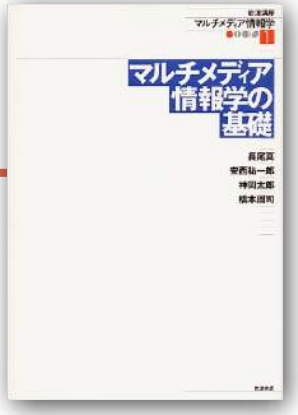
Cognitive Media Processing #2

Nobuaki Minematsu

Title of each lecture



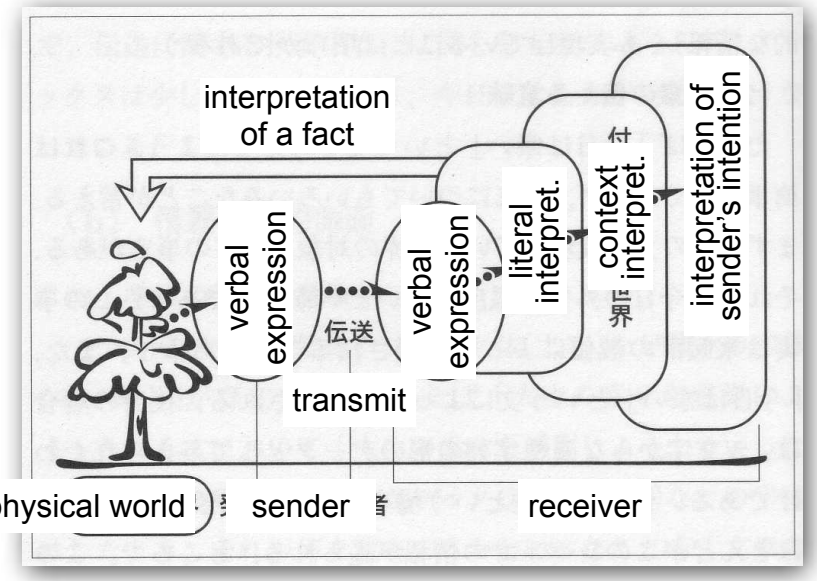
- Theme-1
  - Multimedia information and humans
  - Multimedia information and interaction between humans and machines
  - Multimedia information used in expressive and emotional processing
  - A wonder of sense - synesthesia -
- Theme-2
  - Speech communication technology - articulatory & acoustic phonetics -
  - Speech communication technology - speech analysis -
  - Speech communication technology - speech recognition -
  - Speech communication technology - speech synthesis -
- Theme-3
  - A new framework for “human-like” speech machine #1
  - A new framework for “human-like” speech machine #2
  - A new framework for “human-like” speech machine #3
  - A new framework for “human-like” speech machine #4

Menu of the last lecture

- The term of “information” used in human communication.
  - Two kinds of definition of information (C. Shannon vs. this lecture)
  - Data and information - intention of a sender and interpretation of a receiver -
- Various forms of information in human communication
  - Classification of media information
  - Context dependency of information
- Information and knowledge
  - From data to information
    - Knowledge-based cognitive processing
  - Unconscious processing
    - Your brain creates your world but you cannot be aware of the brain’s processing.
  - Various forms of information and conversion between them
    - Recognition and synthesis: abstraction and embodiment
  - Logical information and expressive (感性, KANSEI) information
    - Behaviors and information processing of autistics

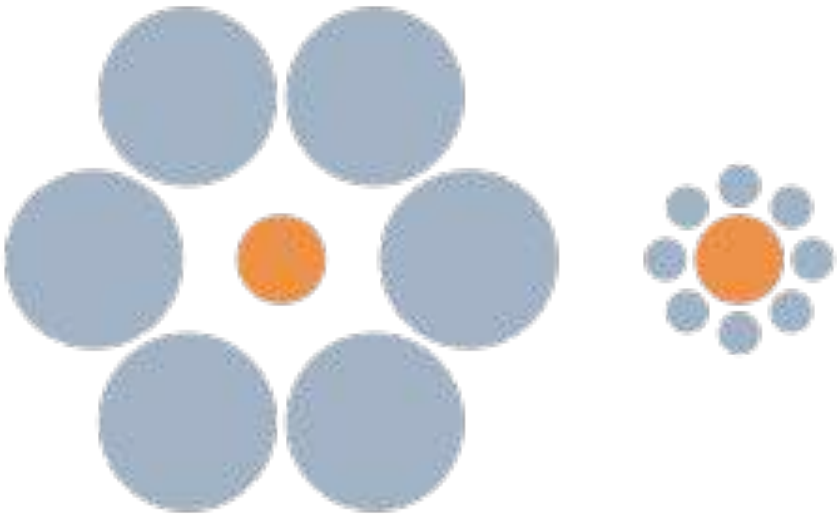
Forms of info. in human communication

- Qualitative aspect of information - intention and interpretation -
  - A message in the form of text
    - Interpretation often requires understanding the *context* of the message including a sender's intention as well as the (literal) *content* of the message.
  - "It's cold this morning."
    - From statement of a weather fact to "I want a cup of hot coffee."
    - Proper interpretation of a message depends on the context where the message is made.
  - High-context language and low-context language
    - High-context : less verbally explicit communication, less written/formal information
    - "Can you pass me the salt?" "Yes, I can."

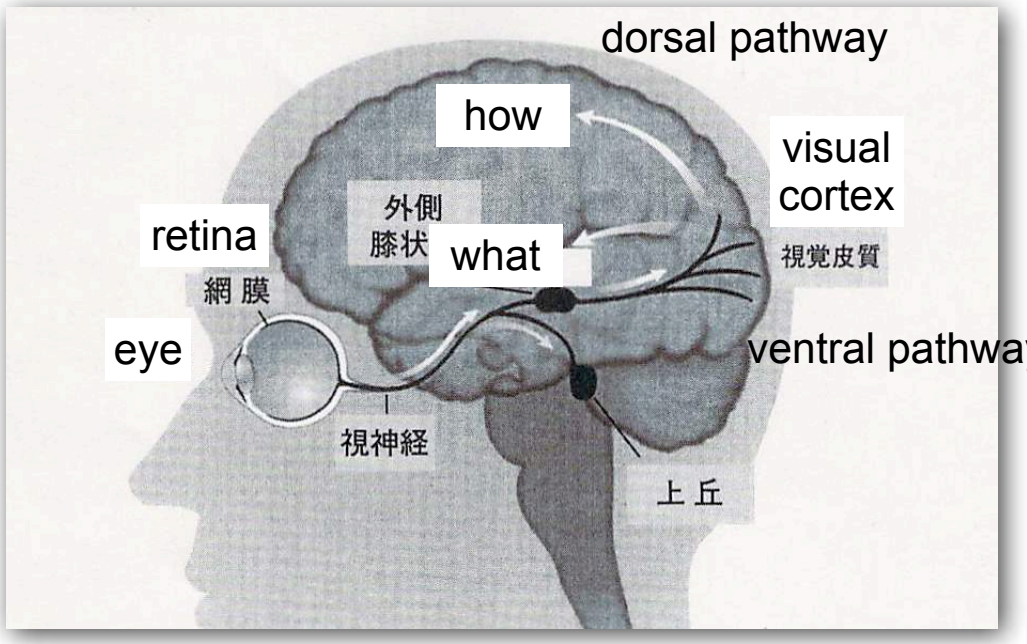


Implicit knowledge

- Unconscious processing
  - Difficult to notice consciously what is being done in the brain “unconsciously”.
    - Ebbinghaus illusion
      - When you pick up one of the circles, is the distance bet. the two fingers different bet. the circles?
      - Your mind is easily tricked but your fingers *in action* are not tricked.
    - What-pathway and how-pathway in the vision system of the brain
      - A brain damage in the visual cortex makes “conscious” experiences of seeing impossible.
      - But blind individuals can behave properly according to the visual characteristics of nearby objects!

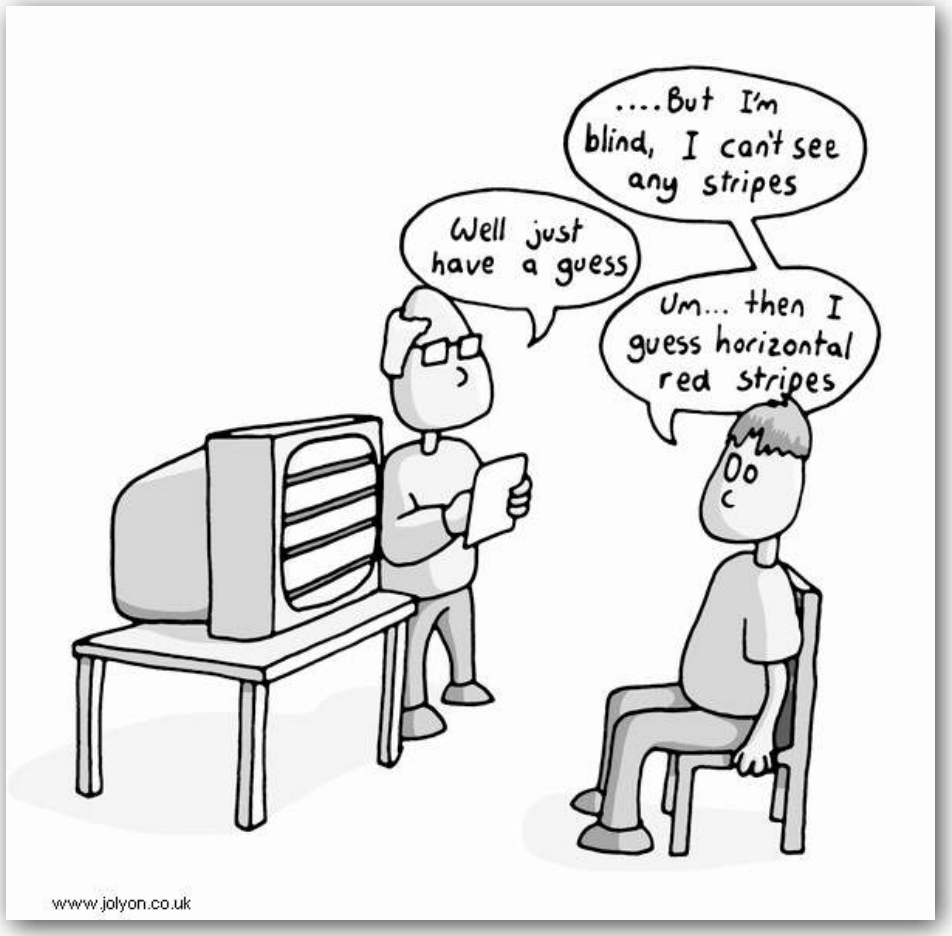


Ebbinghaus illusion



Implicit knowledge

- Unconscious processing
  - Blind sight [L. Weiskrantz'86]



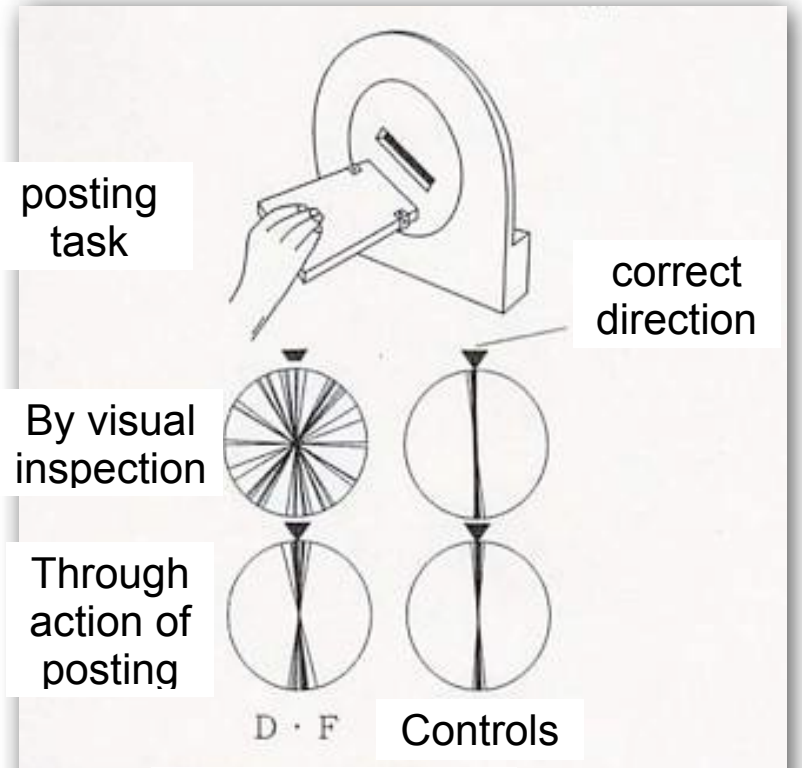


図 1・5 Unconscious action

D.F. has a severe brain damage on the visual cortex but no damage on the cortex associated with handling things. She cannot guess (consciously) the hole direction by visual inspection but can guess (unconsciously) through action of posting.

