Towards Artificial Communication Partners with a Multiagent Mind Model Based on Mental Image Directed Semantic Theory

Masao Yokota
Fukuoka Institute of Technology
Japan

1. Introduction

In recent years, there have been developed various types of real or virtual robots as artificial communication partners. However, they are to play their roles according to programmed actions to stimuli and have not yet come to understand or imitate delicate mental functions of their human partners such as *Kansei*, one of the topics in this chapter. Kansei evaluates non-scientific matters such as art, music, natural scenery, etc. by Kansei words (e.g., 'heart-calming', 'fantastic', 'grotesque') (Fukuda et al, 1998; Sugita et al, 2003; Butz et al, 2005) and 'Artificial Kansei' namely 'Kansei for robots' is expected to play a part in 'artificial or robotic individuality' (Tauchi et al, 2006).

The author has proposed a human mind model consisting of Stimulus, Knowledge, Emotion and Response Processing Agents, intending the intelligent system IMAGES-M (Yokota, 2005a; Shiraishi et al, 2005) to understand and imitate miscellaneous human mental functions involving *emotion processing* as well as *knowledge processing* originally aimed at. For example, Kansei is defined and realized as tight collaboration of Knowledge and Emotion Processing Agents. This multiagent mind model is much simpler than Minsky's (Minsky, 1986) and its most remarkable feature is that the agents involved communicate with one another by exchanging and computing mental images represented in the formal language $\boldsymbol{L_{md}}$ developed for integrated representation and computation of multimedia information (Yokota, 2005, 2006) while other multiagent systems were designed to employ special programming languages for inter-agent communication (e.g., Labrou et al, 1999; Vieira et al, 2007).

IMAGES-M is originally intended for integrated multimedia understanding as *knowledge processing* for intuitive human-robot interaction such that may happen between ordinary (or non-expert) people and home robots (Yokota et al, 2008). Here, 'integrated multimedia understanding' means especially 'multimedia understanding based on such a knowledge representation common to multiple information media (i.e., natural language, picture, music, gesture, etc.) that can facilitate cross-media operations (Yokota et al, 1984; Eakins & Graham, 1999; Kherfi et al, 2004; Yokota & Capi, 2005a)'. For ordinary people, however, natural language is the most important because it can convey the exact intention of the sender to the receiver due to its syntax and semantics common to its users, which is not necessarily the

case for another information medium such as gesture or so. Therefore, natural language can as well play the most crucial role in intuitive human-robot interaction and actually IMAGES-M is a language-centered multimedia understanding system where its attention mechanism is to be controlled efficiently in a top-down way according to people's suggestions in natural language (Yokota, 2007).

For integrated multimedia understanding intended here, it is essential to develop a systematically computable knowledge representation language (KRL) as well as representation-free technologies (Brooks, 1986) such as neural networks for unstructured sensory/motory data processing (i.e., *stimulus/response processing*). This type of language is indispensable to *knowledge-based* processing such as *understanding* sensory events, *planning* appropriate actions and *knowledgeable* communication with ordinary people in natural language, and therefore it needs to have at least a good capability of representing spatiotemporal events that correspond to humans'/robots' sensations and actions in the real world (Yokota & Capi, 2005b).

Most of conventional methods have provided robotic systems with such quasi-natural language expressions as 'move(*Velocity*, *Distance*, *Direction*)', 'find(*Object*, *Shape*, *Color*)', etc. for human instruction or suggestion, uniquely related to computer programs for deploying sensors/motors as their semantics (e.g., Coradeschi & Saffiotti, 2003; Drumwright et al, 2006). These expression schemas, however, are too linguistic or coarse to represent and compute sensory/motory events in such an integrated way as intuitive human-robot interaction intended here. This is also the case for AI planning ("action planning") which deals with the development of representation languages for planning problems and with the development of algorithms for plan construction (Wilkins & Myers, 1995).

In order to solve this problem, the author has employed the formal language so called 'Language for Mental-image Description ($\boldsymbol{L}_{md}$)' proposed in his original semantic theory 'Mental Image Directed Semantic Theory (MIDST)' (e.g., Yokota, 2005, 2006), the main topic in this chapter.

MIDST is based on the hypothesis that natural language understanding in humans is basically mental image processing and concerns:

(i) Modeling of the human mind as a society of agents;
(ii) Modeling of omnisensory mental image in humans;
(iii) Modeling of conceptualization in humans;
(iv) Designing the formal language $\boldsymbol{L}_{md}$ for mental image description;
(v) Formulation of word concepts as mental images in $\boldsymbol{L}_{md}$;
(vi) Mutual translation between expressions in miscellaneous information media (e.g., natural language, picture, robotic action, etc.) and those in $\boldsymbol{L}_{md}$;
(vii) Computation on $\boldsymbol{L}_{md}$ expressions;
(viii) Formalization of human mental competence and performance as a deductive system in $\boldsymbol{L}_{md}$;

The final goal of MIDST is to realize artificial communication partners with a good capability of intuitive interaction with ordinary people and the items (i)-(viii) above are assumed to be its subgoals. The key idea of MIDST is the model of human attention-guided (i.e., active) perception yielding omnisensory images that inevitably reflect certain movements of the focus of attention of the observer (FAO) scanning certain matters in the world, either inside or outside of the mind. More analytically, these omnisensory images are associated with spatiotemporal changes (or constancies) in certain attributes of the matters scanned by FAO and modeled as temporally parameterized "loci in attribute spaces", so

called, to be formulated in the formal language $\boldsymbol{L_{md}}$. This language has already been implemented on several types of computerized intelligent systems including IMAGES-M (e.g., Yokota et al, 1984; Oda, et al, 2001; Amano, et al, 2005; Yokota & Capi, 2005a). The most remarkable feature of $\boldsymbol{L_{md}}$ is its capability of formalizing spatiotemporal matter concepts grounded in human/robotic sensation while the other similar KRLs are designed to describe the logical relations among conceptual primitives represented by lexical tokens (e.g., Dorr & Bonnie, 1997; Zarri, 1997; Sowa, 2000). Moreover, in $\boldsymbol{L_{md}}$ expression are hinted what and how should be attended to in the world as analogy of human FAO movement and thereby the robotic attention can be controlled in a top-down way (Yokota, 2007), which is the author's answer to the essential issue in robotic imitation, namely, how to control robotic attention mechanism efficiently (e.g., Demiris & Khadhouri, 2006).

The remainder of this chapter is organized as follows. Section 2 presents MIDST, focusing on the multiagent mind model, the omnisensory mental image model and the formal language $\boldsymbol{L_{md}}$ with liguistic or pictorial manifestations for its validation. Section 3 details about grounding natural language expressions in mental images in view of natural language processing by computers. In Section 4, the mental fucntion Kansei is modeled as collaboration of **Kn** and **Em** viewed from artificial or robotic individuality. Section 5 presents a discussion on applications and further developments for the language $\boldsymbol{L_{md}}$ presented in this chapter. Conclusions and planned future work are given in the final section.

2. Mental Image Directed Semantic Theory

2.1 Multiagent mind model

Figure 1 shows the multiagent mind model proposed here, consisting of Stimulus, Knowledge, Emotion and Response Processing Agents. This is a functional model of human central nervous system consisting of the brain and the spine. These agents are to communicate with one another by exchanging and computing mental images represented in the formal language $\boldsymbol{L_{md}}$. Their basic performances are as follows.

1) **Stimulus Processing Agent** (**St**) receives stimuli from the world (**W**) and encodes them into mental images (i.e. encoded sensations) such as *"I sensed something oily."* (if verbalized in English.)
2) **Knowledge Processing Agent** (**Kn**) evaluates mental images received from the other agents based on its memory (e.g. knowledge), producing other mental images such as *"It is false that the earth is flat."*
3) **Emotion Processing Agent** (**Em**) evaluates mental images received from the other agents based on its memory (e.g. instincts), producing other mental images such as *"I like the food."*
4) **Response Processing Agent** (**Re**) converts mental images (i.e. encoded actions such as *"I'll walk slowly."*) received from the other agents into real actions against **W**.

A *performance* ***P*** against a *stimulus X* with a *result Y* at each agent can be formalized as a function by the expression (1).

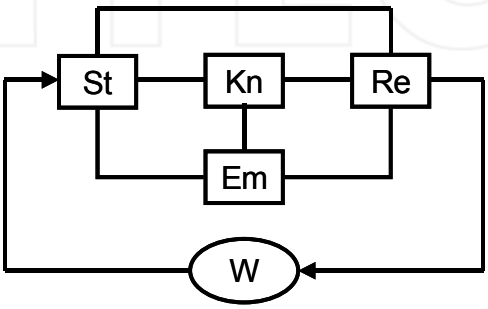
$$Y=\mathbf{P}(X) \tag{1}$$

where

P: a combination of ***Atomic Performances*** defined later in association with ***Attribute Spaces***;
X: a spatiotemporal distribution of stimuli from **W** to **St** or a mental image for another agent.
Y: a series of signals to drive actuators for **Re** or a mental image for another agent.

For example, all the agents are to work during understanding information media such as natural language, picture, music, gesture, etc., sometimes performing Kansei by tight collaboration of **Kn** and **Em** as detailed later while **St** and **Re** are exclusively to work during *reflection* so called.

A performance **P** is assumed as a function formed either consciously or unconsciously, or in other words, either with or without reasoning. In a conscious case, a set of atomic performances are to be chosen and combined according to *X* by a meta-function, so called, '***Performance Selector***' assumed as '***Conscience***'. On the contrary, in an unconscious case, such a performance as is associated most strongly with *X* is to be applied automatically as is in the case of reflection.



Multiagent model of human mind (St: Stimulus Processing Agent; Kn: Knowledge Processing Agent; Em: Emotion Processing Agent; Re: Response Processing Agent; W: World surrounding human mind, including his/her body).

2.2 Omnisensory mental image model and formal language L_{md}

Here are described the omnisensory mental image model and the syntax and semantics of L_{md} in association with the mental image model. In MIDST, word meanings are defined in association with mental images, not limited to visual but omnisensory, modeled as "Loci in Attribute Spaces" so called. See Fig.2-a and assume that the human is observing the phenomenon where the triangular gray object is moving in the sinusoidal trajectory and that its corresponding sensations (i.e., sensory images) are being caused in his/her mind. In this case, the moving triangular gray object is assumed to be perceived as the loci in the three attribute spaces, namely, those of 'Location', 'Color' and 'Shape' in the observer's mind. As easily imagined, attribute spaces correspond with human sensory systems and the loci represent certain sensations of the phenomena outside or inside human minds. From the viewpoint of artifact, an attribute space stands for a certain measuring instrument or sensor just like a chronograph, barometer, thermometer or so and the loci represent the movements of its indicator. The performance of an attribute space is the model of '***Atomic Performance***' introduced in Section 2.1.

These loci are to be articulated by "Atomic Locus" over a certain absolute time interval $[t_i, t_f]$ as depicted in Fig.2-b and formulated as (2) in L_{md}, where the interval is suppressed because people are not aware of absolute time (nor always consult a chronograph).

$$L(x,y,p,q,a,g,k) \tag{2}$$

The expression (2) works as a formula in many-sorted predicate logic, where "L" is a predicate constant with five types of terms: "Matter" (at 'x' and 'y'), "Value" (at 'p' and 'q'), "Attribute" (at 'a'), "Event Type" (at 'g') and "Standard" (at 'k'). Conventionally, Matter variables are headed by 'x', 'y' and 'z'. This formula is called 'Atomic Locus Formula' whose first two arguments are sometimes referred to as 'Event Causer (EC)' and 'Attribute Carrier (AC)', respectively while ECs are often optional in natural concepts such as intransitive verbs. For simplicity, the syntax of $\boldsymbol{L_{md}}$ allows Matter terms (e.g., 'Tokyo' and 'Osaka' in (3) and (4)) to appear at Values or Standard in order to represent their values in each place at the time or over the time-interval. Moreover, when it is not so significant to discern ECs or Standards, anonymous variables, usually symbolized as '_', can be employed in their places (See (23) for example). A logical combination of atomic locus formulas defined as a well-formed formula (i.e., wff) in predicate logic is called simply 'Locus Formula'.

The intuitive interpretation of (2) is given as follows.

"Matter 'x' causes Attribute 'a' of Matter 'y' to keep* (p=q) *or change* (p ≠ q) *its values temporally* (g=Gt) *or spatially* (g=Gs) *over a certain absolute time-interval, where the values 'p' and 'q' are relative to the standard 'k'."

