Put That There:
~~20~~ 30 Years of Multimodal Interaction

Prof. James L. Crowley
Laboratoire Informatique de Grenoble
INRIA GRA Research Centre
Univ. Grenoble-Alpes

MIT Media Lab

Put That There

November 2, 1979

The Architecture Machine

30 Years of Multimodal Interaction

- The International Conference on Multimodal Interaction
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

The Origins of ICMI (a personal View)

Preambule:

Science is performed by communities of scientists that share problems and problem solutions (Paradigms).

Communities emerge, mature, compete, stagnate, and die.

T.S. Kuhn, *The structure of scientific revolutions*. University of Chicago press, 2012.

The ICMI community is a typical, unremarkable, example.

The Origins of ICMI (a personal View)

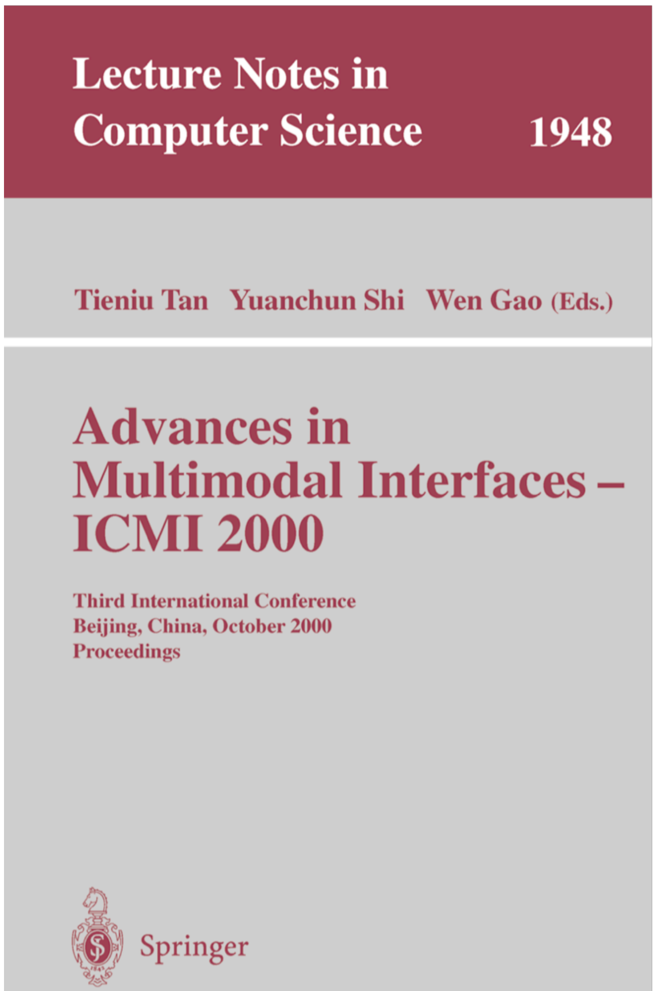
- Workshop on Perceptual User Interfaces – Mathew Turk
  - From GUI to PUI (IJCAI, Chambery 1993)
  - PUI (Banff 1997, San Francisco 1998, Orlando 2001)
- International Conference on Multimodal Interaction
  - 1st ICMI, Beijing 1996
  - 2nd ICMI, Hong Kong 1998
  - 3rd ICMI, Beijing 2000 (Tieniu Tan, Yuanchun Shi, Wen Gao)
  - 4th ICMI, Pittsburgh 2002 (Alex Waibel, Wen Gao)
- ICMI Advisory Board formed in 2003 by Sharon Oviatt
- ACM/CHI Sponsorship 2004

Perceptual User Interfaces - Banff 97

ICMI, Beijing Oct. 2000

Jim, Sharon and Jeff (Provided by Jeff Cohn

3rd International Conference on Multimodal Interaction Beijing, October 14-16, 2000



<u>Session Titles</u>

- Affective and Perceptual Computing
- Gesture Recognition
- Facial Expression Detection, Recognition and Synthesis
- Multilingual Interfaces and Natural Language Understanding
- Speech Processing and Speaker Detection
- Object Motion, Tracking and Recognition
- Handwriting Recognition
- Input Devices and Its Usability
- Virtual and Augmented Reality
- Multimodal Interfaces for Wearable and Mobile Computing
- Sign Languages and Multimodal Navigation for the Disabled
- Multimodal Integration and Application Systems

ACM Records for ICMI 2002 - 2017

Name	Location	Podium/Poster	Acc. Rate	Attendance
ICMI '02	Pittsburgh	87/165	53%	?
ICMI-PUI '03	Vancouver	45/130	35%	170
ICMI '04	State College	43/85	51%	127
ICMI '05	Trento	44/97	45%	102
ICMI '06	Banff	40/81	49%	87
ICMI '07	Nagoya	55/99	55%	143
ICMI '08	Chania	44/92	48%	111
ICMI-MLMI '09	Cambridge, Mass	41/118	35%	153
ICMI-MLMI '10	Beijing	41/100	45%	66
ICMI '11	Alicante	47/120	39%	130
ICMI '12	Santa Monica	44/123	36%	214
ICMI '13	Sydney	49/133	37%	171
ICMI '14	Istanbul	51/127	40%	242
ICMI '15	Seattle	52/127	41%	200
ICMI '16	Tokyo	55/144	38%	242
ICMI '17	Glasgow	65/149	43%	?

Comment on the Topics covered at ICMI

Looking back…

Most papers have been about multimodal perception, new modalities for perception and display and perceptual user interfaces.

Very few papers have presented theories, models, techniques or frameworks for multimodal interaction…. why?

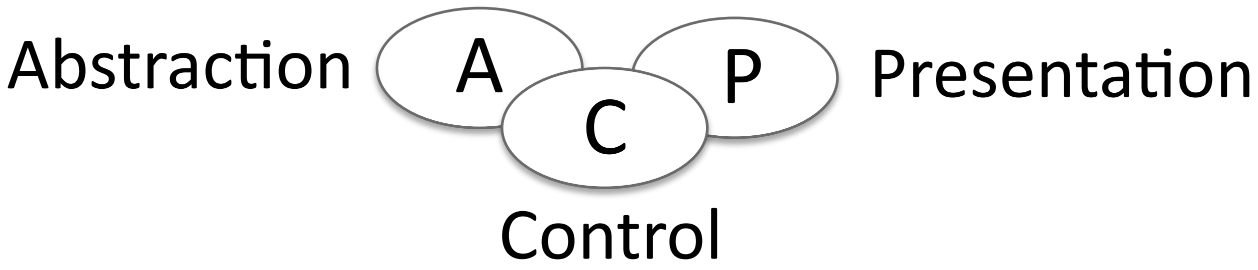
Possible explanation: … lack of established paradigms.

Papers addressing problems within established paradigms tend to receive higher scores in reviews.

30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interactive Systems
  - Presentation – Abstraction – Control
  - PAC vs MVC
  - PAC Component Model for Perception User Interfaces
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

PAC: Presentation-Abstraction-Control

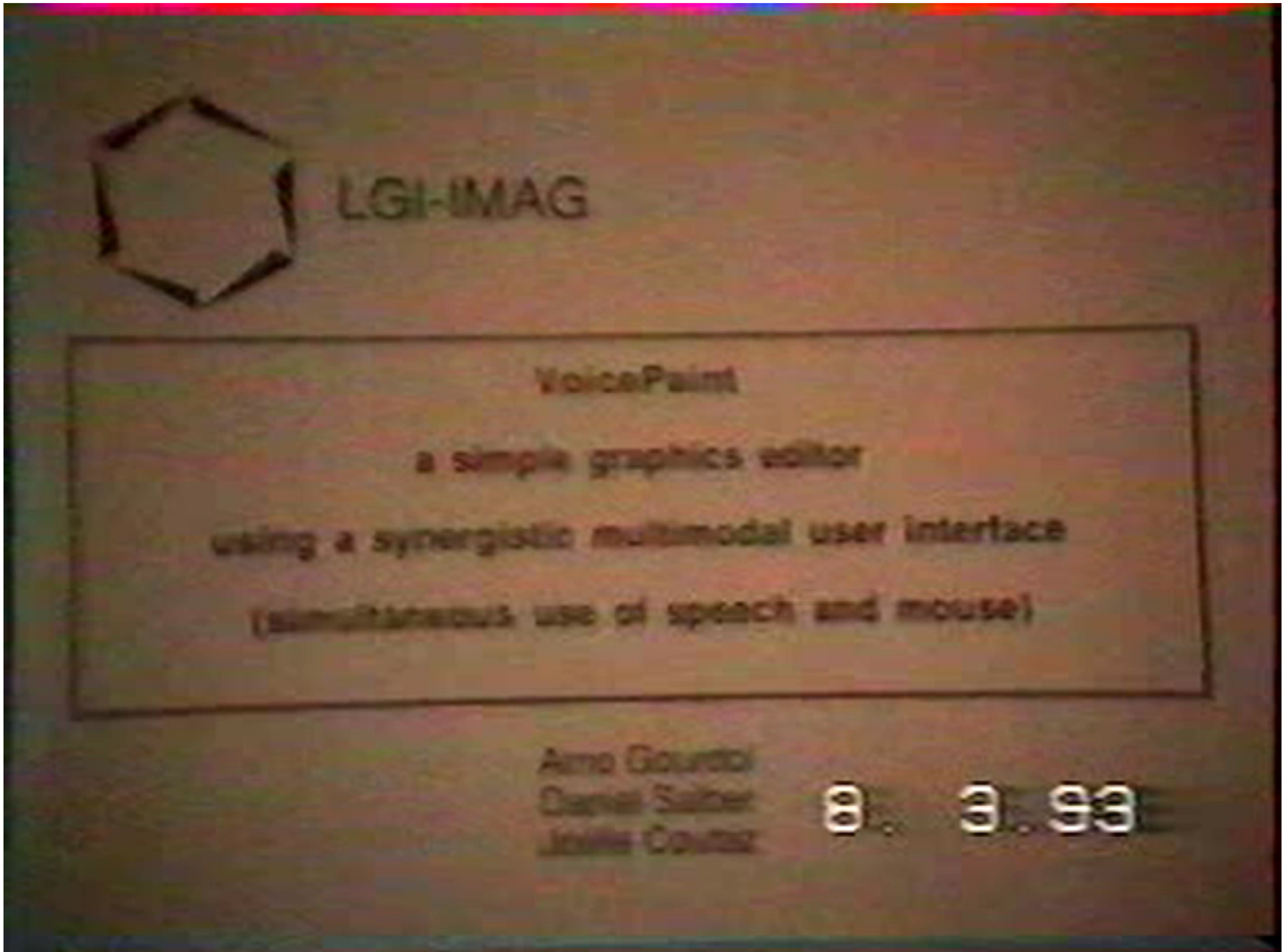


Presentation: Input – Output Rendering
Abstraction: Functional Model
Control: Communication and Coordination.

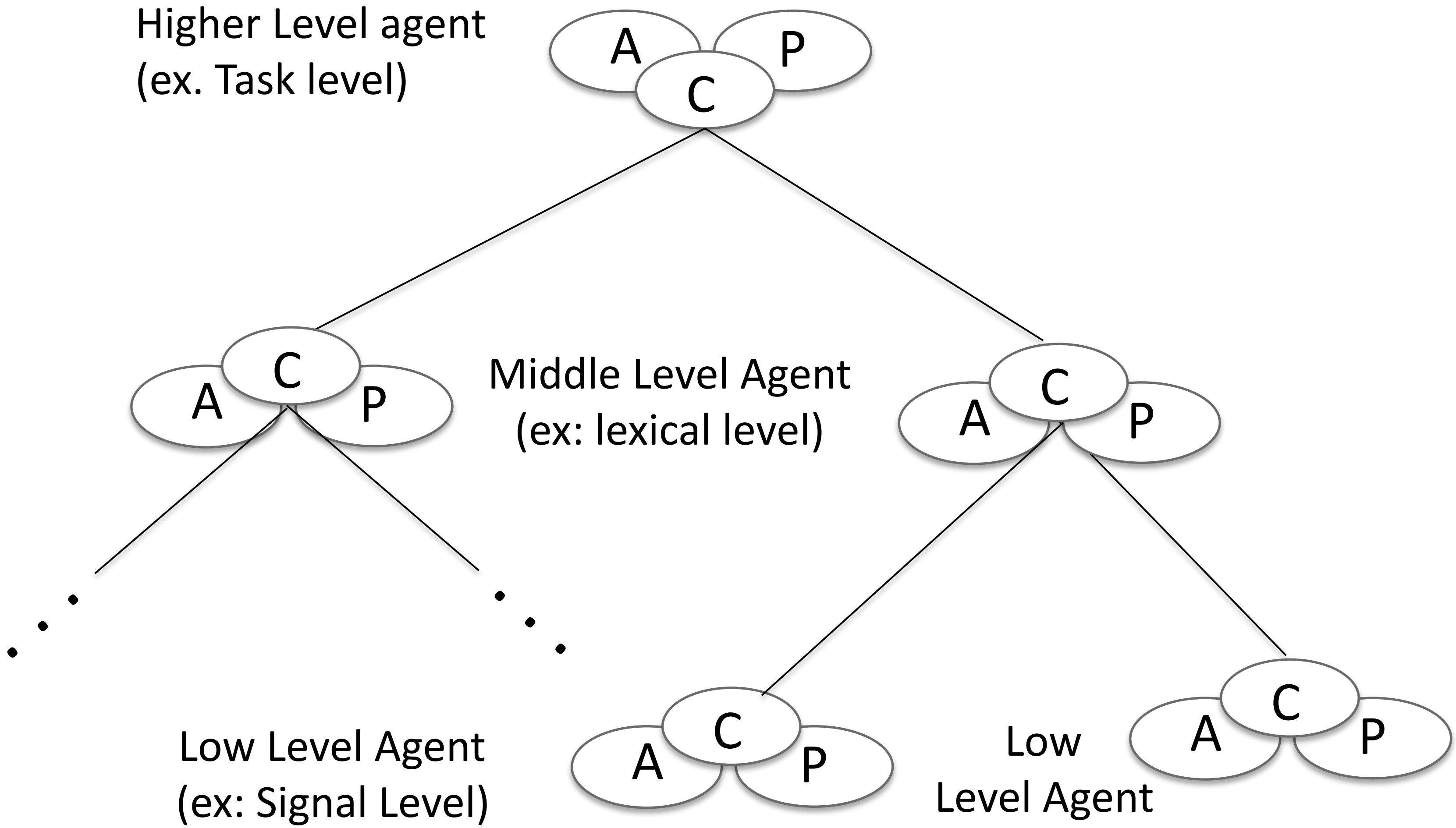
PAC* is an Architectural Model (Design Pattern) for multi-user multi-modal interactive Systems.
PAC Facilitates multimodal interaction (fusion and fission) through hierarchical decomposition.

*Coutaz, J, PAC, An Object Oriented Model for Dialog Design. In *Human–Computer Interaction, INTERACT'87*, pp. 431-436 , 1987.

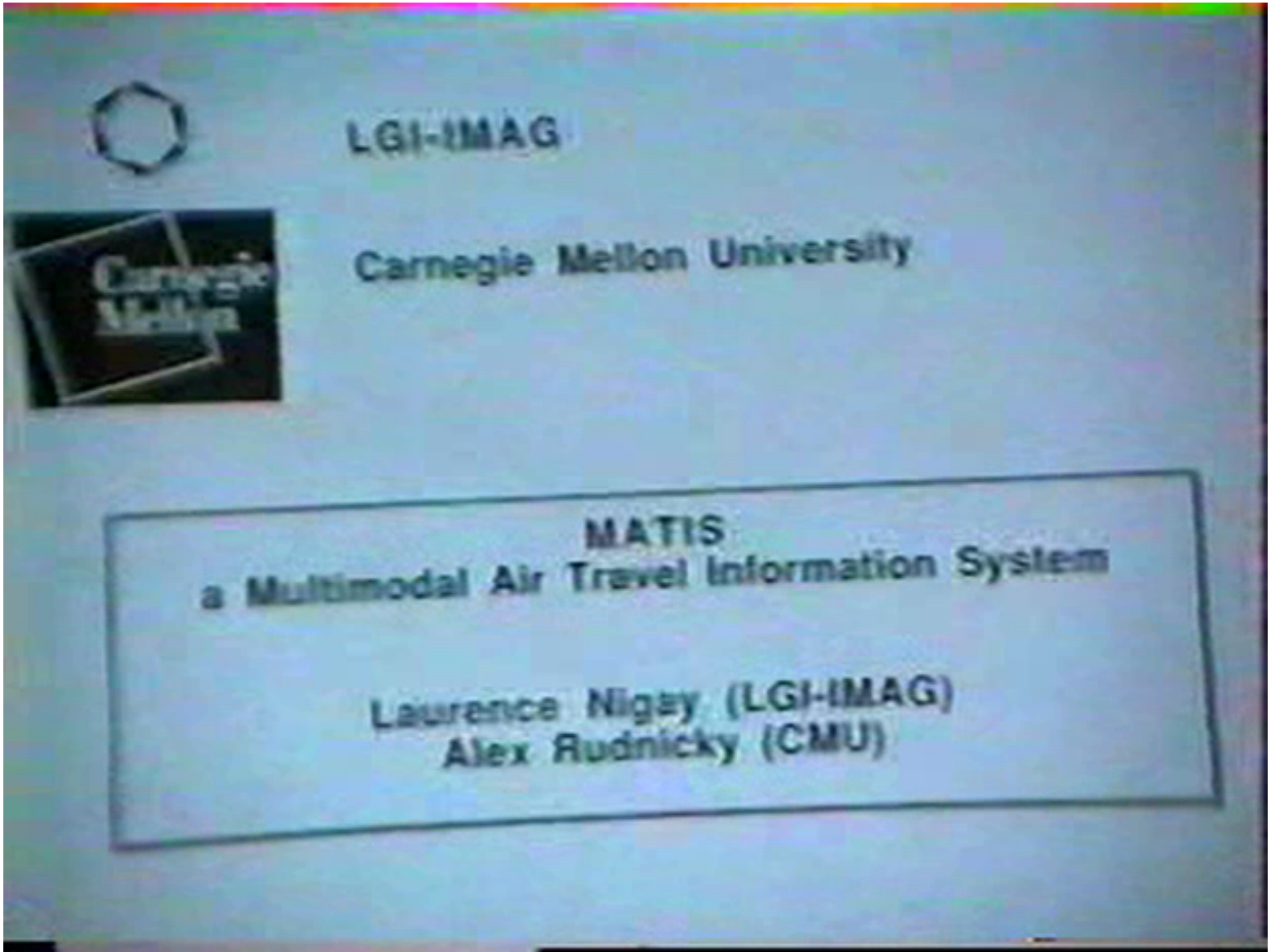
PAC Demo : Voice Paint – 1991



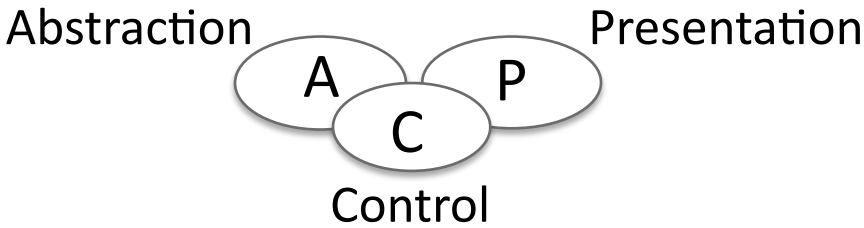
PAC*: Hierarchical Multi-modal Composition

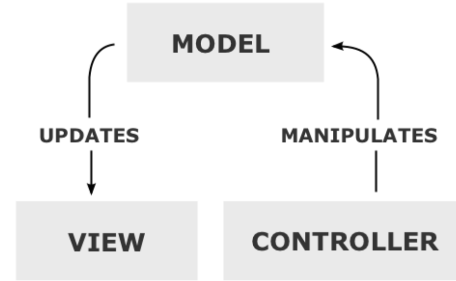


PAC Demo : MATIS (1993)
Multimodal Air Travel Information System



PAC vs MVC





MVC – Model View Controller (A. Goldberg) (Krasner 1988)

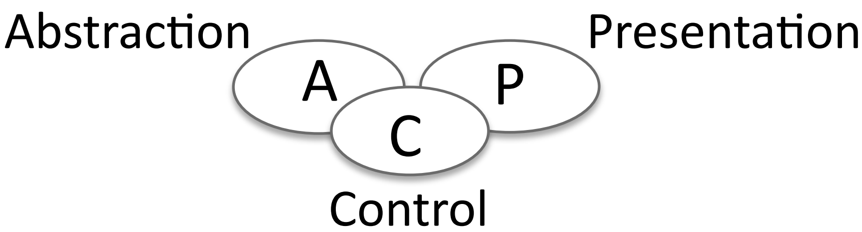
- Model: Dynamic Data Structure for the System
- View: Output display.
- Controller: Converts input to commands for model or view

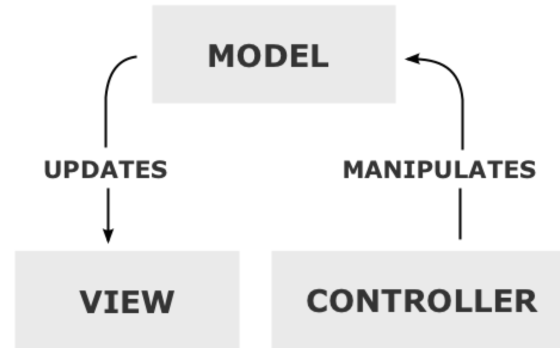
PAC – Presentation Abstraction Control (J. Coutaz 1987)

- Presentation: Input – Output Rendering
- Abstraction: Functional Model
- Control: Communication and Coordination.

*Krasner, G.E. and Pope, S.T.. A description of the model-view-controller user interface paradigm in the smalltalk-80 system. Journal of object oriented programming, 1(3), pp. 26-49. , 1988.

PAC vs MVC





Abstraction (PAC) = Model (MVC)
Presentation (PAC) = View + Controller (MVC)
Control (PAC) =? no equivalence in MVC.

