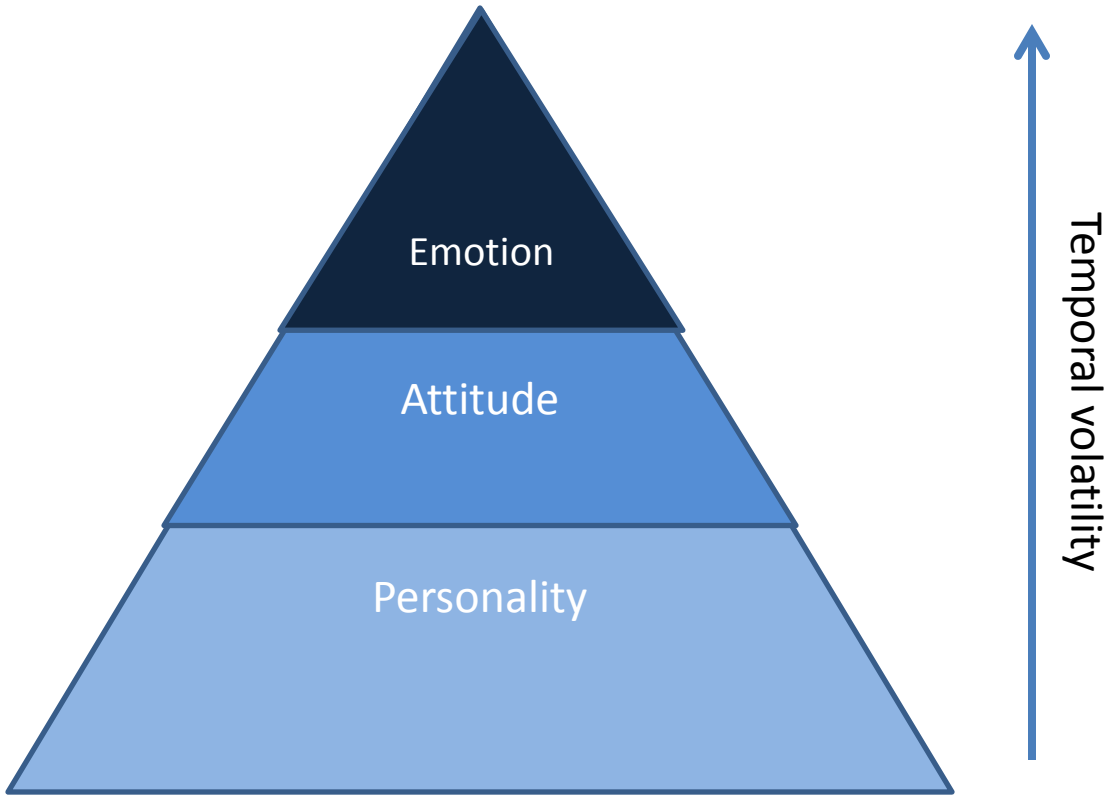
Emotion detection

Nisheeth

Sources of affective responding



Emotions are universally recognized

But how?

Attention to the person or the emotion: Underlying activations in MEG *(Bayle, Bostan & Taylor, 2007)*

Rationale

Facial emotion processing is fast (100ms) and automatic and occurs regardless of whether you attend to the face or not.

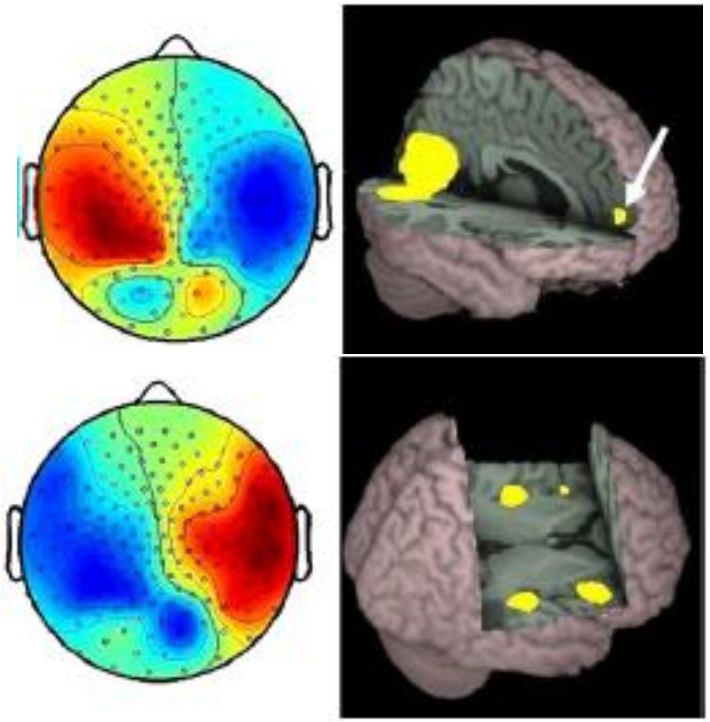
Facial identity is also fast (but slower) and occurs in parallel according to most models

But there is some evidence from schizophrenia suggesting that the parallel (and therefore separate) brain regions interact

Methods and Results

Used MEG and happy/fear/neutral faces

Identity task - press button when 2 identities the same
Emotion task - press button when 2 emotions the same



90ms orbito-frontal response to emotion regardless of attention

170ms right insula response when attending to emotion

Also 220ms activation increase for areas associated with identity processing

Conclusions

Parallel processing verified

Impaired facial emotion recognition and reduced amygdalar volume in schizophrenia *(Chihiro et al, 2007)*

Rationale

Amygdala volume known to be reduced in Schizophrenics

Emotion recognition known to be impaired in Schizophrenia

Direct link between the two not studied (properly) before

Methods

Facial emotion intensity recognition task

Results

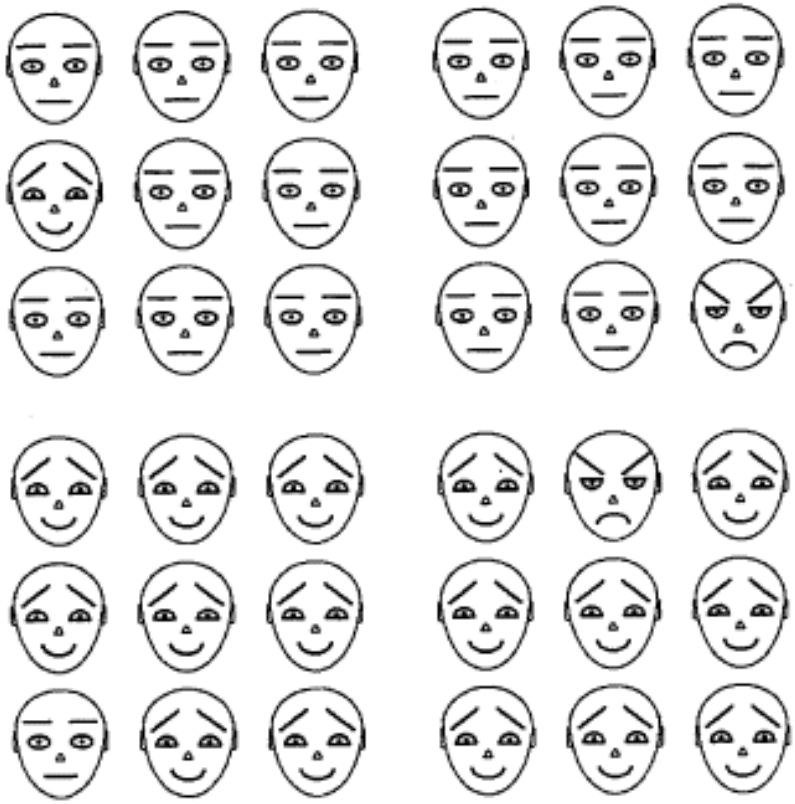
(1) The schizophrenia patients had smaller amygdalar volumes than the healthy controls;

(2) the patients showed impairment in recognizing facial emotions, specifically anger, surprise, disgust, and sadness;

(3) the left amygdala volume reduction in these patients was associated with impaired recognition of sadness in facial expressions.

