# Sign-Language Synthesis for Mobile Environments

Seiji Igi
Communications Research Lab.
2-2-2 Hikaridai, Soraku-gun
619-0289, Kyoto, Japan
igi@crl.go.jp

Mitsue Tamaru
EXA Corporation
580 Horikawa-cho, saiwai-ku
212-8555, Kawasaki, Japan
Mitsue-tamaru@exa-corp.co.jp

Yasunori Yamamoto
National Museum of Ethnology
10-1 Senri Expo Park
565-8511 Suita, Japan
yamty@minpaku.ac.jp

Megumu Ujitani
National Museum of Ethnology
10-1 Senri Expo Park
565-8511 Suita, Japan
ujitani@ minpaku.ac.jp

Shigeharu Sugita
National Museum of Ethnology
10-1 Senri Expo Park
565-8511 Suita, Japan
sugita@minpaku.ac.jp

## ABSTRACT

This paper describes the synthesis of sign-language animation for mobile environments. Sign language is synthesized by using either the motion-capture or motion-primitive method. An editing system can add facial expressions, mouth shapes and gestures to the sign-language CG animation. Sign-language animation is displayed on PDA screens to inform the user of his/her mobile environment.

## Keywords

Sign Language, Computer Graphics, Animation, Mobile Environment, PDA.

## 1. INTRODUCTION

Sign language is the most important communication tool for the hearing impaired [Sac89] and, currently, a number of services are available to assist them. With the dramatic growth of the Internet, the number of sign-language services is expected to increase, but these may be difficult to develop since sign language contains many visual and dynamic cues such as hand motions, facial expressions, and gestures.

The research described here was motivated by the desire to develop an animation system that provides information to the moving hearing-impaired by sign language. This paper describes how we developed both a method to synthesize sign language with facial expressions and gestures, and a practical system to create sign-language content. We also describe a system that detects the location of the PDA user and provides him/her with local information through sign-language animation.



***WSCG POSTERS proceedings***
*WSCG'2003, February 3-7, 2003, Plzen, Czech Republic.*


## 2. METHOD OF SYNTHESIZING SIGN LANGUAGE ANIMATION

### Related work

To date, few methods have been developed that can synthesize sign language by computer [Lee93] and these can be classified into two categories: those that divide a sign into its more basic elements, like phonemes in spoken language, and those that use sign-language motion data acquired from a motion capture system or from magnetic sensors [Sak96] [Kur96]. The second method generally utilizes the actual movements of actors signing a word as one unit, and it can generate more realistic animation than the first kind. However, the cost of acquiring data using the second method is high.

### Method of synthesizing sign language

We use the motion-capture method for Japanese Sign Language [Lu00] that is used to synthesize realistic sign-language animation, and we propose the motion-primitive method [Lu99] for new sign language.

Since the grammar of sign language is different from that of spoken language, synthesis is carried out as follows. Text is linguistically translated into signs, and then the translated signs, facial expressions, and mouth shapes are visually synthesized. Figure 1 is a block diagram showing this synthesis of sign-language animation. The two steps we used are shown in the figure. The actual motion data from the signers is captured and then sign-language animation is synthesized based on the actual motion data gathered in the first step. Engine by motion primitive can also make motion data. In the second step, in addition to the sign-language animation synthesized from motion data, some textured images are used to reproduce facial expressions and mouth movements.