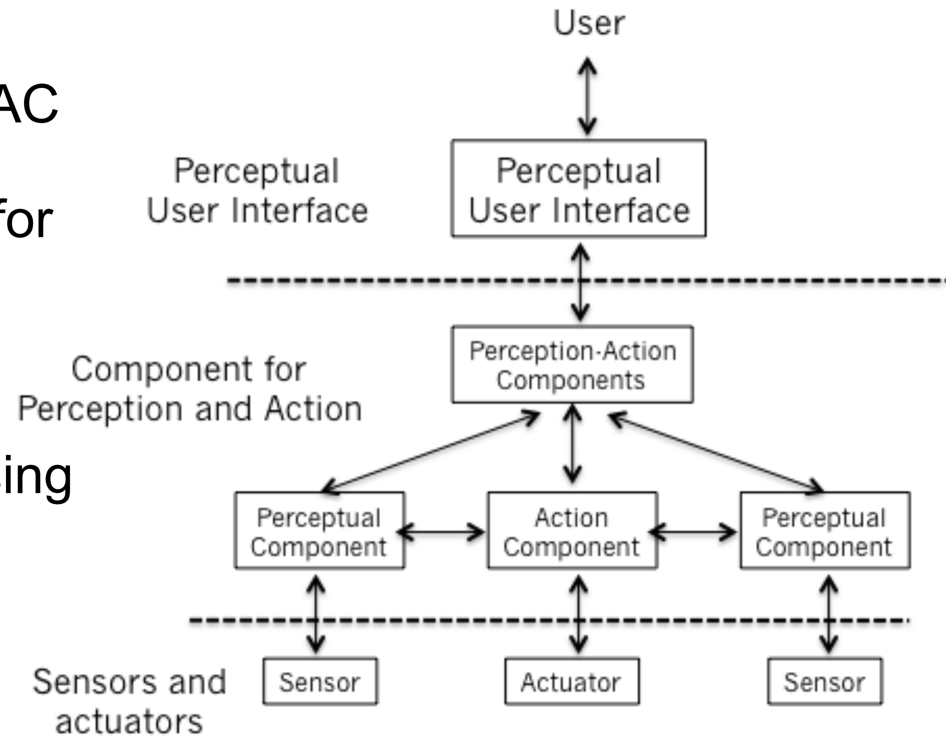
PAC enable hierarchical composition for input and output at multiple levels of abstraction and multiple time scales. (not MVC)

The PAC architecture is now often presented under the name PAC-MVC or new MVC and used for web programming, (Ruby, etc), Apple software (Finder, Carbon) and many other programming frameworks.

Perception for Interaction using PAC

During the 1990's, PRIMA used PAC to develop a series of multimodal perceptual user interfaces (PUIs) for interactive systems.

This lead to an approach to make these systems "Context aware" using Situation Modeling

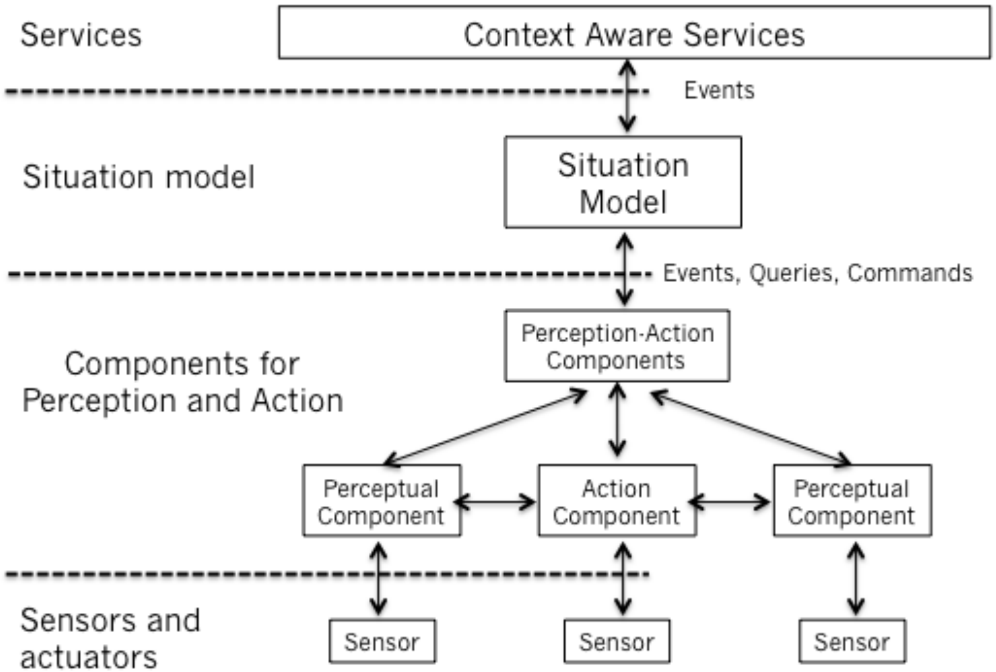


1) J. L. Crowley and J. Coutaz, "Vision for man-machine interaction", EHCI, Grand Targhee, Aug 1995.
2) J. Coutaz, F. Berard and J. L. Crowley, "Coordination of Perceptual Processes for Computer Mediated Communication", FG96 - International Workshop on Face and Gesture Recognition, Vermont, Oct 1996.
3) J. L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications", IEEE Conference on Computer Vision and Pattern Recognition, CVPR '97, St. Juan, Puerto Rico, June 1997.
4) J. L. Crowley, J. Coutaz and F. Berard, "Things that See: Machine Perception for Human Computer Interaction", *Communications of the A.C.M.*, Vol 43, No. 3, pp 54-64, March 2000

30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
    - Multimodal Context-Aware Services
    - Situation Modeling
    - Situated interaction with smart environments.
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from of Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

Context Aware Multimodal Services



1) J. L. Crowley, J. Coutaz, G. Rey and P. Reignier, "Perceptual Components for Context Aware Computing", UBICOMP 2002, *International Conference on Ubiquitous Computing*, Goteborg, Sweden, September 2002.
2) J Coutaz, J. L. Crowley, S. Dobson, and D. Garlan, "Context is Key", *Communications of the ACM*, Special issue on the Disappearing Computer, Vol 48, No 3, pp 49-53 March 2005.

(and about 30 other papers over the last 15 years)

Early Examples of Context Aware Systems

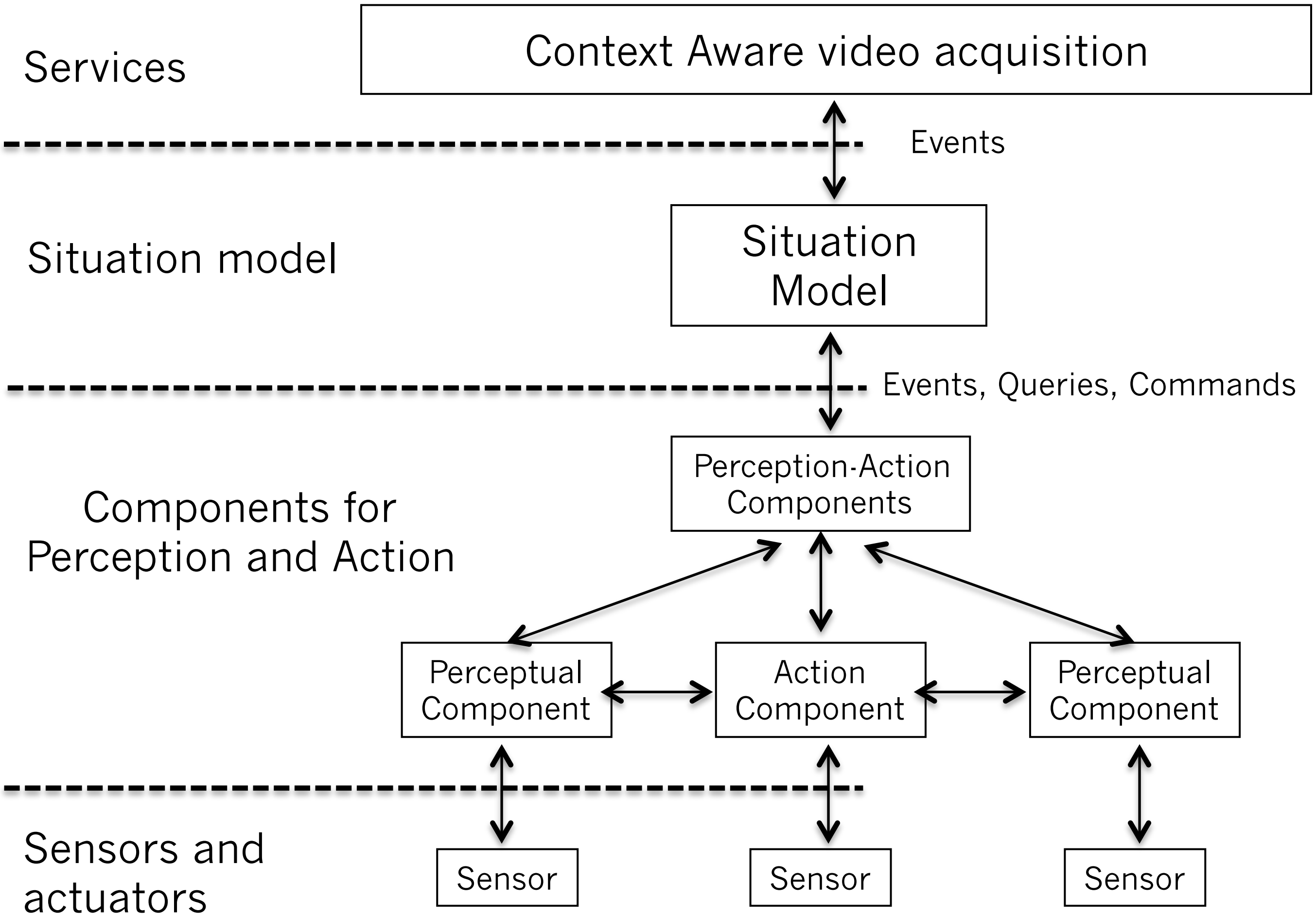
Early examples of situation aware systems

- Privacy filter for Media Space (2000)
- Lecture recording system (IST FAME)
- Activity monitoring for assisted living (ANR CASPER)

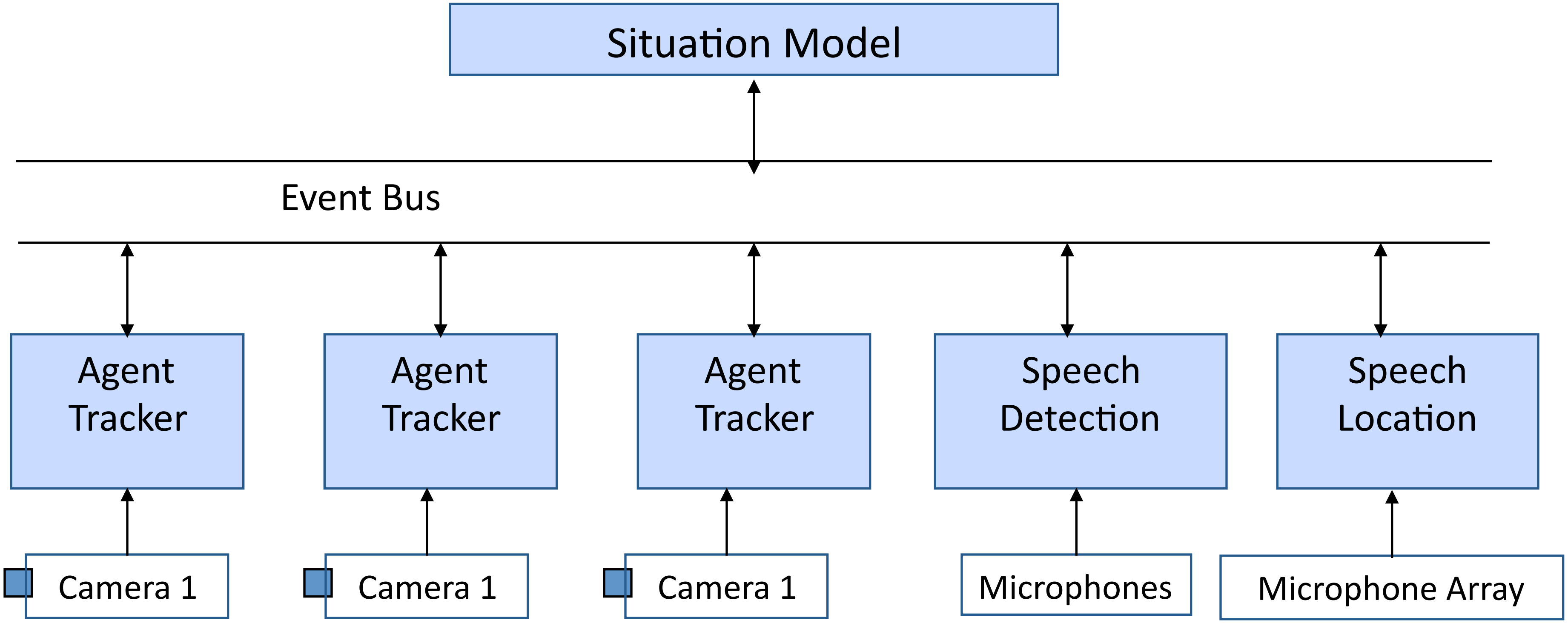
Examples constructed in IST CHIL (multi-modal services)

- Memory Jog (non-obtrusive memory prosthesis)
- Context aware Mobile Phone manager
- Meeting minute recording system

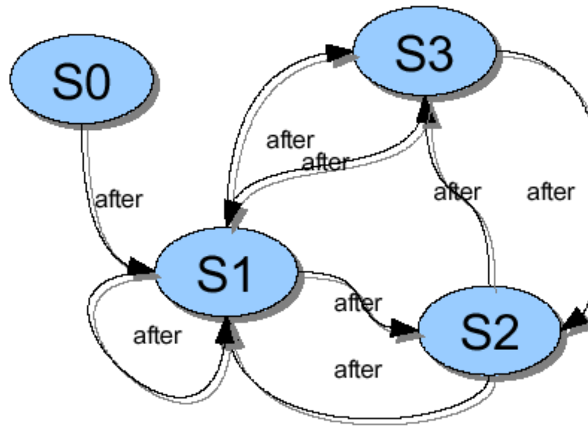
FAME Context Aware Video Acquisition System (2003)



Software Components



Situation Model (State Space)



<u>Situations</u>:

S_0	empty room	=> A_1
S_1	Speaker enters the room	=> A_1
S_2	Speaker speaks	=> A_2
S_3	Audience asks a question	=> A_3

Actions for the System

Recording camera and microphone:

A1 Record wide-angle view of the scene

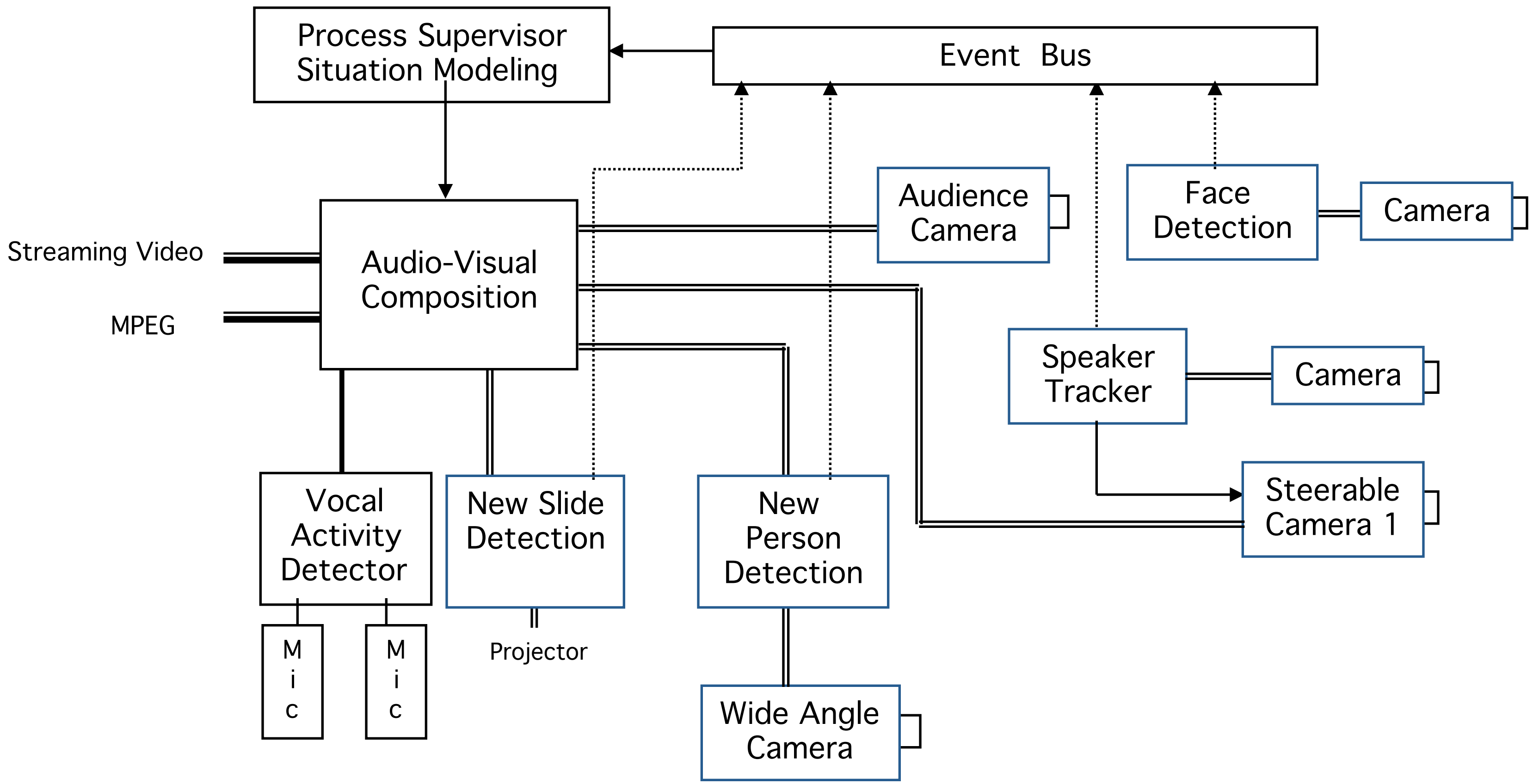
A2 Record the speaker

A3 Record the audience

Situations and behaviors were preprogrammed using a Graphical User Interface

Situation Aware Camera Man (IST FAME – 2003)

Video Acquisition System v2.0 (2006)



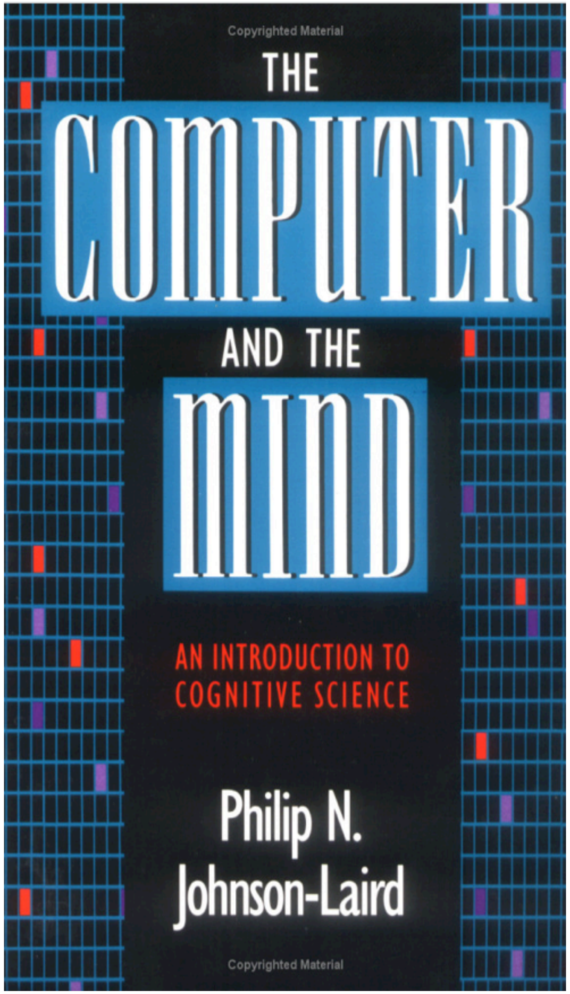
Context Aware Video Acquisition

Automatic Recording of 24 hours of InTech lecture Series at Inria – 2007

Situation Models (1987): Philip Johnson-Laird

Philip N. Johnson-Laird

PhD Psychology, 1967, University College London
Stuart Professor of Psychology at Princeton Univ.
1971-1973: Inst. of Advanced Study, Princeton U.
1973-1989: Laboratory of Exp. Psychology, Univ of Sussex
1989- Applied Psychology Unit, Princeton Univ.



Situation Models:
mental models for natural language and inference.

Johnson Laird proposed Situation Models as a framework to describe human abilities to

1) Provide context for story understanding
2) Interpret ambiguous or misleading perceptions.
3) Reason with default information
4) Focus attention for problem solving

Prima adapted situation models as as theory for construction of context aware systems and services.

Situation Models: as a theory for context aware services

A Situation model is a network of situations with associated behaviors

Situation: A set of relations between entities.

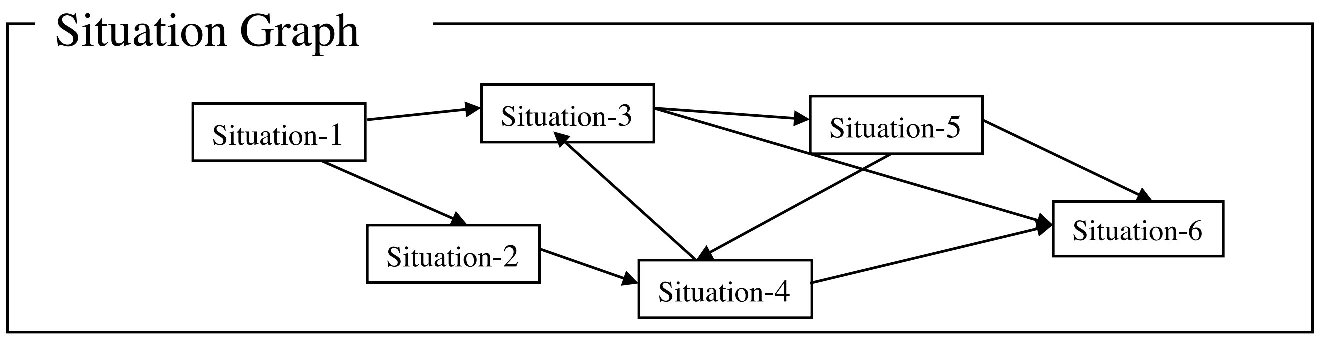
Entities: Any observable phenomena
Ex: People, things, times, places, events

Relations: Predicates. (spatial, temporal, etc). Relations associate entities

Behaviors: Event-Condition-Action rules associated with a situation.

