# Virtual cognitive model for Miyazawa Kenji based on speech and facial images recognition

HAMIDO FUJITA, JUN HAKURA, AND MASAKI KUREMATSU
Faculty of Software and Information Science,
Iwate Prefectural University
020-0193, Iwate
JAPAN
issam@soft.iwate-pu.ac.jp, http://www.fujita.soft.iwate-pu.ac.jp

*Abstract:* - In this paper we a representing a virtual interactive model based on cognitive model of Miyazawa Kenji. We created a computer model based on cognitive thinking of Kenji literature on story telling. The user can interact in real time with Virtual Kenji. The facial gestures been collected and analyzed through Motion capture system consists of six camera. These six cameras set to collect all emotional facial gestures of people who read and practice an recorded assigned Kenji manuscripts for experiment. Each person has 50 markers of 5 mm size attached to all parts of the face (lips, mouth, eyebrow, moustache, eyelash, forehead). The emotional linkage between these facials parts and cognitive emotion been analyzed and recorded. We have proposed a database; called as Facial recognition database based on FACS model, Also we have correspondingly, speech synthesis part that would analyze the emotional part of human speech. These synthesized two parts are been re-constructed on hologram that represents the cognitively the character of Kenji virtual model who has a face with gestures harmonize with a speech and facial images generated by the system. Also, the system interacts with the human user based on collected observed response on human user and inference by the system in real time.

*Key-Words:* - Cognitive modeling, speech and image recognition, motion capture, Maya software, Motion Builder

## 1 Introduction

Virtual modeling has been active in the past years, thanks to computer graphics and virtual reality systems. This paper contributes to present an experimental work on building a virtual system based on Miyzawa Kenji art work observed on by his writing. Miyazawa Kenji born in Iwate, Japan on 1896 http://www.kenji-world.net/english/who/who.html . He has famous children narrative stories with unique characters. There are rhythms in kenji's stories. Kenji's stories are set against the whole of the universe, a world replete with people, animals, plants, the wind, clouds, light, the stars and the sun. All hold discourse together. All are in empathy with one another. This free association between the elements and living things that make up our world is one of the distinguishing features that predominates Kenji's works. The interaction he portrays is never nonsensical, but always animated with an authenticity that rings true to the reader. Such rhythm represents or reflects certain cognitive behavior inside the stories through which the user may interact and feel the emotional or living part that he/she may interact. http://www.kenji-world.net/english/who/rhythm.html

Most of the work of Kenji had been translated into English. http://www.kenji-world.net/english/translat/translat.html, as well as to other languages. In this paper we are building a virtual work based on Kenji artwork. It is an experimental system based on hologram that interact with the human user based on the cognitive based built-up system.

Based on Kenji scripts we have extracted patterns that reflect the emotional feature behind them. Those features can be classified into and be represented by facial images based on Ekman FACS system (section 2). Those patterns reflect the facial(Sec.2) and sound extracted patterns(Sec.3). Those patterns can be specialized and instantiated the data stored in the database. Those generated graphics animated files reflect the facial cognition and sound data reflects the voice emotional model. The extracted patterns come from examining and analysis of the masterpiece of Kenji work. Such analysis is based on ontological reasoning of the text. Such reasoning reflects the recognition of Kenji work by expertise and people who are aware of the Kenji experience and background. Through such analysis, Kenji words and extracted narrative story patterns reflect the emotional representation of the context in the story. Each word reflects its related domain and its connection(relation) with other words reflect the semantical representation of that context.

There are two parts in the system the namely, computer based graphics representing the animated facial image generated by the system and interacted with the user based on the latter cognitive facial recognition. Moreover,

there is speech synthesis part which make and generate the voice of Kenji in harmony with the generated facial images. Also this will interact with the human user based on collected sound patterns that been reflected with a reasoning process capturing the emotional recognition of human user voice that interacting with Kenji based on the

