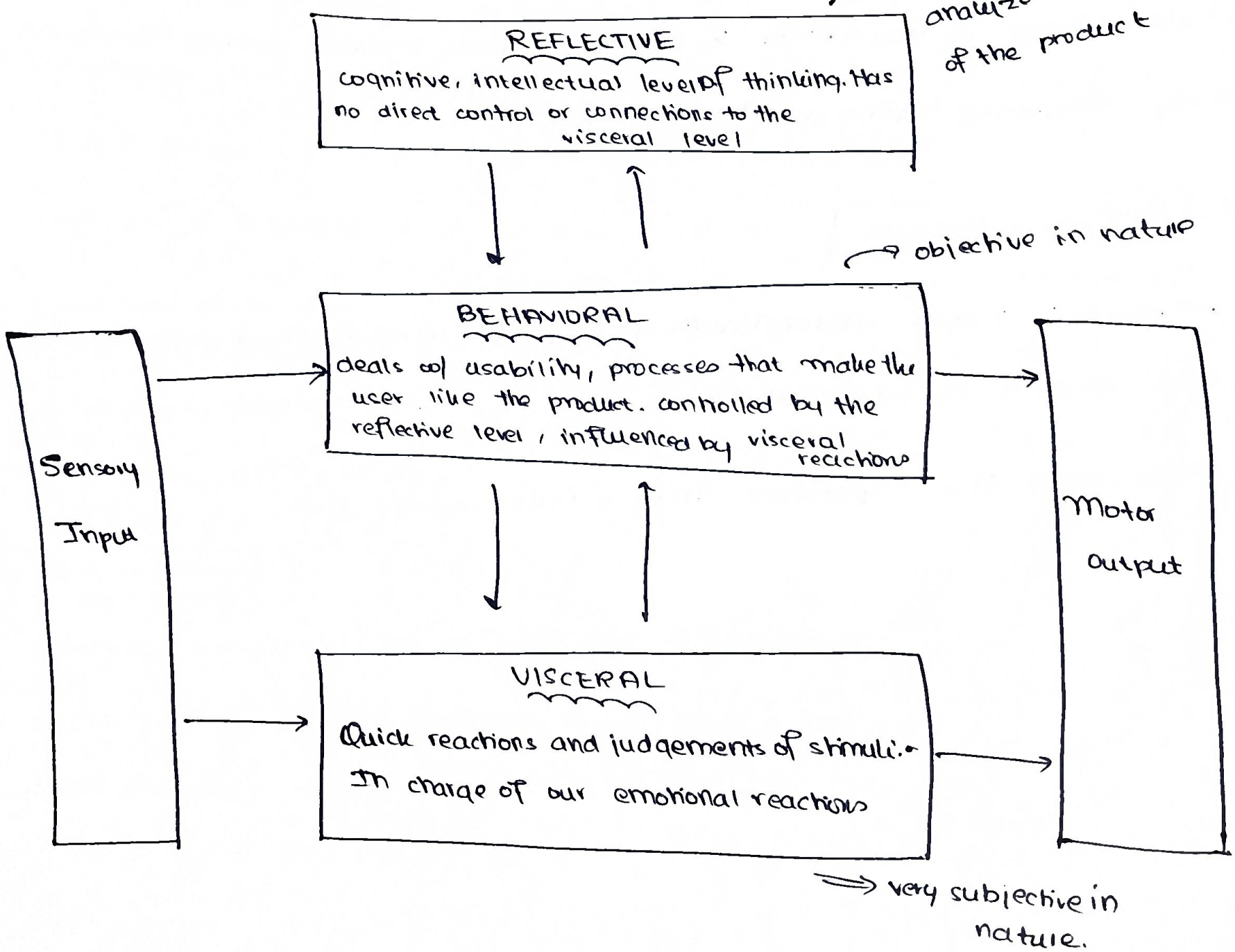


→ If one can elicit strong emotions from users - those emotions can be used to create loyalty or drive a customer to take action. Thus, the <u>emotional design</u> of a product or service affects its success.

→ Designers aim to reach users on 3 cognitive levels - (i) visceral
(ii) behavioral
(iii) reflective

so that users develop only +ve associations w/ products and services.

* Three Levels of Design



* Affective Computing and Emotional Design

A. Visceral Design

→ understand user's feelings in order to improve the first impression

→ eg. first time one bought an iPhone

B. Behavioral Design

→ understand the emotions users would feel during the tasks in both successful and failed attempts

→ can enable a thoughtful & emotional flow process within the design

→ eg. keyboard typing on the macbook

C: Reflective Design

→ enable. better understanding & post-usage emotional experience and enable an emotional bond.

→ eg. recall the experience from a roller coaster ride

Affective Computing?

Week 3

Experimental Design: Affect Elicitation, Research & Development Tools

* Affect Elicitation

* Datasets of emotional expressions

→ are of 3 major types

A. **Acted / Posed Expressions**

→ obtained by asking individuals (actors), to portray emotions

→ easy to collect

→ ecological validity is a concern (since real emotion is not conveyed)

B. **Naturalistic display**

→ ecologically valid

→ notoriously difficult to collect

C. **Induced Emotions**

→ emotional responses are elicited via some stimulus

→ combination of A & B

→ data collection effort is high

→ validity lies between A & B

* Methods of Emotion Elicitation

→ of 2 kinds

A. **Passive / Perception-based**

→ Individuals observe stimuli, such as film clips, images or music that are designed to evoke particular feelings

→ easier to collect & standardize.

B. **Active or Expression based**

→ Individuals are instructed to perform particular behaviors that might naturally evoke different emotions, such as posing facial emotions, adopting body postures or interacting w/ other people

* Methods of passive emotion elicitation

① Images - images could be presented for 10 seconds on a computer screen that is at a fixed distance, with a constant screen resolution, screen brightness & image size

- presentation method is standardized, such that all the individuals have the same viewing experience.

- The images can be selected on the basis of which emotion should be elicited from a database of standard images

(i) International Affective Picture System (IAPS)

(ii) Genova Affective Picture Database (GAPED)

Advantages

(i) non-invasive

(ii) easily accessible

(iii) relatively simple setup

Disadvantages

(i) strength of emotion is lower

(ii) emotional reactions are short and transient

(iii) lack of personalization

↳ individuals w/ a cocaine addiction react differently to pleasant, unpleasant & neutral IAPS image.

→ Images can be used only for reactive modalities (facial expressions, physiological signals) but not for productive modalities (text or gestures).