In (2), when $g=G_t$, the locus indicates monotonic change (or constancy) of the attribute in time domain, and when $g=G_s$, that in space domain. The former is called 'temporal event' and the latter, 'spatial event'. For example, the motion of the 'bus' represented by S1 is a temporal event and the ranging or extension of the 'road' by S2 is a spatial event whose meanings or concepts are formulated as (3) and (4), respectively, where 'A_{12}' denotes the attribute 'Physical Location'. These two formulas are different only at the term 'Event Type'.

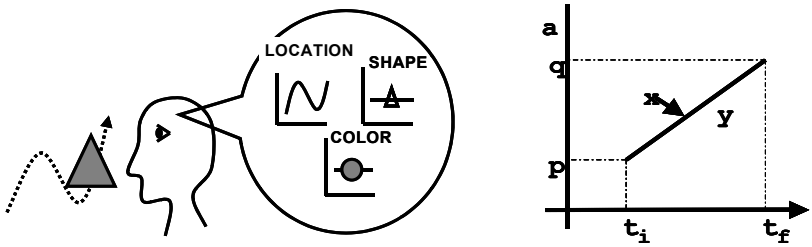
(S1) The bus runs from Tokyo to Osaka.

(S2) The road runs from Tokyo to Osaka.

$$(\exists x,y,k)L(x,y,\text{Tokyo},\text{Osaka},A_{12},G_t,k)\wedge \text{bus}(y) \tag{3}$$

$$(\exists x,y,k)L(x,y,\text{Tokyo},\text{Osaka},A_{12},G_s,k)\wedge \text{road}(y) \tag{4}$$

It has been often argued that human active sensing processes may affect perception and in turn conceptualization and recognition of the physical world while such cognitive processes or products have seldom been formulated for computation (e.g., Leisi, 1961; Noton, 1970; Gardenfors, 2000; Langacker, 2005). The author has hypothesized that the difference between temporal and spatial event concepts can be attributed to the relationship between the Attribute Carrier (AC) and the Focus of the Attention of the Observer (FAO). To be brief, it is hypothesized that FAO is fixed on the whole AC in a temporal event but *runs* about on the AC in a spatial event. Consequently, as shown in Fig.3, the *bus* and FAO move together in the case of S1 while FAO solely moves along the *road* in the case of S2. That is, ***all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal events of FAO***.



(a) (b)

Fig. 2. Mental image model (a) and Atomic Locus in Attribute Space (b)

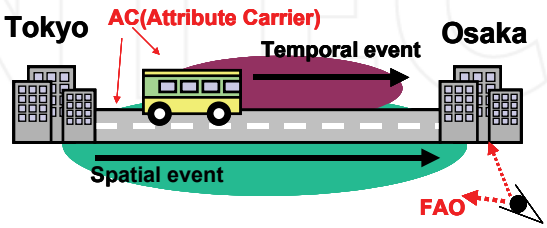


Fig. 3. FAO movements and Event types

2.3 Tempological connectives

The duration of a locus corresponds to an absolute time-interval over which FAO is put on the corresponding phenomenon outside or inside the mind. Such an absolute time-interval is suppressed in an atomic locus formula because it is assumed that people cannot measure the absolute time by any chronograph but a certain relative time (Actually, people do not always consult a chronograph even if they can). MIDST has employed 'tempo-logical connectives (TLCs)' denoting both logical and temporal relations between loci by themselves because these must be considered simultaneously in locus articulation.

A tempo-logical connective K_i is defined by (5), where τ_i, χ and K refer to one of the temporal relations indexed by an integer 'i', a locus, and an ordinary binary logical connective such as the conjunction '$\wedge$', respectively. The definition of τ_i is given in Table 1 from which the theorem (6) can be deduced. This table shows the complete list of topological relations between two intervals, where 13 types of relations are discriminated by τ_i ($-6 \leq i \leq 6$). This is in accordance with Allen's notation (Allen, 1984), which, to be strict, is exclusively for 'temporal conjunctions ($=\wedge_i$)' as introduced below.

$$\chi_1 \, K_i \, \chi_2 \Leftrightarrow (\chi_1 \, K \, \chi_2) \wedge \tau_i(\chi_1, \chi_2) \tag{5}$$

$$\tau_{-i}(\chi_2, \chi_1) \equiv \tau_i(\chi_1, \chi_2) \quad (\forall i \in \{0, \pm1, \pm2, \pm3, \pm4, \pm5, \pm6\}) \tag{6}$$

The TLCs used most frequently are 'SAND ($\wedge_0$)' and 'CAND ($\wedge_1$)', standing for 'Simultaneous AND' and 'Consecutive AND' and conventionally symbolized as 'Π' and '•', respectively. For example, the concepts of the English verbs 'carry' and 'return' are to be defined as (7) and (8), respectively. These can be depicted as Fig.4-a and b, respectively. The

expression (9) is the definition of the English verb concept 'fetch' depicted as Fig.4-c. This implies such a temporal event that 'x' goes for 'y' and then comes back with it. In the same way, the English verb concept 'hand' or 'receive' depicted as Fig.4-d is defined equivalently as (10) or its abbreviation (10′) where ECs are merged into a set.

$$(\lambda x,y)\mathrm{carry}(x,y) \Leftrightarrow (\lambda x,y)(\exists p,q,k)L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p,q,A_{12},G_t,k) \wedge x\neq y \wedge p\neq q \tag{7}$$

$$(\lambda x)\mathrm{return}(x) \Leftrightarrow (\lambda x)(\exists p,q,k)L(x,x,p,q,A_{12},G_t,k)\bullet L(x,x,p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q \tag{8}$$

$$(\lambda x,y)\mathrm{fetch}(x,y) \Leftrightarrow (\lambda x,y)\ (\exists p_1,p_2,k)\ L(x,x,p_1,p_2,A_{12},G_t,k)\bullet$$

$$((\underline{L(x,x,p_2,p_1,A_{12},G_t,k)\Pi L(x,y,p_2,p_1,A_{12},G_t,k)}) \wedge x\neq y\wedge p_1\neq p_2 \tag{9}$$

$$(\lambda x,y,z)\mathrm{hand}(x,y,z).\equiv.(\lambda x,y,z)\mathrm{receive}(z,,y,x)$$

$$\Leftrightarrow (\lambda x,y,z)(\exists k)L(x,y,x,z,A_{12},G_t,k)\Pi L(z,y,x,z,A_{12},G_t,k)\wedge x\neq y\wedge y\neq z\wedge z\neq x \tag{10}$$

$$(.\equiv.\ (\lambda x,y,z)(\exists k)L(\{x,z\},y,x,z,A_{12},G_t,k)\wedge x\neq y\wedge y\neq z\wedge z\neq x) \tag{10'}$$

Such locus formulas as correspond with natural event concepts are called 'Event Patterns' and about 40 kinds of event patterns have been found concerning the attribute 'Physical Location (A_{12})', for example, *start*, *stop*, *meet*, *separate*, *carry*, *return*, etc.

Temporal relations and definition of τ_i†			Allen's notation
χ_1 +·····+ χ_2 +·····+	$t_{11}=t_{21}$ $\wedge t_{12}=t_{22}$	$\tau_0(\chi_1, \chi_2)$	equals(χ_1,χ_2)
		$\tau_0(\chi_2, \chi_1)$	equals(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{12}=t_{21}$	$\tau_1(\chi_1, \chi_2)$	meets(χ_1, χ_2)
		$\tau_{-1}(\chi_2, \chi_1)$	met-by(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{11}=t_{21}$ $\wedge t_{12}<t_{22}$	$\tau_2(\chi_1, \chi_2)$	starts(χ_1, χ_2)
		$\tau_{-2}(\chi_2, \chi_1)$	started-by(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{11}>t_{21}$ $\wedge t_{12}<t_{22}$	$\tau_3(\chi_1, \chi_2)$	during(χ_1, χ_2)
		$\tau_{-3}(\chi_2, \chi_1)$	contains(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{11}>t_{21}$ $\wedge t_{12}=t_{22}$	$\tau_4(\chi_1, \chi_2)$	finishes(χ_1, χ_2)
		$\tau_{-4}(\chi_2, \chi_1)$	finished-by(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{12}<t_{21}$	$\tau_5(\chi_1, \chi_2)$	before(χ_1, χ_2)
		$\tau_{-5}(\chi_2, \chi_1)$	after(χ_2, χ_1)
χ_1 +·····+ χ_2 +·····+	$t_{11}<t_{21}\wedge t_{21}<t_{12}$ $\wedge t_{12}<t_{22}$	$\tau_6(\chi_1, \chi_2)$	overlaps(χ_1, χ_2)
		$\tau_{-6}(\chi_2, \chi_1)$	overlapped-by(χ_2, χ_1)

Table 1. List of temporal relations (†χ_1 and χ_2 exist during [t_{11}, t_{12}] and [t_{21}, t_{22}], respectively)

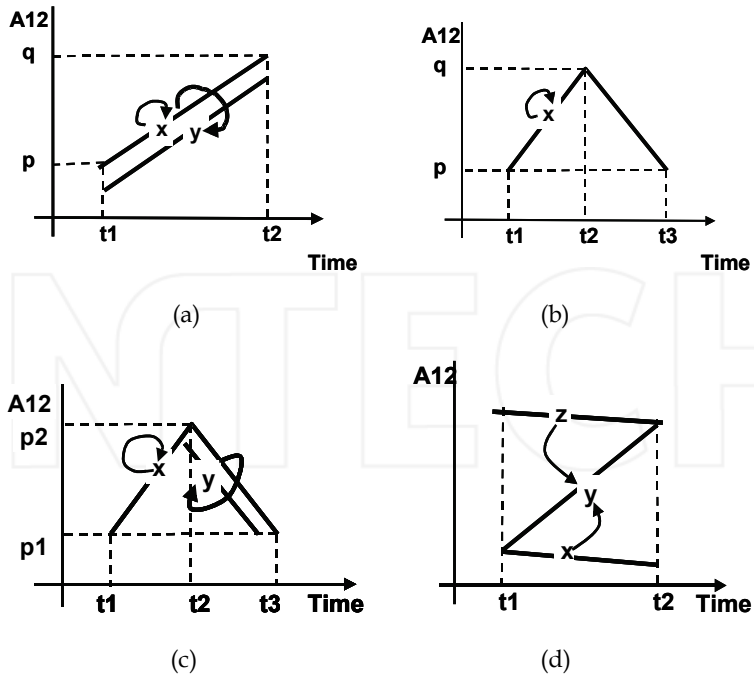


Fig. 4. Depiction of loci: 'carry' (a), 'return' (b), 'fetch' (c) and 'hand/receive' (d)

In order for explicit indication of time points, a very important concept called 'Empty Event (EE)' denoted by 'ε' is introduced. An EE stands only for absolute time elapsing and is explicitly defined as (11) with the attribute 'Time Point (A_{34})' and the standard 'K_{Ta}' denoting *absolute time*, where t_i and t_j are conventionally given as real numbers with the condition $t_i<t_j$. According to this scheme, the duration $[t_a, t_b]$ of an arbitrary locus χ can be expressed as (12).

$$\varepsilon([t_i,t_j]) \Leftrightarrow (\exists x,y,g)\ L(x,y,\ t_i,t_j,A_{34},g,K_{Ta}) \tag{11}$$

$$\chi \Pi \varepsilon([t_a,t_b]) \tag{12}$$

Any pair of loci temporally related in certain attribute spaces can be formulated as (13)-(17) in exclusive use of SANDs, CANDs and EEs. For example, the loci shown in Fig.5-a and b correspond to the formulas (14) and (17), respectively.

$$\chi_1 \wedge_2 \chi_2 .\equiv. (\chi_1 \bullet \varepsilon)\Pi\chi_2 \tag{13}$$

$$\chi_1 \wedge_3 \chi_2 .\equiv. (\varepsilon_1 \bullet \chi_1 \bullet \varepsilon_2)\Pi\chi_2 \tag{14}$$

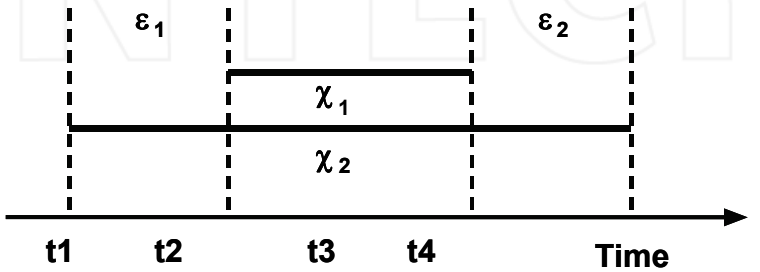
$$\chi_1 \wedge_4 \chi_2 .\equiv. (\varepsilon \bullet \chi_1)\Pi\chi_2 \tag{15}$$

$$\chi_1 \wedge_5 \chi_2 \mathrel{.\equiv.} \chi_1 \bullet \varepsilon \bullet \chi_2 \tag{16}$$