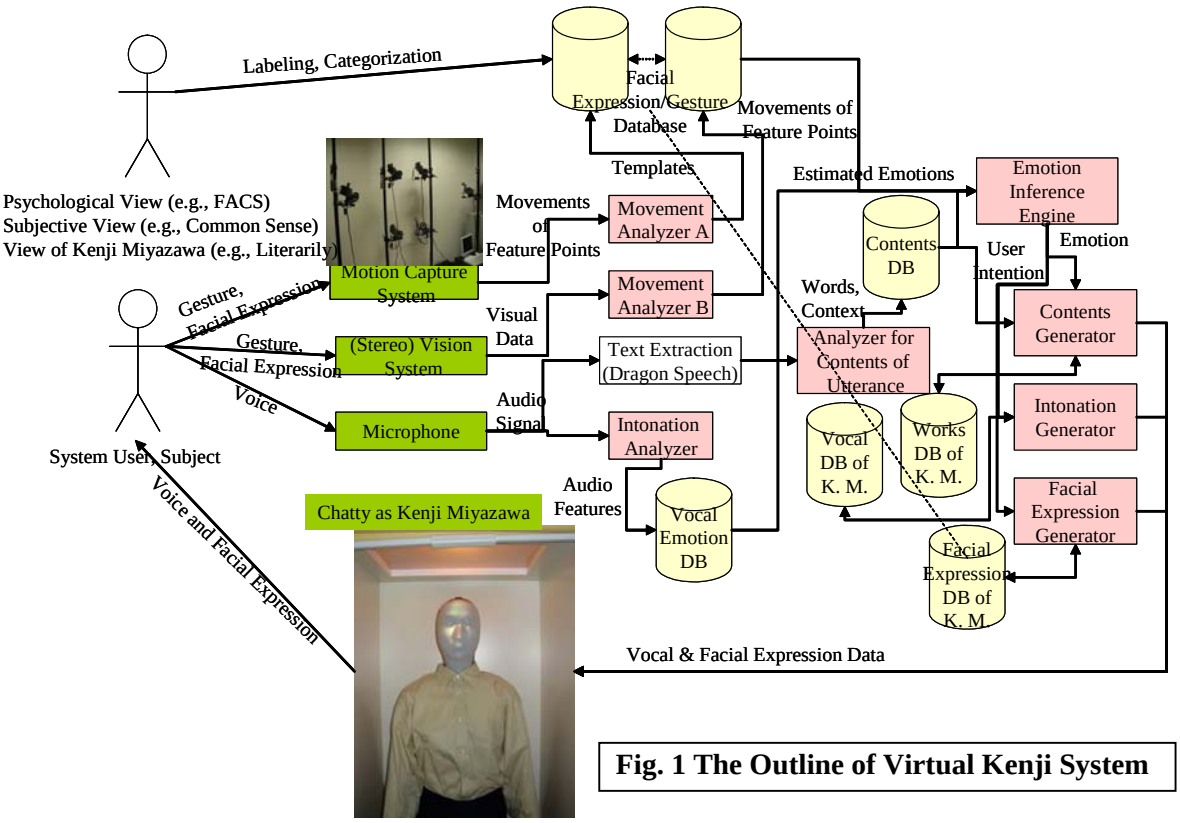



**Fig. 1 The Outline of Virtual Kenji System**

latter speech and face generated scenarios.. The general view of our system design can be seen at Fig. 1.

## 2 The Facial synthesis part of the system__ Recognize and Animate Human Facial Expressions

This section is assigned for the construction of a database based on an automated facial expression analysis method, together with the its uses in facial expression recognition and making facial animation with character's emotional expressions. The automated facial expression analysis method allows users to define the emotional expressions from video sequence. Namely, the defined emotional expressions are automatically identified from the observations of a subject. The actions of the feature points in the defined facial expressions are assumed as caused by a set of system(s) as in [3]. The system identification method (i.e., LSM: Least-Squares Method) is adopted to identify the system. The identified systems are stored in the database, and used both in the recognition and exhibition of the emotional expression by the virtual Kenji system. For the sake of the project, we assume that the user is accustomed to the FACS (Facial Action Coding System) [2], and defines the facial expressions as from the FACS point of view. The main reason for choosing the FACS as definition base is to have the database to exhibit emotional expression of the virtual Kenji. Because the Virtual Kenji ought to face with human user, the expressions should also be comprehensive to that user. The FACS helps to judge which of the six basic emotions is exhibited by the subject. It is considered as universal judgments that can be available across different clutural people. The FACS itself is originally provides a set of still images that depicts the typical examples of facial expressions to be used as templates based on psychological analysis. More recently, however, it is extended to the automatic and spontaneous facial action recognition(e.g., [1]). Our approach differs from that of Movellan et. al., in that we utilize and extract feature points of actions for making animated based expressions used by virtual Kenji. When virtual Kenji should exhibit certain emotional expressions, the set of systems that is labeled as the emotional expression is invoked form the database. The system could estimate the temporal movements of the feature points so that the movements of that points constituting the facial emotional expression come to be available. The movements of the points are now applied to generate the facial movements of the virtual Kenji. The emotional expressions and the facial motion to generate certain utterances are blended in the exhibited

face. This will enable the virtual Kenji to be natural. The detailed mechanisms of the facial expression recognition and animation are described in the next sub-section.

