

# Mobile Affective Inference Recognition: Recent Methods, Applications and Challenges

Ou Changkun

hi@changkun.us

Seminar Media Computer Science, LMU | Munich, Germany | 2018-02-26

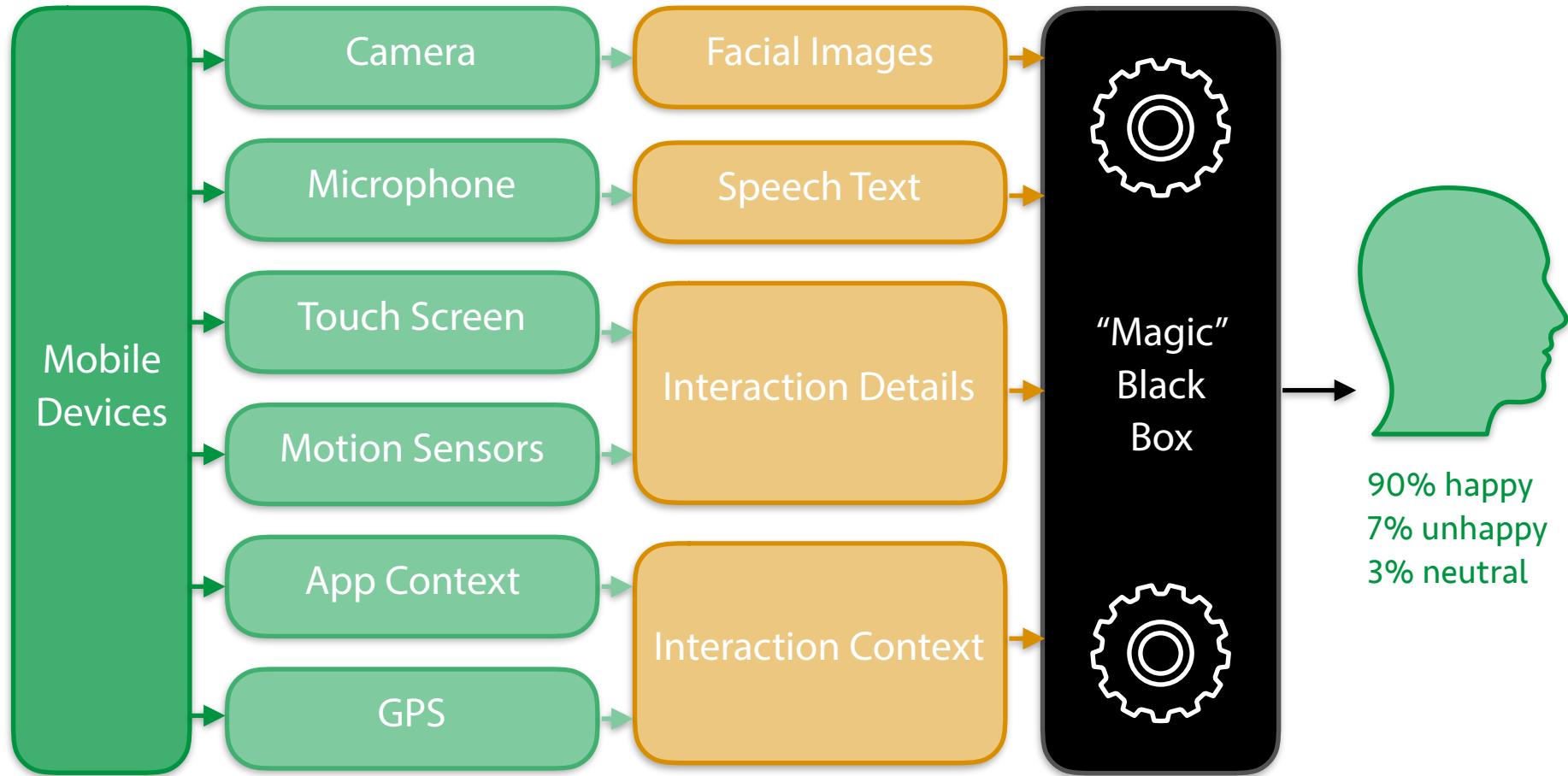
# Agenda

- Motivation
- Methods & Models
  - Vision & Audio aspects
  - Interaction Aspects
  - Sensors Fusion
- Emotion Inference Applications & Emotional System
- Challenges