- Behaviors control perception, action, interaction, and reasoning
- Behaviors can be programmed or learned for each situation.

Situation Models: as a theory for context aware services



<u>Situation Graph</u>: A network of situations with transition conditions

- The situation controls attention (entities and relations to observe)
- Behaviors for each situation specify actions and interaction

A <u>Situation Model</u> specifies the set of entities, relations, situations, state transitions and behaviors for a context.

Examples of Service Constructed with Situation Modeling

Examples constructed in IST CHIL (multi-modal services)

- Memory Jog (non-obtrusive memory prosthesis)
- Context aware Mobile Phone manager
- Meeting minute recording system

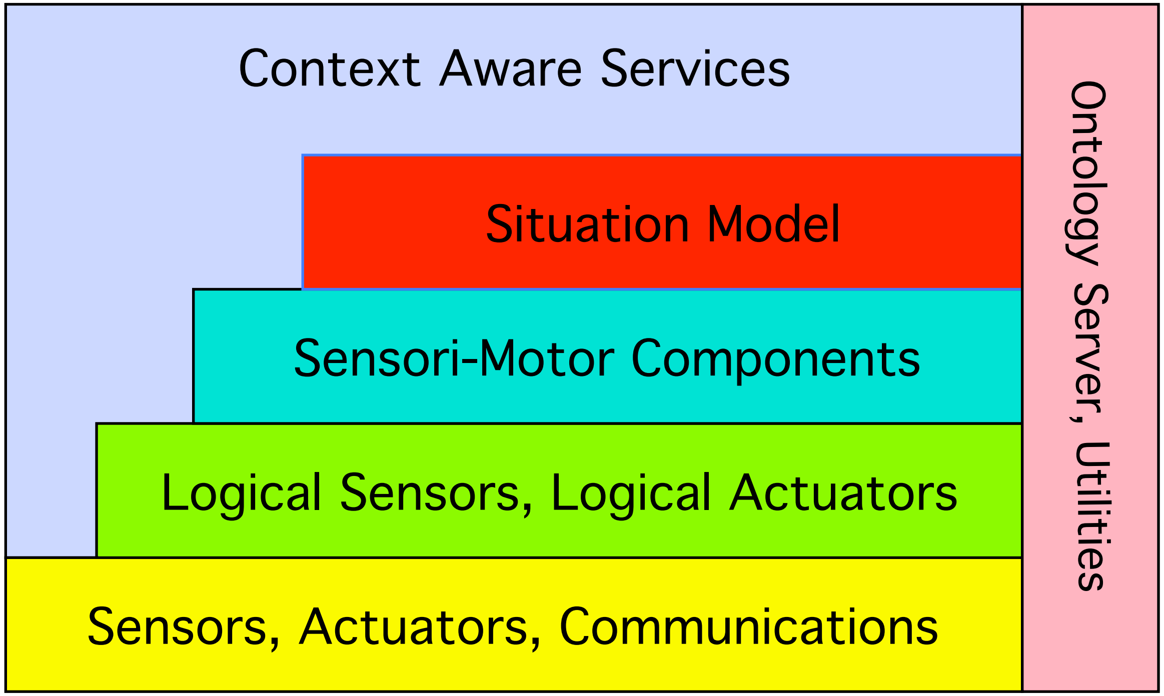
Smart Environments

- Project Casper: Monitoring for autonomous ageing
- Project Cont'act: Interactive environment

Robotics:

- Polite, social interaction with social robots

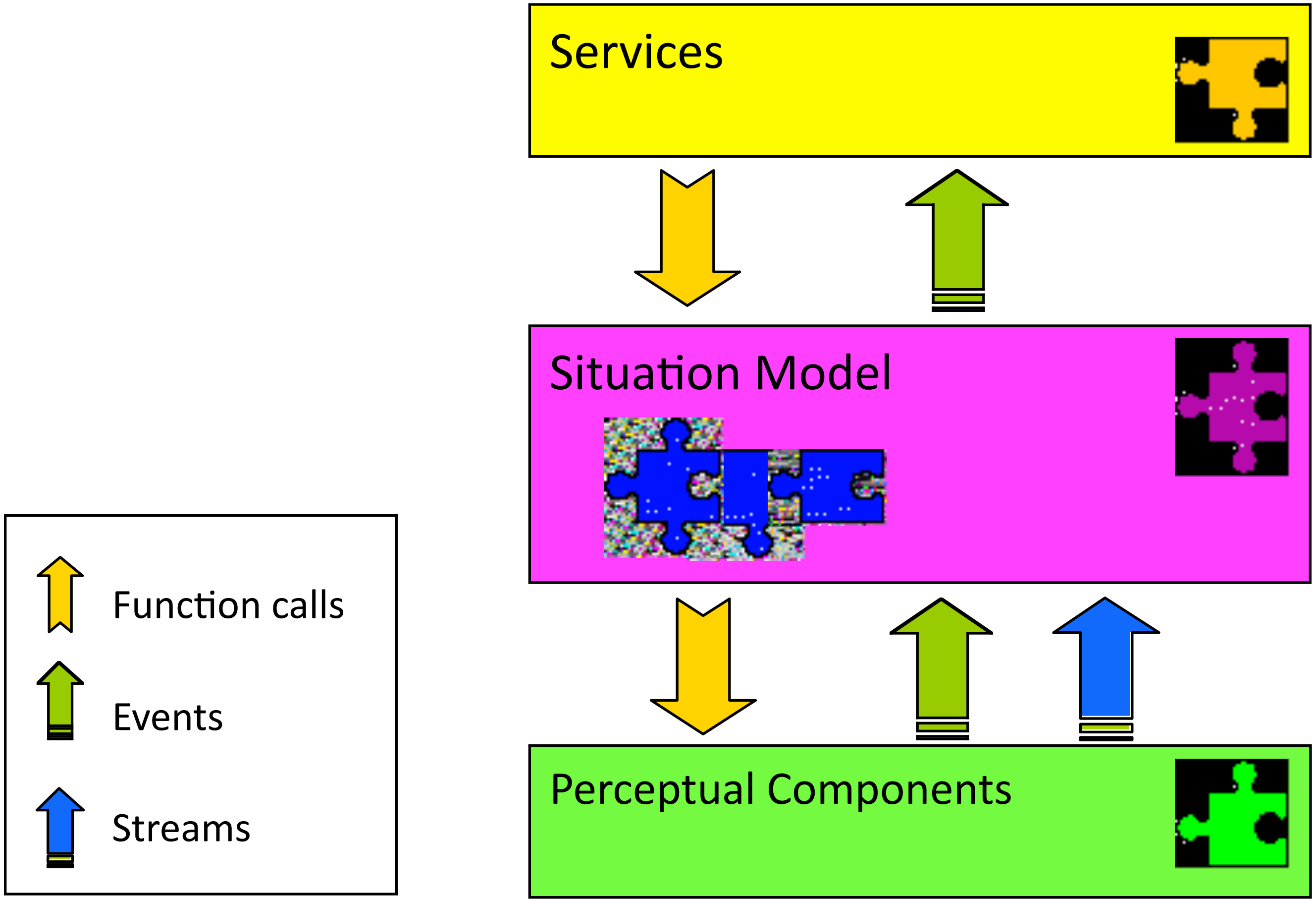
Project IST CHIL (2005 – 2009)
Computers in the Human Interaction Loop.

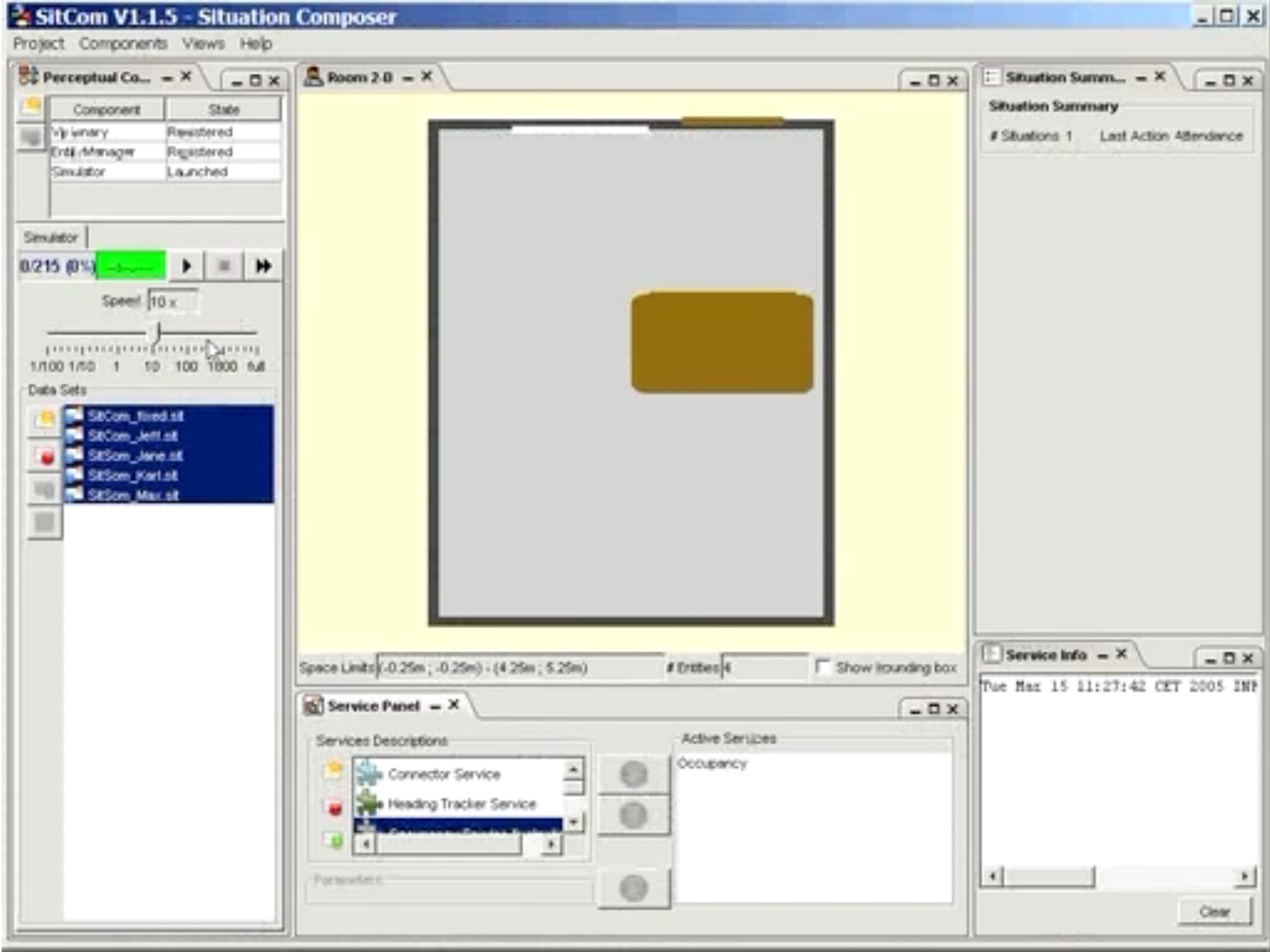


1. O. Brdiczka, J. L. Crowley, J. Curin, J. Kleindienst, "Situation modeling", in Computers in the Human Interaction Loop, A. Waibel (Ed), Springer Verlag no 12, p. 121-132, 2009.

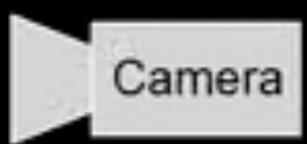
2. J. Soldatos, I Pandis, K Stamatis, L Polymenakos, JL Crowley, "Agent based middleware infrastructure for autonomous context-aware ubiquitous computing services", Computer Communications, Vol 30, Issue 3, Pages 577-591, February 2007.

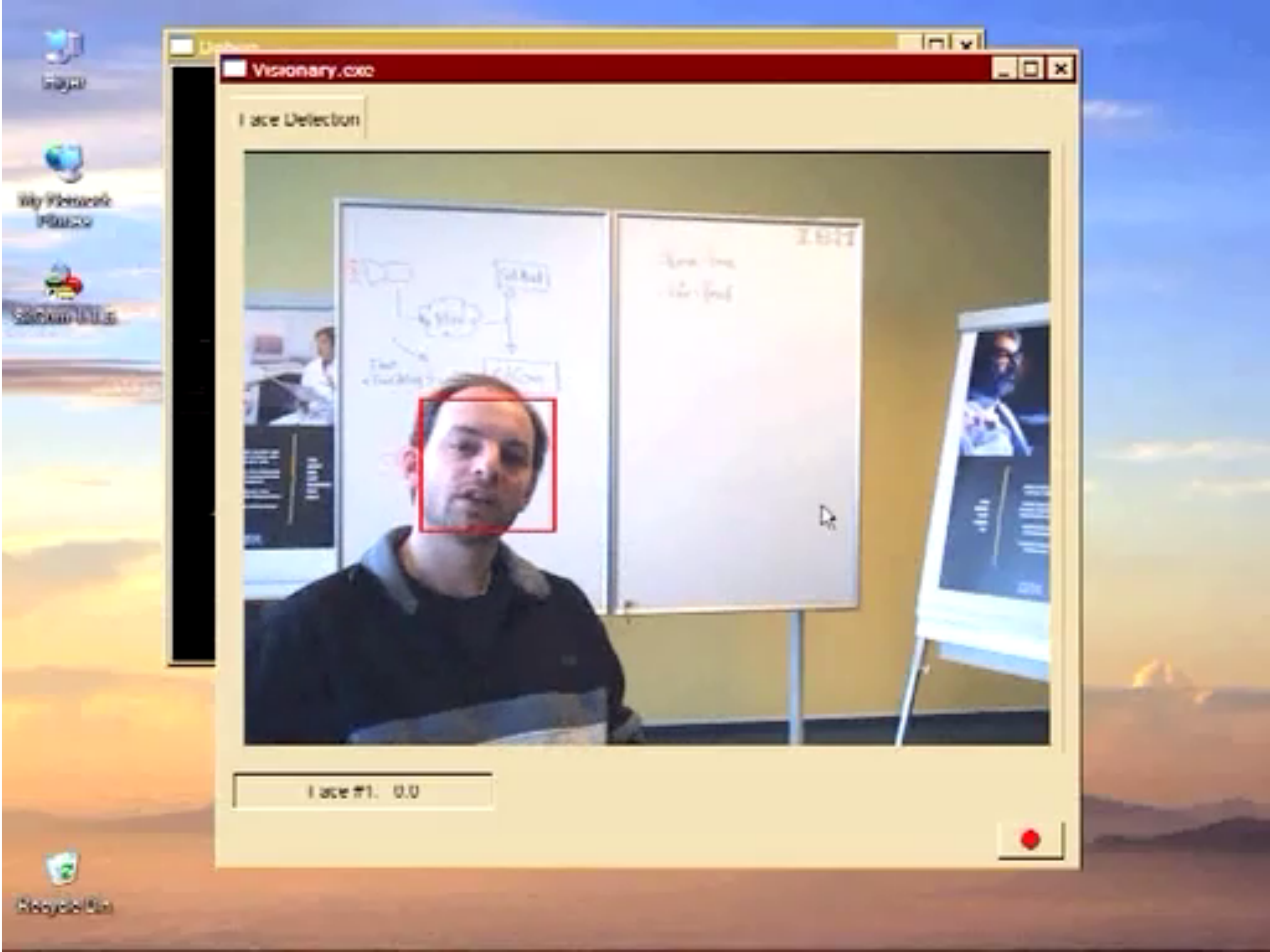
IST CHIL Core: Situation Model



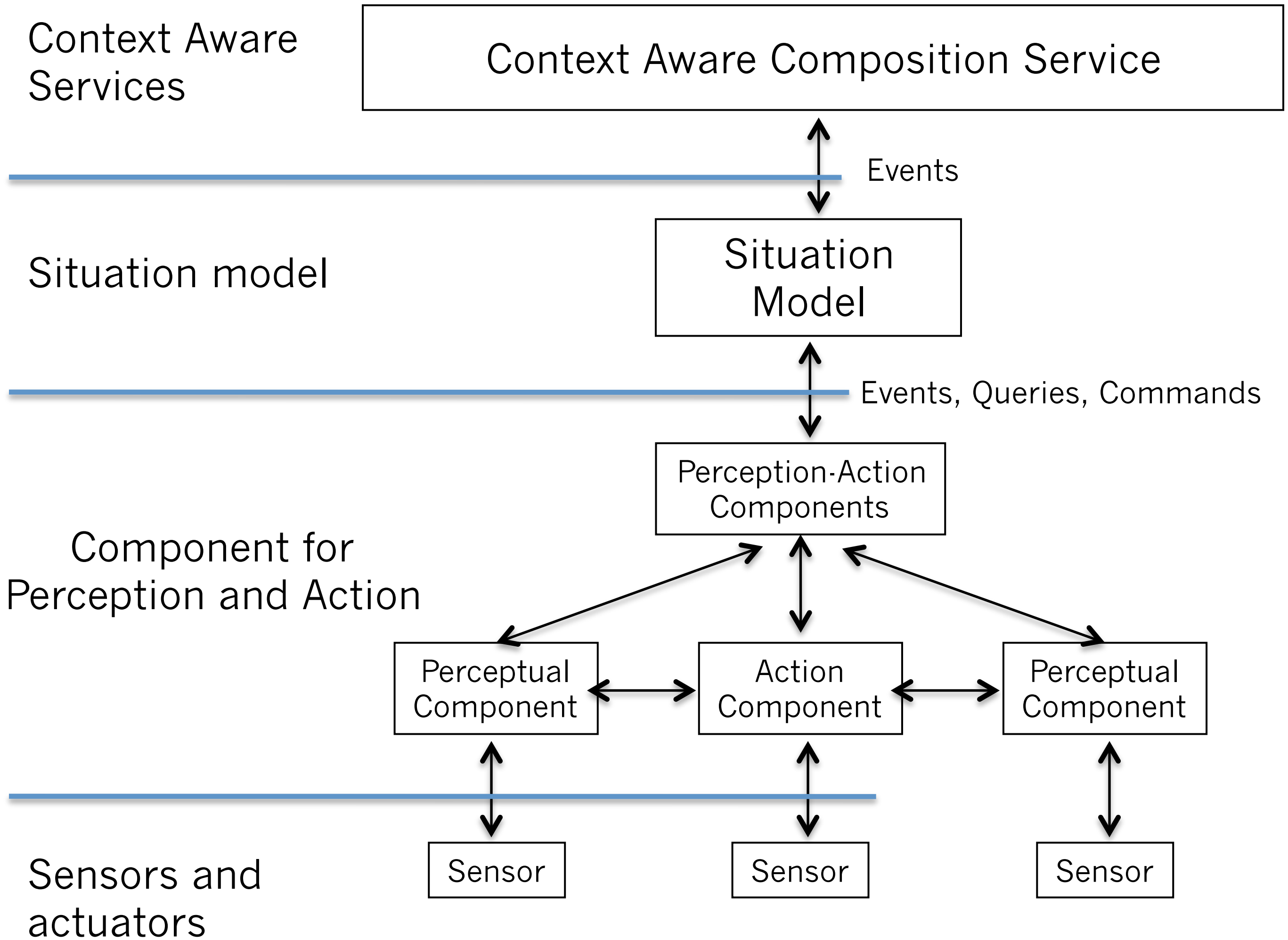

SitCom V1.1.5 - Situation Composer
Project Components Views Help
Perceptual Co...
Component State
Registered
Registered
Simulator Launched
Simulator
Speed 10 x
1/100 1/10 1 10 100 1000 full
Data Sets
SitCom_Jeff.sit
Room 2-B
Situation Summ...
Situation Summary
Situations 1 Last Action Attendance
Space Limits (-0.25m ; -0.25m) - (4.25m ; 5.25m) # Entities 4 Show bounding box
Service Panel
Services Descriptions
Connector Service
Heading Tracker Service
Active Services
Occupancy
Parameters
Service Info
Tue Mar 15 11:27:42 CET 2005
Clear

Using Real Perceptual Components




Visionary.exe
Face Detection
Face #1: 0.0

Cont’act:
Context Aware Interactive Environment



Application to Robotics: Training Aibo to be polite

John is on the phone ⟷ John is facing Mary ⟷ John is talking to Aibo

John is on the phone	John is facing Mary	John is talking to Aibo
AIBO Behaviours	AIBO Behaviours	AIBO Behaviours
Be Quite (10)	Be Quite (5)	Be Quite (-10)
Play (-5)	Play (0)	Play (5)
Talk to John (-10)	Talk to John (-5)	Talk to John (10)
Dance (-10)	Dance (-10)	Dance (5)
Sing (-20)	Sing (-20)	Sing (-5)

Learning for Situated Services:

1) Learn to identify relevant entities and relations (Brdiczka et al 06)
2) Learn network of situations for a context (Zaidenberg et al 06)
3) Learn appropriate behaviors for each situation (Barraquand 08)

R. Barraquand, J. L. Crowley, "Learning Polite Behavior with Situation Models", Third International Conference on Human Robot Interaction (HRI 2008), 12-15 March 2008, Amsterdam, The Netherlands

Recent Examples of Applications

1) Video Surveillance (Startup BlueEye Video – 2003)
2) Customer monitoring (Start up: HiLabs - 2008)
3) Public Event Recording (Startup MeanInFull – 2014)
4) Context aware mobile applications (Start up: Situ8ed 2015)
5) Multi-modal Observation of Kitchen Activities (Thesis N. Aboubakr)
6) Multi-modal observation of chess experts (Thesis T. Guntz)

Lino, the user interface robot (2003)

(used with Permission from B. Krose).

