Personal Media Producer: A System for Creating 3D CG Animation from Mobile Phone E-mail

Koichi Emura
The University of Tokyo / Research Center for Advanced Science and Technology, Tokyo, Japan
Email: emura@mpeg.rcast.u-tokyo.ac.jp

Makoto Yasugi, Toshiyuki Tanaka, Seiya Miyazaki and Sachiko Motoike
Matsushita Electric Industrial Co., Ltd. / Network Development Center, Tokyo, Japan
Email: {yasugi.makoto, tanaka.toshi, miyazaki.seiya, motoike.sachiko}@jp.panasonic.com

***Abstract*— Mobile phone e-mail is increasingly being chosen by consumers as their primary communication tool. Messages exchanged among mobile phone users frequently contain "emoticons (smileys)" and "emoji (pictograms)". Some consumers have started to use services that add a relevant animated image to a message. Consumers will start to exchange animated messages if they are available.**

This paper proposes a method of generating animation sequences from plain text messages without an authoring process. This method utilizes information available in the consumer's ubiquitous computing environment to obtain necessary information not included in the original e-mail.

The results of the preliminary evaluation show that a user needs around two minutes to create a one-minute animation using this method, a 3.6 times improvement in creation efficiency over that of a prior method.

***Index Terms*— Mobile Applications, Mobile Phone E-mail Services, Computer Graphics Animation, Context-aware, Cinematography, Speech Act, Case Frame**

I. INTRODUCTION

A report on the survey "ICT Use by Individuals" [8] says that Japan has an especially high utilization rate of mobile phone e-mail. According to the report, the utilization rate of mobile phone e-mail in Japan is 87.7 % while the utilization rate of PC e-mail is 94.2 %. In contrast, in the United States, the utilization rate of PC e-mail is high at 96.1 %, whereas the utilization rate of mobile phone e-mail is extremely low at 12.4 %. In the Republic of Korea, mobile phone Internet usage is similar to that of Japan, but the utilization rate of mobile phone e-mail is 43.1 %, which is still far lower.

Referring to the survey site "HER STORY" [15], 524 out of 941 females (56 % of them) prefer mobile phone e-mail to PC e-mail or making phone calls as their primary communication tool. The survey goes on to say that 153 out of 225 females in their 20s and 273 out of 471 females in their 30s choose mobile phone e-mail as their primary communication tool, which is 68 % and 58 % respectively. Other reports show the same trend: that mobile phone e-mail will be chosen as a primary communication tool.

Fig. 1 shows a sample mobile e-mail exchange between females in their 20s, which roughly says "Last night I received a love letter from my boyfriend!", "I am going to send an answer", and "It's late. Good night." As you may notice, pictograms are used instead of words or punctuation marks. Emoji are a common characteristic of mobile phone e-mail as well as emoticons.

Mobile phone network operators provide emoji in their proprietary character code area. NTT DoCoMo provides 252 characters, KDDI 487, Vodafone 480. Table I shows some extracts from NTT DoCoMo's list of pictograms [17].

This tendency to make messages visual goes even further.

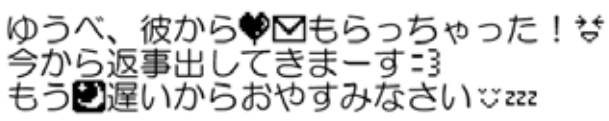

Figure 1. Sample of mobile phone e-mail

TABLE I.
SAMPLE LIST OF PICTOGRAMS

pictogram	definition	pictogram	definition
	Heart		Night
	Mail		Relief face
	Happy face	zzz	Sleepy
	Dash		

Based on "Creating Mobile Animation Messages without Authoring", by K. Emura, M. Yasugi, T. Tanaka, M. Miyazaki and S. Motoike which appeared in the Proceedings of the IEEE Consumer Communications and Networking Conference 2006, Las Vegas, U.S.A., January 2006. © 2006 IEEE.

II. CURRENT ANIMATED MESSAGE SERVICE

A. Deco-mail Service

NTT DoCoMo provides a deco-mail service [16] consisting of ready-made HTML templates with colorful animated GIF pictures. A consumer can select his/her favorite template from the listed menu that matches his/her mood.