Logical and expressive

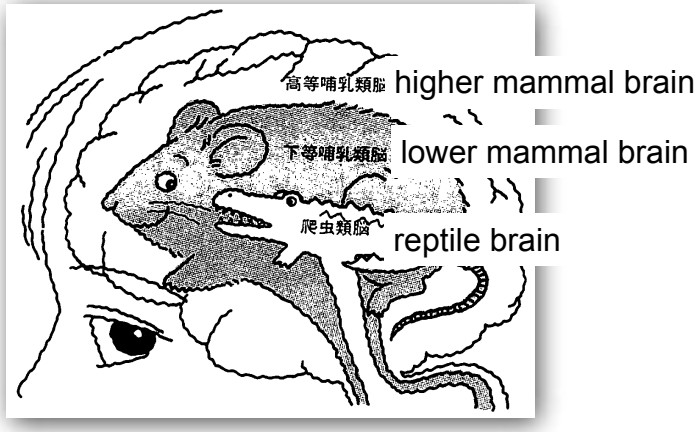
- Logical information and expressive information
  - Logical information
    - Interpretation does not depend on receivers, e.g. objective facts.
  - Expressive (KANSEI, 感性) information
    - Interpretation strongly depends on receivers, e.g. subjective impression.
    - Tastes differ (十人十色).

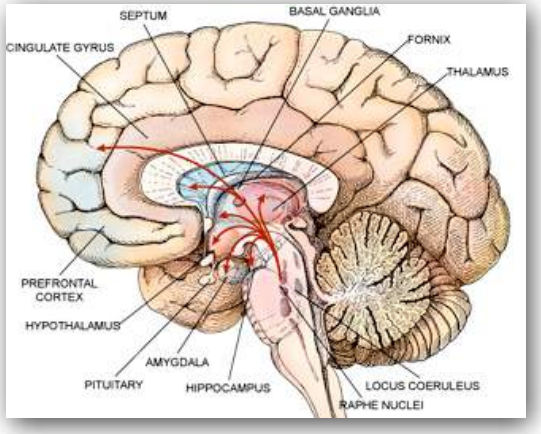
Is Tokyo the capital of Japan?

Which guy do you think is more handsome?

Logical and expressive

- Logical information and expressive information
  - Factors (bases) to describe expressive information
    - Facial expressions (as example)
      - 6 factors of surprise, fear, dislike, anger, happiness, and sorrow
      - A still debatable problem in psychology
    - Theory of mind [D. Premack et. al.'78]
      - The ability to attribute mental states to oneself and others and to understand that others have different mental states than one's own.
        - Different individuals have different minds.
        - Those who don't have theory of mind have difficulty in understanding this fact.
      - One of the theories that explains the cause of autism (自閉症) [S. Baron-Cohen'91]
        - Difficulty in reading the mind of others and understanding that everybody has one's own mind.
        - Difficulty in reading the facial expressions.
        - Abnormality in information processing in the "old" brain.





a, a, a, **a**, *a*, **a**,

etc

Forms of info. in human communication

- Context dependency of information
  - “The lobster at no.18 is furious and about to burst into explosion.”

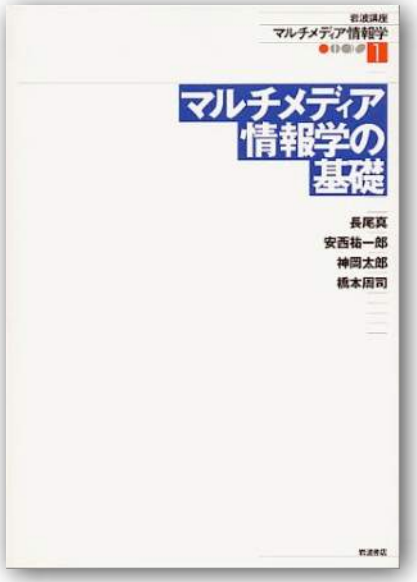
“Can you pass me the salt?”
“Yes, I can.”

Multimedia information and interaction bet. H and M

Nobuaki Minematsu

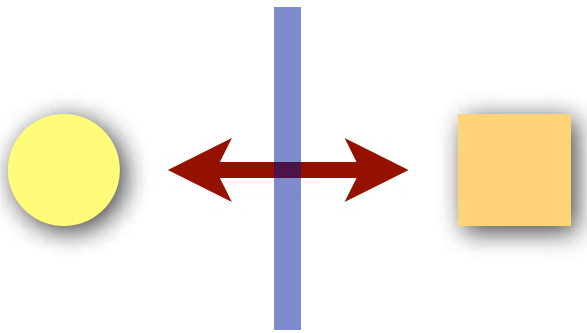
Today's menu

- Interaction and multimedia
  - User-friendliness and reality
- Role of multimedia interface
  - Direct interface and indirect (agent) interface
  - Metaphor and affordance
- Multimodal interface
  - Integration of different forms of input/output modalities
  - Adaptive interface
- Social interaction and multimedia
  - Human-likeness is needed?
  - Expressive (KANSEI, 感性) information and expressive interface
- Summary



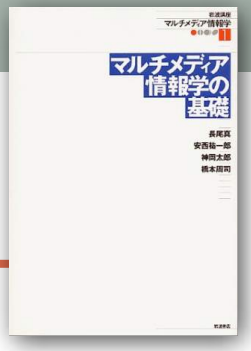
Interaction and multimedia

- Multimedia interface
  - Machine-side view of the interface
  - Capability of processing multiple forms of media info. is realized on machines.
- Multimodal interface
  - User (human)-side view of the interface
  - Multiple modalities based on the human five senses are available.
- Some issues of implementing the interface on machines
  - How to make effective and efficient interface through the use of multiple forms of media information? --> user-friendliness
    - Inadequate use may make the interface more complicated to human users.
  - How to get users to feel something “real” in the interface? --> reality
    - Unconscious processing that enables users to feel something “real”
- Various forms of multimedia/multimodal interface
  - Interface between human and machine
  - Interface between humans through a machine
    - Human communication via. a machine



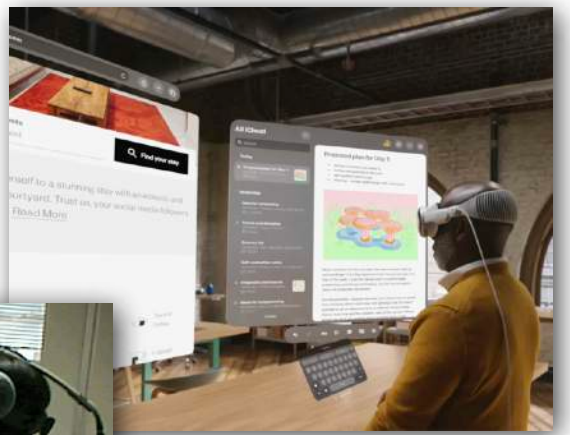
Role of multimedia interface

- Importance of multimedia interface
  - A machine with multiple functions tends to be complicated to users.
    - Requirement of “user-friendly” interface
    - Especially to aged users
  - What is the “user-friendly” interface?
    - Easiness to learn: less time needed to learn how to use that machine
    - Flexibility: capability to adapt (modify) the interface based on users and their context
    - Rapid response time: directly linked to user satisfaction
  - General principle to realize the “user-friendly” interface
    - Good understanding of human cognition and behaviors
    - Deep understanding including unconscious processing done by humans

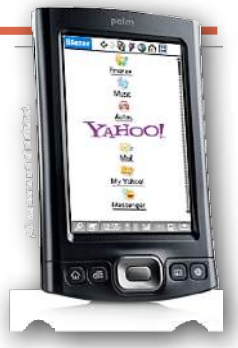


Role of multimedia interface

- Features of machines and devices with multimedia interface
  - Mobile
    - Mobile phones, wearable devices, etc
    - Small size: some difficulty to type text on
  - Ubiquitous
    - Home electronics, devices for handicapped, information traffic system (ITS), environmentally embedded system, etc
    - Technology for intelligent and social infrastructure
  - Virtual
    - Remote control through virtual reality technology / computer art
  - Cooperative
    - Groupware using multimedia interface
    - Cooperative operation among many users
  - Entertainment
    - Personalization of machines



Role of multimedia interface

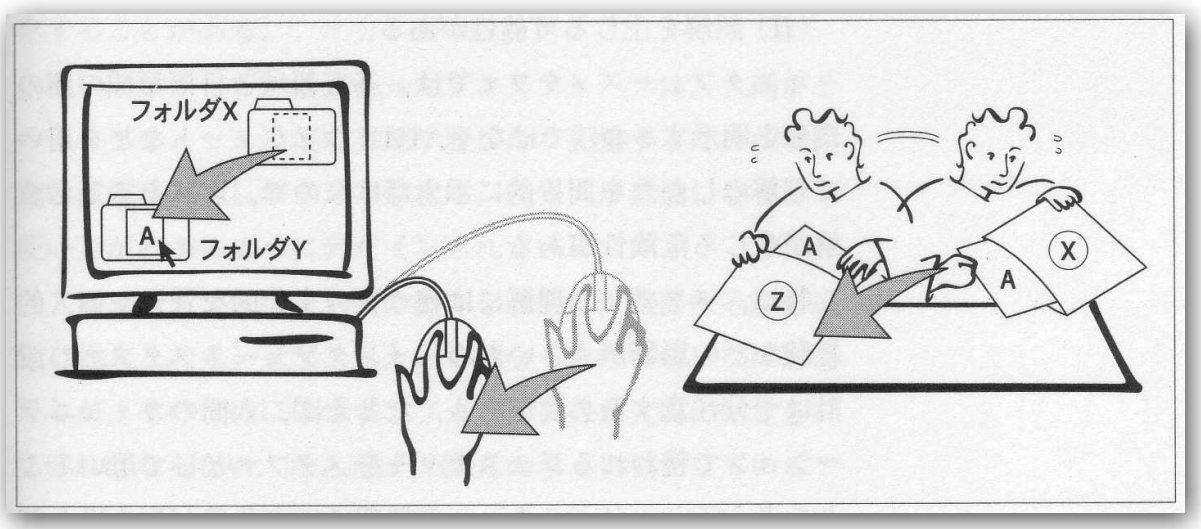


- Interface through direct control
  - Interface that gives users a feeling of directly controlling an object
    - Direct effects caused by users' action to a machine are instantly observed.
      - Word (WYSIWYG) vs. LaTeX
    - Tactile perception of remote things, which is virtually and technically synthesized.
    - Often user-initiative, where users themselves can decide what to do.