Kröse, B.J., Porta, J.M., van Breemen, A.J., Crucq, K., Nuttin, M. and Demeester, E., 2003, November. Lino, the user-interface robot. In *European Symposium on Ambient Intelligence* (pp. 264-274). Springer, Berlin, Heidelberg.

Categories of Context Aware Services

Categories for context aware services can be defined by the nature of interaction with users. (Crowley-Coutaz 2015).

Examples: Tools, Advisors, Media, Affectors…

J. L. Crowley and J. Coutaz, "An Ecological View of Smart Home Technologies", 2015 European Conference on Ambient Intelligence, AMI 2015, Athens, Nov. 2015

Categories of Intelligent Services

Tools: A service used to achieve a goal. The behavior of a tool should be reliable, predictable and robust to environmental conditions.
Example: smart thermostat

Advisors: Observe users and their activities in order to propose possible courses of actions. Should be completely obedient. Should not take initiatives or create unwanted distractions (nag-ware).
Examples: GPS Navigation system giving route advice

Categories of Intelligent Services

Media: Extensions to human perception and experience, for entertainment, communications, and display of information. Can be interactive or simply peripheral, and ideally should provide a sense of immersion.
Examples: Ambient Orb (Rose 14)

Affectors: Services that inspire affection. Affectors can help compensate for a loss of social contact that can result from ageing or hospitalization.
Examples: Nabastag, Paro affective Robot, Nao, Jibo,…

30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
  - Perception-Action-Cognition-Emotion
  - Multi-modal Perception vs Multimodal Interaction
  - The complementary Nature of Cognition and Emotion
- Modeling Comprehension from of Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

What is a modality?

Modality: A channels for sensing and action (including communications).

Examples of modalities:

Sensing: Vision, audition, taste, …

Action: Manipulation, Locomotion, Speech,

Actions can be semiotic (for communications), systemic (for perception) or ergotic (to affect the environment).

Perception, Action, Cognition and Emotion

Perception: Interpretation of sensing through recognition, action, emotion, cognition

Action: Intentional movement to communicate, to sense, or to affect the environment.

Cognition: conscious abstract reasoning to understand phenomena and plan actions.

Emotion: Intuitive (somatic) reaction to a situation that provides rapid response.

P PERCEPTION

A ACTION

C COGNITION

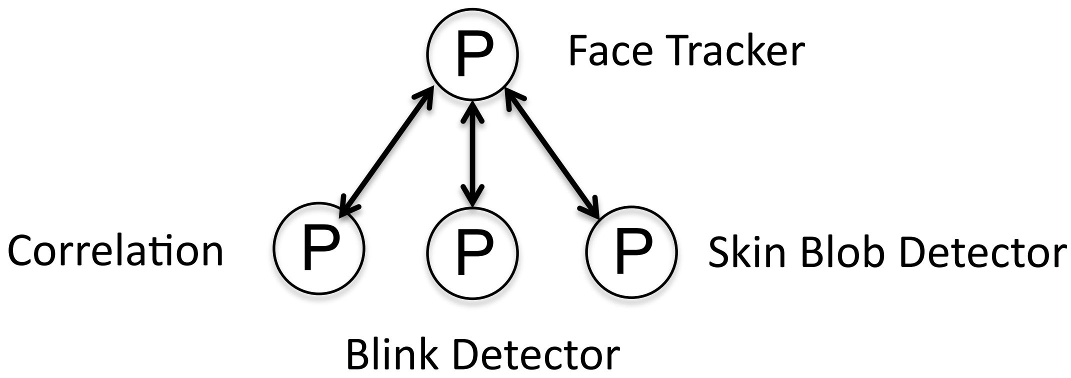
E EMOTION

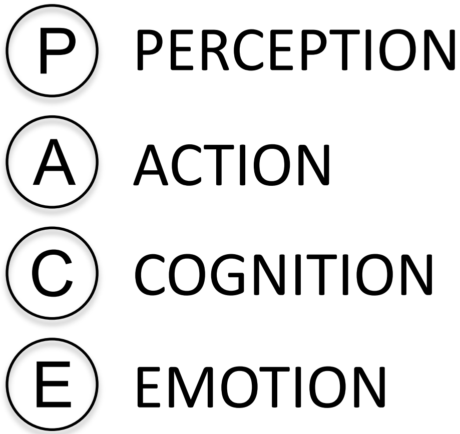
Emotion and cognition guide action and perception.

Hypothesis: Emotion guides cognition.

Multi-modal Perception vs Multi-modal Interaction

Example of <u>Multi-modal Perception</u>:
Face detection and tracking from multiple modalities

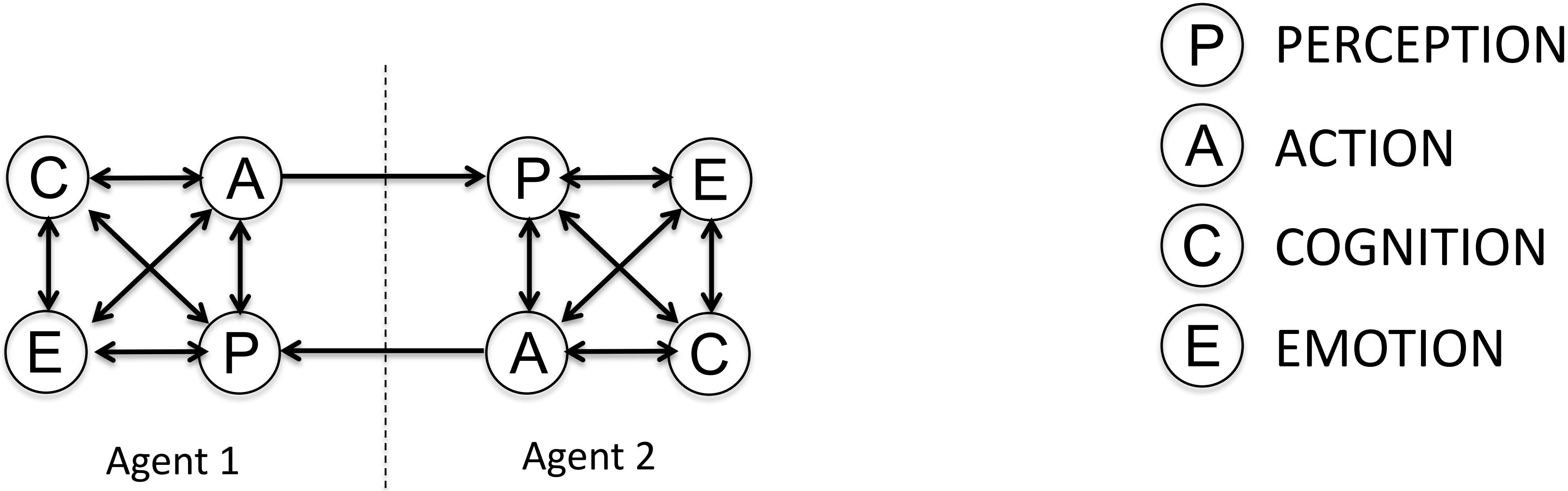




J. L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications", *IEEE Conference on Computer Vision and Pattern Recognition,* CVPR '97, St. Juan, Puerto Rico, June 1997.

Multi-modal Perception vs Multi-modal Interaction

Multi-modal Interaction:



Multi-modal Interaction: Tightly coupled perception-action between two natural or artificial systems.

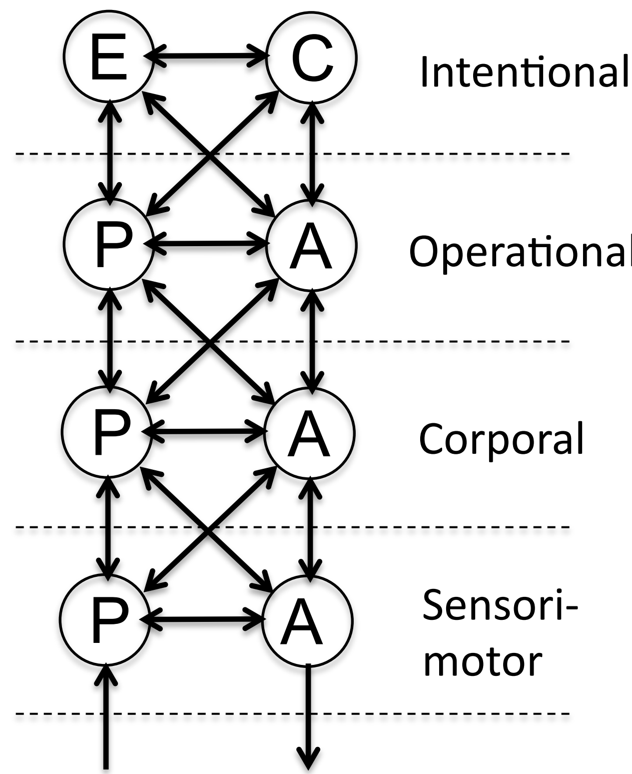
PACE is hierarchical

Intentional: Determines goals and behaviors

Operational: Rapid, automatic, skill level, interaction of perception and action.

Corporal: Controls Movements of the body and its relation to the environment.

Sensori-motor: Sensor and motor signals. (images, sounds, tactile maps…)



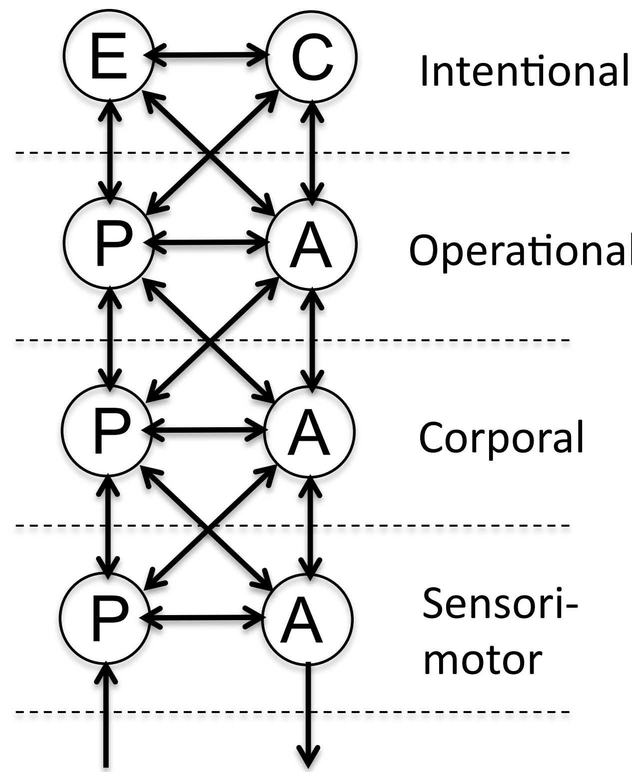
What is the relevance of such a model in the epoch of Deep Learning?

An architectural model is independent of the mechanism used for implementation.

The model describes “what” not “how”.

Models are used to predict and explain. (Kuhn 62).

An architecture can be implemented using classic programming OR using Neural networks.

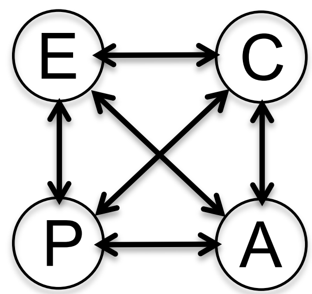


Emotion and Cognition:
Two Complementary Intentional Systems

Emotion:
Fast, Reactive, Predictable.

Enables rapid reaction to threats and opportunities.

Kahneman's System 1?



Cognition:
Slow, Deliberative, Creative

Enables planning, explanation, prediction and understanding.

Kahneman's System 2?

Kahneman D, Egan P. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux; 2011

ANR - CEEGE

Chess Expertise from Eye Gaze and Emotion

James L. Crowley,

Prof. Grenoble INP, Univ. Grenoble Alpes

LIG, CR INRIA GRA

Prof. Thomas Schack

Dept of NeuroCognition

University of Bielefeld

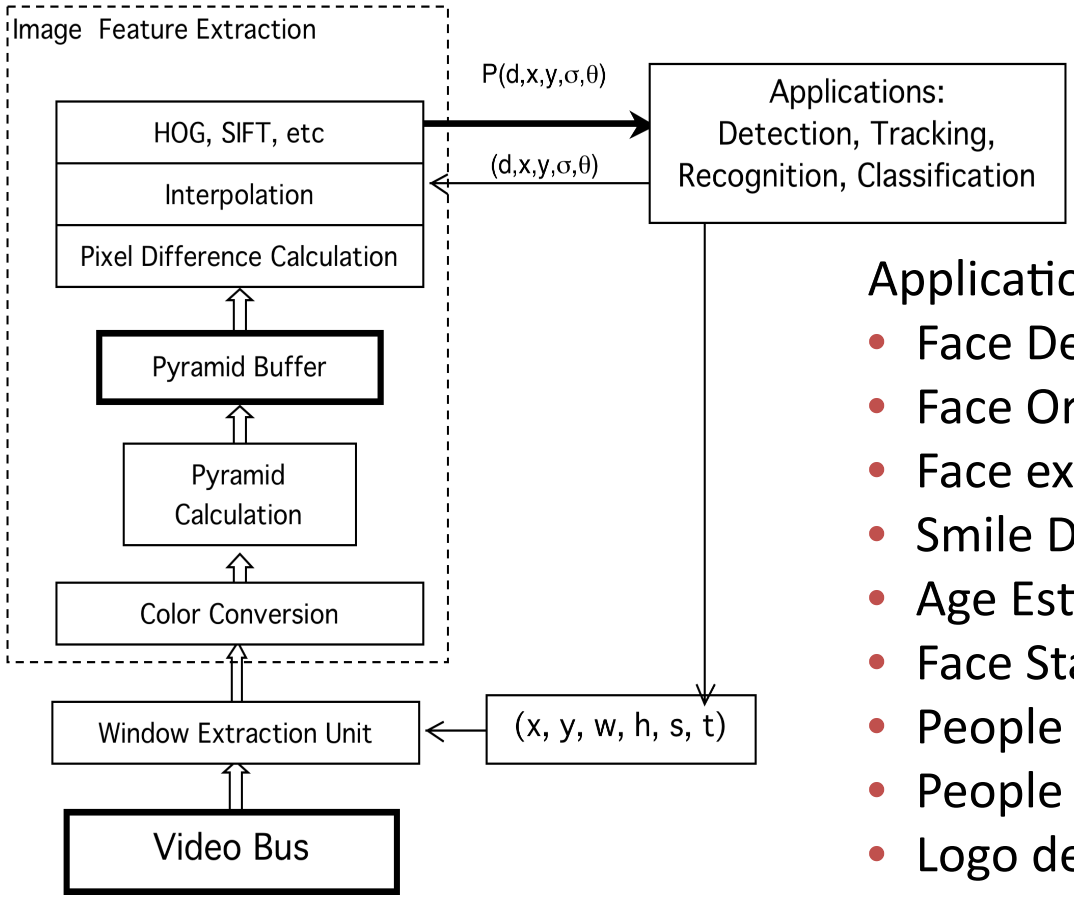
CEEGE: Research Questions

1) What are the most effective techniques to observe and model the emotions of subjects engaged in solving problems?

2) Is it possible to use eye-gaze and emotion to discover and model the understanding and reasoning of a person engaged in solving problems?

3) Are techniques for deep learning more effective than traditional cognitive science for modeling the understanding and predicting the actions of subjects?

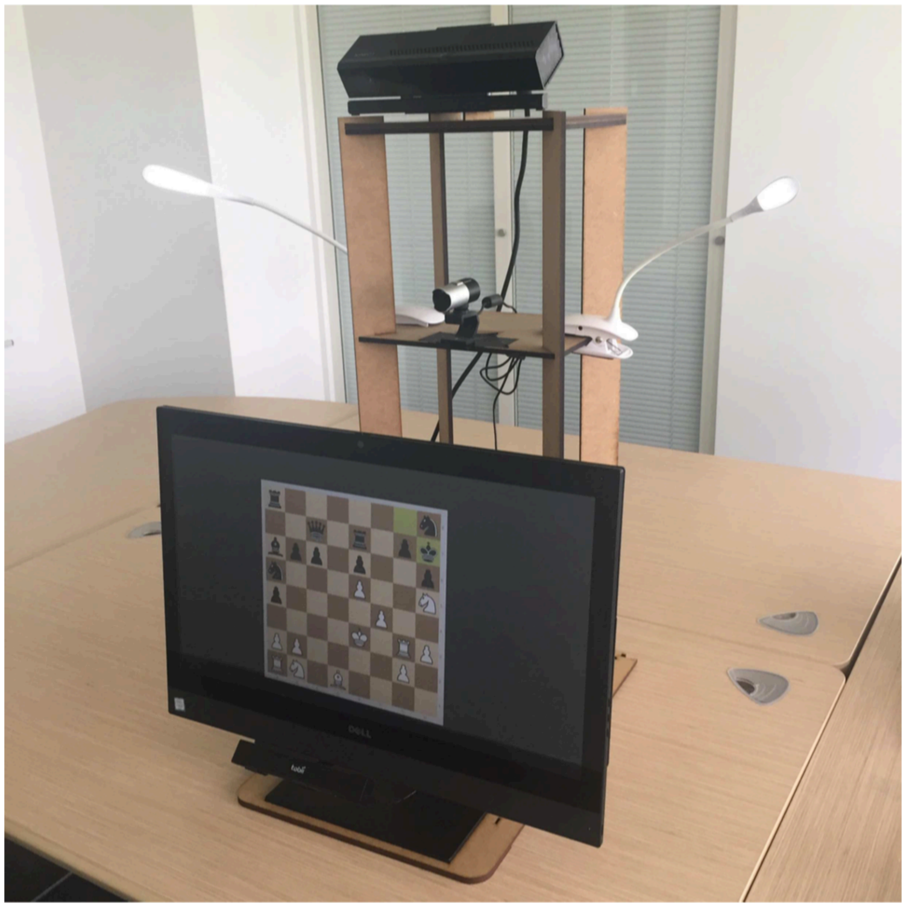
Prima CV (2012)
Real Time Vision Library for Mobile Devices



Applications include:

- Face Detection and tracking
- Face Orientation
- Face expression recognition
- Smile Detection
- Age Estimation
- Face Stabilization
- People Counting
- People tracking
- Logo detection and recognition
- Text detection

CEEGE – Research Instrument



Sensors:

- Interactive Touch-Screens
- Kinect 2.0
- HD Webcam 1080p
- Eyetracker bar (Tobii & Fovio)
- Intensity light control

Software:

- Open Pose (body skeleton)
- Open Face (Emotions)
- Eye-works Fixation, pupil dilation)
- RGBD Sync – In-house synchronous Multimodal recording

Observing Fixation and Attention with a Remote Eye-Tracker

Remote Eye Trackers (Tobii, Fovio) provide real tracking of gaze and fixation.


1488550163.197
lichess.org
Valence
−1
0
1

Body Posture: Kinect, Open Pose

FACS: Facial Action Coding System

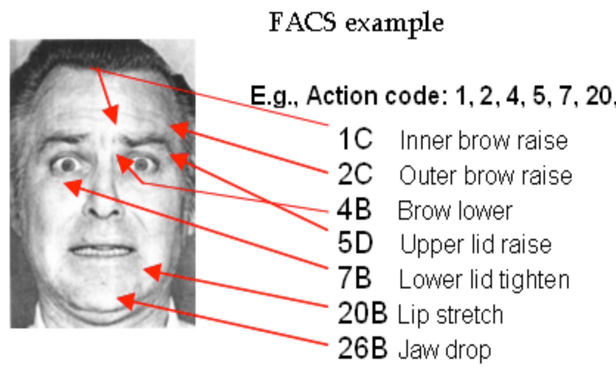


Image provided by UCSD Machine Perception Lab

Facial Action Coding System (FACS) : A system to label human facial expressions, developed by P. Ekman and W. V. Friesen, 1978)

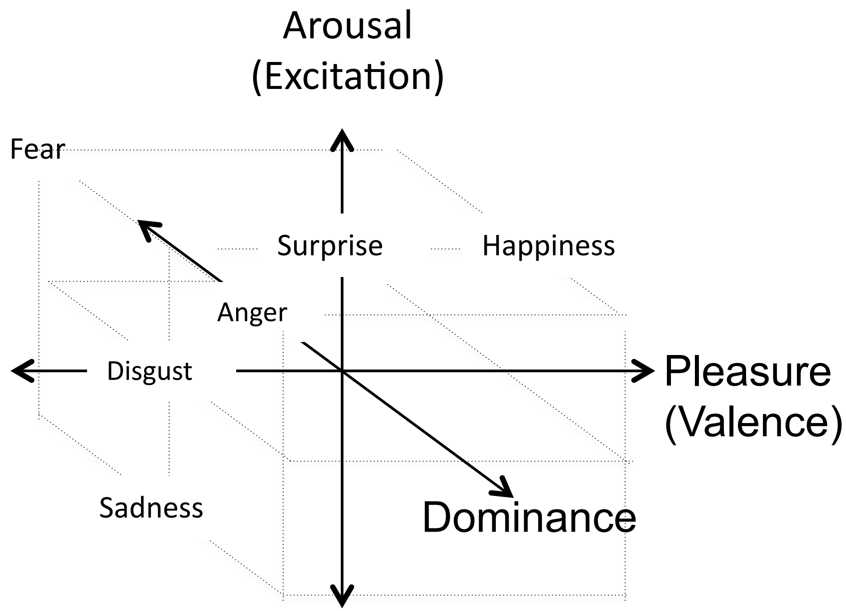
A common standard for recognizing facial expression of emotions

Available in several commercial and Open Source software Packages.