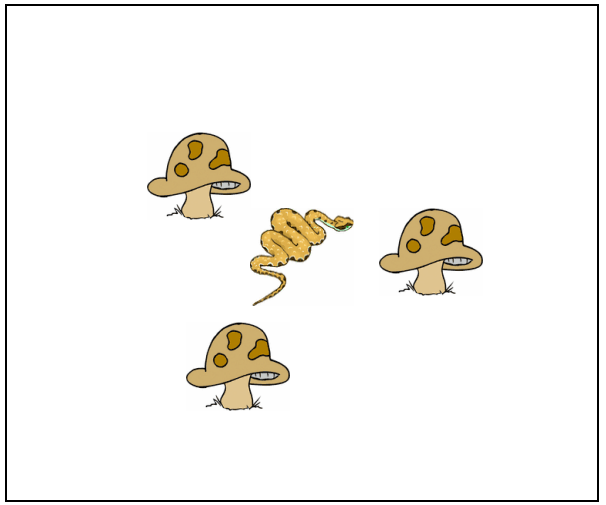
Automaticity of emotion recognition

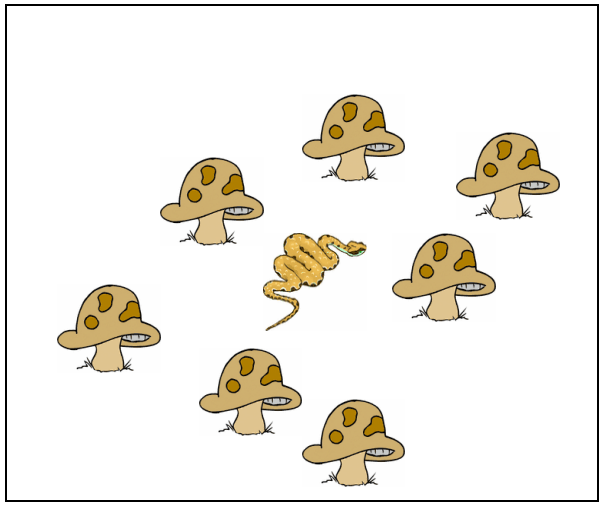
Angry faces are detected much more rapidly than faces depicting non-threatening expressions



Ohman & Mineka, 2001

Attention is driven by fear





Summary

- Emotion detection is built deep into the human information processing machinery
- Very fast and sub-conscious detection of salient emotion cues

AI and emotion detection

- How can a computer detect emotion in a human's behavior?
- Approaches
  - Motor expression (today)
    - Facial detection
    - Speech-based detection
  - Physiological arousal (tomorrow)
    - Galvanic skin response
    - Electromyelograms

Facial Expression Recognition: supervised learning

- Young children learn from parents' facial expressions what is desirable and not

Facial expressions communicate

Babies (10 months) almost only smile in presence of caregiver

Babies look to caregiver and behave according to caregiver response when encountering novel object. E.g. a barking dog or a snake

This is known as social referencing and is also seen in chimpanzee societies

A similar process, observational fear, is seen in monkeys. Infant monkeys show fearful unconditioned response to mother's expression of fear when the mother could see a snake, but the infants could not. That is, infants showed a fear response to the mother's fear response.

Facial expressions allow for rapid communication

They are produced when there is an emotional stimulus and an audience present

Our interpretation of another's emotion modulates our behaviour and vice versa

The ability to recognise emotion expressions appears very early

first few days (neonates)	can distinguish between expressions of happiness, sadness, and surprise
Four- to six-month	show preferences for facial expressions of happiness over neutral and angry expressions
seven months	can distinguish among expressions of fear, anger, surprise, happiness, and sadness

Classes of Expressions

- Joy
- Sadness
- Fear
- Disgust
- Anger
- Neutral

Automated Facial Expression Recognition

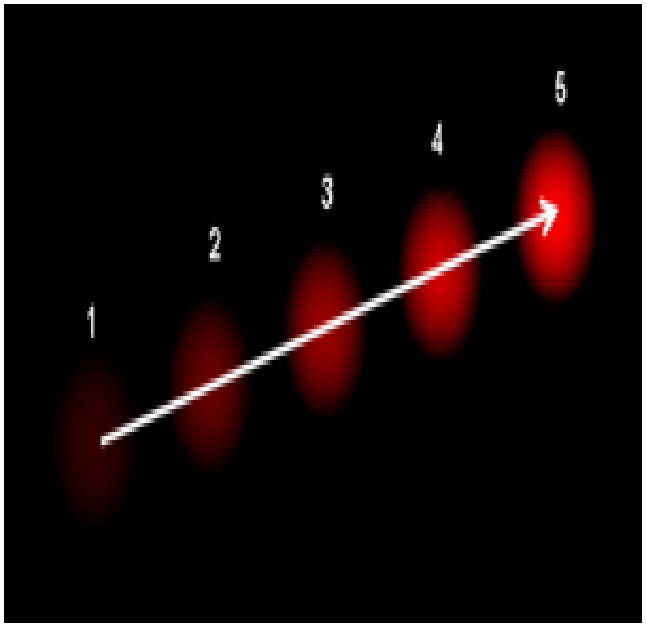
- Affective computing depends on reliably detecting human emotion
- Value proposition: natural ways of communication in person-to-machine interaction.
- Supervised learning via facial expressions.
- More natural (maybe) than keyboard or voice

General Machine Vision

- First step in the process is “vision”.
- After the image is acquired, some preprocessing is done such as to reduce noise, improve contrast.
- Next features are extracted and areas of interest are “detected”
- Finally some high-level processing occurs.

Optical Flow

- Used to capture motion of objects due to relative motion between object and observer.
- Also used to derive “structure” of objects.
- Looks at intensity of “voxels” and tries to solve a set of differential equations.
- Voxels = Volume Pixels = Think Pixels in 3d

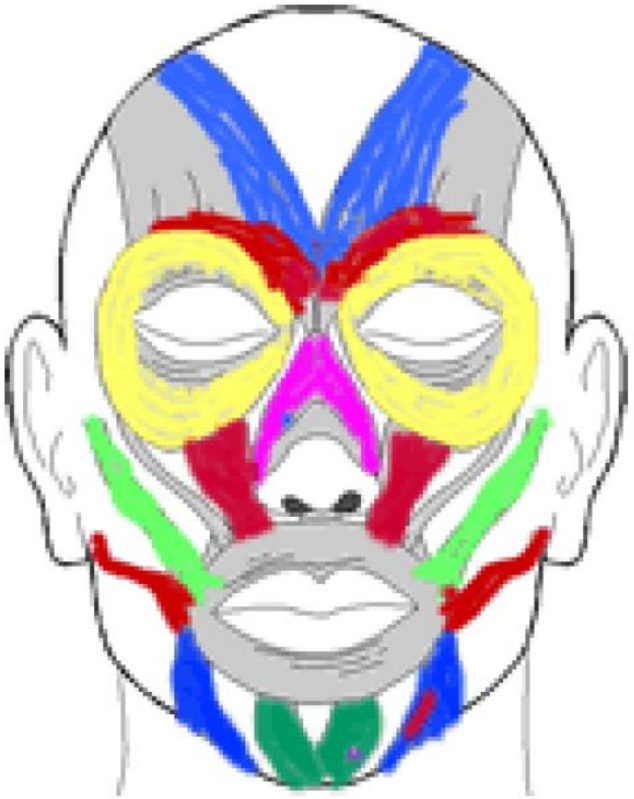


Methods of Facial Reocognition

- Early methods used optical flow to capture movement of features.(Such as facial muscles)
- Modern methods are generally visual feature-based

Early emotion detection method

- EMFACS (Ekman & Friesen,1973)
- Coded muscular features involved in different emotive expressions
- Tried to pick out the same features from videos
- Manual annotation



Weighted Saliency Maps

- Simple example of such a method. Uses pixel intensities of grayscale images.
- Calculates ratio of variance between classes and within a class.
- $\sigma k = \text{VarB}/\text{VarW}$, $k = 1,..., n$.
- VarB=Sum of (ClassMean - OverallMean)^{2,} for all classes and VarW=Sum of (f -MeanofClassof(f))^{2,} for all f. Here n is number of sample points.

