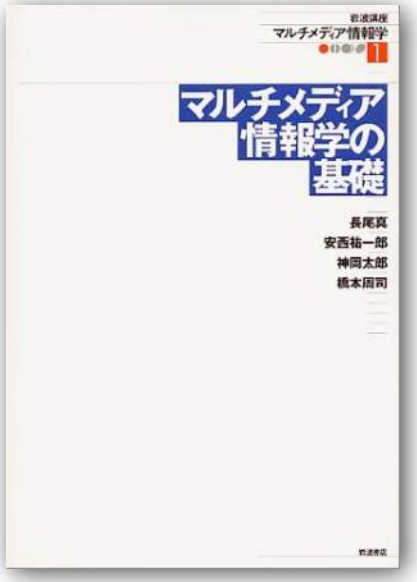
Cognitive Media Processing #3

Nobuaki Minematsu

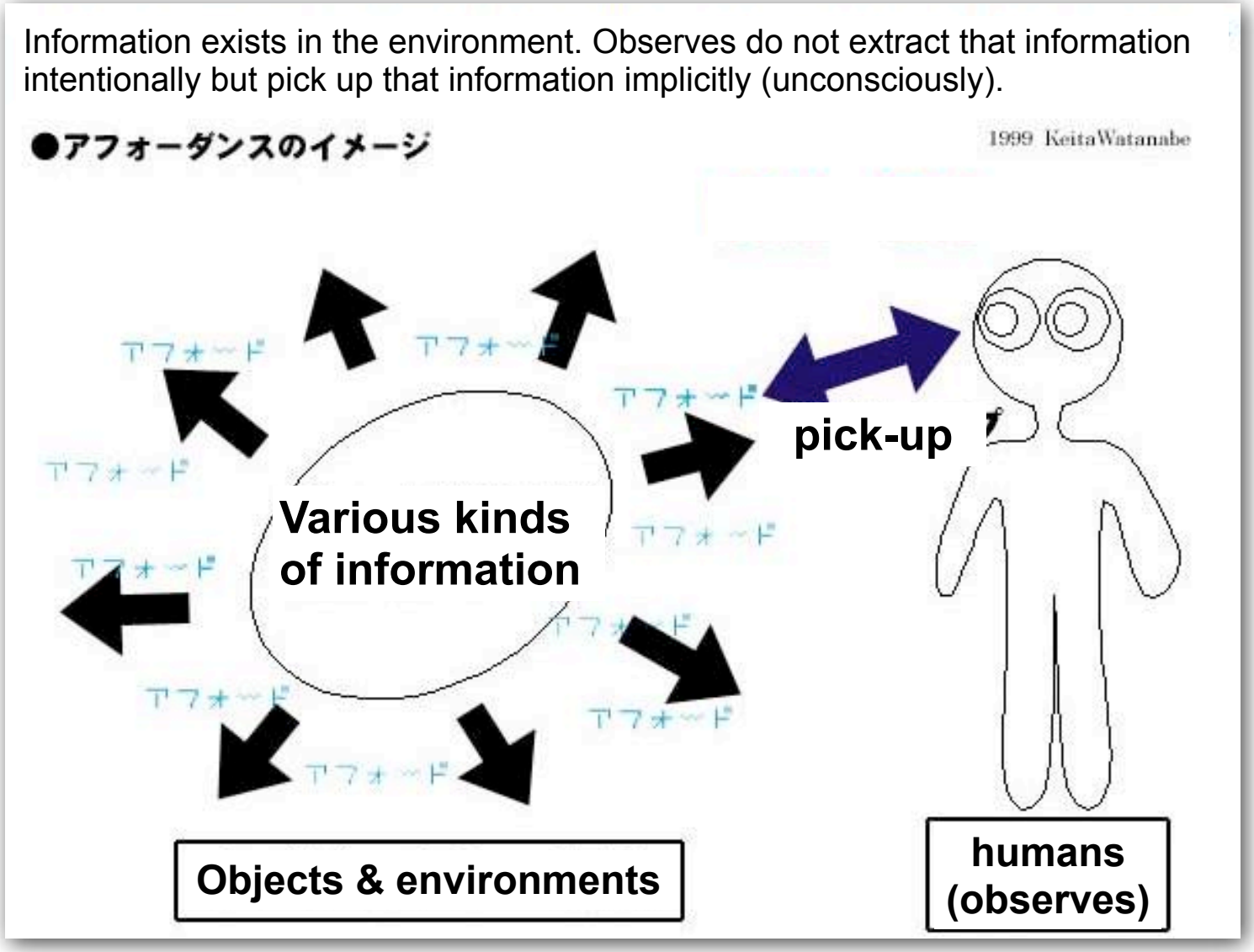
Menu of the last lecture

- 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



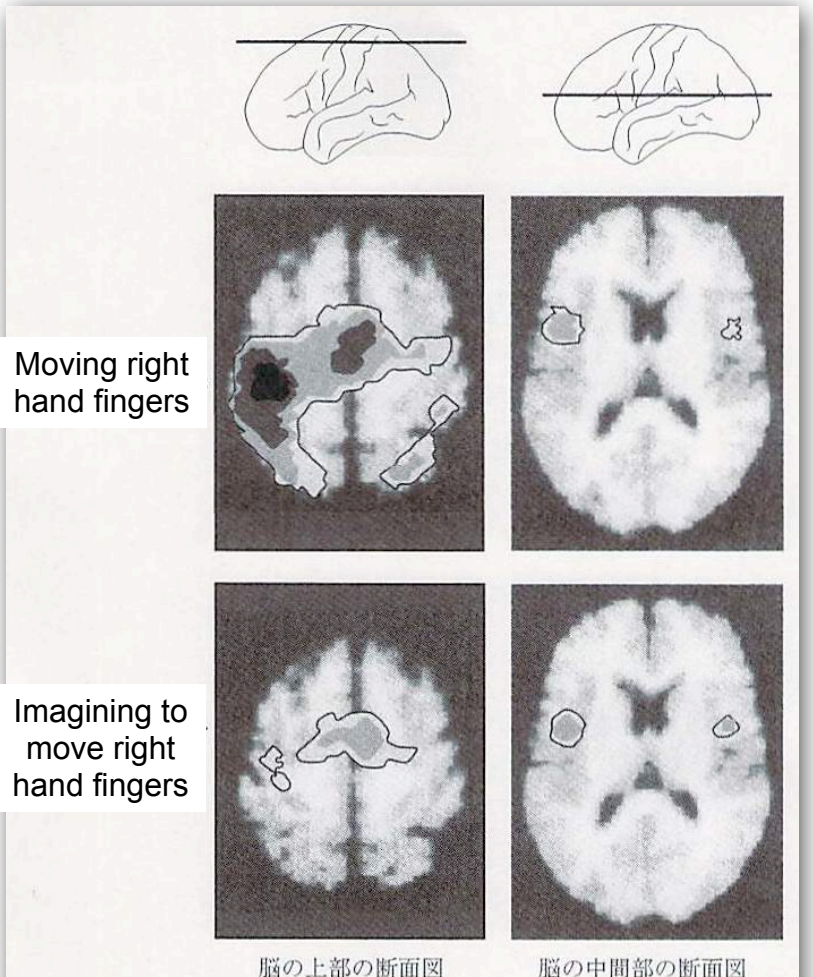
Role of multimedia interface

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

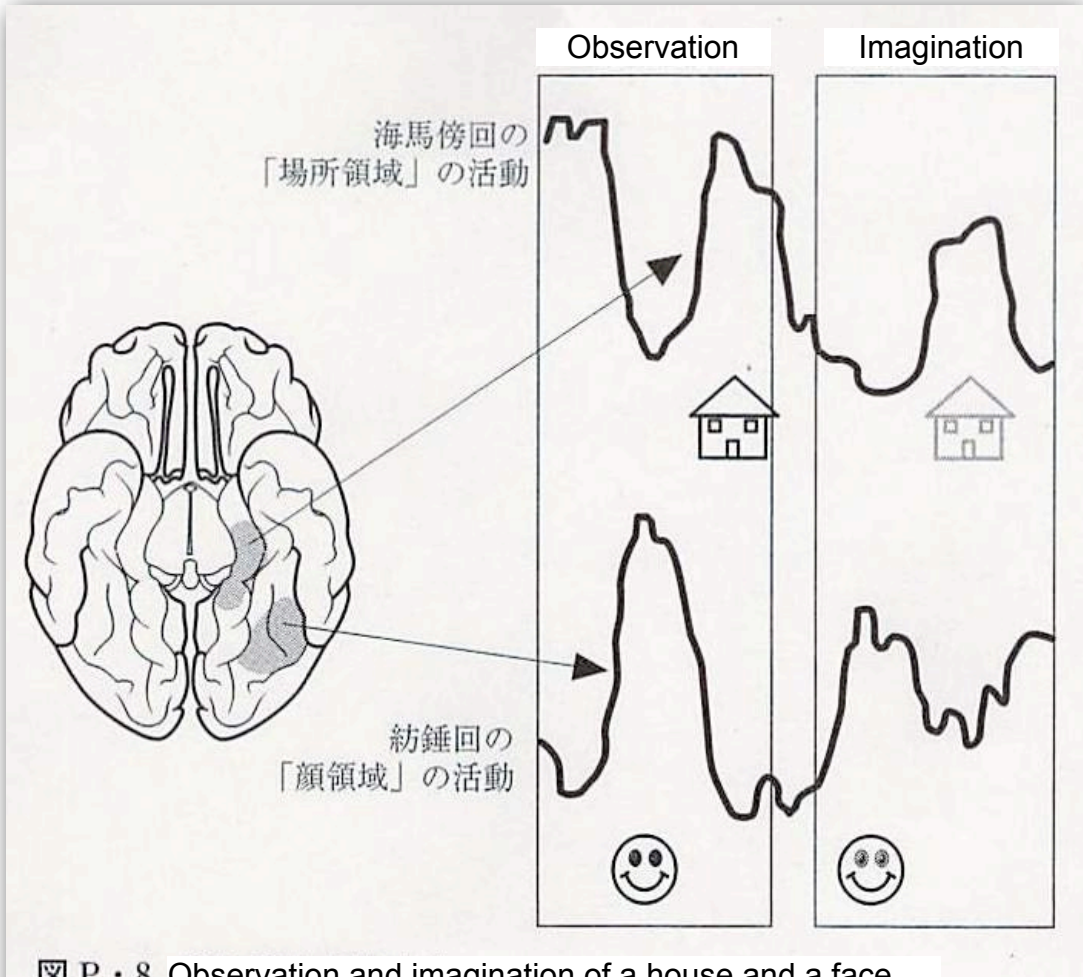


Imagination and execution of an action

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



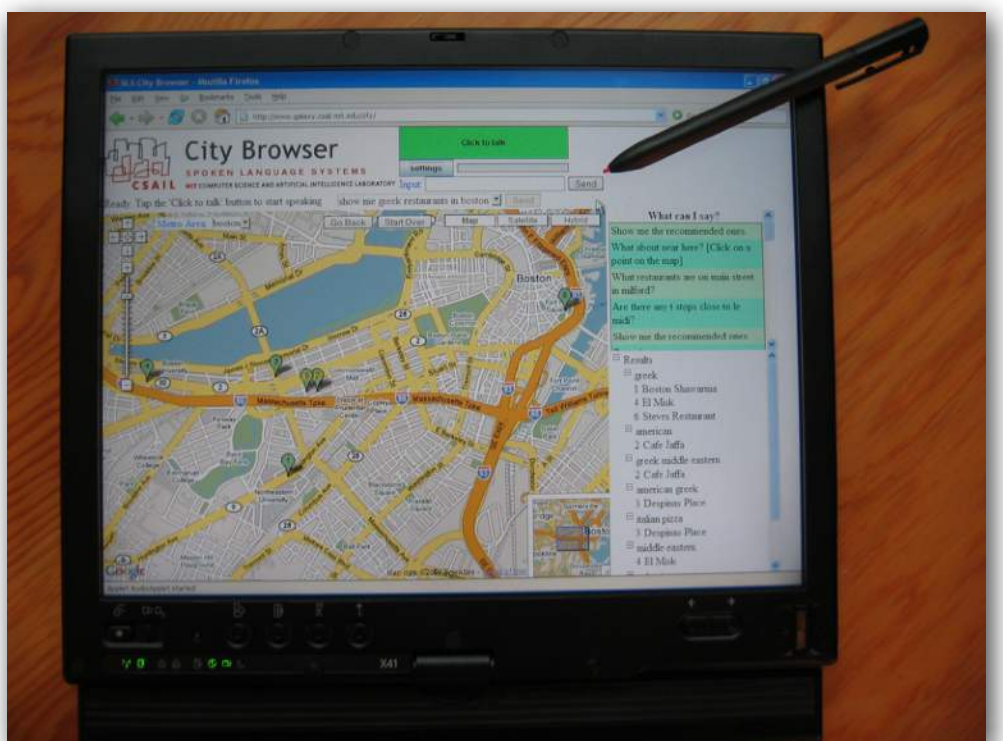
図P・7 Brain activities of real and imaginary motions
イラストは脳の活動を見るために横断面図(上部と中間部)を撮った位置を示す．上の断面図は右手を実際に動かすときの脳活動を，下の断面図は動きを想像したときの脳活動を示す．