(2) Film clips

→ a neutral baseline film is generally shown prior to the presentation of each emotional clip

neutral clip (10 sec.) → emotional clip (1- 2 min.) → self-assessment phase (ask basic qs)

→ short clips ⇒ transient emotions

long clips ⇒ many emotions, but cannot determine which occurred when

→ physical situation should be standardized (20 inch about 5ft away)

→ For each target emotion, 1-2 short clips as homogenous as possible

→ Filmstim is a popular database.

Advantages

(i) capture attention well

(ii) can elicit higher intensity emotion & more complex emotion

(iii) can study emotion latency, rise time, duration, offset → time between action (stimulus) & emotion triggers

Disadvantages

(i) ecological validity

(ii) individuals may have seen the films from which the clips are taken

(iii) not ideal for obtaining data from productive modalities

* Methods of active emotion elicitation

① Behavioral Manipulation

→ Individuals are instructed to adopt particular muscle configurations or behavioral patterns that have been associated with emotional expressions, such as contracting or relaxing facial muscles or exaggerating natural emotional expressions.

→ directed facial action task - method for eliciting an emotion through the manipulation of facial expressions

→ ask participants to think or write about instances from their past when they experienced an emotion such as anger

Advantages

(i) can collect reactive expressions

(ii) effects are strong if the physical behavior associated w/ target emotions are known precisely.

Disadvantages

(i) ecological validity - are such emotions 'pure'?

(ii) one needs to know the physical behavior associated w/ a target emotion (complex emotions like frustration, confusion, engagement etc. are not include) => this limits the target emotions

(iii) poses or actions required can be difficult or complex.

② Social Interaction

→ interact w/ community & analyze emotions

Advantages

(i) realistic & natural

(ii) easy to elicit emotions that are otherwise hard to elicit such as anger or guilt.

Disadvantage

(i) difficult to collect, setup is difficult

* **Experimental Methodology**

* **IRB for Human Research**

→ Research Involving Human Subjects: Institutional Review Board (IRB)

→ IRB is a review committee established to help protect the rights & welfare of human research subjects.

Documents for IRB

(i) a draft / abstract of the proposal ← a clear description of the research methodology or experimental design

(ii) a copy of the informed consent form

(iii) a description of how confidentiality / anonymity will be protected.

(iv) risks and benefits to the subjects

(v) a copy of the recruitment document (ads / flyers)

(vi) data collection instruments (survey, sensors, list of questions)

* **Criteria for IRB Approval**

(i) risks to subjects are minimized / reasonable

(ii) selection of subjects is equitable.

(iii) Informed consent is sought

(iv) plans for monitoring the data is collected

(v) adequate provisions for privacy of individuals & confidentiality of the data

(vi) additional safeguards to protect the vulnerable population.

* Experimental Design

I.

develop concept studies

emotions/ stimuli/ sensors/ modalities

Type of study design

- within-subject (same subjects in diff condns)
- between-subject (diff. set of users in different groups)
- mixed model (a mix of within & between subject model)

make use of the same participants for diff. expt. condns

(control vs treatment groups)

number of groups

determining the sample size

Determine the methods of evaluation

- self-assessments
- interviews
- psychophysiology (refers to how emotions are being expressed in brain signals - EEG signals) heart-rate skin. conductance
- observations
- task performance (also called behavioral data - ask the user to do some secondary activity (like math))

determine the location of study

other equipment types

failures & contingencies

control

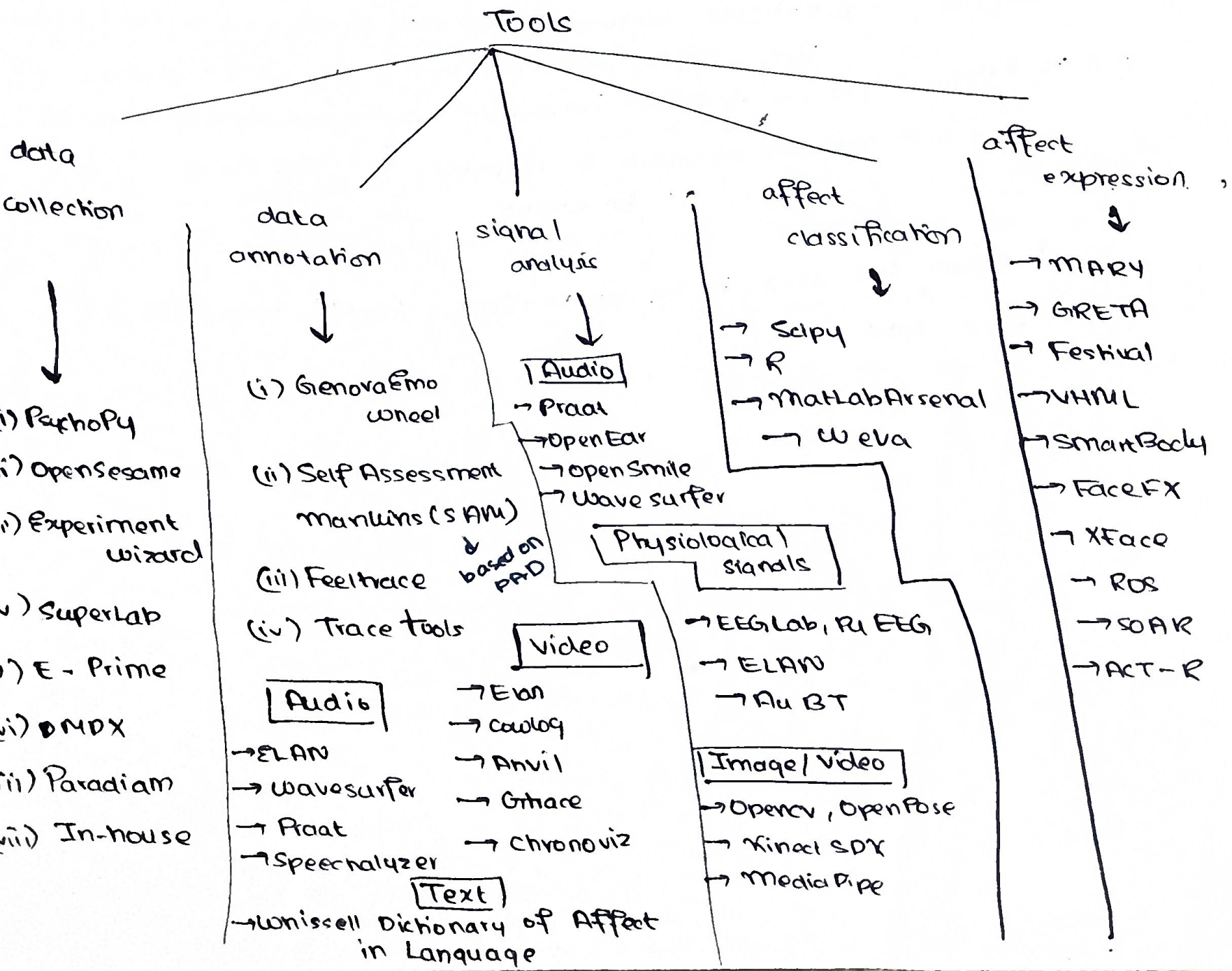
within-subject ⇒ less no. of participants
⇒ longer duration expts

between-subject ⇒ more no. of participants
⇒ shorter duration expt

- develop study protocols & scripts
- determine methods & recruiting participants
  - inclusion criteria
  - exclusion criteria
- Prepare IRB documents
- recruiting participants
- conducting experiments