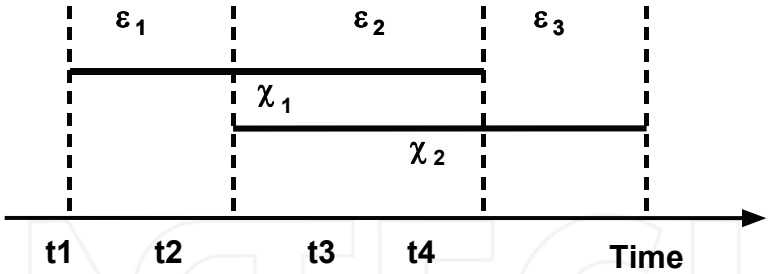
$$\chi_1 \wedge_6 \chi_2 \mathrel{.\equiv.} (\chi_1 \bullet \varepsilon_3)\Pi(\varepsilon_1 \bullet \chi_2)\Pi(\varepsilon_1 \bullet \varepsilon_2 \bullet \varepsilon_3) \tag{17}$$

Employing TLCs, tempo-logical relationships between miscellaneous event concepts can be formulated without explicit indication of time intervals. For example, an event 'fetch(x,y)' is necessarily *finished by* an event 'carry(x,y)' as indicated by the underline at (9). This fact can be formulated as (18), where '$\supset_{-4}$' is the '*implication* ($\supset$)' furnished with the temporal relation '*finished-by* (τ_{-4})'. This kind of formula is not an axiom but a theorem deducible from the definitions of event concepts in the deductive system intended here.

$$(\forall x,y)(\text{fetch}(x,y) \supset_{-4} \text{carry}(x,y)) \tag{18}$$



(a)



(b)

Fig. 5. Tempological relations: (a) during(χ_1, χ_2) and (b) overlaps(χ_1, χ_2)

2.4 Attributes and standards

The attribute spaces for humans correspond to the sensory receptive fields in their brains. At present, about 50 attributes concerning the physical world have been extracted as shown in Table 2 exclusively from Japanese and English words (e.g., Roget, 1975). They are associated with all of the 5 senses (i.e. sight, hearing, smell, taste and feeling) in our everyday life while those for information media other than languages correspond to limited senses. For example, those for pictorial media, marked with '*' in Table 2, associate limitedly with the

sense 'sight' as a matter of course. The attributes of this sense occupy the greater part of all, which implies that the sight is essential for humans to conceptualize the external world by. And this kind of classification of attributes plays a very important role in our cross-media operating system (Yokota & Capi, 2005a).

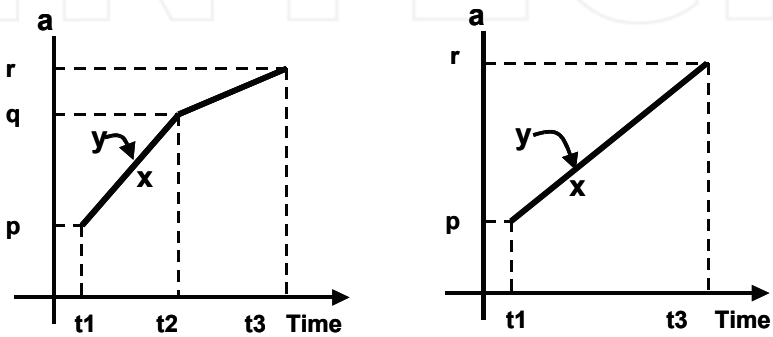
Correspondingly, six categories of standards shown in Table 3 have been extracted after the conventional categorization (Leisi, 1961) that are assumed necessary for representing values of each attribute in Table 2. In general, the attribute values represented by words are relative to certain standards as explained briefly in Table 3. For example, (19) and (20) are different formulations of a locus due to the different standards 'k1' and 'k2' for scaling as shown in Fig.6-a and b, respectively. That is, whether the point (t_2, q) is significant or not, more generally, how to articulate a locus depends on the precisions or the granularities of these standards, which can be formulated as (21) and (22), so called, '***Postulate of Arbitrariness in Locus Articulation***'. This postulate affects the process of conceptualization on a word based on its referents in the world.

Code	Attribute [Property]†	Linguistic expressions for attribute values
*A_{01}	PLACE OF EXISTE NCE [N]	The accident happened in Osaka.
*A_{02}	LENGTH [S]	The stick is 2 meters long.
	…….	
*A_{11}	SHAPE [N]	The cake is round.
*A_{12}	PHYSICAL LOCATION [N]	Tom moved to Tokyo.
*A_{13}	DIRECTION [N]	The box is to the left of the chair.
*A_{14}	ORIENTATION [N]	The door faces to south.
*A_{15}	TRAJECTORY [N]	The plane circled in the sky.
*A_{16}	VELOCITY [S]	The boy runs very fast.
*A_{17}	MILEAGE [S]	The car ran ten miles.
A_{18}	STRENGTH OF EFFECT [S]	He is very strong.
A_{19}	DIRECTION OF EFFECT [N]	He pulled the door.
	……	
A_{28}	TEMPERATURE [S]	It is hot today.
A_{29}	TASTE [N]	The grapes here are very sour.
A_{30}	ODOUR [N]	The gas is pungent.
A_{31}	SOUND [N]	His voice is very loud.
*A_{32}	COLOR [N]	Tom painted the desk white.
A_{33}	INTERNAL SENSATION [N]	I am very tired.
A_{34}	TIME POINT [S]	It is ten o'clock.
A_{35}	DURATION [S]	He studies for two hours every day.
A_{36}	NUMBER [S]	Here are many people.
A_{37}	ORDER [S]	Tom sat next to Mary.
A_{38}	FREQUENCY [S]	He did it twice.
A_{39}	VITALITY [S]	The old man still alive.
*A_{44}	TOPOLOGY [N]	He is in the room.
*A_{45}	ANGULARITY [S]	The knife is dull.

Table 2. Examples of attributes (†S: 'scalar value', N : 'non-scalar value')

Categories of standards	Remarks
Rigid Standard	Objective standards such as denoted by measuring *units* (meter, gram, etc.).
Species Standard	The *attribute value ordinary* for a species. A *short train* is ordinarily longer than a *long pencil*.
Proportional Standard	'*Oblong*' means that the width is greater than the height at a physical object.
Individual Standard	*Much* money for one person can be too *little* for another.
Purposive Standard	One room large enough for a person's *sleeping* must be too small for his *jogging*.
Declarative Standard	Tom is taller *than Jim*. The origin of an order such as 'next' must be declared explicitly just as 'next *to him*'.

Table 3. List of standards



(a) (b)

Fig. 6. Arbitrariness in locus articulation due to standards: Standard k_1 (a) is finer than k_2 (b)

$$(L(y,x,p,q,a,g,k_1)\Pi\ \varepsilon([t_1,t_2]))\bullet(L(y,x,q,r,a,g,k_1)\Pi\ \varepsilon([t_2,t_3])) \tag{19}$$

$$L(y,x,p,r,a,g,k_2)\Pi\ \varepsilon([t_1,t_3]) \tag{20}$$

$$(\forall p,q,r,k)(L(y,x,p,q,a,g,k)\bullet L(y,x,q,r,a,g,k).\supset.(\exists k')L(y,x,p,r,a,g,k')\wedge k'\neq k) \tag{21}$$

$$(\forall p,r,k)(L(y,x,p,r,a,g,k).\supset.(\exists q,k')\ L(y,x,p,q,a,g,k')\bullet L(y,x,q,r,a,g,k')\wedge\ k'\neq k) \tag{22}$$

3. Mind model and natural language

3.1 Spatiotemporal expressions and perceptual processes

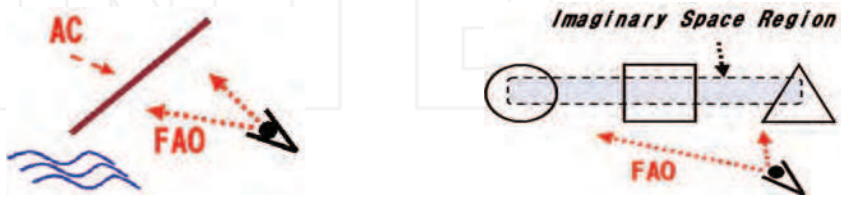
As already mentioned in Section 2, all loci in attribute spaces are assumed to correspond one to one with movements or, more generally, temporal events of FAO. Therefore, an event expressed in $\boldsymbol{L_{md}}$ is compared to a movie film recorded through a floating camera because it is necessarily grounded in FAO's movement over the event. And this is why S3 and S4 can

refer to the same scene in spite of their appearances, where what 'sinks' or 'rises' is FAO as illustrated in Fig.7-a and whose conceptual descriptions are given as (23) and (24), respectively, where 'A_{13}', '↑' and '↓' refer to the attribute 'Direction' and its values 'upward' and 'downward', respectively.

(S3) The path *sinks to* the brook.

(S4) The path *rises from* the brook.

Such a fact is generalized as '***Postulate of Reversibility of a Spatial Event (PRS)***' that can be one of the principal inference rules belonging to people's common-sense knowledge about geography. This postulation is also valid for such a pair of S5 and S6 as interpreted approximately into (25) and (26), respectively. These pairs of conceptual descriptions are called equivalent in the PRS, and the paired sentences are treated as paraphrases each other.



(a) (b)

Fig. 7. FAO movements: 'slope' (a) and 'row' (b) as spatial events

$$(\exists y,p,z)L(_,y,p,z,A_{12},G_{s},_)\Pi L(_,y,\downarrow,\downarrow,A_{13},G_{s},_)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (23)$$

$$(\exists y,p,z)L(_,y,z,p,A12,G_{s},_)\Pi L(_,y,\uparrow,\uparrow,A_{13},G_{s},_)\wedge path(y)\wedge brook(z)\wedge p\neq z \quad (24)$$

(S5) Route A and Route B meet at the city.

$$(\exists p,y,q)L(_,Route_A,p,y,A_{12},G_{s},_)\Pi L(_,Route_B,q,y,A_{12},G_{s},_)\wedge city(y)\wedge p\neq q \quad (25)$$

(S6) Route A and Route B separate at the city.

$$(\exists p,y,q)L(_,Route_A,y,p,A_{12},G_{s},_)\Pi L(_,Route_B,y,q,A_{12},G_{s},_)\wedge city(y)\wedge p\neq q \quad (26)$$

For another example of spatial event, Fig.7-b concerns the perception of the formation of multiple objects, where FAO runs along an imaginary object so called '***Imaginary Space Region (ISR)***'. This spatial event can be verbalized as S7 using the preposition 'between' and formulated as (27) or (27'), corresponding also to such concepts as 'row', 'line-up', etc. Employing ISRs and the 9-intersection model (Egenhofer, 1991), all the topological relations between two objects can be formulated in such expressions as (28) or (28') for S8, and (29) for S9, where '*In*', '*Cont*' and '*Dis*' are the values 'inside', 'contains' and 'disjoint' of the attribute 'Topology (A_{44})' with the standard '9-intersection model (K_{9IM})', respectively. Practically, these topological values are given as 3×3 matrices with each element equal to 0 or 1 and therefore, for example, '*In*' and '*Cont*' are transpositional matrices each other.

(S7) The square is between the triangle and the circle.
(S8) Tom is in the room.
(S9) Tom exits the room.

$$(\exists y,x_1,x_2, x_3,p,q)\ (L(_,y,x_1,x_2,A_{12},G_{s,}_)\Pi L(_,y,p,p,A_{13},G_{s,}_))\bullet(L(_,y,x_2,x_3,A_{12},G_{s,}_)$$

$$\Pi L(_,y,q,q,A_{13},G_{s,}_))\wedge ISR(y)\wedge p=q\ \wedge triangle(x_1)\wedge square(x_2)\ \wedge circle(x_3) \quad (27)$$

$$(\exists y,x_1,x_2, x_3,p)(L(_,y,x_1,x_2,A_{12},G_{s,}_)\bullet L(_,y,x_2,x_3,A_{12},G_{s,}_))\Pi L(_,y,p,p,A_{13},G_{s,}_)$$

$$\wedge ISR(y)\ \wedge triangle(x_1)\wedge square(x_2)\ \wedge circle(x_3) \quad (27')$$

$$(\exists x,y)L(Tom,x,y,Tom,A_{12},G_{s,}_)\Pi L(Tom,x,In,In,A_{44},G_t,K_{9IM})\ \wedge ISR(x)\ \wedge room(y) \quad (28)$$

$$(\exists x,y)L(Tom,x,Tom,y,A_{12},G_{s,}_)\Pi L(Tom,x,Cont,Cont,A_{44},G_t,K_{9IM})\wedge ISR(x)\ \wedge room(y) \quad (28')$$

$$(\exists x,y,p,q)L(Tom,Tom,p,q,A_{12},G_{t,}_)\Pi L(Tom,x,y,Tom,A_{12},G_{s,}_)$$

$$\Pi L(Tom,x,In,Dis,A_{44},Gt,K_{9IM})\ \wedge ISR(x)\ \wedge room(y)\ \wedge p\neq q \quad (29)$$

The rigid topology between two objects as in the 9-intersection model must be determined with the perfect knowledge of their insides, outsides and boundaries. Ordinary people, however, can comment on matters without knowing all about them. This is the very case when they encounter an unknown object too large to observe at a glance just like a road in a strange country. For example, Fig.8-a shows such a path viewed from the sky that is partly hidden by the woods. In this case, the topological relation between the path as a whole and the swamp/woods depends on how the path starts and ends in the woods, but people could utter such sentences as S10 and S11 about this scene. Actually, these sentences refer to such spatial events that reflect certain temporal changes in the topological relation between the swamp/woods and FAO running along the path. Therefore, their conceptual descriptions are to be given as (30) and (31), respectively. For another example, Fig.8-b shows a more complicated spatial event in topology that can be formulated as (32) and could be verbalized as S12.