Ekman’s Six Basic Emotions

Emotion	Action Units
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26
Fear	1+2+4+5+20+26
Anger	4+5+7+23
Disgust	9+15+16

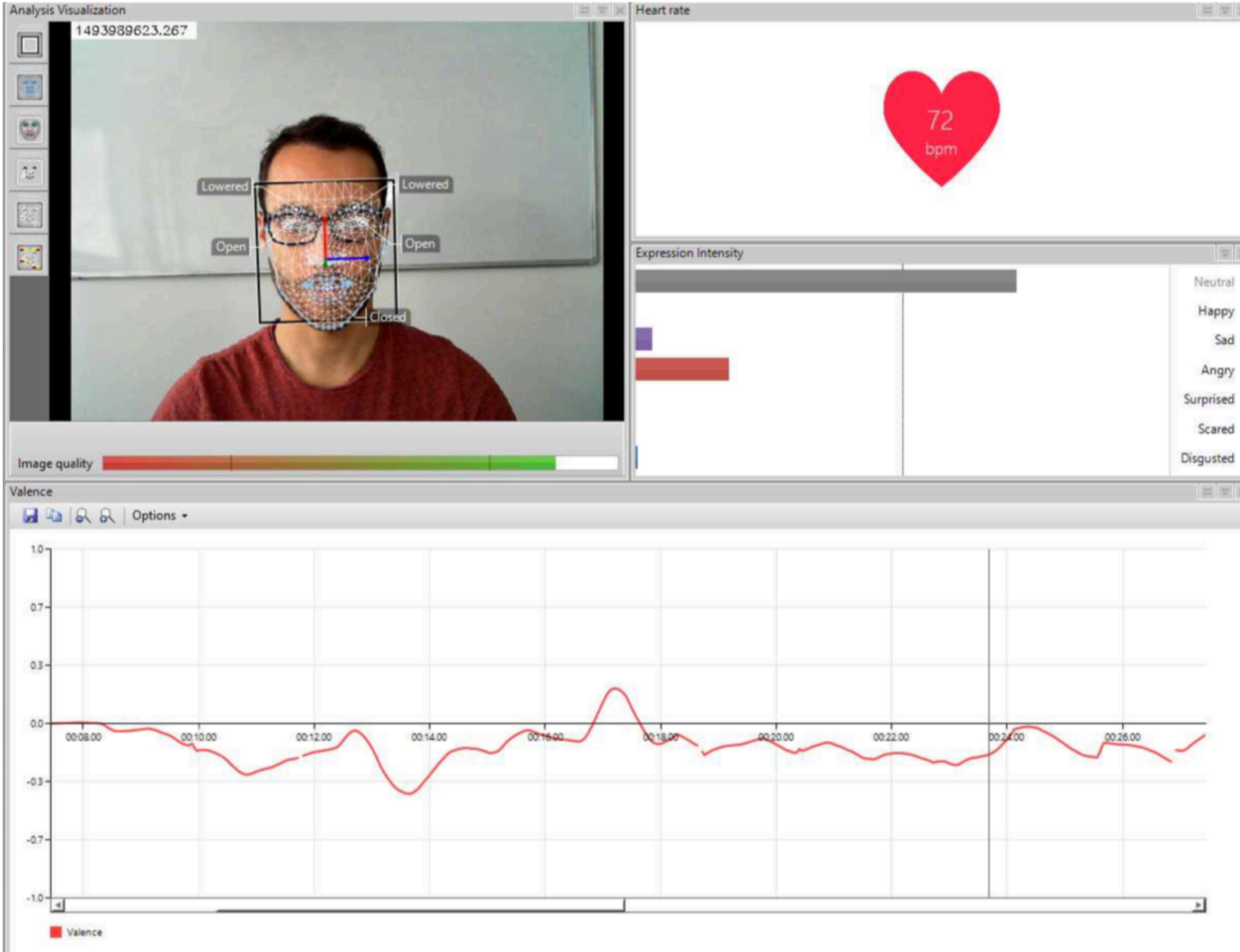
The PAD Model for Emotions



The PAD (Pleasure, Arousal, Dominance) model

1) Pleasure – Displeasure: Valence of an emotion
2) Arousal – Calm: Intensity, physiological excitation
3) Dominance – Submissive: Disposition to assert control.

J. A. Russell , A. Mehrabian, "Evidence for a three-factor theory of emotions", Journal of Research in Personality Vol. 11(3), pp 273-294, Sept 1977.


Analysis Visualization
1493989623.267
Lowered
Lowered
Open
Open
Closed
Image quality
Heart rate
72
bpm
Expression Intensity
Neutral
Happy
Sad
Angry
Surprised
Scared
Disgusted
Valence
Options
1.0
0.7
0.3
0.0
-0.3
-0.7
-1.0
00:08.00
00:10.00
00:12.00
00:14.00
00:16.00
00:18.00
00:20.00
00:22.00
00:24.00
00:26.00
Valence

Fixation and attention

Fixation can be used to measure attention and to predict next move.
What can they tell us about comprehension of the game situation?

J. Le Leoudec, T. Guntz, D. Vaufreydaz and J. L. Crowley, Deep Learning Investigation for chess player attention, prediction using eye-tracking, and game data, Submitted to 2019 ACM Symposium on Eye Tracking Research & Applications, ETRA'2019, Denver, Colorado (under review).

Pilot Study – First experiment – Oct 2017

Objectives:

1. Verify experimental equipment
2. Verify that eye-gaze and emotion correlate to expertise.

Task: 6 time limited tasks of increasing difficulty (Mate in N).

Measurements:

Eye-gaze (Tobii remote), posture (Kinect), Ekman 7 Emotions (Face Reader).

Hypothesis:

- Players would display concentration during problem solving, frustration if unable to solve problem, and pleasure when finding problem solution

Experiments:

Session 1: 21 subjects (14 recordings usable)
Session 2: 9 subjects (8 recording usable).

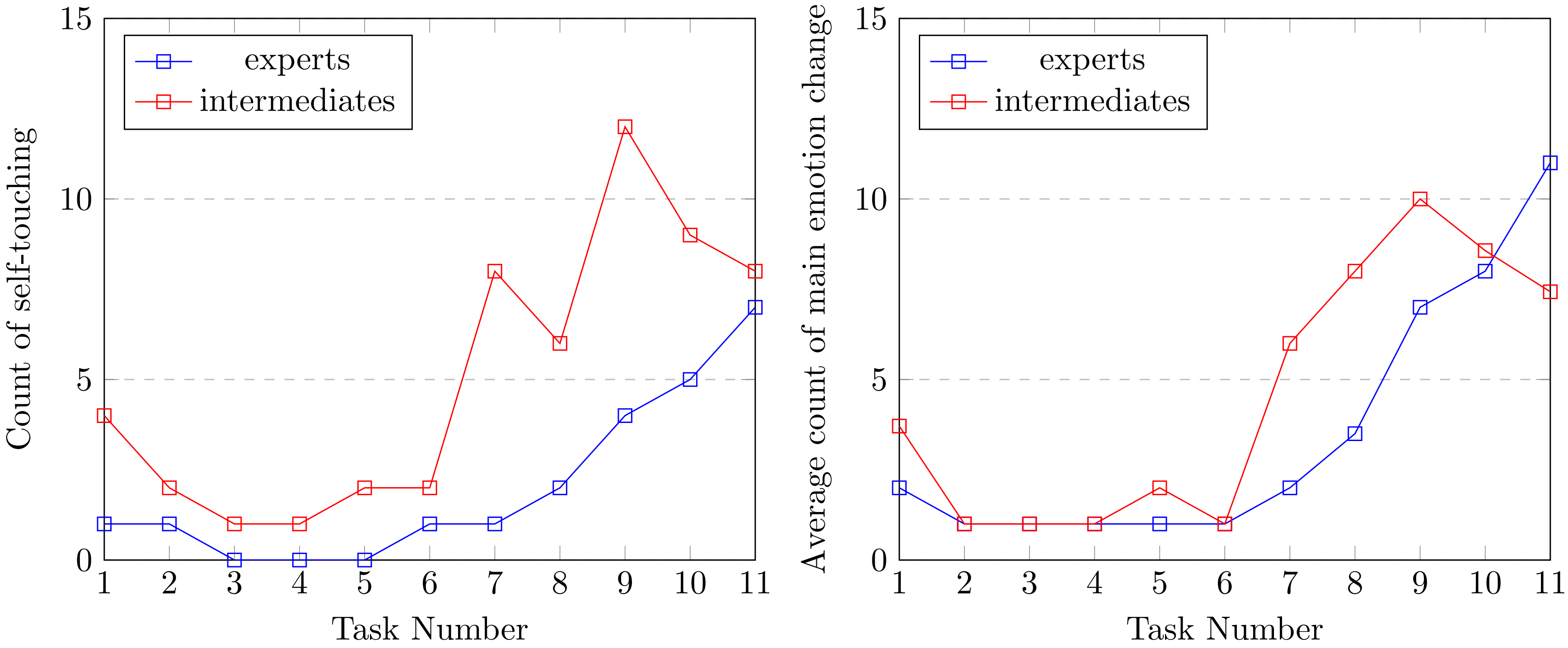
Pilot Study – First experiment – Oct 2017

Results:

1) 22 useful recordings (9 expert, 13 Intermediate).
2) Surprising result: Self touching and rate of change in emotion increased from a neutral emotion during reactive play of easy problem to frequent touching and rapid changes in emotion as the problems became more challenging.

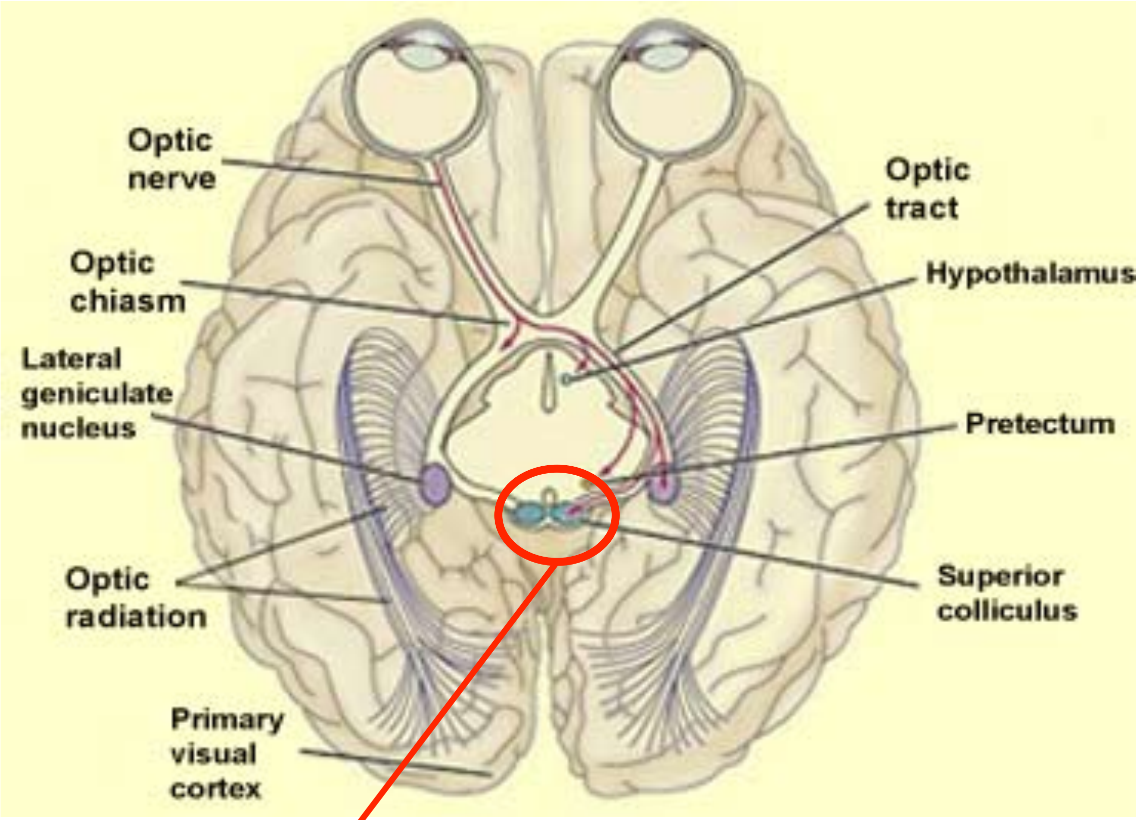
WHY !?

Physiological reaction to problem difficulty



Number of self-touches (left) and number of changes in emotion (right) for intermediates and experts engaged in solving the 11 problems.

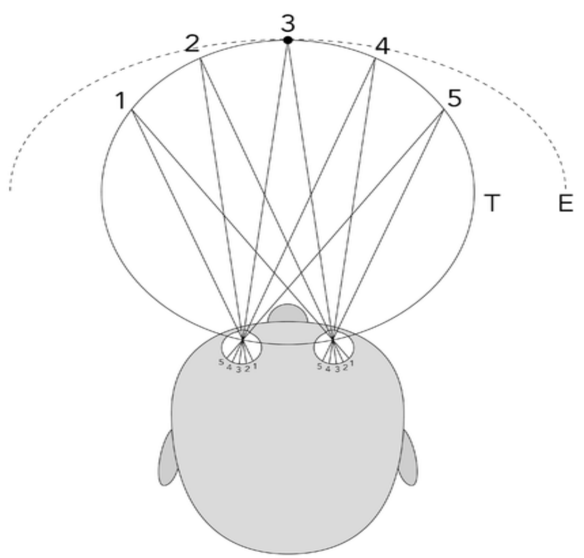
The Physiology of Fixation and Attention

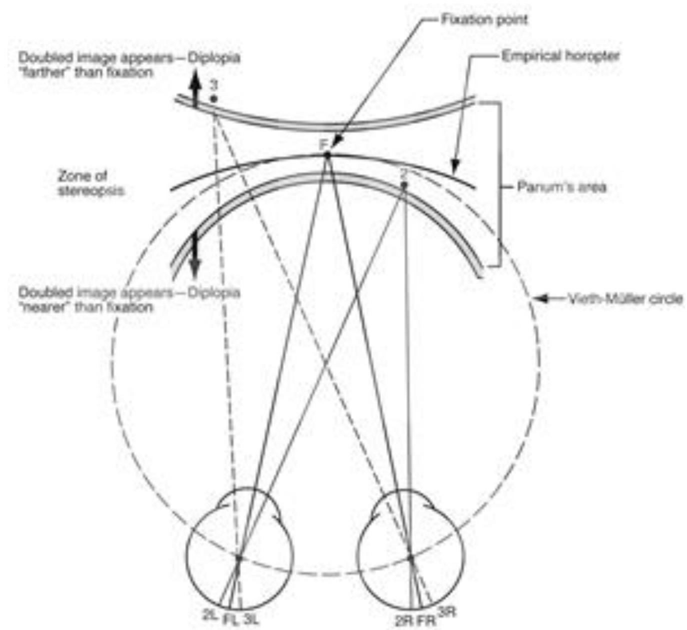


Superior Colliculus:

- 7 Layer filter with input different brain regions
- Controls vergence, version and cyclotorsion.

The Superior Colliculus Controls the Horopter



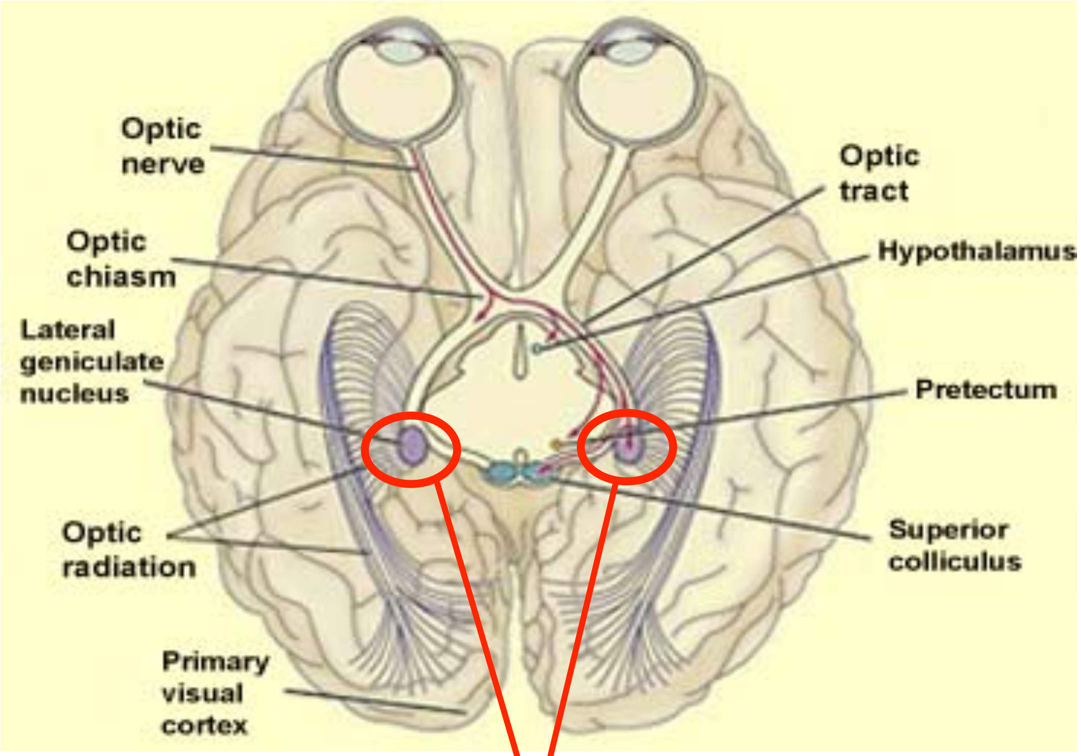


<u>The Horopter</u>: The Locus of Fixation.

Points in space that project to the same position in both retina.

Visual stimuli outside the horopoter are un-attended.

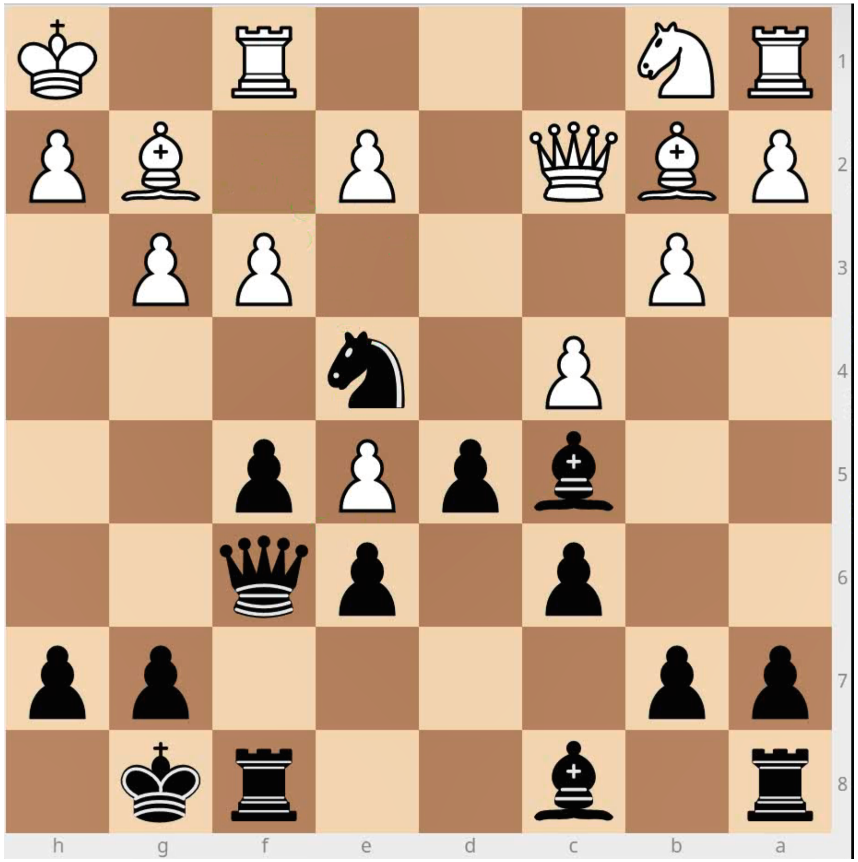
The Physiology of Fixation and Attention



Lateral Geniculate Nucleus (LGN).

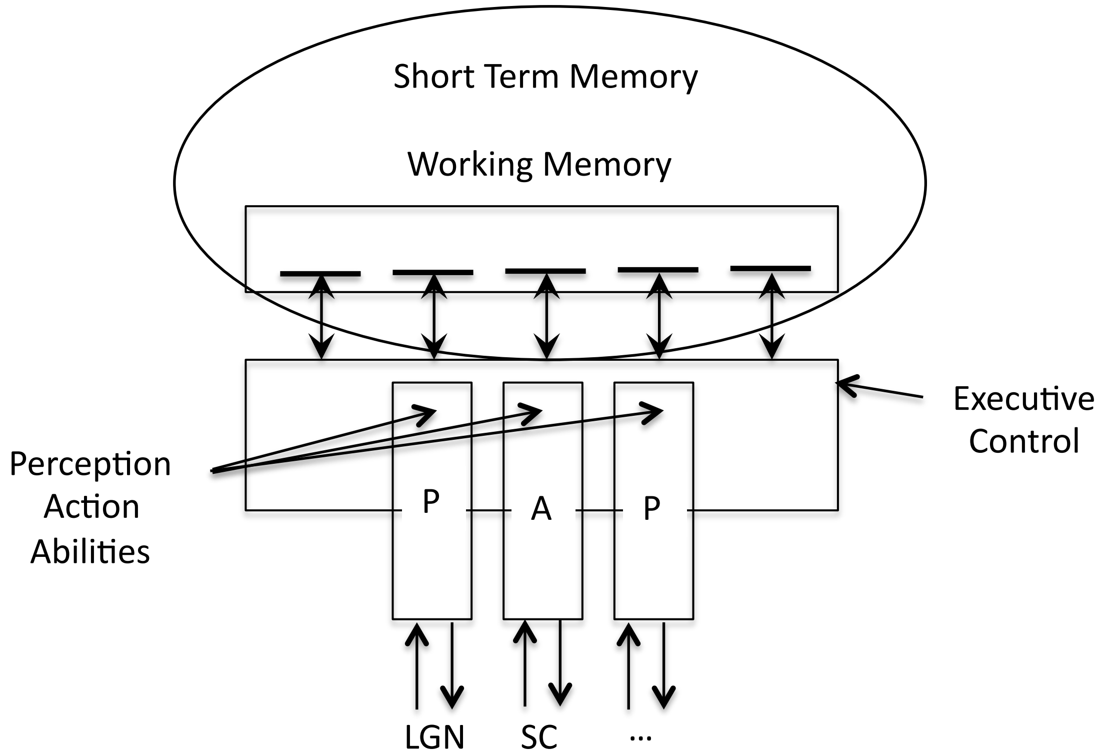
- Filter for Attention: The LGN Suppresses non-attended visual information
- The LGN relays a filtered retinotopic map to the visual cortex.

