WT7 A Survey on Psychology - Connecting Perception, Language, and Memory studies with Computer Vision

カストナー・マークアウレル（名大）　土屋志高（早大）　大角耕介（中部大）
秋山諒（奈良先端大）　河合里咲（東大）　池田宗也　（同大）

Psychology

- **Studying the human mind and its functions**
  - Behavior, thoughts, and feelings
  - Other than sociology, focus on individuals
- **Many cross-disciplinary applications**
  - Education, Law, Medicine, …
  - Most disciplines *interact with* or *develop something for* humans
- **Methodology**
  - Surveys / user studies
  - Behavioral experiments

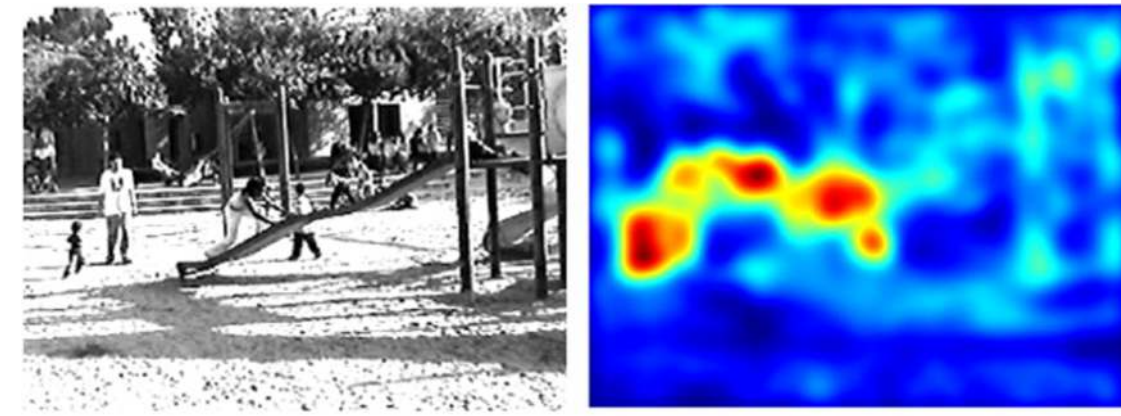We look at three main themes: Studies related to **Language**, **Perception** and/or **Memory**

Opportunities for CV

Semantic Gap and Human Perception

- **Understanding data**
  - Semantics based on human perception
  - Multi-modal approaches
  - Improving perception- / language-models
  - Gain better understanding of AI models (XAI)
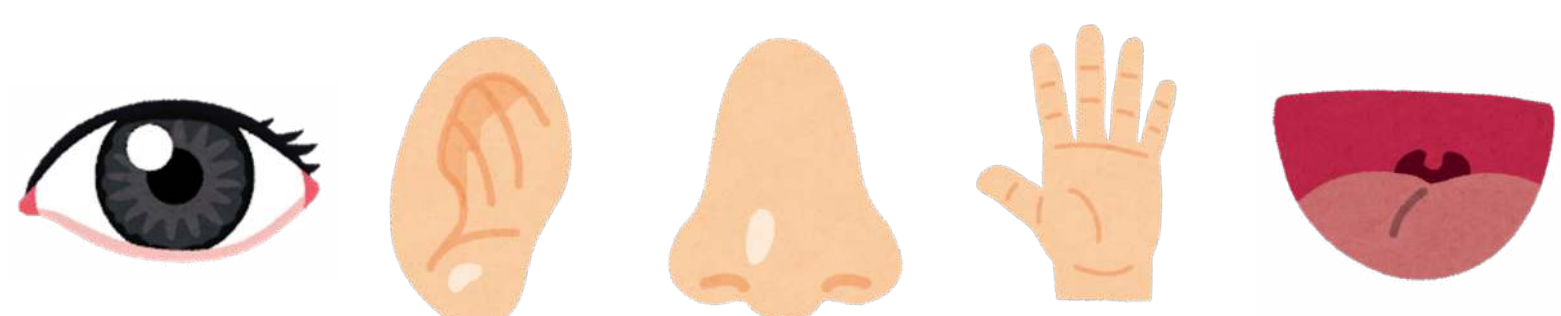- **Finding new applications**
  - Virtual / augmented reality
  - Saliency-based approaches