B. Animated Greeting Card Service

FunMail provides an Animated Greeting Card Service [14]. It creates a visual greeting message by picking a matching visual image for a text message that a consumer types in. This system uses keywords extracted from the input text to select the matching animated image.

C. Deru Moji 3D Pictogram Display

Vodafone provides Deru Moji 3D Pictogram Display [18]. 3D animations corresponding to certain pictograms, emoticons, and the vocabulary extracted form the input text will fill the screen, popping out and spinning around in the e-mail received.

The above visualization services attach a relevant picture or pictures to the text typed in, to make content fun and exciting. What consumers actually want differs, however. Consumers are already making the content itself visual by using emoji, and they want to convert the text into animation [6].

This paper explores methods of creating animated messages that correspond precisely to the content of the consumer's typed-in text.

III. PROBLEM S TO BE SOLVED

Two problems, at least, need to be solved when creating an animated message from the text typed in by a consumer.

The first problem is that messages carry too little information - they only rarely contain descriptions of places, persons, or objects. This knowledge is tacitly shared by the sender and the message recipient or recipients. Without this information, the animated message cannot have an appropriate background, persons in the appropriate shapes, or objects in the correct colors.

The second problem is that animated messages are hard to see or boring without good direction. Without direction, a scene of two people talking may become a continuous image where two people are standing still in the center of the picture.

IV. RELATED WORK

Previous research has tried to solve these problems through templates.

Sakamoto et al. [11] proposed the system "ComicDiary" that creates a comic strip based on the records of the tour of an exhibition. No text input is necessary. The comic, composed of 12 frames, is automatically generated in diary style. The system selects a template based on the user's attributes. For example, if the user is a child, a template with a simple structure and easy-to-understand expressions is selected.

Hayashi et al. [5] proposed the Automatic Production Engine (APE), which converts input scenario descriptions into a TVML (TV program Making Language) [19] script. They prepared templates for conversation scenes in a TV program. For example, a news program template defines two characters as newscasters who stand or sit near each other, not moving, and talk in turn. Camerawork directions are also described in the template.

These work fairly well within their limitations. The generated animation is sometimes better than the animations created by a novice non-professional user.

However, extending the applicable area is quite difficult using these approaches. It is impractical to expect a service provider to prepare a variety of templates of the many situations in which consumers will wish to send messages.

For this reason, this paper takes a different approach to solving the problem, by making a scenario from the input text with information obtained from the ubiquitous computing environment around the consumer.

Arbrecht et al. [1] proposed a facial-animation-from-text system that combines synthesized speech and non-verbal facial animation sequences by interpreting input sentences and emoticons. Although this system shows a simple but effective method to interpret emoticons for generating appropriate facial expressions, it is hard to apply emoticons to other expressive directions such as camerawork.

Persson [10] proposed Expressive SMS, an avatar-based messaging application that allows users to concatenate and annotate simple animations. His study reported that the meaning of the created animations heavily depended on the shared context between sender and recipient, with the most common usages relating to expressing emotion, attitude, or opinion in relation to some phenomena, and that the face is important as a conveyer of expressiveness.

V. SYSTEM OVERVIEW

We have implemented a prototype system called the Personal Media Producer (PMP), based on the concept of converting mobile phone e-mail to animation. The PMP system currently runs on Java Standard Edition with Java3D. In this section, we describe the system overview and show output examples.

The graphic user interface (GUI) of the PMP system shown in Fig. 2 consists of an animation window, message window, emoticon button, emoji button, play/stop buttons and a send button.

Clicking the emoticon/emoji buttons cause menus to pop up, allowing the user to add emoticons/emoji into the message window. These menus are shown in Fig. 3. When an emoticon/emoji is chosen, the user returns to the message window where the emoticon/emoji has been inserted. Playing the message means that the system plays an animation corresponding to each sentence sequentially.