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．

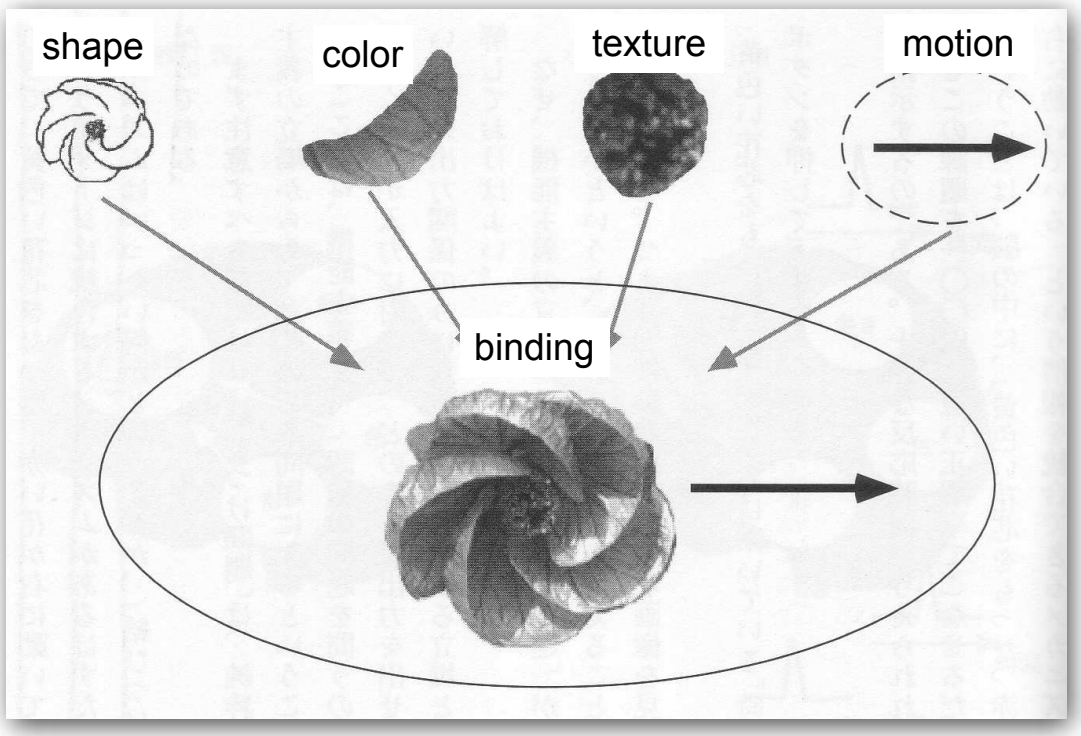
Multimodal interface

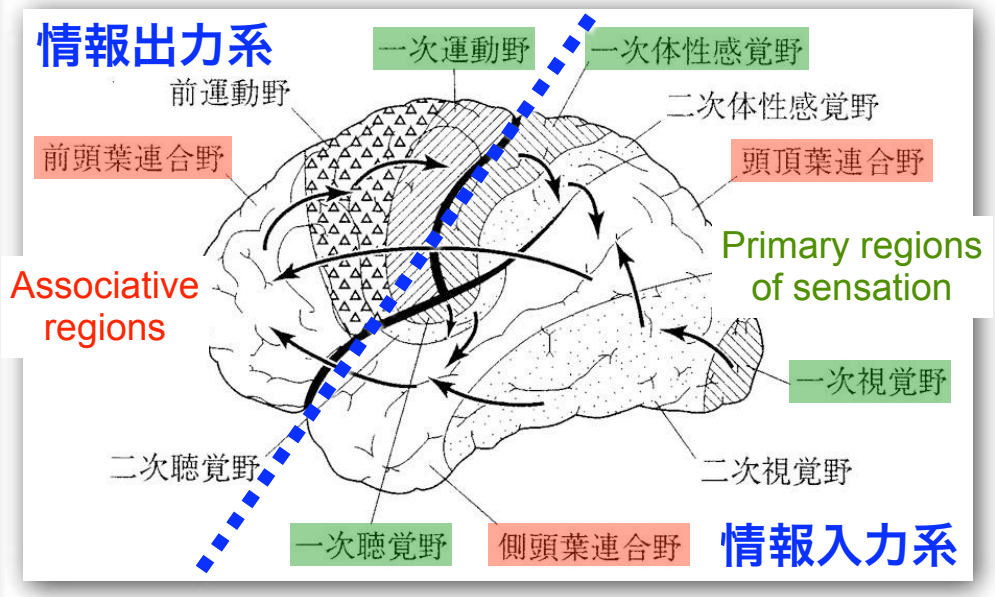
- Examples of the multimodal interface
  - Integration of various input modalities
    - keyboard (text), pointing device, speech, touch screen, still/moving images, etc.
  - How to integrate inputs of different modalities?
    - Temporal and spatial integration of inputs through different modalities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



Multimodal interface

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



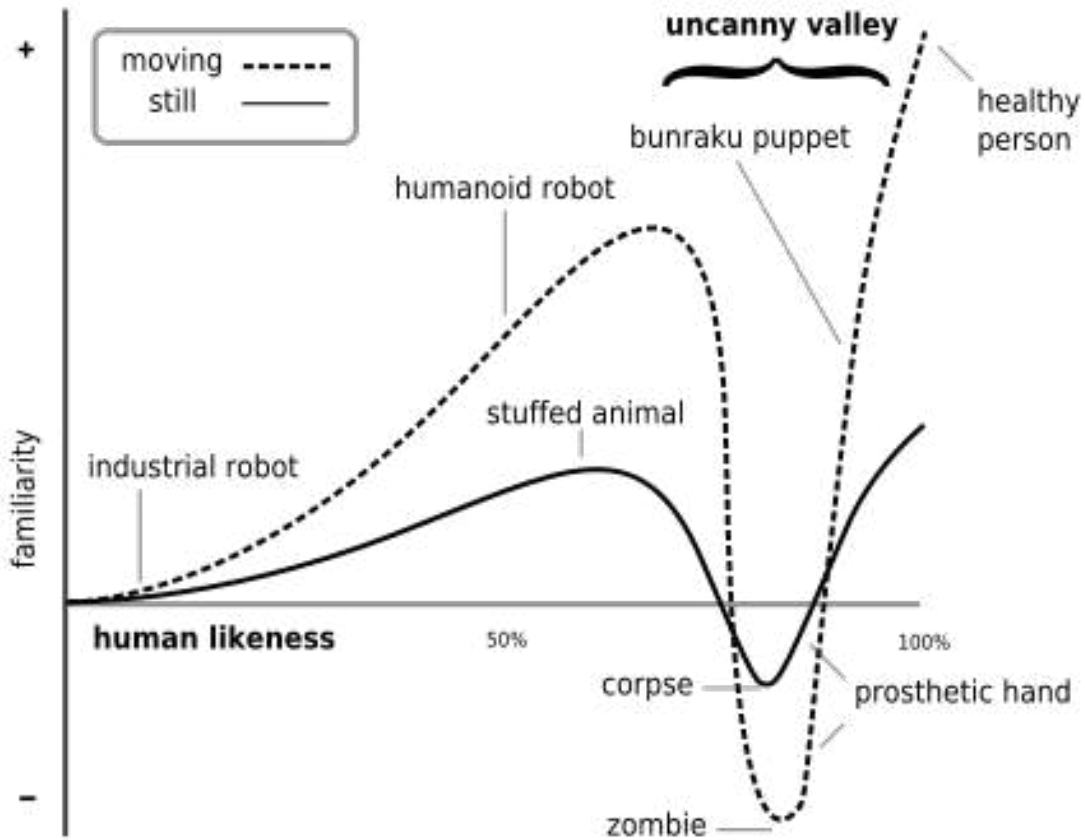


Social interaction and multimedia

- What is social interaction?
  - Interaction caused in the context of social relations
    - One individual has to play various social roles due to social environments.
      - Associate 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 multimedia

- The uncanny valley

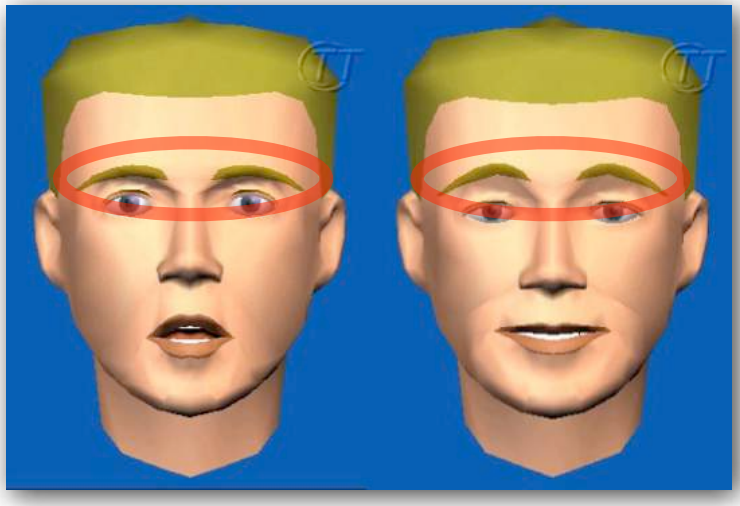


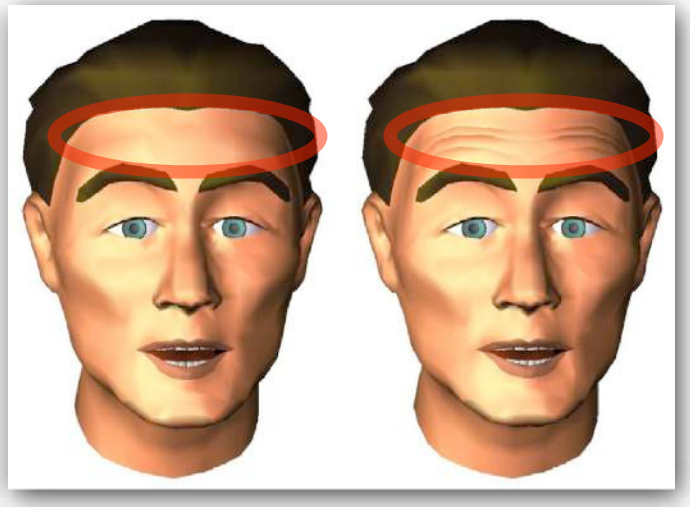
Social interaction and multimedia

- 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 noticing 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.

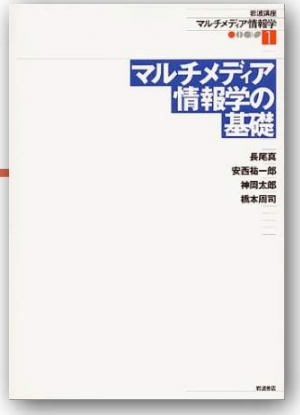
Social interaction and multimedia

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





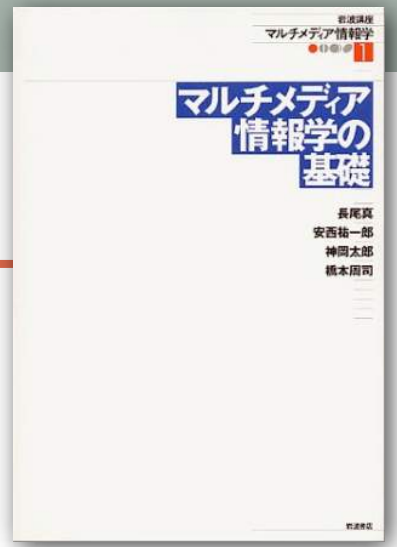
Title of each lecture



- 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 sensation - 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

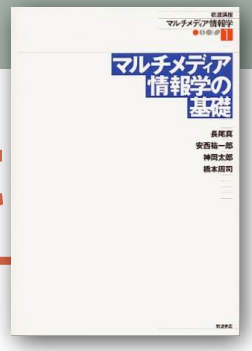
Multimedia information used in expressive and emotional processing

Nobuaki Minematsu



Today's menu

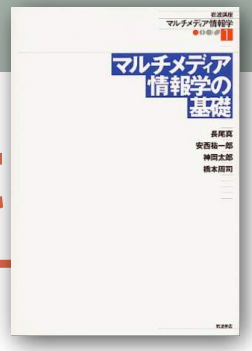
- Emotional and expressive information processing
  - Definition of emotional and expressive information
  - Quantitative measurement of the information
  - Emotion and the brain -- which region of the brain deals with the emotion?
- Emotional information and multimedia sensory information
  - The five senses and emotion
  - Integration of information transmitted using different kinds of media
- Transmission or communication of emotional information
  - Encoding, transmission, and decoding of the information
  - PUI, visualization, facial expressions, etc
  - Some examples of emotional systems
- Summary
- Announcement on the next lecture
  - The first assignment is given in the next lecture.



Emotional and expressive informat

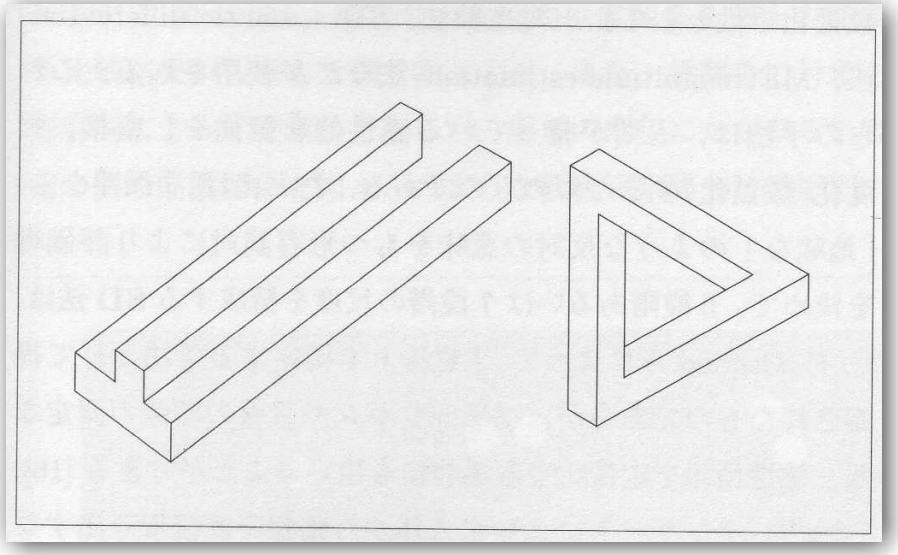
- History of information engineering
  - 1st generation: measurement and extraction of the specific aspects from temporal and/or spatial patterns of physical signals (raw materials)
    - Signal processing, quantitative, analogue, differential equations, etc
  - 2nd generation: conversion from the extracted patterns to symbolic representations for logical processing of these symbols
    - Pattern recognition, qualitative, logical programming, knowledge, inference, digital, etc
  - 3rd generation: processing of what humans perceive without explicit logical judgment or thinking. 情 of 知・情・意　　善悪と快不快
    - Emotional processing, cognitive science, psychology, subjective, etc

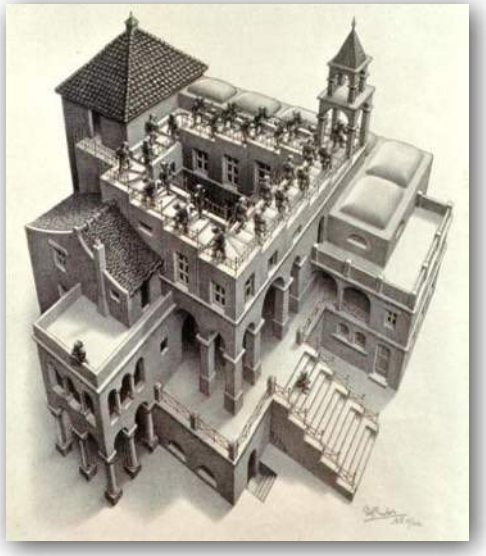
layer	channel or media	law	area	scale of reality
physical	light, sound, force, etc	law of mechanics and electricity	signal processing	explainability
			virtual reality	causality
logical	language, symbols	logic and syntax	knowledge processing	provability
	algebra		artificial intelligence	logicality
emotional & expressive	music, art	subjectivity	emotional info. processing	subjective consistency
	expression	like and dislike	artificial (virtual) emotion	



Emotional and expressive informat

- Expressiveness (感性) and information
  - Expressiveness (emotion) : states and actions of the mind affected by stimuli received through some physical channels from the outer world.
    - Very subjective and dependent on context and receivers' characteristics
    - Recent advances of media technology try to deal with expressive information.
  - Functions of the expressive (emotional) mind
    - The expressive mind helps human (logical) judgment and understanding of the outer world, which is done through physical channels of different media.
    - This help or support makes the judgment and understanding very efficient.
      - But this help or support is very implicit and unconscious (prelogical).



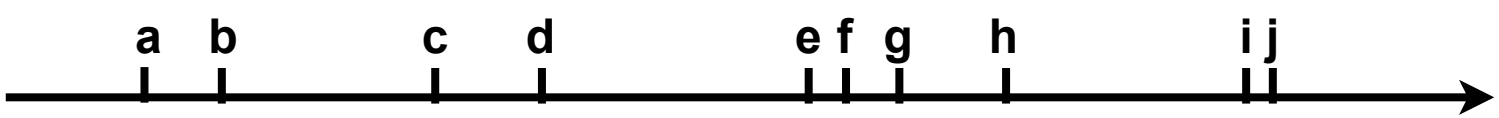


Examples of illusionism

If we are totally logical, nobody must be tricked by these pictures.

