Conversational Agents with Emotion and Personality: Mind (Brain Internal States)

August 12th, 2019

Soo-Young Lee
Director, Institute for Artificial Intelligence
School of EE / Brain Science Research Center
Korea Advanced Institute of Science & Technology

sylee@kaist.ac.kr, http://ki.kaist.ac.kr

Contents

Background

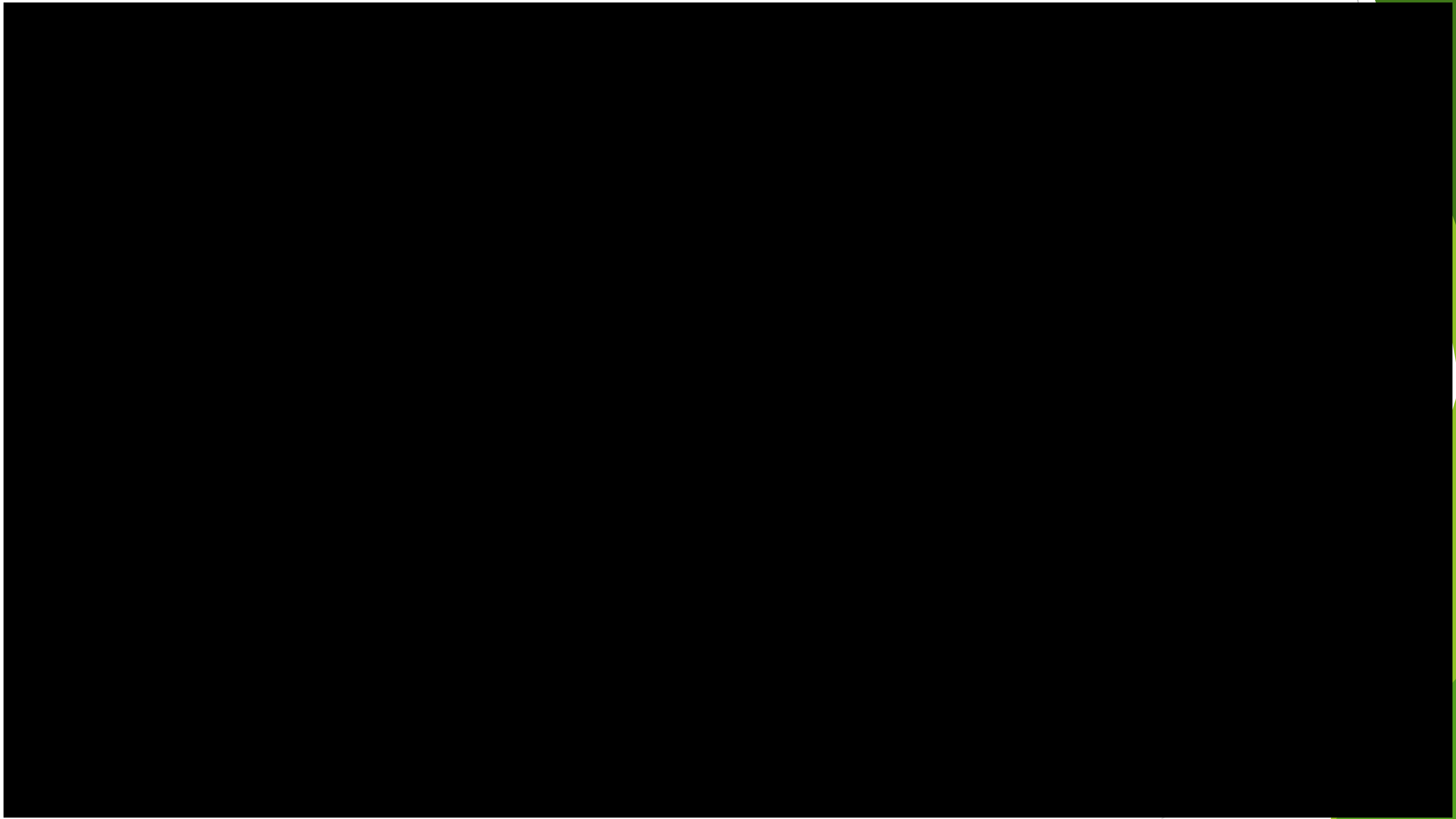
Smart Speaker and Beyond

- From Voice Control and Q&A Devices
- Via Personal Assistant
- To Digital Companion (Office Mate)

Personal Assistant: Artificial Secretary (Braintech'21: 1998-2008)

- Dual Goals
  - Understand brain information processing mechanism
  - Develop Personal Assist
    (or Artificial Secretary)

Emotional Conversational Agent (June 2016-April 2019)



Companions We Need at Office and Home

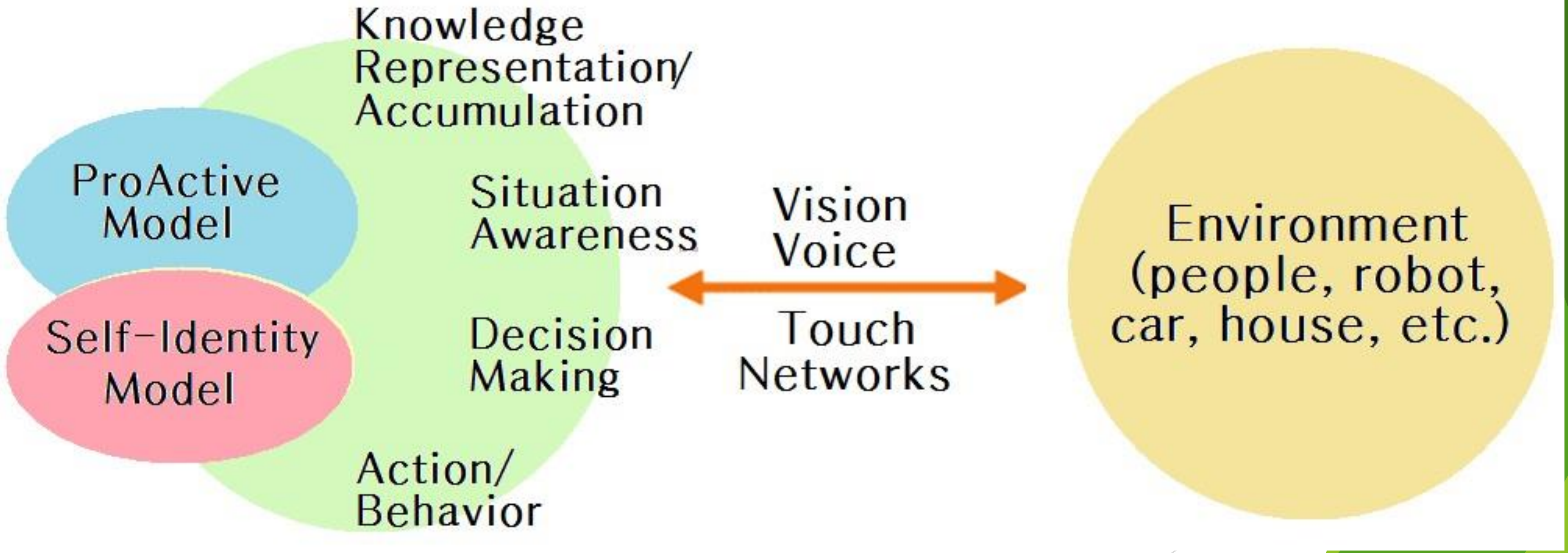
- We want intelligent companions who understand me and situations well and respond accordingly at any time at any place.
- Personal Companion or Office Mate
  - from pets to companions



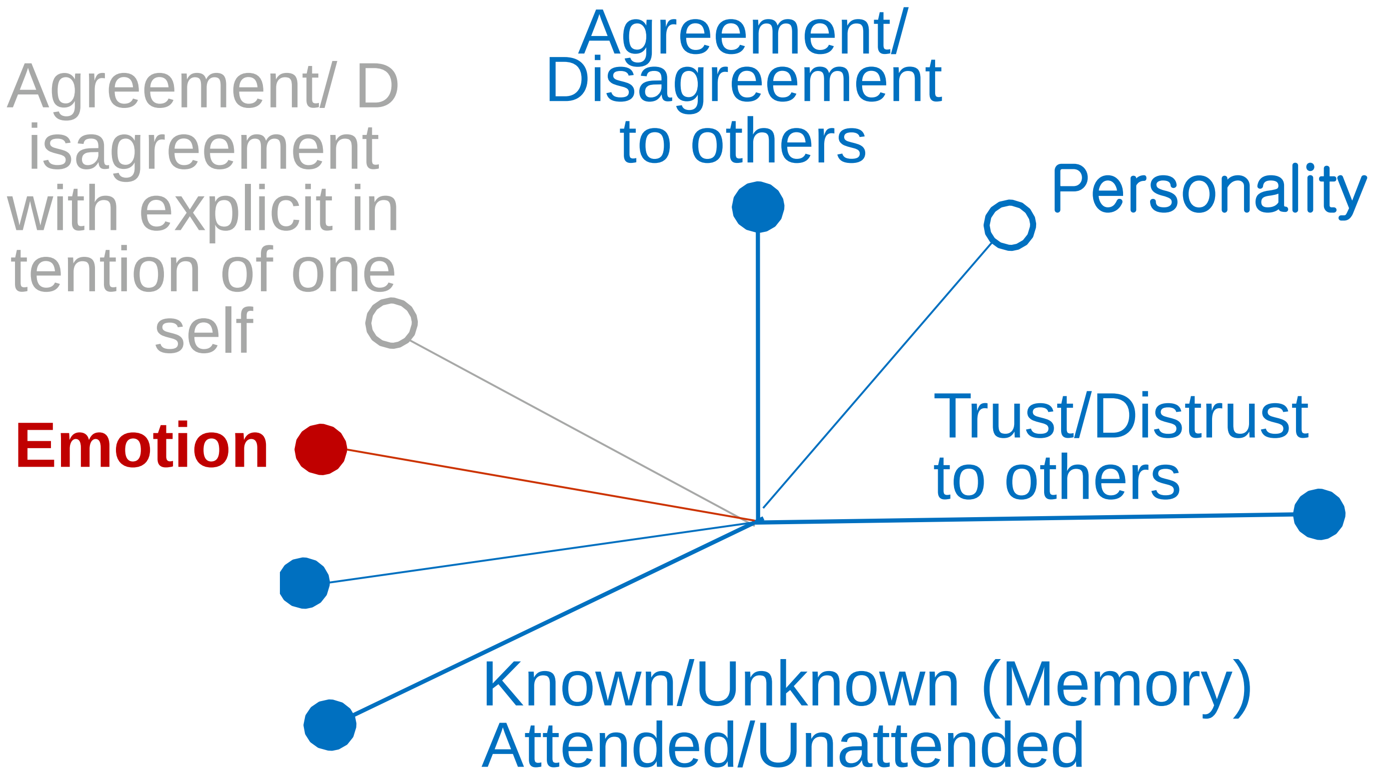


Beyond Personal Assistant: Digital Companion

- Everywhere (Home, Automobile, Office, etc.)
- Personality (not one-for-all)
- Interaction with context/emotion/intention/situation



Mind: Brain Internal Space

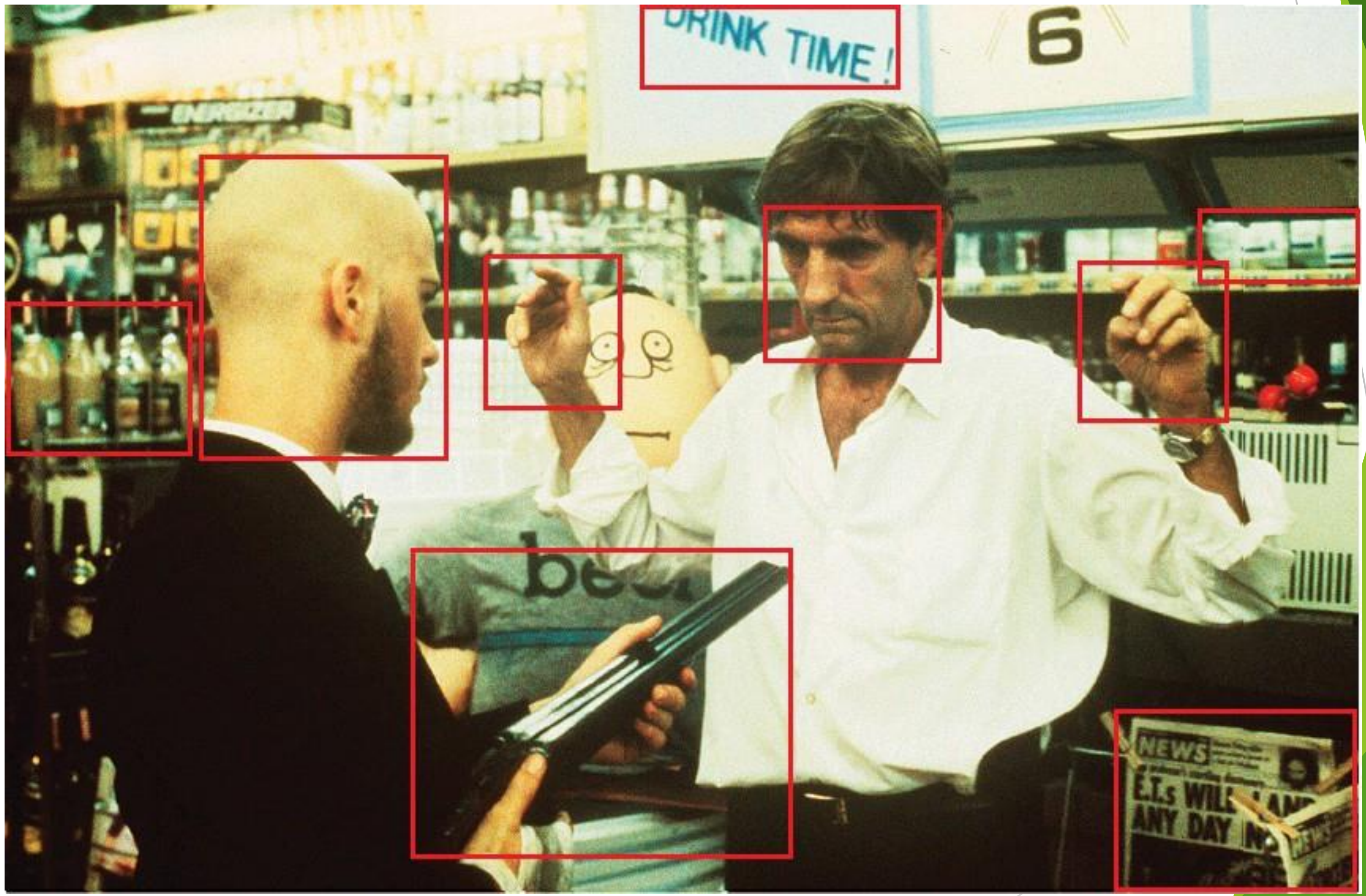


Time Dynamics Intention Trust **Emotion** Memory Ethics Personality

Fast ⟷ Slow

Situation Awareness

- Needs both explicit and implicit information



(IEEE Spectrum, June 2008)

Teach AI to understand and respond to human mind

Decision/Action and Mind/Environments

- Human decision making is different from person to person, and from time to time.
  - affected by internal states (mind) which may have **temporal dynamics** and unknown environments.

$$Action[n]=f(Audio[n],Video[n],Mind[n],Environment[n])$$

$$Mind[n+1]=Mind[n]+g_1(Mind[n],Audio[n],Video[n],Action[n])$$

$$Environments[n+1]=Environments[n]$$
$$+g_2(Environments[n],Audio[n],Video[n],Action[n])$$

- Develop Human-Agent Interaction based on internal state models. (Game Theory / Theory-of-Mind)

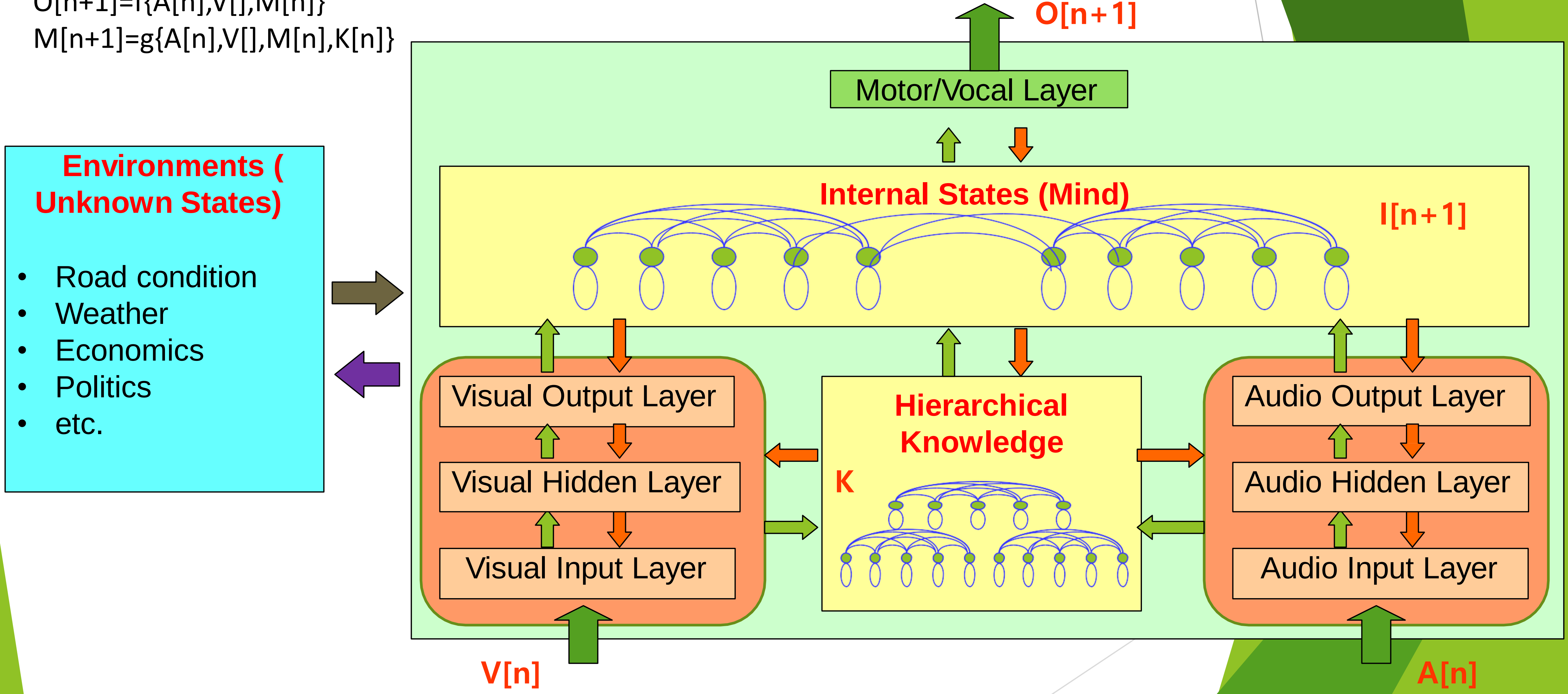
Environments: Unknown Space

- Road condition
- Weather
- Economy
- Politics
- etc.




