

Generation of modifier representation in sign animation

M Murakami¹, T Kuroda², Y Manabe³ and K Chihara⁴

^{1,3,4}Graduate School of Information Science, Nara Institute of Science and Technology,
8916-5 Takayama, Ikoma, Nara 630-0101, Japan

²Kyoto University Hospital, Department of Medical Informatics,
54 Kawahara, Shogoin, Sakyo, Kyoto 606-8507, Japan

^{1,3,4}{mikako-m, manabe, chihara}@is.aist-nara.ac.jp, ²Tomohiro.Kuroda@kuhp.kyoto-u.ac.jp

^{1,3,4}<http://chihara.aist-nara.ac.jp>, ²<http://www.kuhp.kyoto-u.ac.jp/Official/medinfo>

ABSTRACT

Most of the Japanese-to-JSL (Japanese Sign Language) translators utilize CG (Computer Graphics) animation to present translation results. However, foregoing systems cannot produce natural and legible sign expressions for the Deaf, because they are based on word-for-word translation. Our research aims to produce animation including proper modifier representation, which can express nuances of original sentences. This paper proposes a method to include modifier representation into animations automatically.

1. INTRODUCTION

Sign Language is a native language for the Deaf. To establish smooth communication between the Deaf and the hearing people, many researches have been developing systems, which can translate Japanese and JSL mutually (Sagawa et al, 1996; Kato et al, 1997).

Foregoing Japanese-to-JSL translation systems mostly use 3D CG animation, called sign animation, in order to present Sign Language. Most of sign animations are concatenations of short animation corresponding to given word. Although transitions between short animations are smoothly interpolated, conventional sign animations cannot express the changes of motion duration and speed. It is indispensable to include these elements into sign animation; otherwise, sign animation comes quite dull expressions without any modulation for the Deaf, like a robot voice for the hearing people. Kurokawa et al (1997), Igi et al (2001) are developing translation system, which enables editors to add modulation on sign animations manually. Automatic generation method of adding modulation into sign animation is important factor for establishing automatic Japanese-to-JSL translation system in the near future.

This paper takes up hand movement mainly, and proposes the method to include modifier representation into sign animation by means of determining patterns of motion.

2. PROPOSED METHOD TO INCLUDE MODIFIER REPRESENTATION

2.1 Modifier Representation

JSL is expressed via many channels such as hand motion, finger alphabet, facial expression, mouth movement and nod et cetera. All elements have important meaning to convey what signer's want to say. A sign word can be expressed by totally different hand motion and expressions to give some modulations.

Two types of modulation; emotional expression and modifier representation can be found on JSL. Emotional expression is given by signer's feeling. For example, when people get angry, they speak with asperity. Such expressions don't appear on written sentence. To the contrary, modifier representation shows qualifiers to modulate meaning of other words. Thus, modifier representation actually appears on written sentence as a word. However, foregoing studies consider modulation as emotional expression, and modifier

representation isn't being taken into consideration with sign animation. This seems the one reason why sign animations are not natural and legible. This paper focuses on modifier representation and proposes the method to include it into sign animation.

Oblique arrow in Fig. 1 shows typical process flow of foregoing system that translates into JSL from Japanese. Firstly, spoken Japanese sentence is converted into a train of words by morphological analysis. For example, the system translates the sentence 'this book is very cheap' into concatenation of four words 'this', 'book', 'very' and 'cheap'. Secondly, each Japanese word is replaced with corresponding sign word one by one, and synthesizes them smoothly.

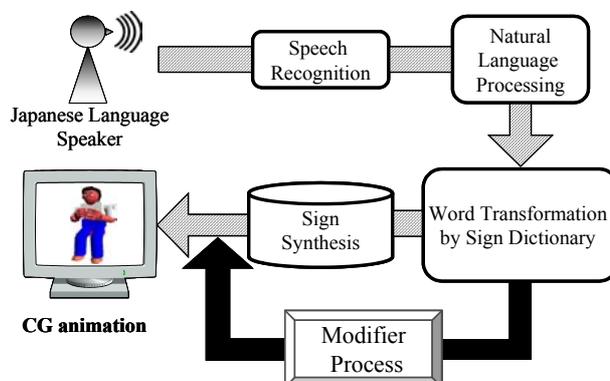


Figure 1. Structure of Translation System Including Modifier Representation.