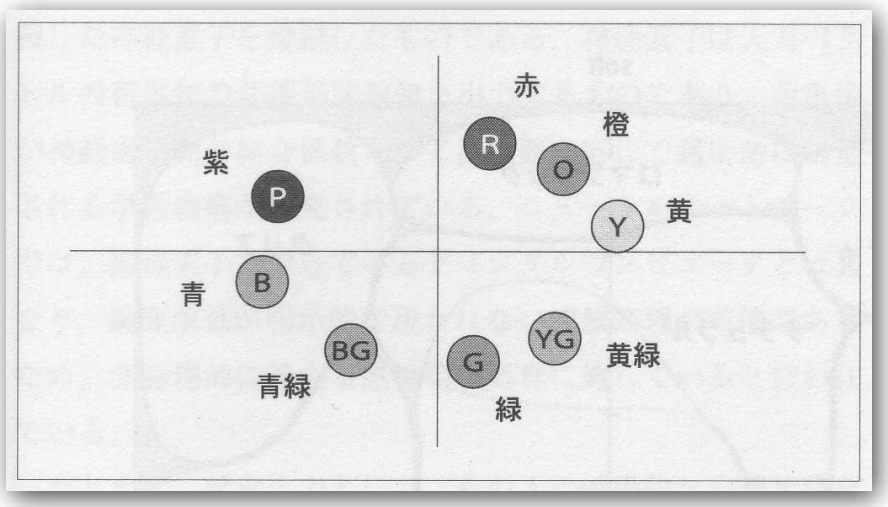
Quantification of expressive info

- Quantitative representation of expressive information
  - Mapping from a physical media space into an expressive media space.
    - Use of questionnaire (subjective judgment) to assign some quantitative scores to emotional expressions such as expressive arts and products.
      - Quantification of expressiveness
    - ME (Magnitude Estimation) method
      - Subjects give scores of magnitude to a given stimulus by comparing it to the standard stimulus.
      - Mapping between scores and a set of stimuli is done by assuming the power law (べき法則) between physical observation and our sensation.
    - Paired comparison (一対比較法)
      - Out of N stimuli, subjects have to compare all the possible pairs (${}_N C_2$) and select one which has a larger magnitude.
        - The number of comparisons is proportional to N^2.
      - All the stimuli can be aligned (plotted) on a single scale of the property of interest.
    - Multi-dimensional scaling (MDS)
      - Subjective difference is quantitatively measured between any pair of the N stimuli of interest.
      - An N x N distance matrix can define a geometrical shape in a multi-dimensional space.
      - This shape is projected nonlinearly onto a two-dimensional plane.

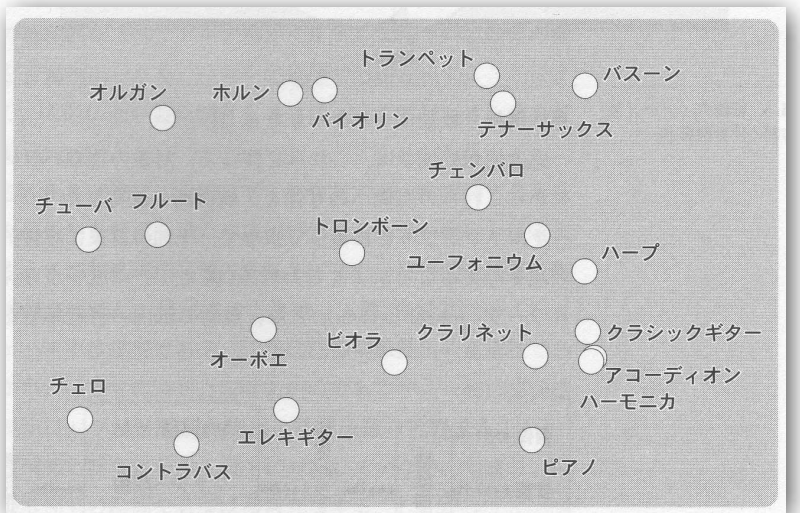


Quantification of expressive info

- Quantitative representation of expressive information
  - Two examples of MDS

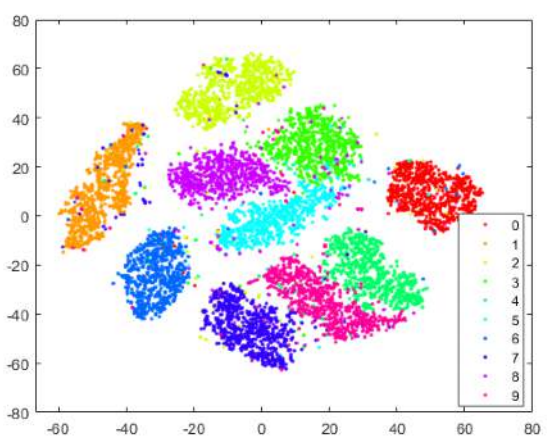


MDS of colors



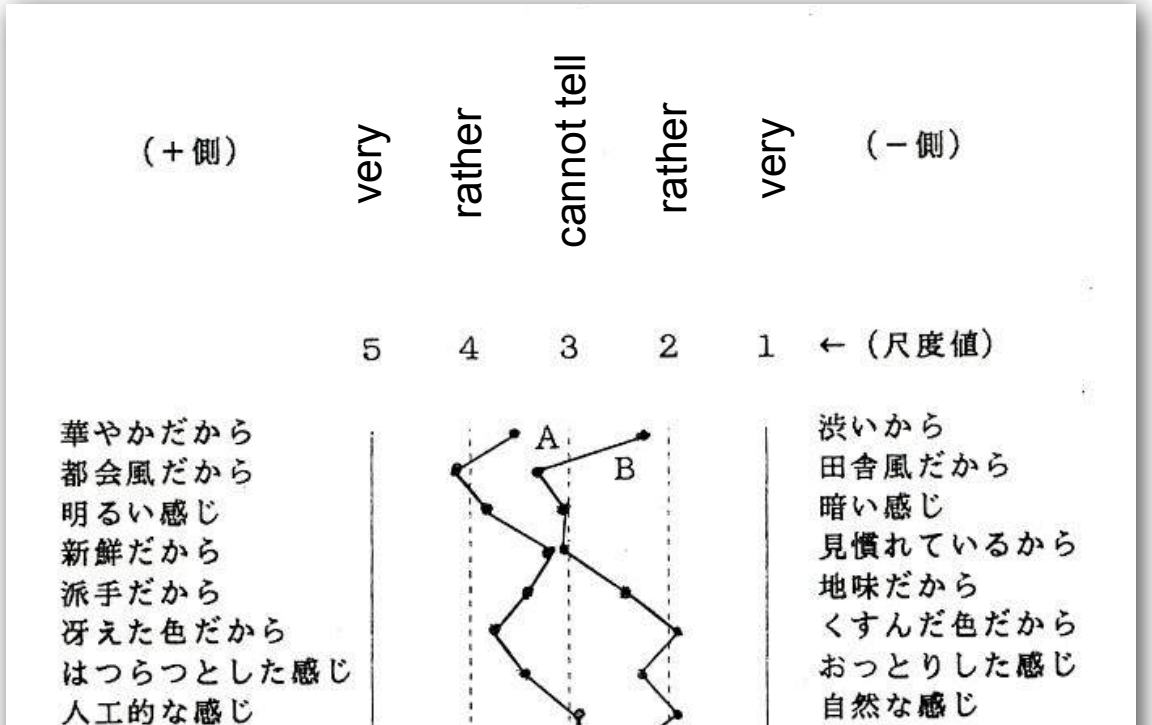
MDS of timbre of musical instruments

  - t-SNE (t-distributed stochastic neighbor embedding)
    - https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding



Quantification of expressive info

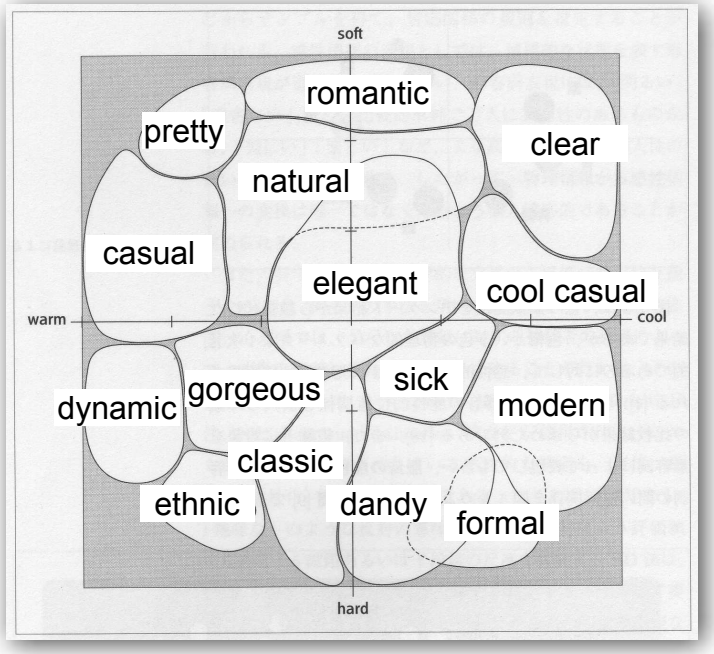
- Quantitative representation of expressive information
  - Mapping from a physical space into an expressive space
  - SD (Semantic Differential) method
    - Preparation of many pairs of adjectives representing totally opposite meanings
      - light -- dark, cold -- hot, beautiful -- ugly, small -- large, high -- low, etc.
    - N pairs of adjectives = an N-dimensional space
      - Can build an N-dimensional space for expressive representation
      - Any stimulus is represented as a point in the N-dimensional space.

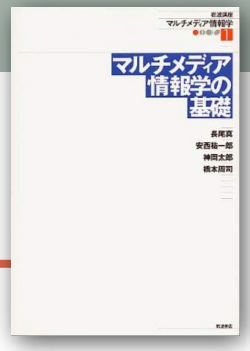


Examples of adjectives used to represent subjective impression of a given piece of clothing.

Quantification of expressive info

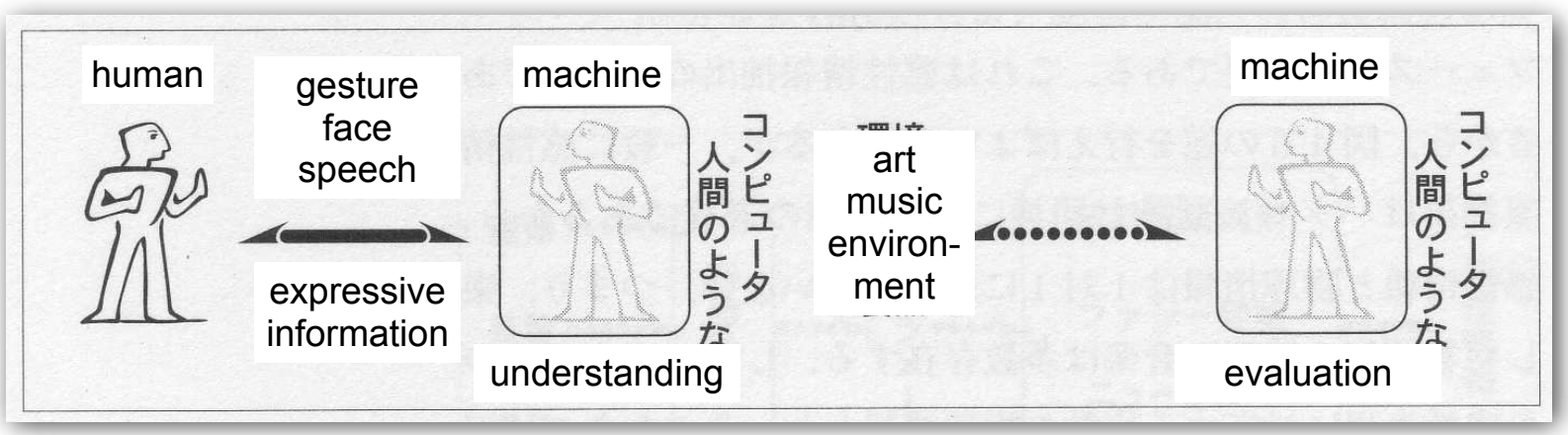
- Quantitative representation of expressive information
  - Mapping from a physical space into an expressive space
  - SD (Semantic Differential) method
    - Preparation of many pairs of adjectives representing totally opposite meanings
      - light -- dark, cold -- hot, beautiful -- ugly, small -- large, high -- low, etc.
    - N pairs of adjectives = an N-dimensional space
      - Can build an N-dimensional space for expressive representation
      - Any stimulus is represented as a point in the N-dimensional space.
  - Factor analysis
    - Dimension reduction to determine a fewer number of very essential dimensions
      - The meaning of the dimensions is considered by researchers.
      - 1st dimension: soft -- hard
      - 2nd dimension: warm -- cold
      - 3rd dimension: .....

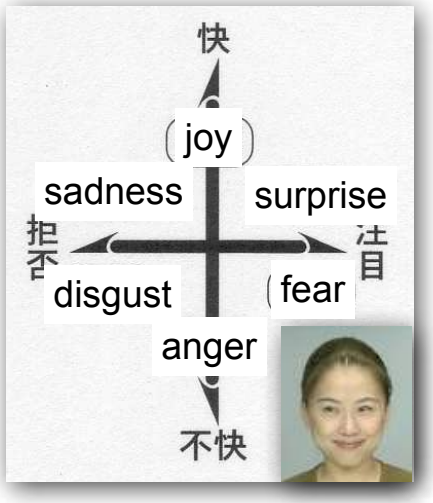




Expressive info. processing

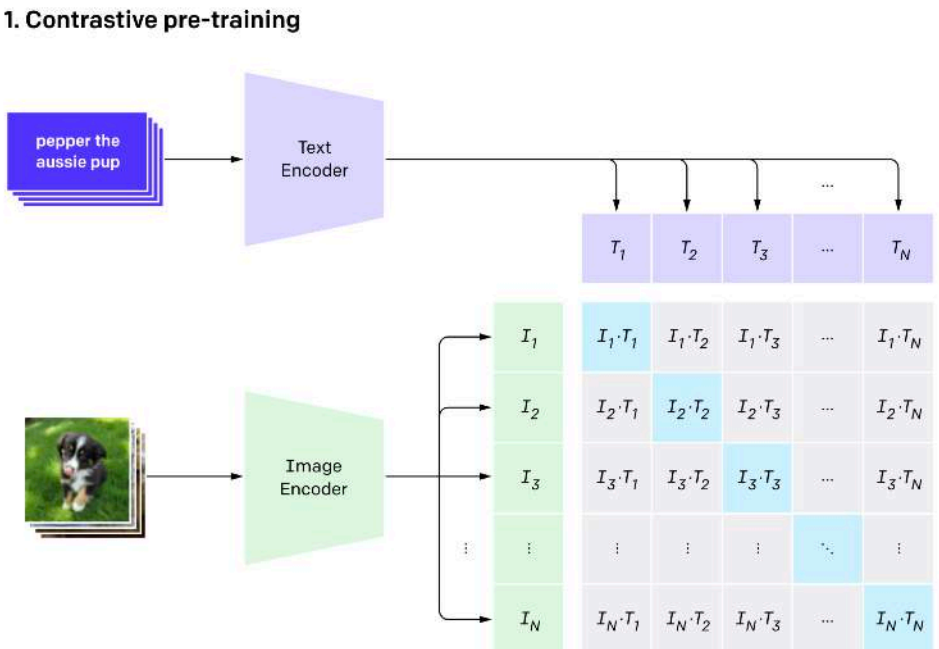
- Development of expressive info. processing systems
  - Systems that can detect expressive information given from human users
  - Systems that can detect expressive information given from the environment and show the information to human users.
    - Output methods : facial expressions, music, expressive utterances, etc
- A typical example of emotional machines
  - Input = facial expressions, output = adjectives to represent the faces
  - Input = expressive utterances, output = adjectives to represent the utterances
    - Six fundamental emotions of anger, fear, disgust, joy, sadness, and surprise
    - Visual or acoustic features associated with each adjective are extracted and modeled.