(S10) The path enters the swamp/woods.
(S11) The path exits the swamp/woods.

$$(\exists x,y,z)L(_,z,p,q,A_{12},G_{s,}_)\Pi L(_,x,y,z,A_{12},G_{s,}_)\Pi$$

$$L(_,x,Dis,In,A_{44},G_s,K_{9IM})\ \wedge ISR(x)\ \wedge\{swamp(y)/woods(y)\}\ \wedge path(z)\ \wedge p\neq q \quad (30)$$

$$(\exists x,y,z)L(_,z,p,q,A_{12},G_{s,}_)\Pi L(_,x,y,z,A_{12},G_{s,}_)\Pi$$

$$L(_,x,In,Dis,A_{44},G_s,K_{9IM})\ \wedge ISR(x)\ \wedge\{swamp(y)/woods(y)\}\ \wedge path(z)\ \wedge p\neq q \quad (31)$$

(S12) The path cuts the swamp twice (as shown in Fig.8-b), passing p_1 outside, p_2 inside, p_3 outside, p_4 inside and p_5 outside the swamp on the way.

$$(\exists x,y,z,p_1,\ldots,p_5)L(_,z,y,x,A_{12},G_s,_)\Pi((L(_,x,p_1,p_2,A_{12},G_s,_)$$

$$\Pi(L(_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(_,x,p_2,p_3,A_{12},G_s,_)\Pi L(_,z,In,Dis,A_{44},G_s,K_{9IM}))$$
$$\bullet(L(_,x,p_3,p_4,A_{12},G_s,_)\Pi L(_,z,Dis,In,A_{44},G_s,K_{9IM}))\bullet(L(_,x,p_4,p_5,A_{12},G_s,_)\Pi$$

$$L(_,z,In,Dis,A_{44},G_s,K_{9IM})))\wedge path(x)\wedge swamp(y)\wedge ISR(z) \tag{32}$$

Lastly, consider such somewhat complicated sentences as S13 and S14. The underlined parts are deemed to refer to some events neglected in time and in space, respectively. These events correspond with skipping of FAOs and are called 'Temporal Empty Event' and 'Spatial Empty Event', denoted by 'ε_t' and 'ε_s' as EEs with g=G_t and g=G_s at (11), respectively. The concepts of S13 and S14 are given by (33) and (34), where 'A_{15}' and 'A_{17}' represent the attribute 'Trajectory' and 'Mileage', respectively.

(S13) The *bus* runs 10km straight east from A to B, and *after a while*, at C it meets the street with the sidewalk.

(S14) The *road* runs 10km straight east from A to B, and *after a while*, at C it meets the street with the sidewalk.

$$(\exists x,y,z,p,q)(L(_,x,A,B,A_{12},G_t,_)\Pi L(_,x,0,10km,A_{17},G_t,_)\Pi$$

$$L(_,x,Point,Line,A_{15},G_t,_)\Pi L(_,x,East,East,A_{13},G_t,_))\bullet\varepsilon_t\bullet \tag{33}$$

$$(L(_,x,p,C,A_{12},G_t,_)\Pi L(_,y,q,C,A_{12},G_s,_)\Pi L(_,z,y,y,A_{12},G_s,_))$$

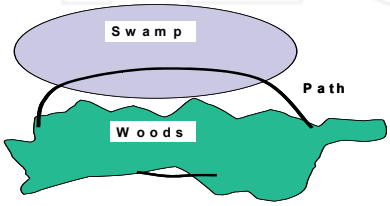
$$\wedge bus(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q$$

$$(\exists x,y,z,p,q)(L(_,x,A,B,A_{12},G_s,_)\Pi L(_,x,0,10km,A_{17},G_s,_)\Pi$$

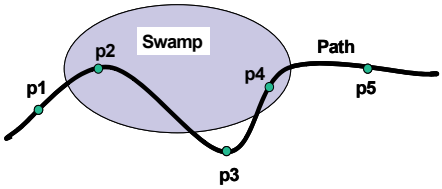
$$L(_,x,Point,Line,A_{15},G_s,_)\Pi\ L(_,x,East,East,A_{13},G_s,_))\bullet\varepsilon_s\bullet \tag{34}$$

$$(L(_,x,p,C,A_{12},G_s,_)\Pi\ L(_,y,q,C,A_{12},G_s,_)\Pi L(_,z,y,y,A_{12},G_s,_))$$
$$\wedge road(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q$$

From the viewpoint of cross-media reference as integrated multimedia understanding, the formula (34) can refer to such a spatial event depicted as the still picture in Fig.9 while (33) can be interpreted into a motion picture.



(a)



(b)

Fig. 8. Delicate topological relations: (a) the path partially hidden by the woods and (b) the path winding inside-outside-inside-outside of the swamp

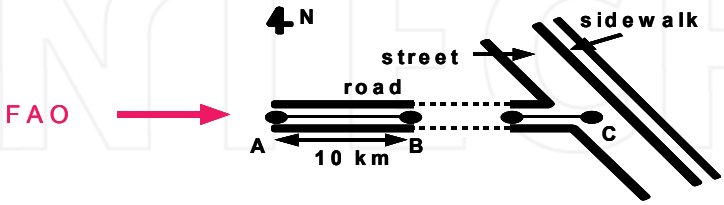


Fig. 9. Pictorial interpretation of the formula (34) and the movement of FAO involved

3.2 Conceptualization

It is well known that, from the cognitive viewpoint, there are two types of mental images, namely, (a) perceptual images and (b) conceptual images. The former are live images of the current world and the latter are recalled ones often in association with tokens such as words.

Ideally, a word concept should be associated with such a conceptual image that is abstract enough to represent the perceptual image of every matter referred to by the word. It is, however, practically impossible for an individual to obtain such a conceptual image because such instances or referents are usually too numerous for him/her to encounter and observe. In this sense, our conceptual image for a word is always imperfect or tentative to be sometimes updated by an exceptional instance just like a 'black swan'.

It is generally assumed that a word concept is an abstraction on properties and relations of the matters involved such as locations, shapes, colors, functions, potentialities, etc. In MIDST, a word concept is to be represented as an abstract locus formula resulted from generalization on the locus formulas of a number of matters referred to by the word.

Figure 10 illustrates the mental process of conceptualization on the word 'conveyance', where a set of its referents ***Sr*** (={*Matter_1*,…, *Matter_n*}) are generalized by abstraction and formulated as (35). The underlined part of this formula implies that matter 'z' includes two matters 'x' and 'y' in its 'Place of existence (A_{01})'. As easily imagined, the variable 'z' denotes a certain referent generalized so as to represent any member of ***Sr***.

This process consists of three stages as follows. Firstly, the attributes other than 'Place of existence (A_{01})' and 'Physical location (A_{12})' are discarded. Secondly, the concrete objects 'human', 'book', etc. and their concrete attribute values are replaced by the variables 'x', 'y', 'z', 'p', 'q', etc. And finally, the relationships '≠' and '=', the most essential for this concept, are placed among these variables. The equalities (or inequalities) in 'Physical location' are determined at the precision of the standard represented by the variable 'k'.

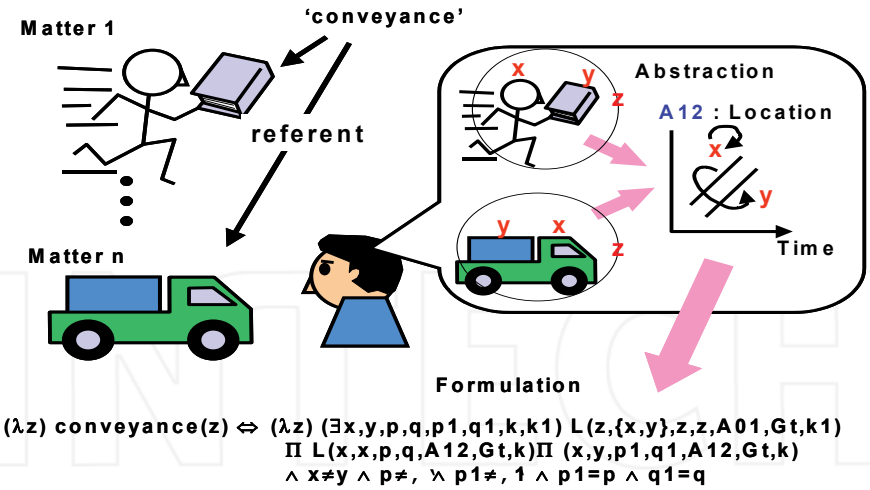


Fig.10. Conceptualization: the process of abstraction on referents of the word 'conveyance'.

$$(\lambda x)\text{conveyance}(z) \Leftrightarrow (\lambda z)(\exists x,y,p,q,p_1,q_1,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}$$

$$\Pi L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p_1,q_1,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q\wedge p_1\neq q_1\wedge p_1=p\wedge q_1=q$$

$$\Leftrightarrow (\lambda z)(\exists x,y,p,q,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}$$

$$\Pi L(x,x,p,q,A_{12},G_t,k)\Pi L(x,y,p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q$$

$$\Leftrightarrow (\lambda z)(\exists x,y,p,q,k,k_1)\underline{L(z,\{x,y\},z,z,A_{01},G_t,k_1)}\Pi L(x,\{x,y\},p,q,A_{12},G_t,k) \wedge x\neq y \wedge p\neq q \tag{35}$$

For another example, the matter called 'snow' can be conceptualized as (36), where '_', as defined by (37), stands for the variable bound by the existential quantifier, reading 'Snow is powdered ice attracted from the sky by the earth, melts into water,…' (Hence forth, refer to Table 2 for undefined attributes.)

$$(\lambda x)\text{snow}(x)\Leftrightarrow(\lambda x)(\exists x_1,x_2,\ldots)((L(_,x,x_1,x_1,A_{41},G_t,_)\prod L(\text{Earth},x,\text{Sky},\text{Earth},A_{12},G_t,_))$$

$$\wedge L(_,x,x_1,x_2,A_{41},G_t,_) \wedge \text{powder}(x_1)\wedge \text{ice}(x_1)\wedge \text{water}(x_2) \wedge\ldots) \tag{36}$$

$$L(\ldots,_,\ldots) \leftrightarrow (\exists x)L(\ldots,x,\ldots) \tag{37}$$

For a more complicated example, the concept of 'umbrella' can be represented as (38), reading 'At raining, a human puts an umbrella in line between rain and himself/herself in order not to get wetter, …'

By the way, the concepts of 'rain' and 'wind' can be given as (39) and (40), reading 'Rain is water attracted from the sky by the earth, makes an object wetter, is pushed an umbrella to by a human,…,' and 'Wind is air, affects the direction of rain,… ,' respectively.