- Interface agent
  - Muti-function machines = difficult to use them directly
    - Agent = autonomous software that can operate those machines for naive users.
      - Often system-initiative, where a system guides a user to fulfill some specific tasks
      - Customizable / adaptive / autonomous
  - Problem
    - A machine is usually viewed as a black box.
      - Good balance between direct interface and indirect (agent) interface.

Role of multimedia interface

- Creation of user-friendliness through metaphor
  - Indication of a function by metaphor
    - Operations in a familiar domain are used as metaphor in an unfamiliar domain.
      - Experiences of sending “postal” mails help us learn how to send “electronic” mails.
      - Desktop metaphor
        - File, folder (drawer), trash box

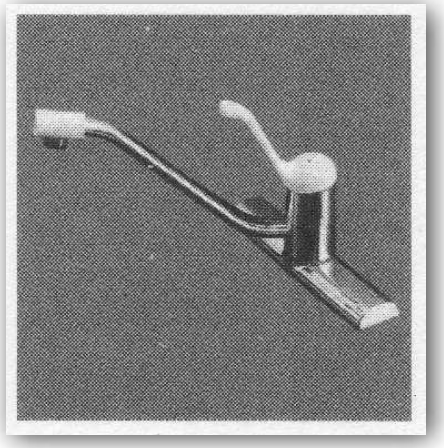


Role of multimedia interface

- Metaphor does not always work correctly.
  - Confusion in understanding metaphor
    - Ejection of a CD-ROM = throwing a CD-ROM away into the trash box?
  - Reasons of misunderstanding
    - Differences in culture and/or experience between users and developers
      - Inevitable when using metaphor interface
    - Developers' care sometimes turns out to be unwanted care.
      - Interface should be customizable due to users' characteristics.

Role of multimedia interface

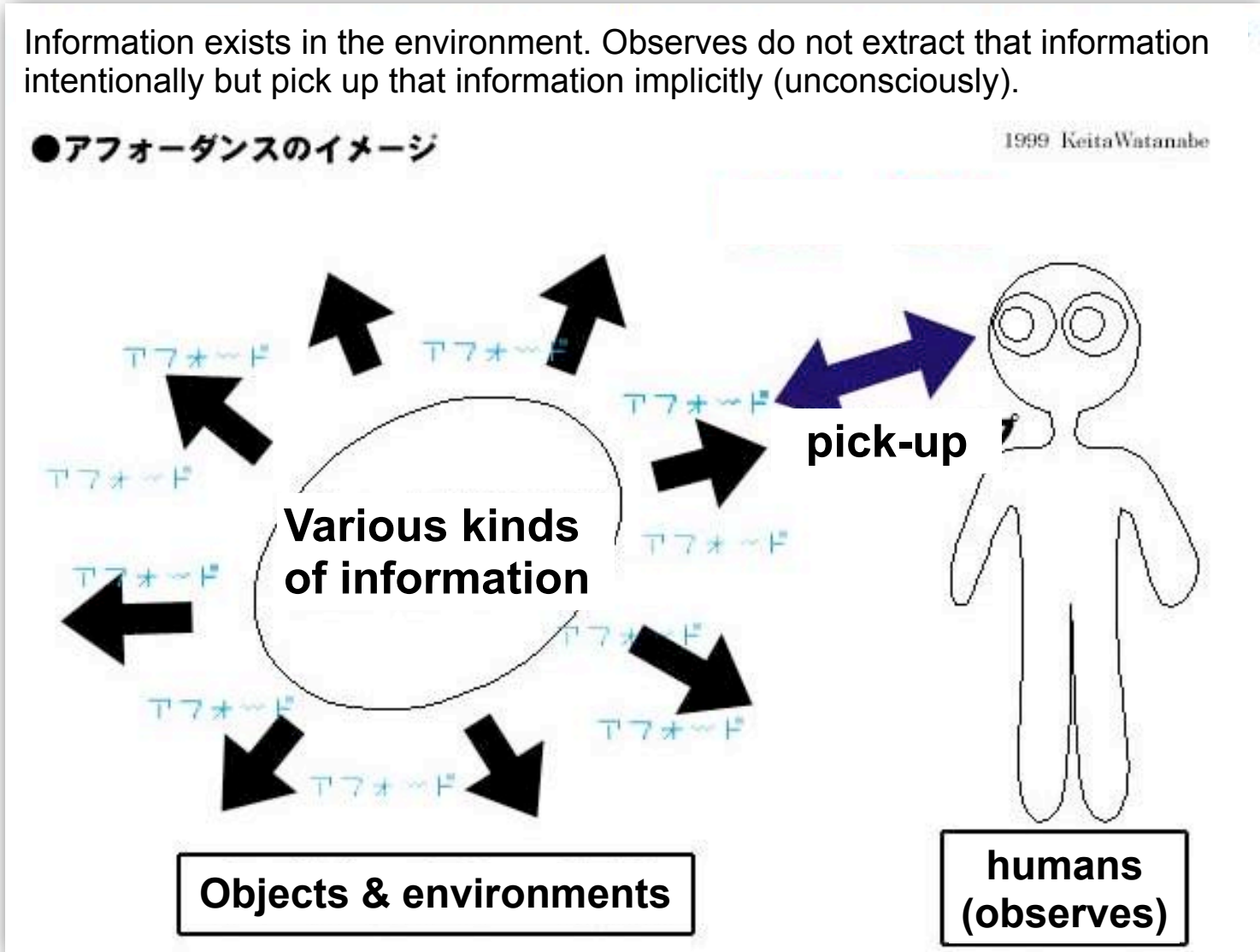
- Creation of user-friendliness through affordance
  - Operations or actions that an object can accept are sensed as the object's attributes.
  - Those attributes are often implicitly afforded to users by that object (affordance).
    - Affordance induces users to adequate operations to that object.
    - Originally proposed by J. Gibson, who is a professor of ecological psychology (1979)
  - Machines with good affordance
    - Appearance of those machines tells uses implicitly how to use them.
      - No explicit learning is required on how to use it and/or handle it.



Role of multimedia interface

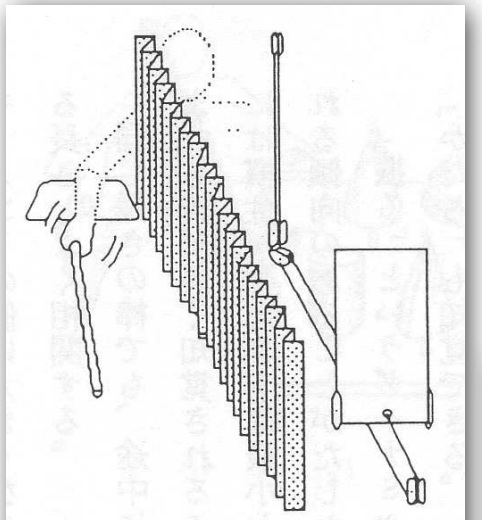
- Affordance defined in ecological psychology (生態心理学)

Information exists in the environment. Observes do not extract that information intentionally but pick up that information implicitly (unconsciously).

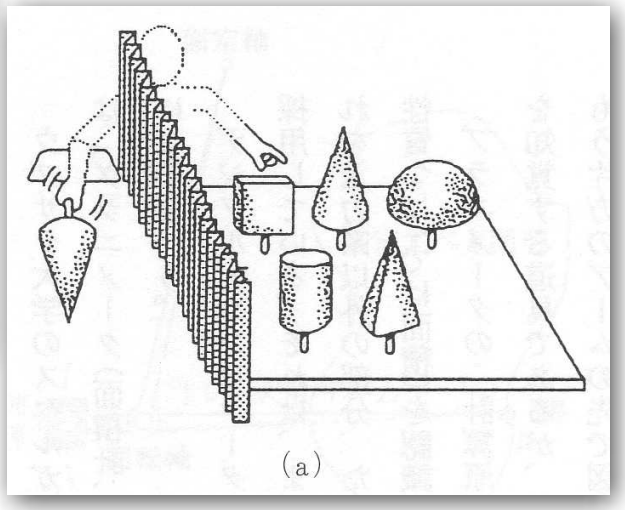


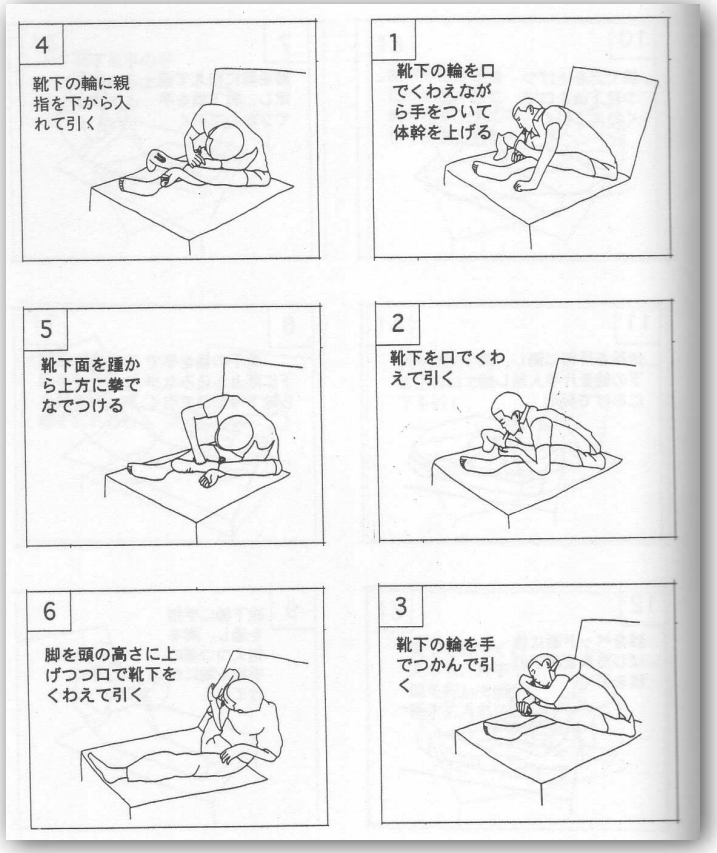
Role of multimedia interface

- Affordance defined in ecological psychology (生態心理学)
  - Information (attributes) that the environment tells implicitly.
  - The question is whether you can “pick up” affordance adequately.
    - Picking up is often done unconsciously and it is difficult to describe affordance explicitly.
    - Affordance study observes precisely human behaviors of picking up affordance.