The Physiology of Fixation and Attention



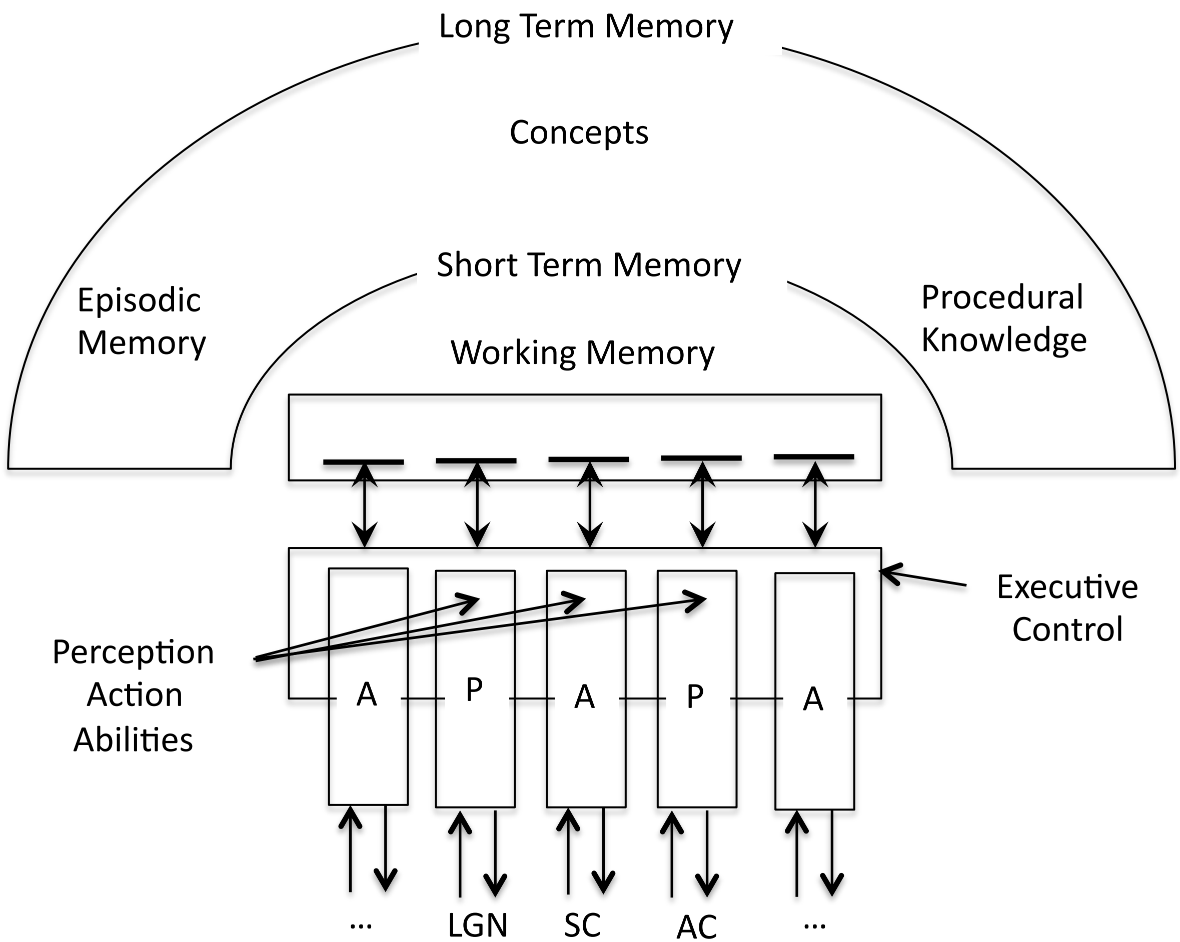
The LGN suppresses visual information

Cognition is limited by Working Memory



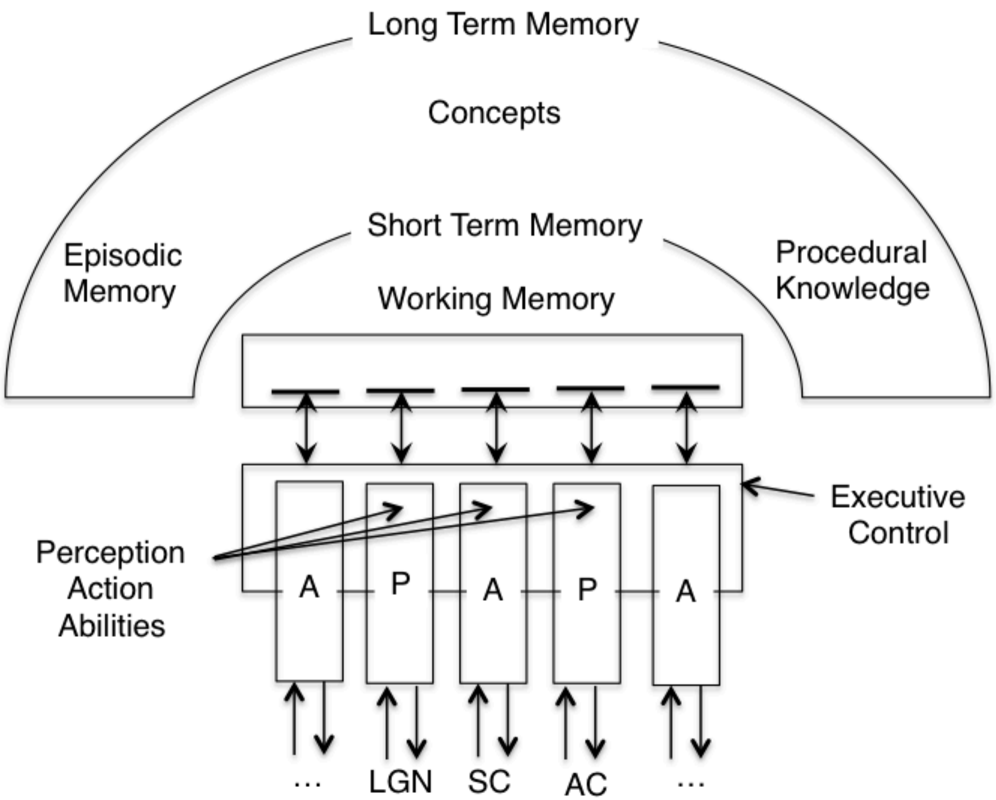
G. A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81-97, 1956.

Working Memory (WM) associates perception with Long Term Memory (LTM) and Short Term Memory (STM)



N. Cowan, Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*(2), 197–223, 2014

WM associates perception with STM and LTM



Working memory elements are called “entities”.

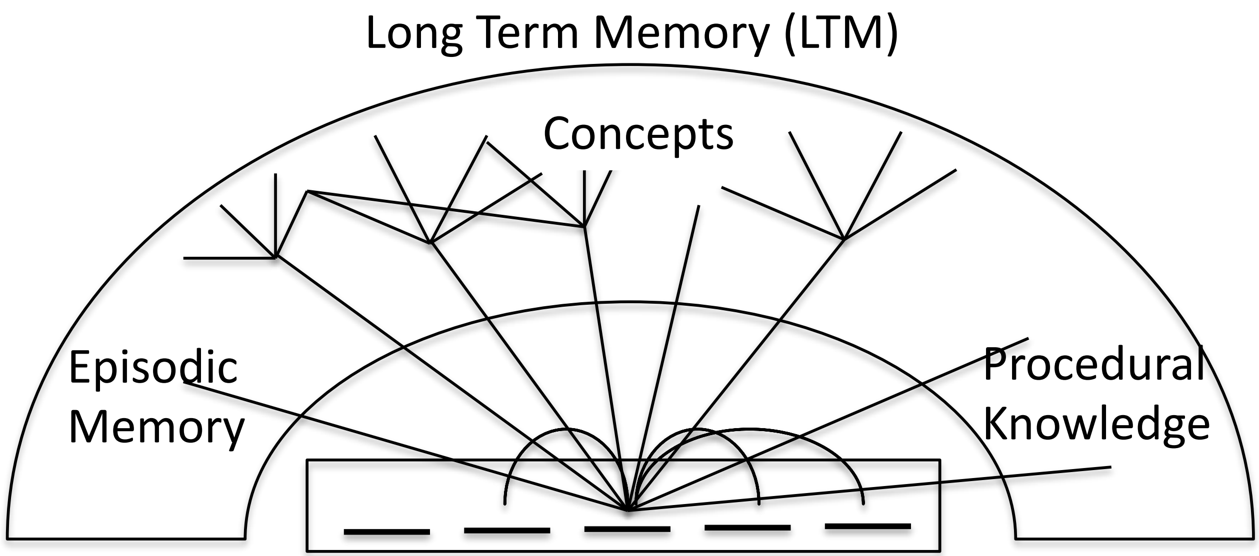
WM entities represent perceived or recalled phenomena.

Entities can be associated with

- Perception and action
- Episodic memories
- Concepts
- Procedural knowledge
- Reasoning Knowledge
- Other forms of memory

W. Kintsch, *Comprehension: A paradigm for cognition*, Cambridge university press, 1998.

Spreading Activation from WM to LTM



Hebbian model for association of entities from working memory with concepts, Procedural knowledge, Episodic memory in Long Term Memory Association

J. R. Anderson, A spreading activation theory of memory, Journal of Verbal Learning and Verbal Behavior, Volume 22, Issue 3, Pages 261-295, June 1983

An Information Processing Model for Comprehension in Chess

<u>Concepts</u>: Mental constructs generalized from particular instances.

Concepts are basic elements of cognition.
Concepts are modeled with Frames.

<u>Frames</u>: Abstract schema for concepts.

<u>Schema for a Frame</u>:

- A unique ID
- A name (optional)
- A set of relations to other concepts
- Meanings: episodic memories as examples
- Roles: Actions that are enabled or prevented by the concept.

M. Minsky, (1975) A Framework for Representing Knowledge, in: Patrick Henry Winston (ed.), The Psychology of Computer Vision. McGraw-Hill, New York), 1975.

Concept Frames for Chess Entities

<u>Frames</u> provides schema for representing concepts as entities in WM.

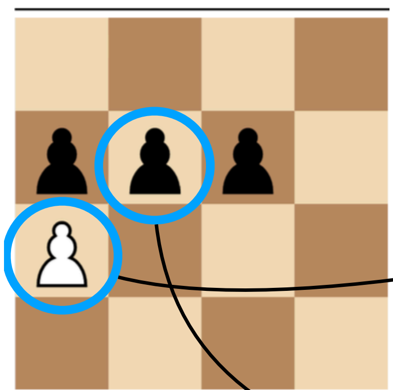
Frames encode relations between entities and Long Term memory
(relations in a frame are internal and immutable)

Example: Concept Frame for a Chess piece.

```
(ChessPiece (piece-ID)
    (kind (one-of (king, queen, bishop, knight, rook, pawn)))
    (color (one-of (black white)))
    (position (row (range 1 to 8)   (column (range a to h))
    (actions (move-procedures))
)
```

F. Gobet, and H Simon, (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory*, *6*, 225-255, 1998

Beginners reason with pieces



a) wall of pawns

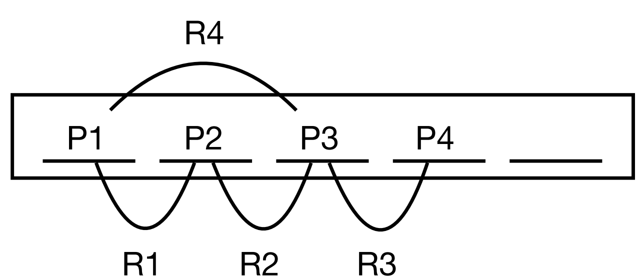
```
( ChessPiece ( P1 )
        ( kind  ( pawn ) )
        ( color ( white ) )
        ( position ( A ; 4 ) )
        ( actions ( pawn_moves ... ) )
)
```

```
( Relation ( R1 )
        ( Name ( Threatened ) )
        ( Kind  ( offensive ) )
        ( Subject ( P1 ) )
        ( Object ( P2 ) )
)
```

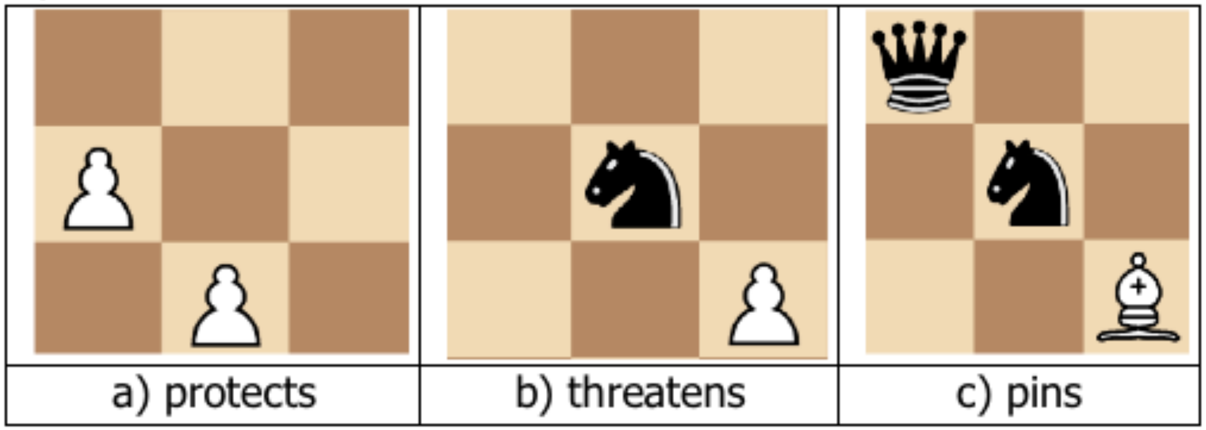
```
( ChessPiece ( P2 ) ( . . . ) )
( ChessPiece ( P3 ) ( . . . ) )
( ChessPiece ( P4 ) ( . . . ) )
```

```
( Relation ( R2 ) ( . . . ) )
( Relation ( R3 ) ( . . . ) )
( Relation ( R4 ) ( . . . ) )
```

Working Memory:



Experts reason with chunks.



Chess chunks are concepts for configurations of pieces, or configurations of simpler chunks. Chunks are composed hierarchically.

Chess chunks associate configurations with actions, roles, and meanings.

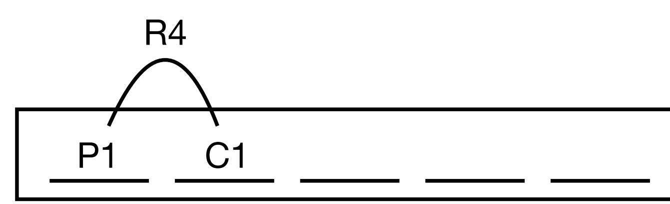
Examples:

a) Protects (WP1, WP2)

b) Threatens (WP1, BK1)

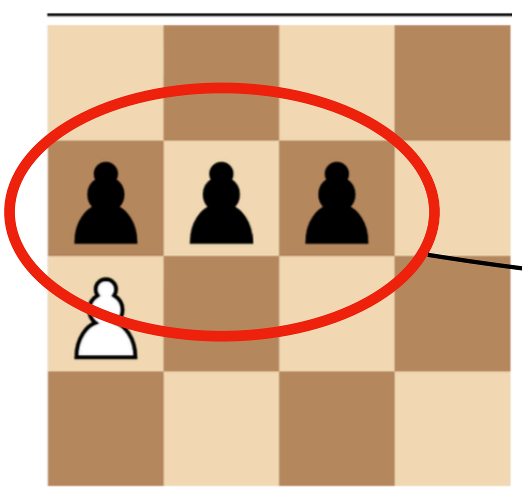
c) Pins(WB1, BQ, BK1)

Working Memory:



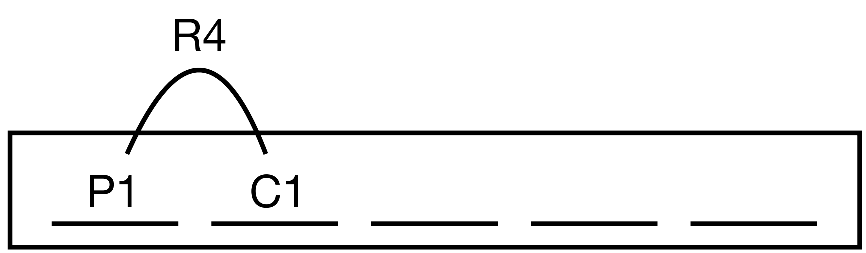
Chess Chunks: Compound Concepts

Chunks reduce WM load by replacing several entities with a single compound entity.
Experts reason with several thousand chunks.

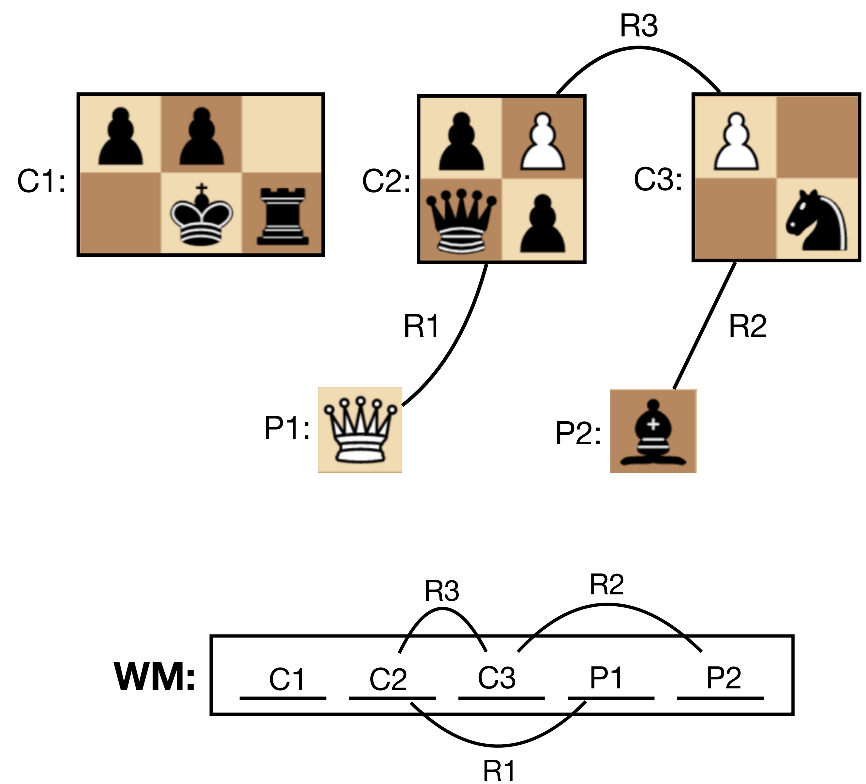


(Concept Wall-of-Pawns
(Color (black, white))
(Kind (Defensive))
(Composed of (list-of (Pawns))
(Position (row)
(Relations (Beside(P1, P2), (Beside(P2, P3))
(Roles (blocks (opponent pawns), screens (own pieces))
)

Working Memory:



Despite Chunking, WM is rapidly over-loaded

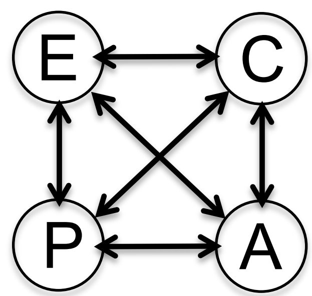


Emotion and Cognition:
Two Complementary Intentional Systems

<u>Emotion</u>:
Fast, Reactive, Predictable.

Enables rapid reaction to threats and opportunities.

Kahneman's System 1?



<u>Cognition</u>:
Slow, Deliberative, Creative

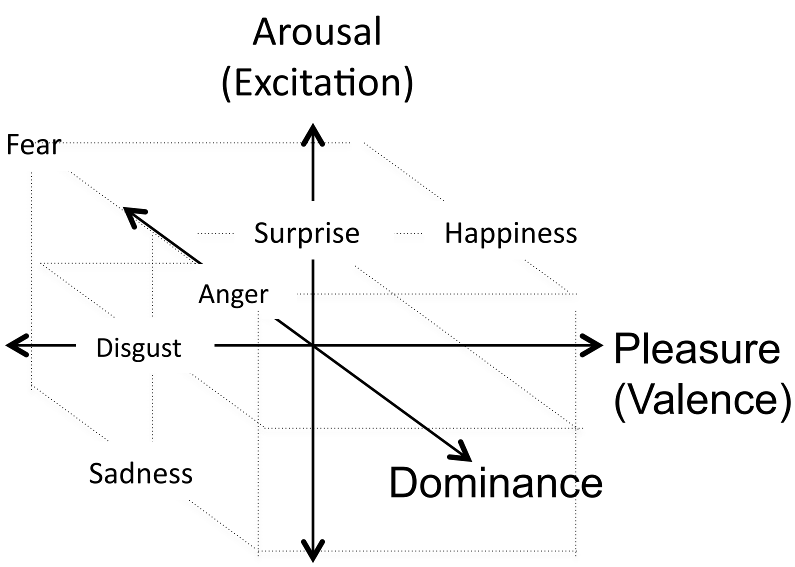
Enables planning, explanation, prediction and understanding.

Kahneman's System 2?

Kahneman D, Egan P. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux; 2011

Emotions substitute experience for reason

Human emotions are displayed by physiological signals learned from past experience with similar situations.



PAD model (Pleasure, Arousal, Dominance)

<u>Hypothesis</u>:
Pleasure, Arousal and Dominance are associated with concepts by experience

For any situation, (P, A, D) express:

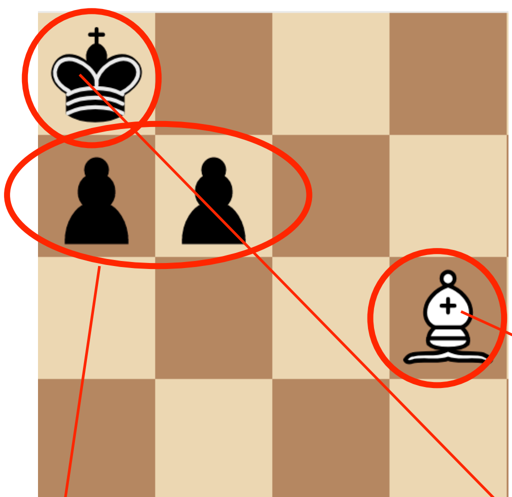
Pleasure: Frequency of positive (or negative) outcomes

Arousal: Imminence of opportunity or threat.