Internal States : Mind

$O[n+1]=f\{A[n],V[],M[n]\}$
$M[n+1]=g\{A[n],V[],M[n],K[n]\}$



3 approaches to solve real-world problems

- If you or others KNOW how to solve the problem,

  Just solve the problem with best existing methods.
- If NOT,

  If there exists ENOUGH DATA,

  Use existing Deep Learning models.

  (You may need refine system parameters adaptively.)

  If SOME data is available,

  Develop new model(s), collect data, and improve the model for the problem. (You may need combine the **human approaches / domain knowledge** and neural network theory.

  If NO data is available,

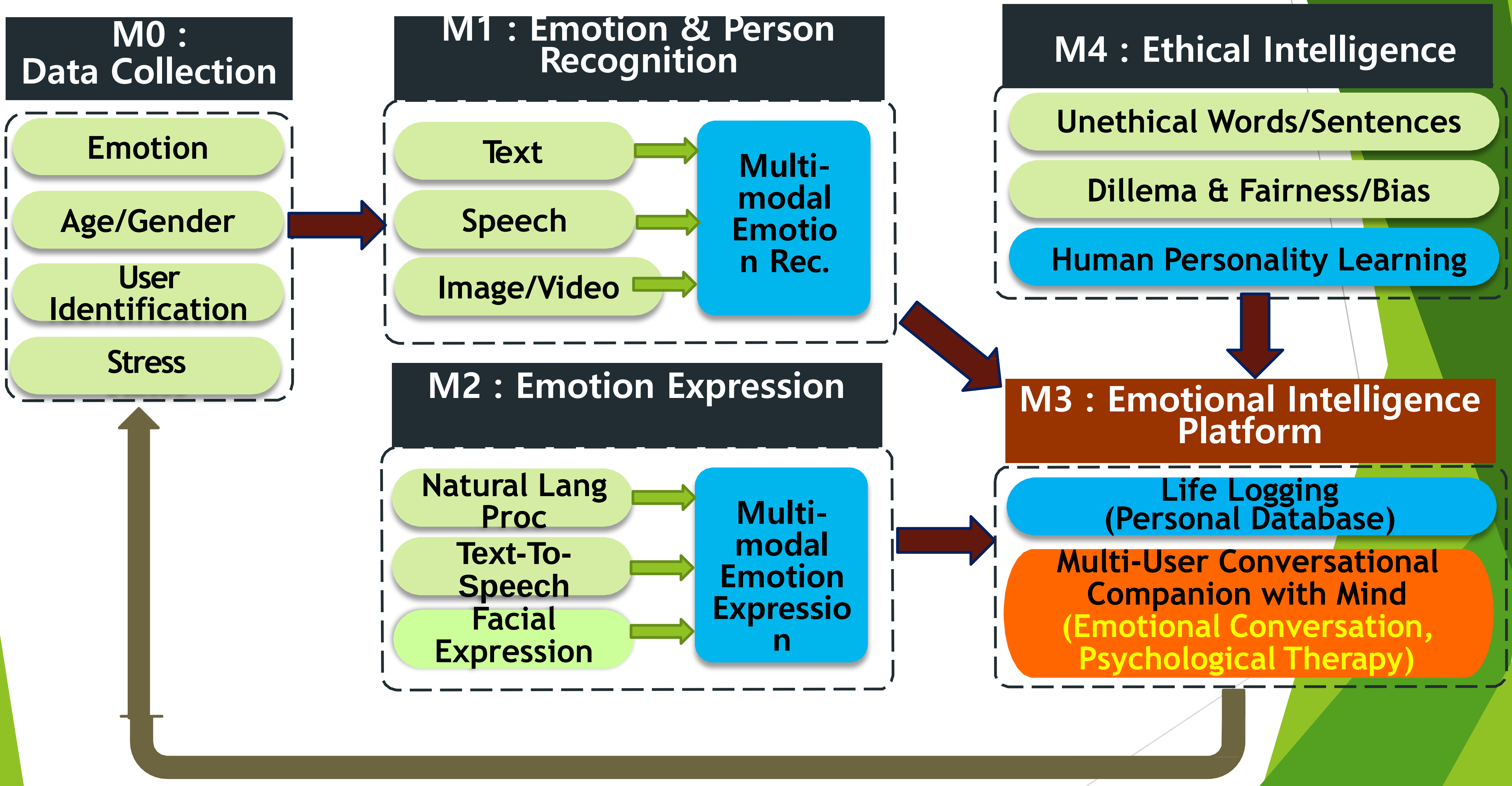
  Conduct cognitive science experiments to find the knowledge.

Emotional Conversational Agents

Companion with Emotional Intelligence

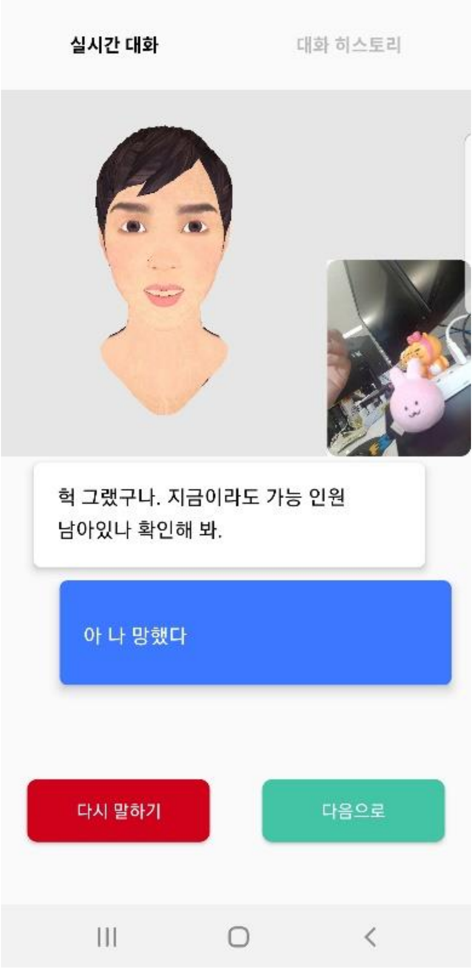
- AI Agents with whom people may fall in love and like to work at office.

Research Modules



ECA Testbed

Android APP



Data Collection

Emotion Recognition from Text

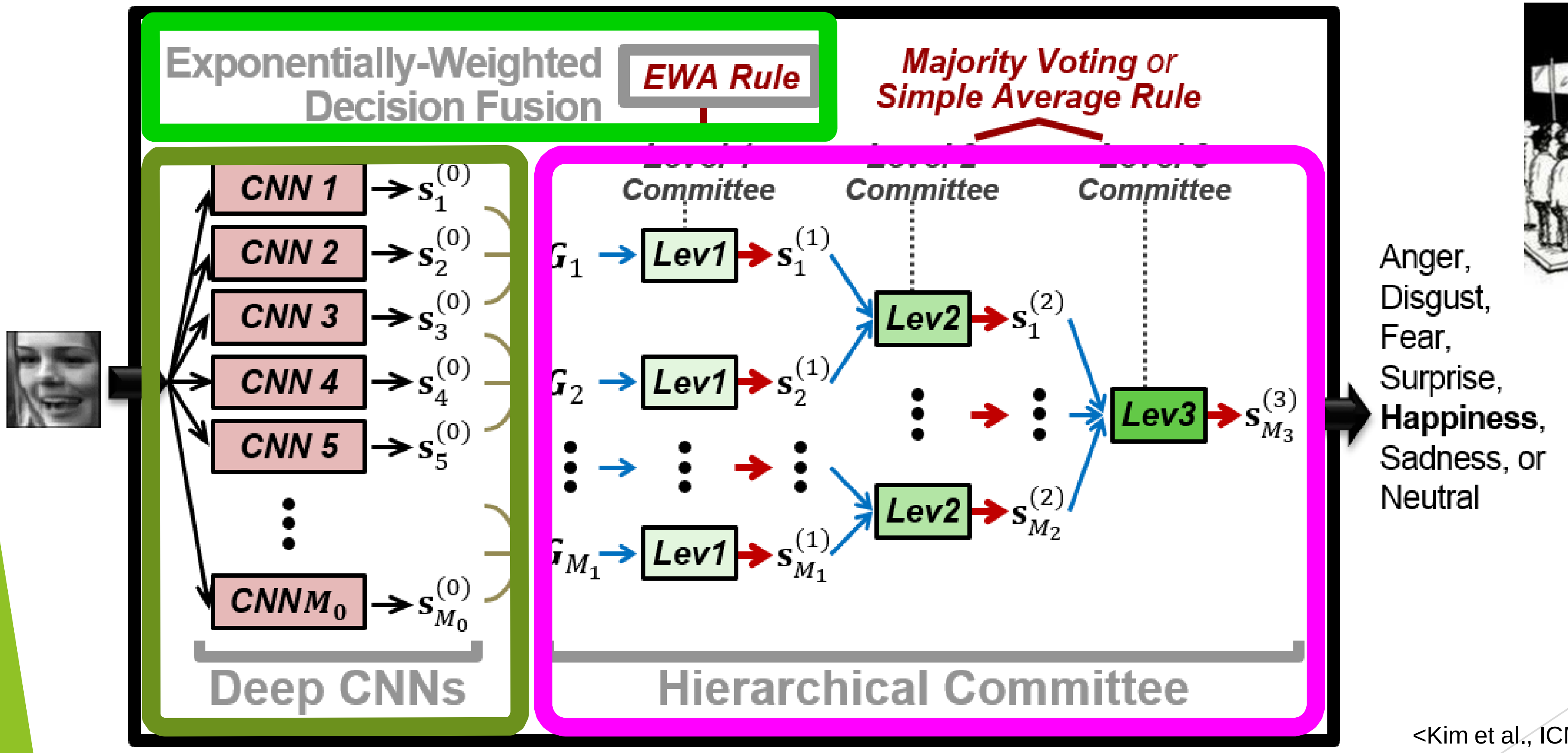
- Dual attention mechanism: local and global
- From essay to conversation
- Accuracy (6 classes + neutral): 78 – 88 % (with ensemble)

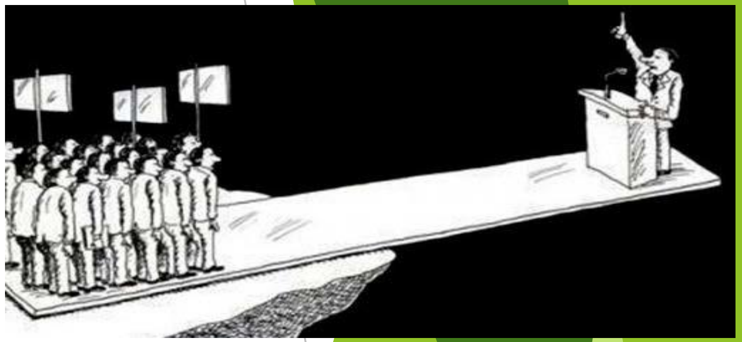
Recognition from Images

- Emotion
- Gender
- Age
- Stress
- Speaker

Facial Expression Recognition in the Wild
(1st Ranked, EmotiW2015)

- Advanced Committee with diverse CNNs and hierarchical structure





<Kim et al., ICMI'15>
<Kim et al., J. Multimodal User In., 2016>

Facial Expression Recognition in the Wild
(Image-based session @ EmotiW'15 challenge)

- 7-class FER of movie scenes, # (training, validation, test) images = (958, 436, 372) + external training data (~35,000)

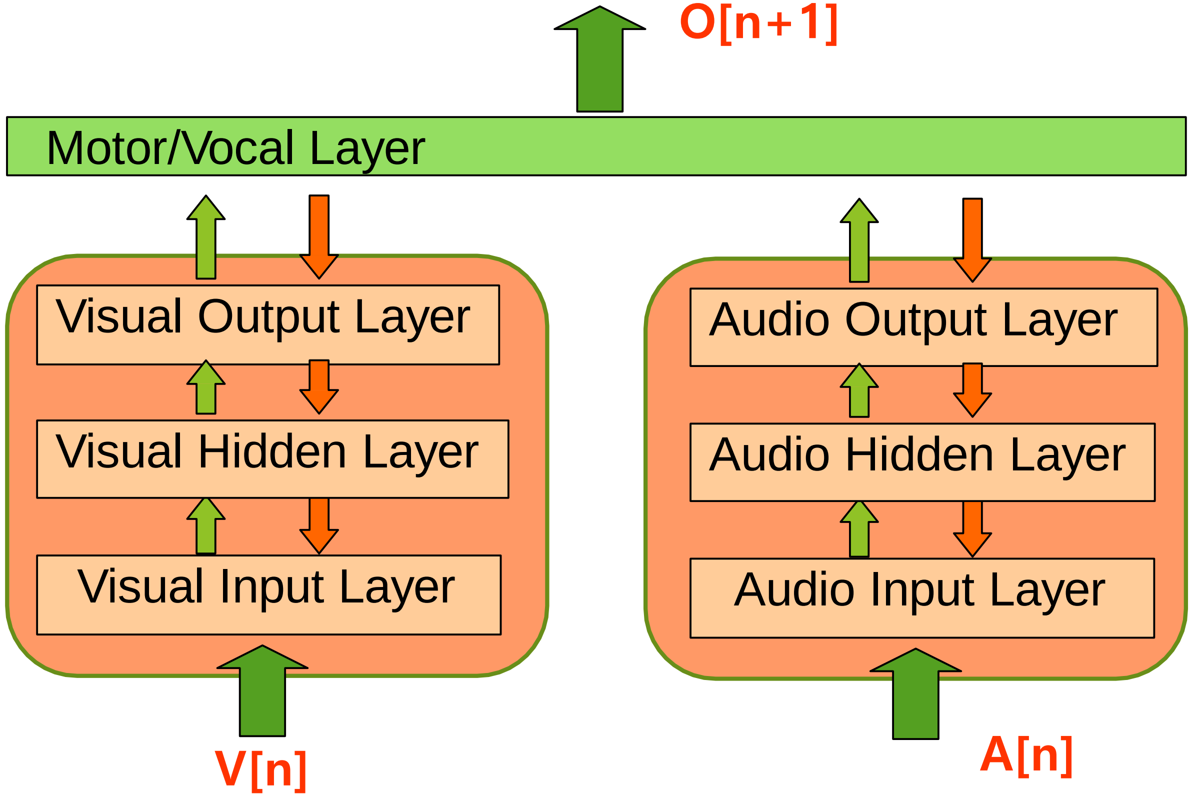
Accuracy (%)	
{LPQ-pHOG} + rbfSVM: **baseline**	39.1
The Best Single Deep CNN	57.3
Single-Level Committee w/ Simple Ave. Rule: **conventional**	58.3
Single-Level Committee w/ Exp Weight Rule	60.5
Hierarchical Committee w/ {Exp Weight, Simple Ave., Majority Vote}	**61.6**

Recognition from Speech

- Emotion
- Speaker
- Stress

- Disentangling different speech features
  - Phoneme
  - Emotion
  - Personality
  - Etc.