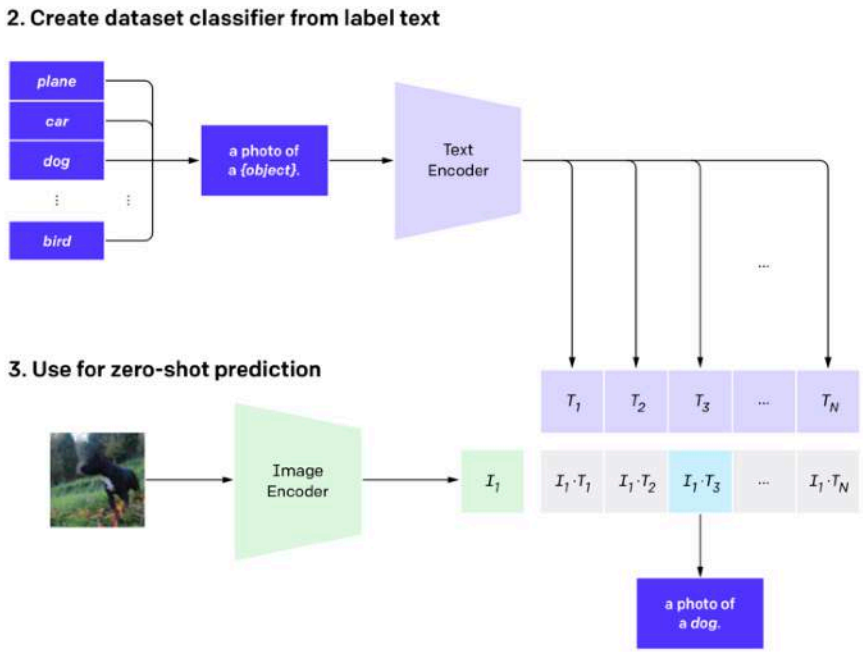


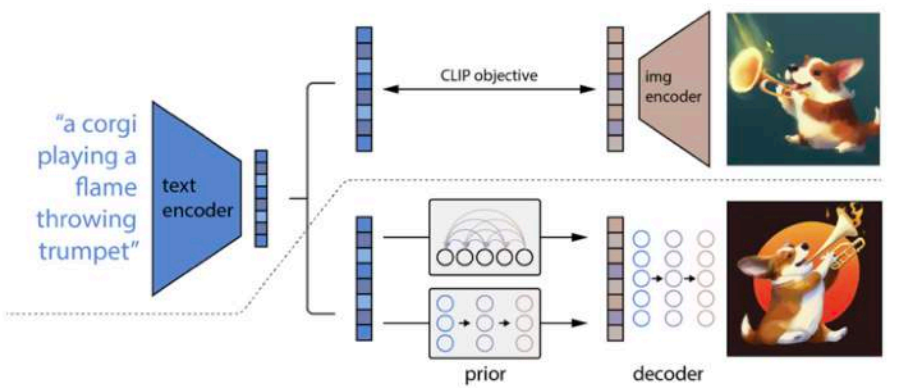


Images generated from language input

- Deep Learning for mapping between images and their descriptions
  - CLIP : Connecting Text and Images



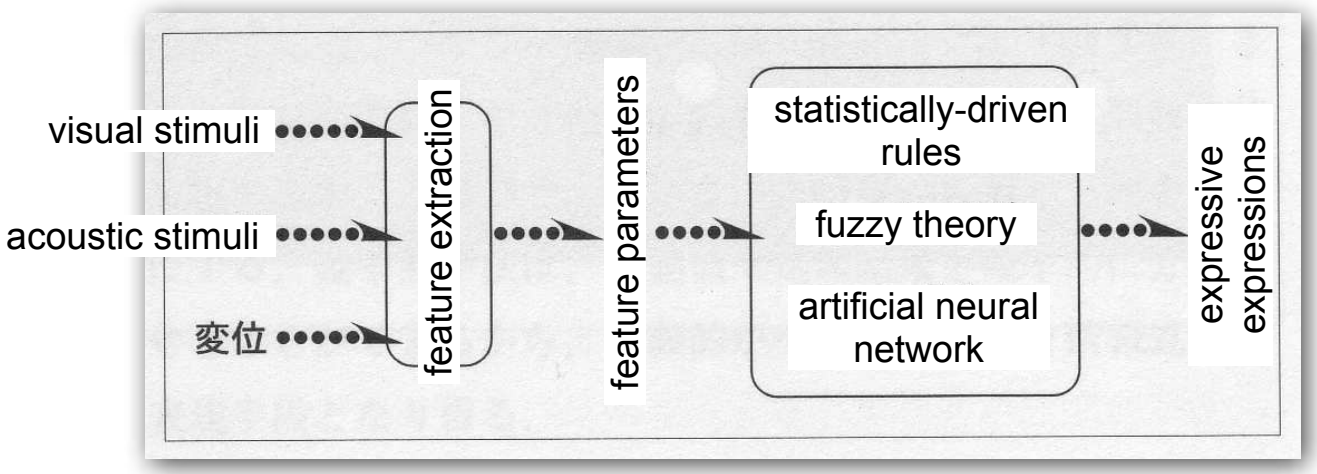


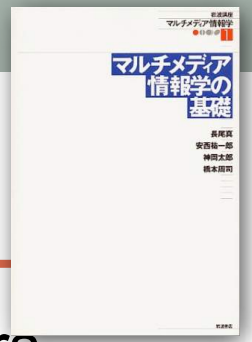


https://openai.com/research/clip

Expressive info. processing

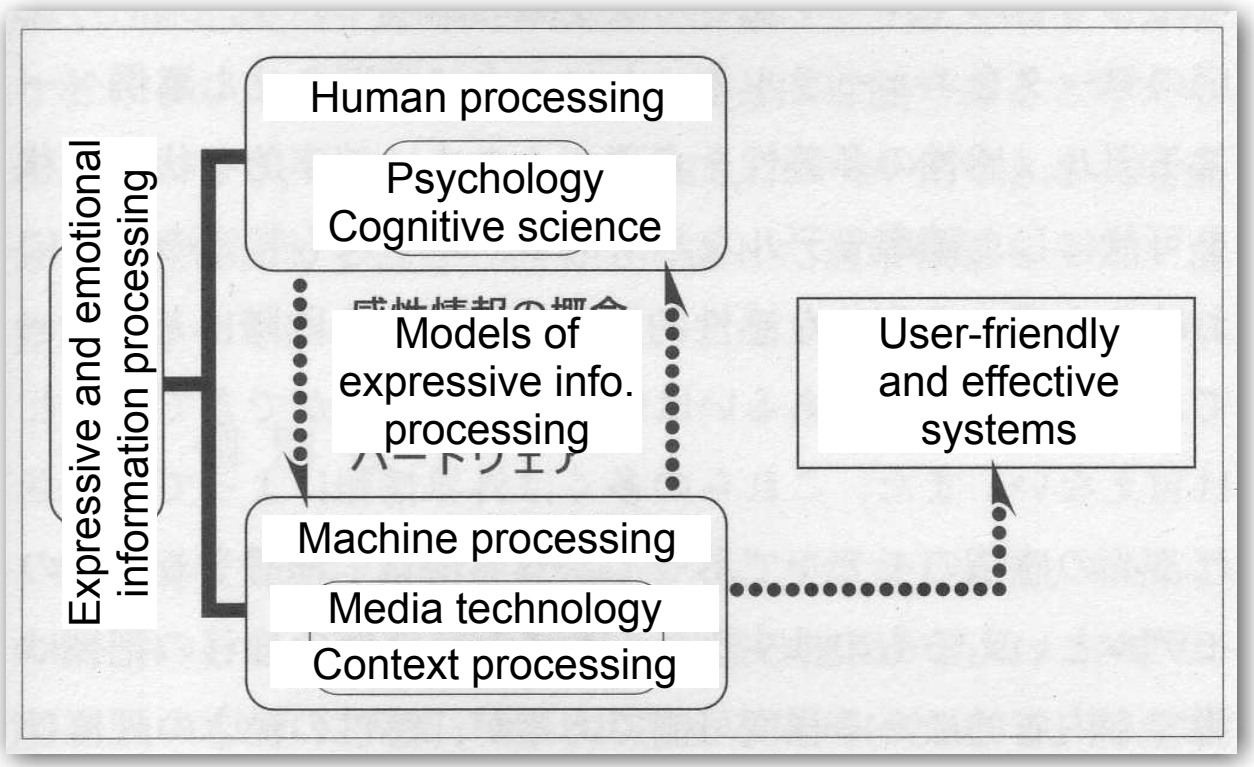
- A typical example of emotional machines
  - Emotional stimuli input --> expressive adjectives output
    - Does this mean that machines can “understand” facial expressions?
    - Or are these machines just pretending to understand facial expressions?
  - Symbolic representation of expressive information
    - Is it really symbolic and discrete? It should be continuous and analogue?
- Conversion from expressive words (symbols) to some physical features
  - “I want to listen to a warm, delicious, and tall piece of musical performance.”
  - A set of words can generate multiple candidates.
    - Dependency on context and user characteristics.

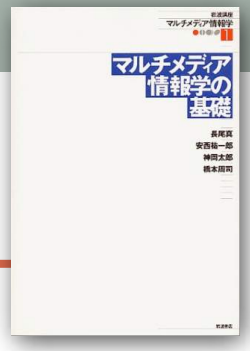




Expressive info. processing

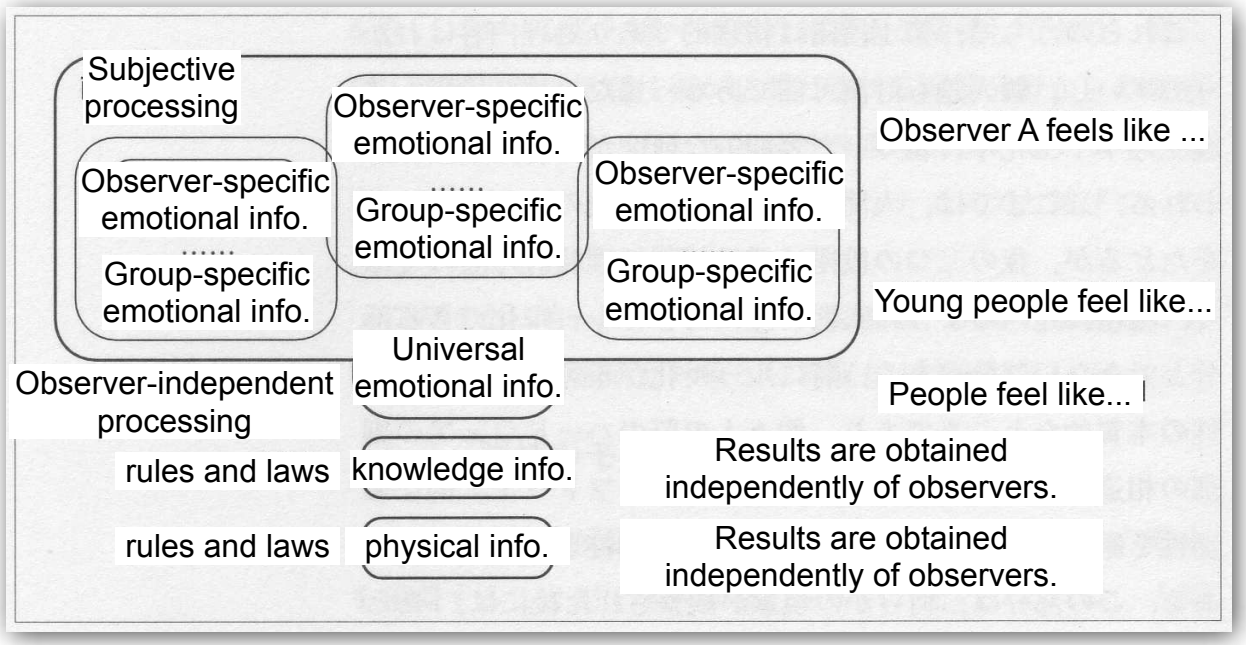
- Toward developing more human-like expressive info. processors
  - Integration of two kinds of knowledge is required.
    - Knowledge on media technology and that on human (cognitive) science
  - How does the brain process expressive and emotional information?
    - It can be a good model for machine processing.





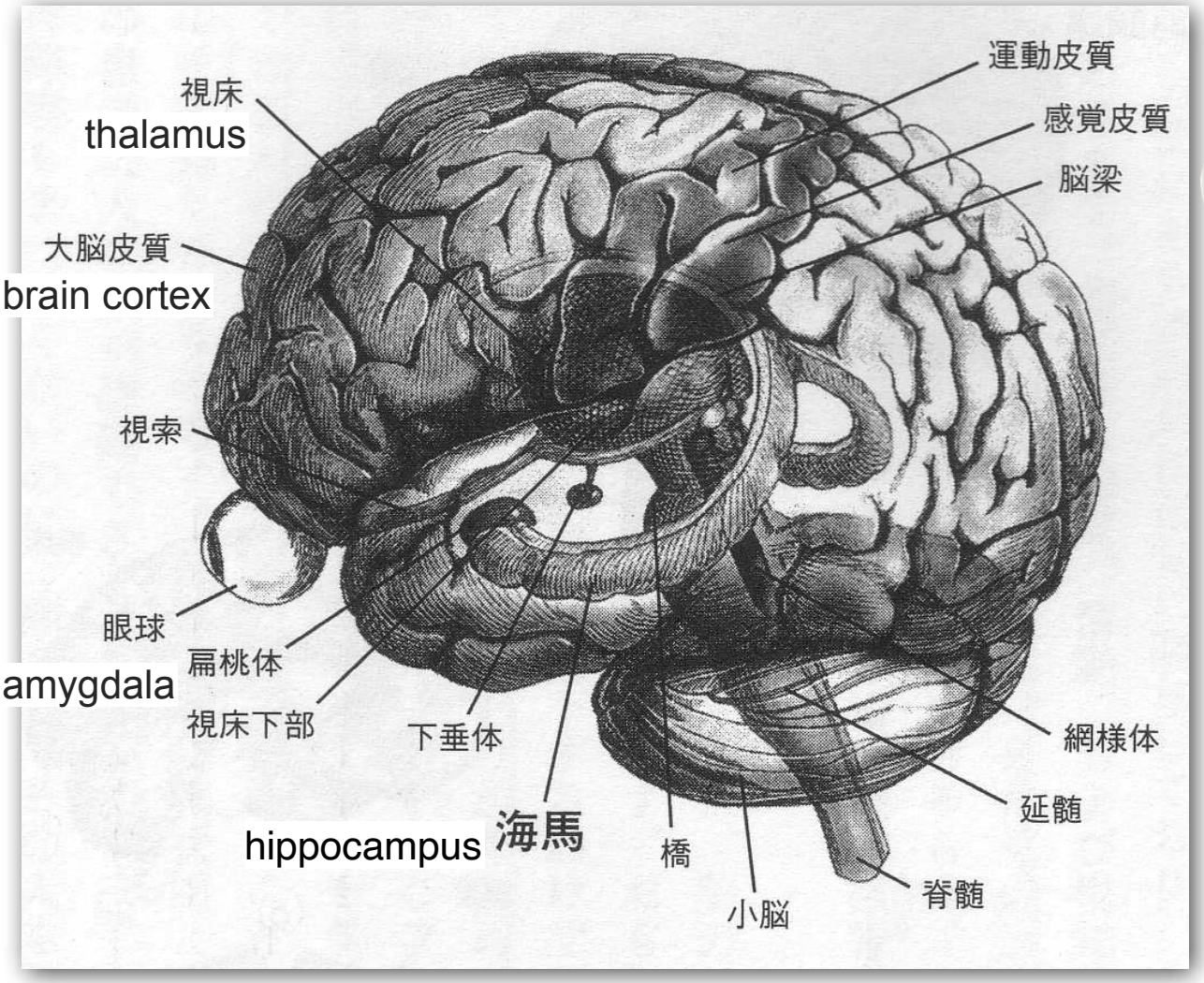
Expressive info. processing

- Model of emotional actions by humans
  - Prediction of human emotional actions requires a good model in machines
    - Rule-based models: input-output relations are described by rules.
    - Example-based models: input-output examples (correlations) are modeled by NN.
    - Statistical models: emotional ambiguity is modeled by probabilistic state transition.
      - These models are driven by inputs and generate (predict) output actions.
  - Individual differences of human emotional actions
    - The prediction models have to deal with individual differences.
      - Hierarchical structure of variability.