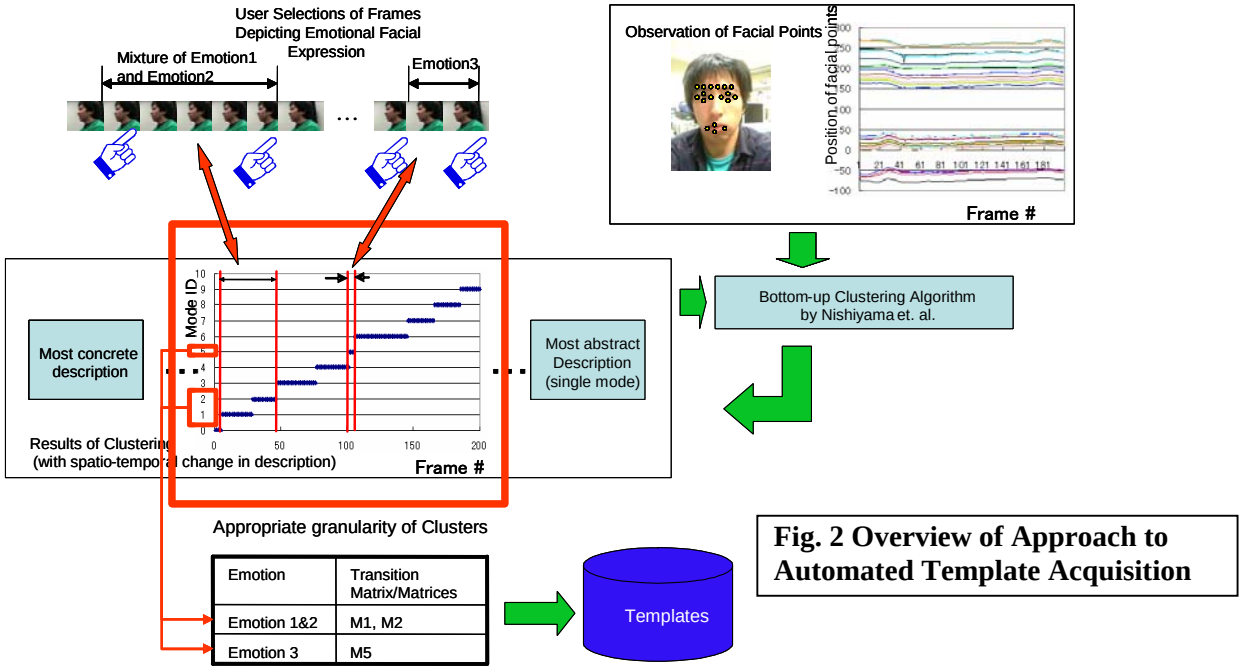



**Fig. 2 Overview of Approach to Automated Template Acquisition**

Fig. 2, shows the overview of our approach. As shown in the figure, the first step to detect emotional facial expressions defined by a user based on FACS is assigning the sequences of frames that express desired emotions from a sample sequence of facial expressions observed in the subject's common sense regular lifestyle. This step is followed by automated clustering by means of the algorithm proposed in [3]. The algorithm identifies the systems that produce the every movement of the facial points. The duration of the motion that can be identified by the same identifier is called mode. The algorithm, then, tries to merge a couple of durations that the same identifier can estimate the movements with the minimum error. This bottom-up clustering continues until the uniform system identifier is determined. Namely, the algorithm as it is finally, generates the identifier that identifies not only the user defined duration, but aslo outside the duration on the same mode.

To extract templates for particular facial expressions, this paper uses the user definition of the desired expressions. Namely, the abstraction process is continued until identifier(s) uniquely identify the duration that contains the user defined facial expressions. Thus, we can get the set of unique identifiers to represent these expressions. We store the (set of) identifiers together with user defined labels of the emotions represented as templates.

With these templates, we detect the user defined expressions from the user's facial expression under consideration.

To extract the user defined facial expressions, template for the particular expression is automatically generated in the proposed approach. The template here is a set of the transition matrices described in the previous section. For this aim, an appropriate cluster level is required. Namely,

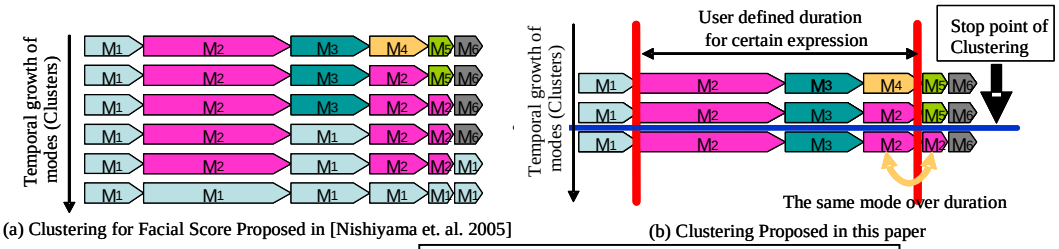



**Fig. 3 Difference in Clustering Method**

a mode, or a set of modes, that is necessary or sufficient to identify the expression should be defined automatically. This section is assigned to describe the generation method of templates, and the scene extraction based on the generated templates.

### 2.1 A Template Generation Method

The basic idea underlying the template generation is defining a termination condition on the clustering. Because, the proposed method is assigned to the user defined sequence of frames that to be extracted as typical facial expression. The mode(s) that is necessary and sufficient to identify the expression can be determined. Fig.3 depicts

the basic idea together with the difference between the proposed method and in the Facial Score. As shown in the figure, the clustering continues until the appearance of the mode that estimates the movements of the facial points across the user defined duration. The mode(s), belongs to the user defined duration, just before the termination condition has invoked, is/are defined as a template. In the case of depicted in the figure, modes $<M_2, M_3, M_2>$ are the template. As shown in this example, the sequence of the modes can also be treated as a template.

