The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

Toward multimodal user interface for intuitive interaction with Buddhism statue ordering system

Tzu-Hsuan Huang Kaoru Sugita Masao Yokota

Fukuoka Institute of Technology
(Tel: 81-92-606-5897, Fax: 81-92-606-8923)

yokota@fit.ac.jp

Abstract: The authors have been developing a Buddhism statue ordering system which enables customers to use affective words such as peaceful and divine, intending to specify the facial expressions of their favorite statues. Such affective words, however, are bound to vague and ambiguous specification of their demands. This paper proposes an innovative two-staged specification method allowing customers to elaborate their demands by graphical inputs following *Kansei* words.

Keywords: artificial *Kansei*, multimodal interface, product customization

1 INTRODUCTION

For ordinary people, natural language is the most important among the various communication media because it can convey the exact intention of the sender to the receiver due to its syntax and semantics common to its users [1]. This is not necessarily the case for another medium such as gesture and therefore natural language can as well play the most crucial role in intuitive human-robot interaction [1]. Natural language, without any doubt, is an extremely powerful means for people to express their ideas of various kinds in abstract, i.e., conceptual level. However, for example, it is not so powerful to describe natural scenery in concrete as picture. This is also the case for certain customer servicing systems allowing users to specify their demands for the favorite products in natural language. Recently, the authors have been developing a Buddhism statue ordering system which enables customers to use affective words such as *peaceful* and *divine*, intending to specify the facial expressions of their favorite statues. Such affective words, however, are bound to vague and ambiguous specification of their demands. This paper proposes an innovative two-staged specification method allowing customers to elaborate their demands by graphical inputs following affective words, so-called *Kansei* words.

The remaining sections of this paper are as follows. Firstly, the multimodal interface of the Buddhism statue ordering system is overviewed in Section 2. Secondly, Section 3 describes the methodology of semantic definition of *Kansei* words as membership functions. Thirdly, Section 4 and 5 present linguistic and graphical expression processing, and lastly, Section 6 concludes this paper.

2 MULTIMODAL DEMAND SPECIFICATION

The Buddhism statue ordering system under development is expected to customize prototypes according to the user's demands specified both linguistically and graphically as shown in Fig.1. The interactions between the user and the system are largely such as follows.

(Step 1) User asks the system for the favorite statue in natural language such as 'Extremely peaceful'.
(Step 2) System retrieves candidate prototypes such as Fig.2 (a)-(c) from Raw Image Database to display.
(Step 3) System retrieves line-drawn faces of the candidates such as Fig.3 (a)-(c), respectively, from Sketch Database to display through a certain line-drawing tool.
(Step 4) User retouches the sketches through the drawing interface.
(Step 5) System reedits and customizes the prototype image data according to the graphical demand specification.
(Step 6) User continues the interaction by returning to Step 1 or Step 4. Otherwise, User terminates the interaction.

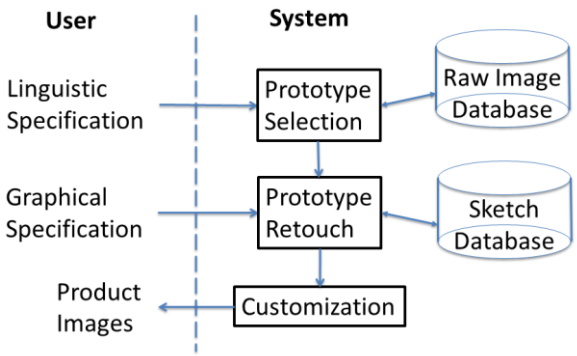
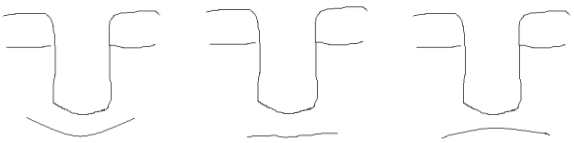