Emotions and the brain

- Which region of the brain processes the emotion?
  - Deep structure of the brain



Location of the two amygdalae

From the old brain to the new brai

- Reptile brain --> old mammal brain --> new mammal brain
  - Emotional processing: old brain
  - Logical processing: new brain
    - Language areas: on the cortex on the left hemisphere.
      - Broca's area: the region linked to speech production.
      - Wernicke's are: the region linked to speech understanding.
      - If these areas are damaged, the patient may lose the capability of using spoken language.

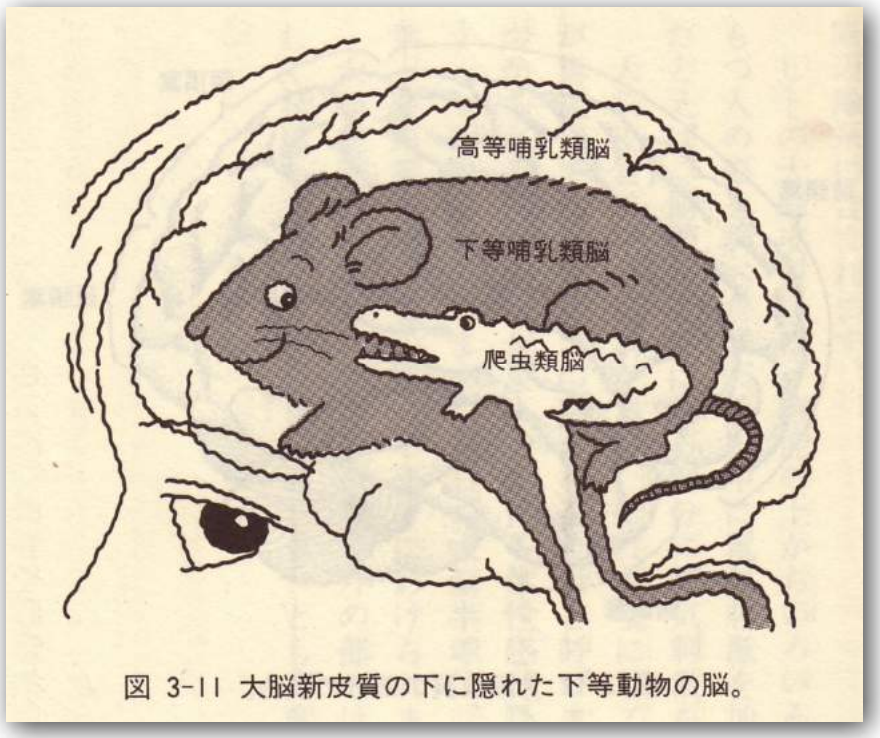
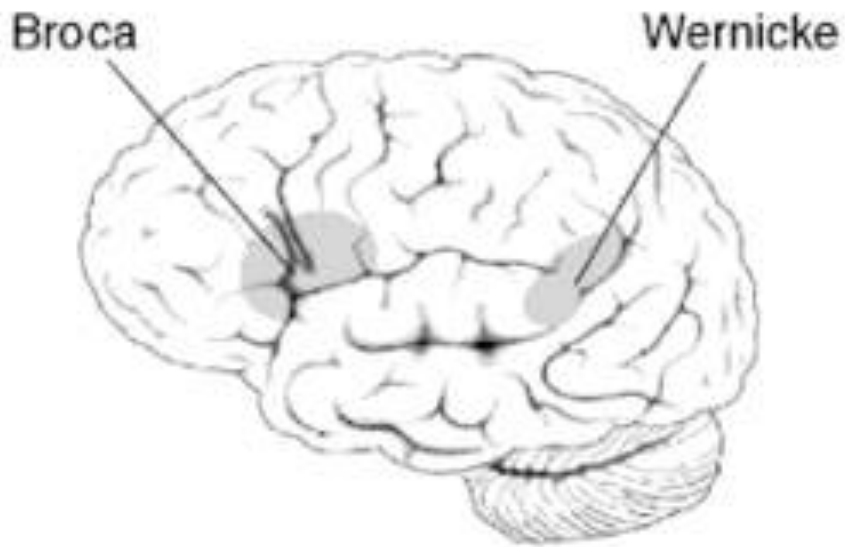
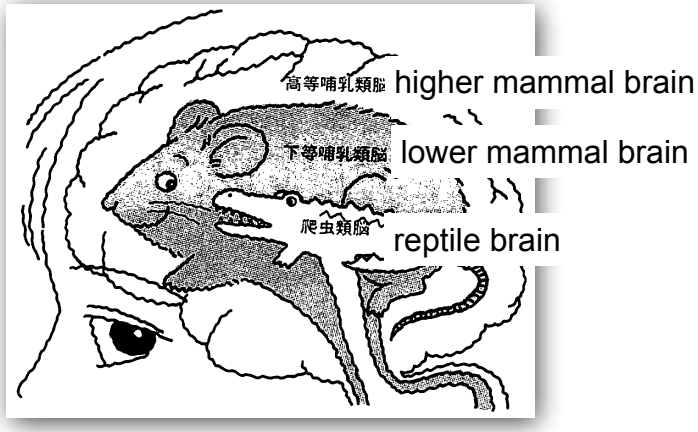


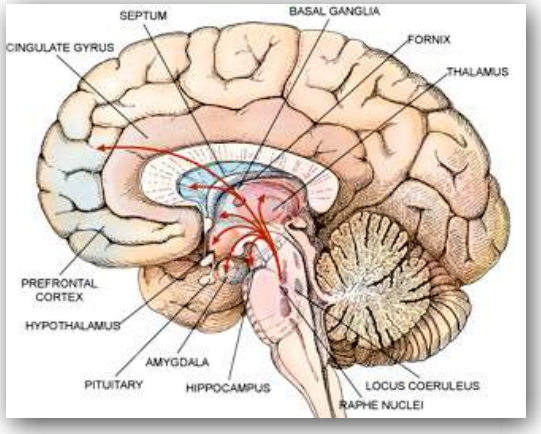
図 3-11 大脳新皮質の下に隠れた下等動物の脳。



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

From the old brain to the new brai

- Low road and high road
  - Sensory inputs are collected to reach Thalamus (視床).
  - Neural activities are transmitted to the cortex (new) and to amygdala (old).
  - Processing in the old brain is faster than that in the new brain.

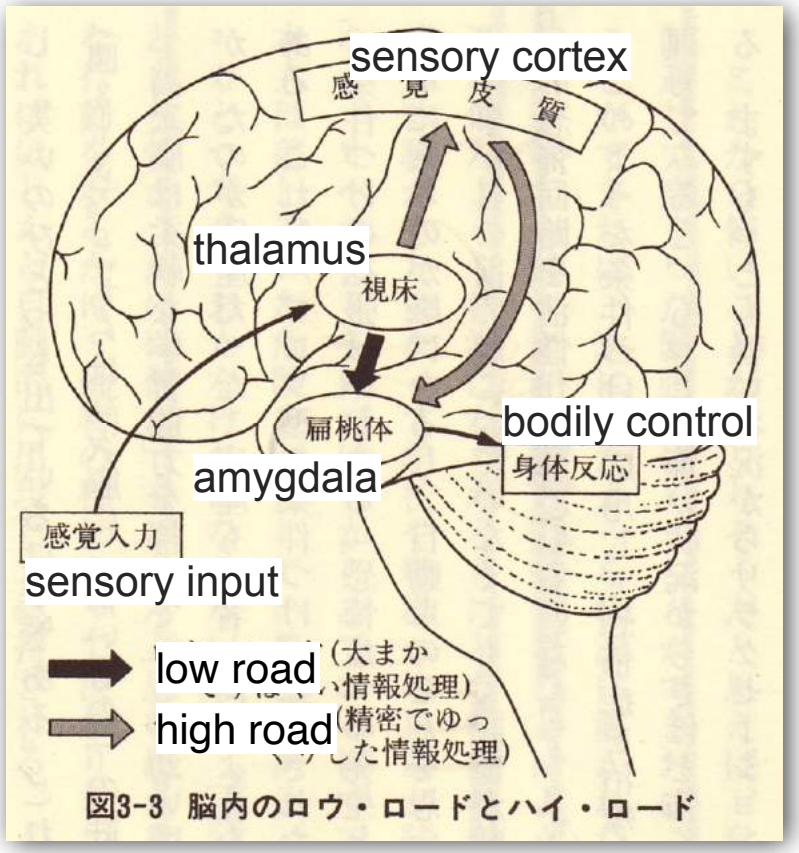
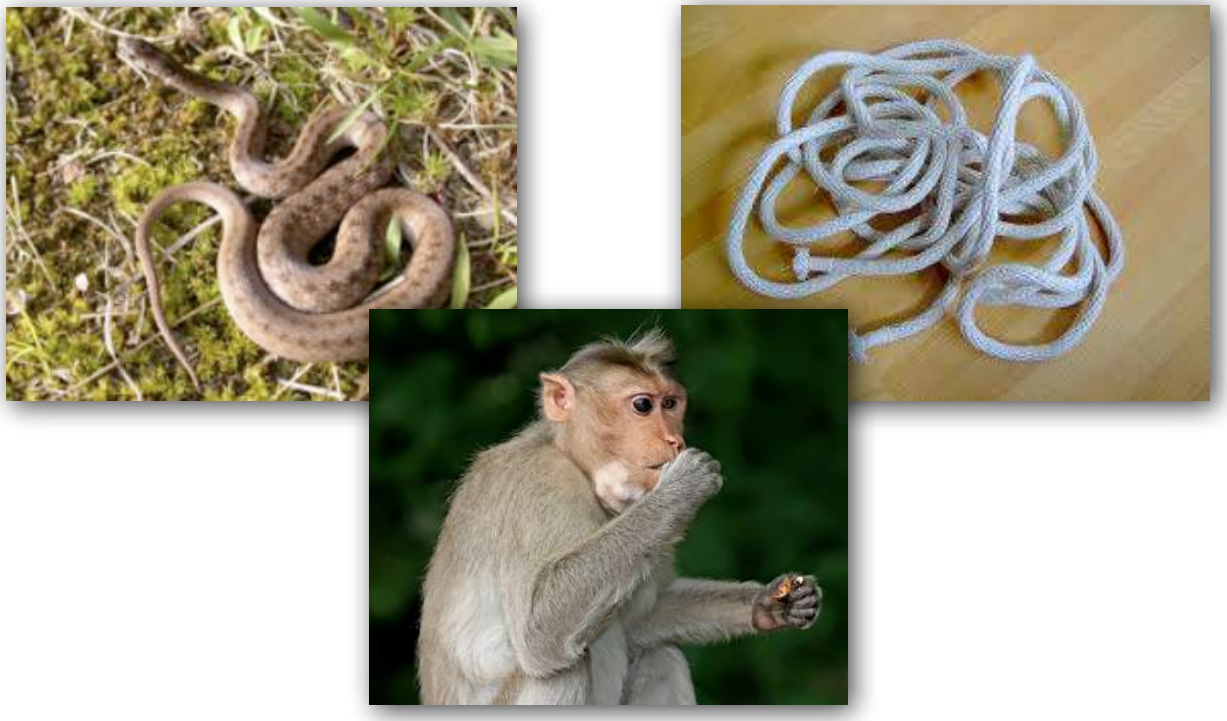


図3-3 脳内のロウ・ロードとハイ・ロード



Monkeys are frightened at a swinging/moving rope because the rope reminds the monkeys of snakes. This is the case even when the monkeys have never seen snakes in their lives so far.

Why can Ichiro make so many hits

- Only with slow processing on the cortex, he can make no hit!!
  - Time required for processing on the cortex is longer than time required for the ball to reach the home base.

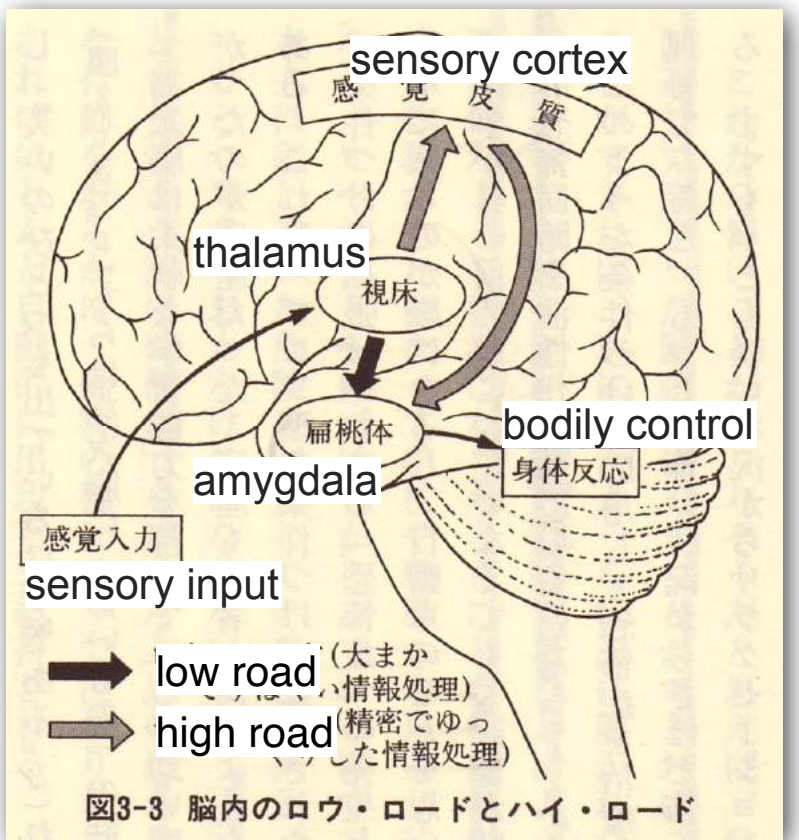
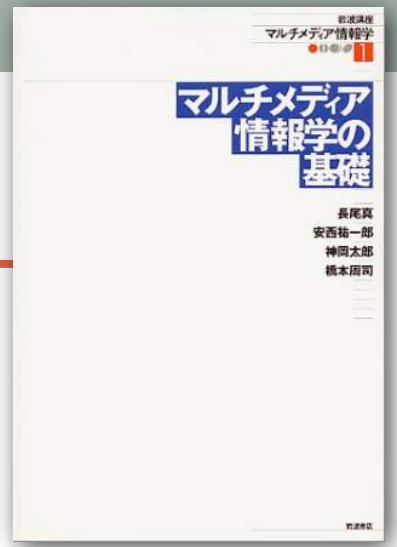