Each template is assigned to the user expression definition as a label. A set of labeled templates is then treated as the Facial Expression Database. Namely, the data structure of the expression elements $E_i$ in the database is as follows:

$$E_i = <\mathbf{M_i}, l_i> \quad (1)$$

where, $\mathbf{M_i} = <M_j, \mid j \in D_i>$, $M_j$ is a mode in the user defined duration, $l_i$ and $D_i$ is a label for the expression.

### 2.2. Recognizing Facial Expression with Facial Expression Database

The template here is the transition matrix or a set of transition matrices that corresponds to mode(s). The extraction process is executed by estimating the movements of the facial points from the previous coordinates of the facial points with the template by using Equation (2).

$$\mathbf{FP^{esti}(t)= M_jFP(t-1)+f_j+\omega_j(t)} \quad (2)$$

Where $\mathbf{FP^{esti}(t)}$ is a set estimated x-y coordinates of facial points at time t by means of identifier $\mathbf{M_j}$, $\mathbf{f_j}$ is a bias term,$\omega_j(t)$ is a process noise of the system. To recognize facial expressions, every facial element in the database is treated as a candidate. The movements of the facial points in the expression under consideration are estimated by means of every $E_i$ in Equation (1). The estimation errors between the actual movements of each facial point and the estimated ones are compared. The label(s) coupled with the mode(s) that can estimate the expression with the error value lower than a certain threshold is given to the exhibited expression. Namely, the expression is recognized as labeled emotion.

### 2.3 Animating Facial Expression with Facial Expression Database

When the Kenji system interacts with the system user, the system is to infer the reaction to the user. The reaction contains not only the contents of dialog, but the facial expression suitable for the situation. The Facial Expression Database is diverted in making animation with the facial model of the virtual Kenji system. Namely, the mode(s) labeled as facial expression is reused. With Equation (2), the movements of the facial points can be estimated. This means that the mode can provide the movements of facial points of Kenji system to represent the labeled emotion. The movements of the facial points is then applied to the computer graphically modeled surface of the Kenji's face.

### 2.4 An Experiment on Recognition Ability

To confirm the possibilities of the proposed methods for recognizing facial expressions, a brief experiment is conducted. The database is constructed with facial expressions by a single subject and a single emotion, i.e., 'happy', is aimed to be recognized. The subject is asked to watch a Japanese comedy on DVD with 3 mm makers attached on the face for facial feature extraction. Although the motion capture system can observe movements of the points in 3-dimensional space, the depth information is neglected for simplicity. The experiment has two phases: template acquisition phase, and recognition phase using the acquired templates. In template acquisition phase, the observation is carried out for about four minutes, and we manually choose 200 frames out of the four minutes observation where we can consider the subject is expressing some emotion. Within the 200 frames, we define the frames where the subject is expressing the emotion, 'happy', as shown in Fig. 4.

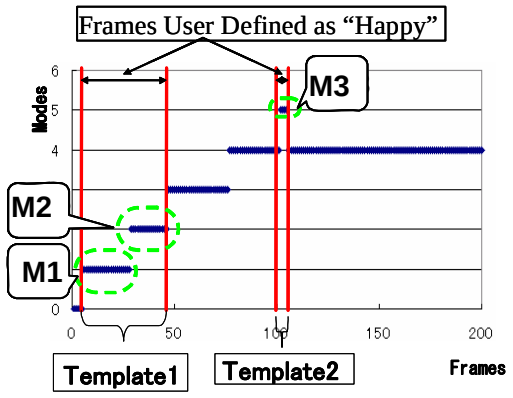



Figure 4, Acquired Templates form User Definition

With the proposed method, we have obtained two expression elements: one consists of two modes, and the other consists of single mode. In our experiment, some expressions that ought to be classified as the same emotion, i.e., 'happy', according to FACS, are distinguished. The reason for the distinction is that we would like to confirm the identification ability for slight change in the expression of the proposed method. This leads to the two expression element for the same facial expression. As shown in Fig.4, the modes of the defined duration are nicely distinguished from the other durations.

Then, we have applied the expression elements in the recognition phase. In the recognition phase, 30 minutes of observation is conducted, and we have obtained 32,000 frames of the facial point movements by means of the motion capture system. To recognize the objective expression from the observation, an estimation of the movements of the facial feature points is calculated by adopting Equation (2) to the acquired templates. The estimation error for each frame and the observed values are compared. The label of expression element with

lower estimation error than certain threshold is the result of recognition. A result of the experiment is depicted in Fig. 5 on which the dots on the line at estimation error 1 are the frames that a human observer judged out as 'happy'. The estimation errors around the dots are abruptly decreased (circled area in the figure). Thismeans that the method can recognize the facial expression 'happy' almost as same as human observer.

