Gridspace

IAP Program 2023
Emotion, Dialog Acts, Personality and Lying

January 31st 2023

Human Speech

So Far...

- Audio, Text and Language Modelling
- NLP models
- ASR -> Dialog System -> TTS system

Human Speech

So Far...

- Audio, Text and Language Modelling
- NLP models
- ASR -> Dialog System -> TTS system

What about....

- Speech **Content?**

Road Map

- Emotion
- Dialog Acts
- Deception
- Health
- Mental Health
- Personality

Road Map

Textual Features

- Emotion
- Dialog Acts
- Deception
- Health
- Mental Health
- Personality

Road Map

Textual Features
Audio Features

- Emotion
- Dialog Acts
- Deception
- Health
- Mental Health
- Personality

Road Map

Textual Features
Audio Features
Visual Features

- Emotion
- Dialog Acts
- Deception
- Health
- Mental Health
- Personality

0.
What is Emotion

Emotion vs Semantic Content

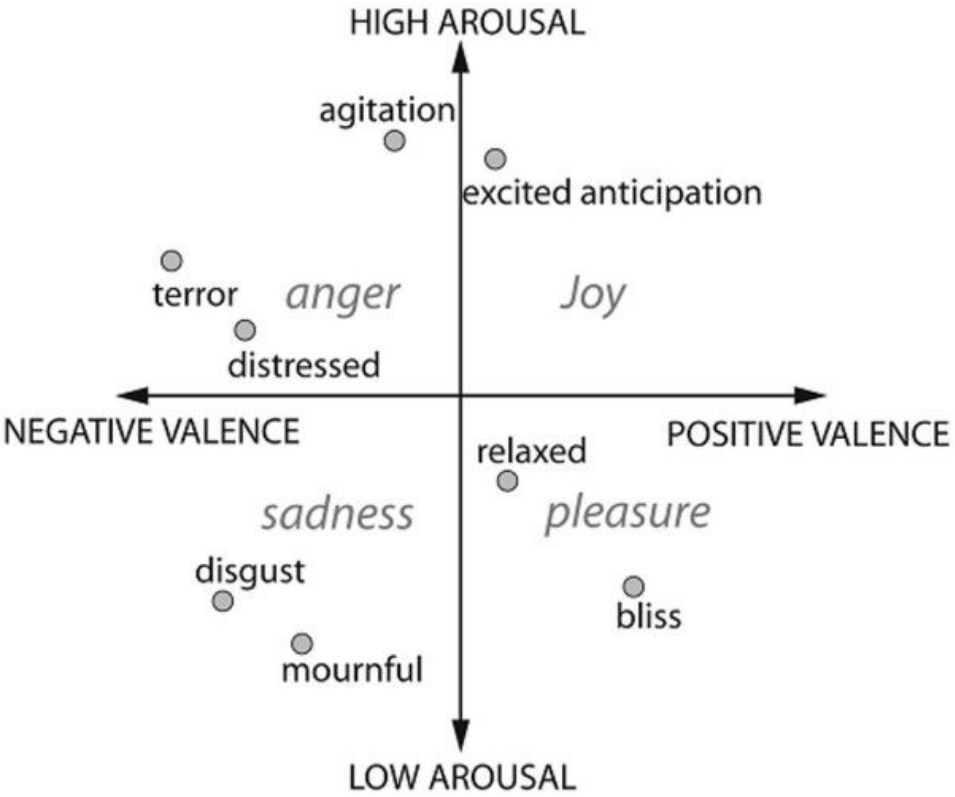
Semantic analysis:

- text based content
- What are they *saying*

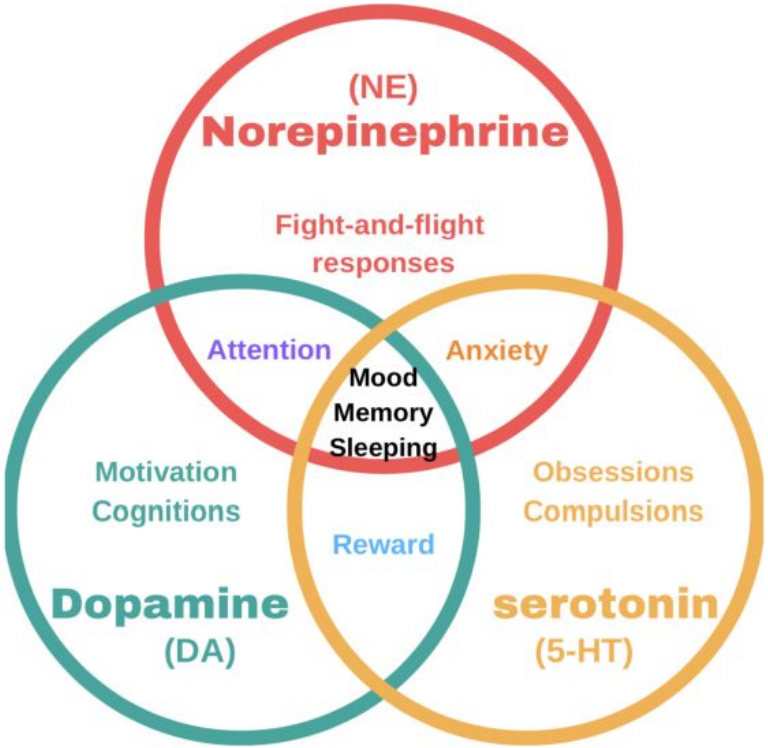
Emotional state:

- Non-linguistic and linguistic cues
- What are they *feeling*

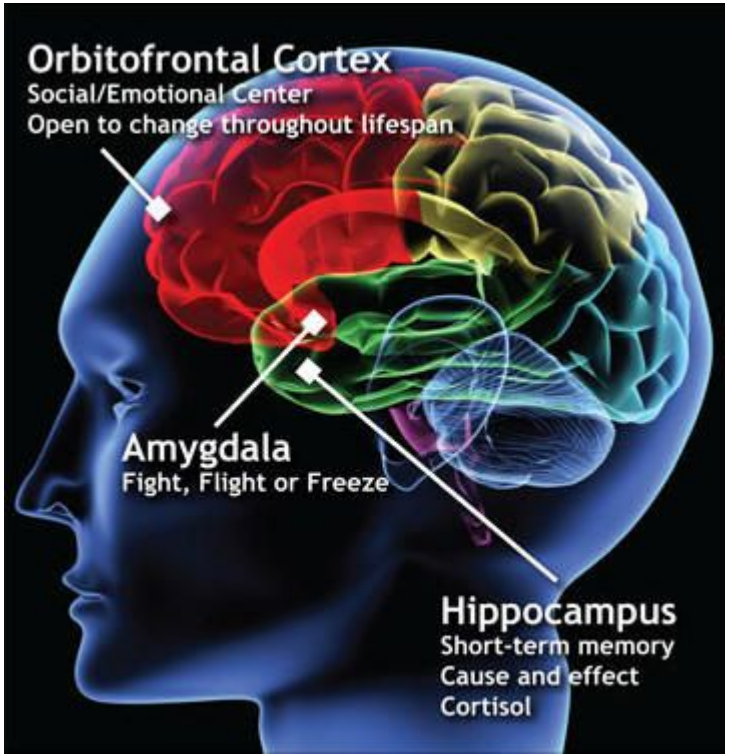
Russell's Circumplex Model - 1980



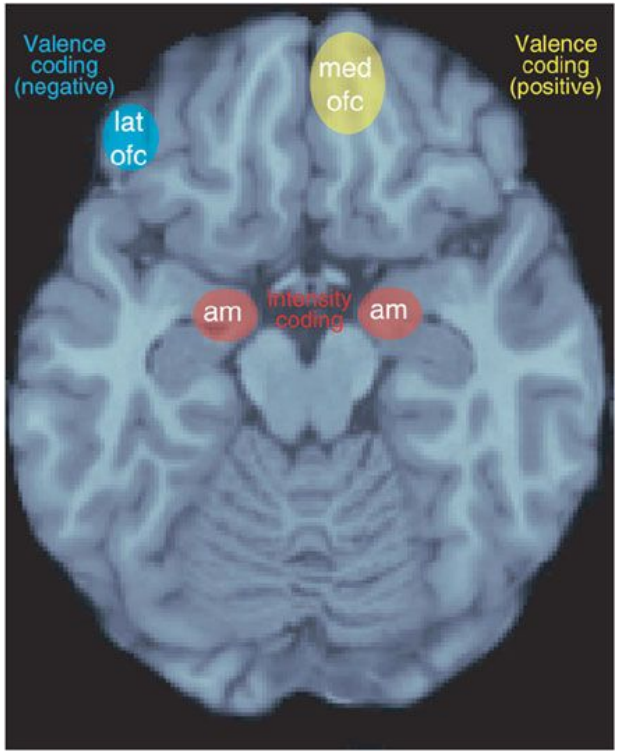
Emotion - A chemical inconvenience



Emotion - A chemical inconvenience



Emotion - A chemical inconvenience



Emotion - A biological advantage

Emotion - A societal advantage

Bot-world

- Why do we care about emotion?
  - Humans respond to emotional tone before language
  - “It sounds robotic” = “it sounds monotone”

Bot-world

- Why do we care about emotion?
  - Humans respond to emotional tone before language
  - “It sounds robotic” = “it sounds monotone”
- A bot should be able to:
  - Detect emotion in the same way a human can
  - Change action depending on that emotion

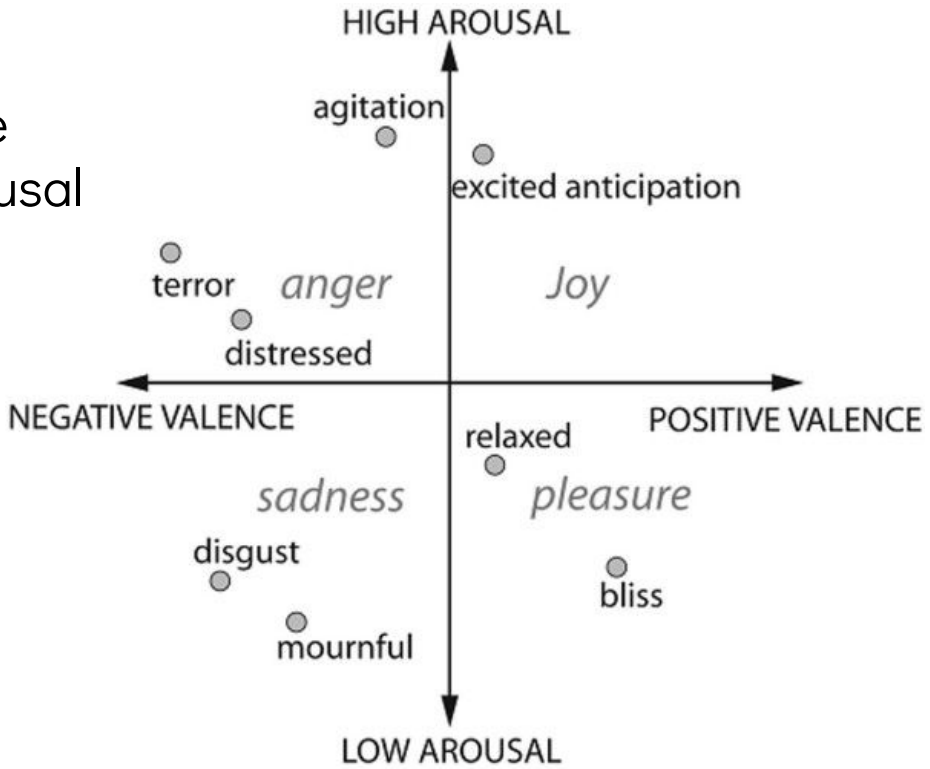
2.
Turning it into a Vector
(because ~maChINe lEaRNinG)

Three Views

- Dimensional
  - Continuous space

Three Views

- Dimensional
  - Continuous space
  - E.g. Valence - Arousal



Three Views

- Dimensional
  - Continuous space
  - E.g. Valence - Arousal
- Categorical
  - Distinct, independent categories
  - E.g. angry, disgusted, worried, happy, ...

Three Views

- Dimensional
  - Continuous space
  - E.g. Valence - Arousal
- Categorical
  - Distinct, independent categories
  - E.g. angry, disgusted, worried, happy, ...
- Componental/Cognitive Appraisal Theory
  - Emotion is a function determined at time of evaluation of incoming stimulus, based on personal relevance.
  - E.g. a function of valence, novelty, goal relevance, goal congruence and coping potential.

Three Views

- **Dimensional**
  - Continuous space
  - E.g. Valence - Arousal
- **Categorical**
  - Distinct, independent categories
  - E.g. angry, disgusted, worried, happy, ...