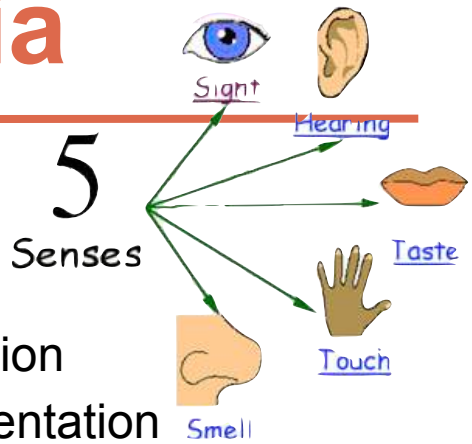
図3-3 脳内のロウ・ロードとハイ・ロード



Today's menu

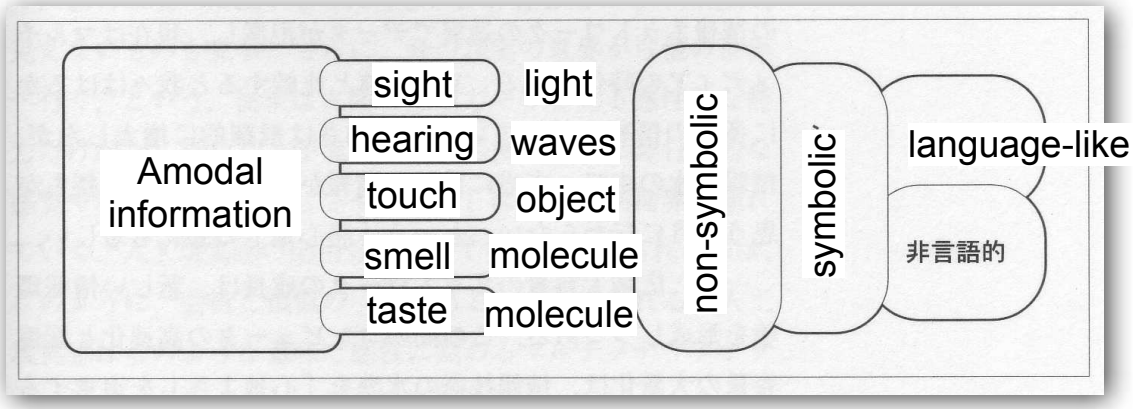
- Emotional and expressive information processing
  - Definition of emotional and expressive information
  - Quantitative measurement of the information
  - Emotion and the brain -- which region of the brain deals with the emotion?
- Emotional information and multimedia sensory information
  - The five senses and emotion
  - Integration of information transmitted using different kinds of media
- Transmission or communication of emotional information
  - Encoding, transmission, and decoding of the information
  - PUI, visualization, facial expressions, etc
  - Some examples of emotional systems
- Summary
- Announcement on the next lecture
  - The first assignment is given at the end of the next lecture.

Emotions and multimedia



- Expressive information and the five senses
  - Sight, hearing, touch, (taste, and smell)
    - Symbolic representation and non-symbolic representation of information
    - Symbols can be organized by some syntax into a higher-level representation
    - Modality-dependent and modality-independent information
  - Information exchange via. a combination of different media (forms)

sense	media	method of expression	interface
sight	light	painting, gesture	physical
hearing	sound waves	speech, music	
touch	force, mechanics	body motions	
taste	molecule, temperature	dishes, drinks	chemical
smell	molecule	dishes, cosmetic	

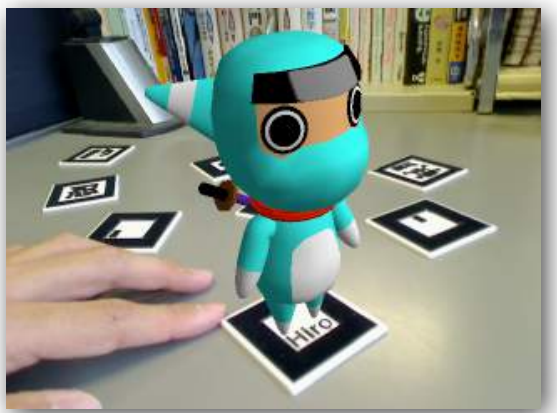


Emotions and multimedia

- Examples of virtual sensory input from / output to the five senses

Grasping a thing virtually

Dancing in a virtual world (Wii)

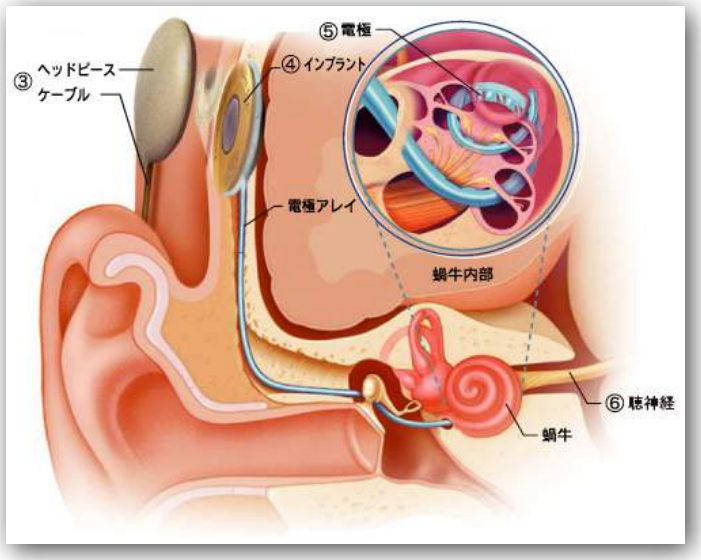


Emotions and multimedia

- Practical applications to support handicapped individuals
  - Machines (mechanical sensors) and the brain (the nerves) are connected.

This artificial arm/hand can sense the heat!!

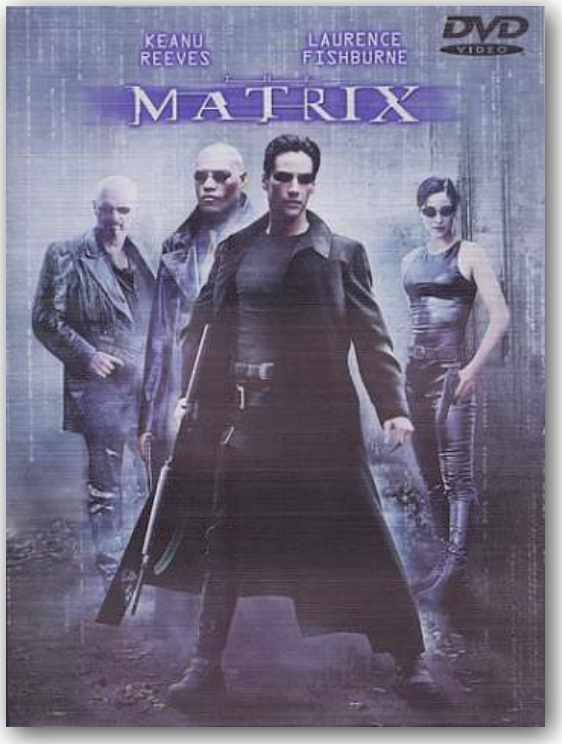
This wheelchair can be controlled by “wishing.”



Cochlear implant

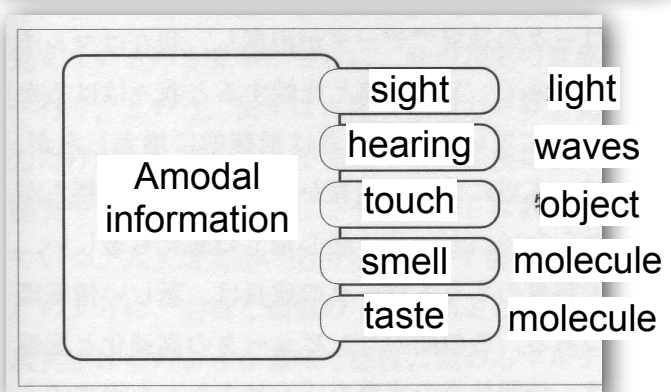
Emotions and multimedia

- Can this turn out to be the real world?



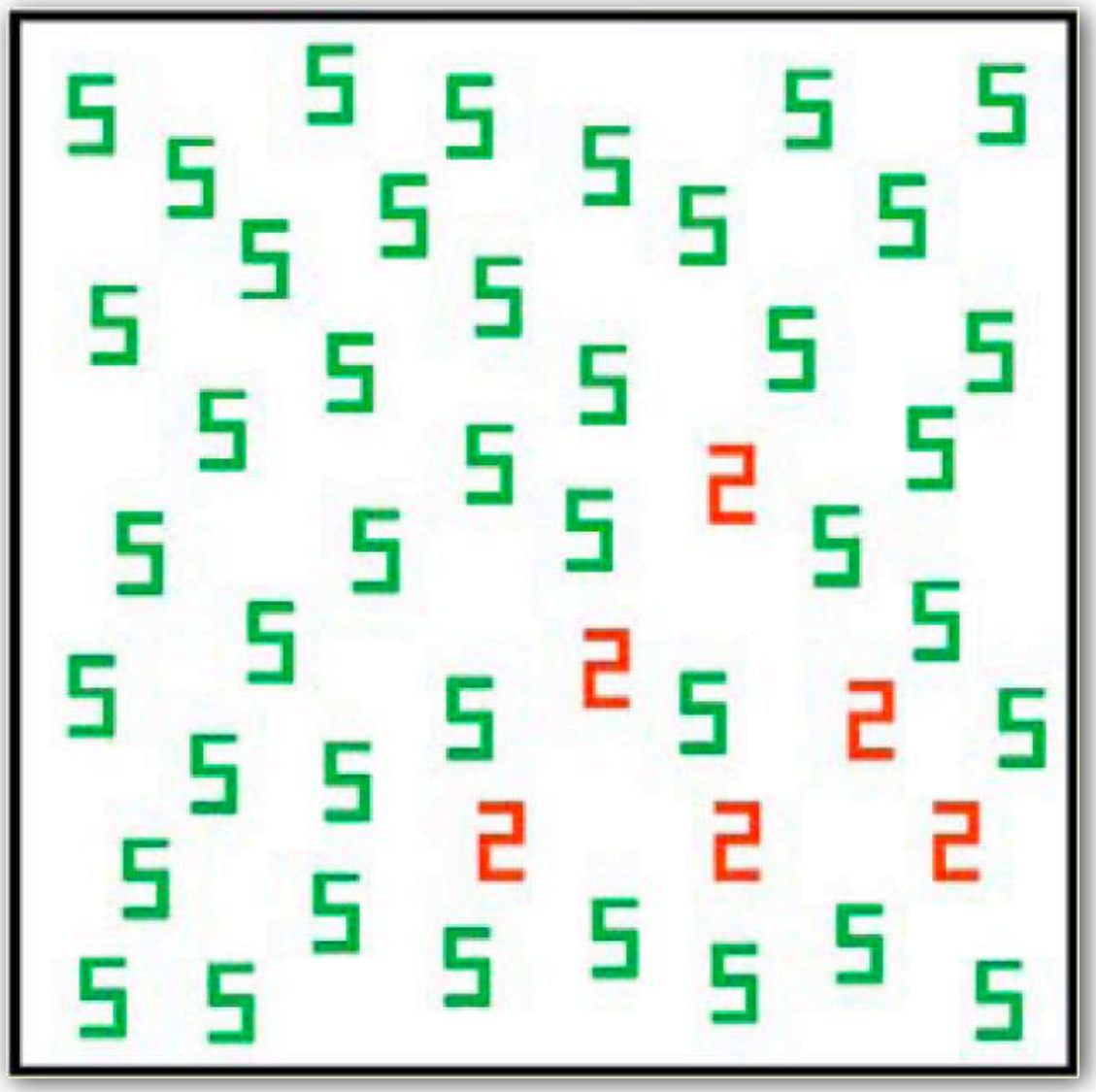
Emotions and multimedia

- Interaction between information of different physical media
  - Multiplier effect (synergism) by different media expressions
    - film and BGM, letters and pictures, etc
  - An interesting example of integrating information transmitted via. different channels
    - McGurk effect (McGurk and MacDonald'76)
      - [ba] (audio) + [ga](visual) = [da]
    - Wall painted in aqua blue (水色)
      - Color stimuli cause a sensation of "cool".
      - The skin's sensation is changed by colors.
  - Synesthesia (共感覚)
    - Looking at numbers causes a sensation of colors.
    - Looking at visual objects causes a sensation of tastes.
    - Babies' sensations are like synesthesia.
      - Sensations should not be divided into a specific number of categories??



Emotions and multimedia

- What is “synesthesia”?

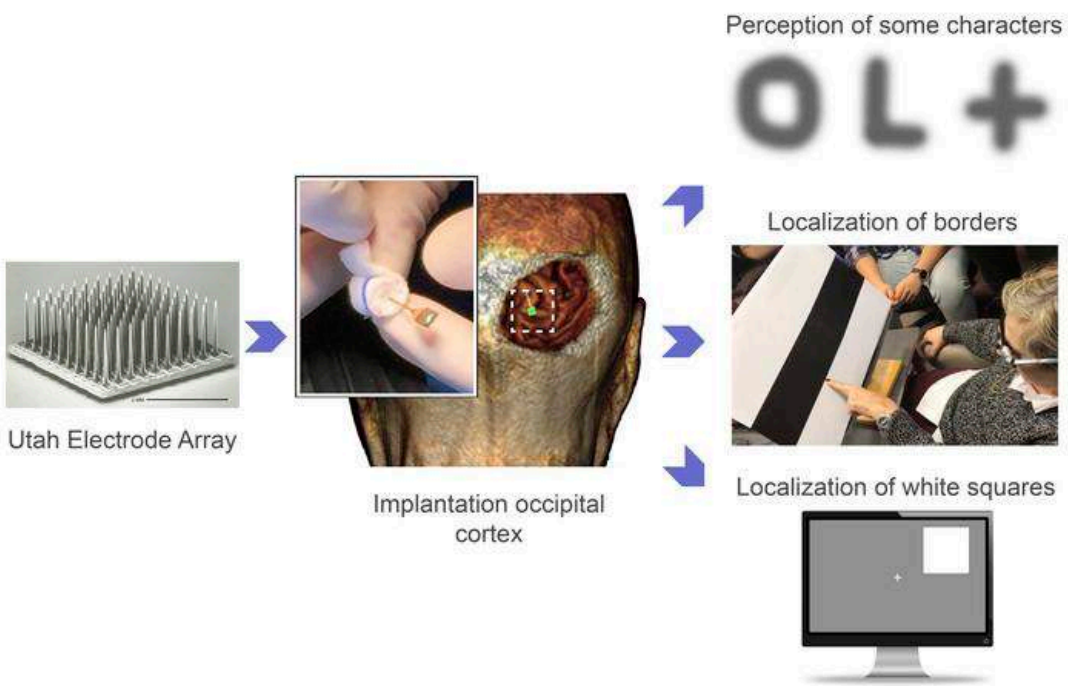


Integration or replacement of sensation

- “I can see through my tongue.”
  - CCD camera’s output (images) are transmitted to the tongue as electric signals.
    - Plasticity of the brain (脳の可塑性)
  - Blind individuals can see objects.
    - Individuals who were born as blind can see for the first time in their lives!!
    - Our sensation of “seeing” is the same as their sensation of “seeing”?