* Research and Development Tools

→ Tools used in affective computing are categorized into 5 types



Data Mining

→ WEKA
→ AutoML
→ R
→ SAS
→ HTK
→ SciPy
→ MATLAB Arsonal

* Tutorial on Psychopy

Psychopy - an open-source Python tool that is widely accepted to create experiments in neuroscience & experimental psychology research.

→ provides a GUI interface

→ has 3 main building blocks for constructing experiments

(i) stimulus - pre-made custom templates to display various stimuli; like shapes, vids / pictures, there is also a response components which records mouseclicks key press

(ii) routines
↳ includes stimulus & response as an event

(iii) Loops

↳ can be used to repeat routines, stimuli, recording of face data.

→ Deep learning based FER systems :

- highly reduce the dependence on face physics-based models and other pre-processing techniques
- enable end-to-end learning to occur in the pipeline directly from the input images.
- the network itself finds relevant locations- later layers will extract finer features

Drawback

→ need a large amount of training data

→ more energy, increased computational complexity

Type of expressions = macro > micro

* Macro Expressions → anger, fear, neutral, sad, surprised

→ used to capture obvious/ universally known facial expressions

→ They can be visually observed through facial landmarks which are salient points in facial regions such as end of the nose, ends of the eyebrows, mouth etc.

→ last between 1/2 sec - 4 seconds

→ they match the content and tone of what is said

Drawbacks - what is considered to be 'universally' exclaimed emotions, are not always the same.

* Micro Expressions

→ more spontaneous and subtle facial movements that occur involuntarily

→ tend to reveal more genuine and underlying emotions in a short period of time.

→ often misinterpreted or missed altogether - occur 1/2 a second or less

→ unconciously reveal a concealed emotions

* Facial Action Coding System

→ a system based on facial muscle changes and can characterize facial actions to express individual human emotions.

* Action Units

Category	AUs
Happy	12, 25
Sad	4, 15
Fearful	1, 4, 20, 25
Angry	4, 7, 24
Surprised	1, 2, 25, 26

each emotion is associated with a few action units (facial muscles)

* Feature Extraction Methods

(1) Geometric Features

→ The relationship between facial components is used to construct a feature vector for training.

→ Two types of geometric features based on the position and angle of 52 facial landmark points
  ↳ compute Euclidean distance

Advantages - ① low complexity
② low storage requirements

Drawbacks

→ shape or geometric features alone are insufficient - since they might correspond to multiple AUs. - however, this can be resolved by considering appearance or texture information.

② Appearance Features

→ They represent changes in skin texture such as wrinkling and deepening of facial furrows and pouching of the skin.

Techniques : (i) use raw pixel-intensity values - prone to lighting conditions

(ii) more robust methods
Gabor wavelets or magnitudes

(iii) Histogram of oriented gradients

(iv) SIFT (Square Invariant Feature Transform)

→ local features are extracted
→ run the SIFT facial point detector
→ take the region around the point and create a histogram
→ then learn features using an ML model.

divide the face into non-overlapping blocks.
compute gradient in X and Y direction
create a histogram
post-processing - normalization

Drawbacks ① higher computational complexity
② higher storage requirements.

→ In SIFT, there are n data points for each face - so there would be multiple histograms for the entire dataset.

→ These histograms have to be organized in order for a classifier to learn them.

→ Use pooling methods - like the Bag of words approach.

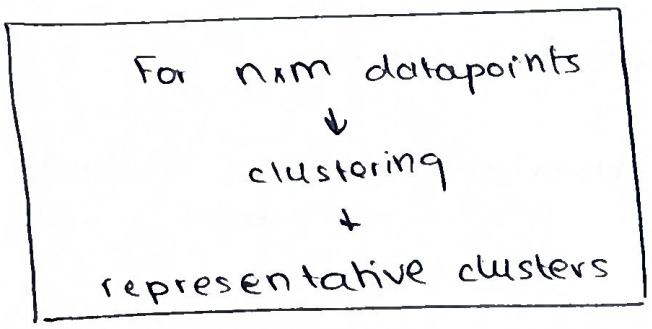
Bag of Words Approach

→ There are n histograms, which are around n-locations on the face.

→ Let each histogram correspond to one word., each face is a bag

→ Create a pooling mechanism, so that the n-histograms can be created & compressed into one histogram.

→ For all the training data m



called vector quantization

→ Take each sample at a time, which has n interest points create a histogram w/ the frequency of occurrence of representative clustering for each face

→ This method would work on videos as well, by applying clustering on each frame.

↳ disadvantage - does not study relationship between frames

(3) Motion Features

→ uses motion as a cue for expression recognition, especially for subtle expressions

Types of Motion Features

(i) **optical flow** - studies the flow (position & velocity) of a pixel as it moves from one frame to another

for N frames, obtain N-1 optical flow frames

the same bag of words approach can be used for further preprocessing.

(ii) **Motion history images** - used to observe human action recognition

- create a single image from a series of frames, flatten into one frame - observe change in pixel at a given location

(iii) **Local binary patterns** - take one block, from the same position on different frames

spatial component:

- from the volume, extract the subvolume
- compute the local binary pattern.
  - ↓
  - compare each pixel with its neighbor, if intensity value of pixel is larger than neighbor, mark as 1, else mark as 0.
- Do it for the 8 neighbors, and make an 8-bit code.
- convert it into a decimal
- do this for all the points, and make a histogram

temporal aspect
- divide the volume into 3 orthogonal planes (x_t, y_t, z_t)
- from each plane, compute the local binary patterns, combine this to make histograms

→ LBP is also used for textures & micro expressions

* FER Databases

A. The Extended Cohn-Kanade Dataset (CK+)

→ has video sequences of both posed & non-posed emotions

→ each video shows a facial shift from the neutral expression to a targeted peak expression.