$$(\lambda x)\text{umbrella}(x) \Leftrightarrow (\lambda x)\ (\exists x,x_1,x_2,y,p,q_1,q_2,\ldots)\ (L(_,x_1,p,p,A_{13},G_t,_)\prod$$

$$((L(_,y,x_1,x,A_{12},G_s,_)\bullet L(_,y,x,x_2,A_{12},G_s,_))\prod L(_,y,p,p,A_{13},G_s,_)\prod$$

$$L(x_1,x_2,q_1,q_2,A_{25},G_t,_) \wedge \text{ISR}(y) \wedge \text{rain}(x_1) \wedge \text{human}(x_2) \wedge \sim(q_1<q_2)\ \ldots) \tag{38}$$

$$(\lambda x)\text{rain}(x) \Leftrightarrow (\lambda x)(\exists x_1,x_2,x_3,x_4,p,q,\ldots)L(_,x,x_1,x_1,A_{41},G_t,_)\prod$$

$$L(\text{Earth},x,\text{Sky},\text{Earth},A_{12},G_t,_)\prod L(x,x_2,p,q,A_{25},G_t,_)\prod L(x_3,x_4,x,x,A_{19},G_t,x_3)$$

$$\wedge\text{water}(x_1)\wedge\text{object}(x_2)\wedge\text{human}(x_3)\wedge\text{umbrella}(x_4)\wedge(p<q)\ldots \tag{39}$$

$$(\lambda x)\text{wind}(x) \Leftrightarrow (\lambda x)(\exists x_1,x_2,p,q,\ldots)L(_,x,x_1,x_1,A_{41},G_t,_)\wedge\text{air}(x_1)\wedge$$

$$(L(x,x_2,p,q,A_{13},G_t,_)\wedge\text{rain}(x_2)\ldots \tag{40}$$

3.3 Knowledge of word meanings

A word meaning M_W is defined as a pair of 'Concept Part (C_p)' and 'Unification Part (U_p)' and is formulated as (41).

$$M_W = [C_p{:}U_p] \tag{41}$$

The C_p of a word W is an $\boldsymbol{L_{md}}$ expression as its concept while its U_p is a set of operations for unifying the C_ps of W's syntactic governors or dependents. For example, the meaning of the English verb 'carry' can be given by (42).

$$[(\lambda x,y)(\exists p,q,k)L(x,\{x,y\},p,q,A_{12},G_t,k)\wedge x\neq y\wedge p\neq q\text{: ARG(Dep.1,x); ARG(Dep.2,y);}] \tag{42}$$

The U_p above consists of two operations to unify the first dependent (Dep.1) and the second dependent (Dep.2) of the current word with the variables x and y, respectively. Here, Dep.1 and Dep.2 are the 'subject' and the 'object' of 'carry', respectively. Therefore, the sentence '*Mary carries a book*' is translated into (43).

$$(\exists y,p,q,k)\ L(\text{Mary},\{\text{Mary},y\},p,q,A_{12},G_t,k)\wedge\text{Mary}\neq y\ \wedge p\neq q\ \wedge\text{book}(y) \tag{43}$$

Figure 11 shows the details of the conversion process of a surface structure (text) into a conceptual structure (text meaning) through a surface dependency structure.

For another example, the meaning description of the English preposition 'through' is also given by (44).

$$[(\lambda x,y)(\exists p_1,z,p_3,g,k,p_4,k_0)\ (\underline{L(x,y,p_1,z,A_{12},g,k)}\bullet L(x,y,z,p_3,A_{12},g,k))\Pi$$

$$L(x,y,p_4,p_4,A_{13},g,k_0)\ \wedge p_1\neq z\ \wedge z\neq p_3\text{: ARG(Dep.1,z);}$$

$$\text{IF(Gov=Verb)} \rightarrow \text{PAT(Gov,(1,1))}; \text{ IF(Gov=Noun)} \rightarrow \text{ARG(Gov,y);]} \tag{44}$$

The U_p above is for unifying the C_ps of the very word, its governor (Gov, a verb or a noun) and its dependent (Dep.1, a noun). The second argument (1,1) of the command PAT indicates the underlined part of (44) and in general (i,j) refers to the partial formula covering from the ith to the jth atomic formula of the current C_p. This part is the pattern common to both the C_ps to be unified. This is called 'Unification Handle (U_h)' and when missing, the C_ps are to be combined simply with '$\wedge$'.

Therefore the sentences S15, S16 and S17 are interpreted as (45), (46) and (47), respectively. The underlined parts of these formulas are the results of PAT operations. The expression (48) is the C_p of the adjective 'long' implying 'there is some value greater than some standard of Length (A_{02}),' which is often simplified as (48').

(S15) The train runs through the tunnel.

$$(\exists x,y,p_1,z,p_3,k,p_4,k_0)\ (\underline{L(x,y,p_1,z,A_{12},G_t,k)} \bullet L(x,y,z,p_3,A_{12},G_t,k))$$

$$\Pi\ L(x,y,p_4,p_4,A_{13},G_t,k_0) \wedge p_1 \neq z \wedge z \neq p_3 \wedge \text{train}(y) \wedge \text{tunnel}(z) \tag{45}$$

(S16) The path runs through the forest.

$$(\exists x,y,p_1,z,p_3,k,p_4,k_0)\ (\underline{L(x,y,p_1,z,A_{12},G_s,k)} \bullet L(x,y,z,p_3,A_{12},G_s,k))\Pi$$

$$L(x,y,p_4,p_4,A_{13},G_s,k_0) \wedge p_1 \neq z \wedge z \neq p_3 \wedge \text{path}(y) \wedge \text{forest}(z) \tag{46}$$

(S17) The path through the forest is long.

$$(\exists x,y,p_1,z,p_3,x_1,k,q,k_1,p_4,k_0)\ (L(x,y,p_1,z,A_{12},G_s,k) \bullet L(x,y,z,p_3,A_{12},G_s,k))\Pi$$

$$L(x,y,p_4,p_4,A_{13},G_s,k_0)\Pi L(x_1,y,q,q,A_{02},G_t,k_1) \wedge p_1 \neq z \wedge z \neq p_3 \wedge q > k_1 \wedge \text{path}(y) \wedge \text{forest}(z) \tag{47}$$

$$(\exists x_1,y_1,q,k_1) L(x_1,y_1,q,q,A_{02},G_t,k_1) \wedge q > k_1 \tag{48}$$

$$(\exists x_1,y_1,k_1) L(x_1,y_1,\text{Long},\text{Long},A_{02},G_t,k_1) \tag{48'}$$

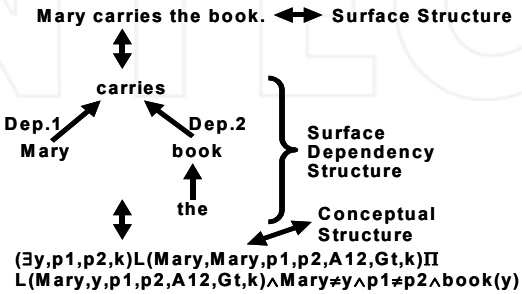


Fig. 11. Process of semantic interpretation of text

3.4 Fundamental semantic computations

Every version of the intelligent system IMAGES can perform text understanding based on word meanings as follows.

Firstly, a text is parsed into a surface dependency structure (or more than one if *syntactically* ambiguous). Secondly, each surface dependency structure is translated into a conceptual structure (or more than one if *semantically* ambiguous) based on word meanings. Finally, each conceptual structure is semantically evaluated.

The fundamental semantic computations on $\boldsymbol{L_{md}}$ expressions are performed to detect semantic anomalies, ambiguities and paraphrase relations in texts. Semantic anomaly detection is very important to cut off meaningless computations. For example, consider such a conceptual structure as (49), where 'A_{39}' is the attribute 'Vitality'. This locus formula can correspond to the English sentence 'The desk is alive', which is usually semantically anomalous because a 'desk' does never have vitality in the real world projected into the attribute spaces.

$$(\exists x)L(_,x,\text{Alive},\text{Alive},A_{39},G_{t},_)\wedge \text{desk}(x) \tag{49}$$

This kind of semantic anomaly can be detected in the following process. Firstly, assume the concept of 'desk' as (50), where 'A_{29}' refers to the attribute 'Taste'. The special symbols '*' and '/' are defined as (51) and (52) representing 'always' and 'no value', respectively.

$$(\lambda x)\ \text{desk}(x) \leftrightarrow (\lambda x)\ (\ldots L^*(_,x,/,/,A_{29},G_{t},_) \wedge \ldots \wedge L^*(_,x,/,/,A_{39},G_{t},_) \wedge \ldots) \tag{50}$$

$$X^* \leftrightarrow (\forall[p,q])X\ \Pi\ \varepsilon([p,q]) \tag{51}$$

$$L(\ldots,/,\ldots) \leftrightarrow \sim(\exists p)\ L(\ldots,p,\ldots) \tag{52}$$

Secondly, the postulates (53) and (54) are utilized. The formula (53) means that ***<u>if one of two loci exists every time interval, then they can coexist</u>***. The formula (54) states that ***<u>a matter has never different values of an attribute with a standard at a time</u>***.

$$X \wedge Y^* .\supset. X\ \Pi\ Y \tag{53}$$

$$L(x,y,p_1,q_1,a,g,k)\ \Pi\ L(z,y,p_2,q_2,a,g,k)\ .\supset.\ p_1=p_2 \wedge q_1=q_2 \tag{54}$$

Lastly, the semantic anomaly of 'alive desk' is detected by using (50)-(54). That is, the formula (55) below is finally deduced from (49)-(54) and violates the commonsense given by (54), that is, " *Alive* ≠ / ".

$$(\exists x)L(_,x,\text{Alive},\text{Alive},A_{39},G_{t},_)\ \Pi\ L(_,x,/,/,A_{39},G_{t},_) \tag{55}$$

This process above is also employed for dissolving such a syntactic ambiguity as found in S18. That is, the semantic anomaly of 'alive desk' is detected and eventually 'alive insect' is adopted as a plausible interpretation.

(S18) Look at the insect on the desk, which is still alive.

If a text has multiple plausible interpretations, it is semantically ambiguous. For example, S19 alone has two plausible interpretations (56) and (57) different at the underlined parts, implying 'Jack with the stick' and 'Tom with the stick', respectively.

(S19) Tom follows Jack with the stick.

$$(\exists x)(L(Jack,Jack,p,q,A_{12},G_t,_)\underline{\Pi L(Jack,x,Jack,Jack,A_{12},G_t,_)})\bullet L(Tom,Tom,p,q,A_{12},G_t,_)\wedge p\neq q \tag{56}$$

$$(\exists x)L(Jack,Jack,p,q,A_{12},G_t,_)\bullet(L(Tom,Tom,p,q,A_{12},G_t,_)\underline{\Pi}\ \underline{L(Tom,x,Tom,Tom,A_{12},G_t,_)})\wedge p\neq q\wedge stick(x) \tag{57}$$

Among the fundamental semantic computations, detection of paraphrase relations is the most essential because it is for detecting equalities in semantic descriptions and the other two are for detecting inequalities in them. In the deductive system intended here, if two different texts are interpreted into the same locus formula, they are paraphrases of each other. For example, the sentence '*Mary goes with a book*' is interpreted into (58) which is proved to be equivalent to (43), the semantic description of '*Mary carries a book*'. In the process of this proof, the axioms (59) and (60) concerning the syntax of $\boldsymbol{L_{md}}$ are utilized.

$$(\exists y,p,q,k)L(Mary,Mary,p,q,A_{12},G_t,k)\Pi L(Mary,y,Mary,Mary,A_{12},G_t,k)\wedge p\neq q\wedge book(y) \tag{58}$$

$$(\forall x_1,x_2,x_3,x_4,p,q,a,g,k)\ (L(x_1,x_2,p,q,a,g,k)\Pi\ L(x_3,x_4,x_2,x_2,a,g,k)$$

$$.\equiv.\ L(x_1,x_2,p,q,a,g,k)\Pi L(x_3,x_4,p,q,a,g,k)) \tag{59}$$

$$(\forall x_1,x_2,x_3,p,q,a,g,k)(L(x_1,x_2,p,q,a,g,k)\Pi L(x_1,x_3,p,q,a,g,k) .\equiv.L(x_1,\{x_2,x_3\},p,q,a,g,k)) \tag{60}$$

For another example, S14 and S20 below can be proved to be paraphrases each other by employing '***Postulate of Reversibility of a Spatial Event (PRS)***'.

(S20) The road separates at *C* from the street with the sidewalk and, after a while, runs 10km straight west from *B* to *A*.