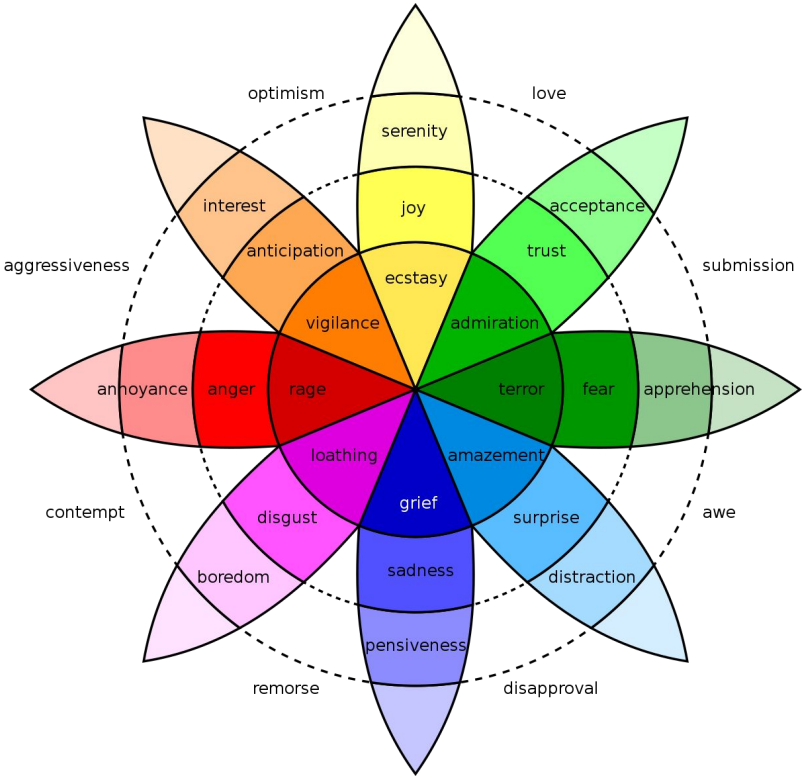
Tomkins - 1962

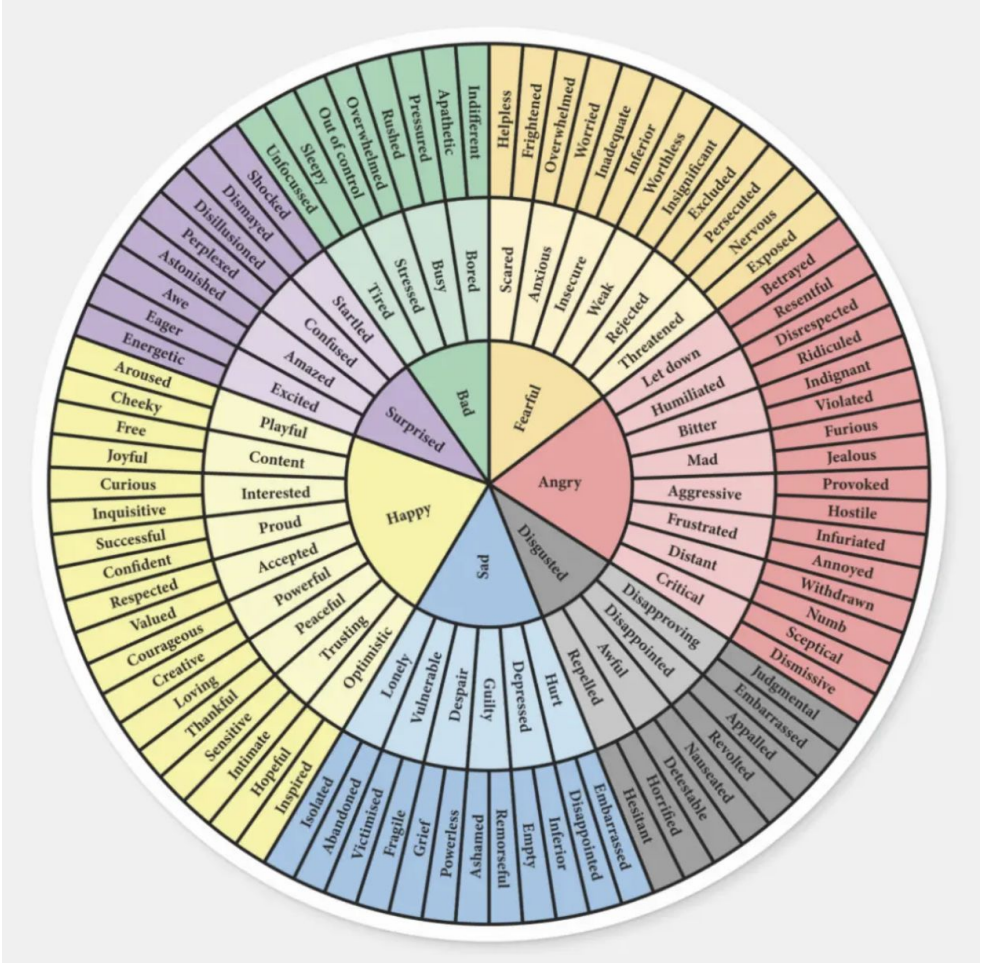
- “ limited number, genetically preprogrammed into the brain, and triggered by changes in stimulation”
  - **Excitement**
  - **Joy**
  - **Surprise**
  - **Distress**
  - **Disgust**
  - **Anger**
  - **Shame**
  - **Fear**

Izard and Ekman's Theories

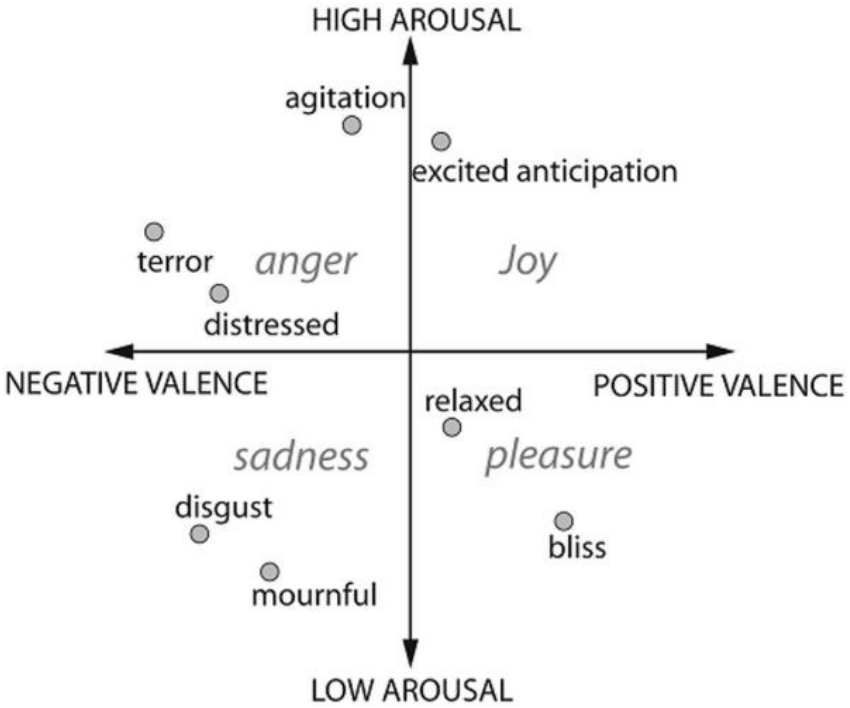
Ekman	Izard
Anger	Anger
Disgust	Contempt
Fear	Disgust
Happiness	Fear
Sadness	Guilt
Surprise	Interest
	Joy
	Sadness
	Shame
	Surprise