Fig.1. Multimodal intuitive interaction in Buddhism statue ordering system

(a) DN1R (b) DN2R (c) DN3R

Fig.2. Examples of raw images of statues



(a) DN1S (b) DN2S (c) DN3S

Fig.3. Face sketches of the statues in Fig.2

3 SEMANTIC DEFINITION OF KANSEI WORDS AS MEMBERSHIP FUNCTIONS

The meaning of a *Kansei* word is characterized by a set of membership functions defined respectively at 5 dimensions of the Attribute space of *Kansei*, each of which corresponds with the 5 primitive emotions, namely, 1) Anger, 2) Disgust, 3) Anxiety, 4) Happiness, and 5) Superiority [2]. In this study, each of the membership functions for a *Kansei* word W is given in such a way that it is characterized by a trapezoid, symmetric at the vertical line passing (C, 0) on a primitive emotion axis, with the height 1, the top length T, and the bottom length B as shown in Fig.4. Therefore, the membership function $f_{ki}(x)$ for i-th dimension of a word W_k is defined uniformly as (1), where x denotes some degree of a certain primitive emotion. Hereafter, for the sake of simplicity, each trapezoid is normalized as $T=1$ and $B=4$, which necessarily implies that the meaning of each *Kansei* word is representable simply by the coordinate consisting of Cs for all the primitive emotions. Each C ranges over [-9, +9] on its corresponding axis [2] and is called MFI (Membership Function Identifier) here.

$$
\begin{aligned}
f_{ki}(x) &= 0 && (x < C_{ki} - B_{ki}/2),\\
&= \frac{2}{(B_{ki} - T_{ki})}(x - C_{ki} + B_{ki}/2) + 1 && (C_{ki} - B_{ki}/2 \le x < C_{ki} - T_{ki}/2),\\
&= 1 && (C_{ki} - T_{ki}/2 \le x < C_{ki} + T_{ki}/2),\\
&= \frac{-2}{(B_{ki} - T_{ki})}(x - C_{ki} - T_{ki}/2) + 1 && (C_{ki} + T_{ki}/2 \le x < C_{ki} + B_{ki}/2),\\
&= 0 && (C_{ki} + B_{ki}/2 \le x).
\end{aligned} \tag{1}
$$

According to this definition and its normalization, such primitive *Kansei* word concepts as *fury*, *favor*, and *joy* can be defined as Table 1 (tentative) because they are represented simply in terms of the unit vector for each corresponding primitive emotion. On the other hand, the meanings of such complex *Kansei* words as *divine* and *peaceful* are defined as certain complicated vectors in the *Kansei* space as shown in Table 2 (tentative). Figure 5 shows the membership functions assigned to several *Kansei* words of 'Happiness', namely, the 4th dimension of the Attribute space of *Kansei*.

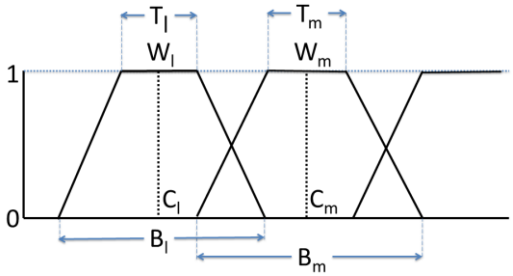


Fig.4. Fuzzification of *Kansei* word meanings

Table 1 Primitive *Kansei* words and MFIs†

Word	1) Ag	2) Dg	3) Ax	4) Hp	5) Sp
Fury	+8.5	0	0	0	0
Anger	+7				
Displeasure	+5				
Irritation	+3				
Upset	+2				
Annoyance	-2				
Scare	-4				
Fear	-6				
Terror	-8.5				
Nausea	0	+8.5	0	0	0
Disgust		+7			
Dislike		+5			
Disfavor		+2			
Favor		-2			
Liking		-5			
Love		-7			
Ardor		-8.5			
Despair	0	0	+8.5	0	0
Discouragement			+6		
Anxiety			+4		
Unease			+2		
Ease			-2		
Anticipation			-4		
Encouragement			-6		
Assurance			-8.5		
Ecstasy	0	0	0	+8.5	0
Joy				+7	
Happiness				+5	
Gladness				+4	
Cheer				+3	
Contentment				+1	
Discontent				-1	
Gloom				-2	
Unhappiness				-5	
Sorrow				-6	
Distress				-7	
Anguish				-8.5	
Awe	0	0	0	0	+8.5
Respect					+7
Admiration					+5
Sympathy					+2
Pity					-2
Disregard					-5
Disrespect					-7
Misery					-8.5