Language

- **Connecting Human and Language**
  1. Psycholinguistics: Individuals
  2. Neurolinguistics: Brain processes
  3. Sociolinguistics: Groups and cultures
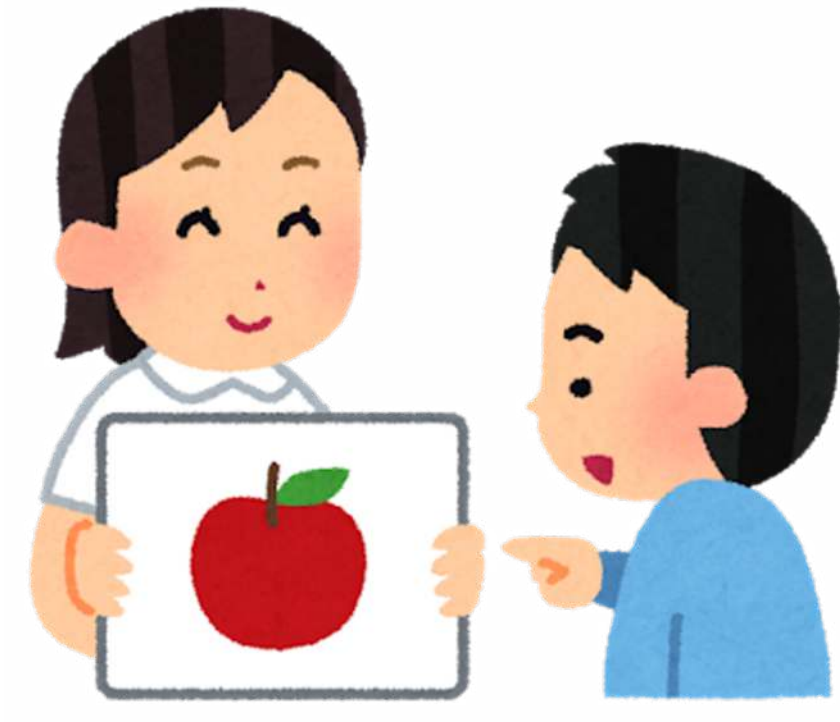
Psycholinguistics looks at **Language Production**, **Language Comprehension,** and **Language Acquisition.**

- **Language Comprehension: How does the human parse words and grammar?**
  - Large implications for language understanding (NLP)
  - E.g. analyzing imageability / concreteness of terms[1]
    - Compare perception of terms
  - E.g. looking at eye movement while reading text in different difficulties[2]

There was an old lady from Bath;
Who waved to her son down the path;
He opened the gates,
And bumped into his mates,
Who were Gerry, and Simon, and/ Garth.

Copyright R. Filik

- **Language Production: Looking at word choice problems, Connection of Thoughts -> Language**
  - Interesting for multi-modal approaches when creating captions or descriptions

Perception

- **Types of Human Perception**
  - Perception of Time
  - Perception of Motion
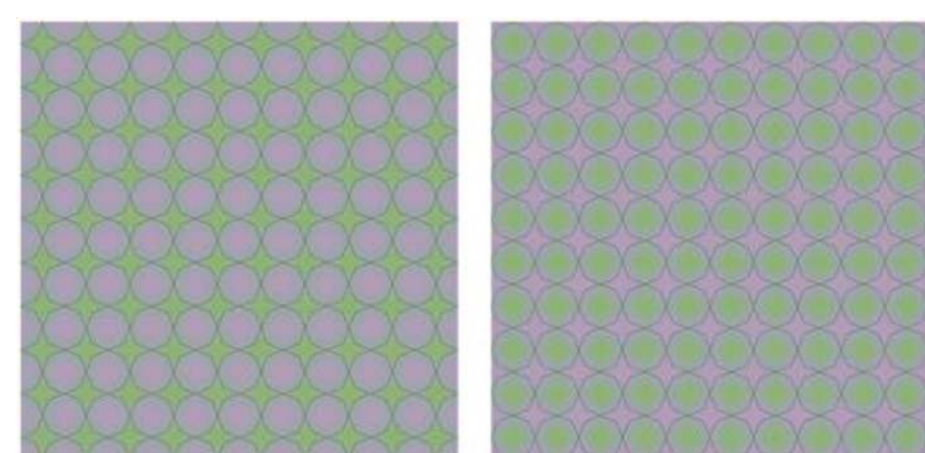 - Static sense
  - …

Five Senses

We focus on **Vision**. Understanding the human vision system is important for future work in CV.

- **Layers of Human Vision**

  Cognition　: Interpretation of unknown things
  Recognition : Understanding based on knowledge
  Perception　: Sensation of Light (color, brightness)