Plutchik's Theory: Psycho-Evolutionary

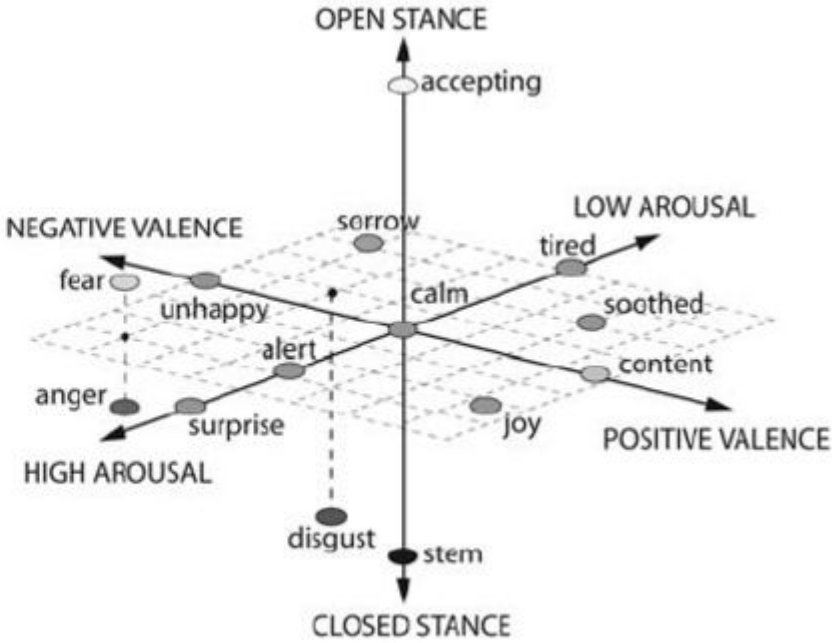




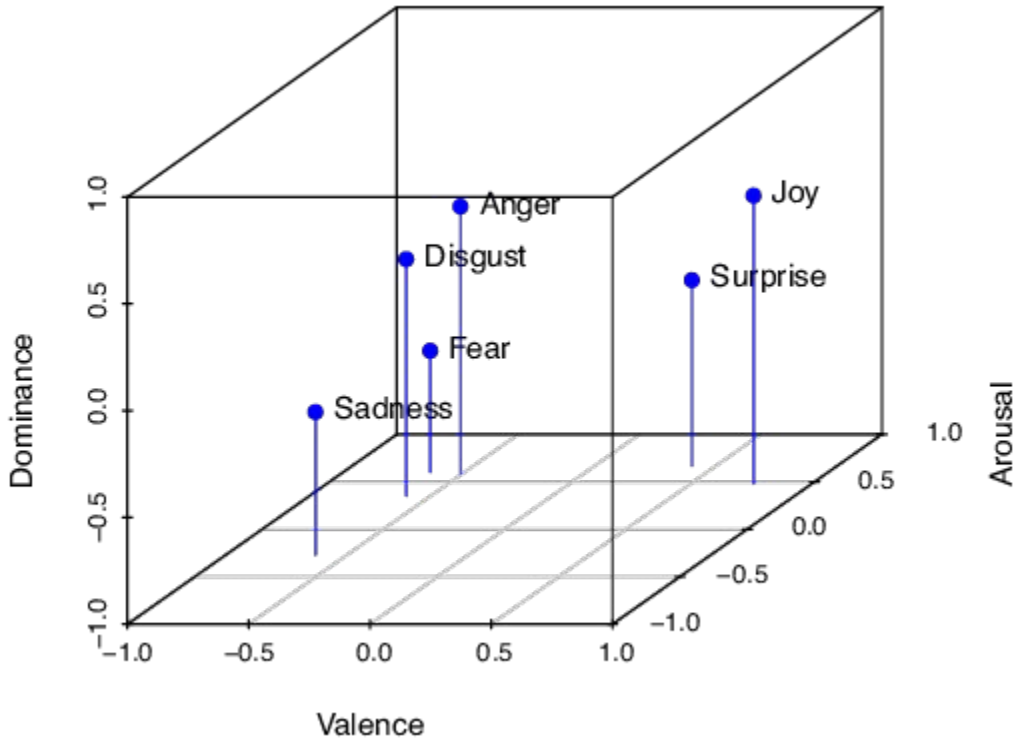
Valence-Arousal Model (Dimensional)



Body Language

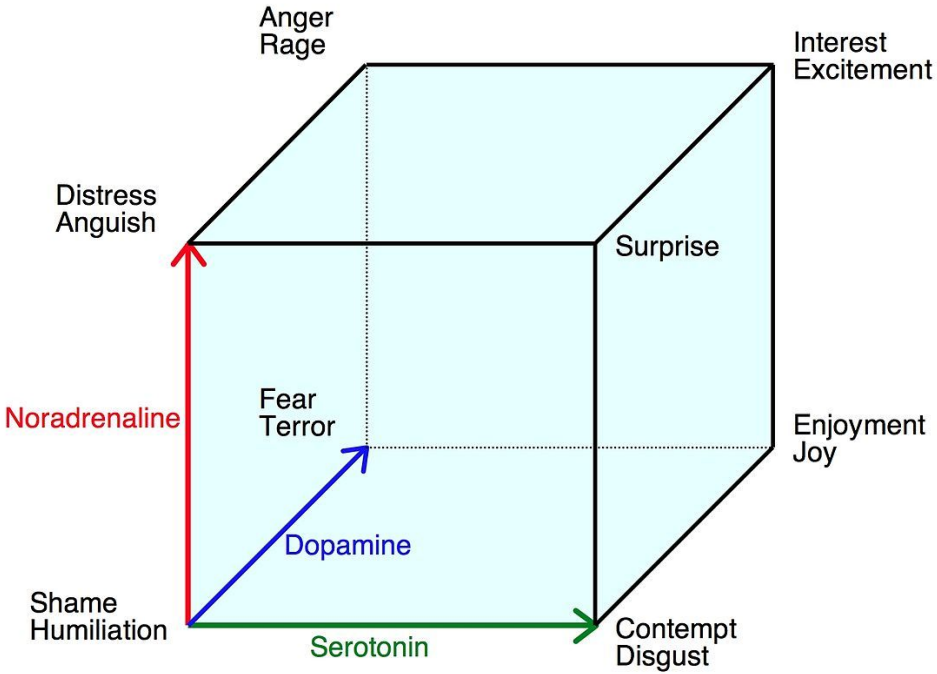


VAD Model

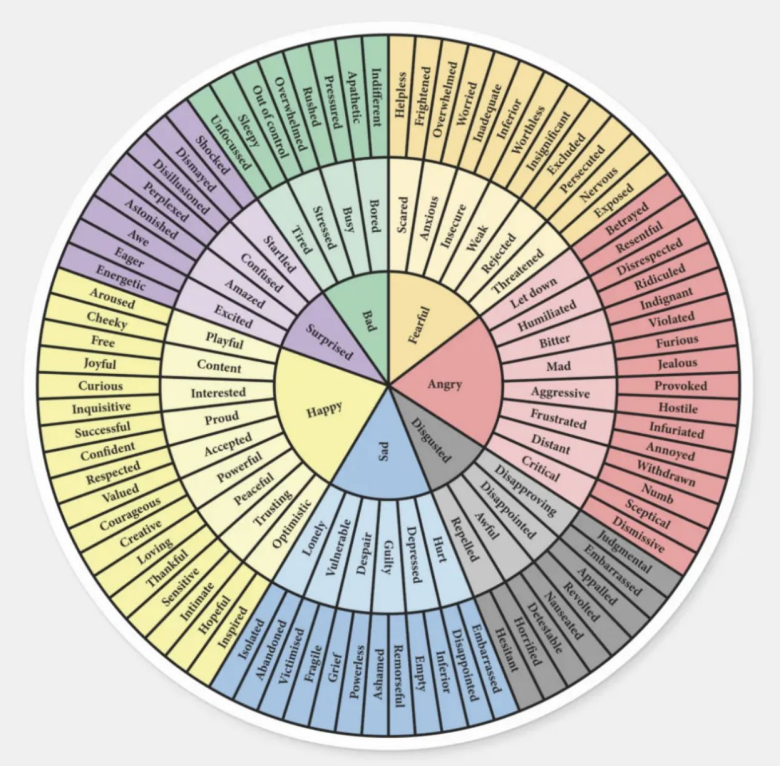


Lovheim Cube of Emotions:

Monoamine Neurotransmitters



How many dimensions do we need?