# Motivation

- What is *Emotion Inference*?
  - Emotion Inference (aka *Emotion Recognition*) is the process of identifying human emotion, mostly from **facial expression**;
  - Recently, there is no inference step anymore.
- Emotion Recognition
  - A simple fact: Human's affect *expressed by various channels in a specific context*;
  - Emotion recognition relies on massive labeled channel data;

# A General Framework



# Facial-based Method & CNN Models

- **Facial expression** is the most important channel for emotion expression;
- Convolutional Neural Network is the recent **breakthrough** model that rules the entire computer vision area;
- Basic Idea: **Modeling human vision process.**

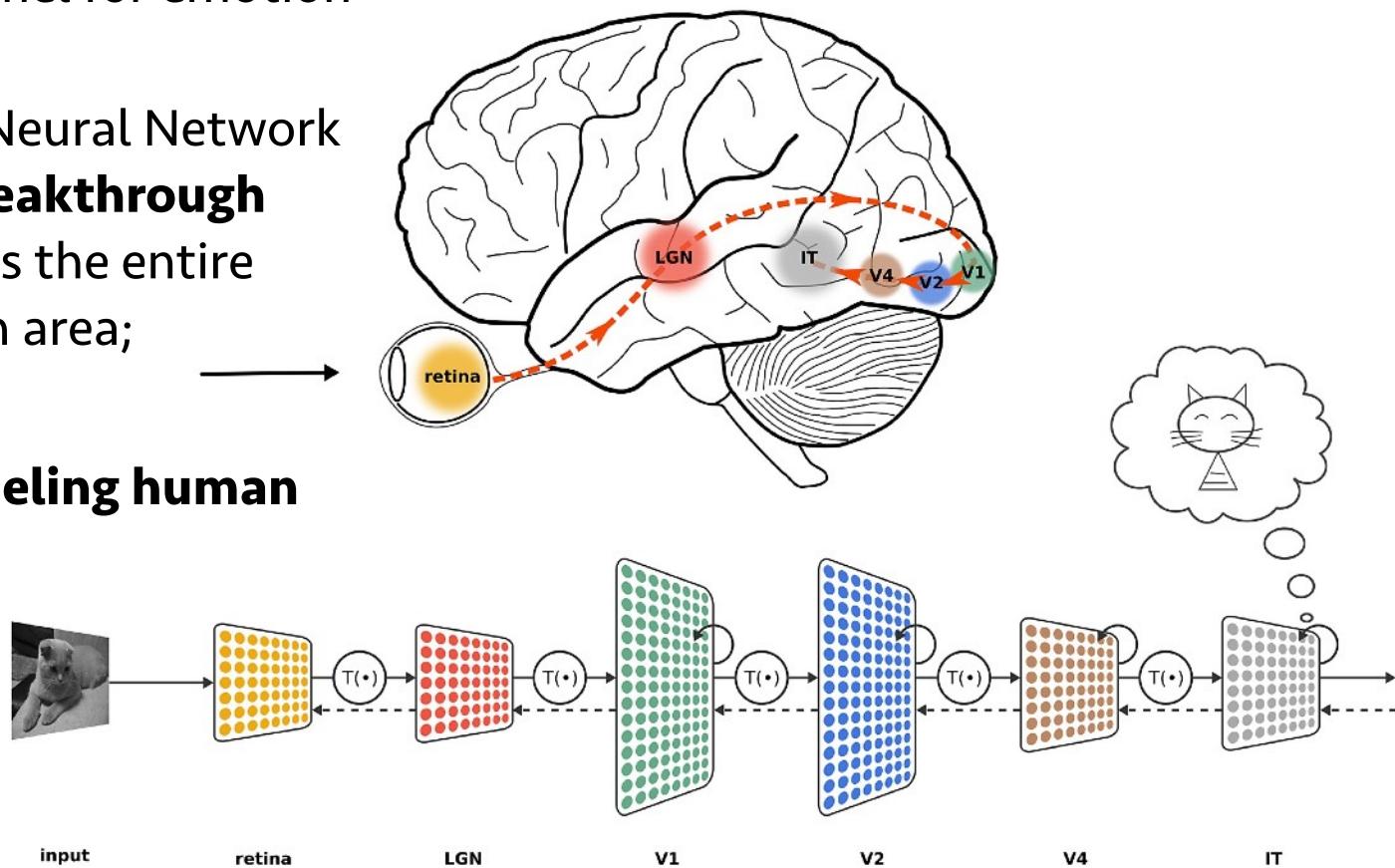


Image source: [https://figshare.com/articles/Ventral\\_visual\\_stream/106794](https://figshare.com/articles/Ventral_visual_stream/106794)

# “Manifold” Conjecture

- Random pixels **never** observes a face.
- Objects or “knowledge” *somewhat* has its structure in high dimensional space
- For example, faces are concentrated as a “manifold” in a  $28 \times 28 \times 1$  dimensional space

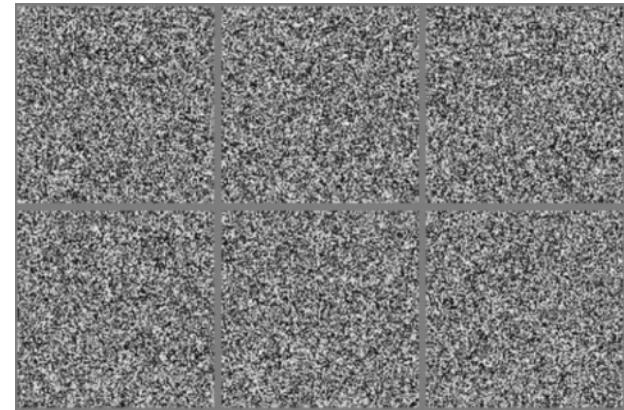


Image source: [Goodfellow et al. 2016, Deep Learning, MIT Press]

# Facial Emotion Inference Steps

- Step0: Prepare your dataset
- Step1: Targeting Face
- Step2: Emotion Recognition