†Ag=Anger, Dg=Disgust, Ax=Anxiety, Hp=Happiness, Sp=Superiority

Table 2 Complex *Kansei* words and MFIs

Word	1) Ag	2) Dg	3) Ax	4) Hp	5) Sp
Peaceful	0	0	0	+1	+1
Divine	-5	0	0	+1	+8
Reverential	-1	-1	+1	-1	+7
Noble	0	0	0	+1	+5
Valiant	+7	+5	+4	-7	-7
Jingoistic	+5	+5	+2	-5	-6

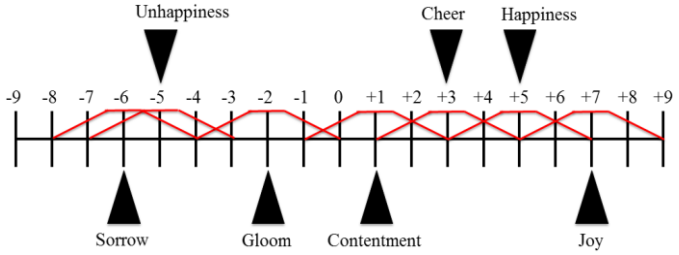


Fig.5 Membership functions of 'Happiness'

4 LINGUISTIC EXPRESSION PROCESSING

The syntax of *Kansei* expression K is defined by (2) as a set of production rules of a context-free grammar. According to this definition, for example, such a considerably complex *Kansei* expression as (4) can be generated as well as rather simple ones such as (3). Moreover, such an extreme expressions as (5) can be as well generated as one of the terminal expressions for D^nW (n≥0), the generalized form of the structure A.

$$\begin{aligned}&\{K \rightarrow A \mid (K) \mid \textit{not } K \mid K \textit{ and } K \mid K \textit{ or } K,\\ &A \rightarrow W \mid DA,\\ &W \rightarrow \textit{peaceful} \mid \textit{noble} \mid \ldots,\\ &D \rightarrow \textit{extremely} \mid \textit{very} \mid \textit{much} \mid \textit{more} \mid \textit{less}\\ &\quad \mid \textit{slightly} \mid \textit{a little} \mid \textit{scarcely} \mid \ldots\}\end{aligned} \tag{2}$$

$$\textit{peaceful, more noble, a little fearful} \tag{3}$$

$$\textit{not very valiant and (peaceful or awful)} \tag{4}$$

$$\textit{very very very extremely peaceful} \tag{5}$$

Simultaneously, the semantics of *Kansei* expression is given in terms of membership functions although, as easily imagined, most of the expressions generated from the above rules are semantically anomalous or of no significance.

Firstly, the semantic interpretation $Sem(W_k)$ of *Kansei* word W_k is formally represented as (6) while it is conventionally given as an ordered set of membership functions as shown in Table X and Y. This formalization reads that the value of the *i-th* primitive emotion at W_k, V_i is equal to $f_{ki}(x)$ *with* C_{ki} and that all the 5 equations are combined with the logical AND denoted by $\wedge$.

$$\begin{aligned}&Sem(W_k) \equiv (V_1=\mu(x,C_{k1}) \wedge V_2=\mu(x,C_{k2}) \wedge \ldots \wedge V_5=\mu(x,C_{k5})),\\ &\textit{where } \mu(x,C_{ki}) \textit{ is } f_{ki}(x) \textit{ with } C_{ki}.\end{aligned} \tag{6}$$

Secondly, such a structure as $D_1D_2\ldots D_nW_k$ (n≥1) is semantically interpreted so that the combination of adverbials $D_1D_2\ldots D_n$ modify (6) by replacing each C_{ki} with (7). Here, $deg(D)$ is a function to assign a certain degree to the adverbial D as shown in Table 3 (tentative) in order to magnify or deduce each primitive emotion identified by C_{ki}.

$$C_{ki} \cdot (deg(D_1) \cdot deg(D_2) \cdot \ldots \cdot deg(D_n)) \tag{7}$$

Table 3 Adverbials and degrees

D	$deg(D)$	D	$deg(D)$
Extremely	10	Scarcely	1/20
Very	3/2	Slightly	1/10
Much	3/2	A little	1/5
More	11/10	Less	10/11