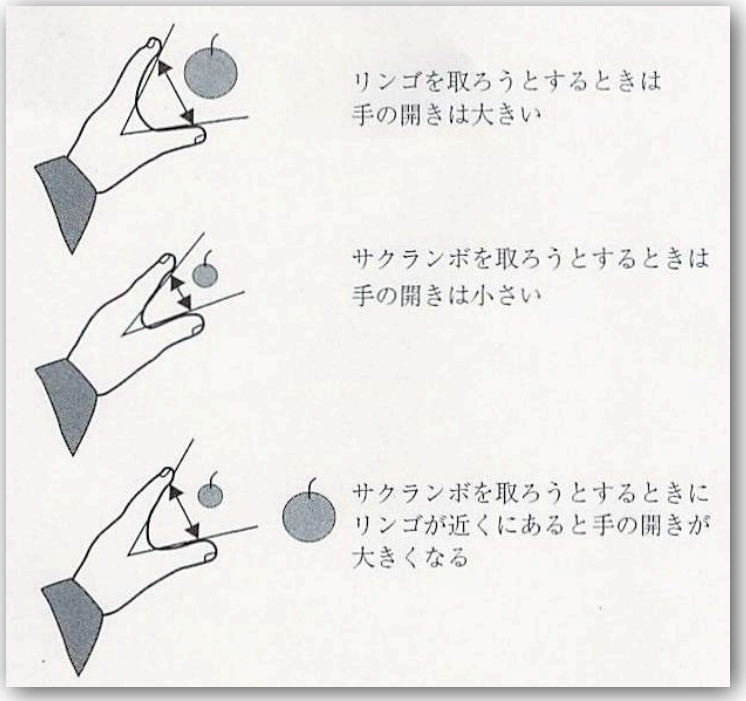
Perception of length of an object by shaking and swinging that object.





Affordance and neuron activities

- Intentional pinch and unintentional pinch
  - When a thing that one can pinch comes into one's sight,....
    - Castiello shows experimentally in a neuroscience study that when such a thing comes into one's sight, brain regions corresponding to pinching behaviors are activated. This is the case even when the observer does not intentionally pinch that thing.
    - "Neuron activities of possible actions caused only by seeing a thing can be considered as what is called affordance proposed by J. Gibson."

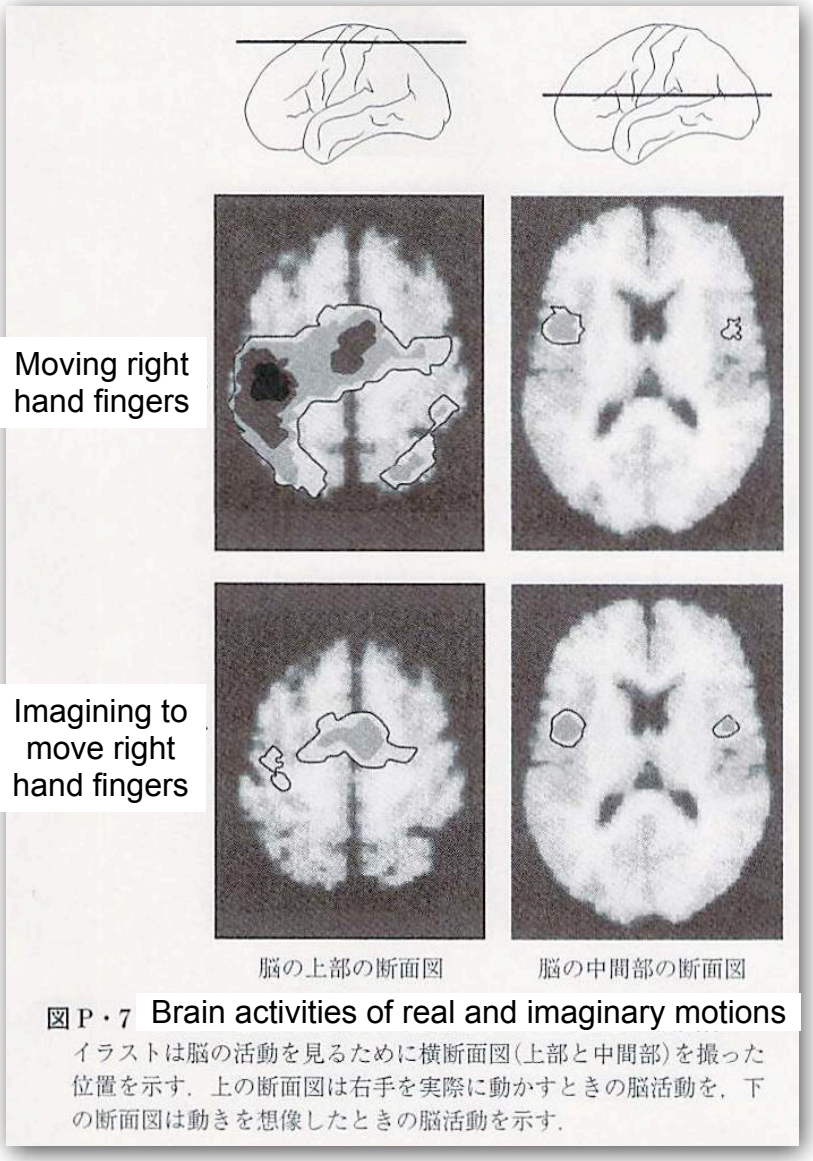


Imagination and execution of an action

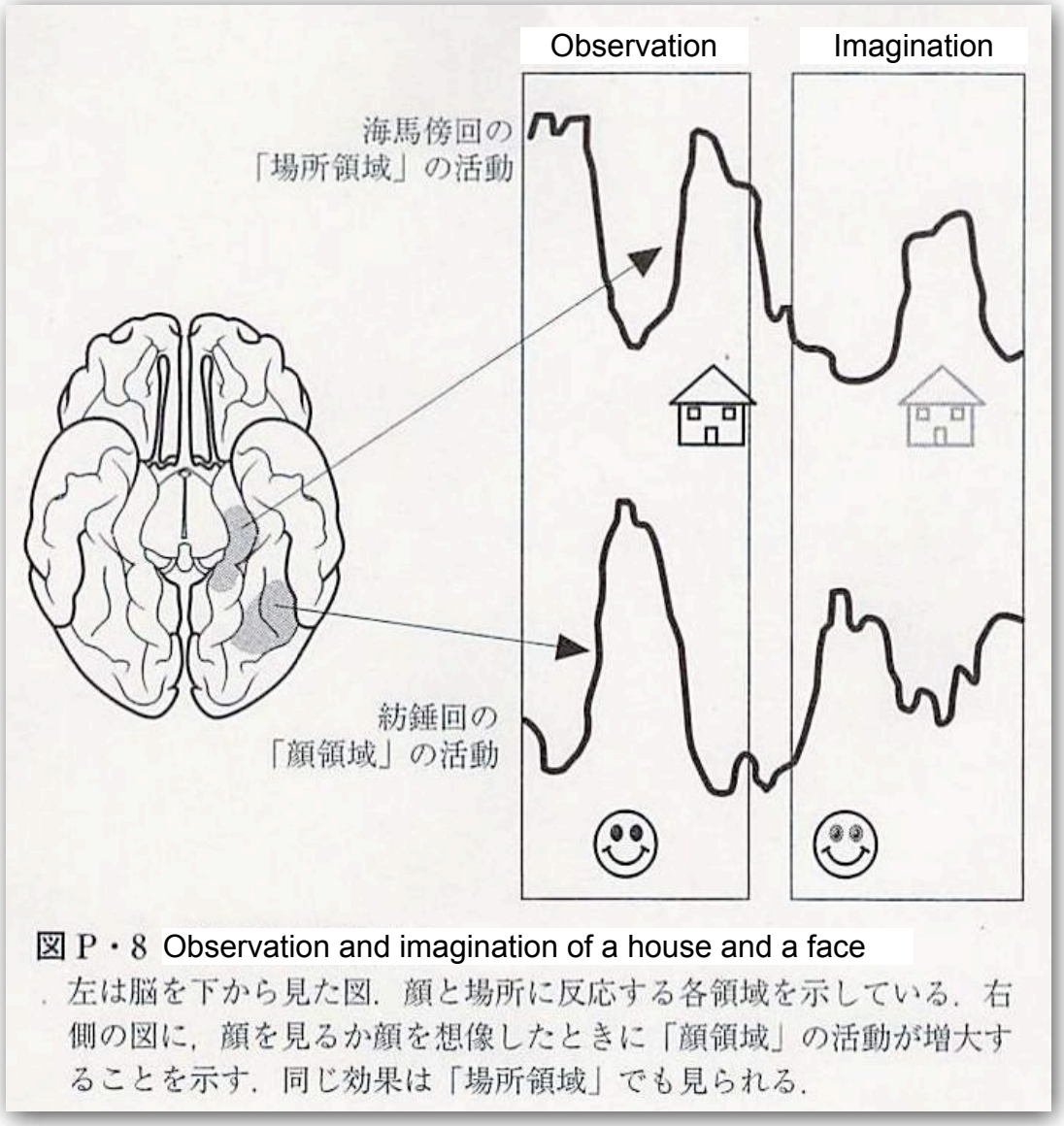
- What is the difference bet. imagining an action and executing that?
  - Similar brain activities are observed for both.
  - Then why we can discriminate between the two processes?
    - If exactly the same activation patterns are observed, discrimination is impossible.
  - Usually, we always imagine (predict) things that are about to happen.
    - Prediction (top-down processing) is always corrected or modified by physical observation (bottom-up processing).
    - No physical observation = no correction = world of only imagination = dreaming
    - No prediction = only physical observation = it become possible to tickle oneself to laugh by using one's own fingers. (One's own fingers are treated as others' fingers)
  - Power of imagination
    - Mental training done by professional athletes
      - Mental training gives almost equal effects to those by physical training.
      - No physical input (observation) leads to no correction.

Imagination and execution of an action

- What is the difference bet. imagining an action and executing that?



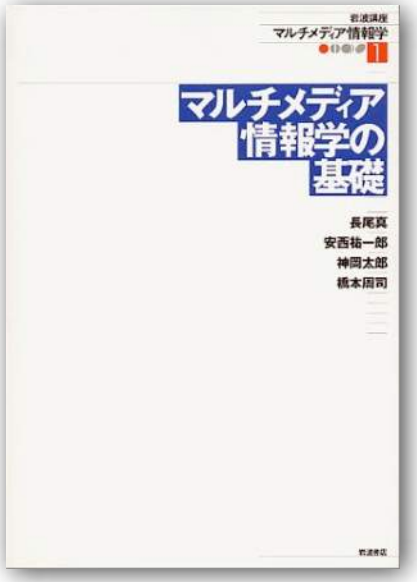
図P・7 Brain activities of real and imaginary motions
イラストは脳の活動を見るために横断面図(上部と中間部)を撮った位置を示す．上の断面図は右手を実際に動かすときの脳活動を，下の断面図は動きを想像したときの脳活動を示す．



図P・8 Observation and imagination of a house and a face
左は脳を下から見た図．顔と場所に反応する各領域を示している．右側の図に，顔を見るか顔を想像したときに「顔領域」の活動が増大することを示す．同じ効果は「場所領域」でも見られる．

Today's menu

- Interaction and multimedia
  - User-friendliness and reality
- Role of multimedia interface
  - Direct interface and indirect (agent) interface
  - Metaphor and affordance
- Multimodal interface
  - Integration of different forms of input/output modalities
  - Adaptive interface
- Social interaction and multimedia
  - Human-likeness is needed?
  - Expressive (KANSEI, 感性) information and expressive interface
- Summary