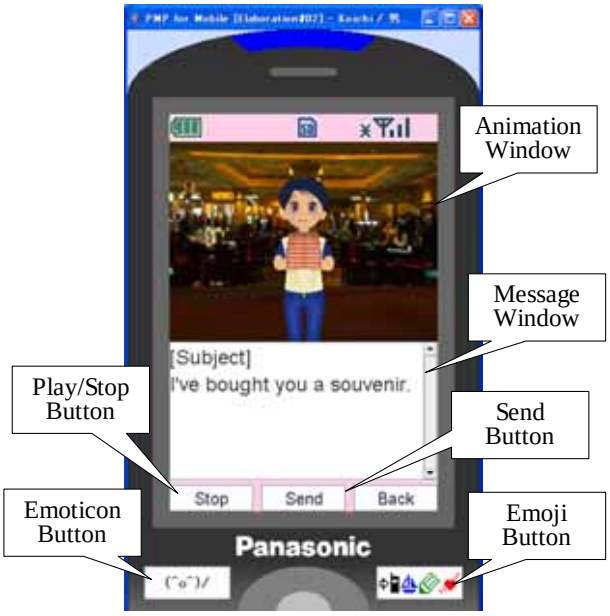

Figure 2. Graphic User Interface of PMP system

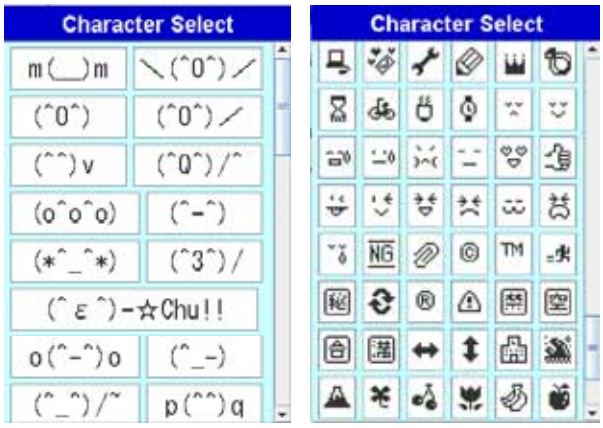

Figure 3. The menus of emoticons and emoji

The process flow of the PMP system is shown in Fig. 4.

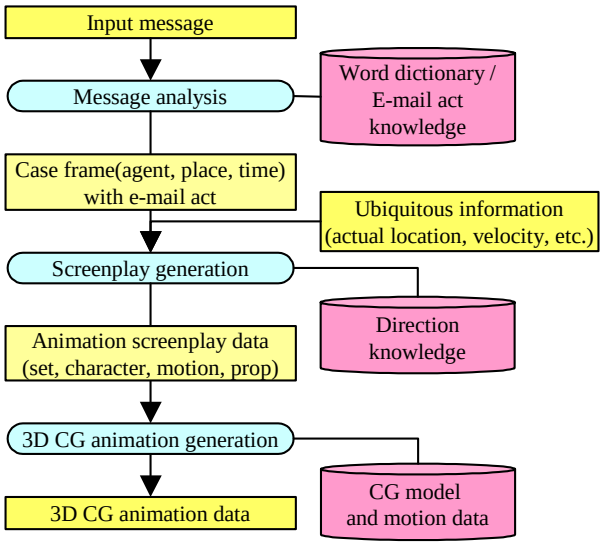

Figure 4. PMP System architecture

The PMP system analyzes the message text as input from the keyboard, creates a case frame, and extracts the acts expressed in the e-mail. Then, it generates screenplay data, converts it into 3D graphical animation data, and, finally, displays the animation. For example, Fig. 6 shows the screenplay data generated by this system from the message shown in Fig. 5.

"Hello. I've bought the souvenir which you asked me to buy. Please wait till I come back."

Figure 5. Example of input message

```
Scene: place="Asakusa"
Super: text="Hello. I've bought the souvenir which you
            asked me to buy. Please wait till I come
            back."
Direction0: who="USER" whatAction="have"
            whatObject="a Japanese fan"
Direction1: who="USER" whatAction="greet"
            whatObject="THE CAMERA"
Direction2: who="USER" whatAction="show"
            whatObject="a Japanese fan"
Direction3: who="USER" whatAction="request"
            whatObject="THE CAMERA"
Direction4: who="USER" whatAction="walk away"
Resource: name="Asakusa" model="asakusa01.jpg"
Resource: name="USER" model="character01.model"
Resource: name="a Japanese fan" model="fan.model"
Resource: name="greet"
            motion="bow_formally.motion"
Resource: name="request"
            motion="bow_informally.motion"
Resource: name="show" motion="show.motion"
Resource: name="walk away"
            motion="walkaway.motion"
```

Figure 6. Generated animation screenplay data

VI. Mobile Phone E-mail-To-Animation Method

Animation screenplay data, direction data, model data, texture data, motion data of a performing character, data of props and other data are necessary for a rendering program to play an animation.