That is, the modifiers $D_1D_2\ldots D_n$ work only to translate the membership functions of *Kansei* word W_k on each corresponding axis. According to this formalization, for example, the meaning of such a *Kansei* expression as '*more noble*' is given as (8) by consulting Table 2 as well.

$$V_1=V_2=V_3=\mu(x,0) \wedge V_4=\mu(x,1.1) \wedge V_5=\mu(x,5.5) \tag{8}$$

Thirdly, the logical expressions '*not K*', 'K_1 *and* K_2', and 'K_1 *or* K_2' are translated into (9), (10), and (11), respectively, '~' and '$\vee$' are the logical NOT and OR, respectively.

$$\sim Sem(K) \tag{9}$$

$$Sem(K_1) \wedge Sem(K_2) \tag{10}$$

$$Sem(K_1) \vee Sem(K_2) \tag{11}$$

Meanwhile, semantic synthesis of membership functions in the logical structures of K is in principle all the same as in the conventional multivalued logic. That is, the equivalences listed in (12) hold for any single or paired membership functions, where *Min* and *Max* are such functions that they take the minimum and the maximum value, respectively, of the two functions at each x.

$$\begin{aligned}&\{\sim(V_i=g(x)) \equiv (V_i=1-g(x)),\\ &(V_i=g(x) \wedge V_i=h(x)) \equiv (V_i=Min(g(x),h(x))),\\ &(V_i=g(x) \vee V_i=h(x)) \equiv (V_i=Max(g(x),h(x)))\}\end{aligned} \tag{12}$$

For example of negation, '*not* W_k', one of the simplest structures, is interpreted as (13) so that '*not*' modifies (6) by replacing each $\mu(x,C_{ki})$ with $1-\mu(x,C_{ki})$, and $\wedge$ with $\vee$.

$$V_1=1-\mu(x,C_{k1}) \vee V_2=1-\mu(x,C_{k2}) \vee \ldots \vee V_5=1-\mu(x,C_{k5}) \tag{13}$$

For stereotypical example of conjunction, the meaning of 'W_k *and* W_j' is synthesized as (14).

$$V_1=Min(f_{k1}(x), f_{j1}(x)) \wedge \ldots \wedge V_5=Min(f_{k5}(x), f_{j5}(x)) \tag{14}$$

For disjunction, exclusively in the case that the two words in 'W_k *or* W_j' are both primitive for the same emotional category like the pair of *happy* and *joyful*, such a semantic synthesis as (15) is possible, where the very same

category is denoted by the suffix *d* and the 4 remainders, by *a*, *b*, *c*, and *d*, respectively.

$$V_s=Max(f_{ks}(x),f_{js}(x))\wedge V_a=V_b=V_c=V_d=\mu(x,0) \quad (15)$$

As easily understood, semantic synthesis in more complicated structures has only to employ the corresponding membership functions synthesized from a certain set of primitive ones, namely, $\{f_{ki}(x)\}$ involved. At conclusion, any *Kansei* expression can be translated into such a disjunctive normal form as (16), where each $g(x)$ denotes a membership function resulted from some conjunctions or negations for each dimension.

$$\vee_{j=1,n}\ (\wedge_{k=1,5}V_k=g_{jk}(x)) \quad (16)$$

Lastly, the candidateship of a statue *S* for a demand *D* is evaluated by the function F_c (Candidateship Function) defined by (17), employing a special function *Hst* to detect the highest peak (i.e., maximum value) of the synthesized function $h(x)$. Each statue image is characterized by a pair of $f(x)$ and MFI as well as a *Kansei* word, and therefore $0\le F_c(S|D)\le 1$. Table 4 (tentative) shows the MFIs of the statues shown in Fig.6.

$$F_c(S|D)=Max(M_1,M_2,\ldots,M_n),$$
$$\text{where}$$
$$M_j=Min(Hst(h_{j1}(x)), Hst(h_{j2}(x)), \ldots, Hst(h_{j5}(x)))$$
$$h_{jk}(x)= Min(f_k(x),g_{jk}(x)),$$
$$Sem(S)\equiv (V_1=f_1(x)\wedge V_2=f_2(x)\wedge\ldots\wedge V_5=f_5(x)),$$
$$Sem(D)\equiv \vee_{j=1,n}\ (\wedge_{k=1,5}V_k=g_{jk}(x)). \quad (17)$$