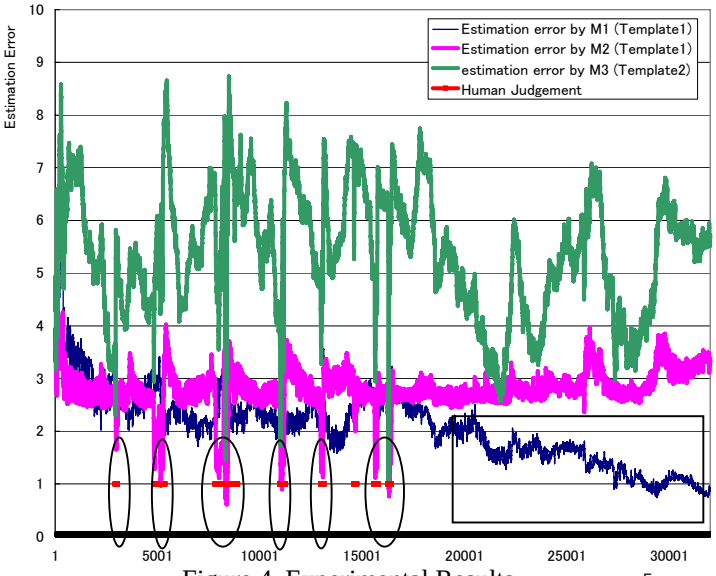



Figure 4, Experimental Results

Although there is another duration that exhibits row estimation errors (area surrounded by a rectangle), they can not be stand as 'happy' because Template 1 consists of two modes, namely M1 and M2. Moreover, we can distinguish the 'happy' expressions by using the templates with appropriate thresholds, because the peaks of the two templates are distinguishable. We apply this method to recognize other emotions of facial expression recognition of Virtual Kenji system..

## 3 The speech synthesis part:

There are some speech synthesis systems in the world. Speech translated by a speech synthesis system have include accent or utterance. But relatd other attributes of speech, for example, volume, speed , tone ,does not change. If a computer is able to change attributes, a computer will be able to speak more natural and users will find emotion in speech generated by computers. Some systems try to change volume and so on. But user should set and tune attributes before computer speaks. While a computer converts text to speech, we changes related attributes. automatically. We change attributes of speech based on properties of text (i.e., emotional featurs). We focus on linguistic information. There are keywords in texts, that important for both writers and readers. If we know keywords, we give them to a computer. If we don't know them, we extract keywords from a text using general information retrieval technique. Using such techniques, for example Term Frequency, we can add a weight reference to words relations in the text. If a computer tries to speak words, which have a weight more than allowed threshold, it turns up the volume and add pause before speaking them. Also, we change attributes of words, which appear in the same sentence with keywords. We change attributes a sentence includes a keyword, too. But changing attribute of words except keywords are more weak than keywords We extract some words, which express emotion from text. We change attributes based on emotion expressed by a word. We have rules about relationships between emotion and voice based on experimental analysis. So we change voice attributes of words based on these rules. And we change attributes of a sentence includes emotional word; like processing a sentence including keywords. We extracts to onomatopoeic from a text. We estimate emtional sound refelcting the onomatopoeic expression and voice change and so on. We focus on the result of parsing based analysis, and feature based grammar. We change attributes first and last sentence in a paragraph. Because these sentence often have most important information. We turn up the volume and slow down. There are some sentences or some words match some rules for changing attributes, we synthesis them. It synthesizes animated, life-like, facial expressions of an individual in synchrony with that individual's speech. The system is speech driven, that is, as an individual speaks the appropriate facial expressions are generated simultaneously. In the database here we collected sound synthesis system that represents the emotional state of sound based on the pitch and amplitude analysis. All sound will be analyzed and categorized according to the emotional state.

A system can read out scripts using speech synthesis technique. So, we have to put accent, intonation, stress, speed, tone and some attribute on scripts before reading out. It is a hard work. So we need a system that supports setting parameters to scripts or sets parameters to scripts automatically. In addition to this, the system doesn't change parameters without stopping. It is good for the system to be able to change parameters during reading out scripts. Because how to speak depends on the environment of this system, user's emotion and so on. So we need to have a method that the system changes parameters for reading out.

Module in Pseudo Miyazawa Kenji System reads out scripts of Kenji Miyazawa. In order to polish up communication with human users, it is necessary to change parameters for reading out automatically. In order to change parameters, we focus on meaning related viewpoints of speech. There are the following seven

viewpoints for speech, Accent, Articulation, Intonation, Phrasing, Prominence, Pause and Rhythm. Speech synthesis technique can support accent and articulation. So we focus on Prominence and Phrasing. Phrasing means that we divide a sentence to some phrases based on syntax and phonology. It looks like chunking. When the system reads out a phrase, it put pause before and after the phrase. Prominence means that the system puts stress on some phrases at reading out scripts. Reading out with prominence helps us to understand what the system says.