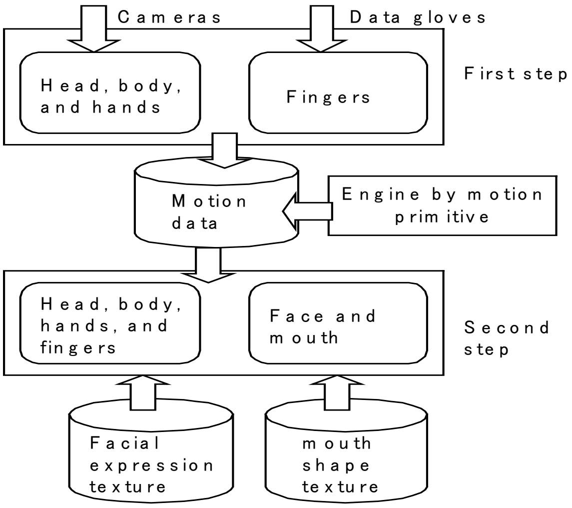



**Figure 1. Method of synthesizing sign language.**

## Acquiring Motion Data

To obtain the motion data of signs and gestures from a signer, we used an optical motion-capture system, which consisted of six cameras and six real-time video-image processors. The motion-capture system was able to capture the movements of his arms, body, and head, but not of his fingers. In sign-language animation, finger motion is also essential, so a pair of data-gloves was used to capture finger motion. Since the motion capture system and the data-glove system work independently, data from these two systems need to be synchronized. In the motion-capture process, motion data from thirty-nine joints including finger joints were obtained at a 60-frames-per-second data rate.

## Synthesis from Motion Primitive

A new method is developed to synthesize motion data based on intuitive motion primitives for new sign language words [Lu99]. The method uses some basic motion primitives (simple movements) to encode the movements of signs. The sign movement is encoded by using five factors; start and end position, motion primitive, hand shape, relationship between the two hands, and frame number.

## Synthesizing Sign Language and Gestures

Assigning motion data to the corresponding joints of a model generates sign-language words and gesture animation. Rotating any joint of the model animates signs and gestures according to the motion data. A sign-language sentence is synthesized by connecting the required words using linear interpolation. Table 1 lists frequent gesture data from a signer.

| Gestures shown in sign-language |
|---|
| *Nod head* |
| *Shake head* |
| *Nod head up and down quickly* |
| *Incline head to the side* |
| *Bow head while making eye contact* |
| *Point index finger* |

**Table 1. Frequently used gestures of signer.**

## Synthesizing Facial Expressions and Mouth Movements

In addition to hand and body motion, facial expressions also play an important role in sign-language conversations. A recent experiment showed that the hearing-impaired monitor the face or mouth of the signer while reading sign language. This finding supports our belief that facial expressions and mouth movements are important factors in synthesizing sign-language animation.

We collected typical facial expressions frequently appearing in sign-language communication by analyzing sign-language videos and interviews with hearing-impaired persons. These facial expressions are used mainly to express emotion or modify hand movements. In our system, these facial expressions are inserted into the sentence animation to supplement hand movements. Figure 2 has some example facial expressions of a signer, and Figure 3 indicates textures categorized into 19 types.

The mouth shapes are synthesized based on Japanese alphabet rules. Uezono et al. [Uez99] proposed a method to simulate the mouth movements in sign-language animation using five Japanese vowel sounds corresponding to the *kana* for “A”, “I”, “U”, “E”,”O”, the nasal articulation “N”, and the affricative articulations “PA”, “PI”, “PU”, “PE”, and “PO”. Japanese *kana* labels are added as the definitions of each signed word, then the vowels and articulations included in these labels are used to automatically determine the mouth shape.

**Figure 2. Typical facial expressions appearing frequently in sign language. The left is a “s*mile with mouth closed*”, and the right is a “*bulging cheek*”.**