Weighted Saliency Maps

- These ratios are then sorted in descending order .
- Above is an example for the top 500 features of each class for a particular sample

Speech Tone Recognition

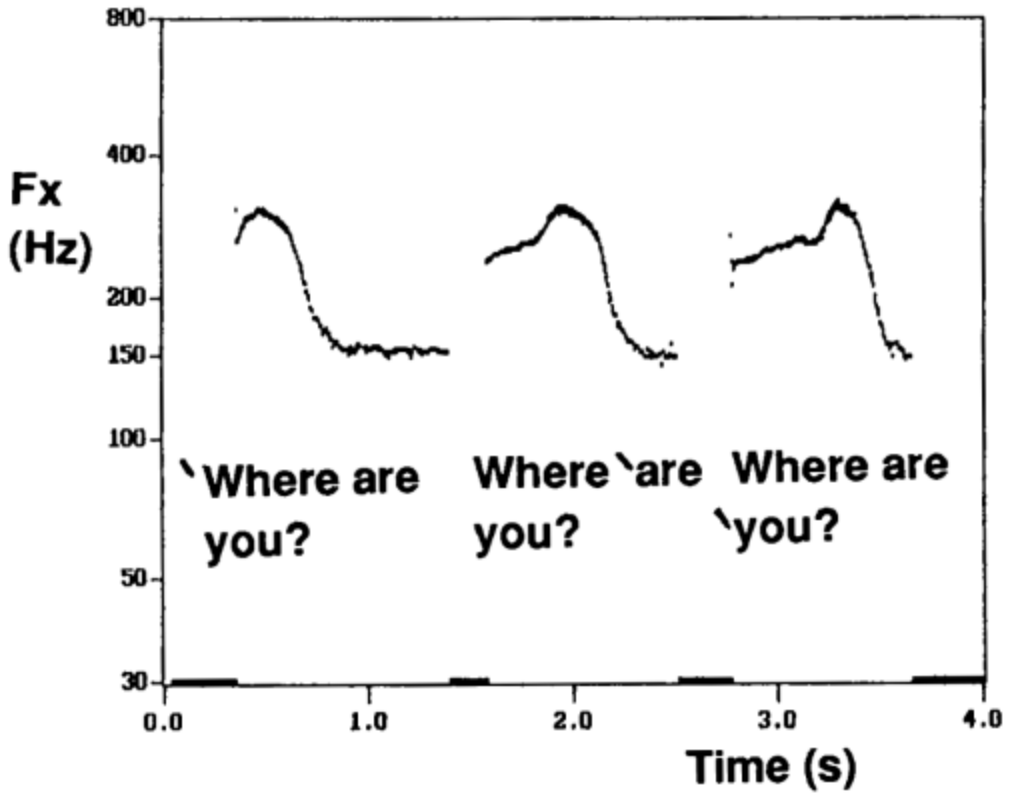
- Why have humanoid robots ?
  - Enjoyable interaction
  - Doesn't require training
  - Easier to teach
- Acoustic patterns contain :
  - Speaker identity
  - Speech content
  - **Speech tone**

Courtesy [7]

Abstraction of the problem

- Classify a given sentence to convey one of:
  - Approval : Good!
  - Prohibition : No!
  - Attention bidding : This looks interesting
  - Soothing : It'll be all right
  - Neutral : This is a book
- Approach: use Prosodic Contours

Using prosodic contours



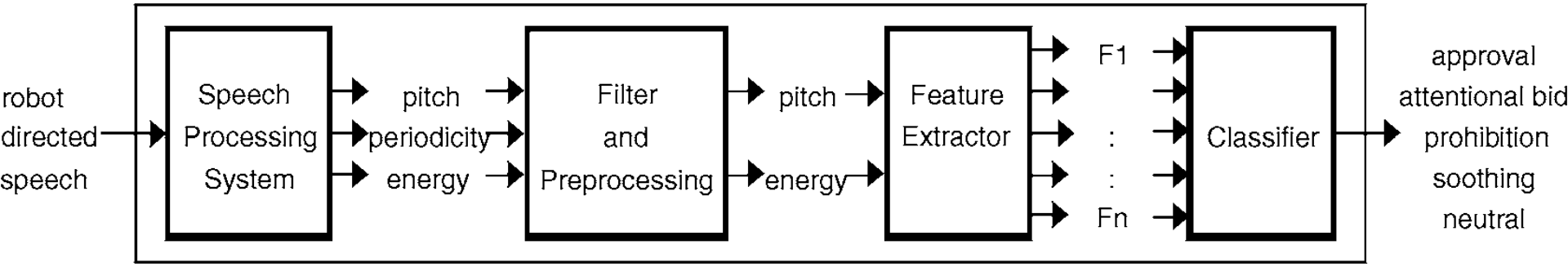
Have to normalize for pitch, length, and a bunch of other parameters

Using prosodic contours

- Intonation estimation
  - Glottal closure frequencies measured by laryngograph
  - Sampled at 1kHz
  - Can be approximated in practical applications
- Power curve estimates
  - Windowed sampling
  - Normalized to mean value of neutral style per speaker
- Voice activity estimation
  - Essentially just a noise filter
  - Trained to a threshold
- ML to the rescue

Courtesy [7]

Algorithm : Classify emotional content in speech



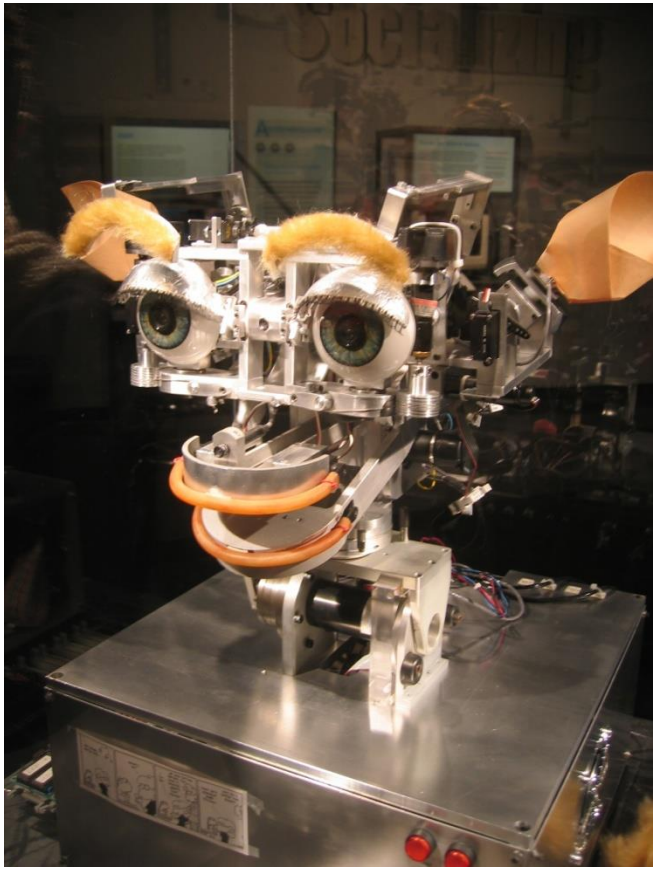
- Processing : tag sample with pitch, energy, percentage periodicity.
- Filter out noise : very high pitches (non-uniform), very low pitches.
- Calculate features (mean,variance of pitch,energy, pitch range)
- Pass to classifier for result.