3.
What are Dialog Acts?

Dialog Act Meaning

"Meaning is use...utterances can only be explained in relation to the activities, or language-games, in which they participate"

- **Wittgenstein 1958**

Type of **Speech Act**

Dialog Act Meaning

Examples:

- Greeting/Closing
- Statement
- Fillers and Pleasantries
- Opinion
- Agreement
- Question
    - Yes/No Question
    - Request for Information

Dialog Act Meaning

Importance for Humans:

- **Societal** cornerstone
- Road Map a conversation
- How To Make Friends And Influence People

Importance in Dialog Systems:

- Conversation Management (more on this on Wednesday)
- Turn taking and signalling
- How To Make Friends And Influence People (for Bots)

4.
Prosodic Features

Prosody and Emotion

- Pitch ~= intensity
- Down/upturn ~= valence
- “Cold” anger vs “hot” anger
- Sarcasm and humour

Prosody and Dialog Acts

This is a question
This is a question?

Prosody and Dialog Acts

This is a question
This is a question?

...enough said

Prosody and Dialog Acts

This is a question
This is a question?
Sarcasm

Swear Words

- Actual words are almost irrelevant
- Highly affective language
- Tone has huge impact

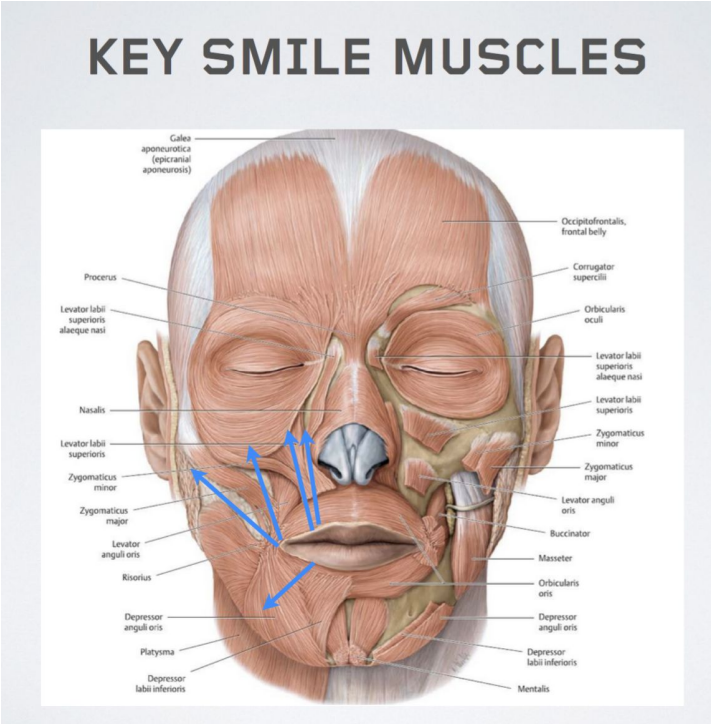
Can I tell if you're lying?

- Uptalk and monotonicity
- "Truthful" vs "untruthful" language
- Use of hyperbole or 'performative' language
- Overexplaining

Cultural Variation

- Different cultures have different prosodic trends
- Tone in a language you don't know
- Multilingual speakers may associate each language differently

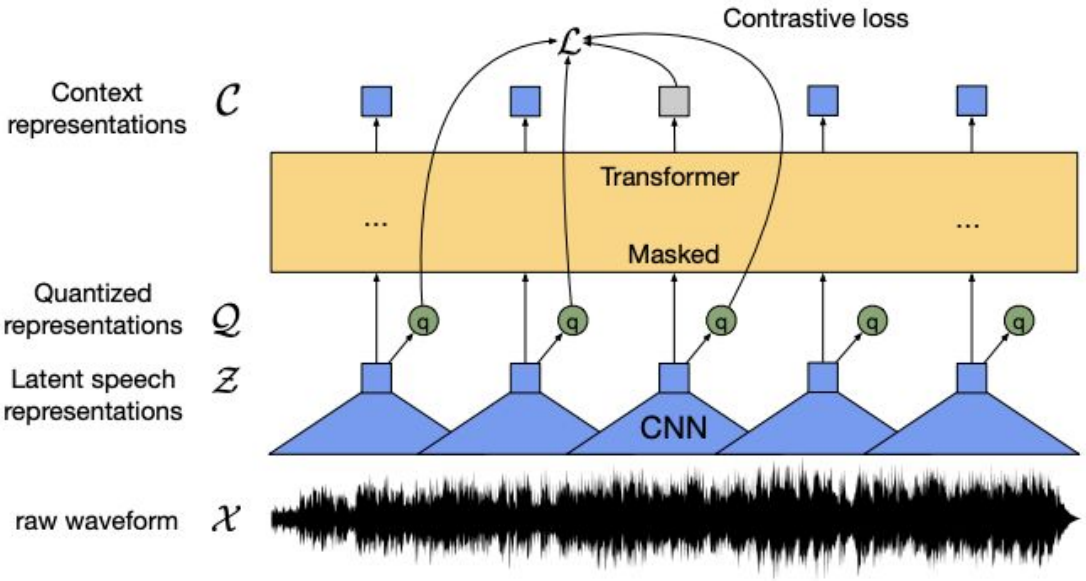
Smiling Demo



“Ability to process prosodic cues for interactional coordination”

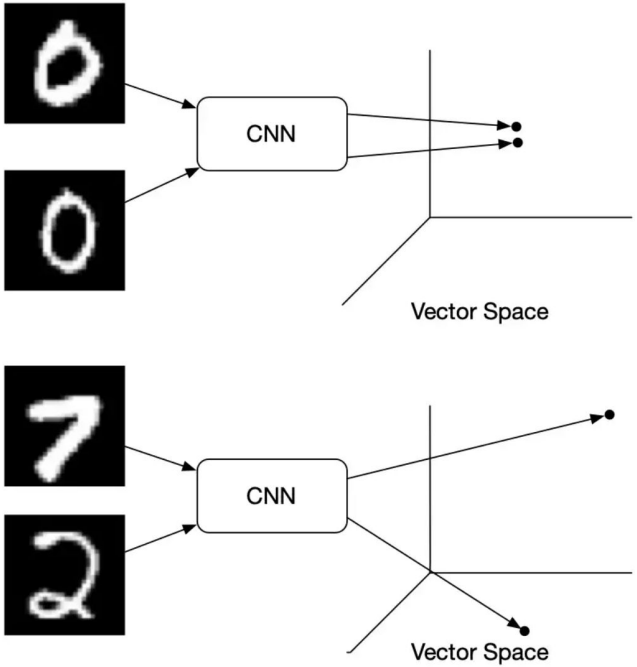
3.
Speech Emotion Recognition (SER)