Dominance: Confidence in ability to control outcome.

Hypotheses:
Emotions drive selection of chunks

Players prefer chunks with high criticality (arousal), valence (positive experience) and dominance (confidence in outcome).



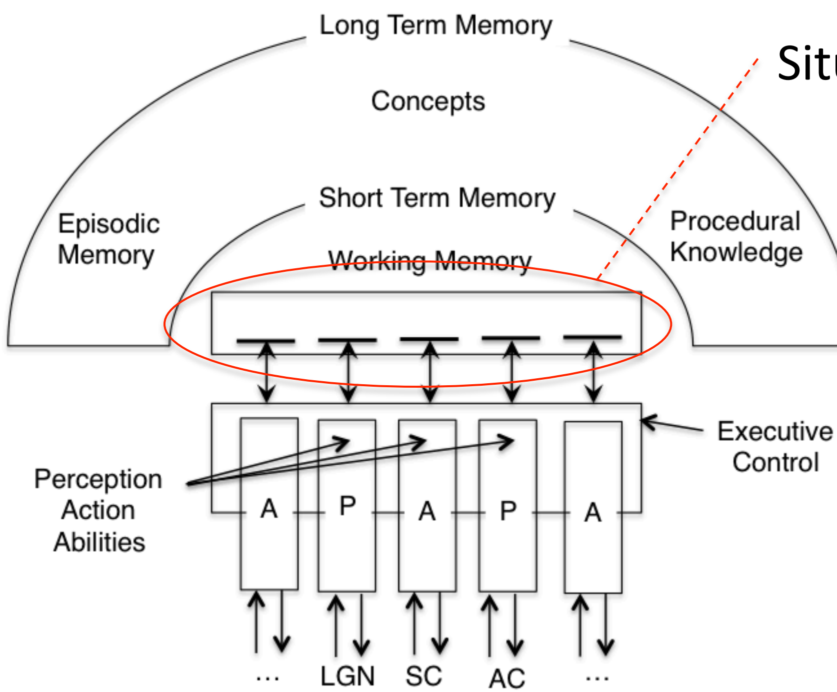
(Concept Trapped-King (BK)
    (Color (black))
    (Kind (Offensive))
    (Composed of (list-of (WB, BK, WP))
    (Blocks (move (BK, WP)))
    (Blocks (move (BK, WB)))
    (Emotion (P +), (A +), (D ++))
)

Wall of Pawns (WP)

Black King (BK)

White Bishop (WB)

Chess Situations: Relations over Chunks



Situation Model

A situation is a set of relations between entities. Relations are external to entities and changeable.

Situation Model schema:

```
(Situation (S-Name)
   (E1  entity-ID) (E2 Entity-ID)
   (R1 (Beside E1 E2))
   (R2 (meanings Episodic-Memory-ID))
   (R3 (Actions Action-ID))
   (Emotions (P) (A) (D))
)
```

Current research hypotheses:

1) Chunks are learned from frequently encountered situations.

2) Emotions (P, A, D) guide the selection of situations and concepts used for reasoning.

T. Guntz, J.L. Crowley, D. Vaufreydaz, R. Balzarini, P. Dessus, The Role of Emotion in Problem Solving: first results from observing chess, *Workshop on Modeling Cognitive Processes from Multimodal Data*, ICMI 2018, Oct 2018.

A Second Experiment

Protocol:

- 7 Tasks of increasing Difficulty (4 Mate-in-N tasks and 3 survival tasks)
- Retroactive Task Explanation (RTE) after each task,

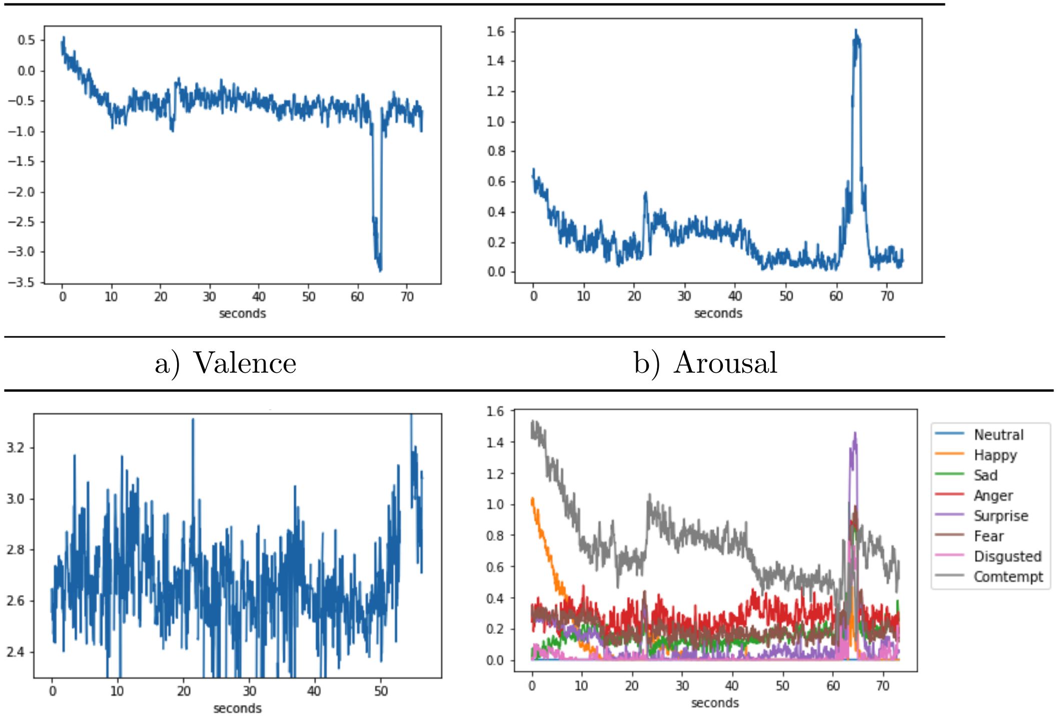
RTE: subject describes understanding of the problem situation.

23 subjects (2 expert, 19 intermediate, 2 beginners).

(Elo ranges 1930 to 2000 and 1197 to 1700).

Measurements: Eye-gaze (Fovio), pupil dilation, FACS, Ekman emotions (Open Face), Self Touches (Open Pose) .

S6 (Elo 1950), Task 8: An impossible survival task

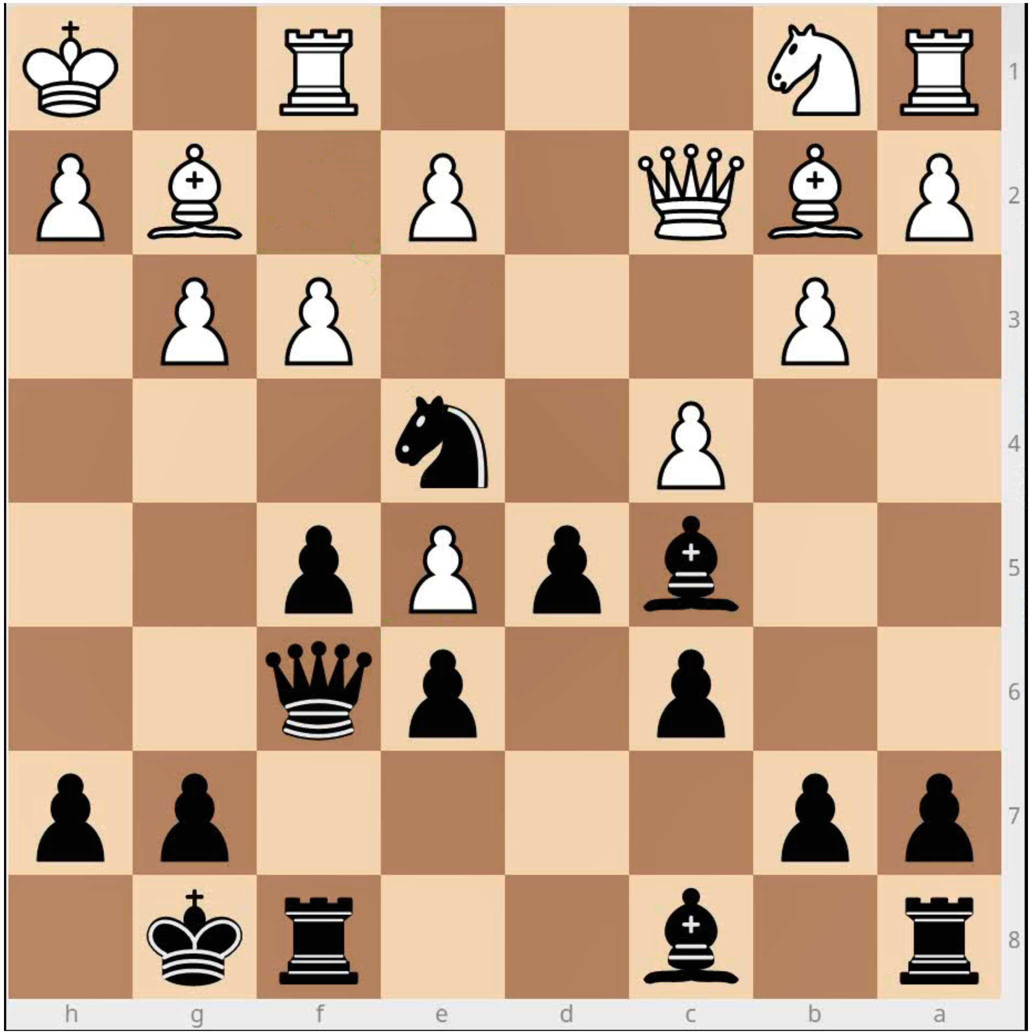


a) Valence b) Arousal

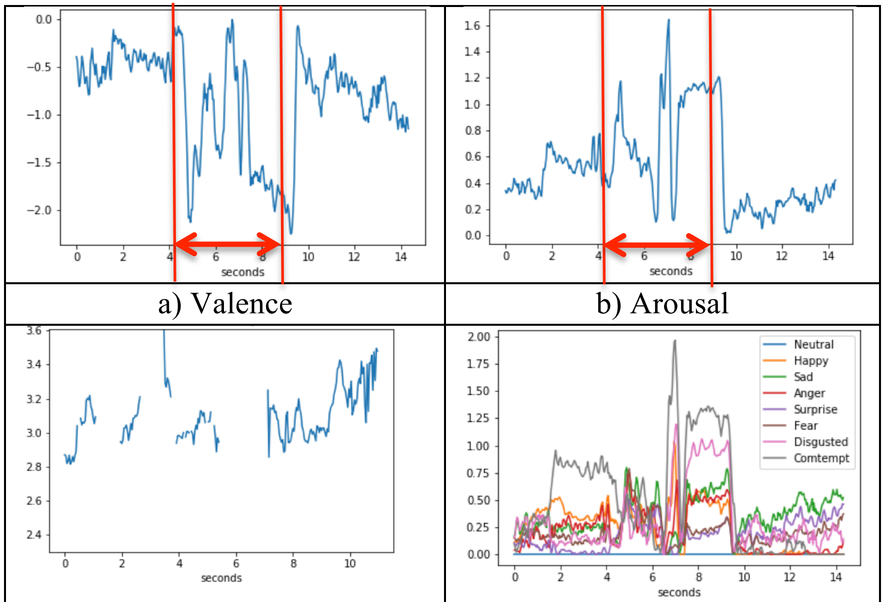
c) Pupil Diameter d) Emotion States

The spike in Valence, Arousal, pupil diameter and fear and disgust (emotion states) corresponds to a self-reported recognition that the situation was hopeless.

Example: Task 4 (mate in 3)

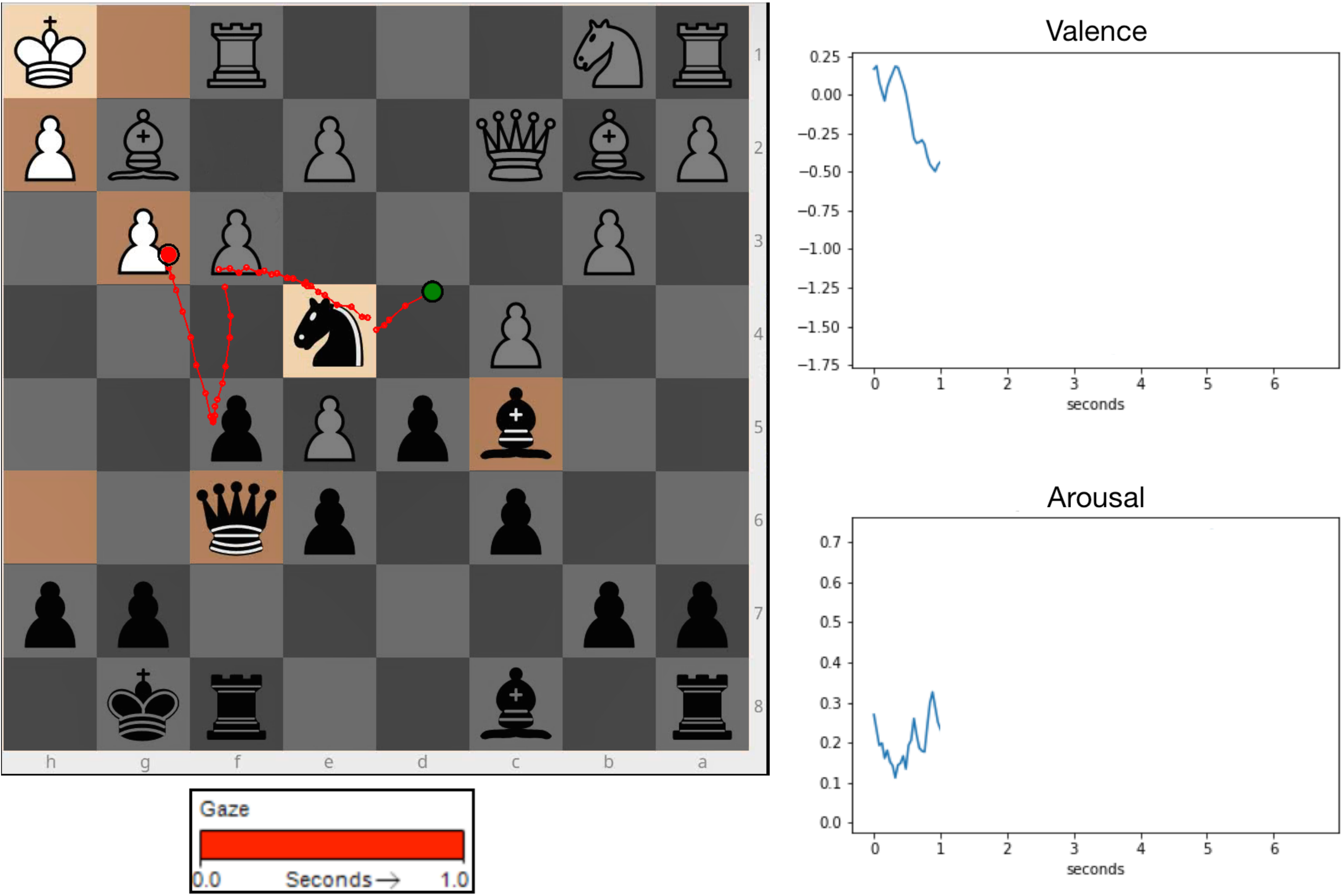


Subject S12 (Expert), Task 4 (mate in 3)

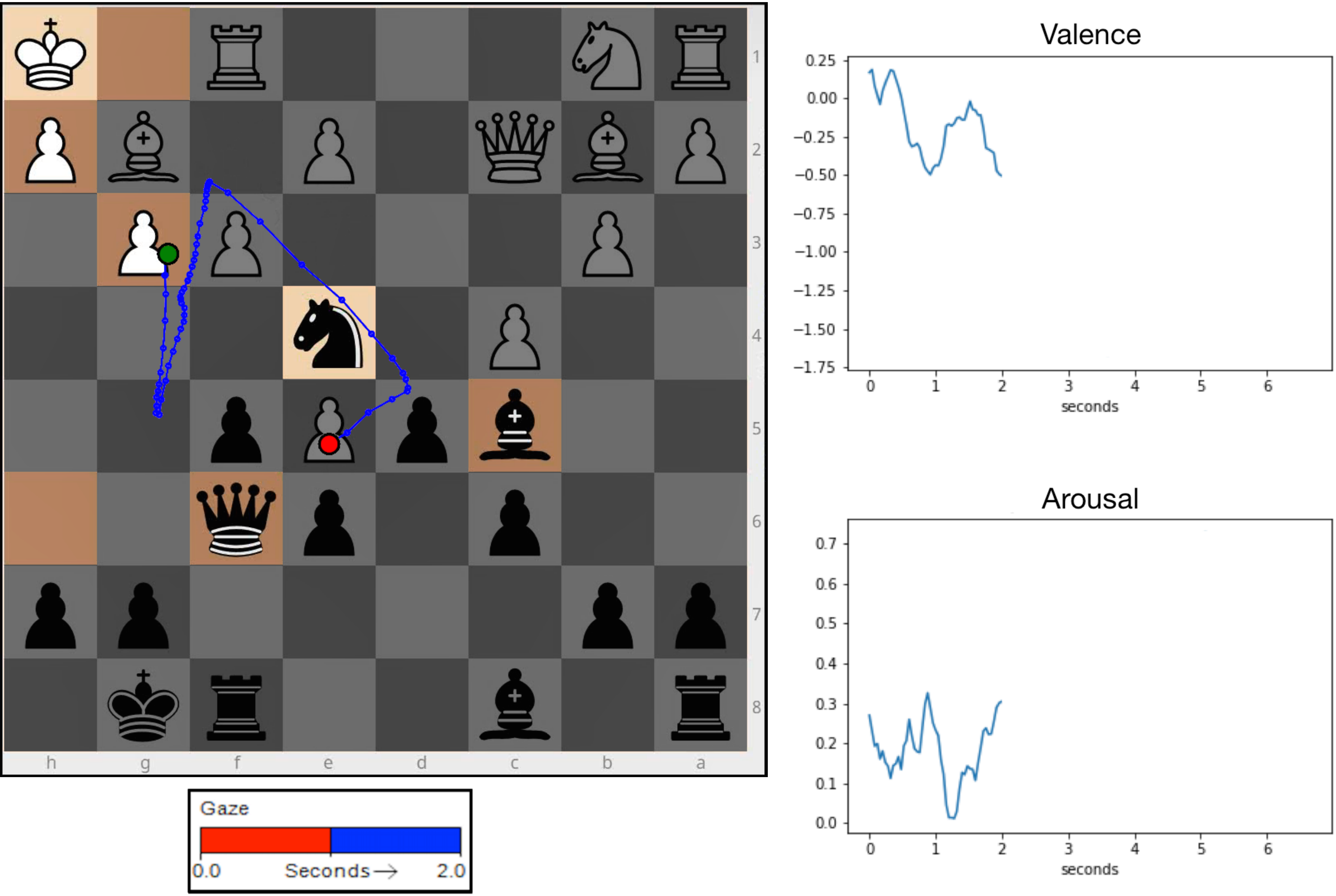


a) Valence
b) Arousal
c) Pupil Diameter
d) Emotion States

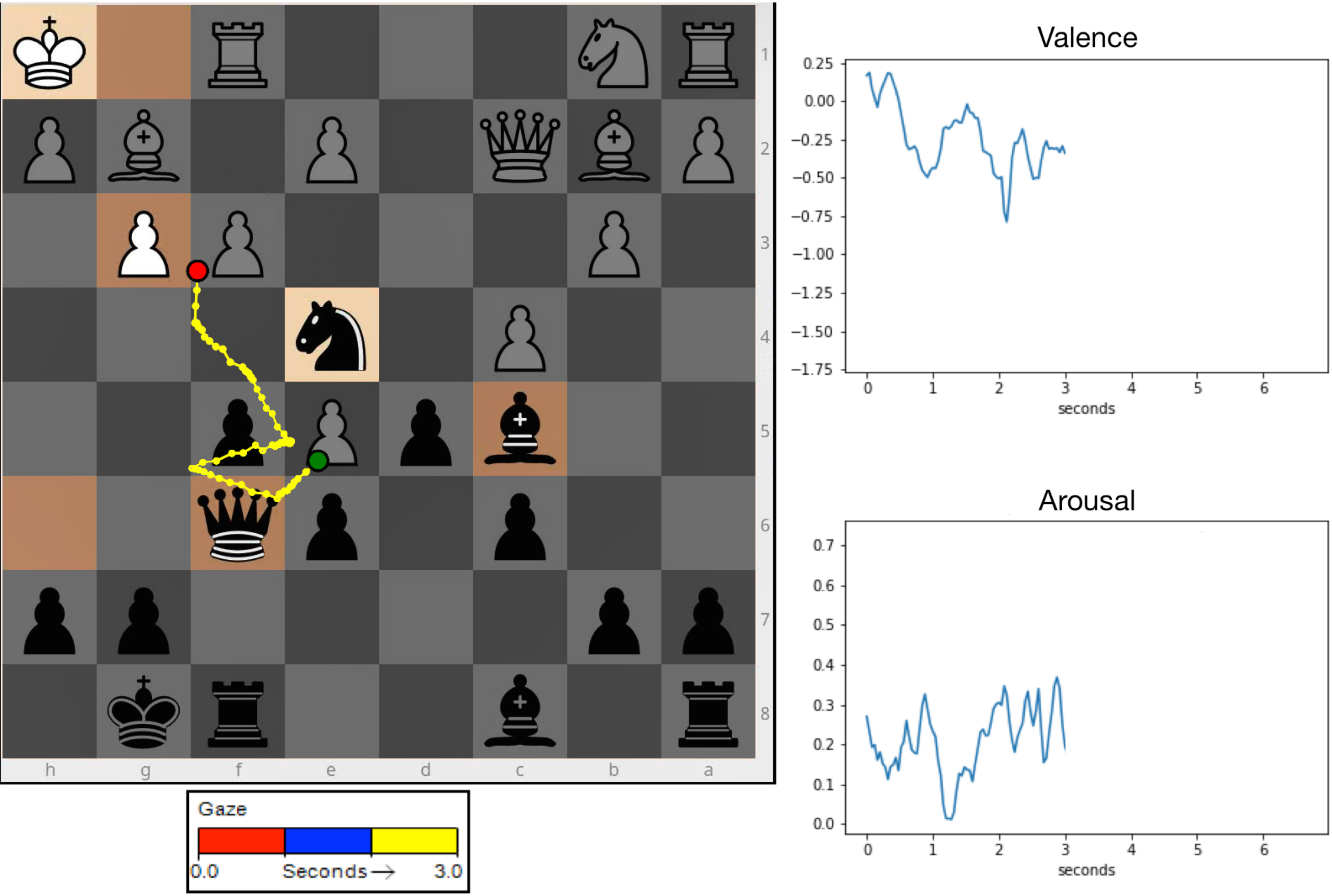
Ok. I see that there is a mate in 3 moves here. It is a kind of pattern because again the king is not able to move at all. There are 3 pieces here, Queen F6, Knight E4 and Bishop at C5."