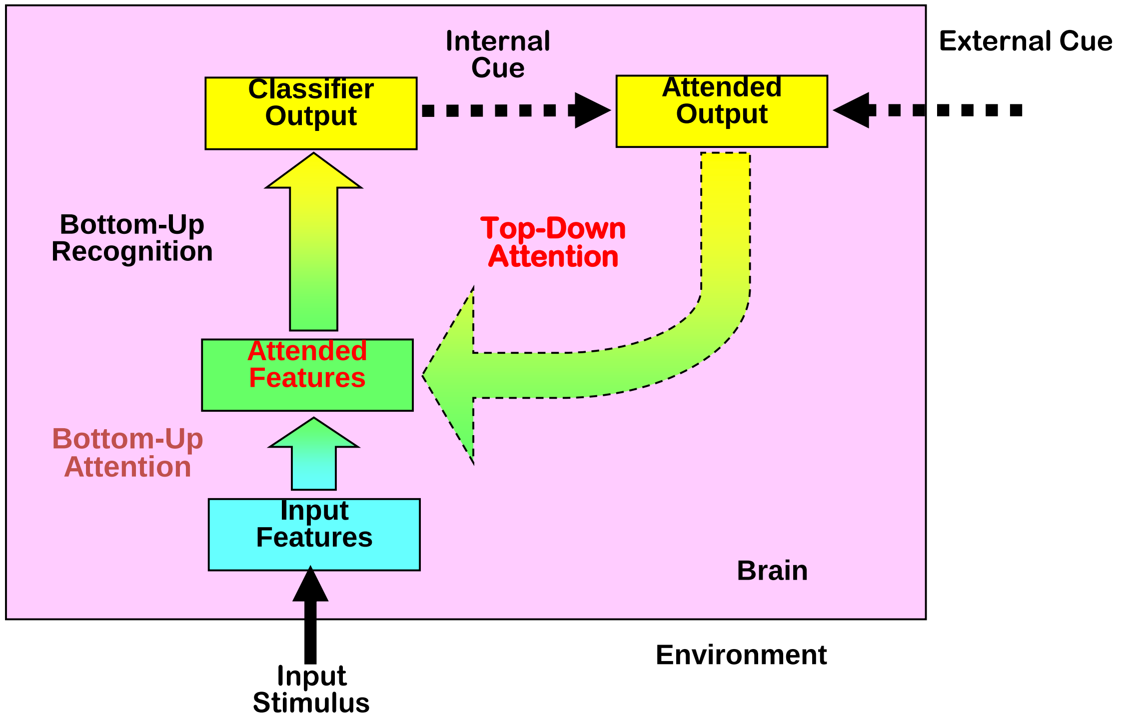
Multimodal Integration with Top-Down Attention

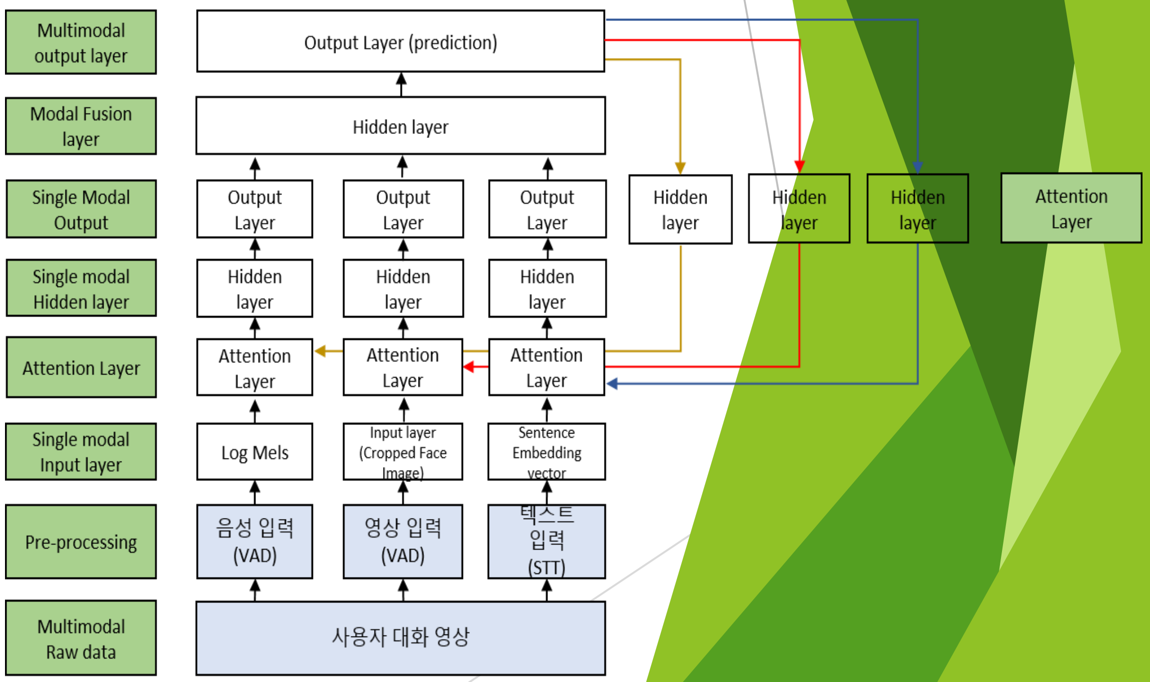


Multimodal Integrated Recognition

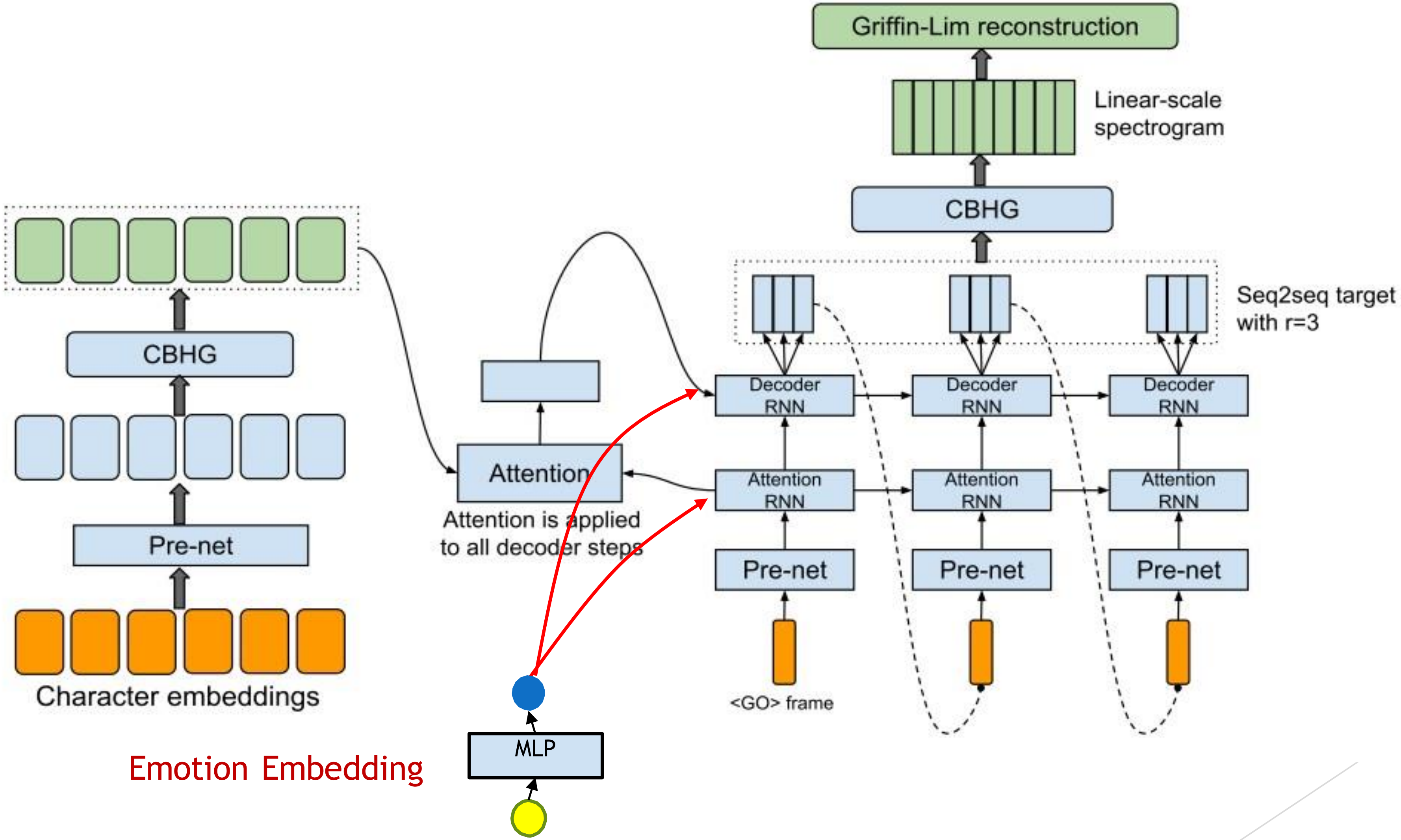
➢ Early Integration, Late Integration, and Attention

- Bottom-Up Attention (Self Attention)
- Top-Down Attention





Speech Synthesis: Emotional TTS (Y. Lee, et al., NIPS Workshop 2017)



Emotional TTS (Y. Lee, et al., NIPS Workshop 2017)

http://143.248.97.172:9000/

- Continuous emotional strength

Suprise

Happy

Disgust

?

Sad

Angry

Fear

감정 세기

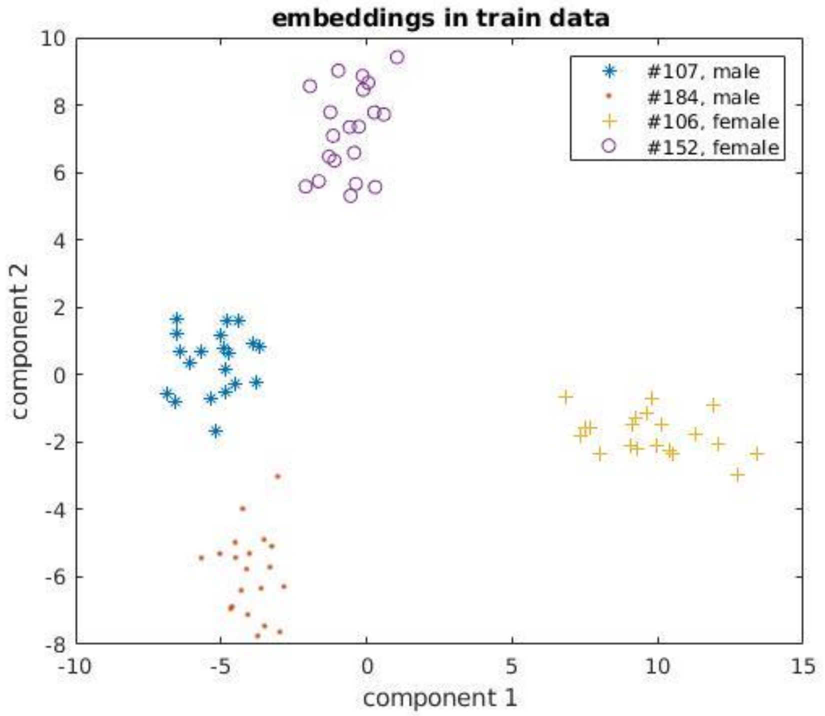
More Controls on Emotional Speech

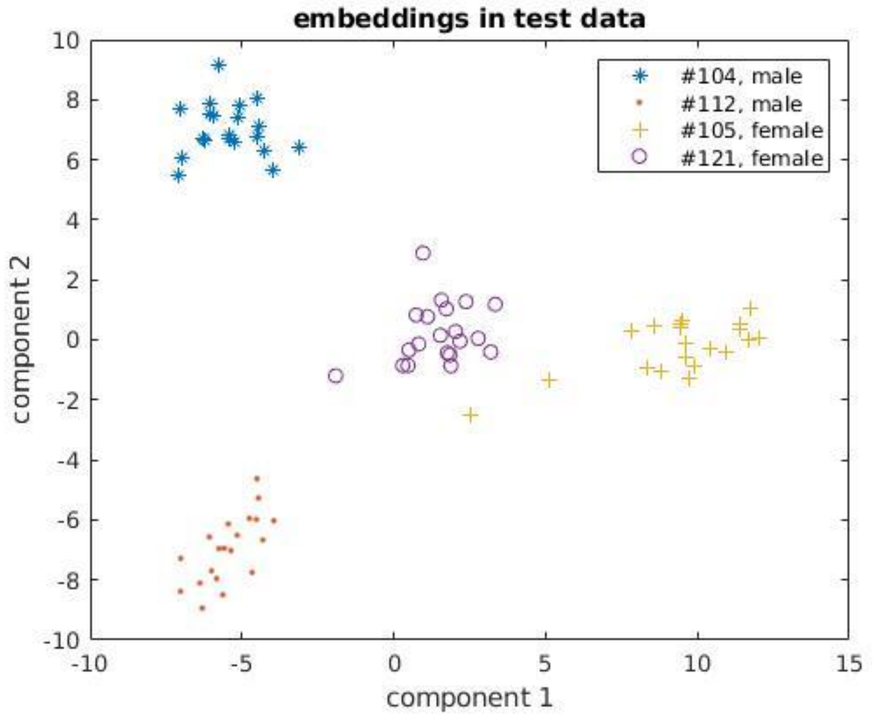
Emotional Strength

Mixed Emotion

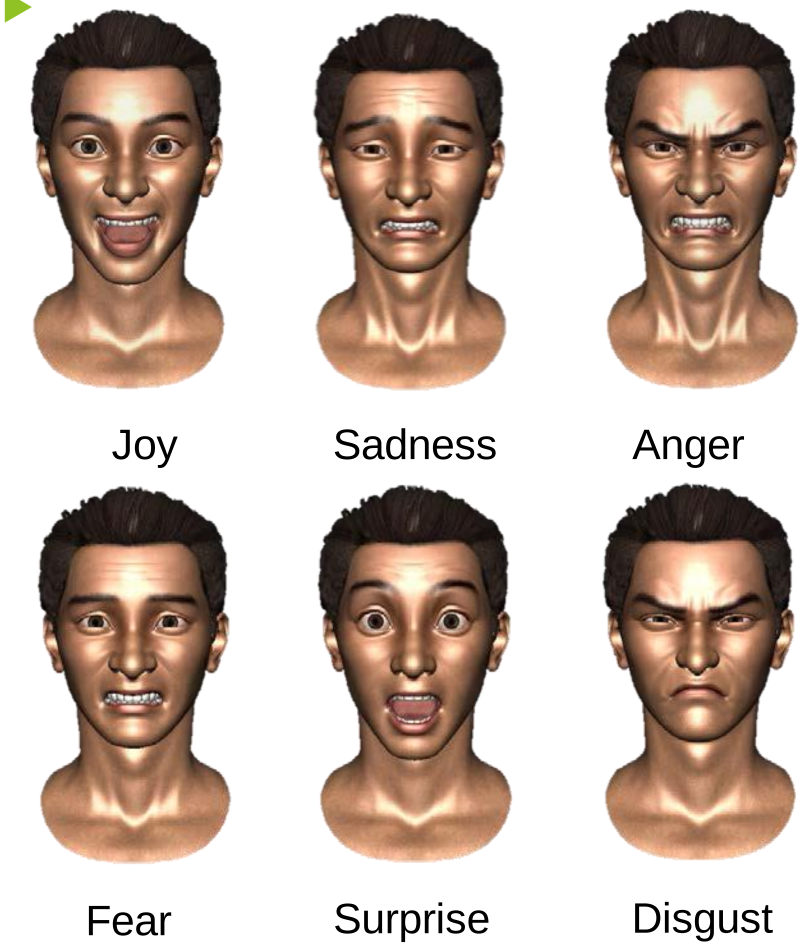
Personalized Voices

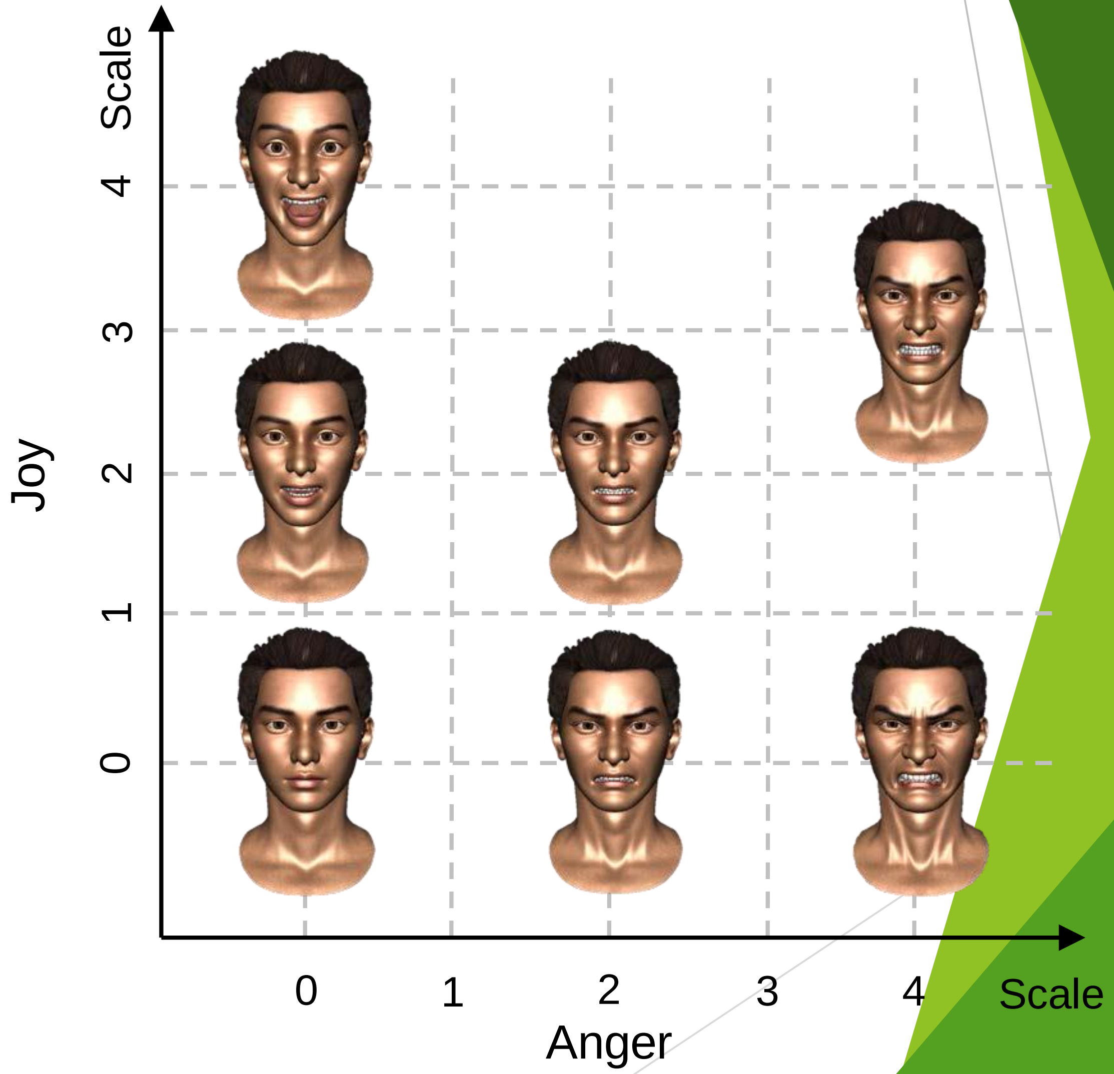
➢ Embedding learning from multiple speakers