Wav2Vec 2.0 Embeddings



- Self supervised training
- Contrastive + diversity loss
- Quantization of latent space
- BERT-style masking for training

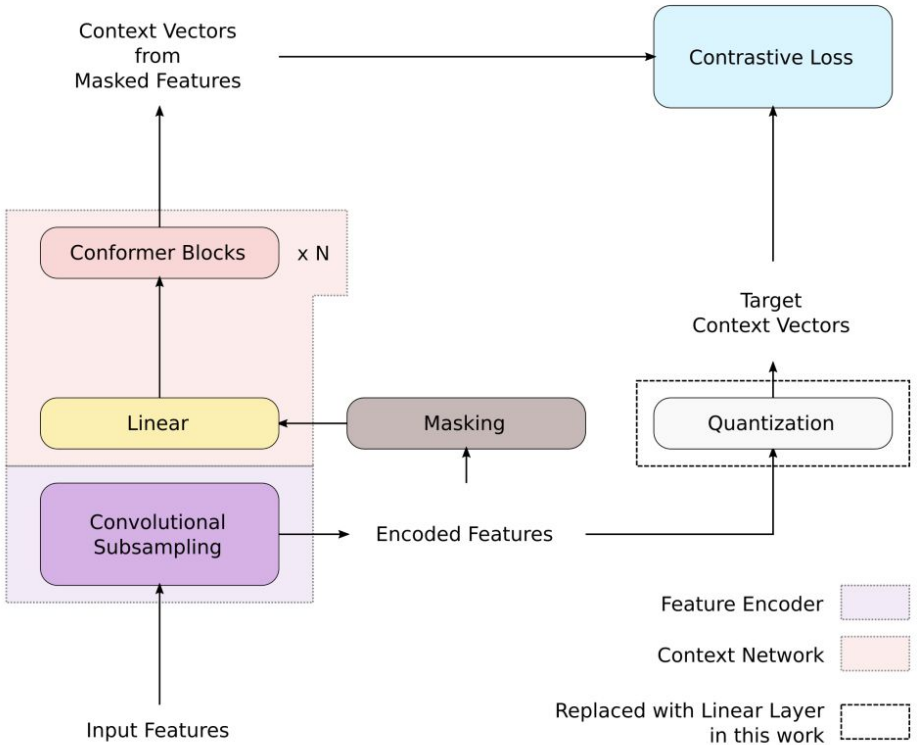
Contrastive Loss



$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \underset{[k \neq i]}{} \exp(\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)},$$

https://towardsdatascience.com/contrastive-loss-explaned-159f2d4a87ec

Wav2Vec 2.0 With Conformer



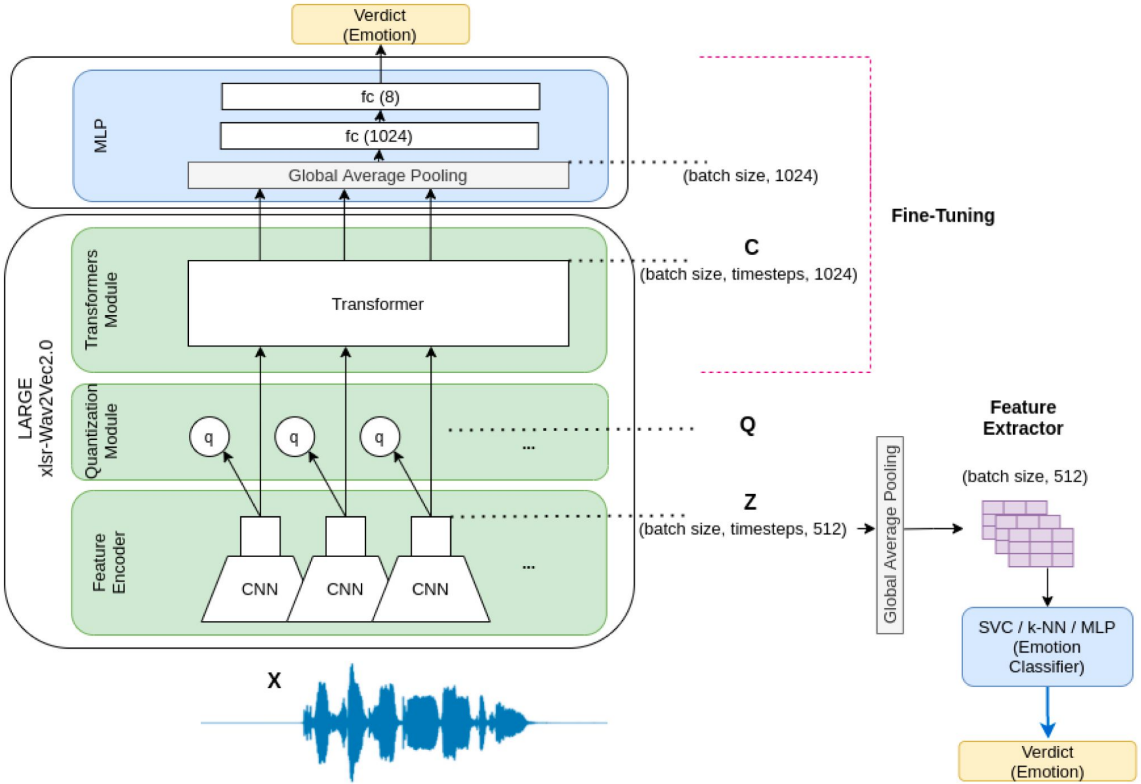
Google Research, Brain Team
{ngyuzh, jamesqin,
danielspark, weihan,
chungchengc, rpang, qvl,
yonghui} @google.com

Categorical Emotion

Some considerations in choosing categories:

- Lexicon has more negative words than positive
- Emotions with most distinct audible features
- Existing theories on categories
- Specific use cases

Classification



C. Zhang and L. Xue, "Autoencoder With Emotion Embedding for Speech Emotion Recognition," in *IEEE Access*, vol. 9, pp. 51231-51241, 2021, doi: 10.1109/ACCESS.2021.3069818.