The postulate PRS can be formulated as (61) using '$\equiv_0$', where χ and χ^R is a perceptual locus and its 'reversal' for a certain spatial event, respectively. These loci are substitutable with each other due to the property of '$\equiv_0$'.

$$\chi^R.\equiv_0.\chi \tag{61}$$

The recursive operations to transform χ into χ^R are defined by (62), where the reversed values p^R and q^R depend on the properties of the attribute concerned. For example, at (34), $p^R=p$, $q^R=q$ for A_{12}; $p^R=-p$, $q^R=-q$ for A_{13}.

$$(\chi_1\bullet\chi_2)^R\Leftrightarrow \chi_2^R\bullet\chi_1^R$$

$$(\chi_1\Pi\chi_2)^R\Leftrightarrow \chi_1^R\,\Pi\chi_2^R$$

$$(L(x,y,p,q,a,G_s,k))^R \Leftrightarrow L(x,y,q^R,p^R,a,G_s,k) \quad (62)$$

According to (62), (34) is transformed into (63) as its reversal and equivalent to the semantic interpretation of S20.

$$(\exists x,y,z,p,q)(L(_,x,C,p,A_{12},G_s,_)\Pi L(_,y,C,q,A_{12},G_s,_)\Pi L(_,z,y,y,A_{12},G_s,_))\bullet\varepsilon_s\bullet$$

$$(L(_,x,B,A,A_{12},G_s,_)\Pi L(_,x,0,10km,A_{17},G_s,_)\Pi L(_,x,Point,Line,A_{15},G_s,_)$$

$$\Pi L(_,x,West,West,A_{13},G_s,_)) \wedge road(x)\wedge street(y)\wedge sidewalk(z)\wedge p\neq q \quad (63)$$

3.5 Natural language understanding

Natural language, as already mentioned, is the most important for intuitive human-robot interaction. In order for comprehensible communication with humans, robots must understand natural language *semantically* and *pragmatically* well enough. As shown in Fig. 12, semantic understanding means associating symbols to conceptual images of matters (i.e., objects or events), and pragmatic understanding means anchoring symbols to real matters by unifying conceptual images with perceptual images. Robot manipulation by verbal suggestion here is defined as human-robot interaction where a human gives a robot a verbal expression of his/her intention and the robot behaviouralizes its conception, namely, semantic and pragmatic understanding of the suggestion, and maybe that repeatedly. As detailed above, semantic understanding is purely symbol manipulation for translation from verbal expression (text or speech) into $\boldsymbol{L_{md}}$ expression and fundamental semantic computations on $\boldsymbol{L_{md}}$ expressions by employing word meanings. On the other hand, pragmatic understanding is rather complicated because unstructured data processing for sensor-actor coordination is inevitably involved as well as pure symbol manipulation.

As also mentioned above, an event expressed in $\boldsymbol{L_{md}}$ is compared to a movie film recorded through a floating camera because it is necessarily grounded in FAO's movement over the event. ***This implies that L_{md} expression can suggest a robot what and how should be attended to in its environment.*** For example, consider such a suggestion as S21 presented to a robot by a human. In this case, unless the robot is aware of the existence of a certain box between the stool and the desk, such semantic understanding of the underlined part as (64) and such a semantic definition of the word 'box' as (65) are very helpful for it. The attributes A_{12} (Location), A_{13} (Direction), A_{32} (Color), A_{11} (Shape) and the spatial event on A_{12} in these $\boldsymbol{L_{md}}$ expressions indicate that ***the robot has only to activate its vision system in order to search for the box from the stool to the desk*** during the pragmatic understanding. That is, the robot can control its attention mechanism in the top-down way indicated in $\boldsymbol{L_{md}}$ to deploy its sensors or actuators during pragmatic understanding.

(S21) Avoid <u>the green box between the stool and the desk</u>.

$$(\exists x_1,x_2,x_3,x_4,p)(\underline{L(_,x_4,x_1,x_2,A_{12},G_s,_)\bullet L((_,x_4,x_2,x_3,A_{12},G_s,_))\Pi L(_,x_4,p,p,A_{13},G_s,_)}\Pi$$

$$L(_,x_2,Green,Green,A_{32},G_t,_)\wedge stool(x_1)\wedge box(x_2)\wedge desk(x_3)\wedge ISR(x_4) \quad (64)$$

$$(\lambda x)\mathrm{box}(x)\leftrightarrow(\lambda x)\mathrm{L}(_,x,\mathrm{Hexahedron},\mathrm{Hexahedron},\mathrm{A}_{11},\mathrm{G}_{t},_)\wedge\mathrm{container}(x) \quad (65)$$

Conventionally, such quasi-natural language expressions as 'move(10meters)' and so on, uniquely related to computer programs, were employed for deploying sensors/motors in robotic systems as their semantics (e.g., Coradeschi & Saffiotti, 2003; Drumwright et al, 2006). These kinds of expression, however, were very specific to devices and apt to have miscellaneous syntactic variants among them such as 'move(10meters, quickly)', 'move(quickly, 10meters, leftward)', etc. for motors and 'find(object, red)', 'find(object, round, red)', etc. for sensors. This is also the case for spatial expressions such as 'left(x,y)' and 'left(x,y,z)', reading 'x is (z meters) to the left of y'. This fact is very inconvenient for communications especially between devices unknown to each other and, what is worse, they, unlike $\boldsymbol{L}_{md}$ expressions, are too linguistic or coarse to represent and compute sensory/motory events in such an integrated way as 'common coding approach to perception and action' (Prinz, 1990).

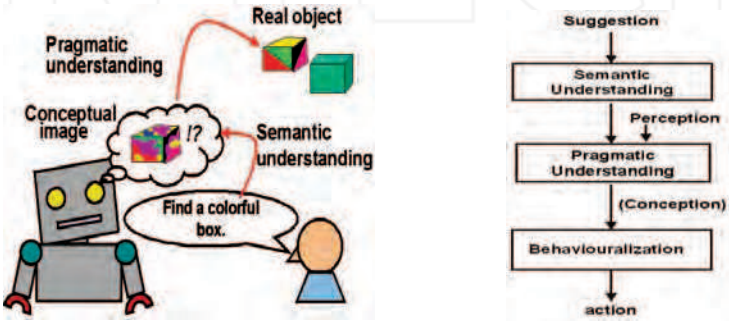


Fig. 12. Semantic and pragmatic understanding of natural language in robots

3.6 Formalization of communication

Here, a piece of communication is identified by a set of messages ($\boldsymbol{M}$) as is formalized in (66). In turn, a message (m) is defined by (67), where D, S, R and B mean the duration, sender(s), receiver(s) and the body of the message, respectively. The body (B) consists of the two elements defined by (68), where E and T mean the event referred to and the task requested or intended by the sender, respectively.

$$\boldsymbol{M}=\{m_1, m_2, \ldots, m_n\} \quad (66)$$

$$m=(D, S, R, B) \quad (67)$$

$$B=(E, T) \quad (68)$$

For example, each item of the message m_0: "Fetch me the book from the shelf, Tom" uttered by Jim during the time-interval $[t_1, t_2]$ is as follows:
$m_0=(D_0, S_0, R_0, B_0)$, $B_0=(E_0, T_0)$, $D_0=[t_1, t_2]$, $S_0=$ *"Jim"*, $R_0=$ *"Tom"*, $E_0=$ *"Tom FETCH Jim BOOK FROM SHELF"*, and $T_0=$ *"realization of* E_0*"*.

In general, without any hint by language, it is not so easy for the receiver (R) to

understand what task (*T*) is intended by the sender (*S*). Therefore, as mentioned in Section 1, IMAGES-M is designed to work as a language-centered (or language-driven) intelligent interface for human-robot multimedia communication. The author has adopted a simple assumuption that there are unique correspondences between the kinds of tasks and the types of sentences as shown in Table 4 while there were proposed more complicated and context-sensitive ones based on speech-act theory (Austin, 1962; Searle,1969).

Sentence type (Examples)	Task (*T*)
Declarative (It is ten o'clock now.)	To believe *E*.
Interrogative ([A] Is it ten o'clock now? [B] What time is it now?)	[A] To reply whether *E* is true or false. [B] To reply what makes *E* true.
Imperative (Show me your watch.)	To realize *E*.

Table 4. Sentence types and Tasks

4. Kansei processing as inter-agent communication

It is well known that emotion in a human can be affected by his/her world, namely, ***W*** in Fig.1. For example, a person's *evaluation* of live image of an object (i.e. image output from **St**) expressed by such words as 'favorite', 'beautiful', 'tasty', etc. can vary depending on his/her emotional bias such as 'hungry', 'depressed', etc. Kansei is one of mental functions with emotion involved but has a more complicated phase than pure emotion originated from instincts or imprinting. For example, sweet jam may be nice on toast but not on pizza for certain people knowledgeable about these foods. For another example, people can be affected on their evaluation of an art by its creator's name, for example, 'Picasso'. These are good examples of Kansei processing as emotional performance affected by knowledge in humans.

Therefore, Kansei can be defined as human emotion toward an object affected by its information for him/her, so called, 'concept', including his/her intellectual pursuits, traditions, cultures, etc. concerning it. In this sense, Kansei is assumed to be reasonable or intelligible among the people sharing such concepts unlike pure emotion. These hypothetical considerations are formalized as (69) and (70).

$$I_P(x) = \mathbf{P_E}(S(x)) \tag{69}$$

$$I_K(x) = \mathbf{P_E}(S(x) \wedge O(x)) = \mathbf{P_E}(S'(x)) \tag{70}$$

where
$\mathbf{P_E}(X)$: Performance of **Em** for mental image X;
$I_P(x)$: Mental image as pure emotion for object x;
$I_K(x)$: Mental image as the result of Kansei processing for object x;
S(x): Live image of object x sent from **St**;
O(x): Concept (i.e., conceptual image) of object x induced by S(x) from knowledge at **Kn**;
S′(x): Unified image of live image and concept of object x.

Figure 13 shows an example of Kansei processing in the mind model, where perceived, induced and inspired images correspond to S(x), S′(x) and $I_K(x)$, respectively, while Fig.14 is for pure emotion with $I_P(x)$ as the inspired image.

These two inspired images $I_K(x)$ and $I_P(x)$ are can be verbalized in **Re** as 'Appetizing!' and 'Fragrant!', as labeled in the figures, respectively. The essential difference between them is assumed to reside in whether or not they are affected by O(x), namely, the concept of 'chocolate cream bread', inferred by **Kn** from the shape, smell, etc. Whereas, pure emotion for an object can be a special case of Kansei processing without knowing or recognizing what it is.

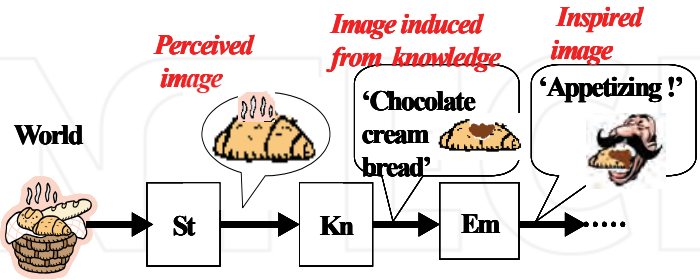


Fig. 13. Example of Kansei processing

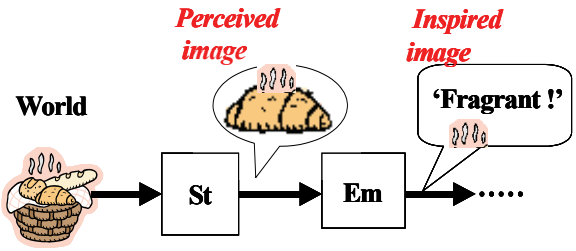


Fig. 14. Example of pure emotion

In MIDST, as already mentioned, the concept of an object *x* is given as an integrated omnisensory mental image of its properties and its relations with other objects involved. For example, the concept of 'chocolate cream bread' can be given by (71), reading that *x* is bread, sweet due to chocolate cream, fragrant of itself, etc., where A_{29} and A_{30} refer to 'Taste' and 'Odour', respectively.

$$(\lambda x)\text{chocolate_cream_bread}(x) \Leftrightarrow (\lambda x)(\exists y,k_1,k_2)L(y,x,\text{Sweet},\text{Sweet},A_{29},G_t,k_1)\Pi$$

$$L(x,x,\text{Fragrant},\text{Fragrant},A_{30},G_t,k_2) \wedge \text{bread}(x) \wedge \text{chocolate_cream}(y) \wedge \ldots \quad (71)$$

Many psychologists have claimed that certain emotions are more basic than others (Ortony & Turner, 1990). The author has assumed that human emotion consists of four primitives such that represent the degrees of 1) Happiness, 2) Aggressiveness, 3) Surprise and 4) Superiority. For example, the degree of Happiness is measured by using such a word set as {anguish, distress, sorrow, gloom, content, joy, ecstasy}, whose each element is possibly ordered on a coordinate axis and fuzzified with a certain characteristic function.

Therefore, the author has assumed Kansei as a certain function to evaluate totally the loci in the attribute spaces of these primitives.

For comprehensible communication with humans, robots must understand natural language *semantically* and *pragmatically*. However, humans and robots can be equipped with sensors, actuators and brains of different competences and performances. Therefore, their vocabularies may well be grounded in quite different sensations, physical actions or mental actions and in turn such a situation may bring inevitably different kinds of semantics to them, so called, "Natural Semantics (NS)" for humans and "Artificial Semantics (AS)" for robots. For example, consider such a scenario as follows.

…A human 'Kate' and a humanoid robot 'Robbie' encounter at the terrace in front of the room where a Christmas party is going on merrymaking. Kate says "Robbie, please fetch me some delicious food from the noisy room." Robbie replies "OK, Kate."….