Results

		INPUT						
		Ang	Fea	Sur	Dis	Joy	Sad	Neu
OUTPUT	Ang	92	-	-	1	2	-	-
	Fea	-	94	9	-	-	-	-
	Sur	-	3	86	-	-	-	-
	Dis	-	-	-	80	-	4	3
	Joy	2	-	-	-	88	-	1
	Sad	2	-	-	10	-	93	1
	Neu	1	-	-	6	7	-	92
	Acc(%)	94.9	96.9	90.5	82.5	90.7	95.9	94.9

KISMET

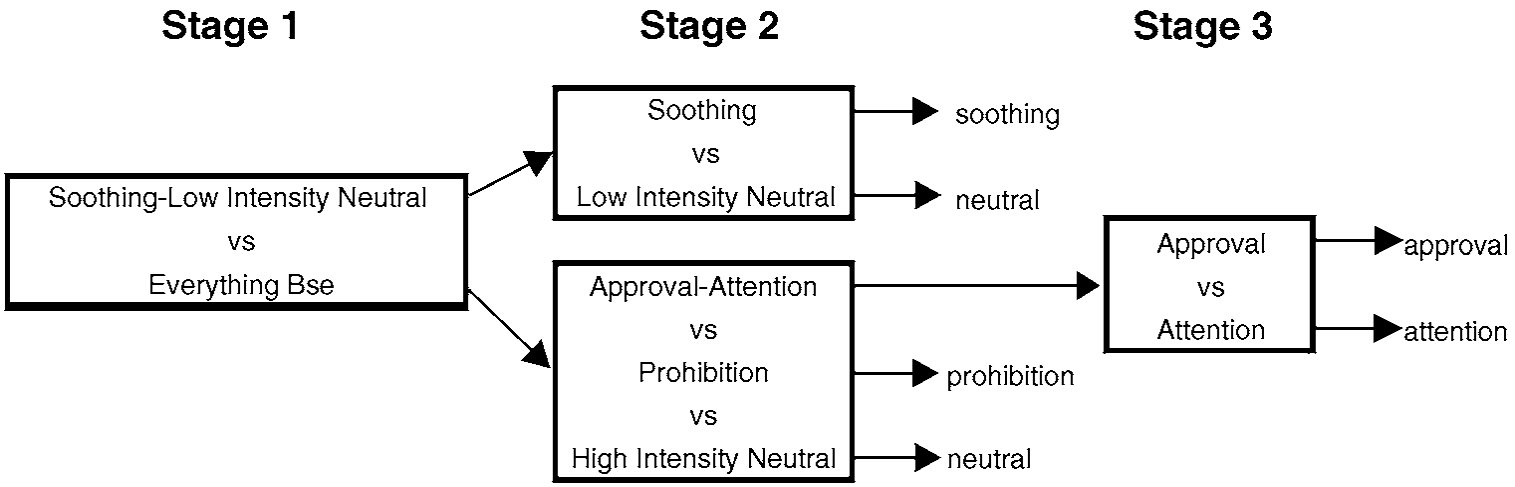
- Adaptive
- Physical emotion display
- Emotion lookup table
- Realistic
- Designed in the early 90s



[A,V,S] Emotion Model

- [Arousal , Valence , Stance] :- A 3-tuple models an “emotion”.
- Arousal:- Surprise at high arousal, fatigue at low arousal
- Valence:- Content at high valence, Unhappiness at low valence
- Stance:- Stern at closed stance, accepting at open stance

5-way classification in KISMET



- Stage 1 : Energy parameters are used to differentiate. (soothing, low-intensity neutral have low mean energy).
- Stage 2:
  - Using prosodic contours, soothing shows a smooth contour, frequency downsweep. Neutral is coarser and flatter.

Classification

- Purely rule-based
  - approval and attention both show high mean pitch, high pitch and energy variance
  - prohibition has low mean pitch but high energy variation
  - neutral shows low energy and pitch variation.
- Many rules really bulky and ad hoc
  - e.g. approval vs attention → both have high energy, and high pitch variation. But in approval, there is an exaggerated rise-fall pitch contour.

KISMET's response to emotion

- Has a synthetic nervous system (SNS) to help react to external stimulus.
- The 'somatic marker' process to tag incoming information with affective content.
  - Arousal : Level of emotional response
  - Valence : Is the stimulus+ve or -ve
  - Stance : How approachable is the percept?
- This information is passed to the 'emotion elicitor'.
- Emotional Elicitor : Each [A,V,S] input contributes to some emotion process. Eg, A large -ve valence might contribute to sad, anger, fear, distress emotions.

Kismet's Emotive Response Table

Prototype	Function of the Associated Behavior	Emotion Associated	Activation Conditions for Kismet
Incorporation	Accept environmental stimulus	acceptance, calm	Acceptance of a desired stimulus
Rejection	Get rid of something harmful already accepted	disgust	Attend to a salient but *undesired* stimulus
Protection	Avoid being destroyed	fear, distress	Appearance of a threatening, overwhelming stimulus
Deprivation	React against important loss	sorrow	Loss of a desired stimulus
Orientation	React to a new or strange object	interest	Appearance of new or *salient* stimulus
Exploration	Explore environment	boredom	Need of a desired yet absent stimulus
Reward	Reinforce beneficial behavior	joy	Success in achieving goal of active behavior
Destruction	Remove barrier to achieve some need	anger, frustration	Delay, difficulty in achieving goal of active behavior
Alert	Startle Response	surprise	Sudden, unexpected stimulus

Response calculation

- The winning emotion process affects the response if its value is above some threshold.
- Two thresholds, one for behavioural response, the other for response through expression (the latter is lower). This indicates that expression leads behavioural response.
  - On praise, first comes interest, and then physical alignment.

Response desiderata

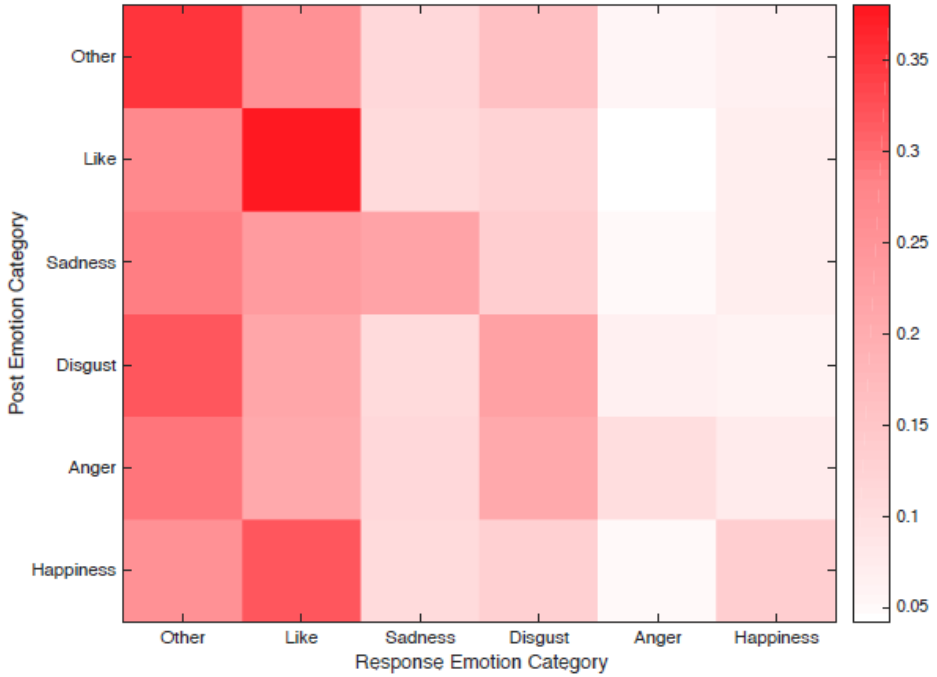
- Aesthetics : Appearance should affect nature of human communication with it.
- Real Time Perfomance : Long delays are not acceptable.
- Voice : Humans should be able to use their natural voice for training. It should be able to recognize a vocalization as having affective content when the intent of the sentence is to approve/prohibit, etc.

Response desiderata

- Unacceptable vs Acceptable misclassification: Shouldn't judge prohibition to be approval, but to judge it as neutral is an acceptable error.
- Expressive Feedback : Respond to emotion to let the person know it has understood.
- Speaker Dependence vs Independence: Former for personalized bots, latter for those that need to interact with many people.