Table 4 Statue images in Fig.6 and their MFIs

Statue	1) Ag	2) Dg	3) Ax	4) Hp	5) Sp
DT	0	0	0	8	1
DN	0	0	0	-2	2
AS	6	2	5	-3	-1
FM	8	4	6	-5	2

(a) DT (b) DN (c) AS (d) FM

Fig.6. Examples of images of statues

5 GRAPHICAL EXPRESSION PROCESSING

The graphical expressions to be processed in the system here are limited to sketches of the faces of statues. A sketch here is to show only how the main parts (i.e., mouth, nose, eyes, and eyebrows, here) of the face lie, for example, lying gradient, direction, curvature, extent, etc. [3] each of which is called 'Attribute' (of a matter such as face part) in general [1]. The syntax of face sketch *F* is defined by the set of production rules (18).

$$\{F\rightarrow HDCDG,\ H\rightarrow B_lRB_r,\ C\rightarrow E_lPNPE_r,\ G\rightarrow M,$$
$$B_l\rightarrow b_{l1}|b_{l2}|\ldots,\ B_r\rightarrow b_{r1}|b_{r2}|\ldots,\ E_l\rightarrow e_{l1}|e_{l2}|\ldots,$$
$$E_r\rightarrow e_{r1}|e_{r2}|\ldots,\ N\rightarrow n_1|n_2|\ldots,\ M\rightarrow m_1|m_2|\ldots,$$
$$D\rightarrow d_1|d_2|\ldots,\ P\rightarrow p_1|p_2|\ldots\} \quad (18)$$

This definition consists of non-terminal symbols such as follows, respectively representing a set of real sketches of face parts or their placements denoted by terminal symbols (i.e., *b, e, n, m, d,* and *p*).

H, C, G: The upper, middle and lower part of the face, respectively,
B_l, B_r: The eyebrow left and right from viewers, respectively,
E_l, E_r: The eye left and right from viewers, respectively,
N, M: The nose and the mouth, respectively,
D, P: Downward and rightward placement of the following part, respectively.

The users of the system are allowed to rearrange what are represented by the terminal symbols through a certain graphical interface as shown in Fig.1. Figure 7 shows several examples of such rearrangements.

On the other hand, the semantics of face expression can be defined in terms of $f(x)$ and MFI as well as *Kansei* expression while, in general, nonlinguistic expression such as picture cannot bear definite meaning comprehensible for every one [1].

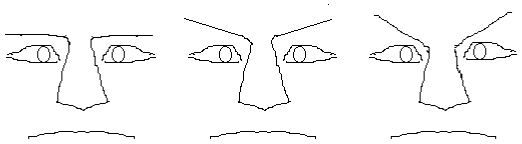


Fig.7. Faces with different eyebrows ($\in B_l$ or B_r)

6 CONCLUSION

Here was presented an innovative specification method based on semantic processing of both linguistic and graphical demands in the Buddhism statue ordering system. Our proposed method interactively allows Users to gradually approximate System's proposals to their real intentions unlike others [e.g., 4] while it is still on the way to practical implementation.

REFERENCES

[1] Yokota, M.: "Subjective Knowledge Representation for Intuitive Human-Robot Interaction Based on Mental Image Directed Semantic Theory", in (Ed. Thomas S. Clary) Knowledge representation, Nova Publishers, 2012.
[2] Huang, T., Sugita, K. and Yokota, M.: "Toward artificial *Kansei* based on Mental Image Directed Semantic Theory", Artificial Life and Robotics: Vol.17-2, pp. 186-190, 2012.
[3] Ekman, P. & Friesen, W.: The facial action coding system: A technique for the measurement of facial movement, Consulting Psychologists Press. 1978.
[4] Sugita, K., Ishida, T., Miyakawa, A. and Shibata, Y.: "*Kansei* Retrieval Method using the Quantitative Feature of Traditional Japanese Crafting Object," Proc. of The 6th International Workshop on Multimedia Network Systems and Applications (MNSA'2004), 112-117, 2004.