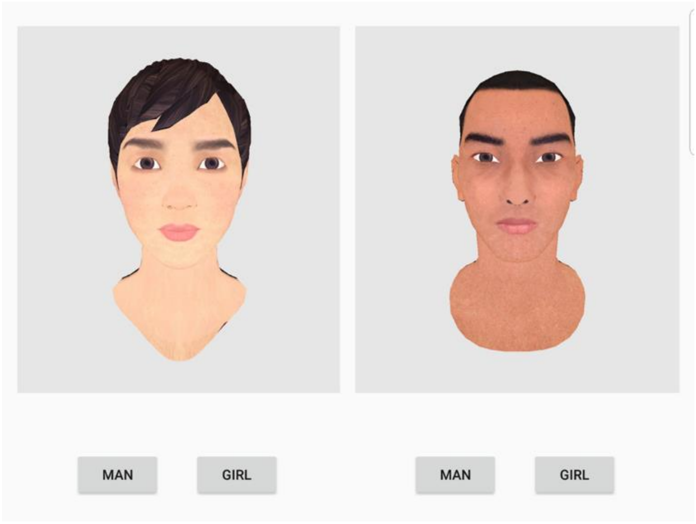


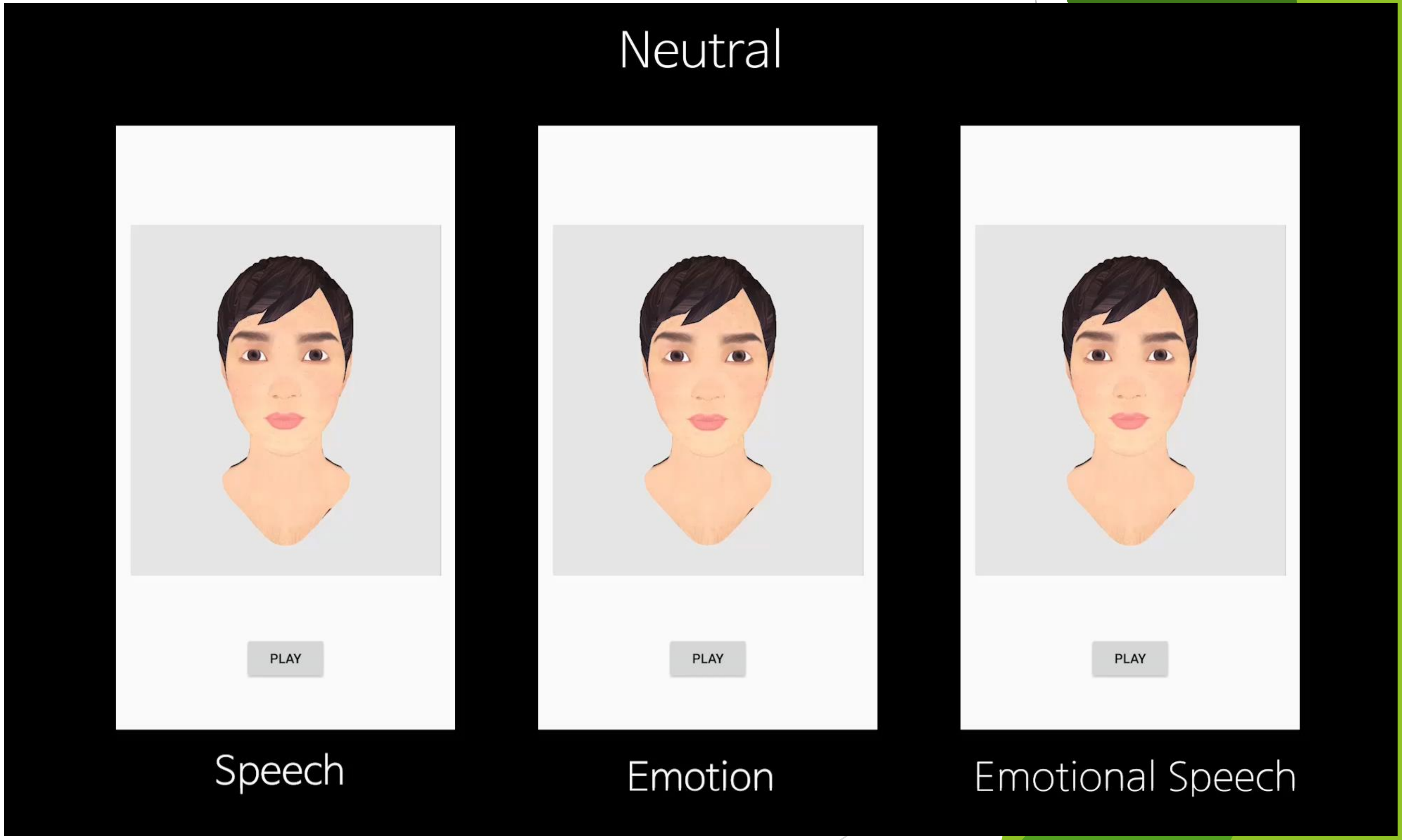
Emotional Facial Expression (Prof. JY Noh)





Facial Expression Synthesis

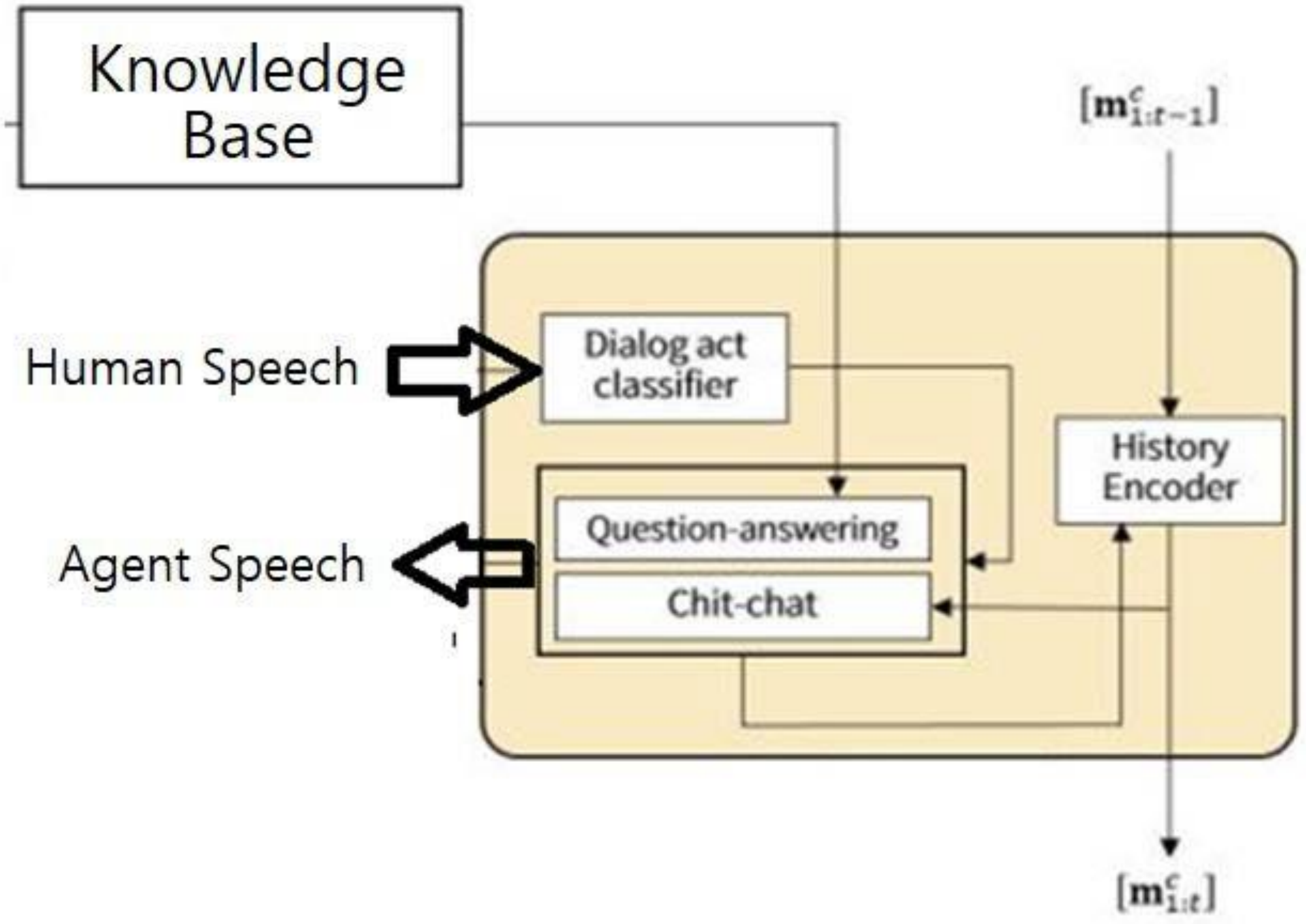




Dialogue Generator

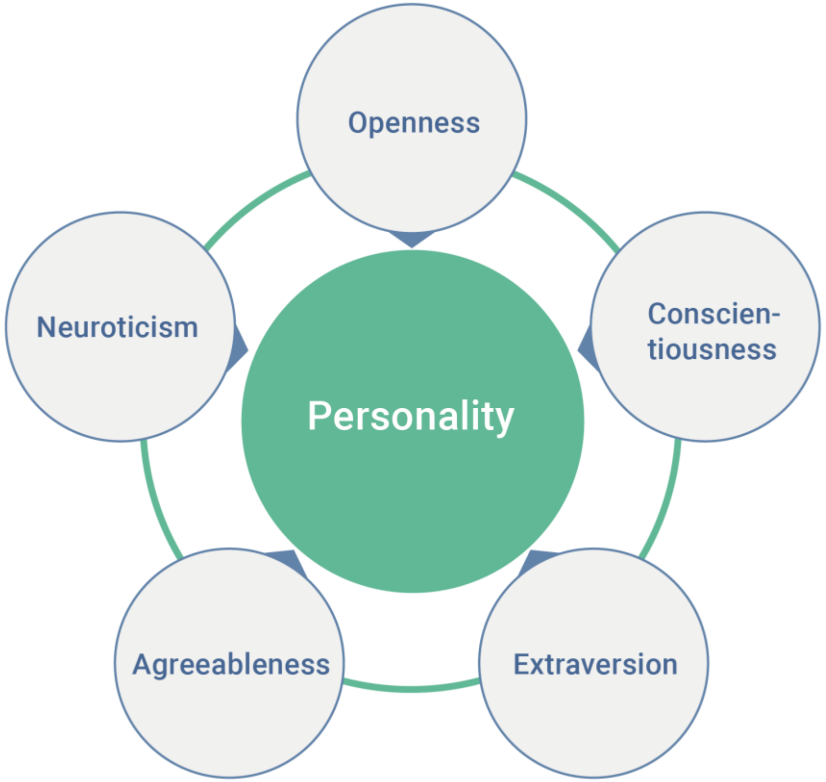
- Chit-Chat
- HappyTalk

Chaotbot with Chit Chat (3rd rank at NIPS2017 ConvAI Competition)



Current Approach

- Combine rule-based and learning-based chatbots
- Personalize with previous conversations
  - Big 5 personal traits



Ethics for Conversational Agents

- Unethical words
- Fairness/Bias
- Dilemma

- Learning human goals from interactions!

Ethics for Conversational Agents

➢Unethical words

Mar 24, 2016

Ethics for Conversational Agents

- Unethical words
- Fairness/Bias
- Dilemma

- Learning human goals from interactions!

Generic Approach: Learning Human Life Goals

- It is impossible to handle each ethical issue separately.
  - Failure of Rule-based Expert Systems
- Each AI companion be different.
  - **Learning** Life Goals from Mentor(s), i.e., Human Companion
- Human has option to use or not-use AI companion.
  - If choose to use, he/she will be responsible to the concequences.

Summary

- Emotion and Personality for Conversational Agents
  - Multimodal Recognition
  - Multimodal Generation
- Human Life Goal Learning

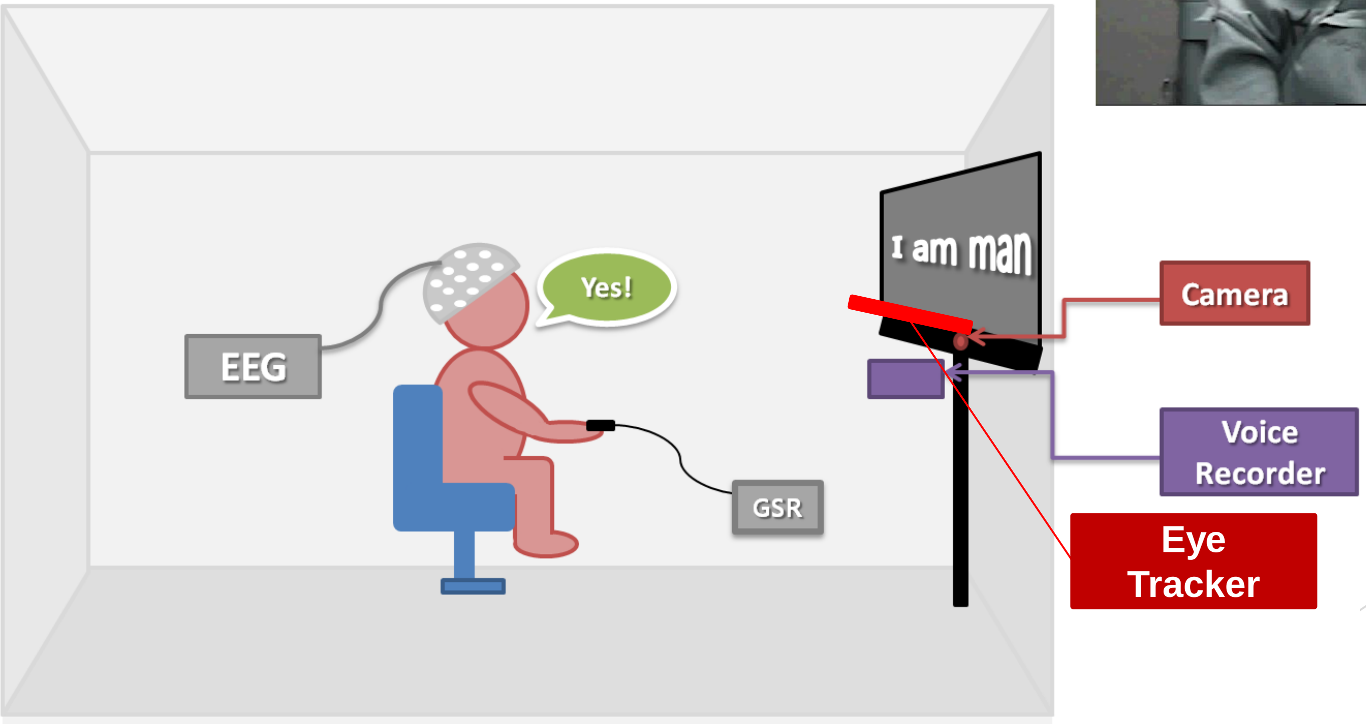
Understanding Mind: Human Internal States