However, as this method is a verbatim translation, cannot produce natural sign animation. In former example, the Deaf don't represent the word 'very'. They express 'cheap' with exaggerated gestures to convey the meaning 'very'. Thus, the Deaf add adverbial meanings by changing expression. This research adds modifier process shown by black arrow. In case where the qualifier appears on the sentence, adjectival movements change based on the degree of emphasis.

2.2 Classifications of Modifier Representation

To standardize the labels of JSL and to handle signs easily on computers, the romanized notational system for JSL is proposed. This system is named Signdex and recorded 545 words as version 1 (Kanda et al, 1997). We picked adjectives recorded on Signdex V.1 as object. Changes of the motion are examined, in case these adjectives are modified by degree adverb like 'very'. We are concerned here only with words that move simply like linear motion, to extract modifier representation more easily. This research ignores the modifier representation given by facial expressions.

Table 1. Classification example of each adjective emphasized by degree adverb.

Signdex	Sample Sentence	"Range"	"Speed"	"Frequency"
aSAI (shallow)	This pond is shallow.	○		
uTSUKUSHII (beautiful)	This flower is beautiful.	○		
oSOI (slow)	That car is slow.		○	
oMOI (heavy)	This stone is heavy.	○		
oMOSHIROI (interesting)	This program is interesting.			○
Tanoshii (pleasant)	Excursions are pleasant.			○
tOOI (far)	Your house is far.	○		
nAGAI (long)	This bar is long.	○		
hAYAI (fast)	That car is fast.		○	
fUKAI (deep)	This pond is deep.	○		
yASUI (cheap)	This book is cheap.	○		

It becomes clear that native signers express degree adverb by varying movement's, "Range", "Speed" and "Frequency" of adjectives. For example, 'very far' is expressed by extending movement's "Range" of 'far'. The proposed method introduces several grammatical rules and varies three features in sign animation. Where grammatical rule is to develop a certain pattern about degree adverb. For instance, it assumes that 'very' doubles "Range" of 'far'. In case 'very far' appears on sentence, the system produces animation of 'very far' using animation 'far' with rule "Range" and parameter '2'. Thus, definition of rule at each set of

degree adverb and adjective makes dictionary small, because there is no necessity to register animation corresponding with all emphasized expression. The example of a classification result is shown in Table 1. Even the same words in Japanese, sometimes there are some different expressions in JSL. In that case, we select the expressions, which are recorded in Signdex V.1.

2.3 Analysis of data

Original sentences (which include only adjective like sample sentences shown in Table 1) and emphasized sentences (which are modified by ‘very’) are acquired by motion capture system, S-TEL. This system obtains whole upper body motion using three sensor units; two gloves and a headset. The gloves obtain hand posture and 6D trackers attached to wrists and head obtain positions and orientations of wrists and head (Kuroda et al, 1998).

An informant wears sensors and expresses sentence given by Japanese text. Number of informants is one. In addition, he isn’t told that the experiment pays attention to the movements of hands mainly, and expresses JSL sentences as usual. Fig. 2 shows the example of acquired animation.

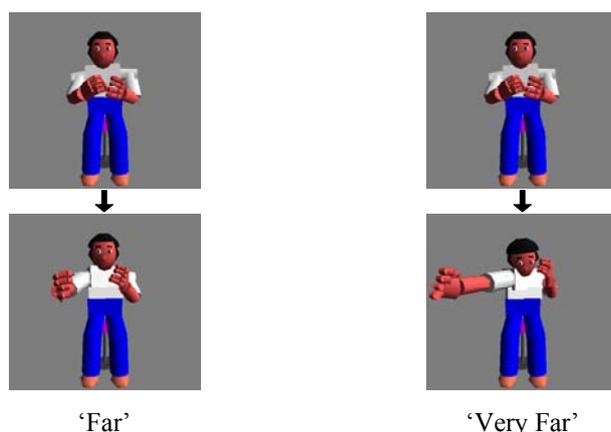


Figure 2. Example of Captured Sign Animation.