Response specifications

- Empirical statement-response data from NLPCC 2014 dataset
  - 23105 emotion annotated sentences from Weibo



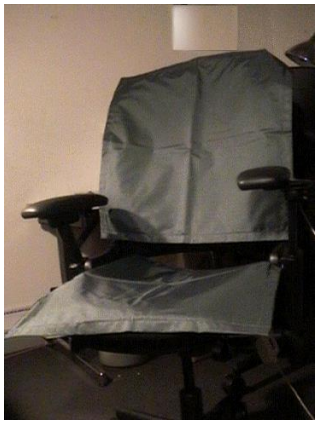
(Zhou et al, 2017)

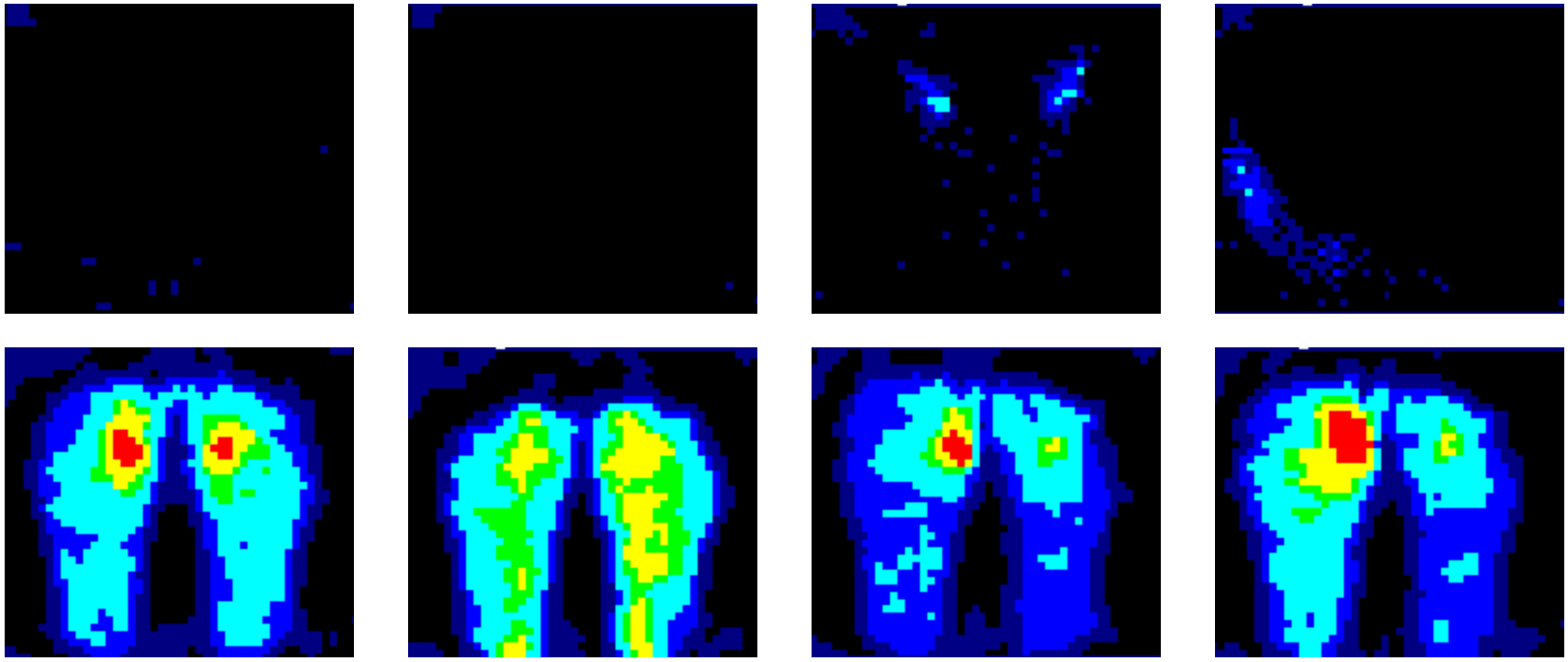
Dealing with embodiment

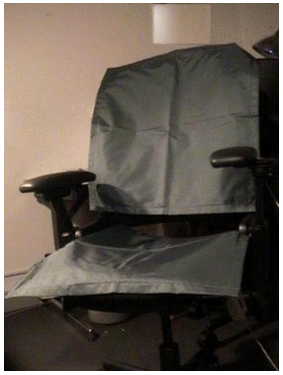
DETECTING EMOTIONS FROM THE BODY

Posterior inference (Mota & Picard, 2003)

Emotion detection using pressure sensors







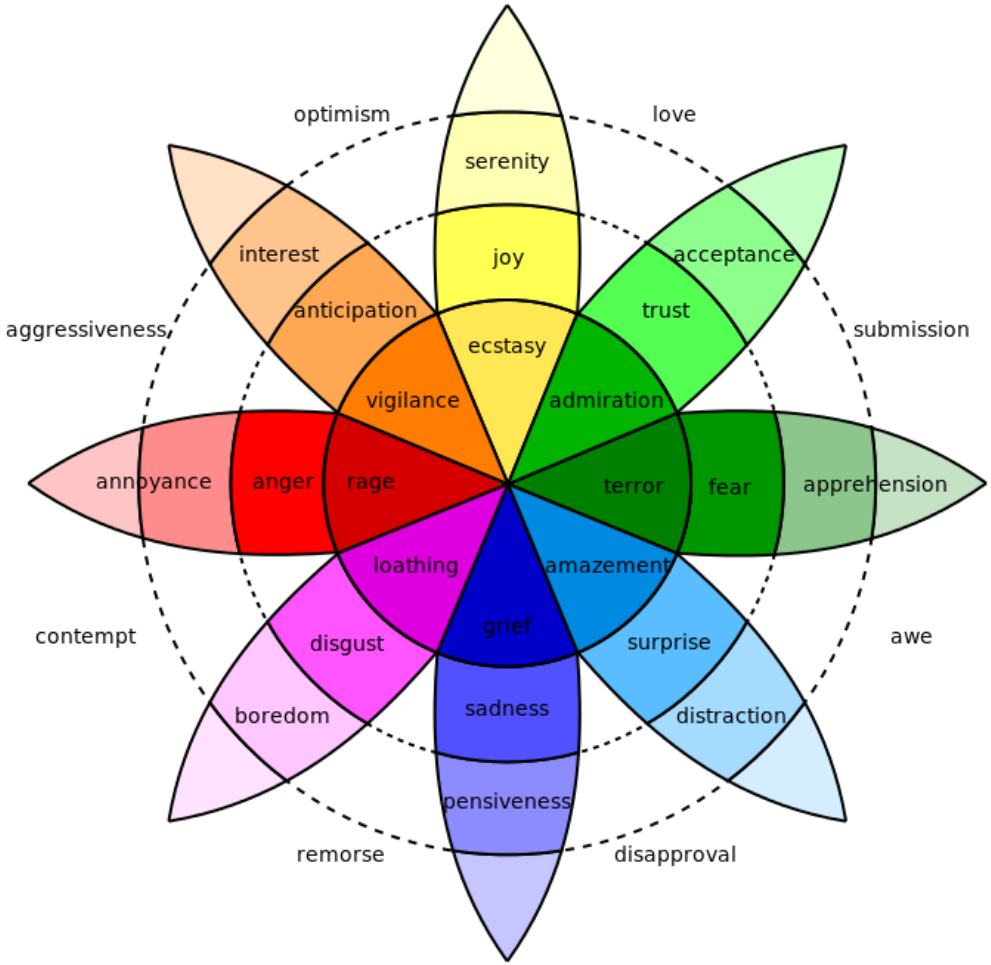
What can the sensor chair contribute toward inferring the student's state: *Bored vs. interested?*

Results (on children not in training data, Mota and Picard, 2003):