Animation screenplay data describes the information of the place, and lists a sequence of actions. Fig. 7 shows an example of this type of data (summary). Fig. 8 shows a snapshot of the animation generated from that data.

```
Place:      Akihabara
Direction1: A man (the mobile message sender) raises
            his left arm cheerfully.
            He has a big, yellow paper bag in his hands.
Direction2: He shows the bag.
Direction3: He waves his hand to the message recipient.
Direction4: He runs toward a station.
```

Figure 7. Example of animation screenplay data (summary)

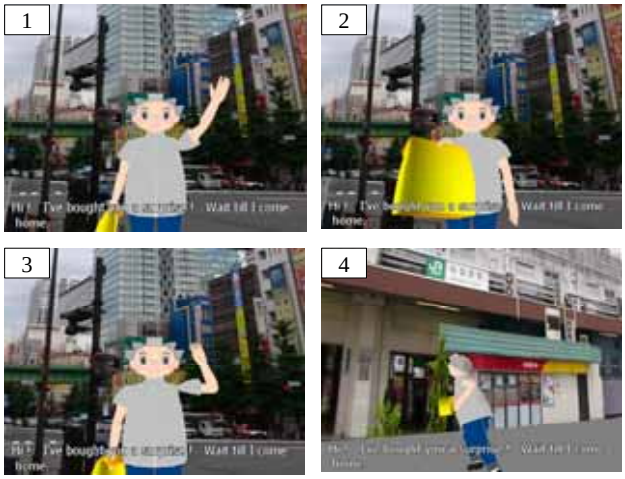

Figure 8. Snapshot of the generated animation

A. *Interpolation of implicit information*

Our objective is to establish a method for generating animation screenplay data from the e-mail message shown in Fig. 9. The message itself lacks sufficient information to do this. For example, the direction to use a photo of Akihabara as background set data is not included in the mail message, and it does not have any indication of how a performer should act in the animation.

"Hi! I've bought you a surprise! Wait till I come home."

Figure 9. Example of a mail message

Current mobile phones have numerous functions other than making phone calls. Internet access, GPS, electronic money, and PIM data storage functions are common features of the latest mobile phones. The information generated, collected, and stored by these functions is valuable. For example, geographical information from the GPS function can be used to set the background data, and the data on what was bought, stored in the electronic money function, can be used to select the prop that the performing character has in his hand.

We call this information that can be acquired from the ubiquitous computing environment, 'ubiquitous information'. Fig. 10 shows the typical roles of ubiquitous information.

"Location" acts on the selection of CG background data.
"Velocity" acts on the selection of a performer's action.
"Purchase" acts on what a performer has in their hand.
"Mail-to" acts on how to direct the animation contents.

Figure 10. Typical roles of ubiquitous information

Fig. 11 is an example of ubiquitous information that can be valuable for converting a message text into the animation.

To create animation screenplay data, two kinds of conversion rules are also needed.

(1) Rules for guessing consumer intention, which will affect the action decision process.

For example, the word "Hi!" indicates the sender's informal relationship with the message recipient, and the phrase "buy a surprise" indicates that the performing character should not show the exact shape of the gift he has bought.

(2) Rules for generating detailed motions from each action.

For example, an informal greeting action must be either expressed by the motions "raising one's right arm", and/or "waving one's hand."

Ubiquitous information might also affect which conversion rule has to be used when generating animation screenplay data. For example, in a case where the "purchase" is a toy and "mail-to" is the sender's daughter, the intension "do not show the exact shape" is converted to the action "carrying a paper bag."

Location? -> Akihabara.
:To use a photo of Akihabara as CG background data
Velocity? -> fast.
:To show that the performer is going to board a train.
Purchase? -> a talking yellow mouse toy.
:To use a big yellow paper bag to show that the performer is carrying something.
Mail-to? -> a daughter.
:To direct animation like cartoon, making his face "big and surprising" :-)

Figure 11. Example of ubiquitous information

This mechanism converts e-mail message text to animation screenplay data. As a result, a mobile animation message is generated without the need for an authoring phase.