B. Compound Emotion (CE)

→ has basic and compound emotions, eg. disgustedly surprised

→ has 22 categories

C. Denver Intensity of Spontaneous Facial Action Database (DISFA)

→ has 130,000 stereo video frames at high resolution

D. Binghamton University

→ has a 100 subjects

→ each subject performed seven expressions in front of the 3D face scanner.

→ each expression sequence contains about 100 frames

Acted Facial Expressions in the Wild

→ uses method actors from movies

→ collected using subtitle analysis

Flow

Extract close-captioned subtitles → recommender system → recommended clip → labeled

* Group Emotion Recognition

→ try to understand the overall perceived emotion of the group.

→ helps understand cohesion

→ can ~~use~~ try to identify universal emotions or valence and (or) arousal.

Factors Identifying Emotions in a Group

→ Faces being less occluded (clearly visible)

→ large faces with smiles

→ large smiles of people in the center of the group

→ some attractive people in image

→ age of some/ a particular person

→ large no. of people smiling

→ The factors that identify emotions in a group are classified into 2:

(i) **Top-down affect** - neighbors, scene, group

(ii) **Bottom-up affect** - individual expressions; attributes like attractiveness, age, gender, large faces, center faces, occlusion

* Bai's Scene Context Model

→ a theory that suggests that:

(i) one first looks at an image from a low-resolution holistic representation, similar to a scene description

then,

(ii) looks at the detailed object-level representation, face analysis etc.

* Group Representation

Steps - When one is interested in top-down & bottom up

Top Down

(1) object detector - detect faces

(2) consider group of faces to be a fully connected graph - vertices V are the faces, Edges E represent the link between 2 people. The weight of an edge is the distance between 2 faces. (calculate euclidean distances from the center of faces)

(3) Make a minimum-spanning tree (to find neighbors) - use Prim's algorithm.

θi = relative face size

$\theta i = \frac{Si}{\text{avg. neighborhood size}}$ → face size from object detection model

→ can also compute distance from the centroid,

indicates weightage of individual in a group photograph

$$d_i = \|c_i - e_i\| \quad (c_i = \text{centroid})$$

e_i → location

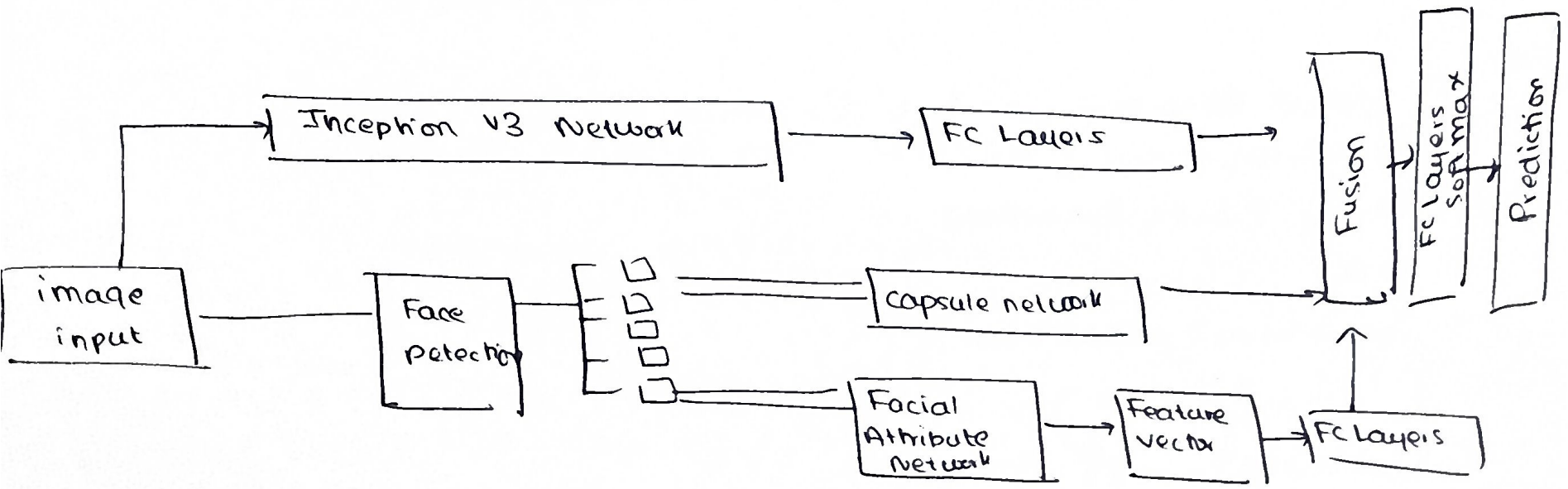
Bottom-Up

(1) use automatic facial emotion recognition (AFER) to compute levels of emotion (eg. Hi to indicate level of happiness)

(2) use θ_i or d_i to add/remove significance based on social weightage

A simple Group Emotion Model (GEM) could thus be of the form: $\frac{\sum H_i \theta_i d_i}{n}$

* Deep Learning for Group Facial Emotion Recognition



* Future work for Group Emotion Recognition

→ video based group-level emotion

→ audio-video fusion

→ large datasets with specific context

datasets

HAPPEI

Group Affect Dataset

(GAF)

* Applications of FER

① Physical Pain – McMasters UNBC Dataset

overlay pain level with video and match with expression

② Depression and Psychological Distress

↓ consider using other modalities like voice as well

↳ for unipolar depression - a person may have psychomotor retardation ⇒ facial emotion frequency and intensity is mellowed down

③ Deception Detection – look for micro expressions

④ Drowsy Driver Detection

⑤ In-class attention

* Limitations of Automatic FER

(i) real-time systems : computational complexity

(ii) illumination

(iii) occlusion

(i) self-occlusion

(ii) external occlusion

(iv) subtle face emotions

(v) individual variability (intra-class variation)

(vi) Ethical Issues

- fairness
- accountability
- transparency

* Tutorial

dataset - Ryerson Audio-Visual Database of Emotional Speech and song

emotions - calm, happy, sad, angry, fearful, surprise and disgust

levels of intensity - normal, strong

modalities - (i) audio-only

(ii) audio-video

(iii) video-only

Attributes - modality, vocal channel, emmotion, emotionintensity, statement, repetition, actor.