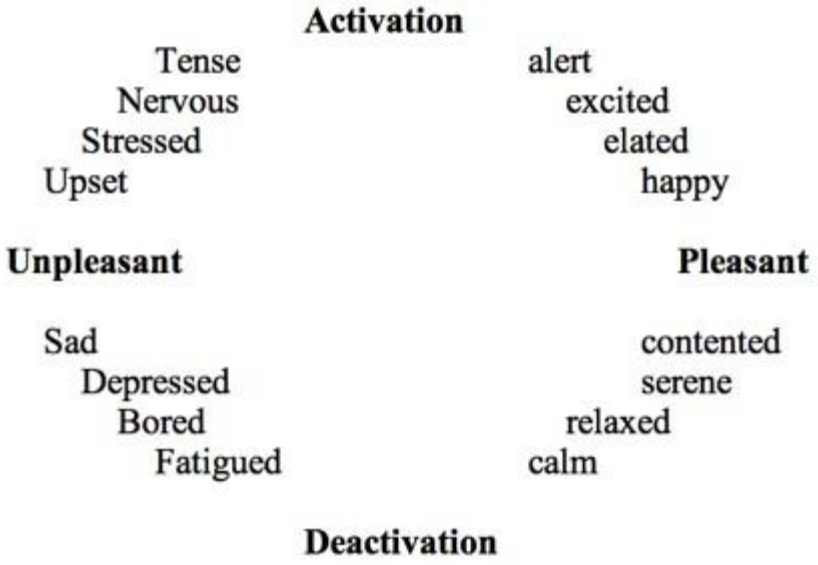
9-state Posture Recognition: 89-97% accurate

High Interest, Low interest, Taking a Break: 69-83% accurate

Models of emotion



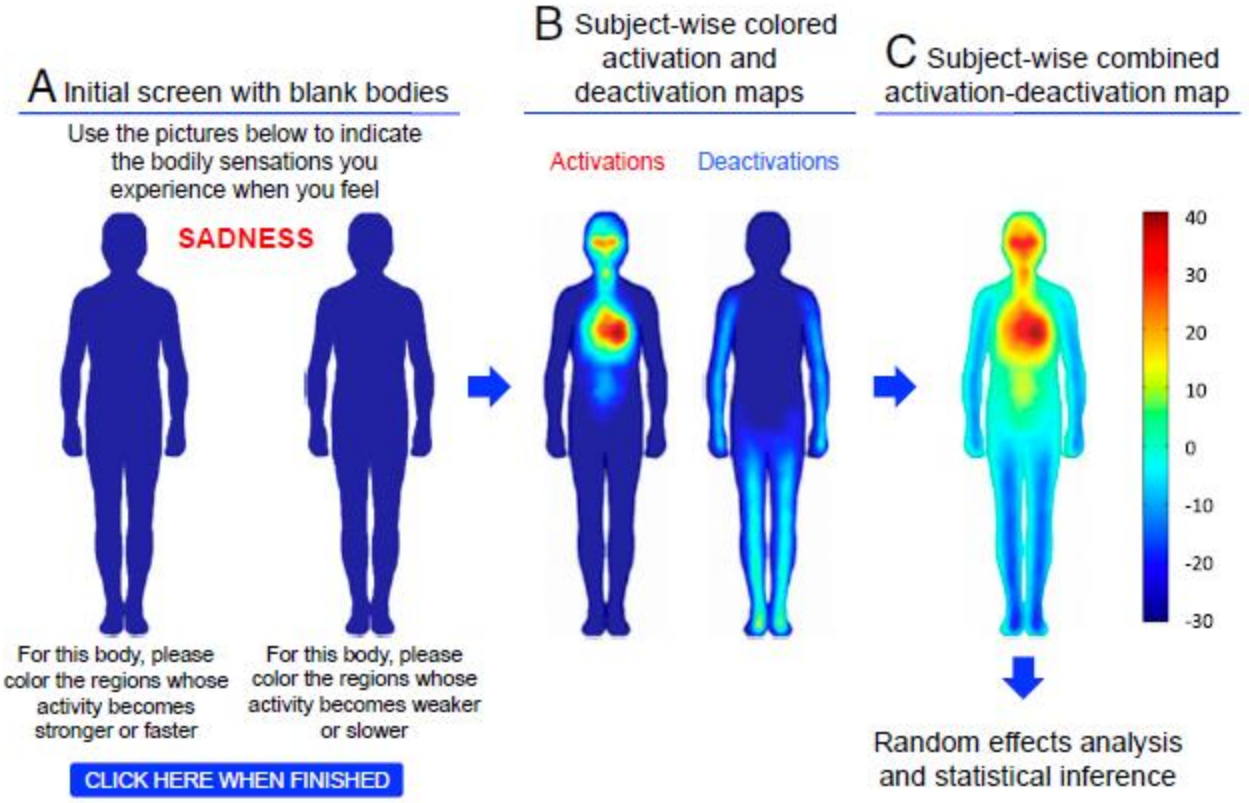
Plutchik's model



Theories of emotion

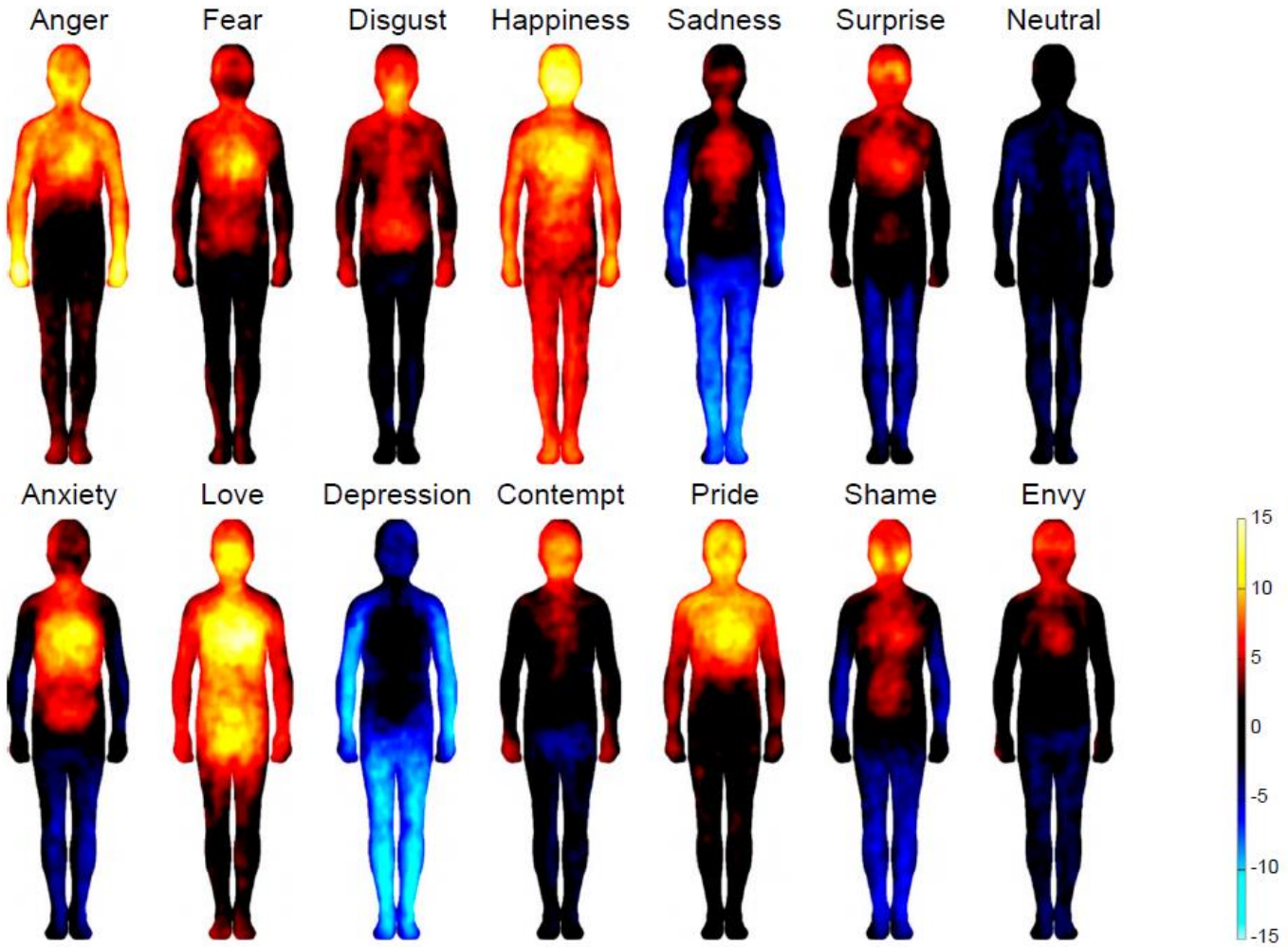
- James-Lange → Stimulus causes physiological response, perception of bodily change is emotion
- Cannon-Baird → Non-physiological independent basis of emotion
- Schacter → Physiological response directs cognitive situational appraisal, ends up being confounded with emotional state

Are emotions situated in the body?