In "Direction 1" in Fig. 6, for instance, the "USER" character action is "greet". Since the mail address of the message recipient is classified as a "COMPANY MEMBER," the system associates the motion data "bow_formally.motion" with the polite, or formal, version of the action "greet". On the other hand, if a message recipient's address is classified as a "FAMILY MEMBER," the system associates the motion data "bow_informally.motion" which behaves informally with the same action, "greet". In "Direction 0" and "Direction 2", the system replaces the object "the souvenir" with "a Japanese fan" from the purchase record of the "USER". Since the input message does not include any location information, the system decides the place "Asakusa" based on the user's location information.

B. *Screenplay evolution by using e-mail acts*

The message analysis process classifies the input message into six classes of e-mail acts that we newly defined to generate case frames [4]. The e-mail acts are 'REPORT', 'REQUEST', 'QUESTION', 'INVITE', 'AFFECT' and 'WISH', and are designed based on an 'illocutional' act, which is a part of speech act theory [2]. An e-mail act corresponds to an act performed by writing something in a mail message. For example, the utterances "I've bought the souvenir" and "Please wait" in the input

message are classified into the e-mail acts ‘REPORT’ and ‘REQUEST’ respectively. The case frames ‘buy (agent = ”I”, object = ”the souvenir”, place = ”n/a”, time = ”past”)’ and ‘wait (agent = ”you”, object = ”me”, place = ”n/a”, time = ”future”)’ are generated from the written utterances.

The screenplay generation process utilizes the e-mail act to create and insert new case frames. Because the e-mail act is ‘REPORT’, the case frame ‘buy (agent = ”I”, object = ”the souvenir”, place = ”n/a”, time = ”past”)’ is replaced with the new case frame ‘show (agent = ”I”, object = ”the souvenir”, place = ”n/a”, time = ”n/a”)’ to report that the user bought the souvenir in the past. When the e-mail act is ‘QUESTION’ as another example, the case frame ‘buy (agent = ”I”, object = ”the souvenir”, place = ”n/a”, time = ”past”)’ is replaced with the case frame ‘go_into (agent = ”I”, object = ”the shop”, place = ”n/a”, time = ”n/a”)’. Since the action ‘have’ is inevitably needed to perform the action ‘show’, the new case frame ‘have (agent = ”I”, object = ”the souvenir”, place = ”n/a”, time = ”n/a”)’ is appended to the head of the scene including the case frame ‘show ()’.

Furthermore, the screenplay generation process associates values extracted from ubiquitous information corresponding to place, object and motion, and finally generates the animation screenplay data.

In “Direction 1” and “Direction 3” in Fig.7, the system replaces the object “you” with “THE CAMERA” based on Direction knowledge [9].

After generating the animation screenplay data, the 3D CG animation generation process loads the CG model data such as character, object and set, and locates them in the scene. Then the characters act according to the skeletal motion data.

Fig. 12 shows the generated graphics made by the animation screenplay data shown in Fig. 6. Although the described content of the input mail message is almost the same as the message shown in Fig. 5, the results are significantly different. The background becomes Asakusa and the character acts formally, with a Japanese fan in his hand. All of these differences are caused by the ubiquitous information.

Figure 12. Snapshot of a result animation

C. *Expressive direction*

Further, high-level information from the message analysis process such as feeling parameters (type and level) is used to generate expressive direction such as facial expression, acting and camerawork. Moreover, emoticons and emoji are translated into a feeling type that triggers expressive direction.

The feeling types consist of three parts. The first part is facial expressive feeling, such as ‘Enjoyment’, ‘Surprise’, ‘Fear’, ‘Anger’, ‘Disgust’, ‘Sadness’, and ‘Contempt’, which are universally expressed as a common expression by all the cultures [3]. The second part is behavioral expressive feeling, such as ‘Energetic’, ‘Lovely’, ‘Trivial’, ‘Confuse’, ‘Heart-warming’, ‘Polite’, ‘Hasty’, ‘Nervous’, ‘Glad’, ‘Expect’, and ‘Lazy’, which have been extracted from the results of analyzing 54 mobile phone e-mails that 10 subjects actually sent and received, and 32 mobile phone e-mails were referred to from the research report of information behavior [12]. The third part is intermodal feeling, such as ‘Bright’ and ‘Dark’, which are known to be common between perceptions of auditory and vision [7].

The feeling level shows strength of the feeling, and takes one of the values 1(weak), 2(fair) or 3(strong).