We cut sign motion corresponding target word from obtained data and analyze it. In this report, we analyze “Range” and “Speed” among the three features.

“Range” is analyzed by PCA (Principal Component Analysis) to examine the amount of change of the data by emphasis. PCA is widely used in signal processing, statistics and neural computing to reduce the dimension of the data. Analysis reveals the direction depending on the movement of wrists and head. Firstly, position information about both wrists and head are analyzed separately, and their trajectories are detected. Secondly, original and emphasized words are mapped on the axis of the first primary component of the original word. Comparing original word with emphasized word, the ratio of the distance is detected.

Table 2. Ratio of “Range” and “Speed”.

Adjective	“Range”			“Speed”
	Right wrist	Left wrist	Head	
shallow → very shallow	5.35	2.25	1.45	0.89
beautiful → very beautiful	1.83	3.49	2.38	1.00
slow → very slow	1.00	1.01	0.57	1.93
heavy → very heavy	1.19	1.40	2.94	1.88
interesting → very interesting	1.06	1.24	8.12	0.78
pleasant → very pleasant	1.86	1.89	9.37	1.00
far → very far	1.75	2.06	6.42	1.00
long → very long	2.16	2.22	1.53	1.68
fast → very fast	0.75	2.03	2.33	0.41
deep → very deep	1.02	0.68	1.16	1.57
cheap → very cheap	2.53	1.30	1.54	1.83

(Original word is made 1)

“Speed” is compared number of frames between original words and emphasized one. Each parameter to emphasize original word using degree adverb ‘very’ are shown in Table 2. The results differ from pre-classifications. Although certain words are classified into only “Range” in forecast, it become clear that these words vary number of frames actually. The same is true of “Speed”.

2.4 Producing of Animation

Using of the obtained ratio, sign animation including modifier representation is generated. To emphasize “Range”, coordinate value of original word is multiplied by each ratio. The first frame of original word is made a standard point (initial position of movement) of emphasized word. To emphasize “Speed”, number of frames is increased or decreased. When speed of motion becomes slow, presenting the same animation at a regular interval increased the number of a frame. To the contrary, when speed becomes fast, cutting out some images at a regular interval reduced frames. Fig. 3 shows the produced sign animation including modifier representations by varying “Range” and “Speed”.

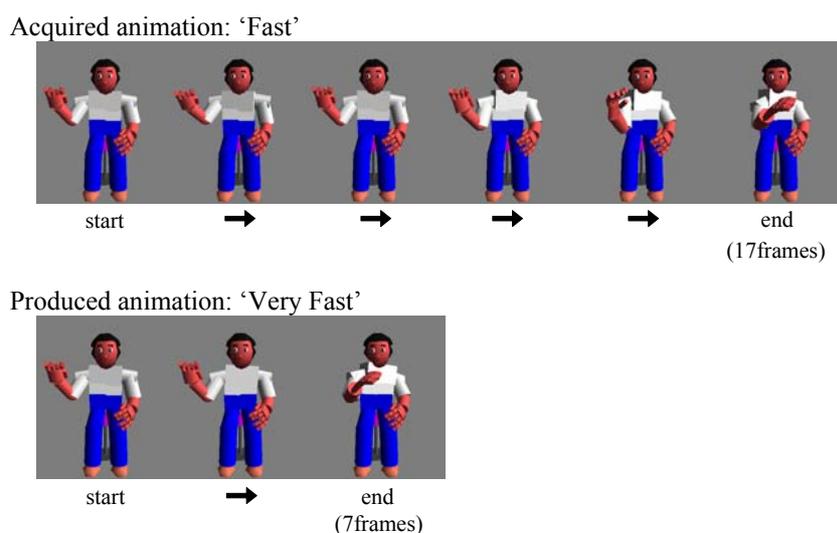


Figure 3. *Generated Sign Animation emphasizing “Range” and “Speed”.*

3. EVALUATION VIA READING TEST

3.1 Legibility of Sign Animation

The experiment aims to confirm that modifier representation is easier to understand than verbatim translation. 23 test subjects participate in this experiment; 1 of them is the Deaf, 4 of them are hearing impaired midway and others is the hearing people who has 1 to 30-years experience in JSL. 18 subjects among 23 are a first time to watch sign animation.