Process Overview

(i) extract frames from video file in Python

(ii) Histogram of Gradients based Emotion Recognition

(iii) usage of Gaussian Naive Bayes (GNB), ~~Latent Dirichlet Allocation~~ Linear Discriminant Analysis (LDA)

Support Vector Machine (SVM)

(iv) Classifying emotions using VGG-16 (Retrained on VGG Face Dataset)

Affective Computing

Week 5

Emotions in Voice

* Tutorial on Emotion Recognition using Speech

Pipeline : (i) preparation of dataset

(ii) selection of suitable and promising features

(iii) designing classification models

Dataset : RAVDESS

- Ryerson Audio-Visual Database of Emotional Speech and Song
- has speech & song by 24 actrs
- emotion classes are - calm, happy, sad, angry, fearful, surprise and disgust
- There are 2 levels of emotional intensity - normal & strong
- Modality formats are - audio only
  - audio - video
  - video - only

Filename Identifiers : modality - vocal channel - emotion - emotional intensity - statement - repetition - actor

Steps : (i) Feature Extraction
- Fundamental Frequency
- Zero Crossing Rate
- Mel Frequency Cepstral Coefficients (MFCC)

(1) ML models - GNB
LDA
SVM
1D-CNN on Raw Data

1. Load data - use librosa.load
2. calculate fundamental frequency - librosa.yin
3. calculate zero crossing - librosa.feature.zero-crossing-rate( )
4. find MFCC - librosa.feature.mfcc( )
5. use train, test, split
6. Fit with GaussianNB, SVC, LinearDiscriminantAnalysis
7. make CNN model add Conv
Maxpooling
Flatten
BatchNormalization
Dropout
Dense

Lecture - Speech-Based Emotion Recognition

* Speech in Affective Computing

→ Video alone may not always convey the right emotion

→ the voices, laughter, tonality etc. can give a clear idea of the emotions in that scenario.

* Applications of Speech in Affective Computing

1. Understanding man-machine interactions - robot understands emotion, and it can give appropriate feedback

2. computer movies and tutorial applications - indexing of videos, eg. look up videos which are happy

3. driver safety via car on-board machine - the mental state of the driver is conveyed to the operating system of the car - find if is inattentive, has -ve emotion - can instruct driver to take a break

4. diagnostic tool for a therapist to treat diseases - predict the intensity of depression

5. Tool for automatic translation system - in which a speaker plays a key role in communication between parties

6. mobile communications

* Difficulties of Analyzing the Emotional State through voice

There are 3 major factors:

A. what is said → speech to text - refers to information of linguistic origin and depends on the way of pronunciation of the words as representative of the language.

B. **how it is said** → duration, means - carries paralinguistic information related to the speaker's emotional state (levels of confidence)

C. **who says it** → culture. - cumulative information regarding the speaker's basic features and identities like age, gender & body size

- also depends on cultural context

* Types of Databases for Voice-based Affect Recognition

There are 3 kinds of databases:

① **Natural Data** - developed on spontaneous speech of real data

- includes the data recorded from the call centre conversations, cockpit recordings during abnormal conditions, btwn a doctor & patient, convos in public places.

② **Simulated (Acted)** - speech utterances are collected from experienced, trained and professional artists

→ better privacy

③ **Elicited (Induced)** - emotions were induced

- give stimuli to respond in a specific manner.

* Existing Databases

(1) AIBO Database - consists of recording from children while interacting with robot.

- contains emotion categories like anger, boredom, emphatic, helpless, ironic, joyful, reprimanding

(2) Berlin Database of Emotional Speech - 10 german sentences acted (recorded) in anger, boredom, disgust, fear, happiness, sadness and neutral.

(3) Ryerson Audio-Visual Database (RAVDESS) - acted statements, in a neutral North-American accent.

(4) IIT KGP - SESC

(5) IITKGP - SEHSC - from All Indio Radio Recordings

globally recognized data sets are: (i) Computational Paralinguistics Challenge (ComPare E)

(ii) EMPATHIC grant databases

* Speech Annotation

→ can use Audino

→ listen to segment, annotate speaker, emotion, metadata etc.

→ check consistency of labelling

* Limitations in Voice-based Affect Analysis

(i) limited work in non-English languages

(ii) limited no. of speakers

(iii) limited natural datasets - privacy concerns

(iv) limited work on emotional synthesis through speech (need for more zero-shot systems)

(v) limited work on cross-lingual emotion recognition

(vi) insights on ML-based methods are available but on DL-based methods it is limited.

Lecture 2 - Analysis and Synthesis of Affective Speech

* Acoustic Feature Extraction

→ called Prosody Features : related to rhythm, stress and intonation. It includes:

(i) fundamental frequency (f_0)

(ii) short term energy

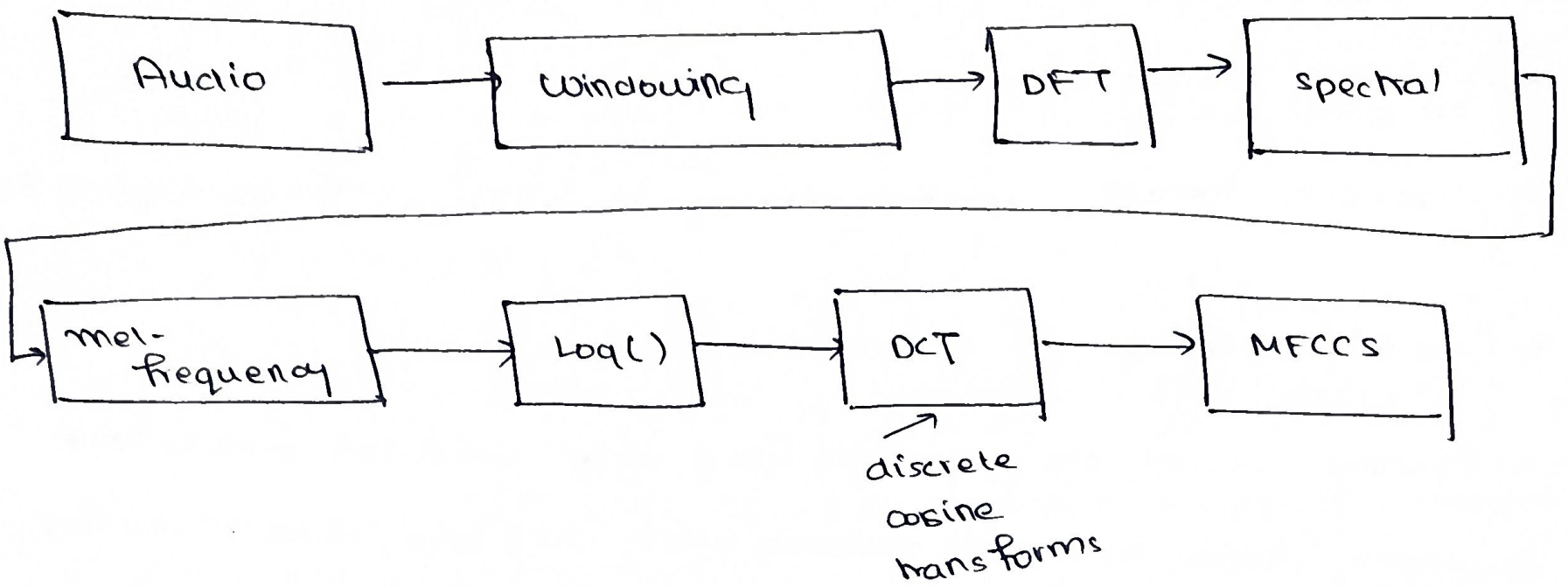
(iii) speech rate; syllable / phoneme rate