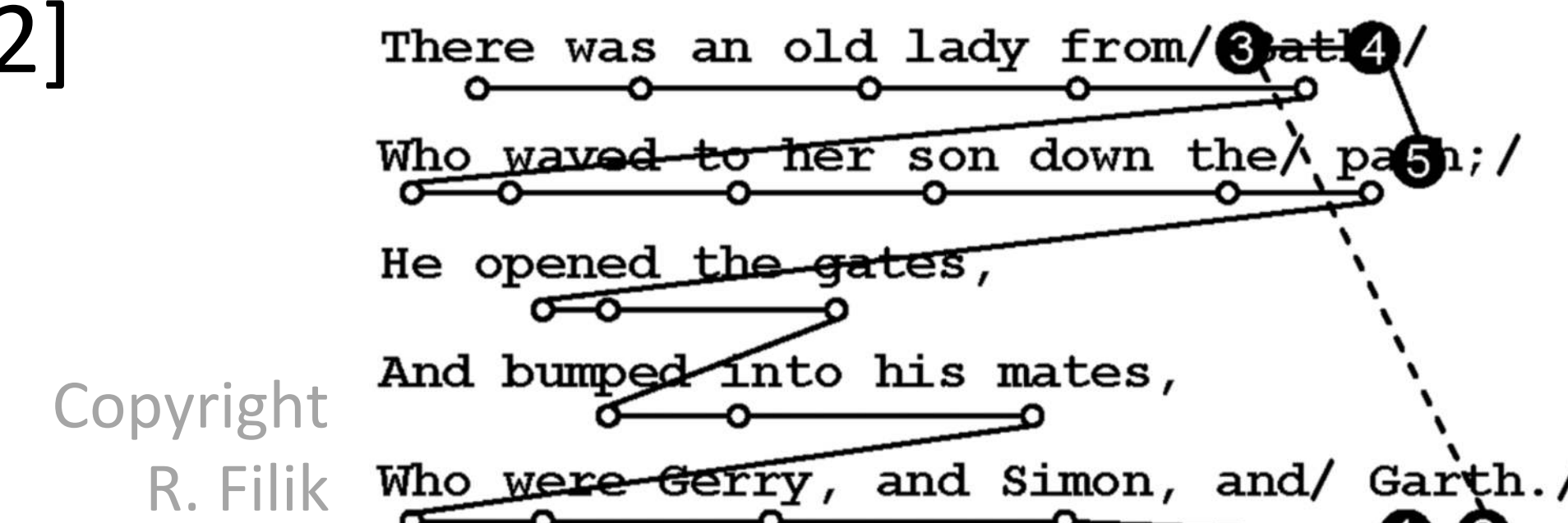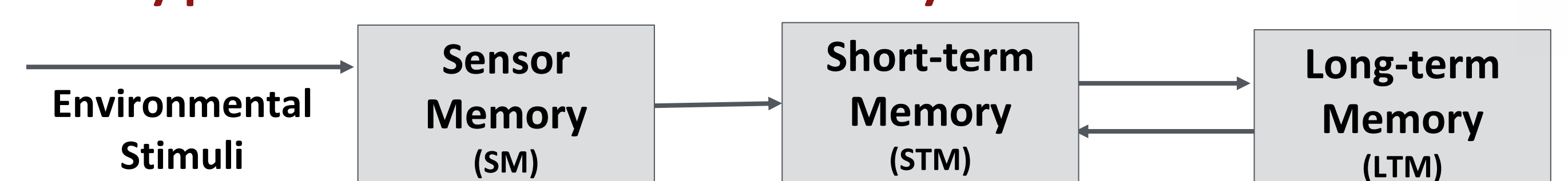- **Parts of Human Vision which are usually not considered in CV**

  - **Estimation of perceived color**
    - E.g. looking at spatial illumination changes[3]
    - E.g. perceived color depends on contour lines[4]
  - **Expressions or other senses affected by color**
    - E.g. face expressions are affected by both the color of a face and the background

Memory

Atkinson and Shiffrin, 1968

- **Types of Human Memory**

Environmental Stimuli → Sensor Memory (SM) → Short-term Memory (STM) ⇄ Long-term Memory (LTM)
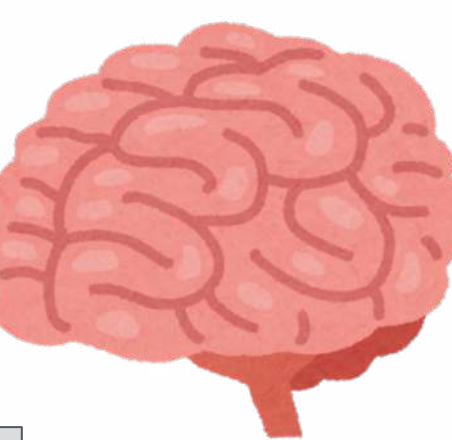
- **Sensor Memory** : Stores information from the senses. It only lasts for a duration of approximately half a second before it is forgotten.
- **Short-term Memory** : Stores information transferred from the SM. It lasts for a very brief time (less than a minute.)
- **Long-term Memory** : Stores information which has been processed in STM. It holds information for longer periods, up to permanently. The LTM contains both explicit and implicit memory.

We focus on **Human Memory Processes** and **Human Memory for Images**. By collaborating these findings with the CV community, it could yield interesting results.

- **Modelling the Human Memory**
  - Models on how the human memory works
    - E.g. Multi-Store Model, Working Memory, Levels-of-Processing
  - Some techniques in CV are an imitation of the Human Memory
    - E.g. Long Short-Term Memory
- **The Human Memory and Image Memorization**
  - Measuring the performance, capacity, and accuracy of the human memory in regards of recognizing images by doing user studies[5]
  - Analyzing image 'memorability' by using CV techniques
    - E.g. predicting image memorability using deep neural networks[6]

[1] Paivio et al., Concreteness, imageability, and meaningfulness values for 925 nouns. 1968
[2] Rayner et al., Eye Movements as Reflections of Comprehension Processes in Reading. 2006
[3] Sérgio et al., Spatial distributions of local illumination color in natural scenes. 2016
[4] Vergeer et al., Flexible color perception depending on the shape and positioning of achromatic contours. 2015
[5] Standing, Learning 10000 pictures. 1973
[6] Khosla et al., Understanding and predicting image memorability at a large scale. 2015