The message analysis process extracts the feeling parameters from the emoticon and the emoji included in the mail sentence based on the co-relation table between feeling parameters and emoticon or emoji which are shown in Table II and Table III

In order to convey the complex feelings that are contained in actual mobile phone e-mail, the co-relation table assigns two or more feeling parameters to each emoticon or emoji. Moreover, feeling parameters can be used as triggers for various kinds of cinematography and facial animations.

The facial animations shown in Fig. 13 are pre-defined. The types of the facial animations are NORMAL, LAUGH, ANGRY, CRY, SURPRISE, RELIEF, GIVEUP, WONDER, WORRY, FEAR, HASTY, NERVOUS, HOPE. Each type has a weight value for each feeling parameter. The range of the weight value is from –3 to 3, and the value of feeling type FT(i) and facial animation FA(j) means that the degree feeling type FT(i) causes the facial animation FA(j).

TABLE II.
FEELING PARAMETERS FOR EMOTICONS

Target	Feeling paramters
m（＿＿）m	Sadness=2
（＾３＾）	Glad=2,Expect=1
（－－）	Polite=2
（＊＾＿＾＊）	Glad=2,Hasty=1
（－＿－＃）	Anger=2,Fear=1
（＾－＾；）	Hasty=3
m（＿＿）m	Sadness=2
＼（＾Ｏ＾）／	Glad=3
（＾－＾）	Glad=2
（＊＾＿＾＊）	Glad=2,Hasty=1
（＾３＾）／	Glad=2,Expect=1
（＾ε＾）－☆Ｃｈｕ！！	Glad=2,Expect=1
ｏ（＾－＾）ｏ	Enjoyment=2,Expect=3
（＾＿－）	Glad=2,Expect=1
（＾＿＾）／～	Sadness=2,Lovely=1

TABLE III.
FEELING PARAMETERS FOR EMOJI

Target	Feeling parameters
	Lovely=2,Glad=1
	Sadness=2,Trivial=1
	Lovely=1,Glad=1
	Glad=2
	Anger=2,Trivial=1
	Trivial=2,Sadness=1
	Confuse=2
	Enjoyment=2
zzz	Heart-warming=1,Trivial=1
	Surprise=2
	Hasty=2,Energetic=1
	Hasty==3
	Energetic=2
	Confuse=2,Lovely=1
	Heart-warming=3

To select a facial animation, the system calculates the total weight of each feeling type FT(i), where i = 1, …, 20 for each facial animation FA(j), and where j = 1, …, 13 in (1), and selects the FA(j) where the total weight $w_{total}(j)$ is the maximum.

$$w_{total}(j) = \sum_{i=1}^{20} w_{FA}(i, j) \times l_{FT}(i) \quad \ldots (1)$$

- $w_{FA}(i, j)$ is the weight of the feeling type FT(i) for the facial animation FA(j)
- $l_{FT}(i)$ is the level of the feeling type FT(i)
- FT(i) = {'Enjoyment', 'Surprise', 'Fear', 'Anger', 'Disgust', 'Sadness', 'Contempt', 'Energetic', 'Lovely', 'Trivial', 'Confuse', 'Heart-warming', 'Polite', 'Hasty', 'Nervous', 'Glad', 'Expect', 'Lazy', 'Bright', 'Dark'}
- FA(j) = {'NORMAL', 'LAUGH', 'ANGRY', 'CRY', 'SURPRISE', 'RELIEF', 'GIVEUP', 'WONDER', 'WORRY', 'FEAR', 'HASTY', 'NERVOUS', 'HOPE'}

VII. EVALUATION AND RESULSTS

We conducted a preliminary evaluation of content creation time, comparing the input and output times for the PMP versus TVML mini [20] systems. TVML mini is a more user-friendly graphical authoring tool based on TVML. While PMP leverages ubiquitous information and knowledge about the sender and receiver, TVML requires the user to input a full scenario, including details of the characters, actions, and objects, as well as the dialog.

The input time was measured as the time from when a subject decided on the written content to when that subject completed inputting the text. The output time was the playback time from beginning to end of the resulting animation. For each system, a single subject input eight example sentences three times each, and we took the average of the averages of the input times for that subject.