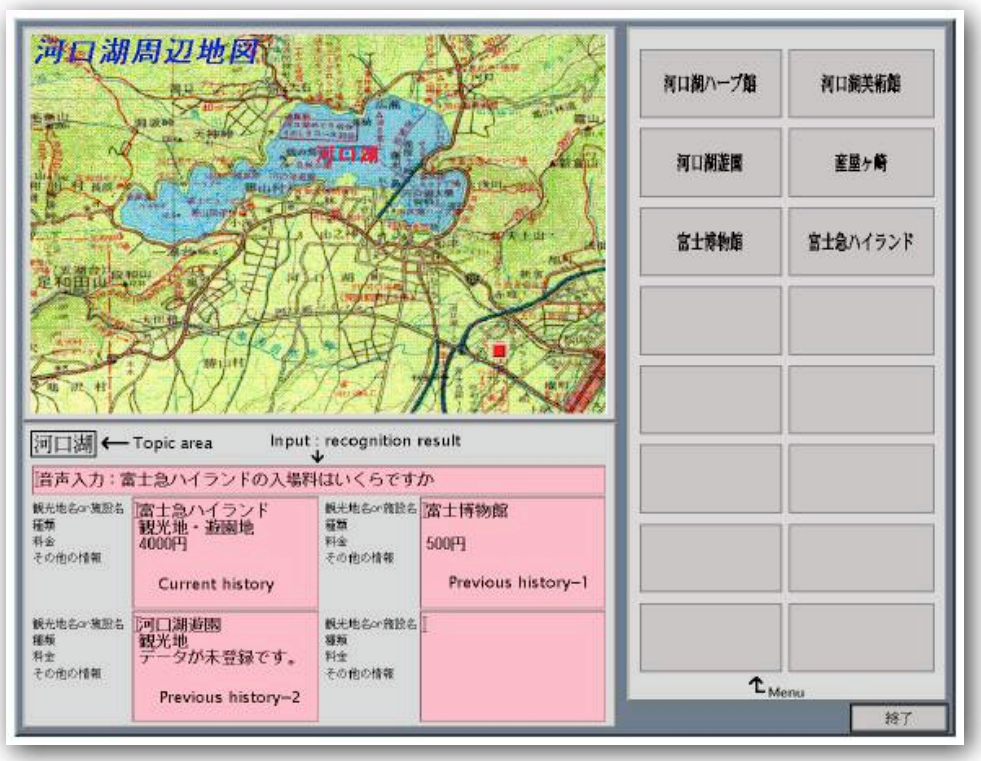


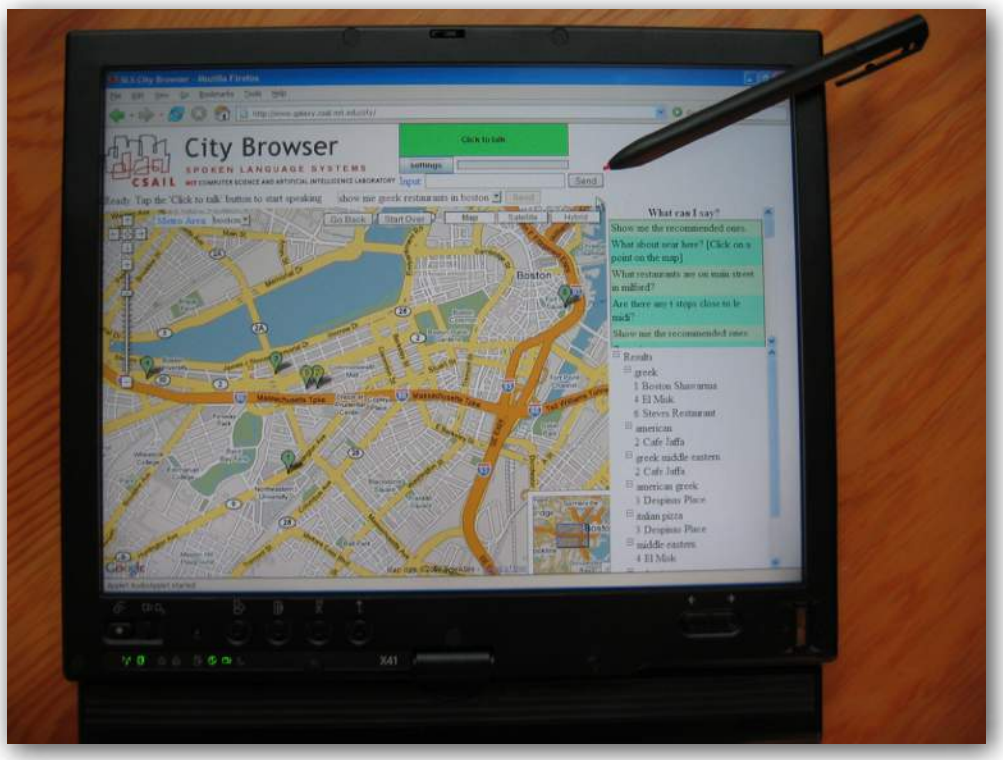
Multimodal interface

- Features of multimodal interface
  - Efficiency and effectiveness
    - Text only / text and speech / text, speech, and images
  - Redundancy and reduced ambiguity
    - Multiple channels between system and user make info. transmission more reliable.
  - Cognitive load imposed on users
    - A good combination of multiple channels can reduce cognitive loads.
  - Naturalness
    - Human-to-human communication often use multiple channels for info. exchange.
  - Variability and customizability
    - Can be modified due to age, gender and tastes of users
  - Synergy
    - Different kinds of information can be transmitted only by combining multimodal channels.
      - Sign languages and facial expressions
    - Complementary and effective use of multiple channels and modalities

Multimodal interface

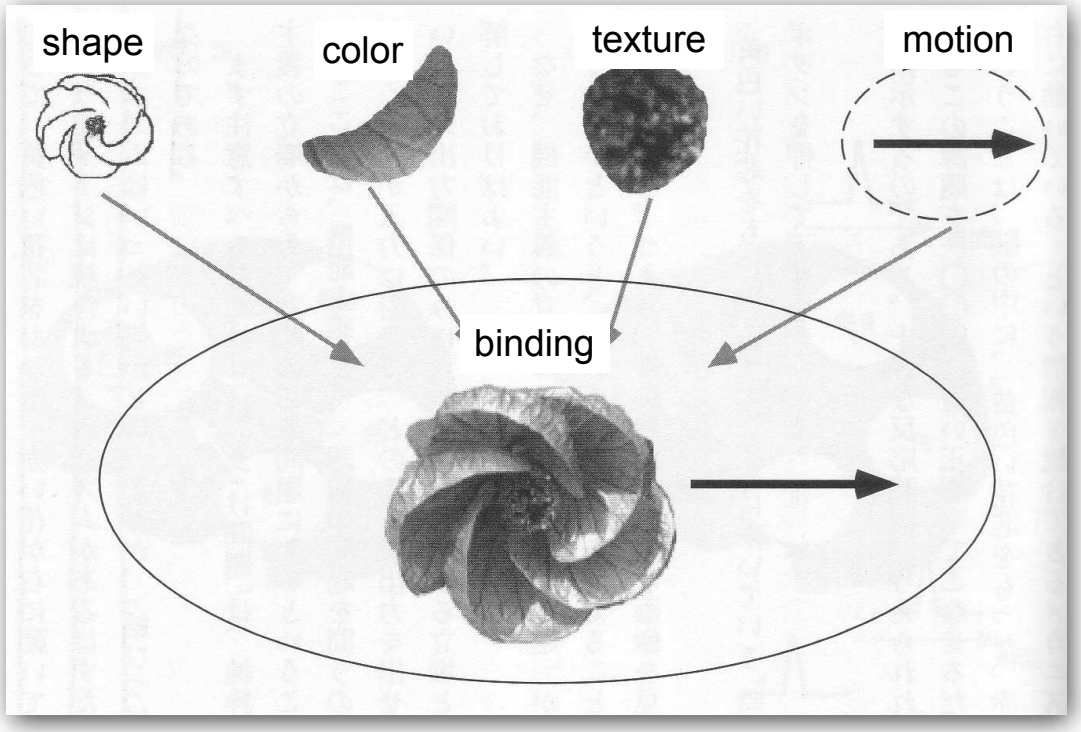
- Examples of the multimodal interface
  - Integration of various input modalities
    - keyboard (text), pointing device, speech, touch screen, still/moving images, etc.
  - How to integrate inputs of different modalities?
    - Temporal and spatial integration of inputs through different modalities
    - How to “bind” them into one?

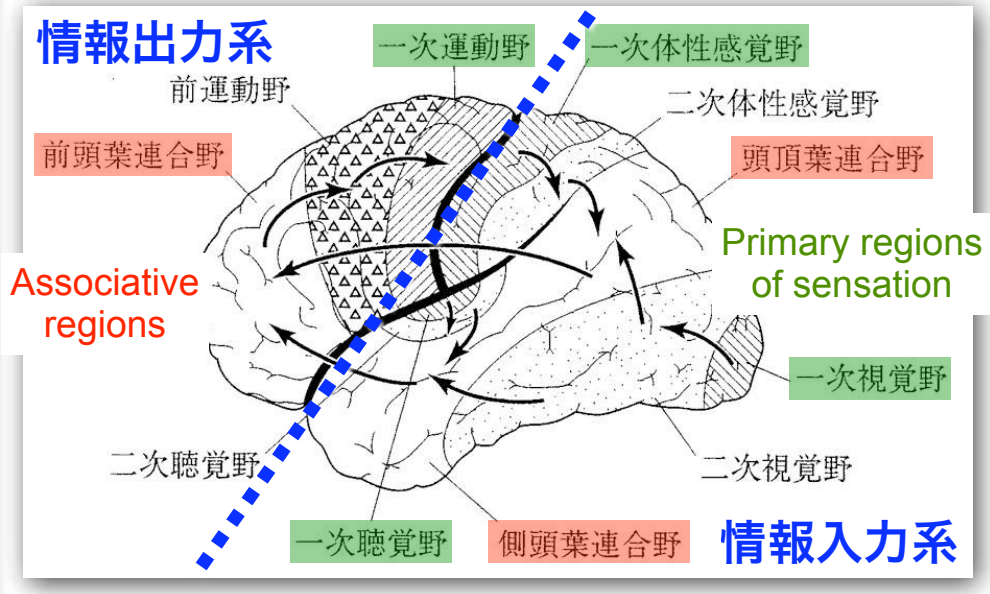


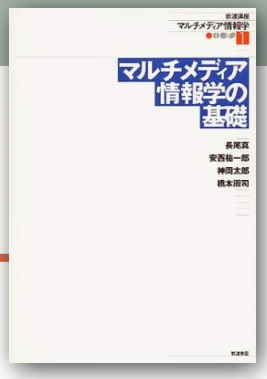


Multimodal interface

- The “binding” problem of the brain
  - Something rounded, red, smooth is moving to the right from the left.
    - Attributes of shape, color, texture, and motion are processed in different regions of the brain.
    - One object is decomposed into separate attributes, which are bound to be one.
    - These attributes are integrated into one “image” on the associative region (連合野).
  - Unconscious processing on the brain