Fig 6 shows overview of reading out module. This module has phrasing process and prominence process. We describe them as follow.

First, we describe Phrasing process. If the number of moras in a sentence is more than threshold, this module tries to divide it to some phrases. We don't divide conversation sentences and short sentences. We regard conversation sentences as good for hearing. Also, reading out with a lot of pause is not good to hear and understand what a system says. So it is not worth to divide short sentences to phrase. This module divides a sentence to some phrases using rules based on text syntax. The reason why we define the threshold is as follow. It is not natural that the system reads out at long time period without breath stop. So the module divides a long sentence to some phrase and put pause before and after them at reading out. The threshold is defined based on the number of moras in Tanka, Tanka is a short poem in Japanese and usually has 31 moras. We decided the threshold is 30. We made 5 simple rules for dividing a sentence based on Japanese grammar. For example, we regard conjunction as end of phrase. A rule has a condition part and a conclusion part.

A condition part consists of morphological strings. A conclusion part of rules says that picks up the phrase (put pause). The system tries to match the condition part of rules to result of morphological analysis. If the condition part is true, it does the action defined in a conclusion part. This module reads out scripts written by Japanese. So we make rules based on Japanese grammar. If the system reads out scripts written by other language, we should make rules based on that language's grammar.

Next, we describe prominence process. In order to put stress on a phrase, the module selects a phrase using rules based on syntax. It picks up onomatopoeia, rheme phrase and theme phrase to put stress. Longman Dictionary says that Onomatopoeia is "the use of words that sound like the thing that they are describing". We think that onomatopoeia is an important point for hearing. So the module picks up onomatopoeia using a dictionary and puts stress on them.

Rheme and Theme are important phrases to understand what someone says. We put stress on these phrases at speaking. So we put stress on rheme and theme phrase. The module tries to find rheme and theme using rules based on

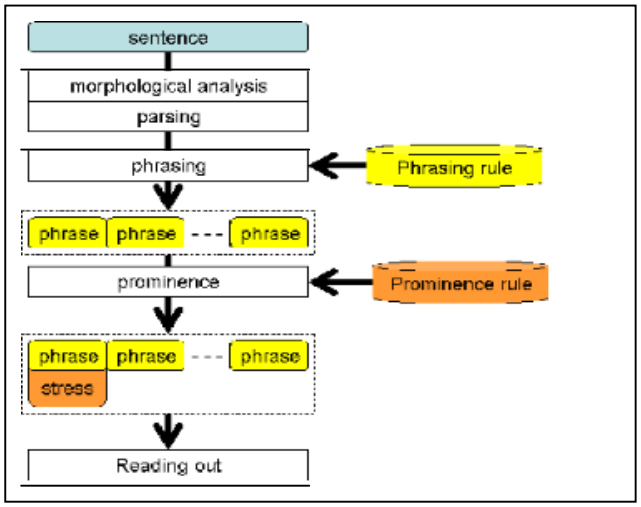


Fig. 6. An overview of Reading out Module

syntax. The structure of these rules is same as one of rules for phrasing. But a conclusion part says that puts (i.e., assigns) stress on a phrase defined in a condition part. The module uses outputs of parser with a dependency grammar at phrasing and prominence process. So the accuracy of phrasing and prominence is influenced the accuracy of parsing and morphological analysis.

In order to show the effectiveness of this idea, we did experiments as follow.

First, the module reads out documents in the 4 patterns. Pattern 1 is that the module doesn't put stress and pause. Pattern 2 is that the module puts stress and pause on documents. Pattern 2 is that the module puts pause. Pattern 3 is that the module assigns stress. When reading out, the module makes the order of the pattern in random. After hearing them, participants answer the following questions, "Which pattern is easiest to hear?", "Which pattern is most difficult to hear?", "Which pattern is most emotional?", "Which pattern is not emotional?", "Which pattern do you like?", "Which pattern do you dislike?". If most participants select pattern 1 is most emotional, we think that phrasing and prominence are good for to be emotional. Next participants assign stress and pause on documents. We compare these documents. Those documents assign stress and pause by a module. We calculate recall and precision. Recall is that the number of characters which assigned stress by a person and the module is divided by the number of characters which assigned as stress by a person. Precision is that the number of characters which assigned as stress by a person and the module is divided by the number of characters which assigned as stress by the module. If these values are high, we think the module is appropriate to assign.