Integration or replacement of sensation

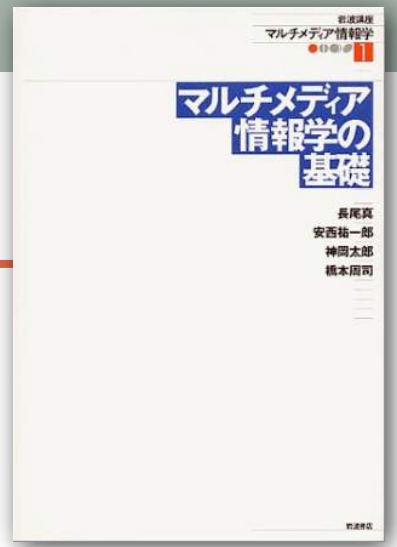
- A digital camera is connected directly to one's visual cortex.
  - "Visual percepts evoked with an Intracortical 96-channel microelectrode array inserted in human occipital cortex," *J. Clinical Investigation*, 2021
    - https://doi.org/10.1172/JCI151331



Wishing can do what?

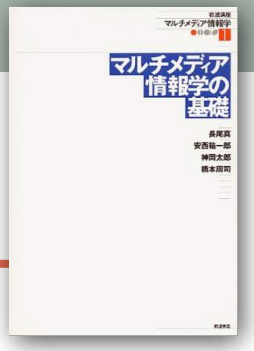
- Practical applications to support handicapped individuals
  - Wishing can control the wheelchair.
  - Wishing can generate spoken language.
    - Applied to those who lost the ability of speaking
    - (articulatory control) in a car accident.

https://wired.jp/article/brain-implants-that-help-paralyzed-people-speak-just-broke-new-records/



Today's menu

- Emotional and expressive information processing
  - Definition of emotional and expressive information
  - Quantitative measurement of the information
  - Emotion and the brain -- which region of the brain deals with the emotion?
- Emotional information and multimedia sensory information
  - The five senses and emotion
  - Integration of information transmitted using different kinds of media
- Transmission or communication of emotional information
  - Encoding, transmission, and decoding of the information
  - PUI, visualization, facial expressions, etc
  - Some examples of emotional systems
- Summary
- Announcement on the next lecture
  - The first assignment is given at the end of the next lecture.

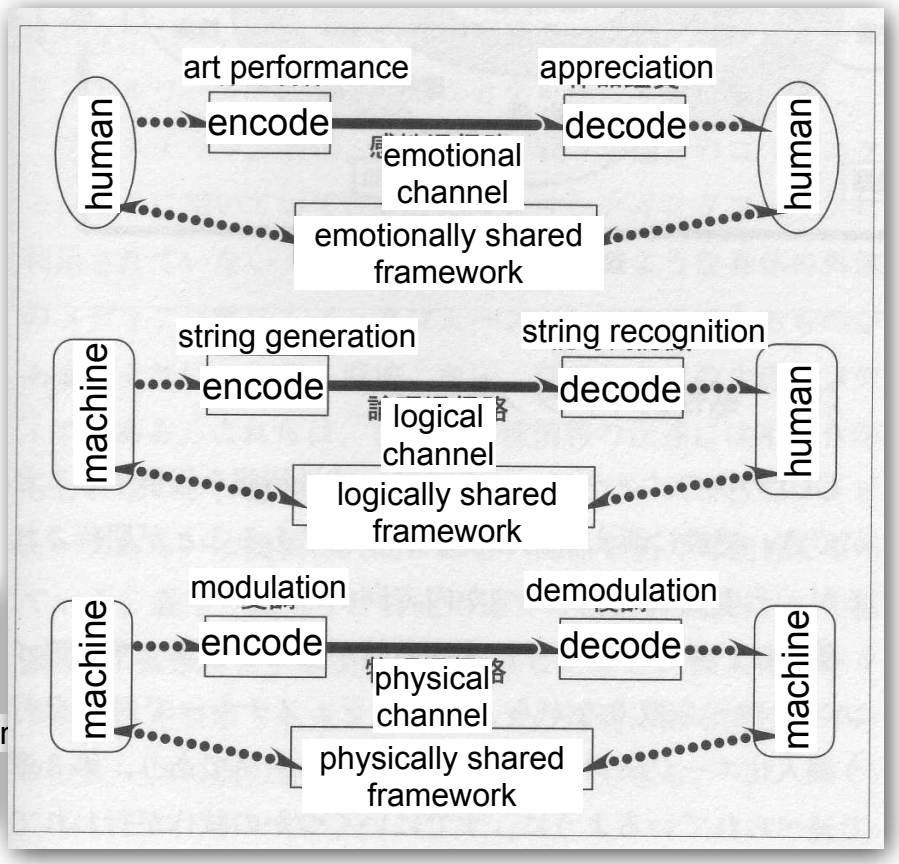


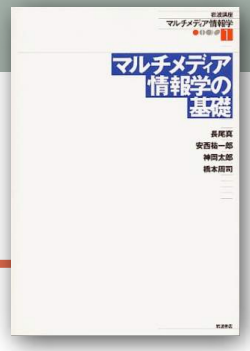
Encoding, transmission and decod

- Layered structure of communication of expressive information
  - The three layers of information and those of communication
    - Physical layer, logical layer and expressive (emotional) layer
    - A shared processing framework is needed between sender and receiver.
      - Physical layer: impedance matching
      - Logical layer: protocol (symbol operation) matching
      - Expressive layer: cultural matching???

layer	channel	law	area	scale of reality
physical	light, sound, force, etc	law of mechanics and electricity	signal processing	explainability
			virtual reality	causality
logical	language, symbols	logic and syntax	knowledge processing	provability
	algebra		artificial intelligence	logicality
emotional	music, art	subjective	emotional info. processing	subjective consistency
	expression	like and dislike	artificial (virtual) emotion	

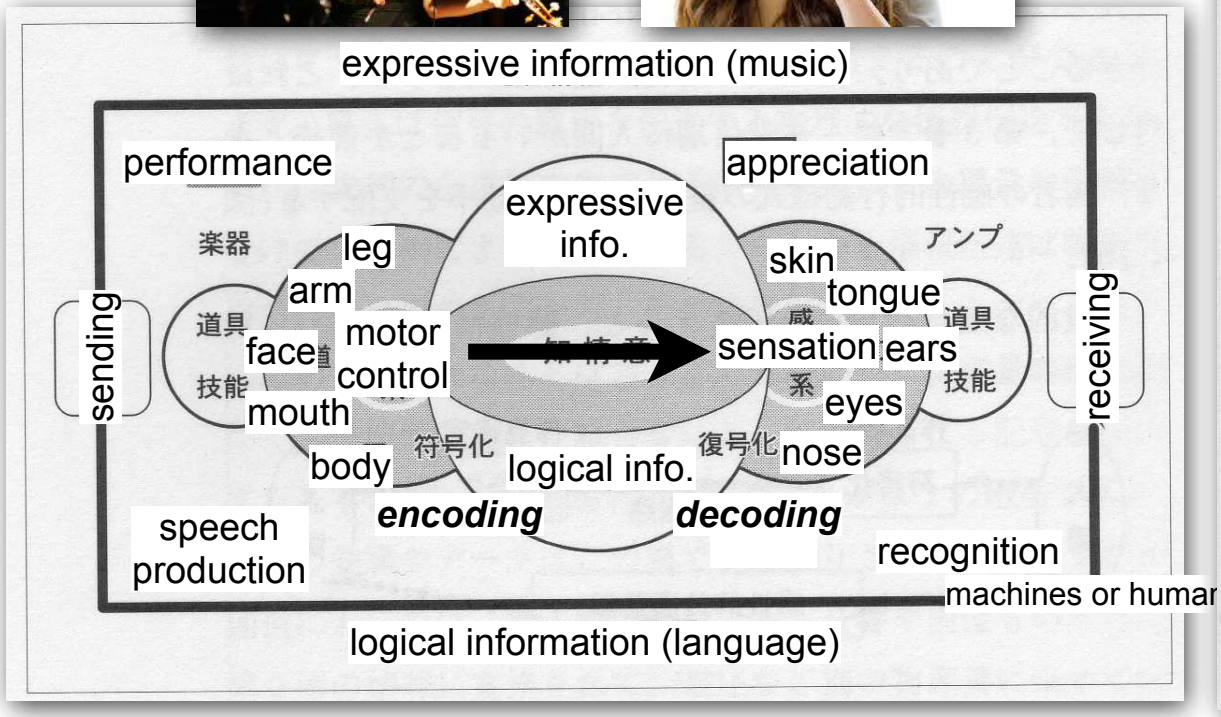
layer	what to transmit	method of reception	effect on reception	terminals
physical	features	observation / conversion	changes due to physical law	machines
logical	symbols	recognition	changes due to syntax or rules	machines or humar
emotional	fea. + sym.	appreciation	changes due to emotions	humans

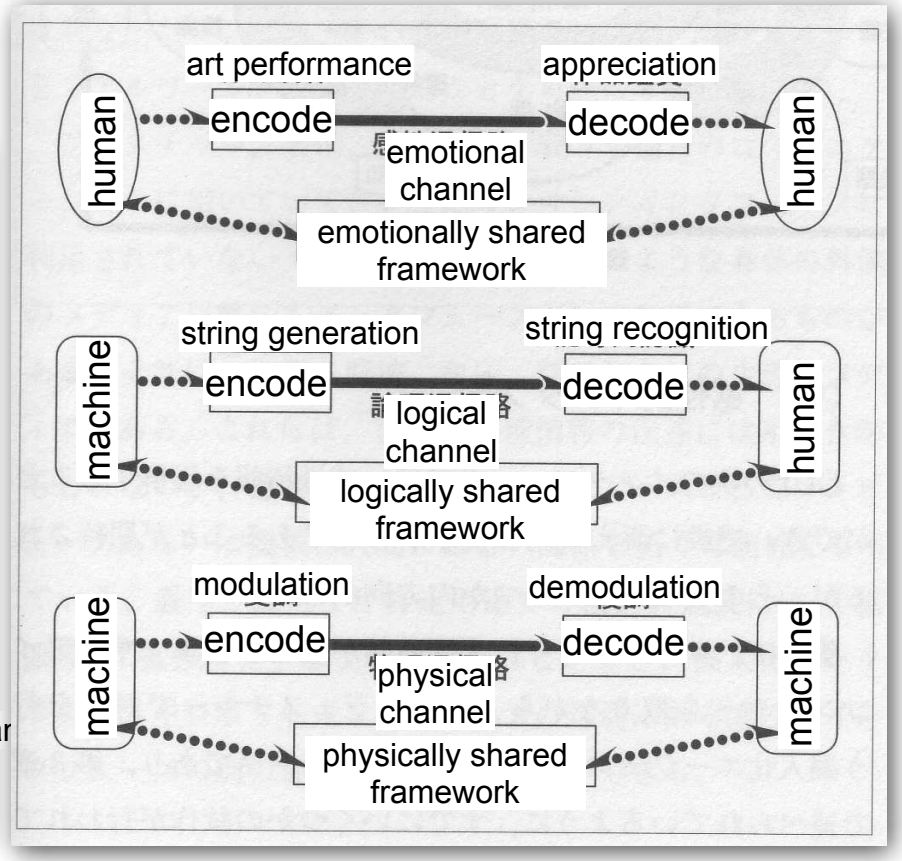




Encoding, transmission and decod

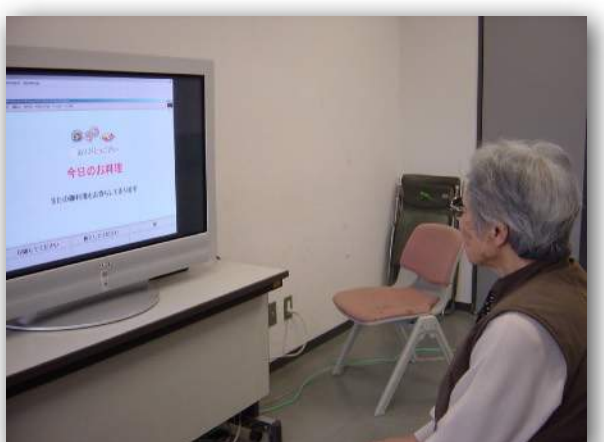
- Encoding and decoding of expressive information
  - Encoding: bodily actions are often used for encoding
  - Decoding: the bodily actions are sensed through the five senses
    - Bodily actions -- patterns of multimedia features -- sensation through the five senses



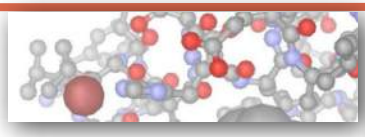


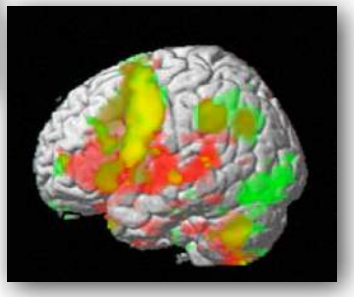
From GUI to PUI

- Expressive and perceptual user interface
  - GUI = Graphical User Interface
  - PUI = Perceptual User Interface
    - Personified agent software (on the screen)
    - Human-shaped robot agent
      - Bodily actions are exchanged between robots and users
    - Physiological features are used for human-machine interface
      - Heat beat rate, blood pressure, brain waves, etc
  - An example of PUI interface
    - Acoustic features of speech are used to estimate the age of a user
      - The computer interface is adapted to the age of users.

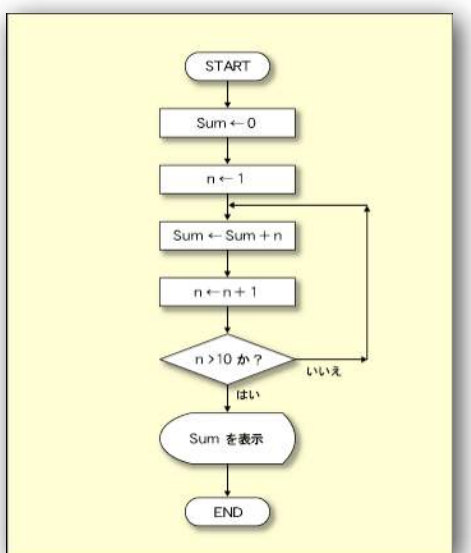


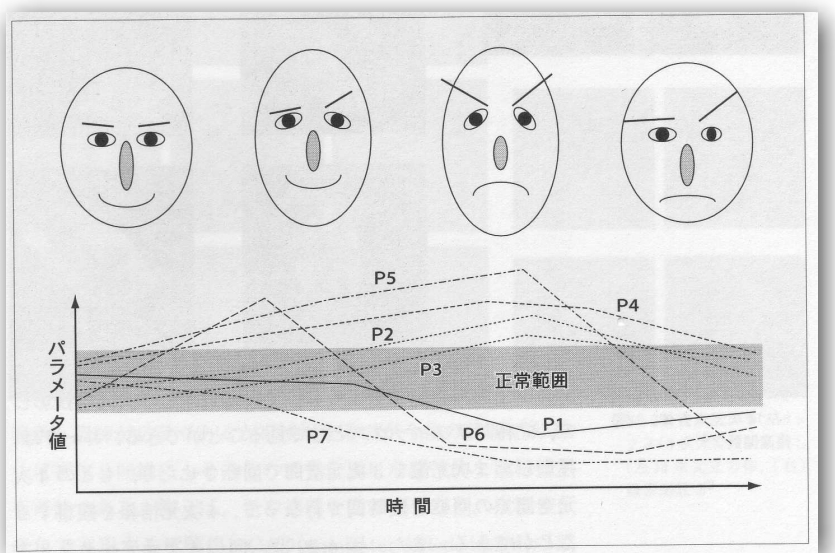
Intuitive understanding using expressions

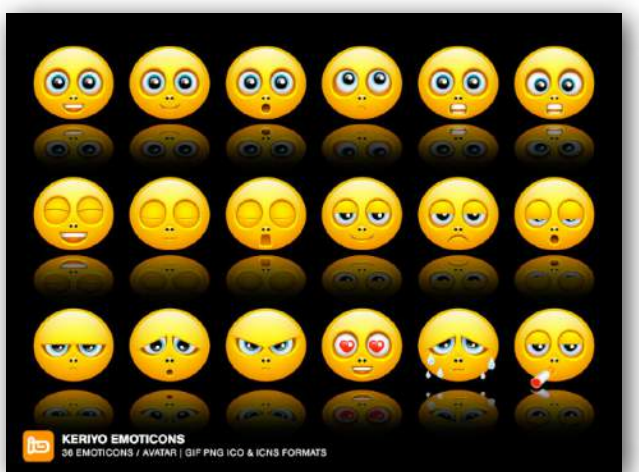




- Visualization of expressive information
  - Can lead to easy and intuitive understanding (intuitive = 直感的)
  - Visualization of information on the physical layer
    - Visualization of molecules and electro-magnetic phenomena on the brain
  - Visualization of information on the logical layer
    - Flowchart of a computer program (algorithm)
  - Visualization of information on the expressive layer
    - Facial display / face icons / acoustic presentation of a painting is possible?