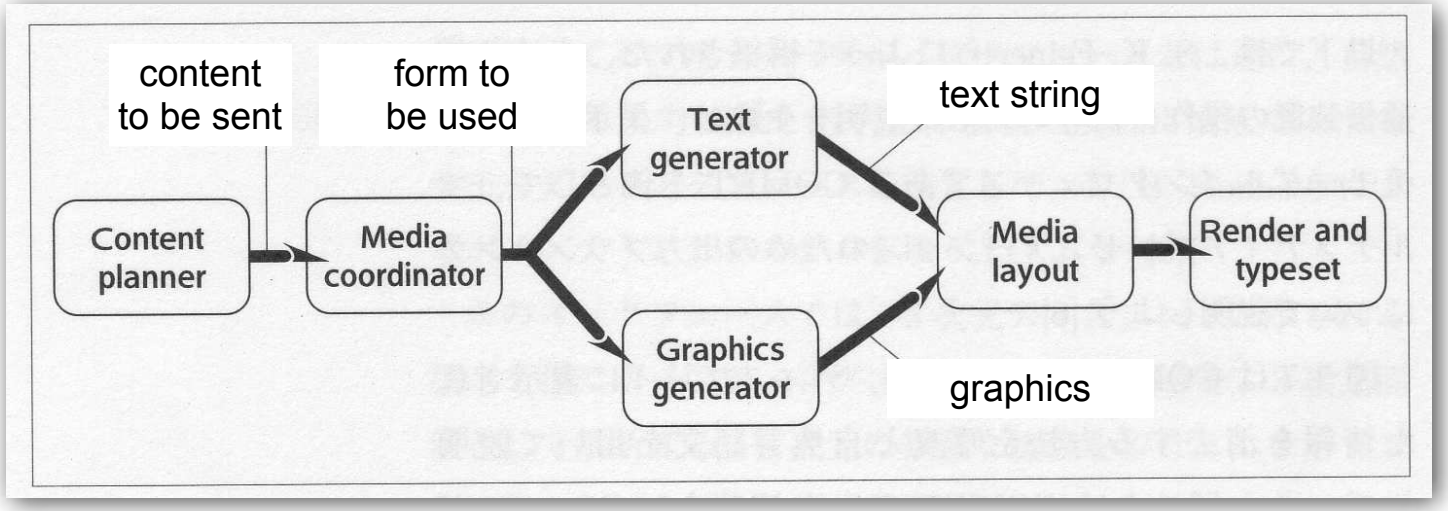






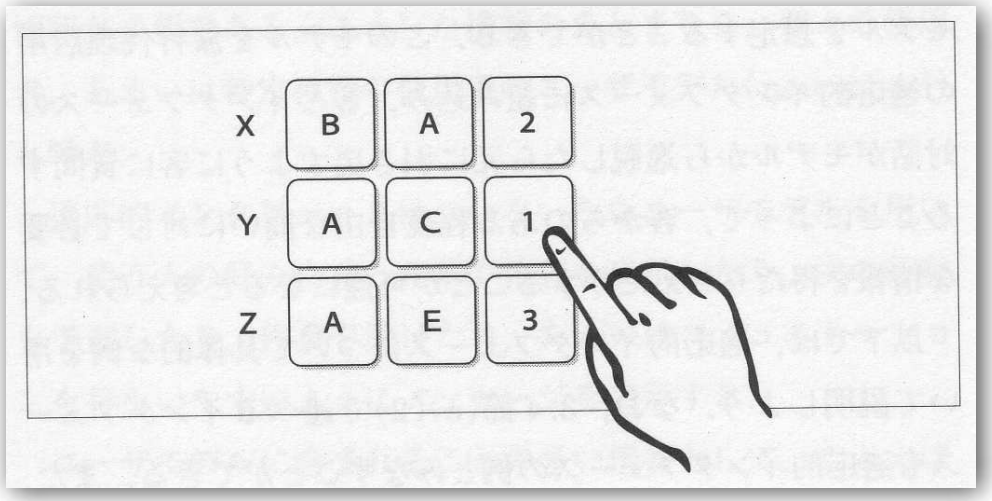
Multimodal interface

- Examples of the multimodal interface
  - Integration of various output modalities
    - Good planning on which channel to use is required before presenting some results.
    - Various factors have to be considered in the planning
      - Amount of text output, size of the screen, environmental noise, etc.
      - Planning should care about personal characteristics of users such as age and gender.
    - Output modules of different modalities have to be driven based on integrated (universal) representation of information



Multimodal interface

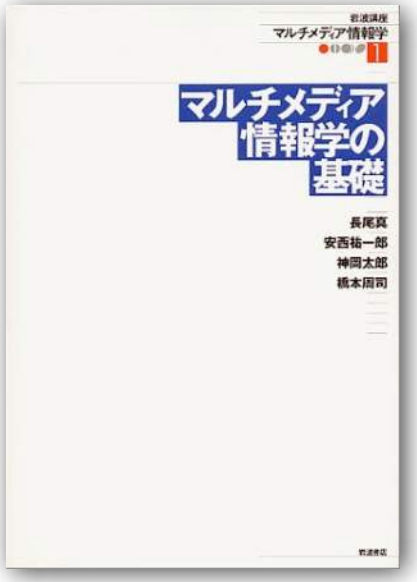
- Examples of the multimodal interface
  - Adaptive interface
    - User model
      - Features of the interface can be modified dynamically depending on users' situation.
      - Static modification based on static features of users such as their knowledge.
    - Dialogue model
      - Task-oriented dialogue sequence templates are prepared and used to interpret user's input.
        - The same action from a user can be interpreted differently depending on the dialogue history
      - Should treat unexpected users' action properly.
        - The templates do not always works well and this unexpected situation has to be solved properly.

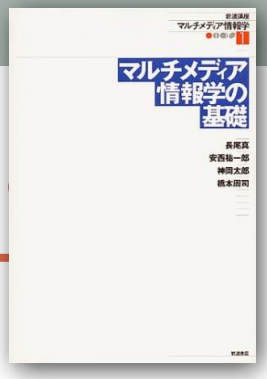


Interpretation of user's actions through spoken language and finger pointing

Today's menu

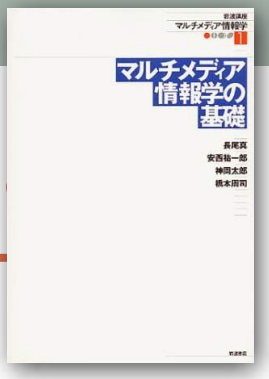
- Interaction and multimedia
  - User-friendliness and reality
- Role of multimedia interface
  - Direct interface and indirect (agent) interface
  - Metaphor and affordance
- Multimodal interface
  - Integration of different forms of input/output modalities
  - Adaptive interface
- Social interaction and multimedia
  - Human-likeness is needed?
  - Expressive (KANSEI, 感性) information and expressive interface
- Summary





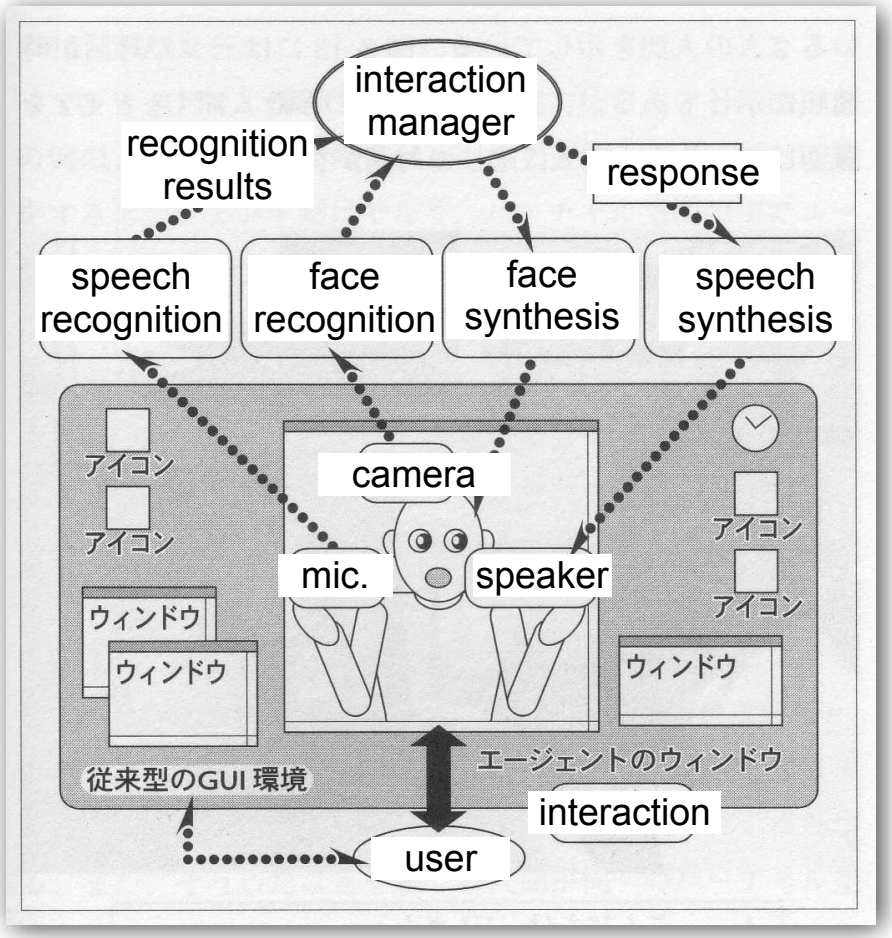
Social interaction and multimedi

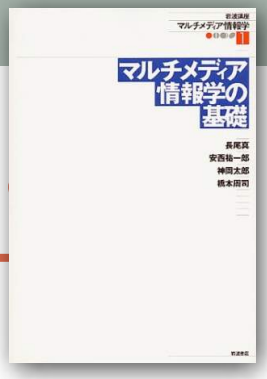
- What is social interaction?
  - Interaction caused in the context of social relations
    - One individual has to play various social roles due to social environments.
      - Professor, committee member, father, husband, adult male, Japanese, etc
    - Interaction bet. an individual and another, bet. an individual and a group, and bet. a group and another.
  - Personification of machines (agents) in the multimedia interface
    - Realization of “social” interaction between a human and a machine
    - What kind of roles can be realized on machines?



Social interaction and multimedi

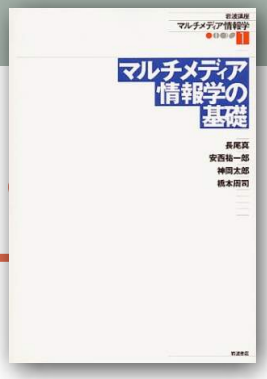
- Personified (anthropomorphic) agents
  - Computer software agents with human appearance
  - From agents on computer screens to human-shaped robots





Social interaction and multimedi

- Avatar agents in a cyberspace
  - A personified agent who take the role of a specific user in a cyberspace.
    - It is “you” in the cyberspace.
  - A virtual world for lots of avatars to communicate with each other in.