- Agreement/Disagreement
- Trust/Distrust
- Preference

Agreement/Disagreement to Others

(S.Y. Dong, et al., Cognitive NS, 2015; IEEE T Cybernetics, 2015)

- fMRI
- EEG (29 scalp and 3 EOG/ECG)
- Eye tracker
- GSR, Video, and Speech

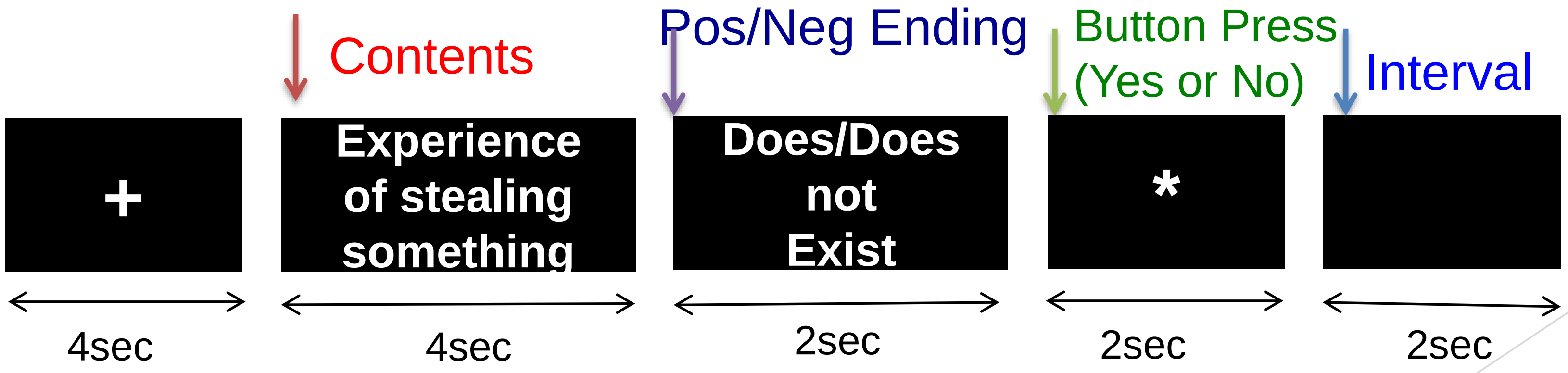


Experiment Design

- Stimulus sentences are all written in Korean
- Each sentence = Contents block + Sentence ending block
- Affirmative/Negative Sentences
- Contents are asking a personal experience/opinion

- English sentence : Subject – Verb – Object
- Korean sentence : Subject – Object – Verb (P/N)

Ex) Given sentence : “I had/had not stolen things”



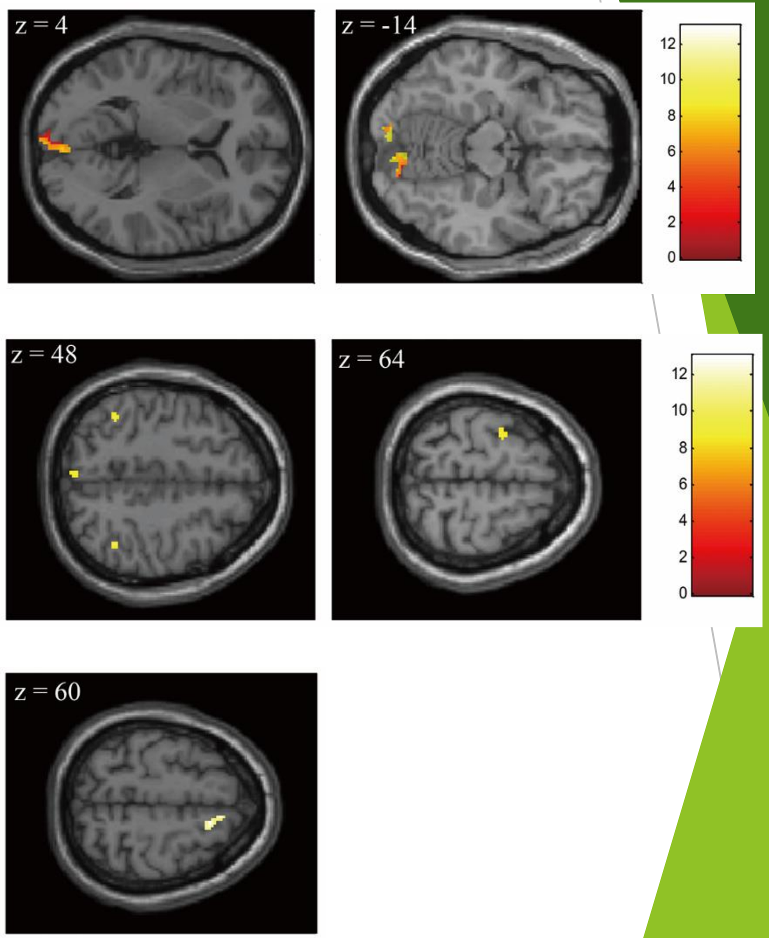
fMRI Results:

Activated regions on Contents vs. Fixations

(a) In both conditions: a small part of the visual cortex in the left and inter-hemispheric occipital lobe (z=4), both sides of lingual gyrus (z=-14)

(b) In the agreement condition: activity in the inferior parietal lobule on both sides, the left precuneus (z=48), and the left middle frontal gyrus (z=64)

(c) In the disagreement condition: activity in the **right superior frontal gyrus** (z=60)



EEG Results

- Three selected features

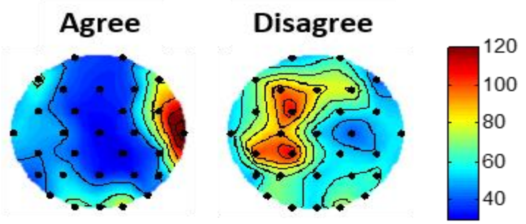
Channel selection based on t-test ($p<0.05$)

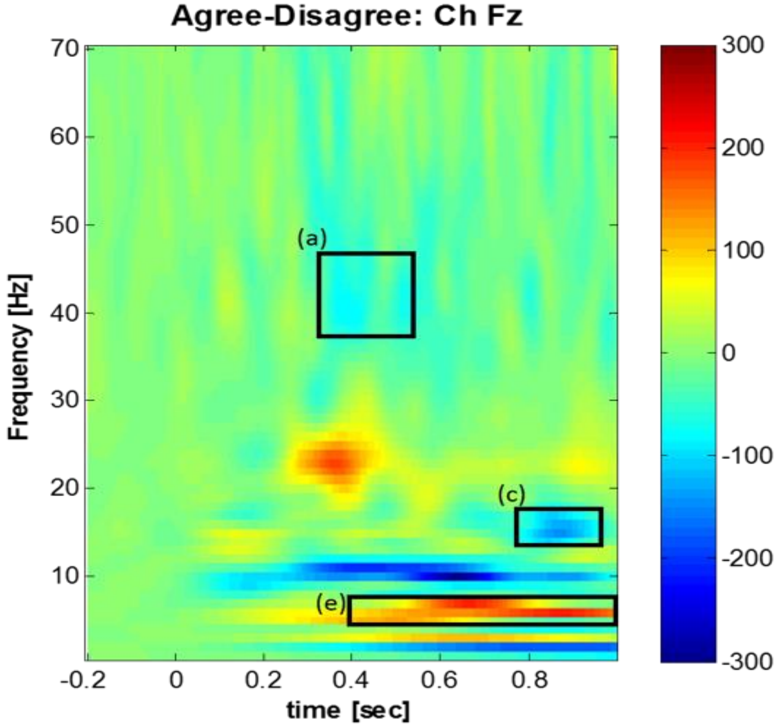
(a) gamma at F3

(c) beta at C4 and FC2

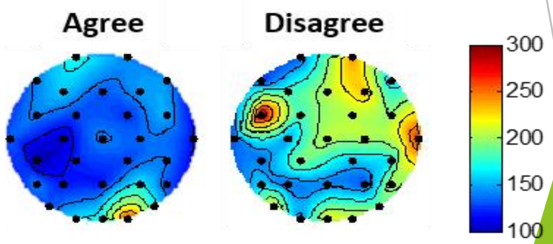
(e) theta at FC5

(a) Gamma scalp topography

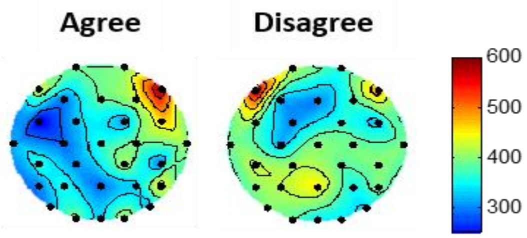




(c) Beta scalp topography

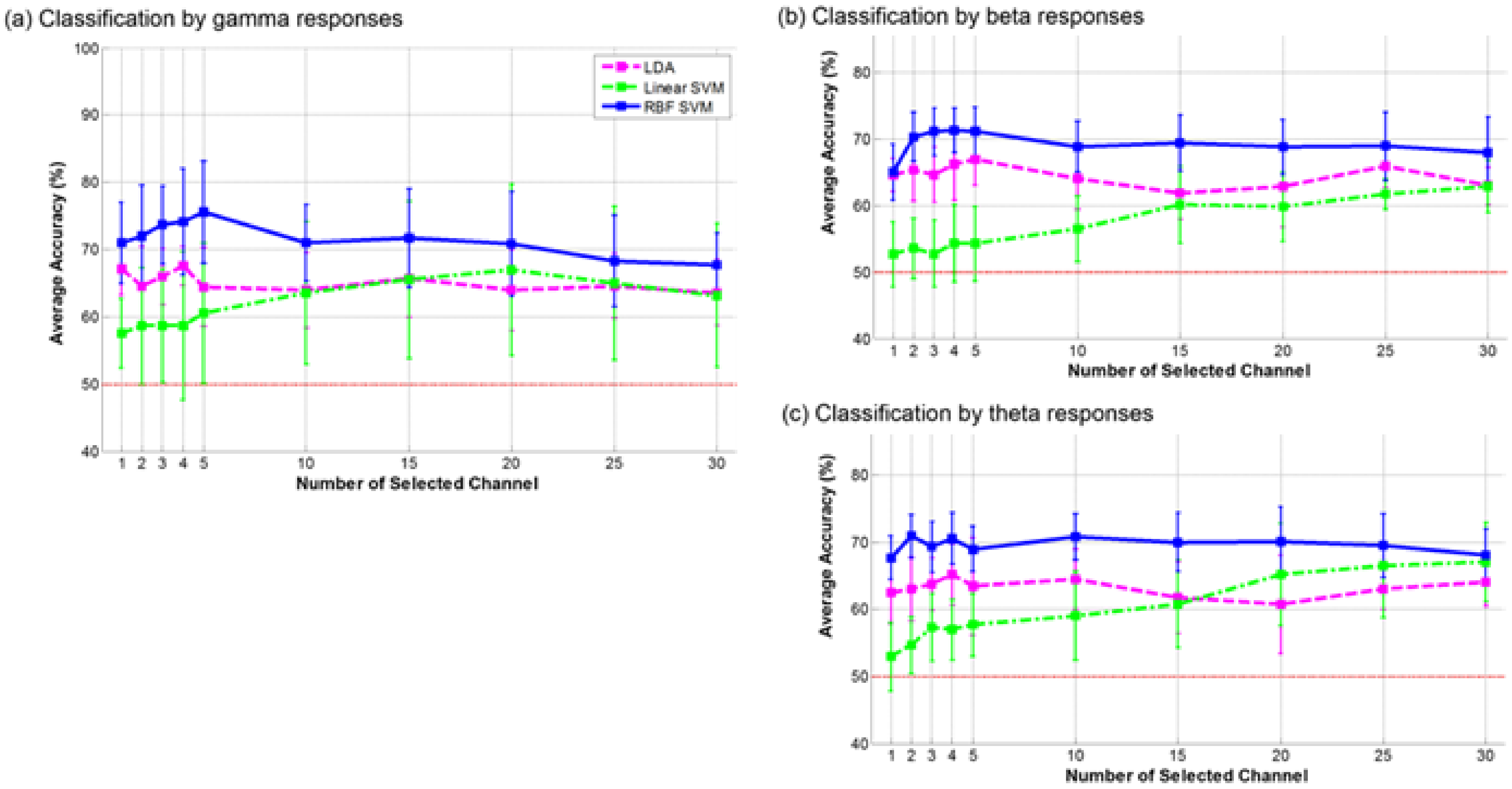


(e) Theta scalp topography



We can do Channel selection based on F-score!

Agreement/Disagreement Test Performance



Trust/Distrust between Human and AI

Trustworthiness

Trustworthiness Space

- Persistence: Consistency
- **Technical Competence**: Capability
- Fiduciary Responsibility: Collaboration or Egoism
- **Human-likeness**: Face, Speech, etc.

Design game-like experiments and measure brain signals

Theory-of-Mind Experiments

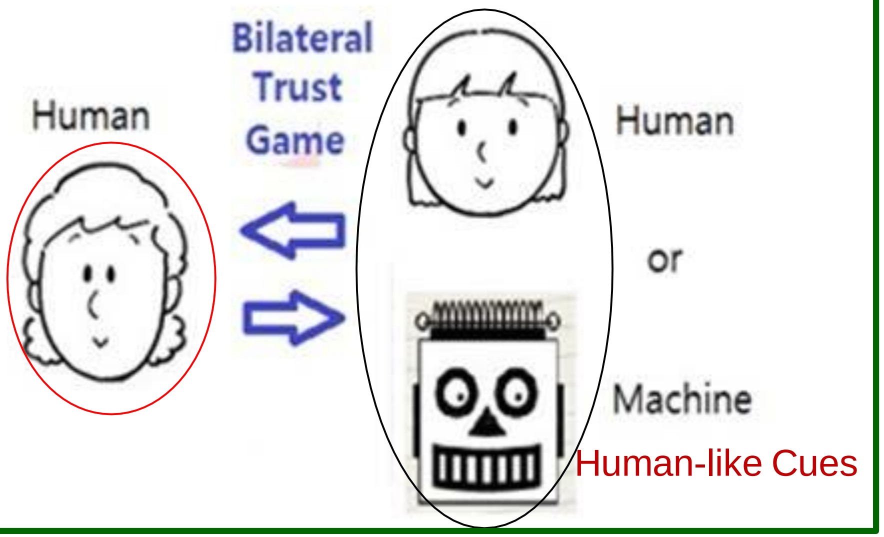
- Technical Competence
  - How far you and AI may consider the future?

(Source: Yoshida et al., 201

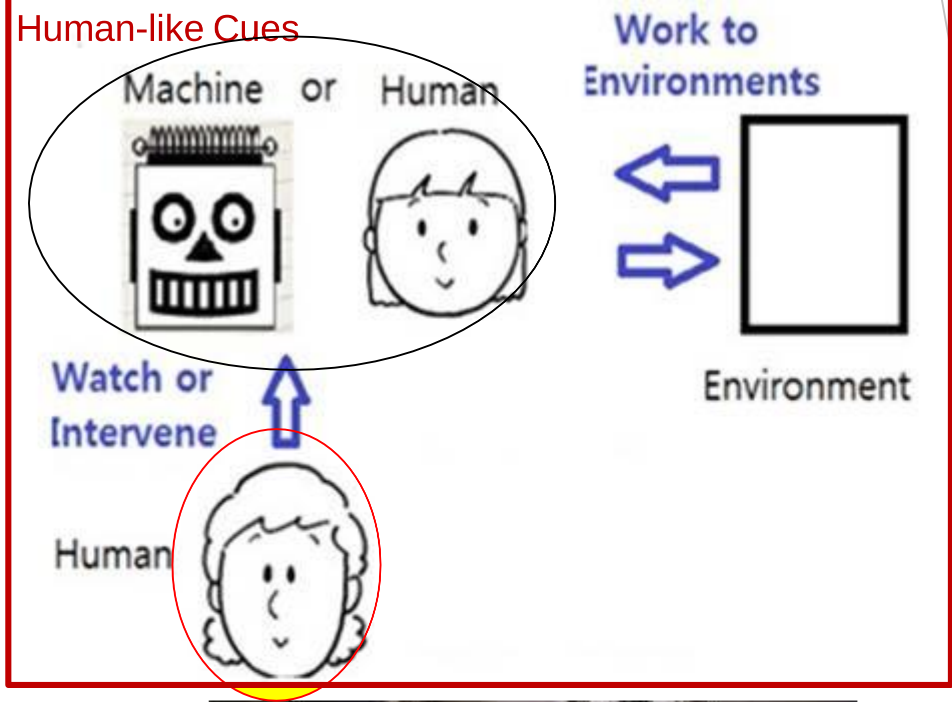
Bi/Uni-lateral Interactions

(E.K. Jung, et al., 2013; S.Y. Dong, et al, in preparation)