→ Spectral characteristics - related to harmonic or resonant structures. It includes:

(i) Mel Frequency Cepstral Coefficients (MFCCs)

(ii) Mel Filter bank energy coefficients (MFBs)

* Steps in computing the Mel Frequency Cepstral Coefficients



* Common Prosody Features

A. **Intensity and amplitude**

→ loudness - measures of energy in the acoustic signal

B. **Fundamental Frequency (f_0)**

→ gives the approximate frequency of the quasio-periodic structure of the voice.

→ pitch = lowest periodic cycle of the acoustic signal → referred to as the perception of f_0

C. **Formant Frequencies (F_1, F_2)**

→ used to analyze voice quality: concentration of acoustic energy around first and second formants.

D. Speech Rate

velocity of speech - number of complete utterances or elements produced per time limit.

E. Spectral Energy

→ related to timbre - relative energy in different frequency bands

* Good Vibrations

→ Positive voices are generally loud with consider variability in loudness, have high and variable pitch, and are high in the first two formant frequencies

→ Variations in pitch show differences between high arousal emotions (joy) and low arousal emotions (tenderness and lust), when compared with neutral vocalizations

* Feature Normalization

→ Angry speech has higher f_0 values than neutral speech, the emotional differences can be blurred by interspeaker differences.

→ Speaker normalization is used to accommodate the differences to to the speakers basic features.

→ For eg. F_0

men: 50 - 250 Hz

women - 100 - 500 Hz

→ Feature normalization can be done by the following methods:

(i) Z-score normalization - each feature will have zero mean and unit variance across all the data

(ii) Min-max approach

(iii) Normal dishibution normalization

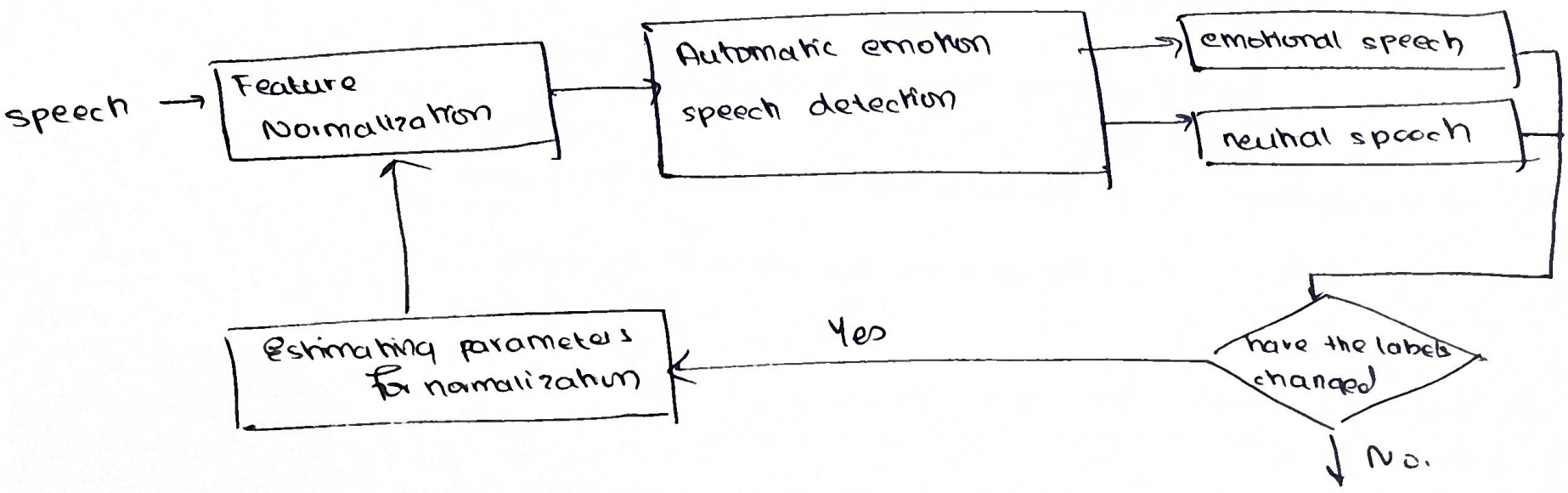
Issue: can adversely affect the emotional discrimination of the features
↓ reduces after normalization

→ To counter this, Iterative Feature Normalization can be used:

* Iterative Feature Normalization

→ Applying a single normalization across the entire corpus can adversely affect the emotional discrimination of the features.

→ Features can be estimated using only neutral (non-emotional) samples
↓
set as the baseline.



Steps

1. Acoustic features without any normalization are used to detect expressive speech (neutral vs. emotional classes)
2. The observations that are labeled as neutral are used to re-estimate the normalization parameters.
3. As the approximation of the normalization parameters improves, the performance of the detection algorithm is expected to improve, leading to better normalization parameters.
4. The process is repeated until the percentage of files in the emotional database that change labels from successive iterations is lower than a given threshold (say 5%).

* ML Techniques for Speech Affect Classification

Traditional
- (i) HMM
- (ii) Conditional Random Fields
- (iii) SVM
- (iv) Random Forest

Deep Learning
- (i) Convolutional Neural Networks
- (ii) Reurrent Neural Networks

* Representations for ML Models

→ Audio can be represented as spectrogram (freq. vs. time) - used in CNN models
→ treat the spectrogram as an image

* Emotion Speech Synthesis

→ text to speech operations, also give emotion class

look at Emotional Prosody Control for Speech Generation

* Challenges in Speech-based affect analysis

(i) intra and inter speaker variability —

- there is a heterogenous display of emotions and differences in the individual's vocal cords
- an emotion can be expressed in a no. of ways & is influenced by context

(ii) understanding which aspects of emotion production-perception are captured w/ acoustic features

(iii) Exhaustive and computationally expensive, limited real-time availability

Affective Computing

Week 6

Emotions in Text

* Applications of Analyzing Emotions in Text

(1) **Sentiment Analysis**

→ text categorization according to affective relevance

→ opinion exploration for market relevance

(2) **Computer - Assisted Creativity**

→ automatic personalized advertising and persuasive communication

(3) **Verbal expressivity in HCI**

→ Affective word selection and understanding are crucial for realizing appropriate and expressive conversations

→ for making question-answering systems

* Emotions through Typography

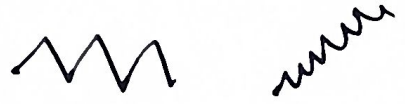
→ Typographers are attuned to subtle features while designing

→ Slight changes in font can convey different emotions - shape and placement of characters - especially gaps and spacing

→ Symmetry of characters and spaces conveys emotions & intent clearly

Poffenberger and Barrows - Line styles also convey emotion

for eg - angry, agitating, furials — angles sloping forward



sad -  — curves sloping down

happy, friendly -  - gentle curve

→ Articles received ratings as being funnier and angrier (satirical) when it was read in Times New Roman in comparison to Arial.