(Nummenmaa et al, PNAS 2013)

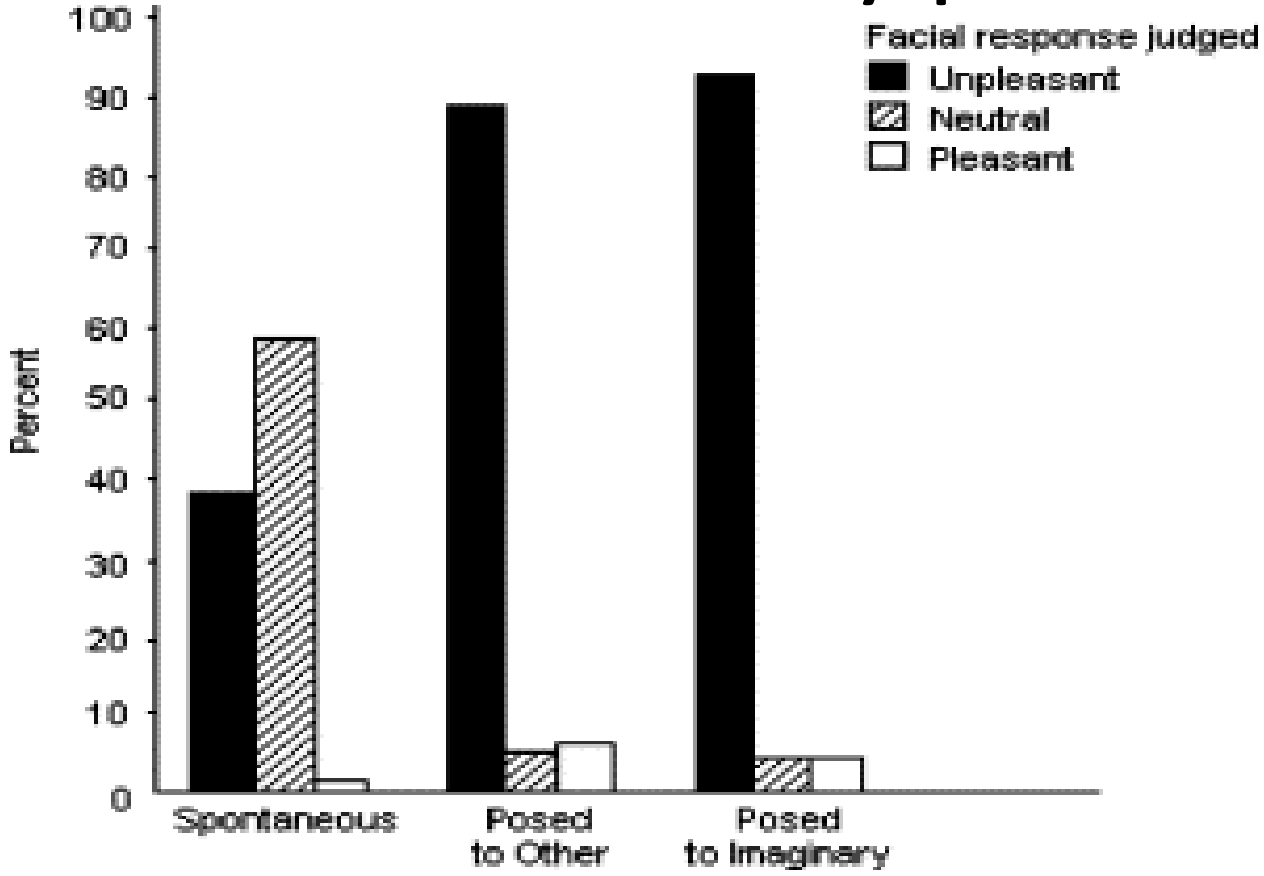
Bodily basis of emotions well-founded



(Nummenmaa et al, PNAS 2013)

Patterns consistent across Finns and Taiwanese

Emotions are heavily performative



In Erickson and Schulkin, 2003

Percentage of facial responses to unpleasant odour classified as unpleasant, neutral, or pleasant in a spontaneous condition, a posed to real person condition, and a posed to imaginary audience condition

The problem

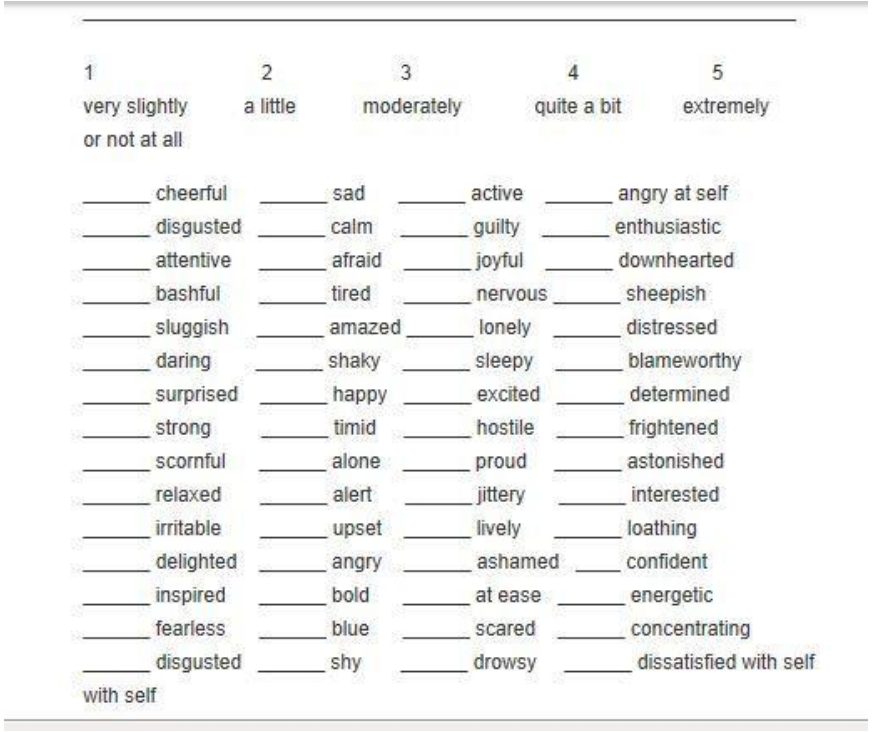
- Facial and speech-based models can detect expressed emotion
- Expressed emotion is heavily performative
- Felt emotion is heavily embodied
- How to
  - Measure felt emotion
  - Correlate with expressed emotion

Possible solutions

- Psychological surveys
- Physiological monitoring
  - Galvanic skin response
  - Blood volume pulse

Surveys

- Ask detailed questions
  - Ask same questions multiple ways
  - Use correlations to judge response authenticity
  - PANAS most common