Social interaction and multimedi

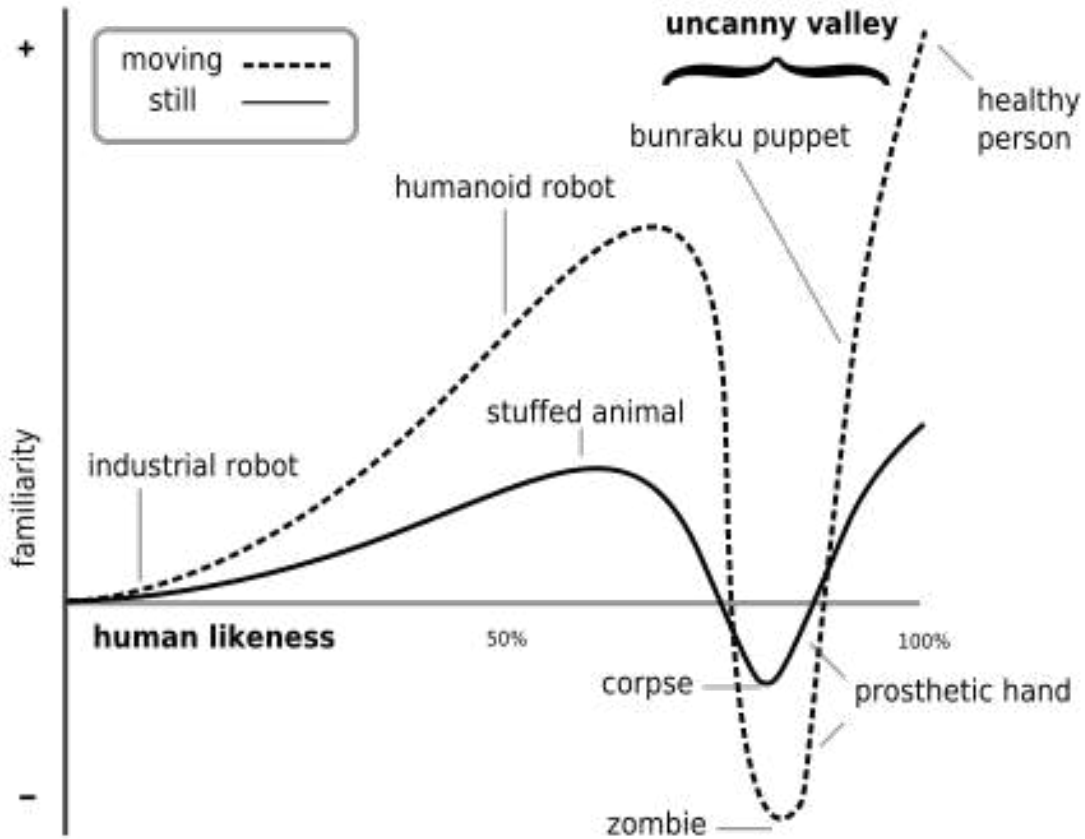
- Some examples
  - Personified computer agent
  - Secretary robot agent
  - A presentation robot
  - Comparison between more than 20 years ago and a few years ago

Social interaction and multimedi

- Features of personified agents
  - Merits
    - Create such an atmosphere that a user feel as if the user is talking to a human.
    - Non-verbal communication is used, which is often found in H-to-H communication.
    - Users can predict better the machine behavior through performance of the agent.
  - Demerits
    - Really human-like? Somewhat unnatural, strange, weird, uncanny (=weird, 不気味)
      - Problem of the uncanny valley
    - Users may use only verbal expressions for explicit and unambiguous communication.
  - The essential question to raise
    - Lots of questions remain to understand human perception and behaviors.
      - In this situation, can researchers (engineers) simulate humans well?
    - The well-know "frame" problem of AI, which is sometimes discussed in relation to autism

Social interaction and multimedi

- The uncanny valley

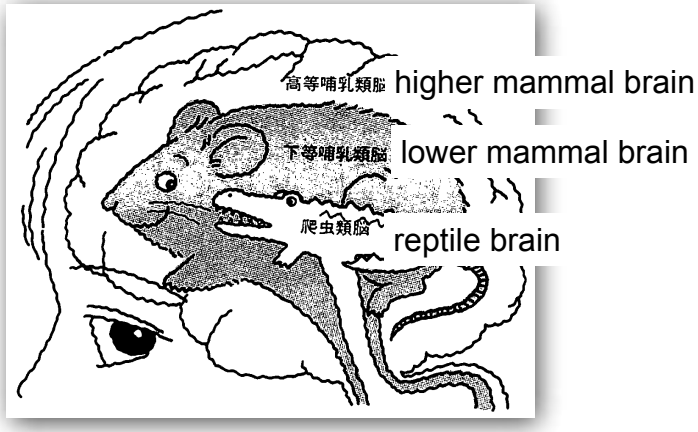


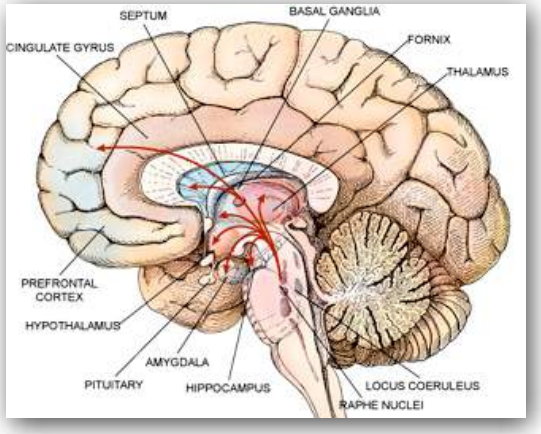
Social interaction and multimedi

- Features of personified agents
  - Merits
    - Create such an atmosphere that a user feel as if the user is talking to a human.
    - Non-verbal communication is used, which is often found in H-to-H communication.
    - Users can predict better the machine behavior through performance of the agent.
  - Demerits
    - Really human-like? Somewhat unnatural, strange, weird, uncanny (=weird, 不気味)
      - Problem of the uncanny valley
    - Users may use only verbal expressions for explicit and unambiguous communication.
  - The essential question to raise
    - Lots of questions remain to understand human perception and behaviors.
      - In this situation, can researchers (engineers) simulate humans well?
    - The well-know frame problem of AI, and autism

Logical and expressive

- Logical information and expressive information
  - Factors (bases) to describe expressive information
    - Facial expressions (as example)
      - 6 factors of surprise, fear, dislike, anger, happiness, and sorrow
      - A still debatable problem in psychology
    - Theory of mind [D. Premack et. al.'78]
      - The ability to attribute mental states to oneself and others and to understand that others have different mental states than one's own.
        - Different individuals have different minds.
        - Those who don't have theory of mind have difficulty in understanding this fact.
      - One of the theories that explains the cause of autism (自閉症) [S. Baron-Cohen'91]
        - Difficulty in reading the mind of others and understanding that everybody has one's own mind.
        - Difficulty in reading the facial expressions.
        - Abnormality in information processing in the "old" brain.





a, a, a, **a**, *a*, **a**,

etc

Social interaction and multimedi

- The frame problem of AI and autism
  - The frame problem
    - Any robot has definite power of computation and, in principle, it has difficulty of handling every possible thing (problem) that can happen in the real world.
    - Humans can ignore many things without consciously dealing with them.
    - “Buy a hamburger in that McDonald shop!”
      - Many trivial but unexpected things can happen but humans ignore these things without being aware that they ignored them.
        - An awareness test
      - Robots can ignore them only by “trying” to ignore them.
    - One of the characteristics of autistics : cannot ignore things
      - “Our brain cannot go through” written by an autistic author.
        - Autism = constipation (便秘) of information
        - Autistics tend to pay attention to any sensory input.
        - Difficult to pick up selectively meaningful inputs only.
      - Similarity in behaviors between robots and autistics.

Robots and autistics

The 4th IEEE International Conference on Development and Learning

Workshop & Tutorial

Workshop (Tuesday, July 19th, 9:30-15:00) workshop page

"Social Cognition: From Humans to Robots"

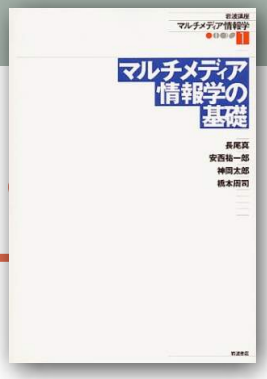
Confirmed speakers:
Prof. Javier Movellan (UCSD)
Dr. Yukiyasu Kamitani (ATR)
Dr. Thierry Chaminade (ATR/UCL)

Tutorial (Tuesday, July 19th, 9:30-15:00)

"The developing child with autism: evidences, speculations and vexed questions"

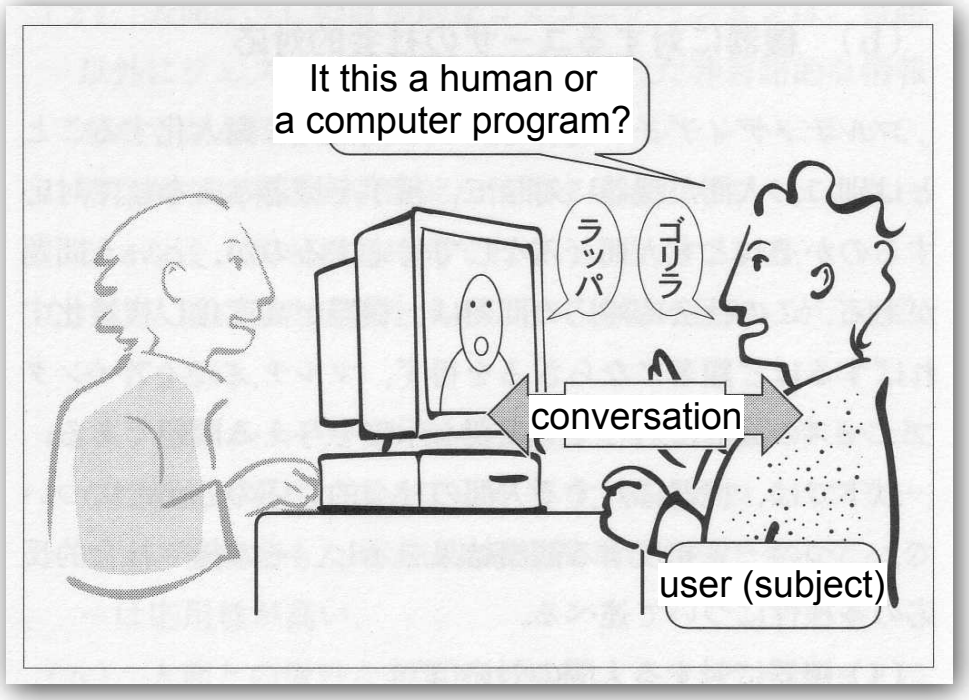
Jacqueline Nadel, PhD, Research director
(French National Centre of Scientific Research, L'Hôpital de la Salpêtrière, Paris, France)

conference. The conference is to be held right after the Ninth RoboCup International Competitions and Conferences, RoboCup-05 (at the same site) and right before the Fifth International Workshop on Epigenetic Robotics (EpiRob-05, in Nara). Several joint sessions/events between these meetings will be organized.



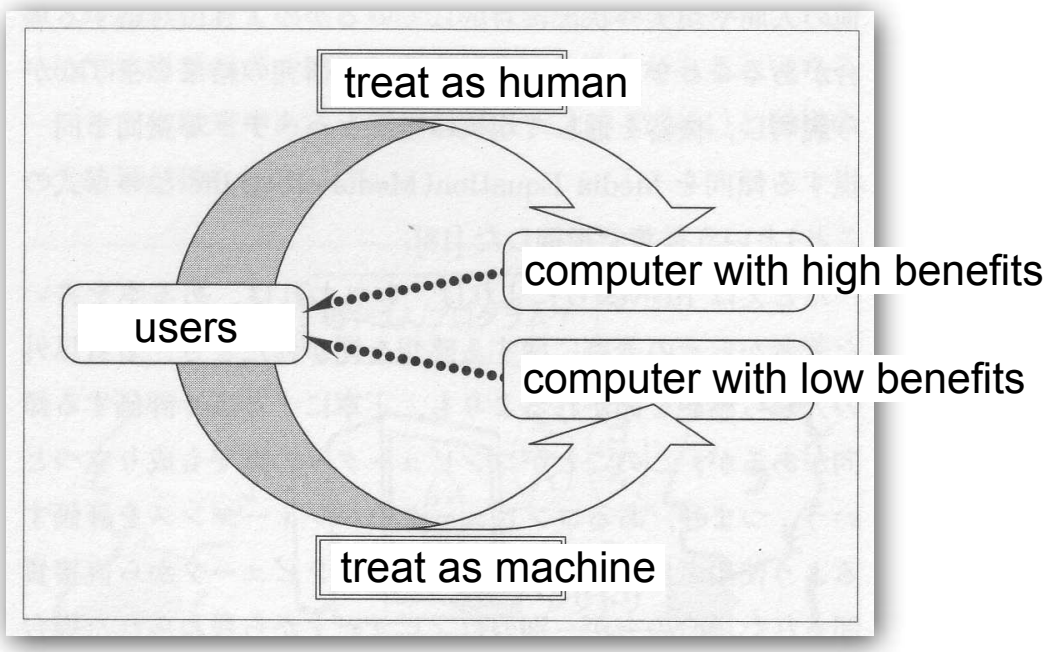
Social interaction and multimedi

- Users' (social) responses to machines
  - Perception of a human operator in a machine
    - Users' responses when they assume that a human operator is controlling the machine at the background.
    - Users' responses when they assume that the machine is working completely automatically.
  - Two extreme cases
    - Non-human appearance with assumption of a human operator
    - Human appearance with no assumption of a human operator



Social interaction and multimedia

- Users' (social) responses to machines
  - Differences in users' responses bet. when perceiving a human and when not
  - Users' active personification of a machine
    - Users tend to treat a machine like a human (living object) more when they receive more benefits from the machine.
      - Personification is often done.
    - Human-shape (appearance) is not always needed.
  - How to make users perceive a human in a machine?