Bilateral Interaction



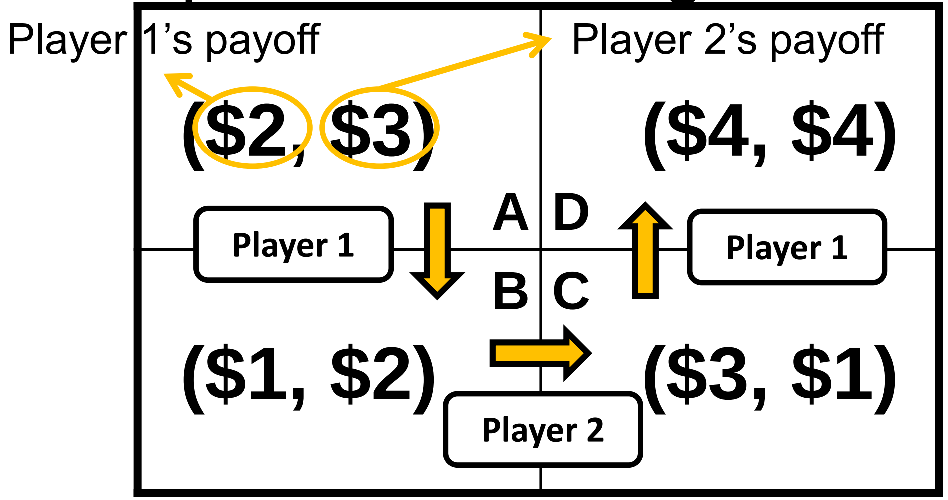
Unilateral Interaction



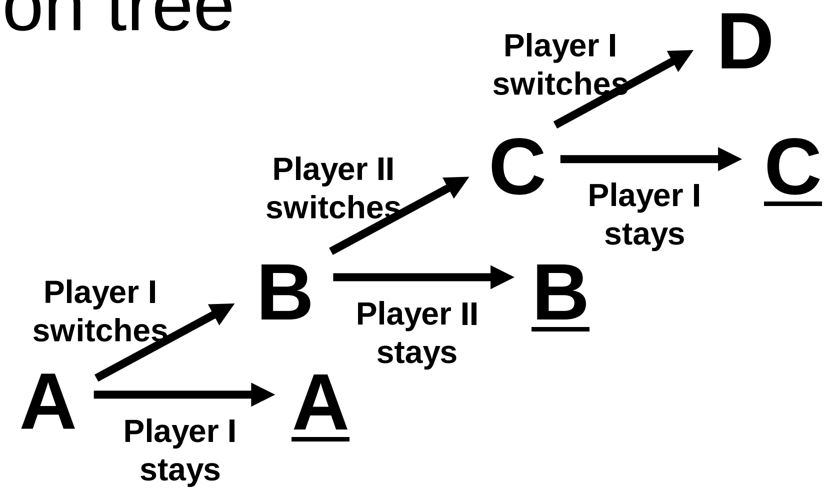
Autonomous Vehicle

Bilateral Experimental Design

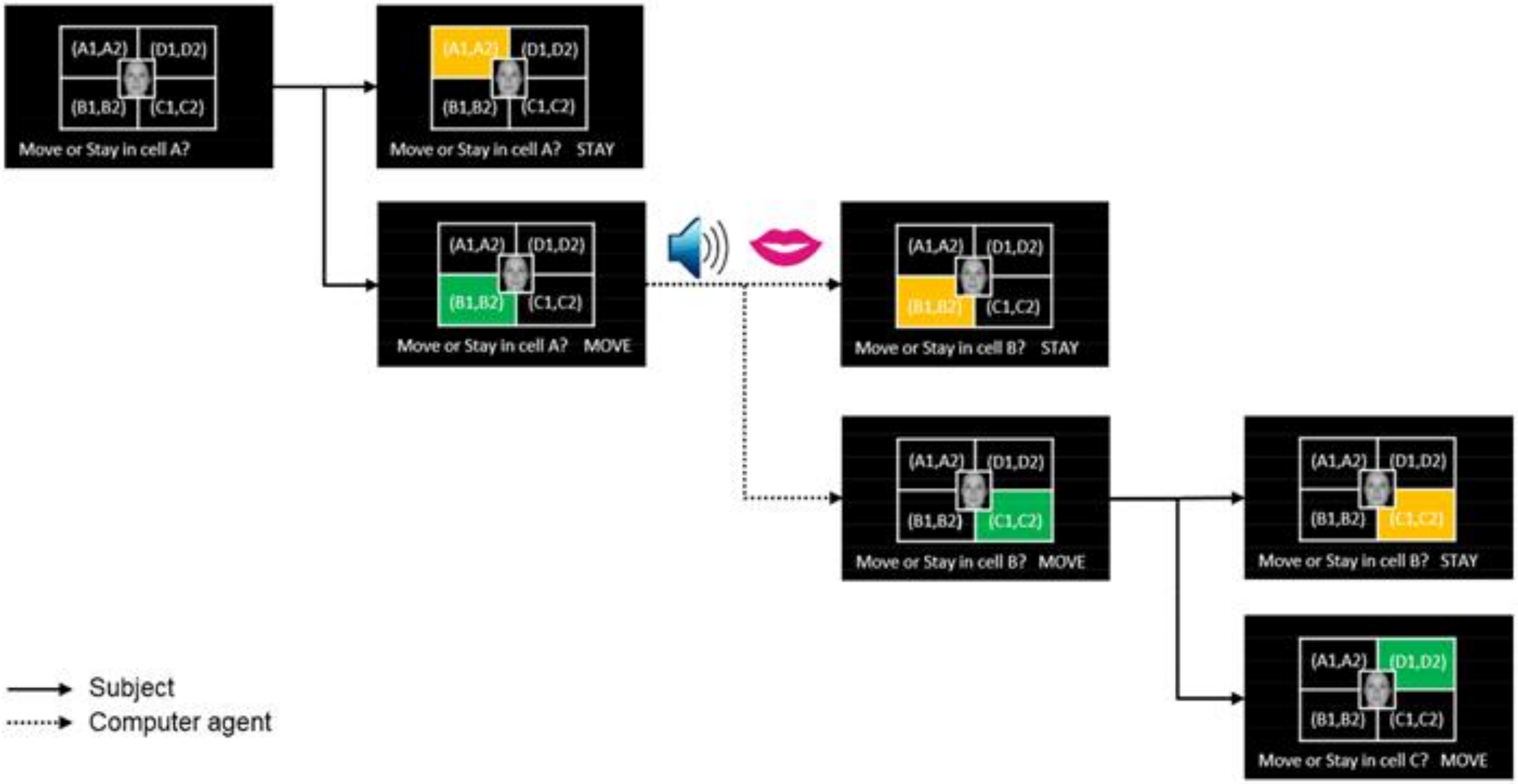
- A 2 × 2 sequential matrix game



- A decision tree



- ➢ A player decides whether to move (switch) or stop (stay) based on payoff in each cell.
- ➢ Player 1: participant
  Player 2: computer agent


(A1,A2)
(D1,D2)
(B1,B2)
(C1,C2)
Move or Stay in cell A?
Move or Stay in cell A? STAY
Move or Stay in cell A? MOVE
Move or Stay in cell B? STAY
Move or Stay in cell B? MOVE
Move or Stay in cell B? STAY
Move or Stay in cell C? MOVE
Subject
Computer agent

Experimental Design (cont'd)

Game types

EGO: "Egoist"

($2, $3) A	D ($4, $4)
($1, $2) B	C ($3, $1)

COL: "Collaboration"

($2, $3) A	D ($4, $4)
($1, $2) B	C ($3, $1)

Reasoning orders: an example game

Myopic (Zeroth-order)

($2, $3) A	D ($4, $4)
($1, $2) B	C ($3, $1)

The opponent will stay (stop).

Predictive (First-order)

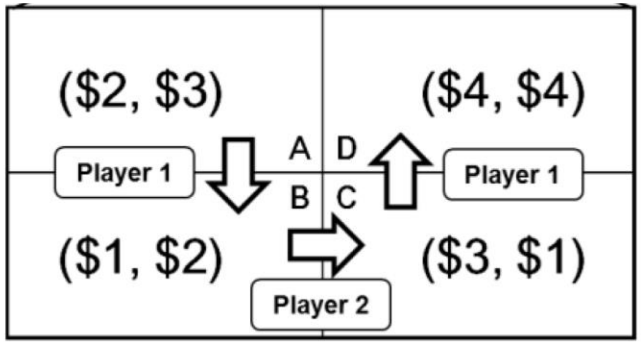
($2, $3) A	D ($4, $4)
($1, $2) B	C ($3, $1)

The opponent will move (switch).

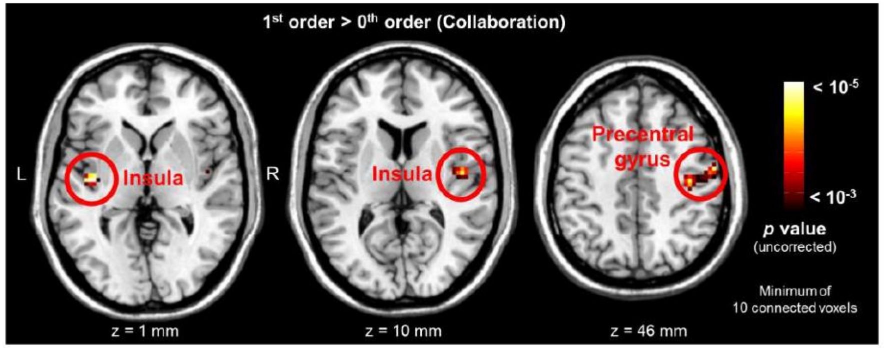
Capability: Prediction Level for Opponent's Action

Experiment goal: Trust level measurement according to opponent's technical ability during Theory-of-Mind game

- **Technical ability:** Myopic (0^{th} order) or Predictive (1^{st} order)
- **Given condition:** Collaboration or Egoism
- **TRUST level:** Expectation of opponent's technical ability (myopic or predictive)

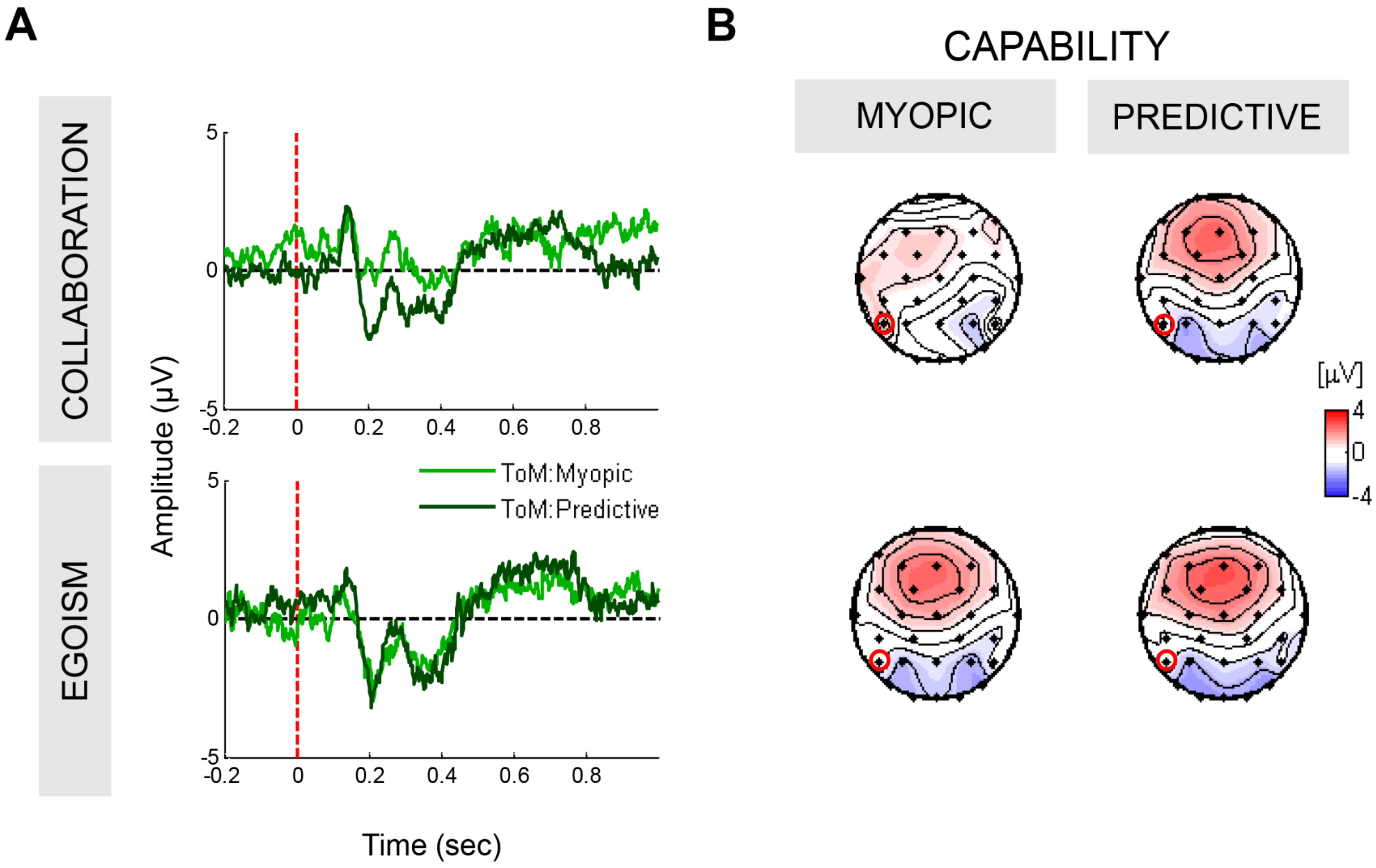


Player1 (P1): Participant (Human)
Player2 (P2): Computerized agent



E.K. Jung, J. Zhang, S.-Y. Lee, and J.-H. Lee, 'A Preliminary Study on Neural Basis of Collaboration: Mediated by the Level of Reasoning', ICONIP2013

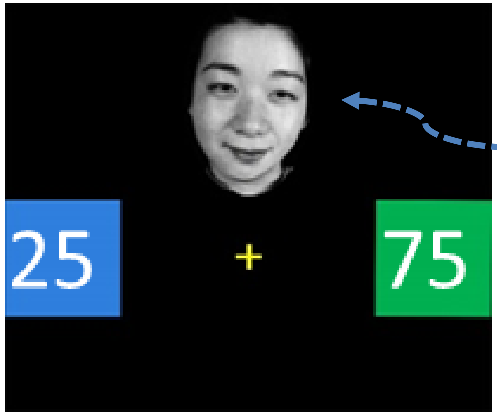
Averaged ERP from ToM Trials



Unilateral Interaction (Player-Supervisor Mode)

(E.S. Jung, et al., 2019; Scientific Reports)

- Iterative game play by machine agent *Player* with human *Supervisor*
  - Human trust on Agent iff Trustworthiness > Risk
- Effect of agent's human-likeness on Trustworthiness
  - {human-faced, robot-faced} agents
- Risk taking personality
  - {Low, Medium, High} risk taking

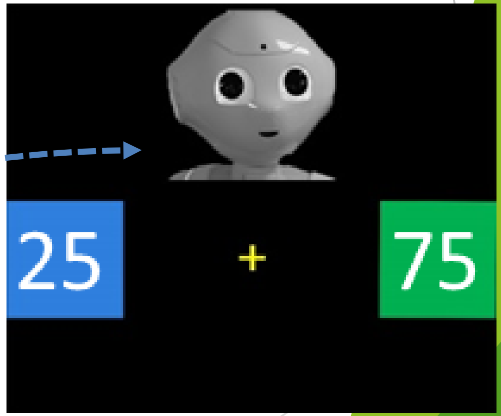


- *Human face*
- *Human voice*
- *Movements*
- *Facial expressions (smile/frown)*