The results obtained through the evaluation are shown in Table IV. The I/O ratio of PMP is 2.1, which means that a user needs around two minutes to create a one-minute animation. PMP's creation efficiency was 3.6 times that of TVML mini, whose I/O ratio was 7.6, and thus PMP provides mobile users a new experience with a significantly lower usage burden.

TABLE IV
AVERAGE INPUT/OUTPUT TIME

	Input time [s]	Ouput time [s]	I/O time ratio
PMP	8.7	4.1	2.1
TVML mini	54.6	7.2	7.6

Figure 13. Example of facial animation

VIII. FUTURE WORK

A. *Ubiquitous information reliability*

There are many types of ubiquitous information, and many methods for searching and retrieving it. The current PMP system uses this information without considering its

primary purpose. Some data may be inauthentic and some unreliable. Using such data may result in a confusing and inconsistent animation. A common criterion for evaluating the reliability of ubiquitous information is needed.

B. Utilizing other ubiquitous information

In the current implementation, ubiquitous information extracted from the mobile terminal device is utilized to create the animation, assuming that it represents what the sender has experienced.

However, other ubiquitous information may be available and can be used. For example, ubiquitous information related to the message recipient could be useful.

C. Preference/Style

When the same message is entered and the same ubiquitous information is available, the current prototype system generates the same animation.

The preferences of the sender or the message recipient should be taken into consideration when generating the animation. Some people prefer a short story, and some prefer a characteristic, or even whimsical, walking style. The former should be reflected in the process of generating animation screenplay data, and the latter in the process of selecting motion data.

D. Déformer

'Déformer' is a dramatic device commonly used in cartoons that grabs the viewer's attention when it is used in an appropriate situation. For example, a surprised feeling would be presented as a big face, or a sad feeling would be presented by darkening the character's surroundings. This technique is very effective in conveying the sender's feelings and emotions.

IX. CONCLUSION

This paper proposes a method of generating animation from text messages without an authoring process. This method utilizes information available in the consumer's ubiquitous computing environment to obtain the information needed to create animations that are not included in the original e-mail message.

The results of the preliminary evaluation show that a user needs around two minutes to create a one-minute animation using the PMP system, a creation efficiency improvement of 3.6 times over that of TVML mini, which shows that this method is far superior to the prior method.

ACKNOWLEDGMENT

The authors would like to express our gratitude for valuable comments and stimulating discussions with researchers at the Digital Eizou Common Specifications Development Project [13], especially Professor Hiroshi Yasuda and Dr. Terumasa Aoki.

REFERENCES

[1] Irene Albrecht, Jorg Haber, Kolja Kahler, Marc Schroder, Hans-Peter Seidel, ""May I talk to you? :-)" - facial animation from text", Proc.of the 10th Pacific Conference on Computer Graphics and Applications (PG'02), pp.77-86, Oct. 2002.

[2] John L. Austin, "How to Do Things with Words", Ed. J. O. Urmson, Cambridge: Harvard UP, 1962.

[3] Paul Ekman, "Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life, Henry Holt and Company, 2003.

[4] Charles J. Fillmore, "The Case for the Case, in Universals in Linguistic Theory", E. Bach and R. Harms (Eds), Holt, Rinehart and Winston, New York, 1968.

[5] Masaki Hayashi, Mamoru Douke, and Narichika Hamaguchi, "Automatic TV Program Production with APEs", The Second International Conference on Creating, Connecting and Collaborating through Computing Conference Proceedings, pp.18-23, Jan. 2004.

[6] Jan Krikke, "Graphics Applications over the Wireless Web: Japan Sets the Pace", IEEE Computer Graphics and Applications, vol. 21, no. 3, pp. 9-15, May/Jun, 2001.

[7] Kinya Maruyama, "Intermodal sensibility interaction appeared in auditory and vision perception", Japanese Journal of Psychology, 1964. (in Japanese)

[8] Ministry of Internal Affairs and Communications, Japan, "ICT Use by Individuals", Information and Communications in Japan - Stirrings of u-Japan - White Paper 2005", pp.12-20, Jun. 2006.

[9] Seiya Miyazaki, Jinhong Shen, Takafumi Yuki, Mitsuru Kaneko, Terumasa Aoki and Hiroshi Yasuda, "Directorial Planning Engine for Automatic Cinematography", SoftCOM2003, Italy, 3059, pp661-665, Oct. 2003.