For a happy end of this dialog, Robbie must have a good knowledge of Kate's NS for Kansei and translate it into its AS appropriately enough to find out the real objects referred to by her words. In this case, Robbie needs at least to interpret Kate's statement as the expression (72) reading "If *Robbie* fetches *Kate* some food delicious ***for her*** from the room noisy ***for her*** (***E₁***), then consecutively it makes *Kate* happier (***E₂***)."

$$\mathbf{E}_1 \supset_1 \mathbf{E}_2, \tag{72}$$

$$\text{where } \mathbf{E}_1 = (\exists x_1,x_2,k_1,k_2,k_3,k_4)\ (L(\text{Robbie},\text{Robbie},\text{Kate},x_2,A_{12},G_t,k_1)\bullet$$

$$(L(\text{Robbie},\text{Robbie},x_2,\text{Kate},A_{12},G_t,k_1)\ \Pi\ L(\text{Robbie},x_1,x_2,\text{Kate},A_{12},G_t,k_1)))$$
$$\Pi(L(\text{Kate},x_1,\text{Delicious},\text{Delicious},B_{08},G_t,k_2)\ \Pi\ L(\text{Kate},x_2,\text{Noisy},\text{Noisy},B_{08},G_t,k_4)$$

$$\wedge\ \text{food}(x_1) \wedge \text{room}(x_2)$$

$$\text{and } \mathbf{E}_2 = (\exists v_1,v_2,k_7)\ L(\mathbf{E}_1,\text{Kate},v_1,v_2,B_{04},G_t,k_7) \wedge v_2 > v_1$$

In (72), the special symbols and their meanings in the expressions are: '$X \supset_1 Y$' ='If X, then consecutively Y', B_{08}='Kansei' and B_{04}= 'Happiness'. As easily understood, 'Delicious' and 'Noisy' in Kansei (B_{08}) should be distiguished from 'Sweet', etc. in Taste(A_{29}) and 'Loud', etc. in Sound(A_{31}), respectively because the former result from total evaluation of the latter. Of course, the values in Kansei greatly depend on the standards (i.e., k_2 and k_4) given by the evaluator (i.e., Event Causer), which are most closely related to 'Individual' or 'Purposive' standard in Table 3.

By the way, Robbie's task is only to make ***E₁*** come true where each atomic locus formula is associated with his actuators/sensors. Of course, Robbie *believes* that he will become happier to make Kate happier as is given by (73) where 'B_{03}' is 'trueness (=degree of truth)' and 'K_B' is a certain standard of 'believability' for Robbie. That is *emotionally* to say, Robbie *likes* Kate as formulated as (74) reading 'Robbie is happy because of Kate' where 'K_H' is a certain standard of 'happy' for Robbie. And then, Robbie comes to believe that he will realize ***E₁*** in future (i.e., *decides* to make ***E₁*** take place) as is expressed by (75). Therefore, this example is also very significant for *intentional* sensing or action of a robot driven by logical description of its belief.

$$(\exists p)L(\text{Robbie},\mathbf{E},p,p,B_{03},G_t,K_B) \wedge p > K_B, \tag{73}$$

$$\text{where } \mathbf{E} = (\mathbf{E}_2 \supset_1 \mathbf{E}_3)$$

$$\text{and } \mathbf{E}_3 = (\exists v_3, v_4, k_8)\ L(\mathbf{E}_2, \text{Robbie}, v_3, v_4, B_{04}, G_t, k_8) \wedge v_4 > v_3$$

$$(\exists p)\ L(\text{Kate}, \text{Robbie}, p, p, B_{04}, G_t, K_H) \wedge p > K_H \tag{74}$$

$$(\exists p, q, t_1, t_2, k)\ L(\text{Robbie}, \mathbf{E}_4, /, p, B_{03}, G_t, K_B) \Pi \varepsilon([t_1, t_2]) \wedge p > K_B, \tag{75}$$

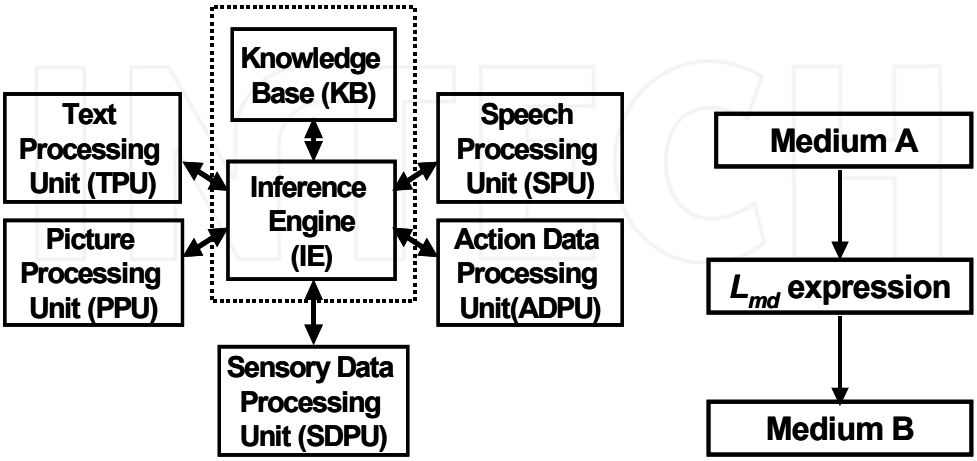
$$\text{where } \mathbf{E}_4 = \varepsilon([t_1, t_2]) \bullet L(\text{Robbie}, \mathbf{E}_1, /, q, A_{01}, G_t, k)$$

For realizing a plausible *Artificial Kansei*, it is most essential to find out functional features of **Em** and to deduce from them such laws that may rule $\mathbf{P_E}$. The author has already obtained a number of postulates concerning $\mathbf{P_E}$ such as Postulate 01-04 below and simulated human-robot dialog based on them (See Fig.17).

Postulate 01: "Some emotion can coexist with another in humans."
Postulate 02: "A desire for something is a belief of becoming happier with it."
Postulate 03: "Something delicious for someone makes him/her happier by its taste."
Postulate 04: "Something noisy for someone makes him/her less happy by its sounds."

5. IMAGES-M as an artificial communication partner

The methodology mentioned above has been being implemented on the intelligent system IMAGES-M shown in Fig.15-a. IMAGES-M is one kind of expert system consisting of Inference Engine (IE), Knowledge Base (KB) and five kinds of user interface: 1) Text Processing Unit (TPU), 2) Speech Processing Unit (SPU), 3) Picture Processing Unit (PPU), 4) Action Data Processing Unit (ADPU) and 5) Sensory Data Processing Unit (SDPU). The pair of IE and KB work as **Kn** and **Em** of the mind model, and the group of user interfaces as **St** and **Re**. As depicted in Fig.15-b, these user interfaces are intended to convert information media and L_{md} expressions mutually in collaboration with IE and KB in order to facilitate various types of cross-media operations such as language-to-picture translation.



(a) (b)

Fig. 15. (a) Configuration of IMAGES-M and (b) cross-media operation via L_{md} expression

H: How does the national road run?
S: It extends between Pref. A and Pref. C via Pref. B.
H: Where does the bus go from the rail way station A?
S: It reaches the town D.
H: What is between the buildings A and B?
S: The railway D.
H: Where do the street A and the road B meet?
S: At the crossing C.
H: Where do the street A and the road B separate?
S: At the crossing C.

Fig. 16. Cross-media operations: A map reproduced from L_{md} expression and Q-A on the map ('H' and 'S' stand for 'Human' and 'IMAGES-M', respectively)

[Dialog 01] ***Postulate 01: "Some emotion can coexist with another in humans."***
H: Tom loves Jane.
H: Whom does he like?
S: He likes Jane.
[Dialog 02] ***Postulate 02: "A desire for something is a belief of becoming happier with it."***
H: Tom wants to go to Tokyo.
H: Does Tom believe to become happier if he goes to Tokyo?
S: Yes, he does.
[Dialog 03] ***Postulate 03: "Something delicious for someone makes him/her happier by its taste."***
H: The food was delicious for Tom.
H: Did it make him happier by its taste?
S: Yes, it did.
[Dialog 04] ***Postulate 04: "Something noisy for someone makes him/her less happy by its sounds."***
H: Mary evaluated the music noisy.
H: Did it make her less happy by its sounds?
S: Yes, it did.

Fig. 17. Examples of postulate-based dialog simulations ('H': human, 'S': IMAGES-M)

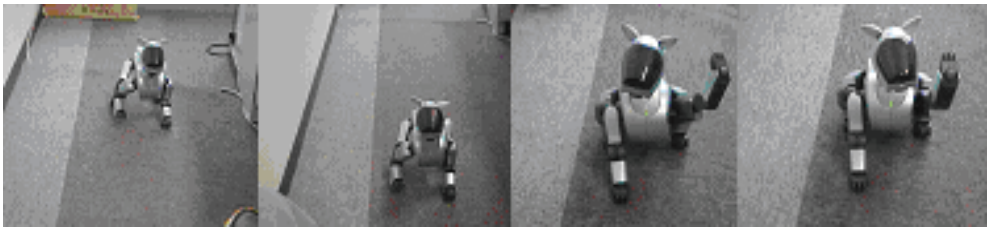


Fig. 18. AIBO (Sony) behaving to the command 'Walk forward and wave your left hand'

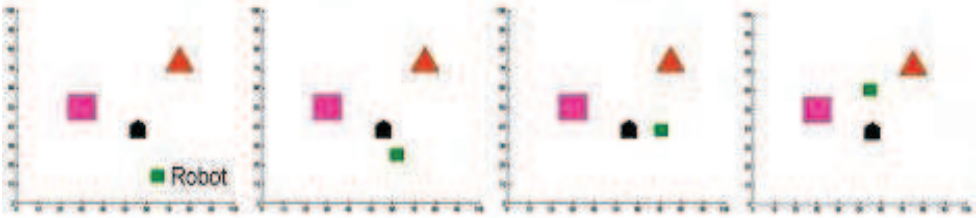


Fig. 19. Simulation in Matlab of the command 'Go to between the rectangle and the triangle, avoiding the pentagon'

Figure 16 presents a map reproduced from $\boldsymbol{L}_{md}$ expression and several transactions during Q-A on it, where IMAGES-M computed temporal and spatial events distinctively very well. In Fig.17 are shown several examples of postulate-based dialog between a human and IMAGES-M performing symbol manipulation, namely, deduction based on the postulates.

Figure 18 shows an example of langage-to-action translation by IMAGES-M. The real robot, AIBO (Sony), cannot perform the two actions (i.e., 'walking' and 'hand-waving') simultaneously and therefore it waved its hand after walking with the transit action (i.e., 'sitting-down') between them inserted, where each action was defined as an ordered set of values in 'Shape (A_{11})' (i.e., time-sequenced snapshots of the action) corresponding uniquely with the positions of their actuators determined by the rotations (i.e., changes in 'Angularity (A_{45})') of the joints (Yokota, 2006). In Fig.19 is presented another example of language-to-action translation for a virtual robot simulated in Matlab, where 'avoid' was defined as 'keep Topology (A_{44}) Disjoint'.

IMAGES-M has so far worked well enough as a language-centered intelligent system. To our best knowledge, there is no other system that can perform miscellaneous cross-media operations, including robot manipulation by natural language, in such a seamless way as ours. Most of computations on $\boldsymbol{L}_{md}$ are simply for unifying (or identifying) atomic loci and for evaluating arithmetic expressions such as 'p=q', and therefore the author believes that the methodology presented here can reduce the computational complexities of conventional ones when applied to the same kinds of problem described here. Recently, $\boldsymbol{L}_{md}$ has been enriched by a general theory of tempological connectives (Yokota, 2008), and another algorithm programmable in procedural languages has been developed in order for more efficient computation of $\boldsymbol{L}_{md}$ expressions (Yokota et al, 2008) while the earlier version of IMAGES-M was programmed in PROLOG with considerable redundancy. This advantage comes from the meaning representation scheme normalized by atomic locus formulas,

which will also facilitate higher-order representation and computation such as described in Section 4 (although it is well known that expressions in higher-order predicate logic without any constraint like this are not computable).

6. Conlusion

The author presented an innovative concept of artificial communication partners with the multiagent mind model based on MIDST and its implementation on IMAGES-M. The competence and performance of the mind model are formalized as a deductive system in the formal language $\boldsymbol{L_{md}}$ employed for many-sorted predicate logic. This deductive system is one kind of appplied predicate logic (i.e., pure predicate logic with certain domain-specific constants (e.g., McDermott, 1982; Shoham, 1989; Kabanza, 1995; Haddawy, 1996)), but the domain-specificness in its syntax and semantics is exclusively related to atomic locus formulas and the essential part of its semantics is subject to their interpretation controled by the family of domain-specific constants, namely, Attributes, Values, Event Types and Standards intended to correspond well with human sensory systems. The author has found the implementation so far a success and come to have such a perspective that the scheme presented here is applicable to various mind models for humans or humanoid robots of different competences and performances simply by controling such a family.