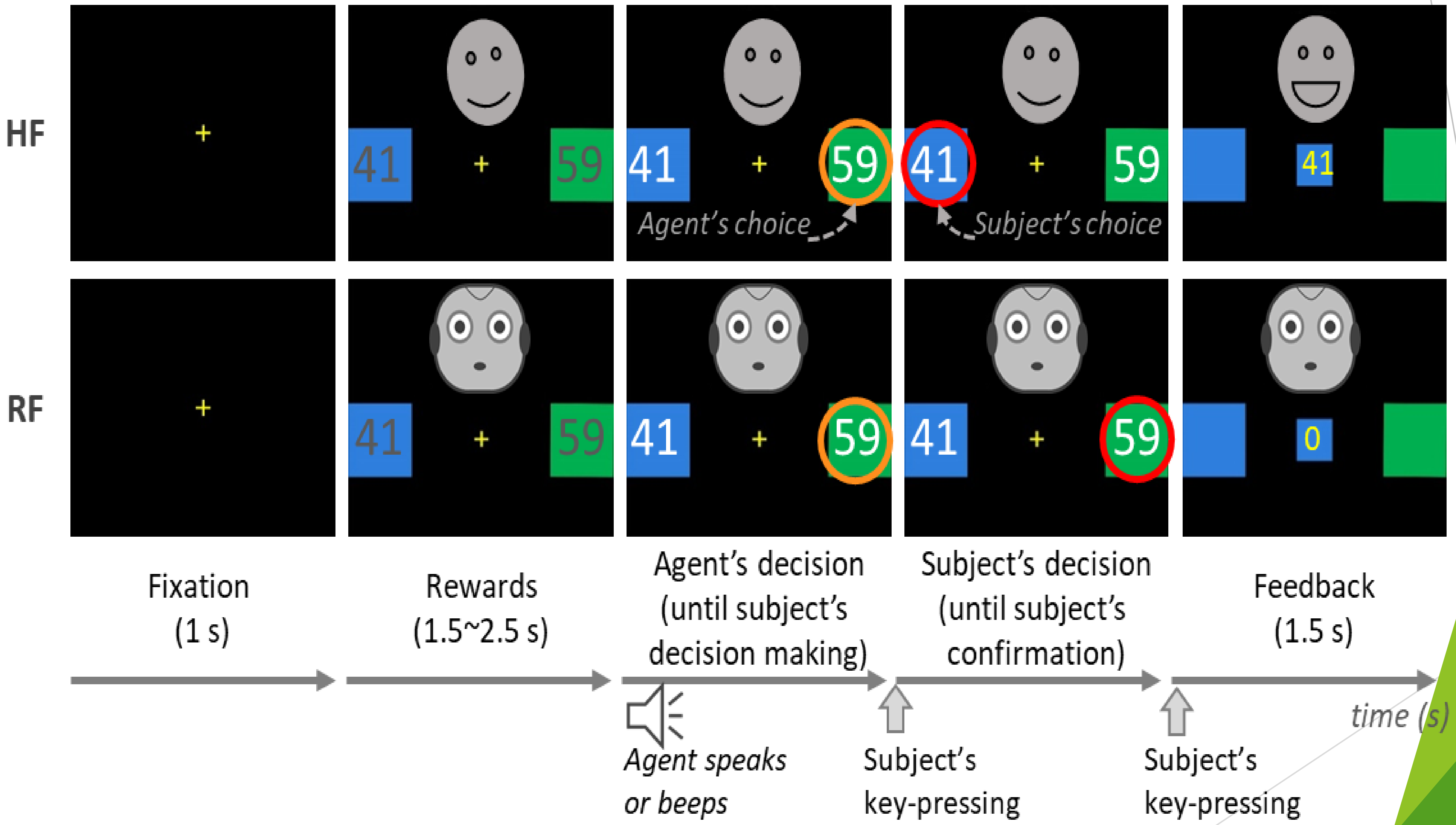
Prob of Correct: 0.75

Prob of Correct: 0.25

- *Robot face*
- *Beep sound*
- *No movement*
- *No emotion revealed*



Experimental Design

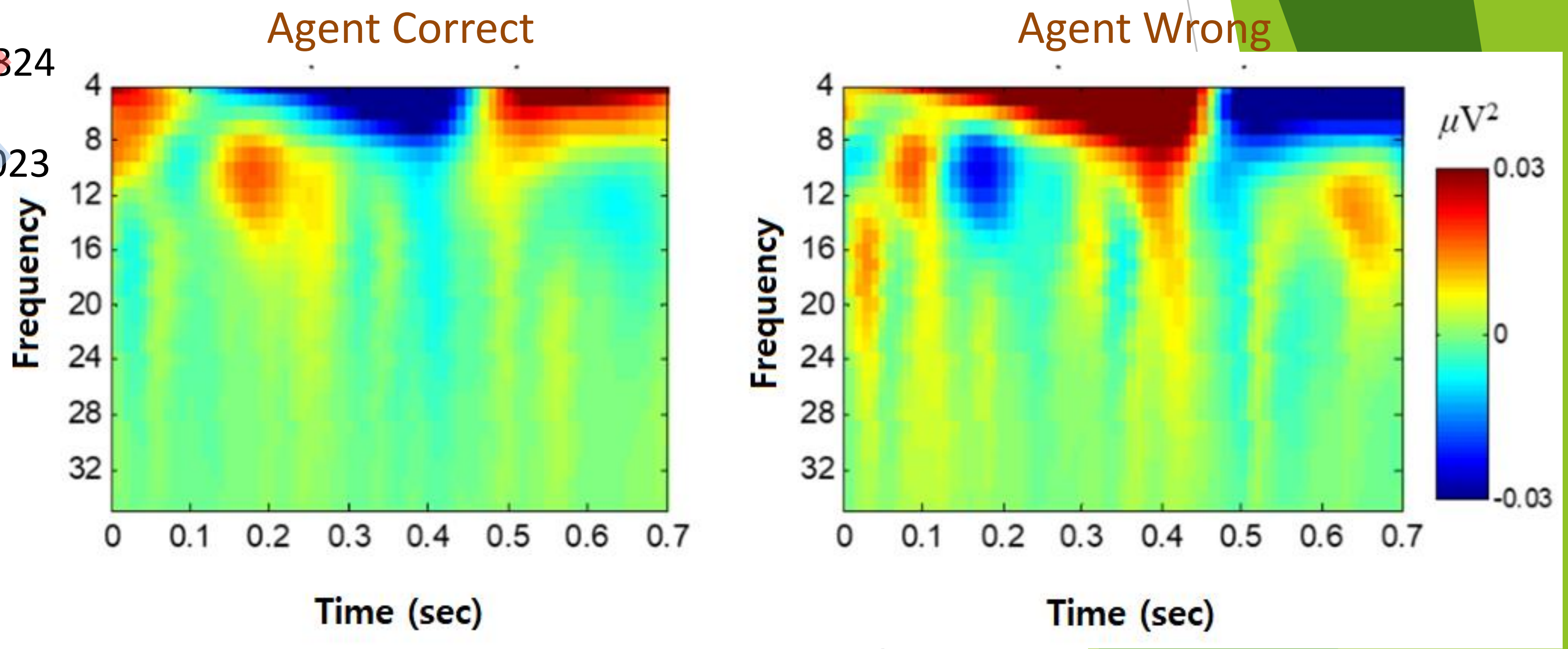


EEG Analysis

- EEG differences due to trust increase/decrease with t-test

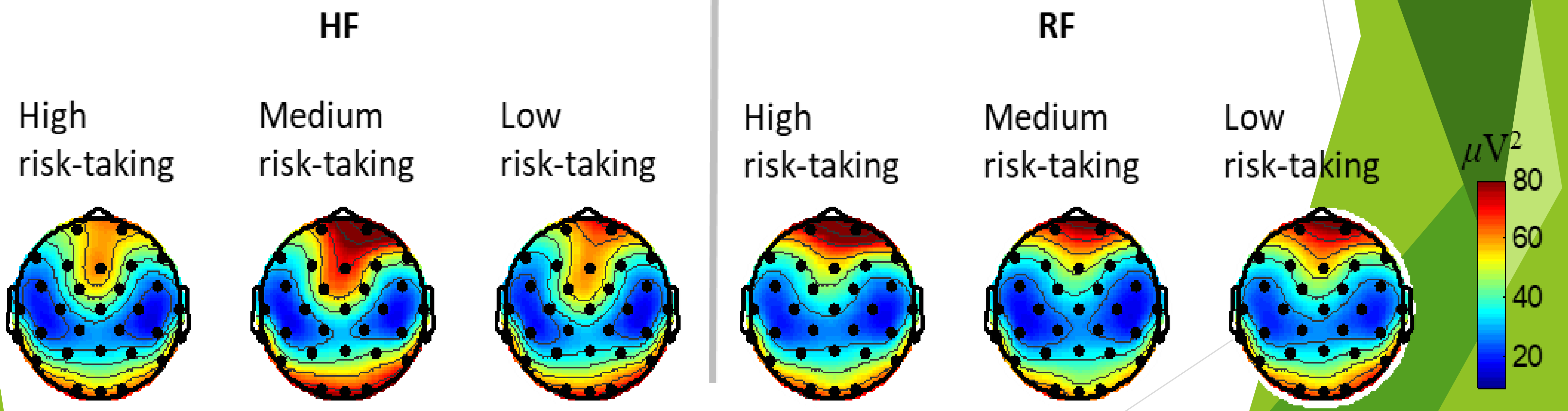
\# of trials

Final answer / **Agent's** answer	Correct (1703)	Wrong (1144)	
Correct	1519	305	1824
Wrong	184	839	1023



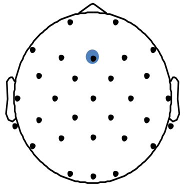
EEG Analysis: Personality Dependence

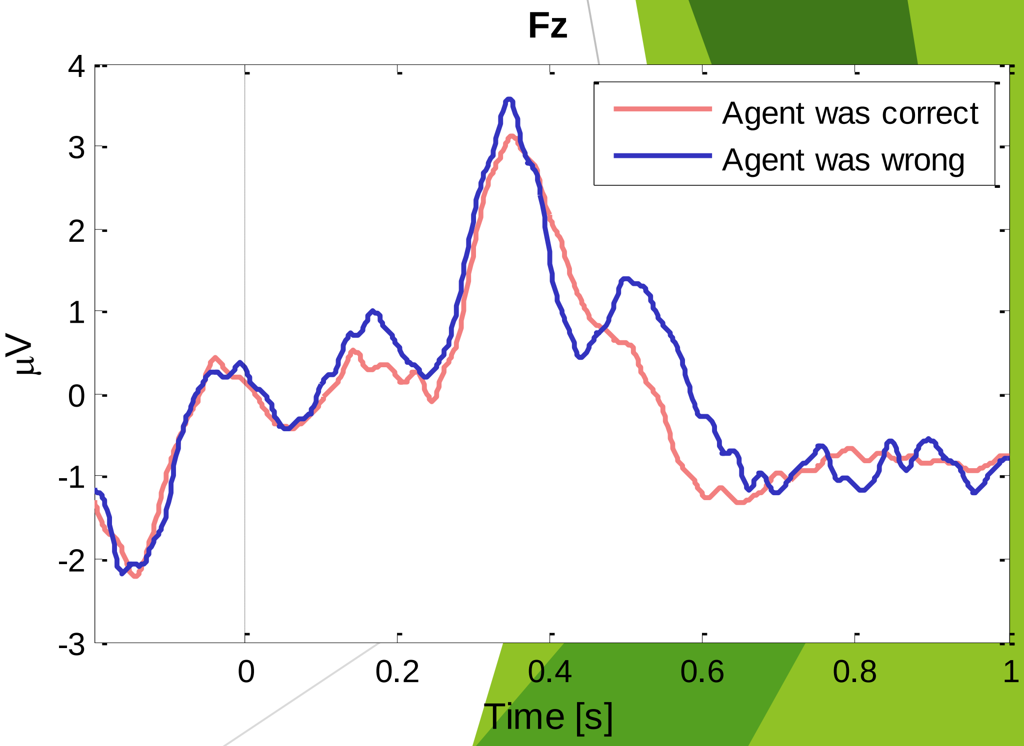
$$g_{\text{blue}} = F(p_{\text{blue}}, \gamma) \cdot r_{\text{blue}}$$
$$F(p_{\text{blue}}, \gamma) = \max[\min[\gamma(p_{\text{blue}} - 0.5) + 0.5, 1], 0]$$
$$\gamma = 0.7\ (High\ Risk\ Taking), 1, 1.5\ (Low\ Risk\ Taking)$$



EEG Analysis

- The number of intervenes on agents represented subjects' implicit trusts
  - More intervenes → low trust level
  - Each subject's intervenes reflected his/her own risk-taking personality
- **Trust changes** during feedback period
  - Different EEG responses





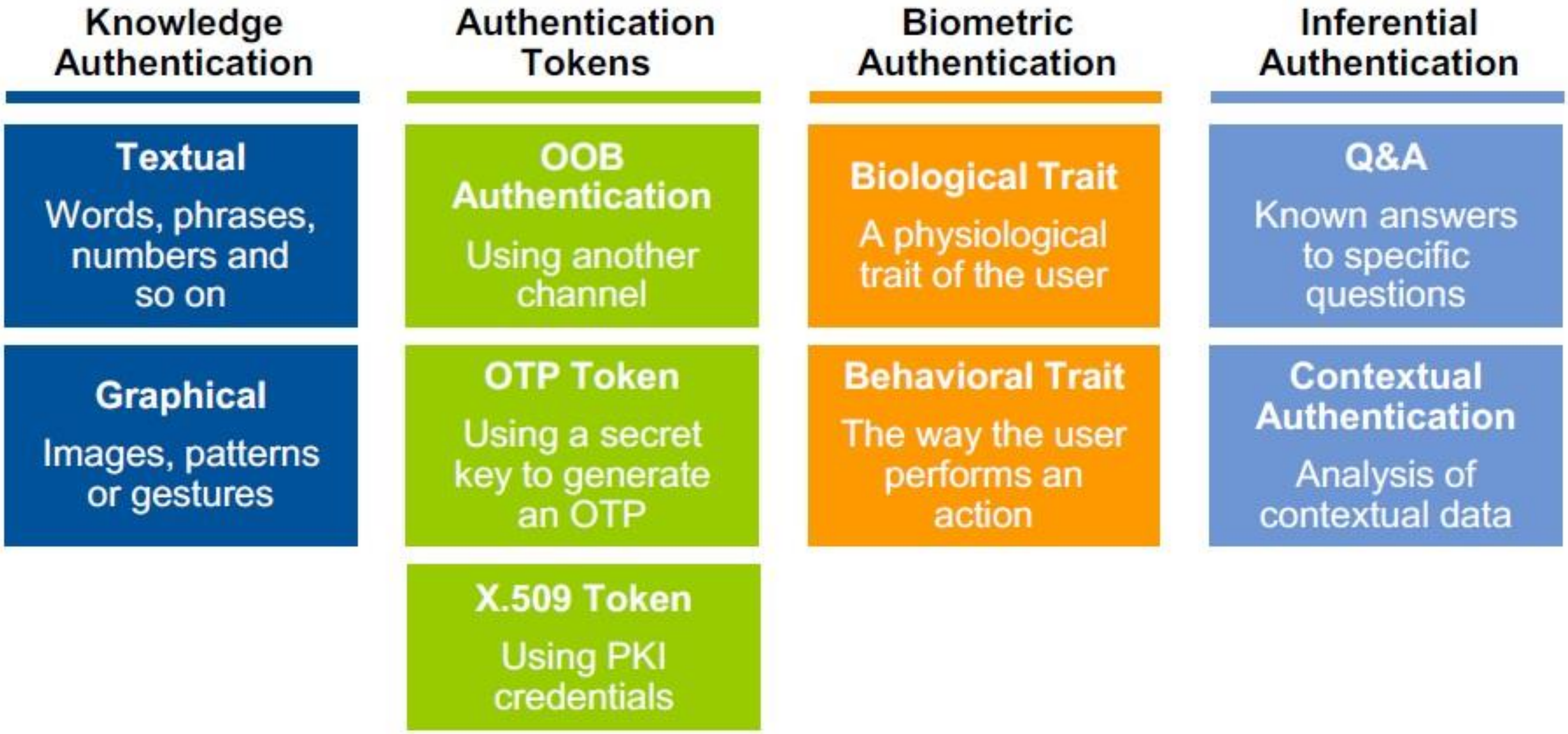
Human Trust on AI

- Human trusts AI more with
  - Similar personality (such as driving style)
  - Human-likeness (such as facial expression and speech)

- Maybe adopted to Human-AI Interfaces
  - For Digital Companion (Office Mate, Silver Mate, etc.), autonomous vehicles, etc.