[10] Per Persson, "ExMS: an Animated and Avatar-based Messaging System for Expressive peer Communication", Proceedings of the International ACM SIGGROUP conference on Supporting group work, pp.31-39, Nov. 2003.

[11] Ryuuki Sakamoto, Yasuyuki Sumi and Keiko Nakao, "ComicDiary: Representing Individual Experiences in Comics Style", The 25th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001), Sketches & Applications, pp.158, Aug. 2001.

[12] The Institute of Socio-information and Communication Studies, "A Research Report of Japanese Information Behavior 2000", University of Tokyo Press, 2001. (in Japanese)

[13] Digital Eizou Common Specifications Development Project, http://www.mpeg.rcast.u-tokyo.ac.jp/DECSDP/

[14] FunMail Animated Greeting Cards, http://www.funmail.com/products_agc.aspx

[15] HER STORY, "Inquiry report on communication tools", http://www.herstory.co.jp/jisya/200403/20040303com_tool.html, Mar. 2004. (in Japanese)

[16] NTT DoCoMo, "What's Deco-mail", http://www.nttdocomo.co.jp/english/p_s/i/deco/about.html

[17] NTT DoCoMo, "List of pictgrams", http://www.nttdocomo.co.jp/english/p_s/i/tag/emoji/e2.html

[18] Vodafone, "Deru Moji 3D Picrogram Display", http://www.vodafone.jp/en/live/mail_function/deru_moji.html

[19] TVML Language Specifications, http://www.nhk.or.jp/strl/tvml/english/spec/index.html

[20] TVML mini, http://www.nhk.or.jp/strl/tvml/english/mini/

Koichi Emura is a Staff Engineer at the Network Development Center of Matsushita Electric Industrial Co., Ltd., Tokyo, Japan and also a Co-operative Research Fellow of Research Center for Advanced Science and Technology at the University of Tokyo. He received his B.E. degree in Electrical and Electronics Engineering from Sophia University, Tokyo Japan in 1991.

He has worked in Matsushita Electric Industrial in Tokyo since 1991 to develop (1) Video-On-Demand system based on MPEG-1 and MPEG-2 that achieves, for the first in the industry, fast forward and rewind playback functions over an ATM network, (2) Content Adaptation and Delivery system using the proposed MPEG-7 Multimedia Description Scheme, and (3) Personal Media Producer system.

Mr. Emura served as the co-chair of the ISO MPEG MPEG-7 Mobile Application Ad-hoc Group from July 2000 to July 2001. He is a member of IEEE, ACM, IEICE, IPSJ, and the SC29/WG11/MPEG-7 subcommittee in Information Technology Standards Commission of Japan. His research interests include content creation and delivery, context-aware content adaptation, mobile augmented reality, and social tagging and filtering.

Makoto Yasugi is a Staff Engineer at the Network Development Center of Matsushita Electric Industrial Co., Ltd. He received B.S. and M.S. degrees, in Informatics from Kyoto University in 1999, 2001, respectively.

He has worked in Matsushita Electric Industrial in Tokyo since 2001, to develop (1) image-processing software for mobile phones, (2) Personal Media Producer system. His research interests include computer graphics.

Toshiyuki Tanaka is a Staff Engineer at the Network Development Center of Matsushita Electric Industrial Co., Ltd.

He received B.S. and M.S. degrees, in Science of Engineering from Shizuoka University, Shizuoka, Japan, in 2000, 2002, respectively.

He has worked in Matsushita Electric Industrial in Tokyo since 2002, to develop (1) Media-Contents retrieval system and (2) Personal Media Producer system. His research interests include natural language processing and multimodal user interface.

Seiya Miyazaki is a Staff Engineer at the Network Development Center of Matsushita Electric Industrial Co., Ltd.

He received B.S., M.S. and Ph.D. degrees, in Electrical Engineering from the University of Tokyo, Tokyo, Japan, in 1999, 2001, 2004, respectively.

Dr. Miyazaki received presentation awards from IIITE in 2002, IVCNZ in 2003, and IIEEJ in 2003. His research interests include speech and image processing, computer graphics and user interfaces.

Sachiko Motoike is a Staff Engineer at the Network Development Center of Matsushita Electric Industrial Co., Ltd. She received a B.S. degree in Mathematical Engineering from the University of Tokyo, Tokyo, Japan, in 1984, and her research interests include natural language processing and artificial intelligence.