Two kinds of sign animations, one is word-for-word translated animation and the other is produced by the proposed method are presented.

[Steps]

1. Japanese sentence is shown by text.
2. Word-for-word translated animation (animation A) is presented.
3. Animation including modifier representation (animation B) is represented.
4. Subjects evaluate legibility of two animations.

There are five selections of answer on questionnaire, that is, “A is much better”, “A is better”, “same”, “B is better” and “B is much better”, and scored from 0 (“A is much better”) to 4 (“B is much better”), and the subjects tick off one of them. Score 0 means that animation A is more legible than animation B, and score 4 means that animation B is more legible. As a result of average, it become clear that animation B is more natural and legible for subjects except for one case, which regarded as the same legibility.

This seems attributed to the factor that the difference does not change according to adjective. For example, a sentence ‘This flower is very beautiful’ got score 2. This result means that there is no difference between two animations. Hereafter, it needs to check about more adjectives.

3.2 Various Modifier Representations

This experiment confirms that proposed method can generate various expressions by varying only several parameters. According to observation in Sign Language Club, the Deaf uses many degree adverbs such as ‘very’, ‘extremely’, ‘a little’, ‘slightly’, and so on. Varying parameters produces some modifier representations corresponding to these degree adverbs. Temporarily, order of degree is defined following.

slightly < a little < (original) < very < extremely

Since Japanese adverbs are quite vague and impossible to order, subjects are asked to reorder the adverbs on his/her own understandings during the test.

Table 3 shows the relations between emphasis level and parameter to give sign animation. Here, R is the ratio obtained by previous experiment explained in 3.1.

Table 3. Parameter of Each Degree Adverb.

	slightly	a little	(original)	very	extremely
Multiplication Parameters	$1/R^2$	$1/R$	1	R	R^2

Four kinds of modifier representation, which have different degree, were produced about each adjective. Three expressions are selected at random from 5 animations including original one. These animations are represented to the same subjects. Fig. 4 shows an example of selected animation.

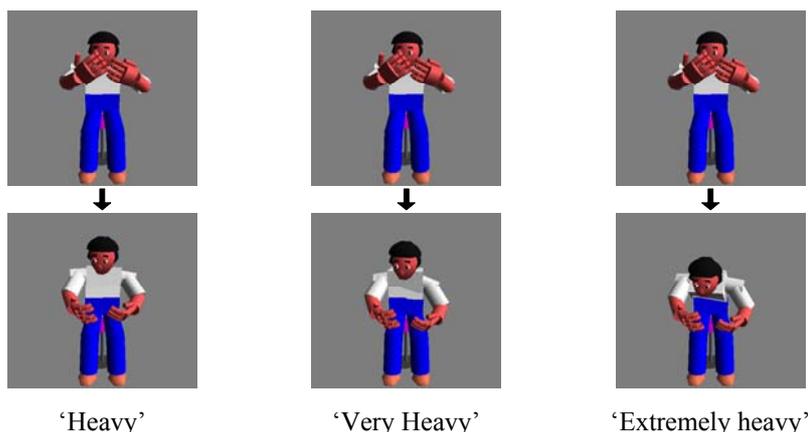


Figure 4. Automatic Generation of Modifier Representation.

[Steps]

1. Three sign animations were represented concurrently to subjects.
2. Subjects connect a degree adverb and each animation with lines in questionnaire (Question 1).
3. After representing all animations, degree adverb is sequenced by subjects (Question 2).

Using a result of question 2, degree of adverb is averaged respectively. Order differs greatly in individuals and there are six patterns.

Experimental results clear that modifier representation is distinguished nicely. Thus, it is possible to produce many expressions by varying some parameters. The subjects couldn't distinguish the degree of word ‘fast’. We carried out measurement of the word ‘fast’, ‘very fast’ and ‘extremely fast’ to find the reason of this result. As a consequence, it become clear that modifier representation of ‘fast’ doesn't change gradually. ‘Very fast’ is expressed by smaller range and fewer frames than ‘fast’. And ‘extremely fast’ is expressed by smaller range and fewer frames than ‘very fast’ in proposed method. Whereas, ‘extremely fast’ is expressed by bigger range and fewer frames than ‘very fast’ in JSL. This is a reason why ‘fast’ cannot be distinguished correctly. It becomes clear that parameters of emphasis on “Range” and “Speed” needed to be changed separately to express various degree adverbs in some words.