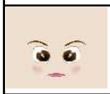
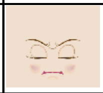
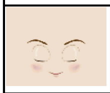
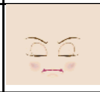
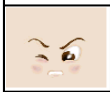
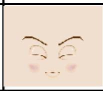
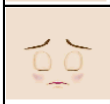
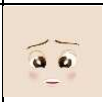
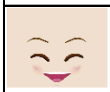
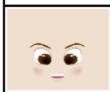
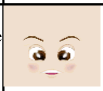
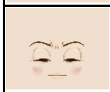
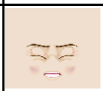
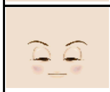
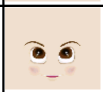
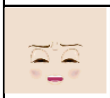
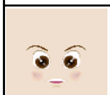
| Texture | Facial Expression & Explanation | Texture | Facial Expression & Explanation |
|---|---|---|---|
|  | *Normal Expression*<br>Usual face |  | *Emphatic Expression 1*<br>To express a large amount of something, big size, etc.. |
|  | *Impassive Expression*<br>To express unconcerned air. |  | *Emphatic Expression 2*<br>To express the meaning of doing best. |
|  | *Painful Expression*<br>To express pain. |  | *Emphatic Expression 3*<br>To express an adjective phrase like really, truly. |
|  | *Unhappy Expression*<br>To express unhappiness. |  | Knitted Brows<br>To modify words like seasoning, colors, etc. meaning "strong" or "dark". |
|  | *Smile*<br>Smiling. |  | *Expression of Failure*<br>To express something that is unsuccessful, not good. |
|  | *Strongly Drawn in Chin*<br>To express an interrogative sentence with surprise. |  | *Drawn up Eyebrows*<br>To express positive feelings, a clear sky, etc.. |
|  | *Narrow Eyes 1*<br>To express dissatisfaction. |  | *Emphasis on the Corner of eye*<br>To express strong feeling. |
|  | *Narrow Eyes 2*<br>To modify words like seasoning, colors, etc. meaning "light". |  | *Interrogative Expression*<br>To express an interrogative sentence. |
|  | *Embarrassed Expression*<br>To express troubling situations such as rainy days. |  | *Happy Expression*<br>To express happiness. |
|  | *Lightly Drawn in Chin*<br>To express the past tense. | | |

**Figure 3. Textures of facial expressions appearing frequently in sign language.**


**Figure 4. Sign language animation displayed on PDA.**

## 3. DISPLAY OF SIGN-LANGUAGE ANIMATION ON PDA

Hearing-impaired people have problems in communicating with their non-handicapped peers. Especially, it is remarkable in the information acquisition after it went out. Therefore, transmitting sign-language-animation information to their PDAs is very useful for them. The PDA is used, because it can be carried easily in mobile environment in spite of small screen size. Figure 4 shows animation displayed on a PDA.

## 4. PROVISION OF INFORMATION CORRESPONDS TO POSITION

We selected an application scene that transmitted content on a museum to a moving hearing-impaired user. Figure 5 shows how the location of the visitor was detected and how information corresponding to the visitor's position was transmitted. The system consists of three systems, namely (1) the information-providing server, (2) the position-detecting server, and (3) the wireless LAN system. The position of a visitor carrying a PDA with an infrared tag transmitting peculiar ID is detected by an infrared sensor installed near an exhibit and the position is sent to the position-detecting server. The PDA receives information concerning the exhibit from the information-providing server. The system provides information on the exhibit that the visitor pays attention to with sign-language animation in company with moving visitor.

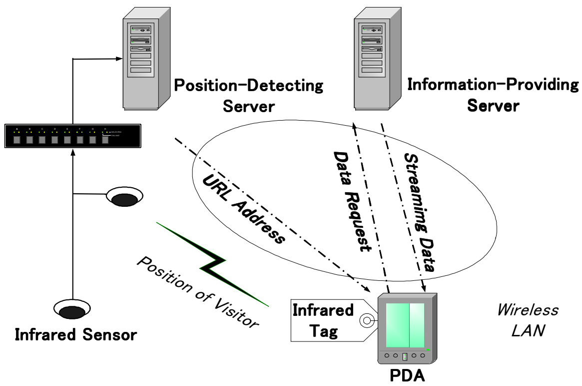



**Figure 5. Method used to detect the location of the visitor and how to transmit information corresponding to the visitor's position.**

## 5. EVALUATION

We evaluated sign-language animation for PDAs using seven hearing-impaired subjects who used sign language in their daily lives. Table 2 summarizes a profile of the subjects. Four are male and three are female. Five persons are in their twenties and the other two are 39 and 57. Almost none of them had ever used PDAs, but almost all used personal computers. The results of the evaluation are summarized in Table 3.

| subjects | age | male /female | time missing hearing | experience of E-mail | experience of PDA |
|---|---|---|---|---|---|
| A | 28 | male | from birth | use | not use |
| B | 20 | female | from birth | use | sometimes |
| C | 25 | male | 2 yrs. of age | not use | not use |
| D | 57 | male | not clear | not use | not use |
| E | 39 | female | 3 yrs of age | use | not use |
| F | 26 | male | from birth | use | not use |
| G | 29 | female | 2 yrs. of age | use | not use |