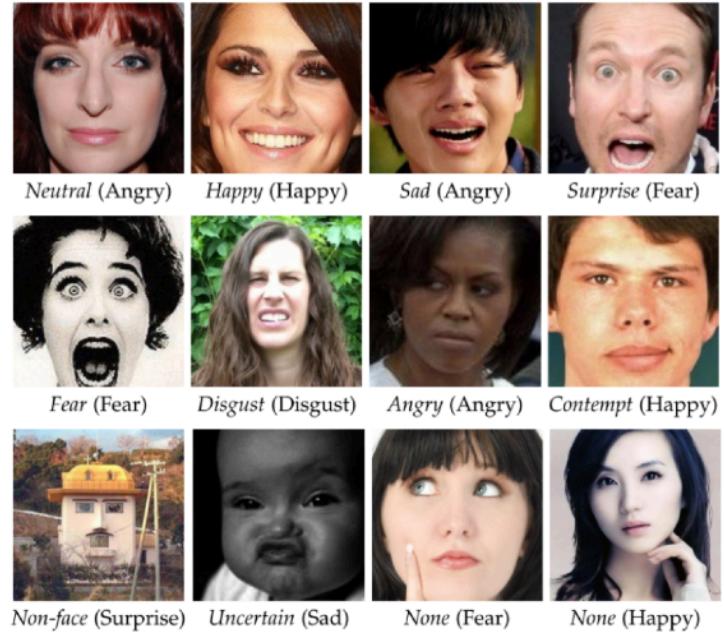
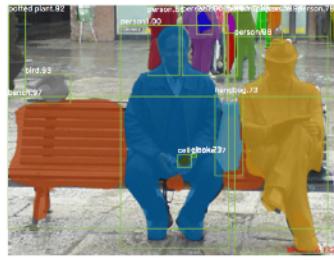
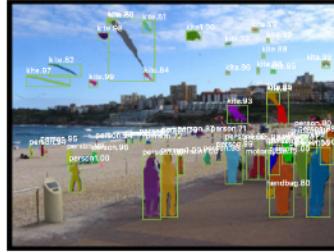
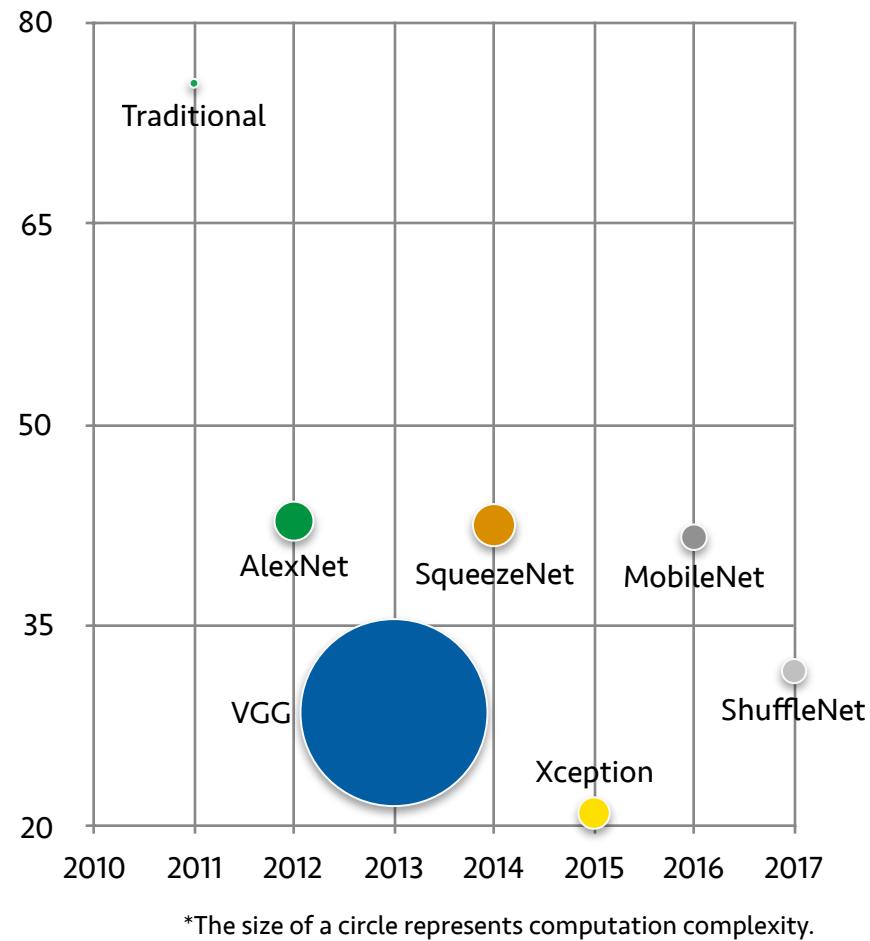


Image source: [He et al., Mask R-CNN, ICCV 2017]