3.3 Discussion

Experimental results make clear that the proposed method can generate more natural and legible sign animations, which express nuances of original sentences. Therefore the proposed method achieved a part of objectives of this research. However, some of the produced animations are still not fully satisfactory. The reason is that generated animation cannot express non-linear changes of motion speed according to modulation. As a result of data analysis, it found that it is necessary to consider the amount of a change in a time target to represent the modulation more naturally. Fig. 5 shows an example of data analysis.

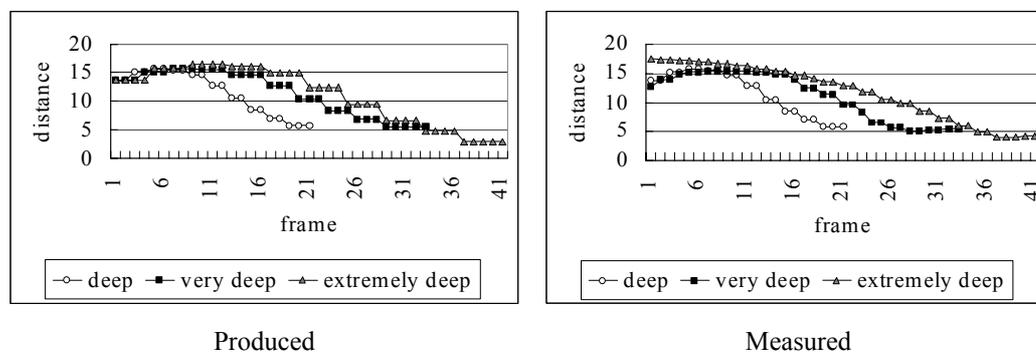


Figure 5. Analysis Result of Sign Animation 'Deep'.

In the next step, a method to express hourly change with a few parameters must be examined. Moreover, we must work on analysis about "Frequency" and consider direction information.

4. CONCLUSIONS

This paper proposes a method to include modifier representation into Japanese-JSL translation system, to improve legibility of sign animation. The analytical results indicate that modifier representation can be divided into some patterns. Based on an observation, modifier representation is classified into three features, "Range", "Speed" and "Frequency". The proposed method produces sign animation by varying these features and introducing several grammatical rules. Moreover, experiment results show that the proposed system can generate sign animation, which convey modifier representation. Additionally, the proposed system can generate various modifier representations.

Acknowledgements: We thank to Sign Language Club "Hureai" in Sakai city. This research is partly supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grants-in-Aid for Scientific Research, 14780638, 2002.

5. REFERENCES

- K Kanda, Y Nagashima, A Ichikawa and M Terauchi (1997), A Proposal of the Standard Labeling System for JSL: Signdex V.1, *HCI International '97*, p. 15.
- H Sagawa, M Ohki, T Sakiyama, E Oohira, H Ikeda and H Fujisawa (1996), Pattern Recognition and Synthesis for a Sign Language Translation System, *JVLC*, 7, 1, pp. 109–127.
- S Igi, R Watanabe and S Lu (2001), Synthesis and Editing Tool for Japanese Sign Language Animation, *IEICE*, J84-D-1, 6, pp. 987–995. (in Japanese)
- T Kuroda, K Sato and K Chihara (1998), S-TEL: An Avatar Based Sign Language Telecommunication System, *IJVR*, 3, 4, pp. 21–27.
- T Kurokawa, K Senba and S Kawano (1997), Synthesis of Facial Movement Images for Sign Language Animation and Their Effect, *HCI International '97*, pp. 1–5.
- Y Kato, K Kanda, Y Nagashima, A Ichikawa and T Kurokawa (1997), A present overview and future problems on Sign Language Engineering, *Human Interface N&R*, 12, 1, pp. 37–44. (in Japanese)