The expressive power of $\boldsymbol{L_{md}}$ was demonstrated with linguistic or pictorial manifestations throughout this chapter. Its most remarkable point in comparison with other KRLs resides in that it can provide terms of physical matters such as *carry*, *snow*, etc. with precise semantic definitions that are normalized by atomic locus formulas and visualized as loci in attribute spaces in both temporal and spatial extents (i.e., temporal and spatial events), which leads to good computability and intuitive readability of $\boldsymbol{L_{md}}$ expressions. As easily imagined by the brief description in Section 4, this is also the case for terms of human mental matters such as *believe*, *like*, *desire*, etc. Among them, the semantic definitions of *belif*, *desire* and *intention* have been studied with special attention focused on inter-agent communication and computationally grounded in Belief-Desire-Intention logic (BDI logic) in conventional multiagent systems (MASs) (e.g., Vieira et al, 2007). These MASs, however, as well as other conventional deductive systems will have to employ another framework for dynamic (i.e., temporally or spatially continuous) processes in the attributes (or propeties) of matters such as "They *come to believe* it *gradually*", namely, *temporal event* of belief that can be easily formulated in $\boldsymbol{L_{md}}$, needless to mention about *spatial event* that has seldom been considered for systematic formulation and computation by others.

The mind model proposed here is much simpler than Minsky's but it can work well enough to compute both physical and mental matters expressed in $\boldsymbol{L_{md}}$. The most important problems remaining unsolved are how to provide each attribute space and how to build its corresponding atomic performance. These problems concern neuroscience (Brooks, 1986) as well as psychology (Yokota, 1988; Fukuda et al, 1998; Sugita et al, 2003) and therefore the author will consider employment of soft computing theories such as neural network, genetic algorithm, fuzzy logic, etc. for their self-organization in the near future.

At last of this chapter, the author would like to acknowledge that this work was partially funded by the Grants from Computer Science Laboratory, Fukuoka Institute of Technology and Ministry of Education, Culture, Sports, Science and Technology, Japanese Government, numbered 14580436 and 17500132.

7. References

Allen, J.F. (1984). Towards a general theory of action and time, *Artificial Intelligence*, Vol.23, No.2, pp.123-154

Amano,M., et al. (2005). Cross-media translation between human motions and texts based on Mental Image Directed Semantic Theory, *Proceedings of IEEE workshop on Multimedia Network Systems and Applications (ICDCS05)*, Columbus OH, pp.707-713

Austin, J. L. (1962). *How to Do Things with Words*, Oxford University Press, London

Brooks, R. A. (1986). A robust layered control system for a mobile robot, *IEEE Journal of Robotics and Automation*, Vol.RA-2, pp.14-23

Butz,A., Fisher,B.D., Krüger,A. & Olivier,P. (Eds.) (2005). *Smart Graphics, Proceedings of 5th International Symposium (SG 2005)*, Frauenwörth Cloister, Germany, August 22-24, 2005, Lecture Notes in Computer Science 3638 Springer

Coradeschi,S. & Saffiotti,A. (2003). An introduction to the anchoring problem, *Robotics and Autonomous Systems*, Vol.43, pp.85–96

Demiris,Y. & Khadhouri,B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, Vol.54, pp.361–369

Dorr,B.& Bonnie,J. (1997). Large-Scale Dictionary Construction for Foreign Language Tutoring and Interlingual Machine Translation, *Machine Translation*, Vol.12, No.4, pp.271-322

Drumwright,E. Ng-Thow-Hing,V. & Mataric,M.J. (2006). Toward a vocabulary of primitive task programs for humanoid robots, *Proceedings of International Conference on Development and Learning (ICDL06),* Bloomington IN, May 2006

Eakins, J.P. & Graham, M.E. (1999). Content-based Image Retrieval, *A report to the JISC Technology Applications Programme*, Institute for Image Data Research, University of Northumbria at Newcastle

Egenhofer,M. (1991). Point-set topological spatial relations, *Geographical Information Systems*, Vol.5, No.2, pp.161-174

Fukuda,M., Sugita,K.and Shibata, Y. (1998). Perceptional Retrieving Method for Distributed Design Image Database System, *Journal of Information Processing Society of Japan*, Vol.39, No.2, pp.158-169

Gardenfors,P. (2000). *Conceptual Space*, MIT Press

Haddawy,P. (1996). A logic of time, chance, and action for repesenting plans, *Artificial Intelligence*, Vol.80, No.2, pp.243-308

Kabanza,F.(1995). Synchronizing multiagent plans using temporal logic specifications, *Proceedings of 1st International Conf. on Multi-Agent Systems (ICMAS-95)*, pp.217-224, June 1995.

Kherfi,M.L., Ziou,D. & Bernardi,A. (2004). Image Retrieval from the World Wide Web : Issues, Techniques and Systems, *ACM Computer Surveys*, Vol.36, No.14, pp.35-67

Labrou,Y., Finin,T. & Peng,Y. (1999). The current landscape of agent communication languages, *Intelligent Systems*, 14 (2), pp.45-52

Langacker,R. (1991). *Concept, Image and Symbol*, Mouton de Gruyter, Berlin/New York

Langacker,R. (2005). Dynamicity, fictivity, and scanning: The imaginative basis of logic and linguistic meaning, In: *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking*, Precher, D. & Rudolf,A.Z. (eds.), pp.164-197, Cambridge University Press, Cambridge

Leisi,E. (1961). *Der Waltinhalt-Seine Struktur im Deutschen und Englischen-*, Quelle & Meyer, Heidelberg

McDermott,D.V. (1982). A temporal logic for reasoning about processes and plans, *Cognitive Science*, Vol.6, pp.101-155

Miller,G.A.& Johnson-Laird,P.N. (1976). *Language and Perception*, Harvard University Press

Minsky, M. (1986). *The society of mind*, Simon and Schuster, New York

Noton,D. (1970). A Theory of Visual Pattern Perception, *IEEE Transaction on Systems Science and Cybernetcs*, Vol.SSC-6, No.4, pp.349-357

Oda,S., Oda,M.& Yokota,M. (2001). Conceptual Analysis Description of Words for Color and Lightness for Grounding them on Sensory Data, *Trans. of JSAI*, Vol.16, No.5-E, pp.436-444

Ortony,A. and Turner,T.J. (1990). What's basic about basic emotions?, *Psychological Review*, Vol.97, pp.315-331

Prinz, W. (1990).: A common coding approach to perception and action, In: *Relationships between perception and action*, Neumann, O. & Prinz, W. (Eds.), pp.167-201, Springer-Verlag

Roget,P. (1975). *Thesaurus of English Words and Phrases*, J.M.Dent & Sons Ltd., London

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge

Shiraishi,M, Capi,G.& Yokota,M. (2006). Human-robot communication based on a mind model, *Artificial Life and Robotics*, Vol.10, No.2, pp.136-140

Shoham,Y. (1989). Time for actions: on the relationship between time, knowledge, and action, *Proceedings of IJCAI89*, Detroit MI, pp.954-959

Sowa,J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks Cole Publishing Co., Pacific Grove, CA

Sugita,K., Miyakawa, A. and Shibata, Y. (2003). Relation Between Kansei Words and the Room Space in Digital Traditional Japanese Crafting System, *Proceedings of International Conference on Advanced Information Networking and Applications (AINA03)*, pp.159-162, May 2003

Tauchi,E., Sugita,K., Capi,G. & Yokota,M. (2006). Towards Artificial Communication Partners with 'Kansei', *Proceedings of IEEE workshop on Network-based Virtual Reality and Tele-existence (INVITE06)*, Wien Austria, May 2006

Vieira,R., Moreira,A., Wooldridge,M. & Bordini,R.H. (2007). On the Formal Semantics of Speech-Act Based Communication in an Agent-Oriented Programming Language, *Journal of Artificial Intelligence Research*, Vol.29, pp.221–267

Wilkins, D.E. & Myers K.L. (1995). A common knowledge representation for plan generation and reactive execution, *Journal of Logic and Computation*, Vol.5, No.6, pp.731-761

Yokota,M., Taniguchi,R. & Kawaguchi,E. (1984). Language-Picture Question-Answering through Common Semantic Representation and Its Application to the World of Weather Report, In: *Natural Language Communication with Pictorial Information Systems*, Bolc,L.(Ed.), pp.203-254, Springer-Verlag

Yokota,M. (1988). A psychological experiment on human understanding process of natural language. *Transaction of IEICE Japan*, Vol.J71D, No.10 (in Japanese) pp.2120-2127

Yokota,M. (2005). An Approach to Natural Language Understanding Based on Mental Image Model. In: *Natural Language Understanding and Cognitive Science*, Sharp,B. (ed.), pp.22-31, INSTICC PRESS

Yokota,M.& Capi,G. (2005a). Cross-media Operations between Text and Picture Based on Mental Image Directed Semantic Theory, *WSEAS Transaction on Information Science and Applications*, Vol.10, No.2, pp.1541-1550

Yokota,M.& Capi,G. (2005b). Integrated Multimedia Understanding for Ubiquitous Intelligence Based on Mental Image Directed Semantic Theory, *Proceedings of IFIP EUC`2005 symposium (LNCS 3823)*, pp.538-546, Nagasaki Japan, Dec., 2005

Yokota,M. (2006). Towards a Universal Language for Distributed Intelligent Robot Networking, *Proceedings of International Conference on Systems, Man and Cybernetics (SMC06),* Taipei Taiwan, Oct. 2006

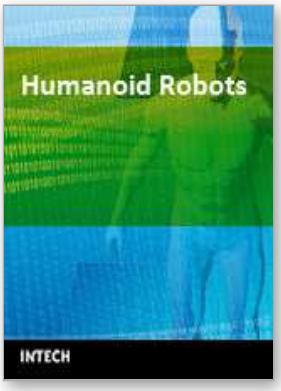
Yokota,M. (2007). A Theoretical Consideration on Artificial Imitation of Human Action Based on Mental Image Directed Semantic Theory, *Proceedings of IEEE workshop on Network-based Virtual Reality and Tele-existence (INVITE07)*, Niagara Canada, May 2007

Yokota,M., Shiraishi,M., Sugita,K. & Oka,T. (2008). Toward integrated multimedia understanding for intuitive human-system interaction, *Artificial Life and Robotics*, Vol.12, No.1-2, pp.188-193

Yokota,M., Abe,Y., Sugita,K. & Oka,T. (2008). Multimedia Description Language $\boldsymbol{L_{md}}$ and Its Application to Integrated Multimedia Understanding, *Proceedings of 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2008)*, Nagoya Japan, Sep. 2008

Yokota,M. (2008). A General Theory of Tempological Connectives and Its Application to Spatiotemporal Reasoning, *Proceedings of 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on Advanced Intelligent Systems (SCIS & ISIS 2008)*, Nagoya Japan, Sep. 2008

Zarri,G.P. (1997). NKRL, a Knowledge Representation Tool for Encoding the Meaning of Complex Narrative Texts, Natural Language Engineering, *Special Issue on Knowledge Representation for Natural Language Processing in Implemented Systems*, Vol.3, pp.231-253



Humanoid Robots

Edited by Ben Choi

ISBN 978-953-7619-44-2

Hard cover, 388 pages

Publisher InTech

Published online 01, January, 2009

Published in print edition January, 2009

Humanoid robots are developed to use the infrastructures designed for humans, to ease the interactions with humans, and to help the integrations into human societies. The developments of humanoid robots proceed from building individual robots to establishing societies of robots working alongside with humans. This book addresses the problems of constructing a humanoid body and mind from generating walk patterns and balance maintenance to encoding and specifying humanoid motions and the control of eye and head movements for focusing attention on moving objects. It provides methods for learning motor skills and for language acquisition and describes how to generate facial movements for expressing various emotions and provides methods for decision making and planning. This book discusses the leading researches and challenges in building humanoid robots in order to prepare for the near future when human societies will be advanced by using humanoid robots.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Masao Yokota (2009). Towards Artificial Communication Partners with a Multiagent Mind Model Based on Mental Image Directed Semantic Theory, Humanoid Robots, Ben Choi (Ed.), ISBN: 978-953-7619-44-2, InTech, Available from: http://www.intechopen.com/books/humanoid_robots/towards_artificial_communication_partners_with_a_multiagent_mind_model_based_on_mental_image_directe

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821