Social interaction and multimedia

- Personified mobile phone
  - Human shape is needed or not?
    - Humanoid mobile phone project (Prof. Ishiguro @ ATR)

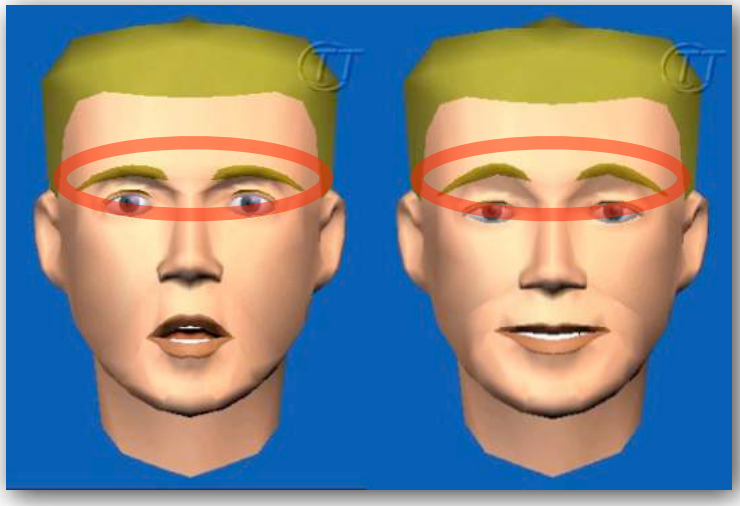
    - Siri, dialogue-based information retrieval system (Apple)

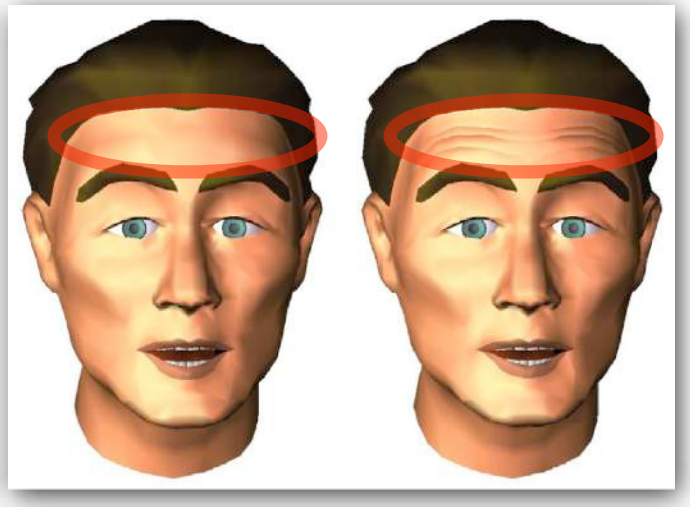
Social interaction and multimedia

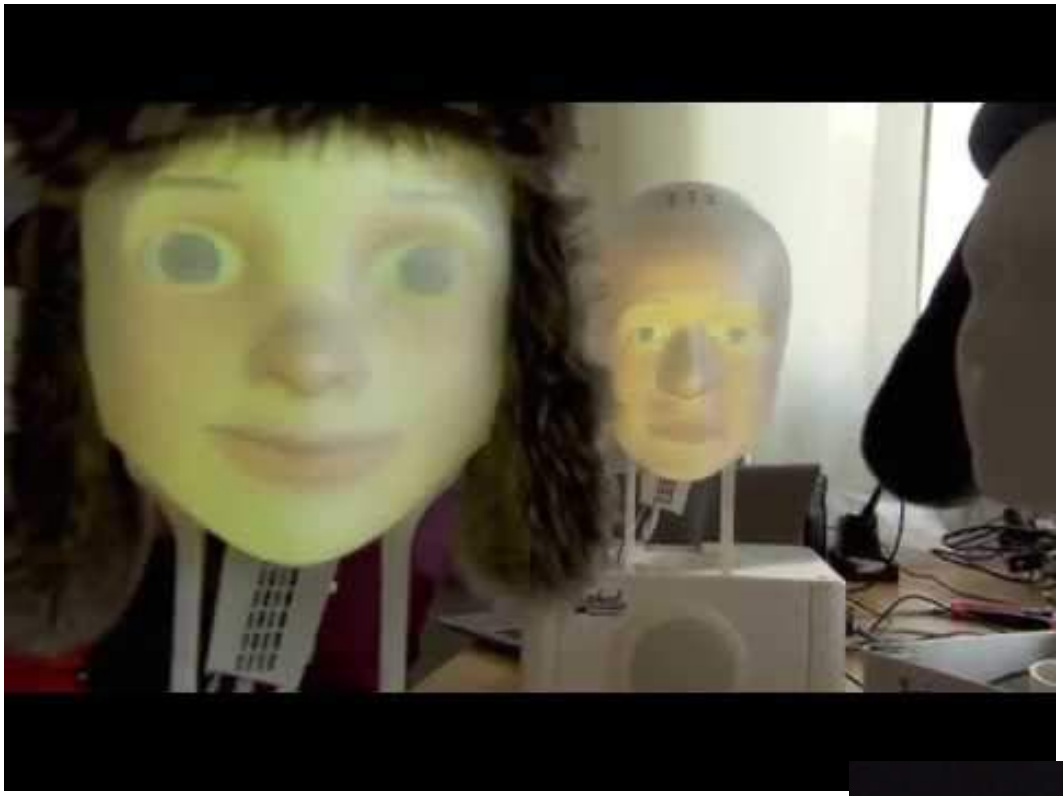
- Expressive (emotional) interaction/interface
  - Sensing users' emotional actions and generating reactions that will change user's emotional state.
  - How to measure emotional actions from users?
    - Physiological and/or physical observation
      - Blood pressure, body temperature, heartbeat, electric resistance of the skin, etc
      - Body motions in gesture and prosodic motions in utterances
      - Lexical choice, style of speaking, etc
  - How to generate emotional responses to users?
    - Symbolically represented emotional statements are converted into responses with different modalities.
      - Use of seven fundamental emotions of anger, fear, disgust, contempt, joy, sadness, and surprise.
      - Context-dependent use of different modalities
      - Good combination of emotional reactions and non-emotional reactions

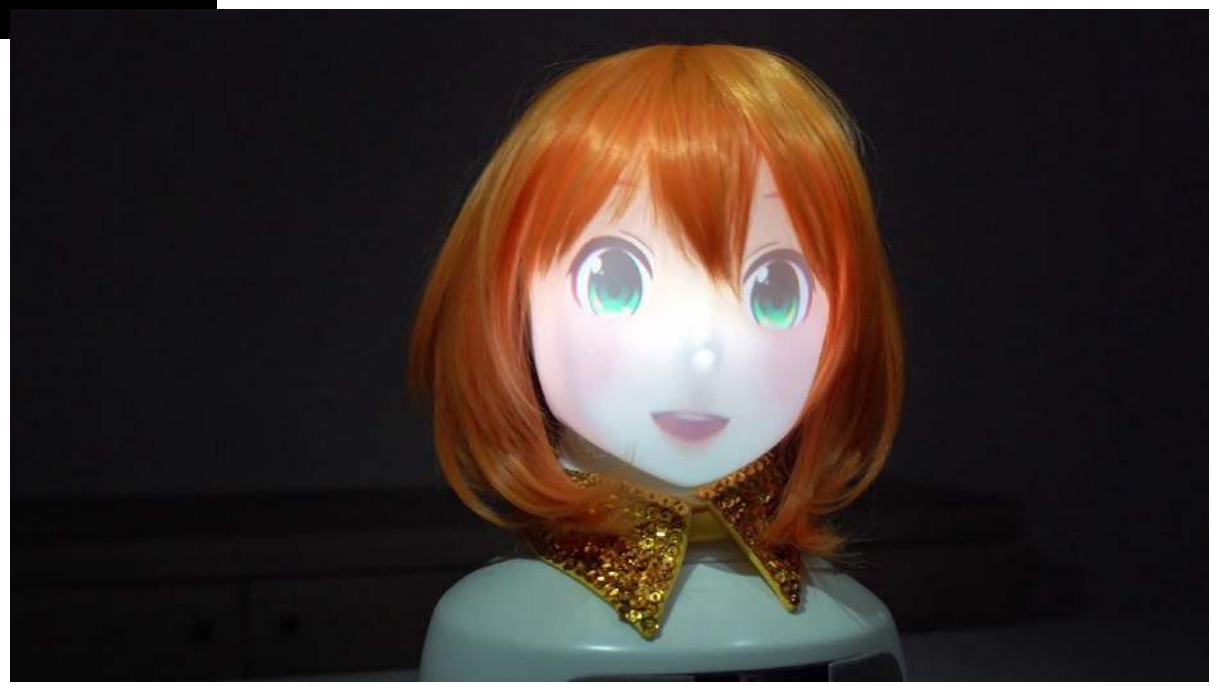
Social interaction and multimedia

- Examples of facial and expressive interface
  - Check eyebrows, view direction, face direction, etc



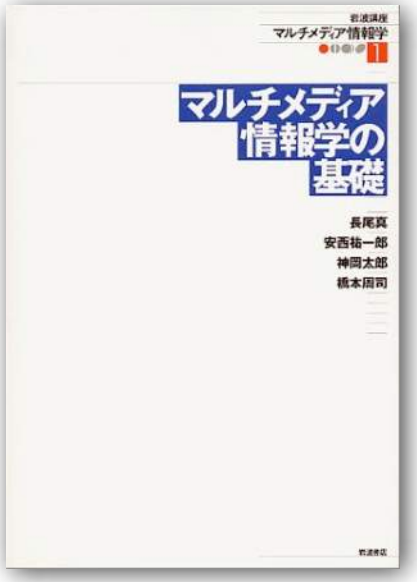






Today's menu

- Interaction and multimedia
  - User-friendliness and reality
- Role of multimedia interface
  - Direct interface and indirect (agent) interface
  - Metaphor and affordance
- Multimodal interface
  - Integration of different forms of input/output modalities
  - Adaptive interface
- Social interaction and multimedia
  - Human-likeness is needed?
  - Expressive (KANSEI, 感性) information and expressive interface
- Summary



Recommended books