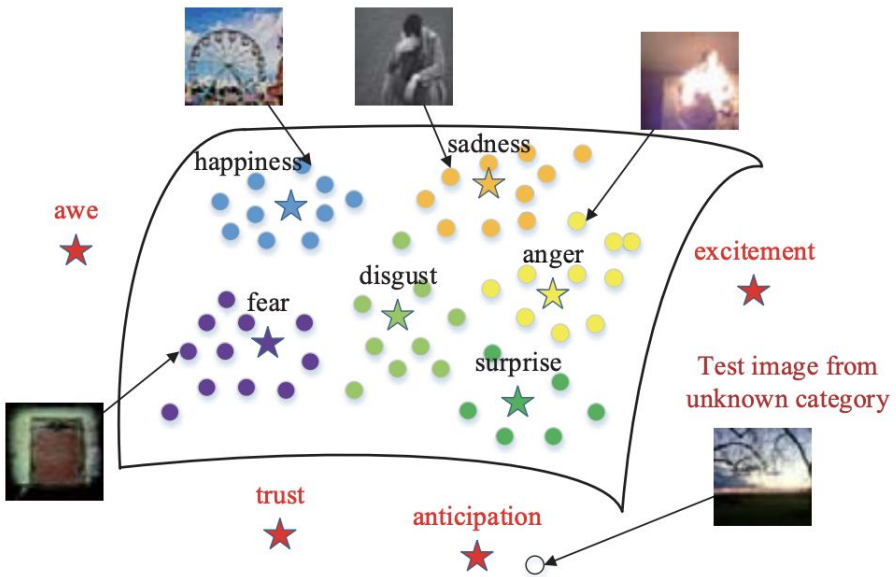
Variations and Improvements

- Teacher-student model
- Incorporating visual embeddings
- Incorporating textual embeddings

Data

Emotion datasets are small
Emotion categories are non-exhaustive
Zero Shot learning!

2019: Zero Shot Learning via Affective Structural Embedding



https://openaccess.thecvf.com/content_ICCV_2019/papers/Zhan_Zero-Shot_Emotion_Recognition_via_Affective_Structural_Embedding_ICCV_2019_paper.pdf

5. Emotion Tracking and Actioning

Maintenance of State

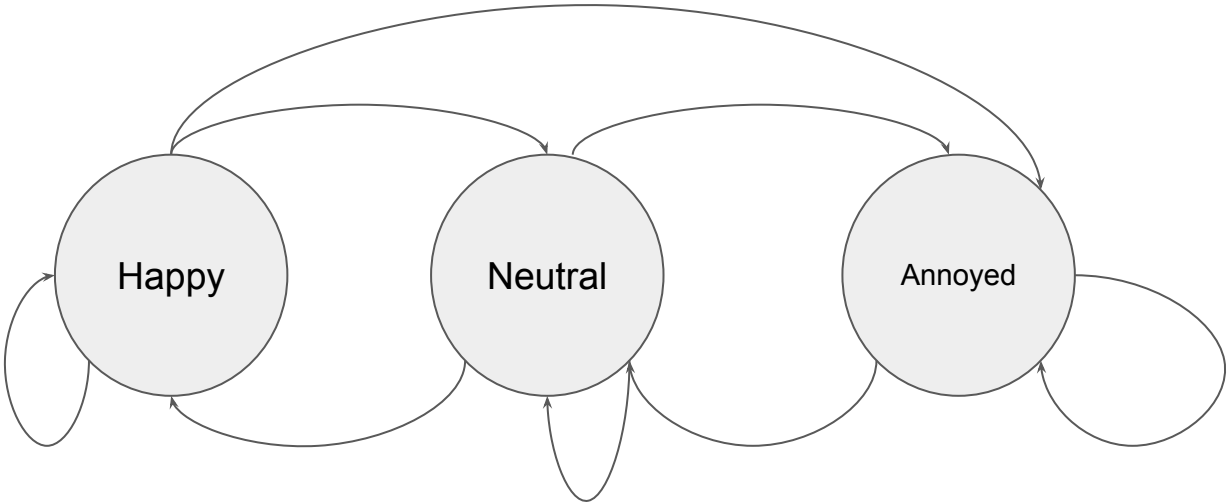
Approach 1: Numerical

- Regularization to prefer outputs with small deviations
- Vector space does not change drastically

Approach 2: Categorical State Space

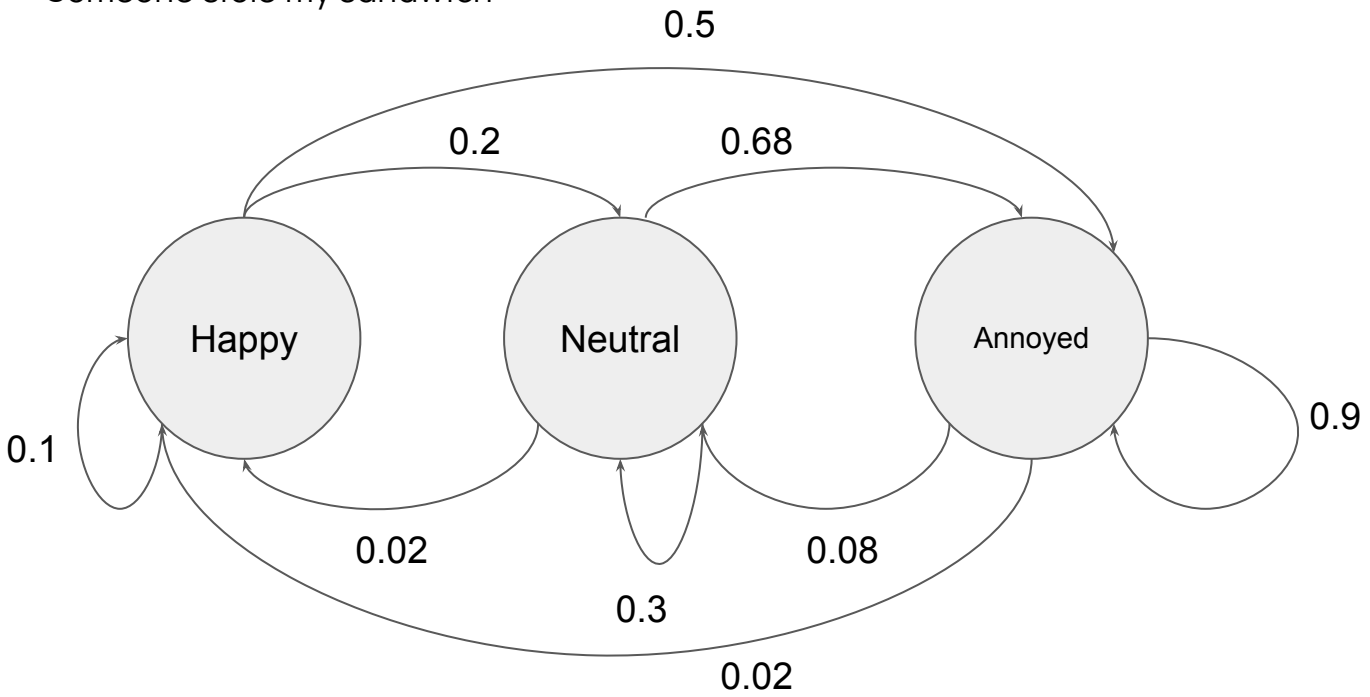
- We exist in an emotion state and, with associated probabilities, change categorical state accordingly (HMM)