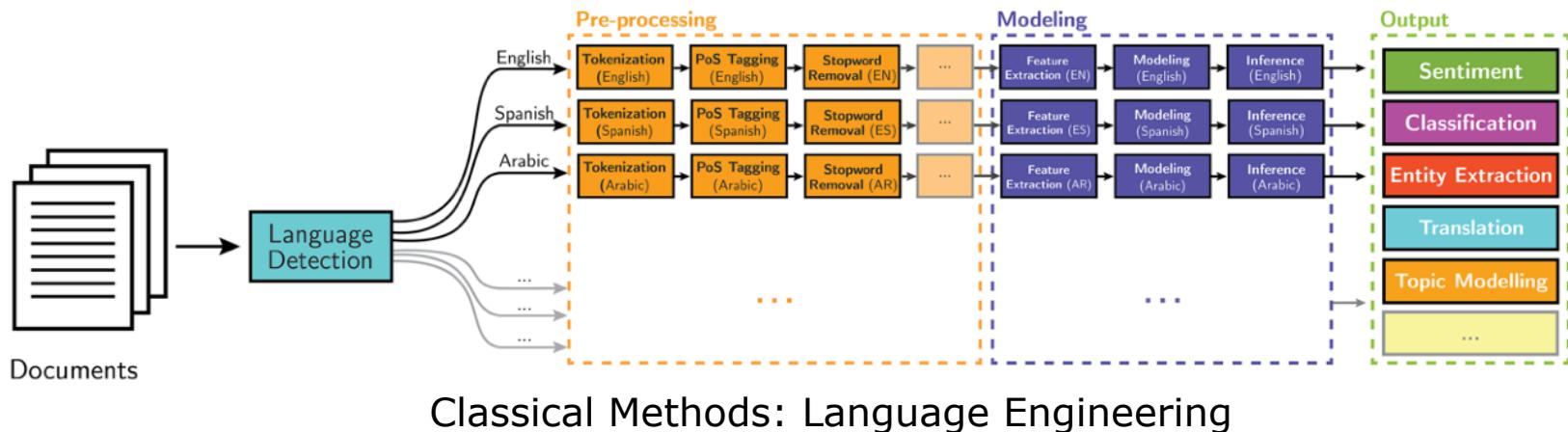
Image source: [<http://mohammadmahoor.com/affectnet/>]

# Recent Contributions: A Comparison

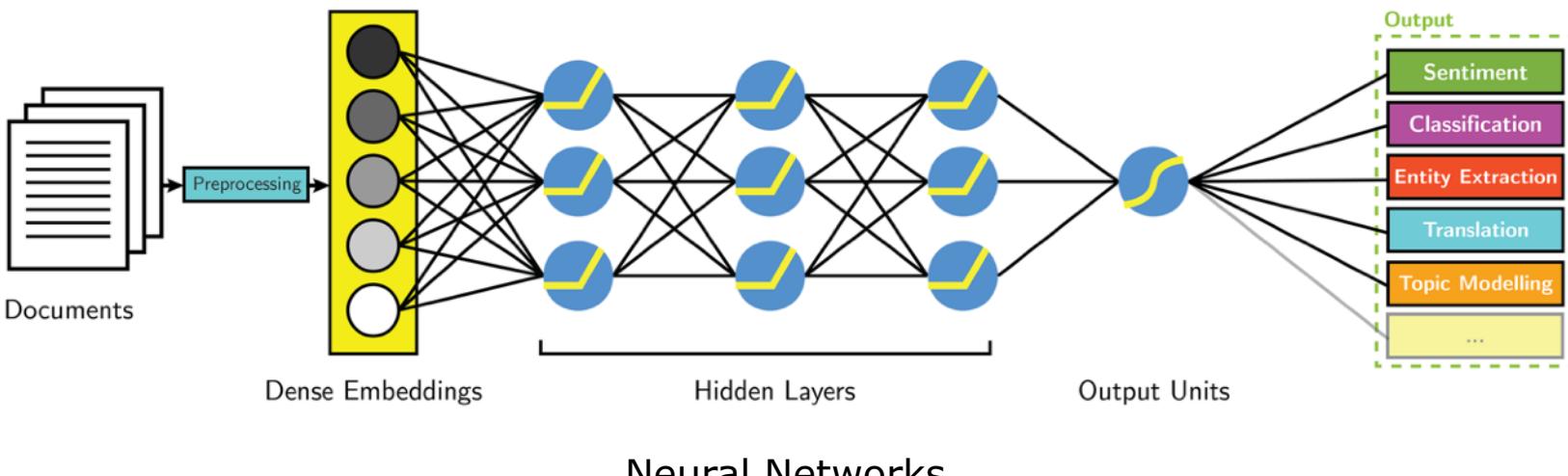
- Classification on Same dataset;
- Performance get improved monthly;
- Balance between recognition performance & model size.