→ It is also observed that good typography elevates mood.

* Aesthetics of Reading

To understand the significance of typography, two tasks were performed

① Relative Subject Duration

→ refers to the participant's perception on how long they have been doing a task (here - reading)

→ With poor typography, participants estimated that they read for lesser time than the actual duration ⇒ more cognitive load

→ With good typography, the underestimated their reading time by a much larger time ⇒ reading was far more easier, the perception of time passing by was faster.

⇒ Good quality typography is responsible for greater engagement during the read task.

(2) Performance in the Candle Task

→ Candle Task = a cognitive performance task - measuring the influence of functional fixedness on a participant's problem solving capabilities.

→ Test is successful with 4/10 participants when the instructions have good typography, 0/9 with poor typography conditions

* Negative/Positive Emotion is not enough

→ Fine-grained emotion annotation is more effective than just positive / negative emotions.

→ For eg. both fear & anger express -ve emotions / opinion. However, anger is more relevant in marketing or socio-political monitoring of the public sentiment.

→ Fearful people tend to have a pessimistic view of the future, while angry people tend to have a more optimistic emotion. Usually, fear is a passive emotion, while anger is more likely to lead to action.

* Categorical / Dimensional Models

→ Dimensional models are scarely used, but can be promising models to represent emotions in textual data.

→ for eg. in order to distinguish between fear and anger:

	Valence	Arousal	Dominance
Fear	Negative	Low	Submissive
Anger	Negative	High	Dominant

* Complexity of Emotions

1. Implicitness

→ emotion expression is very context sensitive and complex

→ Considerable portions of emotion expression are not explicit

eg. 'To be laid off', 'go on a first date' have additional emotional information without specifying any emotional lexicon

→ To overcome this issue:

build a Knowledge base that merges Common Sense and affective Knowledge

eg. spending time w/ friends ⇒ happiness

getting into a car wreck ⇒ anger

2. Metaphors

→ Expressions of many emotions are metaphorical - cannot be assessed by the literal meaning of the expression

Affective Computing

Week 7

[redacted]

Emotions in Physiological Signals

* Emotion and Physiology

→ originates from activity of Autonomic Nervous system

→ pure, unaltered emotions

→ does not require user's attention

→ Recent advancements with wearable technologies allow for hassle free signal acquisition

→ Some common measures are:

(i) BP

(ii) ECG, EEG, EMG GSR

(iii) heart rate, respiration, temperature

* Heart Rate

→ closely linked to arousal

→ indicator of physical activation & effort

→ also indicator of
- fear
- panic
- anger
- appreciation

* Heart Rate Measurement

EGG, EKG

→ monitor electrical changes on the surface of skin

→ very small in μV

Photo-Plethysmography (PPG)

→ measuring pulse signals at various locations on the body attached to fingertip, earlobe.

→ dry sensors, easier to attach

* Cardiac Parameters
  - → Heart Rate
  - → Inter-beat Interval
  - → Heart-Rate Variability
    (Low HRV → aroused, too much stress
    High HRV → relaxed, body tolerates stress)

can compute RMSSD

$$\text{sqrt}\left(\text{mean}\left\{(RR_1 - RR_2)^2 + (RR_2 - RR_3)^2 \cdots\right\}\right)$$

$$n$$

Other Factors - EGG more accurate HR

Emotion is not the only factor affecting heart rate

- Age — Age ↑, HRV ↓
- posture
- breathing frequency

Limitation

cannot find whether arousal = +ve or -ve

* Skin Conductance

→ electrical conductivity of our skin subtly changes whenever we are emotionally aroused

→ called Electrodermal Activity (EDA)
Galvanic Skin Response (GSR)

→ sweating need not even be visible

→ key component in lie detector tests

* GSR Measurement

→ extremely easy to measure as it can be measured anywhere on the body

best places = palms of hand
soles of feet

→ response time between 2-20 μs with a range between 1-3 μs for different individuals

rise time 1 to 3 seconds
recovery time 2 - 10 seconds

* GSR Signals →
- SCL → skin conductance level - slow component
- SCR = skin conductance response
  (emotional reaction to an external event)
  increases amp. temporarily
  use similar inter-stimulation durations +10 sec

* Features from Skin Conductance

- mean peak amp
- sum peak amp
- peak count — MFCC does best
- mean rise time

* Limitation:

1. onset & offset of emotions difficult to identify
2. how to separate SCR & SCL
3. can't tell valence of arousal

must combine GSR w/ FER, EEG & eye-tracking

* Electroencephalography (EEG)

→ When thousands of neurons fire in sync - strong measurable electric field is generated.

→ place electrodes on scalp surfaces

→ EEG reflects how neurons communicate with each other via electrical impulses, and the association w/ cognitive processes such as drowsiness, alertness, approach/avoidance

→ provides excellent time resolution - 200ms

* EEG Electrodes

→ Each electrode placement site has a letter to identify the lobe

prefrontal	Fp	occipital	O
frontal	F	central	C
temporal	T		
parietal	P		

z = midline sagittal plane

→ Even numbered electrodes = RHS of head

odd numbered electrodes = LHS of head

* Frontal Alpha Symmetry = difference between right & left α activity in the frontal regions

→ FAA is used as a proxy for feelings of approach or avoidance.

Increased left frontal activity = anger / joy

Increased right frontal activity = disgust / fear / sadness

* Limitations

→ poor spatial resolution (where did signal come from)

→ Intrusive

→ Training needed

→ always contaminated by artifacts

Affective Computing

Week 8

Multimodal Analysis

→ Multiple modalities ⇒ richer representation → more accurate inference & expression

→ Each modality is expected to provide unique information

Challenges

→ classifiers and fusion methods

→ adequate samples

→ only modest improvements

* Multimodal Affect Recognition

→ Find underlying relationships & correlation between feature sets in different models

→ how much info each modality provides

* Data Collection
  - → spontaneous & subtle
  - → synchronized
  - → affective stimuli labelled simultaneously

* Feature Extraction
  - → use a variable sampling freq
  - → synchronize
  - → Feature selection : optimize the feature space individually, than by a combined feature selection

* Fusion Methods

Early Fusion - concatenate features from multiple cues into one feature vector

- challenging when the no. of features increase & they are all different.