Maintenance of State: Categorical



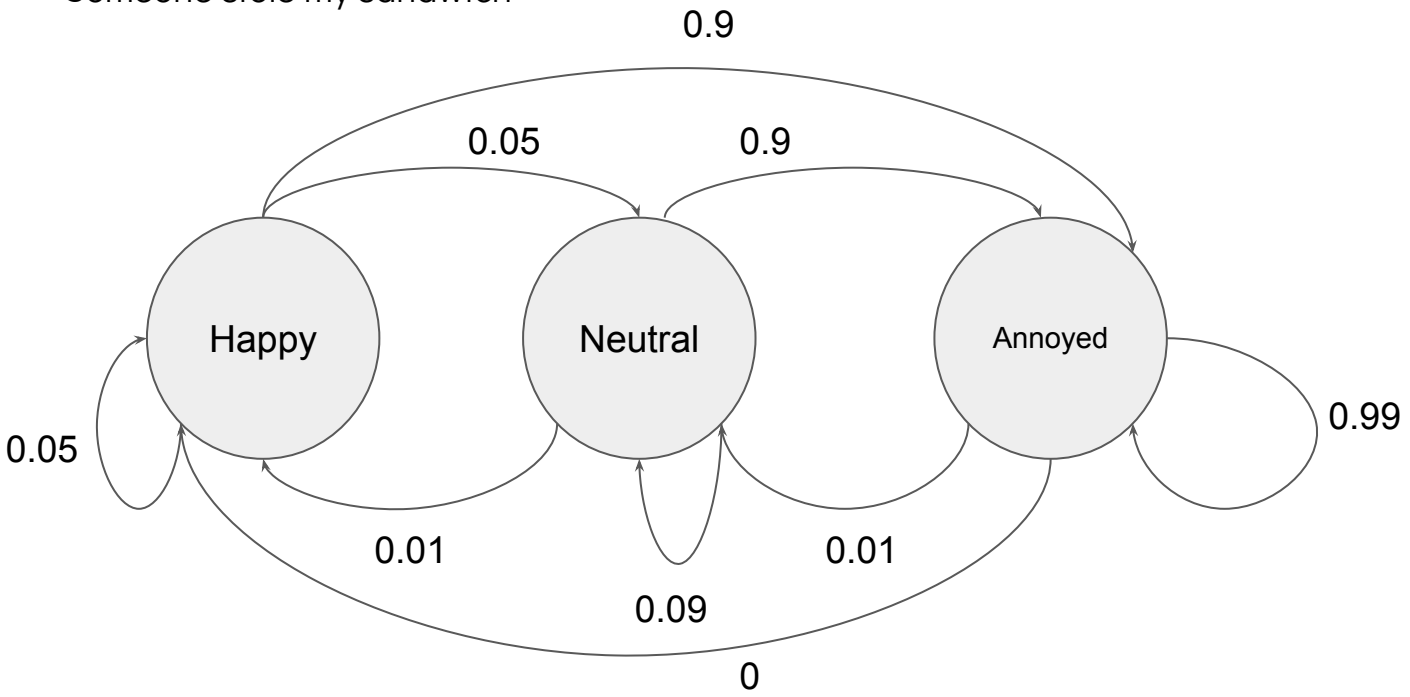
Personality Effect: Average

- “Someone stole my sandwich”



Personality Effect: Easily Irritated

- "Someone stole my sandwich"



Personality Effect

Other things that might affect the probability space:

- Time of day
- Hunger levels
- Type of sandwich
- Person who stole it
- Ease of replacement
- ....

Solomon and Corbit: Opponent-Process Theory

the primary or initial reaction to an emotional event will be subsequently followed by an opposite secondary emotional state.

- pleasure/pain
- depression/elation
- fear/relief

Emotional State & Personality Policy

Does this remind you of anything?

Emotional State & Personality Policy

Does this remind you of anything?

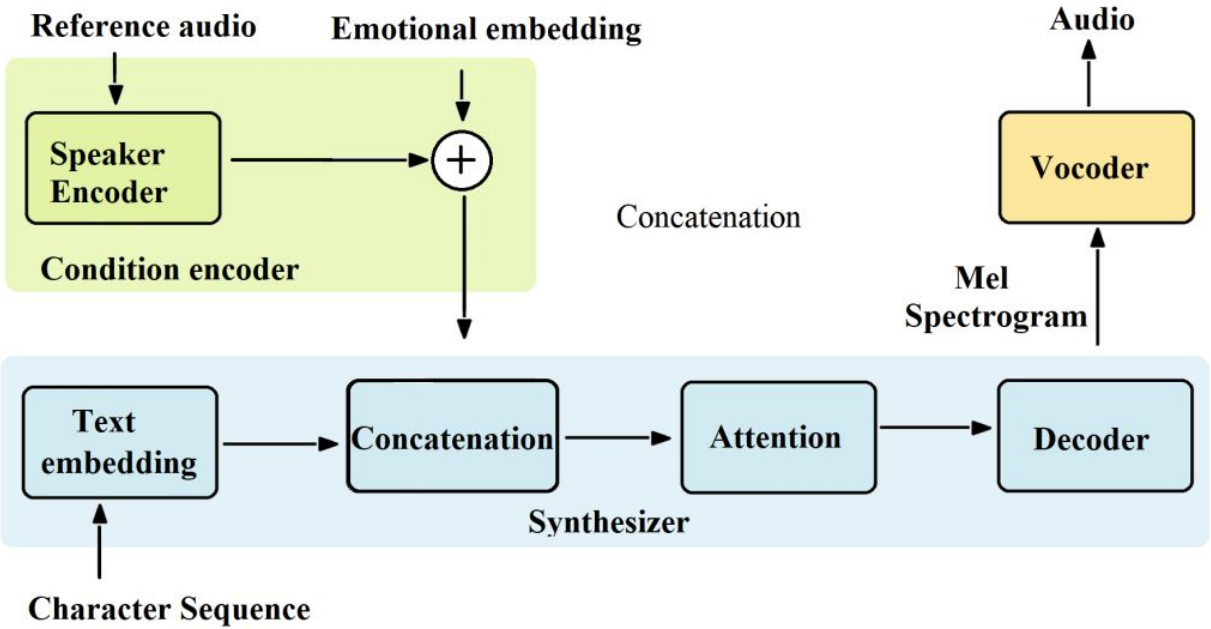
Reinforcement Learning!

- A reward-based mechanism to learn an action policy based on current state
- Simple reward: Happy human is good
- Expanding area of research

Bot-world

A Bot should be able to process and *reproduce* prosodic cues for interactional coordination

Text To Speech



Exercise to the Reader for Next Time:

- Record yourself saying the same phrase with different intonations, emotions and dialog acts (e.g. as a question, concerned, happy etc.). Listen careful to each one and see if you can list what gives each variation its characteristic sound.

- Now take away the words. Do your sentences hold the same meaning?

- Now get a voice bot (alexa, siri, an online app) to repeat the same sentences. What prosodic approach are they taking and how does it affect the interpretation?

Answers from last time:

- How would you parse datetimes in frame-based systems?
    - cascades of regular expressions to implement rule-based approaches
    - read more in Jurafsky chapter 22
- read Professor Richard Sutton's “The Bitter Lesson” (1k words)