# Voice-based Methods & NN Models



Classical Methods: Language Engineering



Neural Networks

Result source: <https://analyticks.wordpress.com/2016/11/07/leveraging-deep-learning-for-multilingual-sentiment-analysis-2/>

# Voice-based Methods

- Voice-based Emotion Inference can be decomposed to two steps:

Result:

3%

Actual label for sample text:  
negative

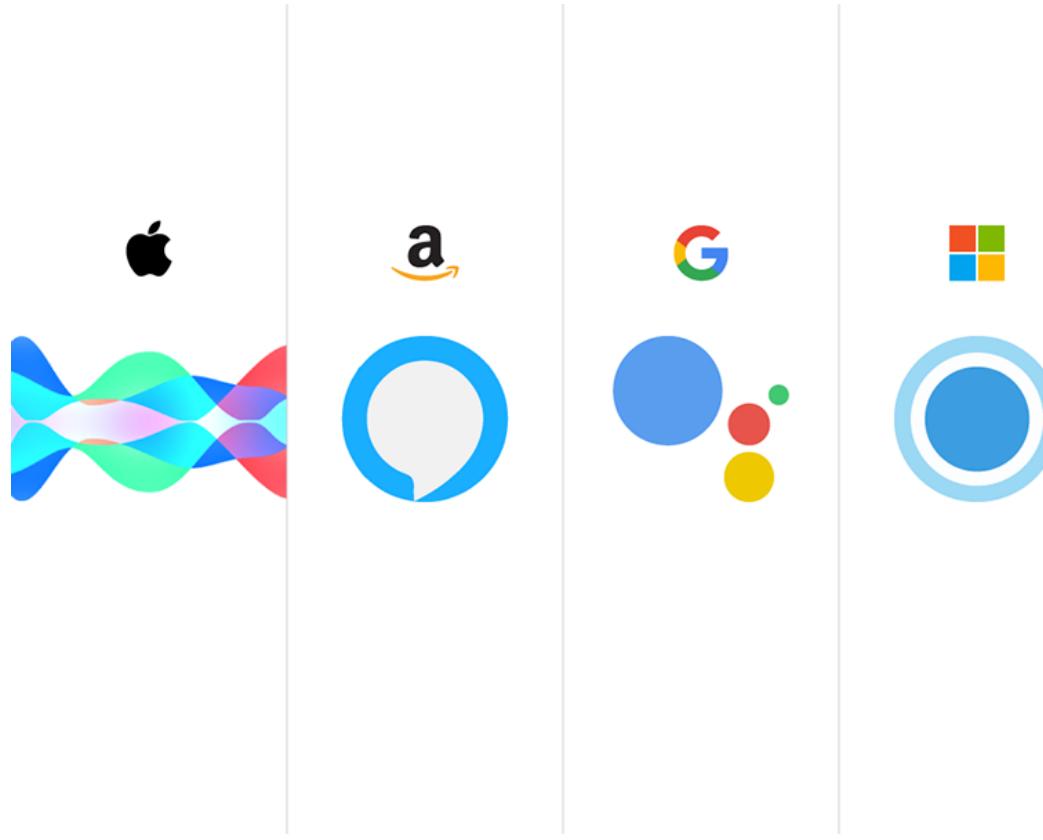
- Step 1: Speech to Text
- Step 2: Text Sentiment Analysis

```
a fugitive on the run a bit like the incredible hulk tv series without the shirt ripping jimmy crosses  
the mob in an entirely contrived way and goes on the run and in an entirely contrived manner finds  
himself working at a catholic reform school have you noticed an oft used description in the last  
sentence entirely contrived is the answer let me repeat for the hard of thinking that this is an  
entirely contrived film where everything relies on coincidence another problem i had was the reform  
school run by the church it's far too compassionate and kind i'm led to believe these type of <OOV>  
make alcatraz look like a country club i'm saying this is a fact but when the head priest looks like  
the spitting image of donald <OOV> you do feel there's a large amount of sugar <OOV> going on br br to  
be honest despite the ridiculous plot twists etc wanted <OOV> t really a bad thriller though it's a  
terribly good one either i never really had the urge to switch it off no matter how contrived it became  
which is an under hand compliment to the movie
```

Result source: Keras examples

# Commercial Success

- Voice Assistants have became consumer products;
- Massive human-level performance language APIs provided by platforms



Icon source: lazear@dribbble

# Interaction-based Methods

- Touch interaction and device motions (gyroscope, accelerometer, etc.) are the commonly used;
- Traditional feature engineering based and promising results are presented;
- All researches only considers three emotions and only in a specific context.