Late Fusion - feature & time dependency are abstract

- each classifier processes its own data stream & o/ps are combined at a later stage.

Soft level - measure of confidence is involved

Hard level - combining mechanism operates on a single hypothesis decision

Slow Fusion - assume conditional independence between modalities

- exploit correlation between modalities

* SEMAINE → sensitive artificial listener

- multimodal dialogue system with the social interaction skills needed for a sustained conversation with a human user

- aims to engage the user in dialogue & create an emotional workout

Semaine Database

→ Spontaneous data capturing the audiovisual interaction between a human and an operator, w/ 4 personalities

Prudence - tempered & sensible

Poppy

Spike - angry - confrontational

Obadiah - sad & depressive

→ all interactions annotated by 2-8 raters

dimensions = Arousal

Expectation

Power

Valence

* Data & Annotations

→ find avg. value for each dimension over all the raters

→ find mean

→ threshold to find binary values

chunk video frames - frame level

word level

Affective Computing

Week 9

Emotional Empathy in Agents / Machines / Robots

* Why Empathetic Agents

→ Empathy is the capacity to understand what other human beings are experiencing

→ Presence of empathetic response by AI leads to better, more +ve & appropriate interactions

→ Works on the assumption that humans prefer to interact w/ machines in the same way they interact w/ other people

→ Attributing familiar human-like qualities to a less familiar non-humanlike entity can make the entity more explainable.

* Anthropomorphic Design

→ Tendency to provide human characteristics to non-lifelike artifacts

→ Anthropomorphism strongly affects: (i) form
(ii) behavior
(iii) interaction

→ Appearance & function of a product impacts how people perceive, interact w/ it

→ Appearance of a robot should match capabilities & user expectations

* Uncanny Valley human likeness vs. familiarity

→ describes the relationship between the human-like appearance & the emotional response it evokes

→ People feal unease & revulsion in response to highly realistic humanoid robots

* Empathy Subprocesses

→ 3 major subprocesses

① Emotional stimulation - an affective response which often entails sharing the emotional state eg. analyzes bot lots of data

② Perspective Taking - cognitive capacity of knowing thoughts & feelings eg. bot responds to user

③ Emotion Regulation - regulating personal distress from the other's pain to allow compassion & helping behavior

eg. bot designed to avoid topics that may triager -ve emotion

* Computational Empathy Analysis

Text use NLP

→ n-gram models

→ max likelihood classifier

Vocal cues → analyze prosodic patterns

→ analyze pitch, energy, jitter, speech segment duration

Facial Expressions → study co-occurrence of facial exp. patterns

→ study gaze patterns – mutual, one-way & mutually averted

→ empathy state in 3 classes – empathy, unconcern, antipathy

* Empathy Simulation

→ Truly empathetic agents are impossible to make.

→ A simulation of human-like behavior can be done –

Methodology

① Empathy Mechanism – internal imitation of expressions

② Empathy Modulation – interpolate perceived emotion

③ Expression of Empathy

* Evoking Empathy

Elements to be considered are:

(i) types of behavior

(ii) appearance & features

(iii) context & situation

(iv) mediating factors

* Theory of Mind

→ capacity to under people's mental states, includes the knowledge that others' mental state may be different from one's own states

→ allows people to infer the intentions of others

→ False Belief Test - designed to assess whether an individual possesses a theory of mind based on ability to attribute false beliefs to others

* Evaluation of Empathetic Response

→ Turing Test - imitation of human behavior

→ Psychological benchmarks: (i) autonomy
(ii) imitation

→ Moral value, accountability

→ Privacy

→ Methodological metrics - questionaires & content analysis

Affective Computing

Week 10

Emotionally Intelligent Machines

① Affective AutoTutor

→ first reactive conversational intelligent tutoring system

→ automatically detects boredom, confusion, frustration & neutral

→ individual diagnoses are combined with a decision level fusion algorithm

→ Reaction: empathetic, encouraging & motivational dialogue.

→ learning gains increased.

② GazeTutor

→ Interventions that monitor periods of waning attention

→ GazeTutor is a multimedia tutor consisting of an animated conversational agent.

* Sensor Scalability

→ Include scalable sensors wherever possible

→ cameras can be in the form of webcams, and are of low cost

→ Motion tracking techniques can be applied to video data, so posture sensor can be replaced w/ web-cams

→ Cameras can be used to monitor heart rate & eye-gaze.

* Accuracy? - How good is good enough?

→ Though naturalistic affect detection has come a long way, some persistent problems are:

(i) Intrusive, expensive & noisy sensors

(ii) Technical challenges from weak signals

(iii) Not enough adequate & relevant trainingdata

(iv) Issues with generalizability

→ Completely perfect affect detectors can never be developed

→ Close the loop when: (i) affect detection is almost perfect

(ii) even moderate degree of recognition is sufficient if the interventions are fail-safe and do not cause harm if incorrect

* Adaptivity - How adaptive is adaptive enough?

Level 0 - No adaptation

→ The system does not alter its behavior in response to the emotional state

→ There is a predefined interaction script

→ eg. Alexa, Siri

Level : Recognizes the need for adaptation

→ The system recognizes that the performance of a particular task could be optimized, but no optimization is performed

→ Indicators could be: (i) negative emotional state
(ii) change in emotional state.

Level 2 - **Single Task Adaptation**

→ A single task is adapted over time to optimize a particular metric

→ This adaptation is achieved by strategic overview of the performance of the system while carrying out the task.

→ eg. Shimi
Intelligent Tutoring Systems

Level 3 : **Multiple Task Adaptation**

→ a set of tasks are adapted over time to optimize a particular metric

→ could involve the reordering of tasks or adaptation of individual tasks

→ Adaptation is the result of accumulated experience.

Level 4 : **Communicated Task Adaptation**

→ Adaptation is carried out between multiple independent agents.

→ The adaptation is communicated between agents & applied individually within each agent.

Parallel Empathy - mirror / align the emotional state of user.

Reactive Empathy - respond to user's emotional cues to alleviate or address their emotional state.

Affective empathy - ability to share feelings of others w/o direct emotional stimulation to oneself

Cognitive empathy - ability to recognize & understand another's mental state.