**Table 2. Profile of the subjects**

| subjects | word by motion capture | word by motion primitive | sentence |
|---|---|---|---|
| A | 91.7 | 80 | 81.8 |
| B | 95 | 63.3 | 100 |
| C | 98 | 63.3 | 54.6 |
| D | 75 | 40 | 54.6 |
| E | 90 | 50 | 45.5 |
| F | 91.7 | 70 | 63.6 |
| G | 100 | 80 | 100 |
| average | 91.6 | 63.8 | 71.4 |

**Table 3. Recognition rate (%)**

## Evaluation Experiment for the Word

Evaluation was carried out for 60 words by the motion-capture and 30 words by the motion-primitive method as shown in Table 3. The evaluation tested whether the subjects could correctly understand the signed words after looking at the sign on the PDA three times. The recognition rate equaled the number of correct words divided by the total number of words times 100. Average recognition for the motion-capture and motion-primitive methods was 91.6% and 63.8%, respectively.

## Evaluation Experiment for Sentences

Evaluation was done on 11 sentences that explained the exhibits at the museum. The sentence with the most words contained 17 and the sentence with the least words contained 6 words. The average number of words in one sentence was 12. The results of evaluation are also summarized in Table 3. We used four criteria; correct answer (meaning of sentence is perfectly understood), almost correct answer (meaning one to three words were mistaken, but the general meaning was grasped), half understanding (half the sentence was comprehended), and wrong answer (or unwritten). The evaluation experiment shows that 44.2% and 71.4% of the sentences are correctly answered and almost correctly answered, respectively.

## Considerations

For words, the recognition rate by the motion-primitive method was lower than that by motion capture. Although it has been reported that the addition of a mouth shape improves the recognition rate [Uez99], the subjects could not lip read due to the small screen of the PDA. In sentences, the average recognition rate was about 70% in spite of some long examples and difficult content that explained exhibits at the museum. The oldest subject said that it was difficult to recognize the animation, as it was too small. Appropriate measures need to be taken to correct this problem in the future.

# 6. CONCLUSION

We developed a prototype system that provided information to moving hearing-impaired persons through sign-language animation displayed on a PDA screen. The animation was synthesized by using both motion-capture and motion-primitive methods. Facial expressions, the shape of the mouth, and gestures were added to the animation. Our evaluation revealed that the subjects could understand about 70% of the sentences used in the experiment. Future problems remaining to be solved are how to improve the display of mouth shapes and facial expressions.

# 7. ACKNOWLEDGEMENTS

We thank all the hearing-impaired subjects for their evaluation of the sign-language animation.

# 8. REFERENCES

[Kur96] Kuroda, T., Sato, K., and Chihara, K. Reconstruction of Signer's Actions in a VR Telecommunication System for Sign Language, Proc. VSMM'96, pp.429-432, 1996.

[Lee93] Lee, J., Kunii, T.L. Computer Animated Visual Translation From Natural Language to Sign Language, J. Visualization and Computer Animation, Vol. 4, pp. 63-78, 1993.

[Lu99] Lu, S., Sakato, H., Uezono, T., and Igi, S. A New Method to Synthesize Japanese Sign Language based on Intuitive Motion Primitives, Proc. HCI99, pp. 441-445, 1999.

[Lu00] Lu, S., Igi, S., and Watanabe, R. Synthesizing Sign-Language Animation for Communication, Proc. SCI2000, Vol. 2, No. 3, pp. 239-246, 2000.

[Sac89] Sacks, O. Seeing Voices – A Journey into the World of the Deaf, the University of California Press, 1989.

[Sak96] Sakiyama, T., Oohira, E., Sagawa, H., Ooki, M. and Ikeda, H. A Generation Method for Real-Time Sign Language Animation, Trans. on IEICE, Vol. J79-D-II, No. 2, pp. 182-190, 1996.

[Uez99] Uezono, T., Sakato, H., Lu, S., and Igi, S. Development and Experiment of Synthesizing System for Japanese sign language animation based on motion-primitives and motion-shapes, T. R. IEICE, HIP98-46, pp. 49-56, 1999.