Features (>8) of Touch Interaction
Deviation in number of strikes
Deviation in number of taps
Mode of strike length
Average of strike length
....



Negative



Neutral



Positive

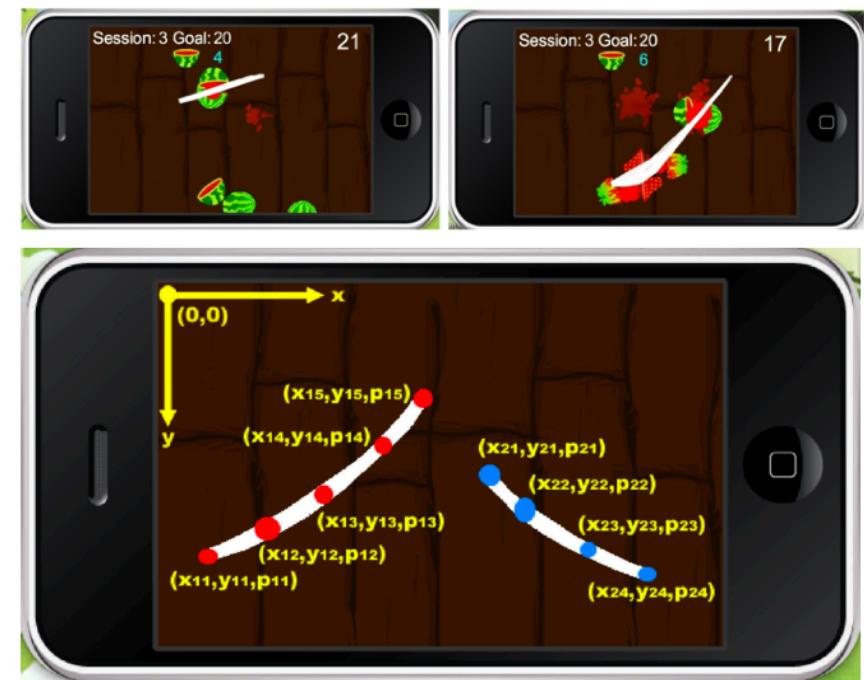


Image source: [Gao et al., What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?, TOCHI 2012]

# Multimodal Method

- 99% literatures based on sensors fusion method consider facial+speech data;
- All channels fusion is not discovered by researches.

# Emotional System

- Two typical application in HCI:
  - Case 1: Spoken Dialogue System
  - Case 2: Adaptive GUI
  - Both are dynamically changing user interfaces based on user's emotions



Video source: <https://youtu.be/lci1NCpe2Aw>

# Challenge: Impermeable Emotions

- Impermeable emotions can not be labeled, e.g. *I am jealous of...;*
- However, some research defend this argument and claims impermeable emotions are trivial and not interested.
- Impermeable emotions has culture difference;



Image source: Microsoft Emotion Recognition

# Challenge: Continues Understanding

- Emotions are not just state;
- Emotions influences each other and transform to others continuously;



Image source: <https://stanchew.wordpress.com/2012/04/23/a-map-of-human-emotions/>

# Challenge: Universal Model

- Facial, Speech models are universal for any case, but interaction detail & context are specific.
  - Is it possible by using one model to learn them all?
- Human emotions identification is not just a classification problem.

# Challenge: Lack of Computations

- Moore's Law is approaching to the end;
- Model compression & delivery;

# Summary

- Mobile emotion recognition becomes dominant than emotion inference;
- Facial & voice channel is the most important channel over all channels, and Neural Networks are recently advances for emotion recognition (>100 different emotions);
- Typical applications of mobile emotion recognition consider Emotion-based UI;
- Emotion-based HCI researches are rarely (Design principle, User Testing, etc.);
- Emotion Inference is a challenging problem & may **not** bring success.