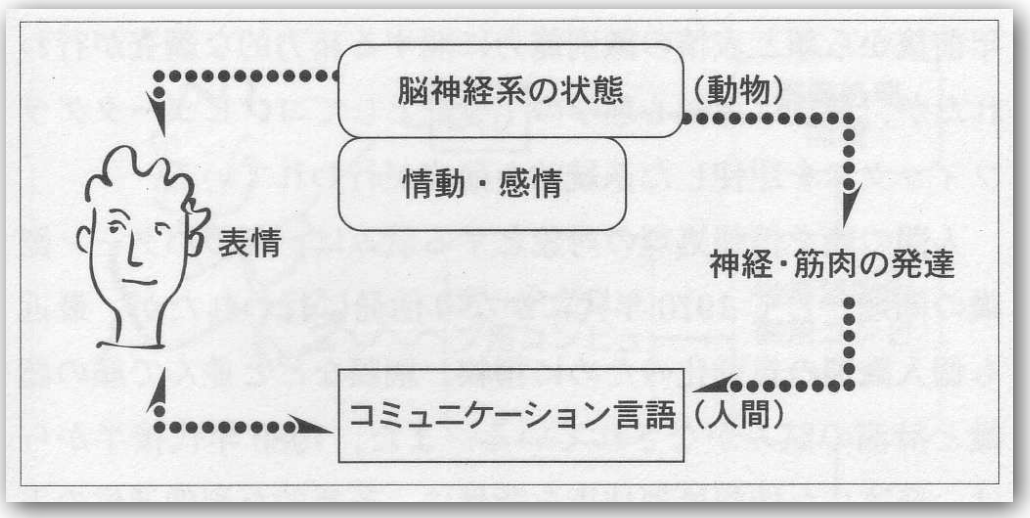






Face! Face! Face!

- Facial expression and expressive information
  - Control of facial muscles = unconscious control
    - This is the reason why an intentional (forced/fake) smile is unnatural.
  - Expressive (and unconscious) information communication using faces.
    - This strategy is possessed only by humans and anthropoid apes (類人猿).
    - Autistic individuals have a severe difficulty on reading facial expressions.
      - Lack of the theory of mind.
    - Typically developed individuals, even young children, are experts of reading faces.
  - Newly born babies (1 hour after birth) have ability of face discrimination
    - Face discrimination ability is inborn, not acquired through experiences.



Face! Face! Face!

- Which are faces?
  - Newly born babies have good interests in face-like icons.
  - But the left icon is an abstract representation of a face, not a real one.
- Good portraits (likenesses)
  - Abstraction or emphasis of the person's characteristics
  - We have a good ability to discover a common pattern between the two images.

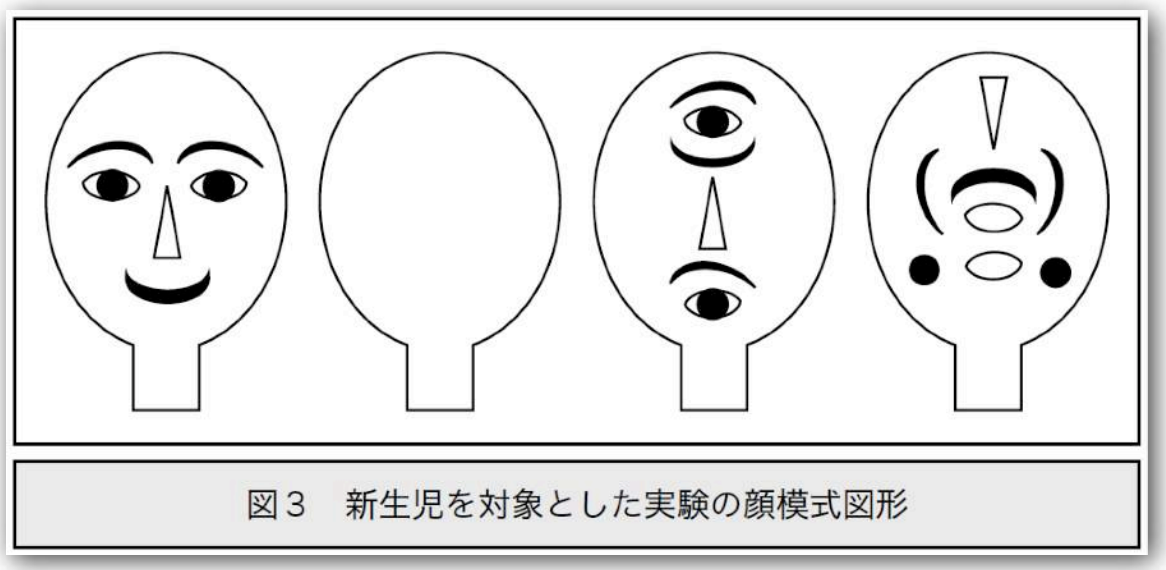
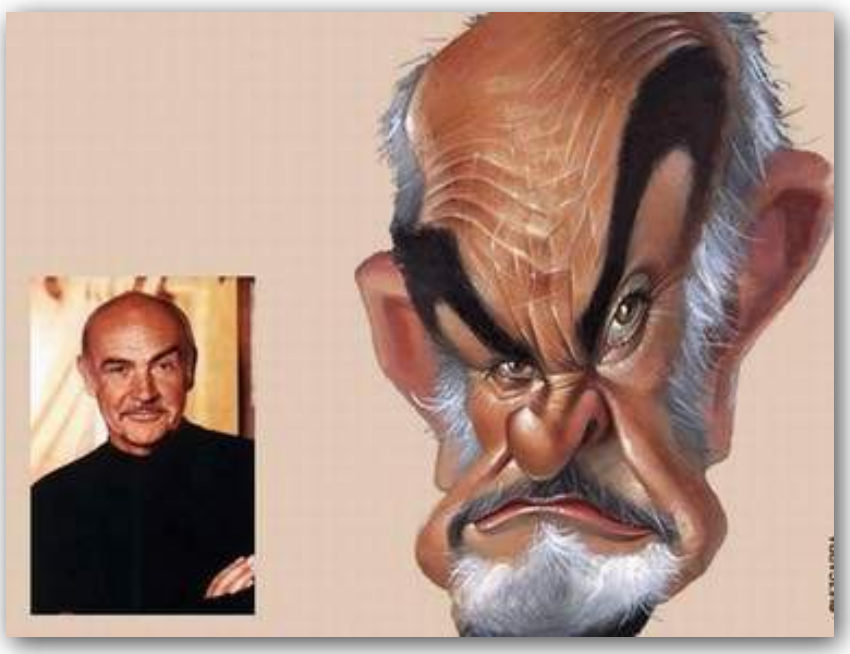
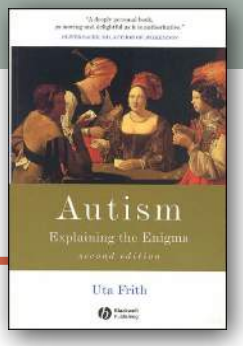


図３　新生児を対象とした実験の顔模式図形





Takako Tokiwa and Yu Darvish

- Looks similar but we can discriminate the two faces.
  - Some autistics claim that they cannot discriminate the two faces.

  - They are very good at detecting local features.
    - Noses, eyes, eyebrows, lips, ears, etc
  - But very poor at detecting holistic and global features.
    - This may be the cause for them to have difficulty in reading facial expressions?

Clinton and Gore

- Looks similar but we can discriminate the two faces.

CORRESPONDENCE

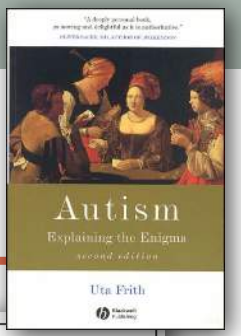
I think I know that face...

SIR — Exactly what do we recognize in a face? Intuition suggests that it is the eyes, nose and mouth — they, after all, are what the dictionary uses to define a 'face'. Portraitists labour to get these features right and poets describe the eyes as sure betrayers of identity. Computer scientists, not to be outdone, have designed vision systems that rely on precise measurements or templates of these 'internal' features to recognize faces[1]. This view often seems justified.

But some images, such as the one shown here, suggest that there might be more to face recognition than just an analysis of the internal facial features. The image

recognition systems might stand to benefit by incorporating, in some measure, the head processing strategy.

Pawan Sinha
Tomaso Poggio
E25-201, Department of Brain and Cognitive Sciences,
Massachusetts Institute of Technology,
Cambridge,
Massachusetts 02142, USA



Face! Face! Face!

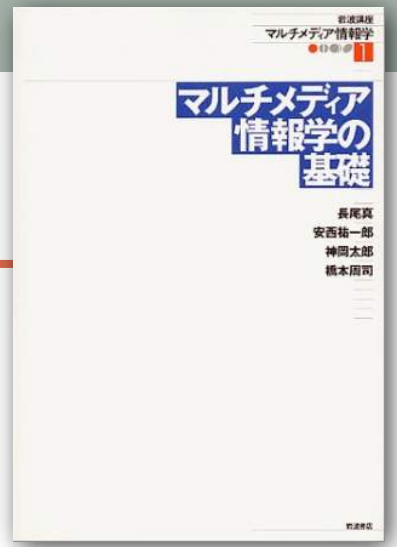
- Some experimental facts

図2　サッチャー錯視（Thompson, 1980）

Elements first?

Holistic patterns first?

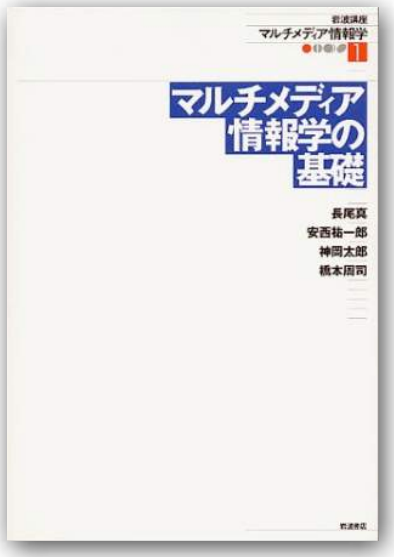
図1　顔と見えるか，果物と見えるか
（M. Moscovitch et al, Journal of Cognitive Neuroscience, 9, 1997）
普通はこの絵を見て，顔と果物を同時に見ることができる．しかし，物体失認の患者はこれに人の顔は見えても，果物を見ることができなかった．反対に相貌失認の患者の場合は，果物ばかりが目立つであろう．

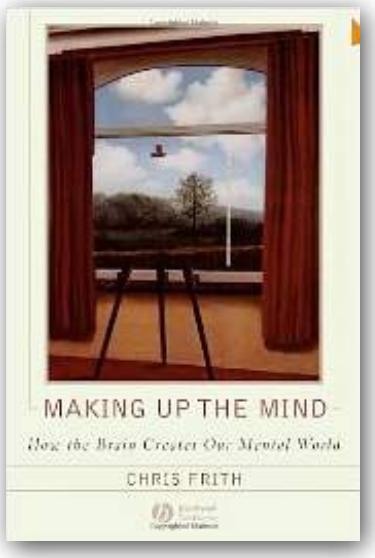


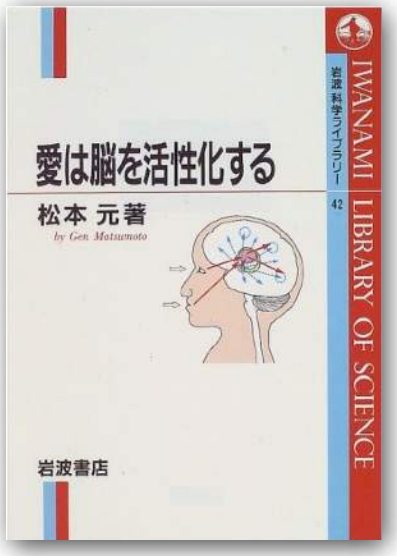
Today's menu

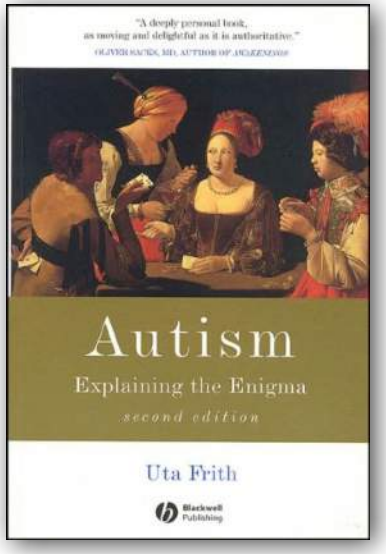
- Emotional and expressive information processing
  - Definition of emotional and expressive information
  - Quantitative measurement of the information
  - Emotion and the brain -- which region of the brain deals with the emotion?
- Emotional information and multimedia sensory information
  - The five senses and emotion
  - Integration of information transmitted using different kinds of media
- Transmission or communication of emotional information
  - Encoding, transmission, and decoding of the information
  - PUI, visualization, facial expressions, etc
  - Some examples of emotional systems
- Summary
- Announcement on the next lecture
  - The first assignment is given at the end of the next lecture.

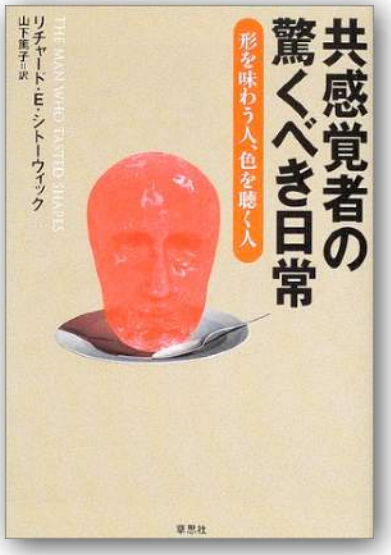
Recommended books











Announcement on the next lecture

- Date: Oct 24 (Tue), 14:55 -- 16:40
- After showing some slides, a 45-min documentary film on synesthesia will be presented. The video files are stored at Google Drive at UTokyo.
  - **https://bit.ly/CMP-D4**
  - Two versions are prepared (English and Japanese)
  - You can select which one to view based on your language performance.

- The 1st assignment is given “before” you view the video.