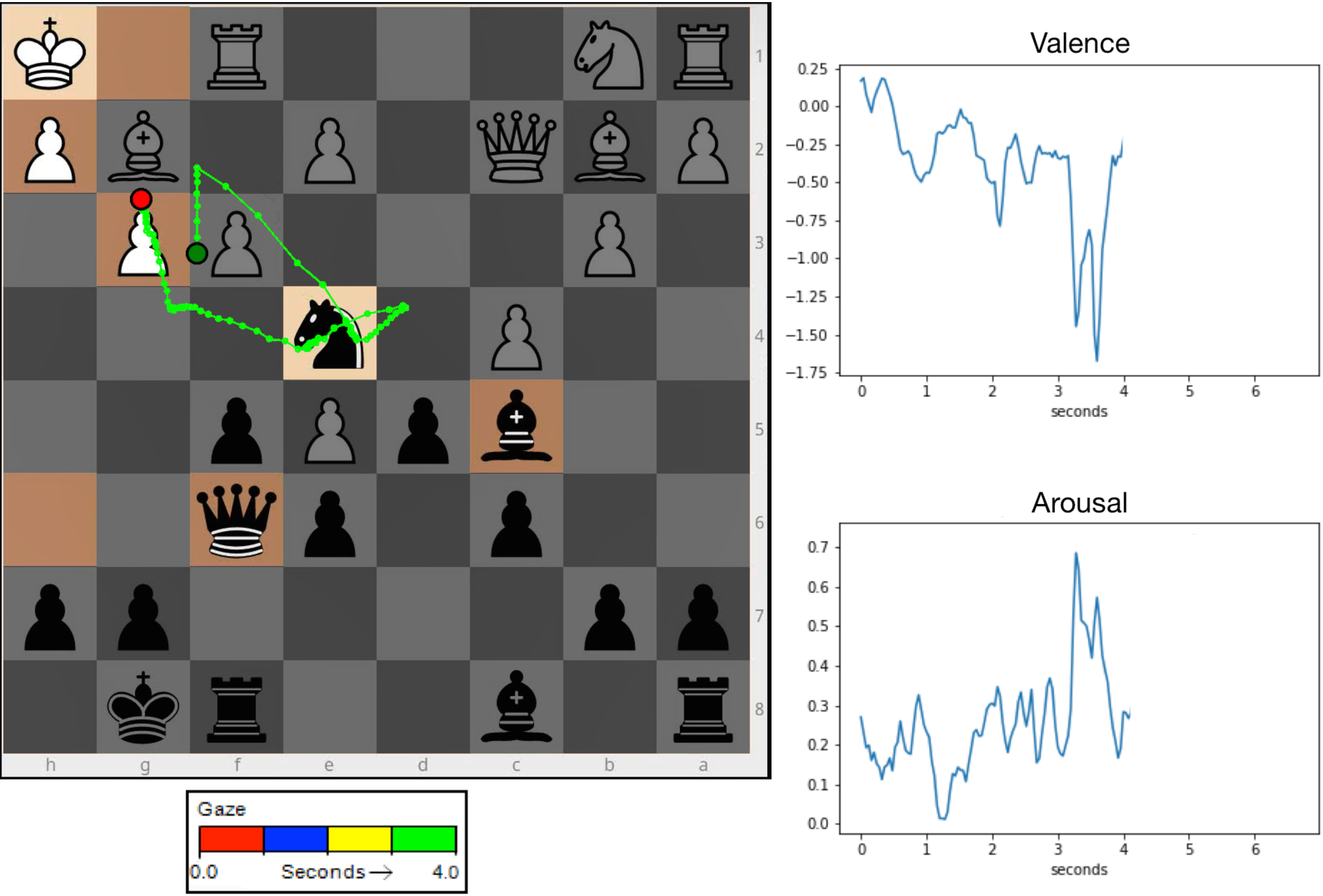
1
2
3
4
5
6
7
8
h
g
f
e
d
c
b
a
Gaze
0.0
Seconds→
1.0
Valence
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
0
1
2
3
4
5
6
seconds
Arousal
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
seconds



h g f e d c b a
1 2 3 4 5 6 7 8
Gaze
0.0
Seconds
2.0
Valence
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
0 1 2 3 4 5 6
seconds
Arousal
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5 6
seconds

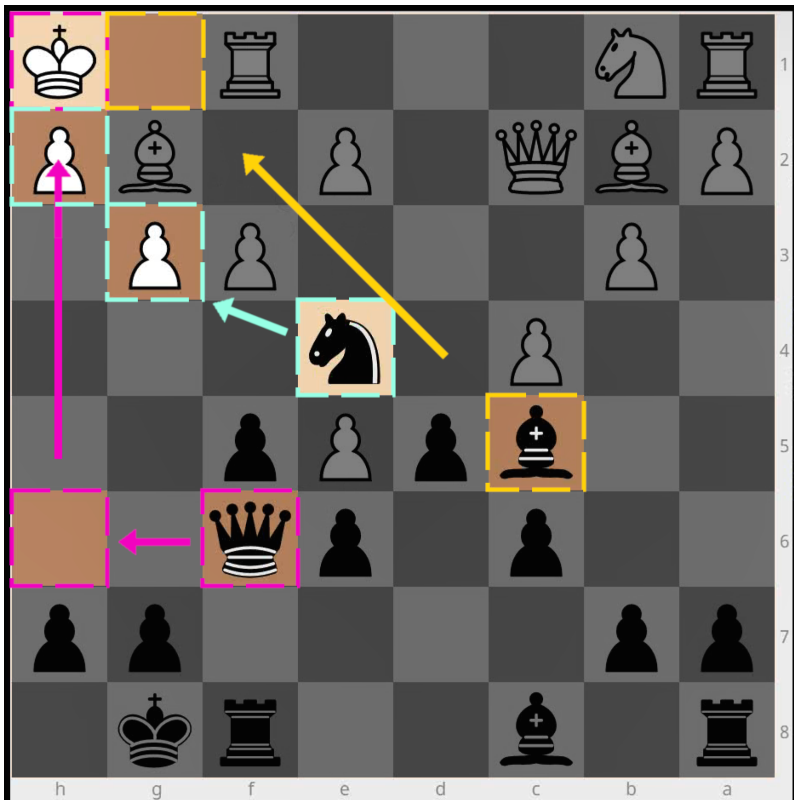


1
2
3
4
5
6
7
8
h
g
f
e
d
c
b
a
Gaze
0.0
Seconds →
3.0
Valence
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
0
1
2
3
4
5
6
seconds
Arousal
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
seconds



1
2
3
4
5
6
7
8
h
g
f
e
d
c
b
a
Gaze
0.0
Seconds →
4.0
Valence
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
0
1
2
3
4
5
6
seconds
Arousal
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
seconds

Retroactive Task Explanation by S12



S12 Task Explanation:

"Ok. I see that there is a mate in 3 moves here.

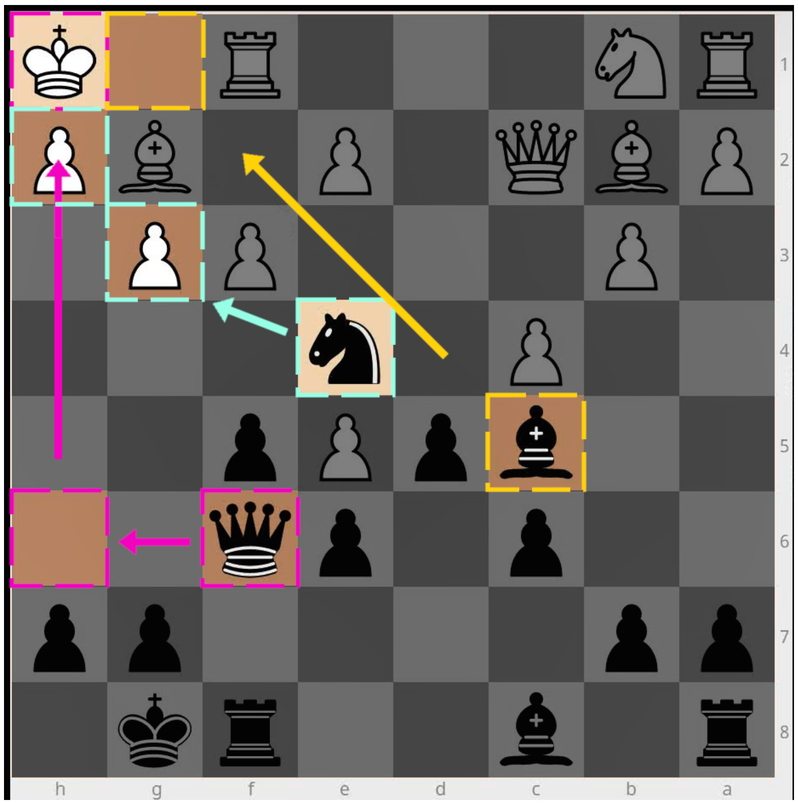
It is a kind of pattern because again the king is not able to move at all.

There are 3 pieces here, Queen F6, Knight E4 and Bishop at C5. So the bishop already controls the only available square of the white king, so there are two pieces.

So first check with the knight then the queen. "

(Knight takes Pawn, check, Pawn H2 at takes Pawn , Queen to H6 Check mate).

Grand Challenge: Automatically Generate Narratives for Player Comprehension.



Chess Concepts:

- C1. Bishop Blocks King
- C2: Knight threatens Pawn
- C3: Pawn takes Knight
- C4: Queen Checks King

Subjects Plan:

S1: Knight (E4) Takes Pawn (G3)

- S2: Pawn (H2) takes Knight (G3)
- S3: Queen to H6 Check-mate

Potential Applications

- Collaborative Intelligent Systems
  - Intelligent Auto-pilot and Drivers Assistant for aircraft, automobiles, trucks, buses, heavy equipment
  - Collaborative robots for manufacturing and service industry
- Training and Education
  - Student Aware intelligent training Pulpit
  - Training for driving vehicles and heavy equipment.
- Socially Aware Service Robots,
- Human Aware personal mobile devices
  - Personal computing,
  - tablet,
  - Smart phone
- Ambient Intelligence
  - Smart Home, furniture, desk, kitchen etc.

SATT Linksium Project MAT / Sym2B

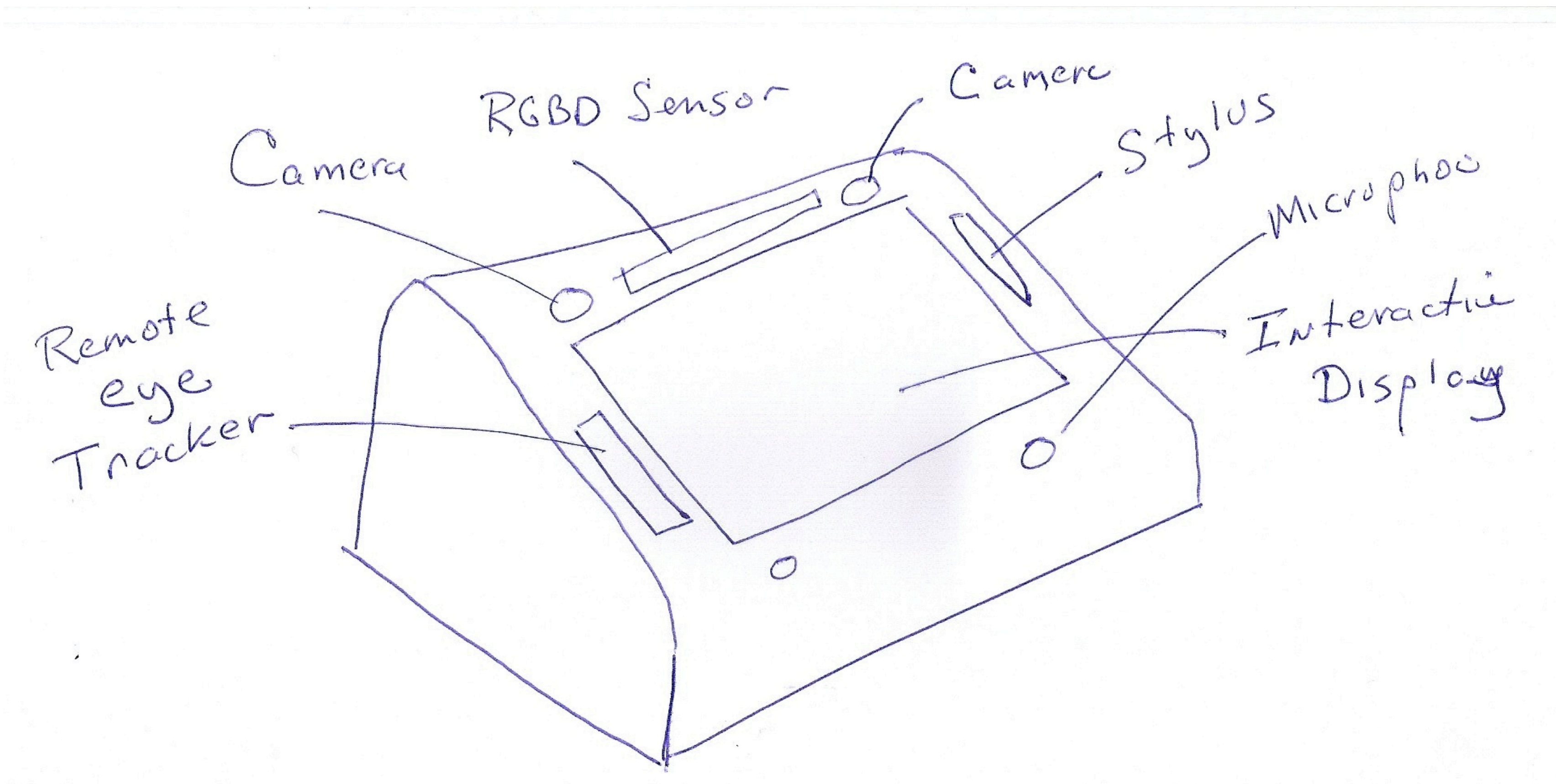
Simulateur avec ou sans mouvement 6-axes

- Un marché mondial annuel d'environ 100 millions d'euros décomposé en :
  - 400 simulateurs de camions
  - 150 simulateurs de bus
  - 100 simulateurs de cars

Training simulator augmented with remote eye-tracking, face expression analysis, pupil dilation and body gesture models.

Student Aware Training Pulpit

Collaboration between the Pervasive team (LIG) and Philippe Dessus and colleagues at LaRAC



30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from of Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

Limitations

For PAC:

- A mature paradigm mis-identified with a competing approach (MVC)

For Situation Models:

- State models and behaviors are constructed by hand.
- Automatic acquisition remains an important challenge.

For ANR CEEGE (Work in Progress!)

- Proposed models explain comprehension and predict behavior. They do not prescribe implementation.
- Chess is much simpler than the real world!
- Very preliminary investigation (TRL 2!) Models have not yet been evaluated. Value will be their usefulness for prediction and explanation.

Open Challenges

1. Multi-modal interaction has a lot to learn from Cognitive science: We need to integrate concepts and theories from human cognition and physiology to build human-aware interactive systems.
2. Learning through interaction: We need a better theory for systems that learn from interactions with users at sensori-motor, corporal, operational, and intentional levels (deep reinforcement learning?)

Some Lessons from the last 30 years
(for students)

- Be open to new ideas. Never stop learning.
  - Sometimes wrong, Always useful.
- Take inspiration from other disciplines.

  (Most successful research is inspired by results from other fields)
- Beware of chasing fads
  - The early birds get the low-hanging fruit.
  - The others get polite rejection letters.
- Follow the fun

  (Cool ideas with no practical use have the most long-term impact)

Special Thanks to: ….

- Joelle Coutaz (For PAC, CARE properties, and everything else…)
- Thomas Guntz for images and data.
- Patrick Reignier, Dominique Vaufreydaz, Philippe Dessus, and the Pervasive Team (for listening to my crazy ideas)
- Laurence Nigay (for ICARE and other good ideas)
- Francois Bérard for digital version of demo films from the 90's.
- Reiner Steifelhagen, Jan Kleindeist, Jan Currin (CHIL)
- Ben Krose (for the Lino film)
- Univ Grenoble Alpes, INRIA Grenoble and LIG for providing a fertile environment for "thinking outside of the box" research.

Bibliography

1. Bolt, R.A., “Put-that-there”, Voice and gesture at the graphics interface . Vol. 14, No. 3, pp. 262-270, ACM, 1980.
2. T.S. Kuhn, The structure of scientific revolutions. University of Chicago press, 2012.
3. Coutaz, J, PAC, An Object Oriented Model for Dialog Design. In Human–Computer Interaction, INTERACT'87, pp. 431-436 , 1987.
4. Krasner, G.E. and Pope, S.T.. A description of the model-view-controller user interface paradigm in the smalltalk-80 system. Journal of object oriented programming, 1(3), pp.26-49. , 1988.
5. J. L. Crowley and J. Coutaz, "Vision for man-machine interaction", EHCI, Grand Targhee, Aug 1995.
6. J. Coutaz, F. Berard and J. L. Crowley, "Coordination of Perceptual Processes for Computer Mediated Communication", FG96 - International Workshop on Face and Gesture Recognition, Vermont, Oct 1996.
7. J. L. Crowley and F. Berard, "Multi-Modal Tracking of Faces for Video Communications", IEEE Conference on Computer Vision and Pattern Recognition, CVPR '97, St. Juan, Puerto Rico, June 1997.
8. J. L. Crowley, J. Coutaz and F. Berard, "Things that See: Machine Perception for Human Computer Interaction”, Communications of the A.C.M., Vol 43, No. 3, pp 54-64, March 2000

Bibliography

9. J. L. Crowley, J. Coutaz, G. Rey and P. Reignier, "Perceptual Components for Context Aware Computing", UBICOMP 2002, International Conference on Ubiquitous Computing, Goteborg, Sweden, September 2002.
10. J Coutaz, J. L. Crowley, S. Dobson, and D. Garlan, "Context is Key", Communications of the ACM, Special issue on the Disappearing Computer, Vol 48, No 3, pp 49-53 March 2005.
11. Johnson Laird. The Computer and the Mind
12. O. Brdiczka, J. L. Crowley, J. Curin, J. Kleindienst, "Situation modeling", in Computers in the Human Interaction Loop, A. Waibel (Ed), Springer Verlag no 12, p. 121-132, 2009.
13. J. Soldatos, I Pandis, K Stamatis, L Polymenakos, JL Crowley, "Agent based middleware infrastructure for autonomous context-aware ubiquitous computing services", Computer Communications, Vol 30, Issue 3, Pages 577-591, February 2007.
14. R. Barraquand, J. L. Crowley, "Learning Polite Behavior with Situation Models", Third International Conference on Human Robot Interaction (HRI 2008), 12-15 March 2008, Amsterdam, The Netherlands
15. Kröse, B.J., Porta, J.M., van Breemen, A.J., Crucq, K., Nuttin, M. and Demeester, E., 2003, November. Lino, the user-interface robot. In European Symposium on Ambient Intelligence (pp. 264-274). Springer, Berlin, Heidelberg.

Bibliography

16. J. L. Crowley and J. Coutaz, "An Ecological View of Smart Home Technologies", 2015 European Conference on Ambient Intelligence, AMI 2015, Athens, Nov. 2015
17. Kahneman D, Egan P. Thinking, fast and slow. New York: Farrar, Straus and Giroux; 2011
18. J. Le Leoudec, T. Guntz, D. Vaufreydaz and J. L. Crowley, Deep Learning Investigation for chess player attention, prediction using eye-tracking, and game data, Submitted to 2019 ACM Symposium on Eye Tracking Research & Applications, ETRA'2019, Denver, Colorado (under review).
19. G. A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63(2), 81-97, 1956.
20. N. Cowan, Working memory underpins cognitive development, learning, and education. Educational Psychology Review, 26(2), 197–223, 2014
21. J. R. Anderson, A spreading activation theory of memory, Journal of Verbal Learning and Verbal Behavior, Volume 22, Issue 3, Pages 261-295, June 1983
22. W. Kintsch, Comprehension: A paradigm for cognition, Cambridge university press, 1998.
23. M. Minsky, (1975) A Framework for Representing Knowledge, in: Patrick Henry Winston (ed.), The Psychology of Computer Vision. McGraw-Hill, New York), 1975.

Bibliography

24. F. Gobet, and H Simon, (1998). Expert chess memory: Revisiting the chunking hypothesis. Memory, 6, 225-255, 1998
25. Kahneman D, Egan P. Thinking, fast and slow. New York: Farrar, Straus and Giroux; 2011
26. J. A. Russell , A. Mehrabian, "Evidence for a three-factor theory of emotions", Journal of Research in Personality Vol. 11(3), pp 273-294, Sept 1977.

30 Years of Multimodal Interaction

- The Origins of ICMI (a personal view)
- The PAC framework for Multimodal Interaction
- Multi-modal Interaction with Context Aware Services
- PACE: A Conceptual Framework for Multimodal Interaction
- Modeling Comprehension from Eye-Gaze and Emotion.
- Conclusions : Limitations, Open Challenges, Lessons.

Put That There:
~~20~~ 30 Years of Multimodal Interaction

Prof. James L. Crowley
INRIA GRA Research Center
Grenoble Institut Polytechnique
Univ. Grenoble-Alpes