In order to do experiments, we made a prototype of the reading out module. We use "Cabocha"[5] as a parser with a dependency grammar and "SMARTTALK" (http://www.oki.com/jp/Cng/Softnew/JIS/sm.html) as a speech synthesizer used in this prototype. Cabocha is a Japanese dependency structure analyzer developed by Graduate School of Information Science, Nara Institute of

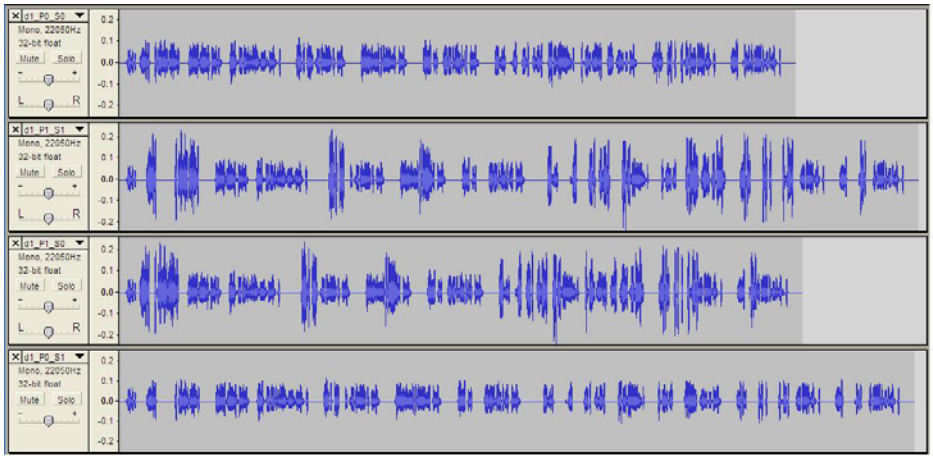
| Table.1: The result of first experiment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pattern | | | Hearing | | Emotional | | Like | |
| No | STRESS | PAUSE | easiest | Most difficult | Rich | Poor | Like | Dislike |
| 1 | NOT PUT | NOT PUT | 71% | 4% | 13% | 34% | 65% | 5% |
| 2 | PUT | PUT | 2% | 40% | 40% | 12% | 14% | 38% |
| 3 | PUT | NO PUT | 6% | 38% | 43% | 7% | 5% | 38% |
| 4 | NOT PUT | PUT | 20% | 17% | 4% | 46% | 16% | 19% |

Science and Technology. SMARTTALK is developed by OKI Electric Industry Corporation. It can specify the volume, the speed, the tone and the intonation of the sound by putting commands, like markup language, into documents to read out. The module translates stress to these commands and puts these commands on documents. The module gives documents with commands to SAMARTTALK. While SMATTALK reads out document, it changes the volume, the tone, the speed and the intonation, automatically. There are 9 participants in this experiment. The module reads out 5 paragraphs extracted from Kenji Miyazawa's novel. We show some sound waves made by the prototype module on Figure.7. These sound waves are made from same sentence by the module. But the module change stress and pause. Top is a pattern 1, which represents that the module have not assigned stress and pause.. Second from the top is pattern 2, which represents that the module assign stress and pause. Third from the top is pattern 3 which represents that the module assign stress. And bottom is pattern 4 which the module assign pause. Pattern 2 and 4 are assigned pauses. So they are longer than others. Pattern 2 and 3 are assigned stress. So their pitches are wider than others.

Table.1 shows the outline of this experiment. The experimental result says that stress contributes to be emotional. But it is difficult for people to hear a document with stress and pause. Most participants dislike this type utterance. We think the reason as follow. The module puts many stresses and pauses on documents and the change of the sound got intense. So these utterances were difficult to hear. We compared the output of this module and documents that assign on stress and pause by human. Table.2 shows the outline of the experimental results. "Major" in table 2 means the result of compare with output of system and documents putted stress and pause by more than half of participants. The experimental result says recall of pause is good. It is that most pauses putted by the module are same as pauses assigned by a participant. But recall of stress and precision are not good. One of the reason is the number of stress assigned by this module is larger than the number of stress assigned by a participant. Table.3 shows the number of stress and pause. The number of stress assign on by this module is about 3 times the number of stress assigned on by a participant. Participants tend to assign on a few stress and pause on a sentence. We guess that participants think a lot of stress and pause weaken their effort and be difficult to hear. The result of experiment for hearing supports this supposition. In addition , we compared pauses assigned by participants and the module and characters assigned on as stress by them. Participant assign on stress on words which express action, for example *shout*, or include exclamation and/or interjection words. The module puts stress on phrases in a document based on syntax structure. There are differences among the module and participants about assignment on stress. Also, the module has a few patterns to express stress. This is one of reason that it is not easy to hear; also, participants don't like this module's produced speech tone or style.

| Table.2 the result of comparing phrasing and prominence | | | | |
|---|---|---|---|---|
| PAUSE | AVERAGE | MAX | MIN | Major |
| RECALL | 82% | 100% | 70% | 88% |
| PRECISION | 24% | 44% | 14% | 16% |
| STRESS | AVERAGE | MAX | MIN | Major |
| RECALL | 46% | 65% | 28% | 38% |
| PRECISION | 11% | 25% | 4% | 4% |

| Table.3 the total number of pauses and stress putted by a system or participants on 5 documents | | | | |
|---|---|---|---|---|
| | MODULE | AVERAGE | MAX | MIN |
| PAUSE | 86 | 26.22222 | 47 | 16 |
| STRESS | 279 | 72.22222 | 186 | 23 |