User Authentication based on Preference

(E.S. Jung. et al., Scientific reports 2017)

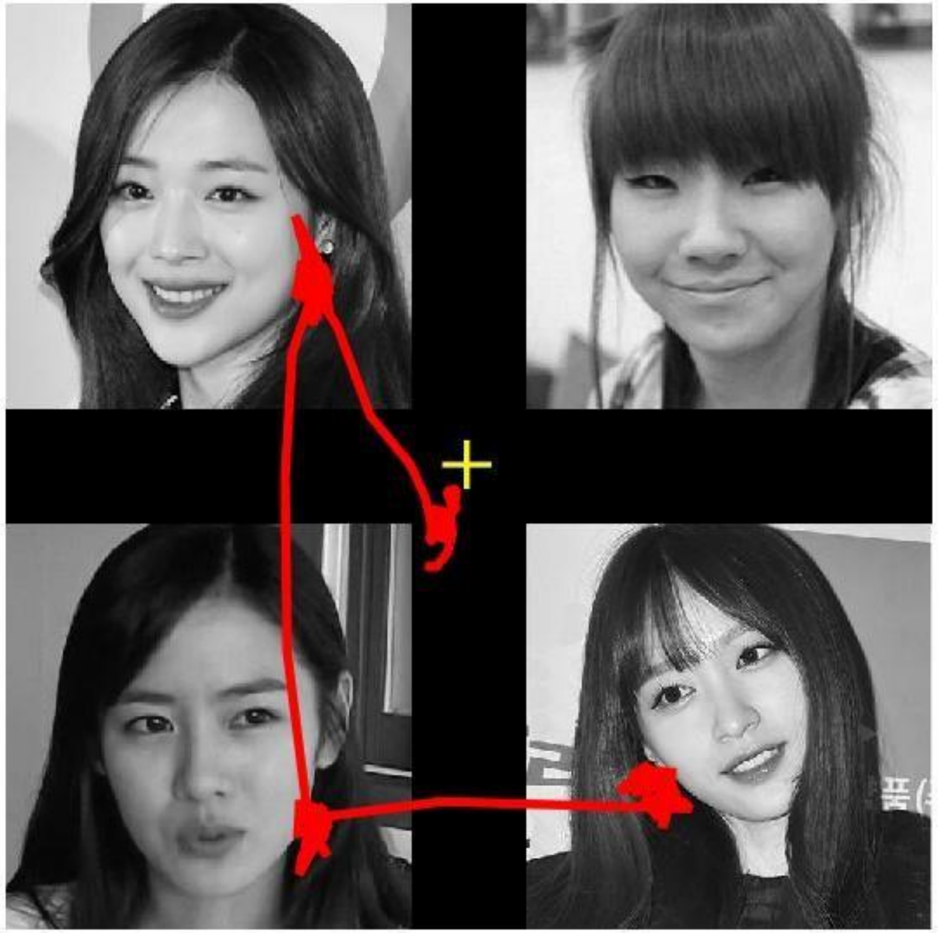


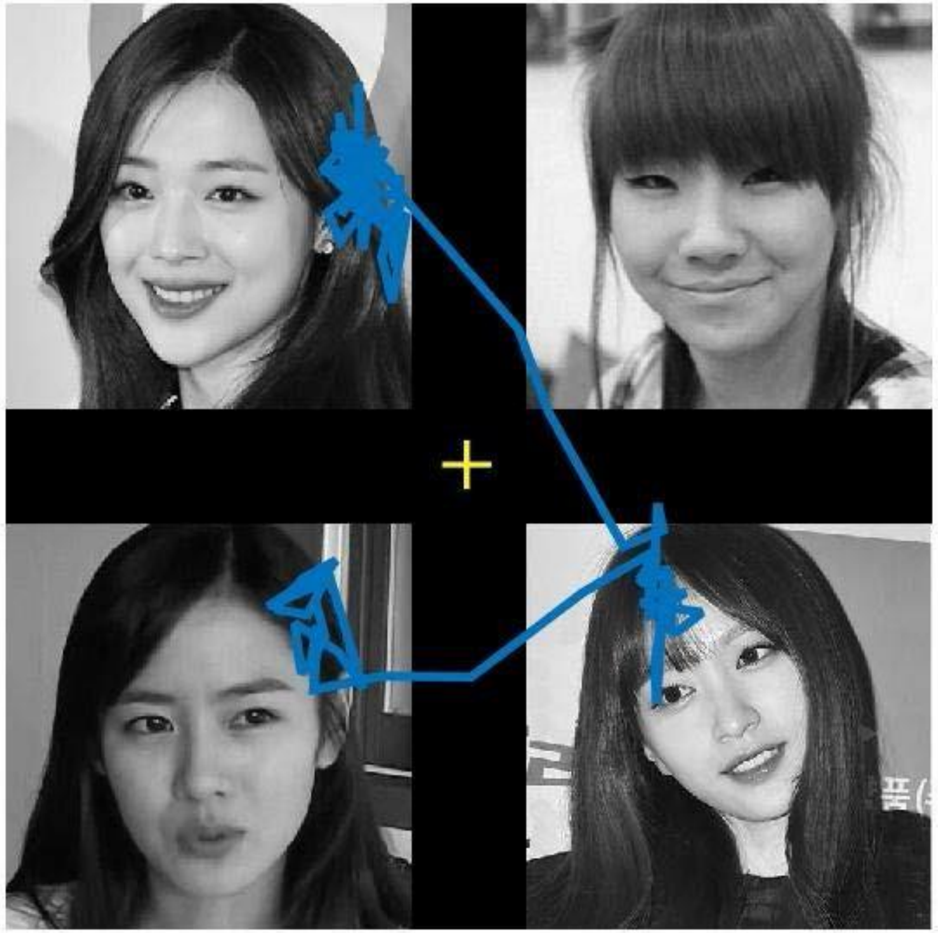
New Safest Authentication Technology

- Inferential Authentication
  - Question by Images
  - Answer by EEG or Eye Tracking

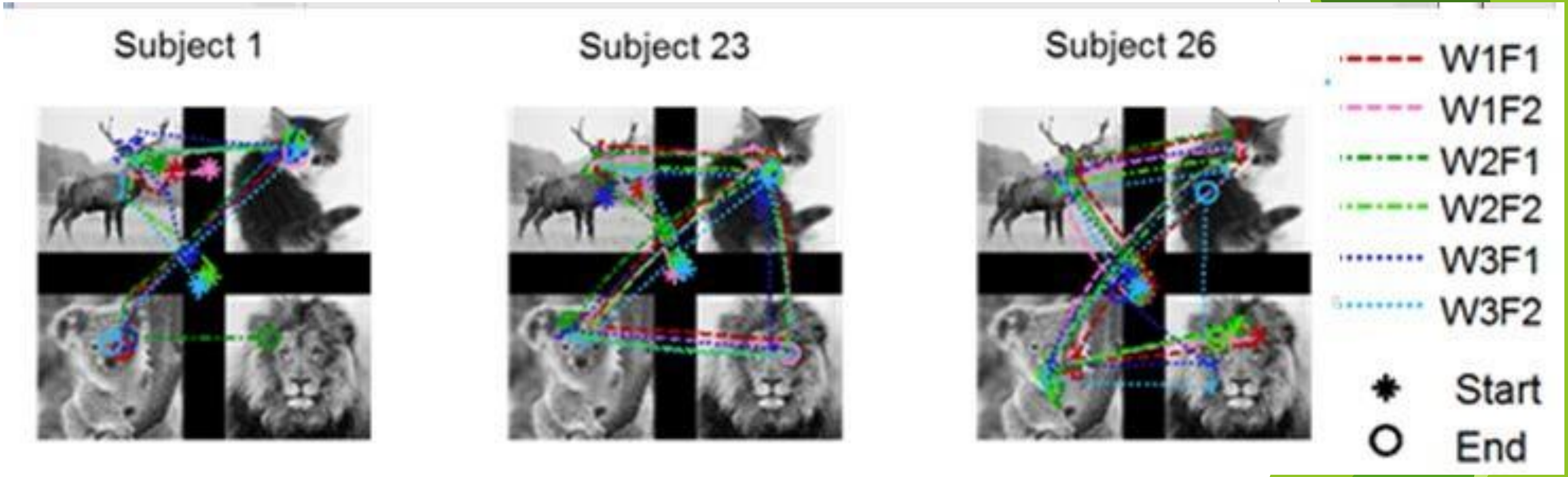
  - Safety: Involuntary responses can not be copied not stolen
  - Accuracy: Multiple Q&A for one authentication

Preference-based Eye Trajectory

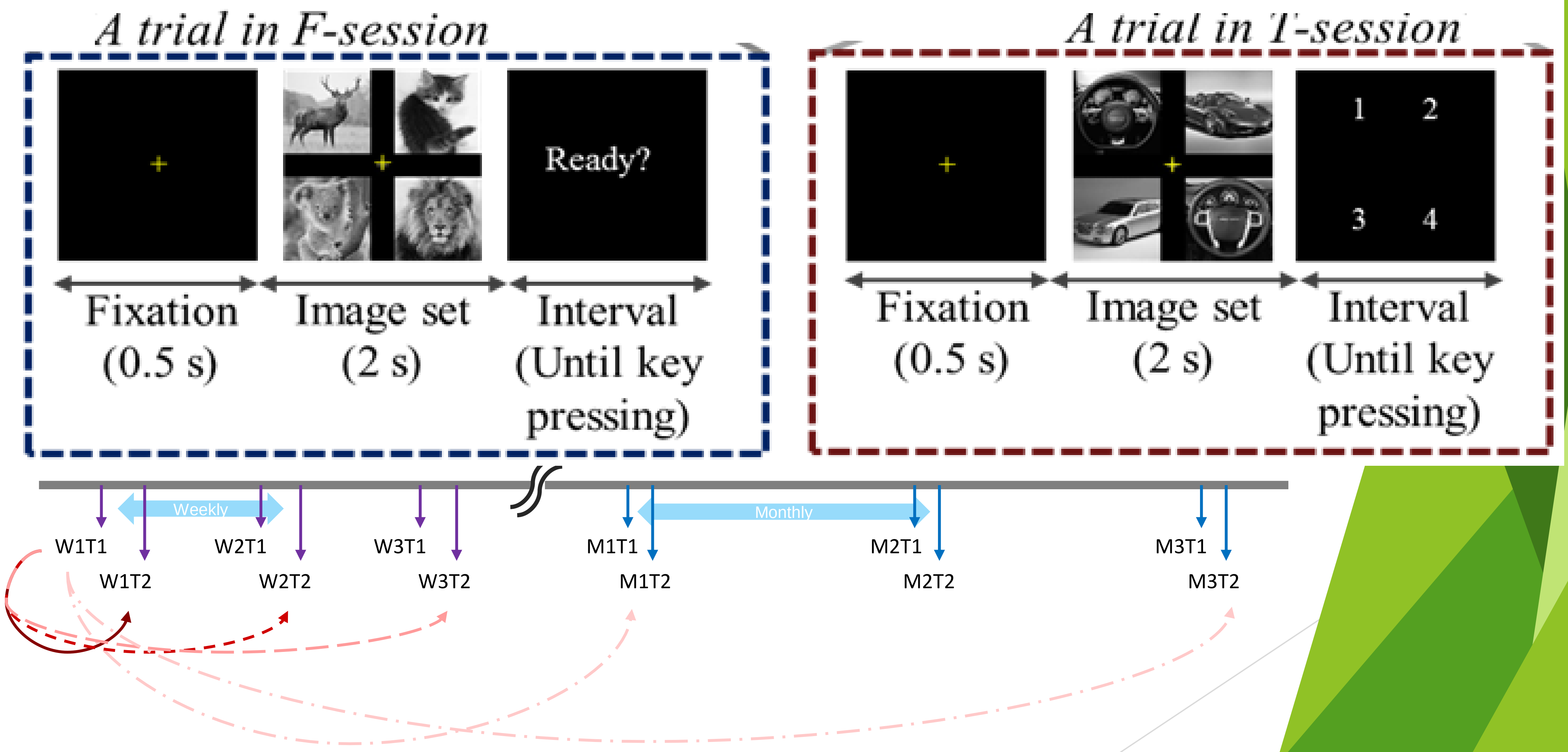




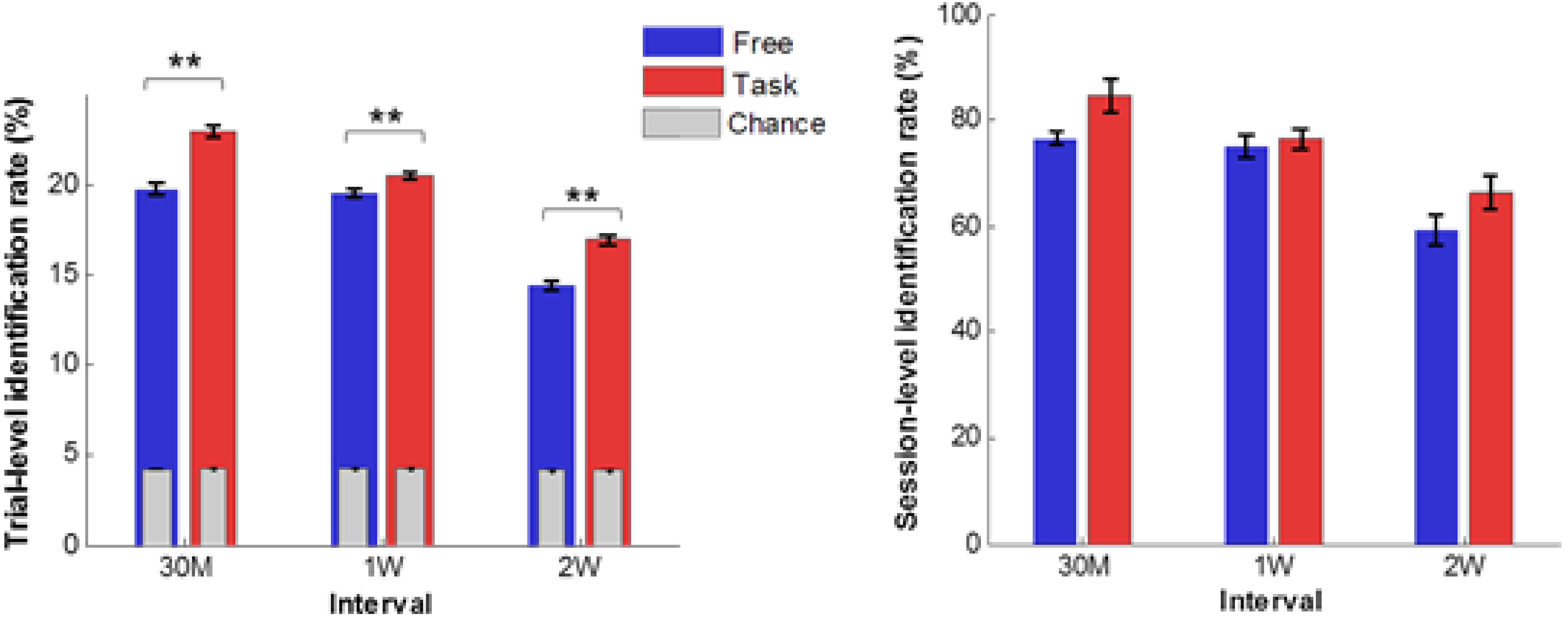
Multi-Image Eye Trajectories



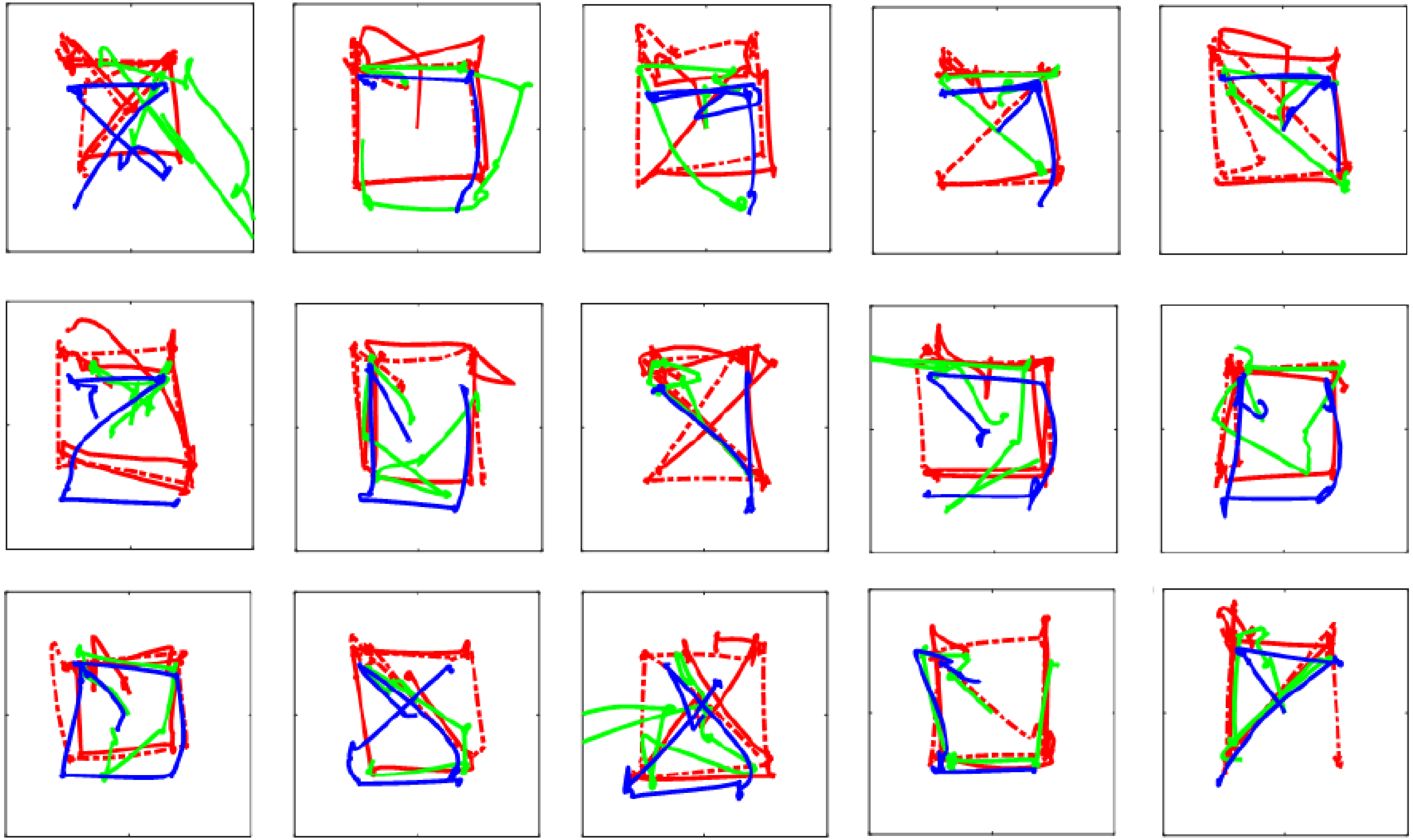
User-Authentication by Eye Tracking (Scientific Report 2017)



Identification Accuracy: Scanpath



Intrusion Experiments



Summary

Next-Generation Office Mates and Data Analytics

- Develop Digital Companions (Office Mate) with **Mind (Internal States) and Environmental States**
  - **Internal states**: personality and experience of human and agents, emotion of agents, trust and binding between human and agents, etc.
  - **Environmental and unknown states**: road condition, economy, politics conditions, social events, etc.
  - Learning internal and environmental states from data
  - Top-down attention for accurate and fair analytics with multimodal integration
  - **Personal and Interactive at Anytime Anywhere**