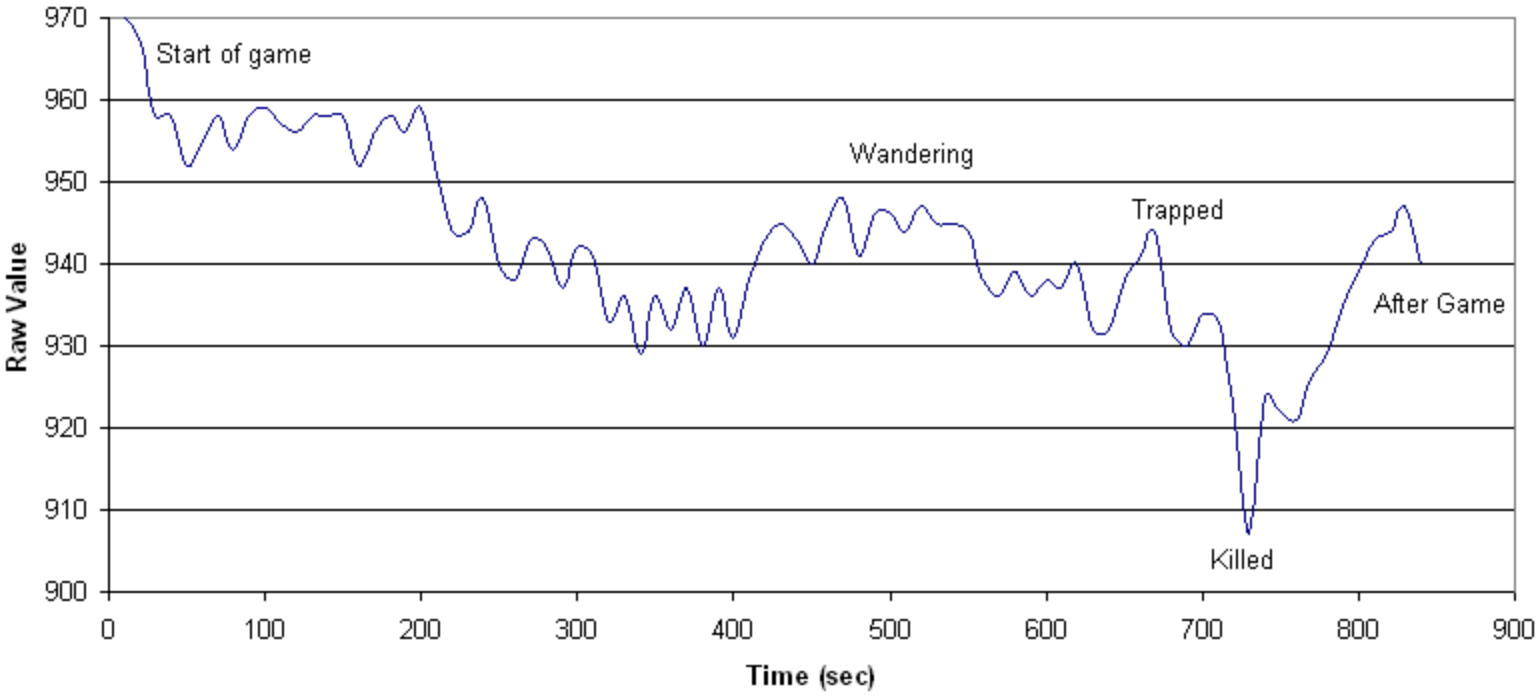


1	2	3	4	5
very slightly or not at all	a little	moderately	quite a bit	extremely

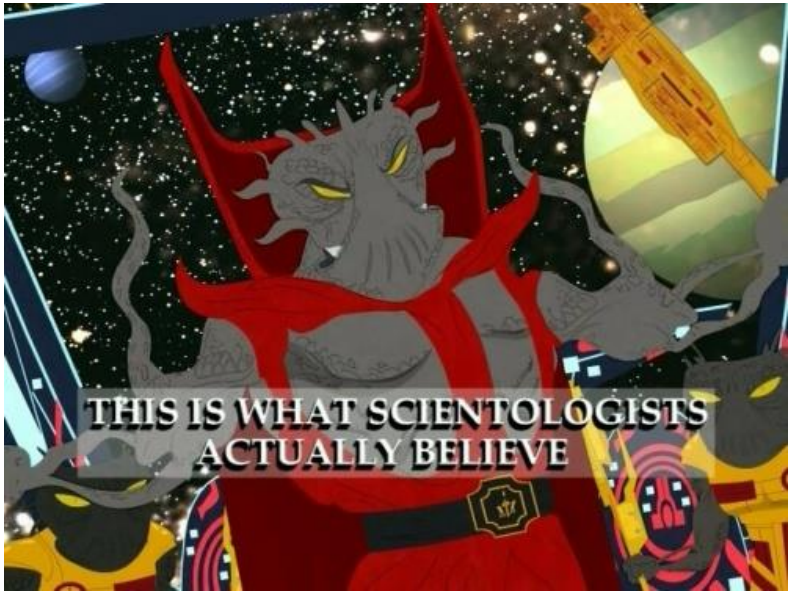
_____ cheerful	_____ sad	_____ active	_____ angry at self
_____ disgusted	_____ calm	_____ guilty	_____ enthusiastic
_____ attentive	_____ afraid	_____ joyful	_____ downhearted
_____ bashful	_____ tired	_____ nervous	_____ sheepish
_____ sluggish	_____ amazed	_____ lonely	_____ distressed
_____ daring	_____ shaky	_____ sleepy	_____ blameworthy
_____ surprised	_____ happy	_____ excited	_____ determined
_____ strong	_____ timid	_____ hostile	_____ frightened
_____ scornful	_____ alone	_____ proud	_____ astonished
_____ relaxed	_____ alert	_____ jittery	_____ interested
_____ irritable	_____ upset	_____ lively	_____ loathing
_____ delighted	_____ angry	_____ ashamed	____ confident
_____ inspired	_____ bold	_____ at ease	_____ energetic
_____ fearless	_____ blue	_____ scared	_____ concentrating
_____ disgusted with self	_____ shy	_____ drowsy	_____ dissatisfied with self

GSR

- Ancient technique
- Measures skin conductance
- Function of sweat gland activity
- Can measured generalized arousal



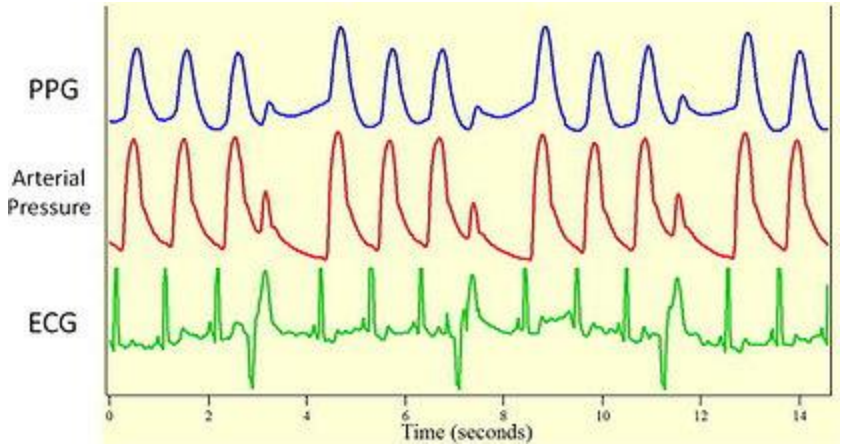
Existing uses



Picard Lab, MIT

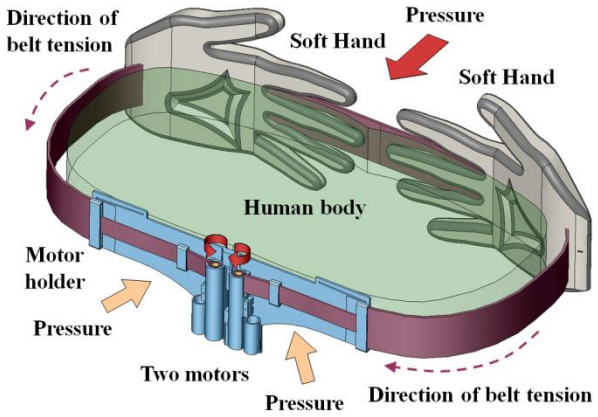
Photoplethysmography

- Blood volume pulse = measure of how much blood is flowing through an organ
- Measured using infra-red light reflection
- Can measure easily at fingers, but also at other body sites
- Similar uses as GSR



Possible applications

- E-learning
- Driver/pilot monitoring
- Neuromarketing
- Athletic performance monitoring
- Quantified self applications
- Digital co-presence



HaptiHug

Summary: emotion detection

- Machine learning methods are hands off
  - Scale well, give good accuracy metrics
  - Not ecologically grounded
- Biophysical methods are intrusive
  - Scale poorly, give good accuracy metrics but poor resolution of emotion space
  - Ecologically well-grounded
- Surveys offer great resolution of emotion space
  - Scale terribly or not at all, very resource inefficient
- Applications
  - Context-sensitivity in human-computer interaction
  - Monitoring human performance/engagement on tasks
  - Remote projection of affect