The experimental result shows that prominence and phrasing are good to read out, emotionaly. But the power of the present module is weak. So we refine the module as in the following, based on the result of analysis experimental results:

1. remove unnecessary pause and stress

2. have more patterns to express stress
3. use other viewpoints, for example semantics, rhythm for reading out

We describe the module has phrasing and prominence process. Phrasing is how to divide a sentence to some phrases and put or assign pause on them. Prominence is how to put stress on some phrase. This idea is based on the viewpoints of human speech. The experimental result shows that phrasing and prominence are good to read out a document more emotionaly. But reading out by the module is difficult to hear. There are some reasons. For examples, the module puts unnecessary stress, or expresses stress in a few extracted patterns. We will enhance the module to read out more natural and emotional based on experimental results.

Generally, people speak with his emotion. Maybe, emotion expressed by facial expression is clearer than emotion included in speech. However, it is significant to extract emotion from speech. Some researchers have tried to extract emotion from speech. Most of them paid attention to the feature of speech, for example fundamental frequency, speech rate, pitch and gain. But we don't have best feature of speech to extract emotion. We think these features to recognize speech. So we should focus on other features. Then, we focus on pause and change of sound wave pattern. Most researchers regard pause as a delimitation of speech and chance to recognize. However, people changes position and length of pause based on his emotion. We guess that pause expresses emotion. So we focus on it to extract emotion from speech. Moreover, we guess that the change of speech pattern appears when emotion appears as well as the facial expression, and tries to extract emotion from the difference compared with a usual pattern. It is difficult to decide features contribute to extract emotion. We decide features based on experimental results.

## 4 Conclusion

This work presents the preliminary results of our cognitive virtual interactive model based on Kenji cognitive model. This is the 1st stage outcome on this project.

Using motion Capture system installed in our laboratory (Fig.1), we have extracted emotions characteristics of Kenji scripts based on emotional analysis of many users who are reading Kenji scripts with emotional harmony with it and feelings. This basically, has been used to extract emotional feature based on common sense knowledge and human science and other analysis parameters. Also, [4] software has been used to confirm the collected or captured emotional features have the same labeling, (label reflect the emotional feature of user or Kenji extracted behavior), as user him/her self does on the video observed/recorder on them when reading the Kenji scripts as shown on Fig.8, which shows a simple labeling for the emotional feature by the users. Scripts text analysis also has been done using emotional analysis. The speech of users has been used to extract the emotional feature as well from it.

We think this system will be used to highlight the motional feature between any system and human user to be feasible and comprehensible, that the user interface enhancement may approach the human level emotionally and cognitively. This approach can participate to make computer be more flexible and friendly. Such softness and flexibility, being reflecting to be more human, can contribute to enhance the interaction between computer and human and make it more ubiquitous for our needs.

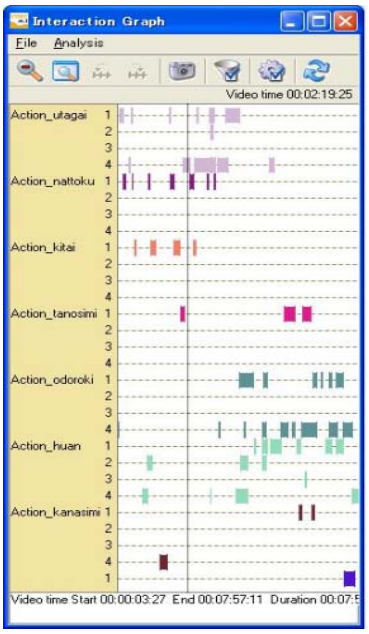



**Fig. 8 Emotional Labeling Using "INTERACT"**

***Acknowledgment'***

We would like to thank all our laboratory students who contributed and participated to build and run the experiment and make this work a real practice. Also, many thanks go to Prof. Tamio Sasaki for providing us with an overview on Kenji literature world, and related artwork analysis.

This research is supported by a research grant by Iwate Prefectural university fund projects.

***References:***

[1] J. R. Movellan and M. S. Bartlett. The next generation of automatic facial expression measurement. In Paul Ekman, editor, What the Face Reveals. Oxford University Press, 2003.

[2] P. Ekman and W.V. Friesen: Unmasking the Face. Prentice Hall, NY.,1975

[3] M Nishiyama, H Kawashima, T Hirayama and T Matsuyama: Facial Expression Representation based on Timing Structures in Faces. IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2005):140-154,2005

[4] INTERACT Trademark software by Mangold co. http://www.mangold-international.com/

[5] Taku Kudo, Yuji Matsumoto :"Fast Methods for Kernel- Based Text Analysis", ACL 2003 (2003)