Japanese sign-language recognition based on gesture primitives using acceleration sensors and datagloves

Hideyuki Sawada, Takuto Notsu and Shuji Hashimoto

Department of Applied Physics, School of Science and Engineering, WASEDA University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169-8555, JAPAN

{sawa, taku, shuji}@shalab.phys.waseda.ac.jp

ABSTRACT

This paper proposes a Japanese sign-language recognition system using acceleration sensors, position sensors and datagloves, to understand human dynamic motions and finger geometry. The sensor integration method realized a robust gesture recognition comparing with a single sensor method. The sign-language recognition is done by referring to a Japanese sign-language database in which words are written as sequences of the gesture primitives. Two recognition algorithms which are the automata algorithm and the HMM are introduced and tested in the practical experiments.

1. INTRODUCTION

Gesture plays an important role not only as nonverbal media for emotional human communication [1][2], but also as a necessary method of communication among people who are hearing impaired as sign language. Sign language has been formed as a language for the verbal communication. Hearing impaired people sometimes need to have communications with able bodied people with the help of a sign-language interpreter. Since the number of interpreters is limited, they cannot always have enough communication when they demand. A computerized interpreter which understands the sign-language automatically is greatly needed as a communication tool in the welfare field. If a computerized system is able to understand the human gesticulation, it will generate a new solution not only as a machine interpreter which translates nonverbal expressions into verbal languages, but also as a flexible human-machine interface which reacts to nonverbal expressions and emotional feelings.

Many attempts to the realization of sign-language recognition have been reported so far [3]-[5]. Most of the approaches introduced in the existing researches can be classified into two categories: one employs wearable devices such as a dataglove and a position sensor, and the other uses an image processing technique. The former approach is suitable for the realtime recognition, since the finger geometry and the hand posture are directly obtained as three dimensional physical data. Constraints by wearing the device and wiring for the data transmission are, however, the problem. On the other hand, the latter approaches are actively studied these days, because the image processing technique lets the users be free from the constraints of the wearable equipment. Great deal of computational processing and the estimation of occluding objects, however, have to be solved by the computing algorithms. The common problem which remains in both approaches is how to extract the starting point of a gesture. Furthermore the dynamic characteristics of gestural motions are hard to be recognized in realtime because the physical data obtained in the both techniques has the positional dimension.

The arm motion and the finger geometry are considered to be the primitive information in understanding gesticulations. Motion can be measured directly as acceleration caused by the applying forces to the body[2][6]. In this study, acceleration sensors and a pair of datagloves are employed for the gesture acquisition. This paper presents a gesture recognition algorithm based on acceleration patterns caused by dynamic movements. The gesture patterns are extended to the classification as gesture primitives for the construction of Japanese sign-language (JSL) recognition system. Two algorithms which are the automata algorithm and the HMM (Hidden Markov Model) are introduced in the sign-language recognition. The HMM has better performance rather than the automata algorithm in the current study. JSL database is also constructed in which JSL words are described as the sequence of gesture primitives.

2. CHARACTERISTIC PARAMETERS FOR GESTURE PRIMITIVE RECOGNITION

2.1 Characteristic parameters Extracted by Acceleration sensor

An acceleration sensor (Nihon Kohden TA-513G 3D acceleration sensor) used in this study is small enough to be attached to any points of the human body. The size is 20x15x12[mm], and the weight is 12.5[g]. It can sense three dimensional accelerations by the piezo-electric devices which cause the change of voltage according to the amount of acceleration applied to the sensor. The sensitivity is about 5mV per 1G in the range between -25G and +25G. The acceleration data are amplified and fed to the computer through an A/D converter as 12 bit binary data.

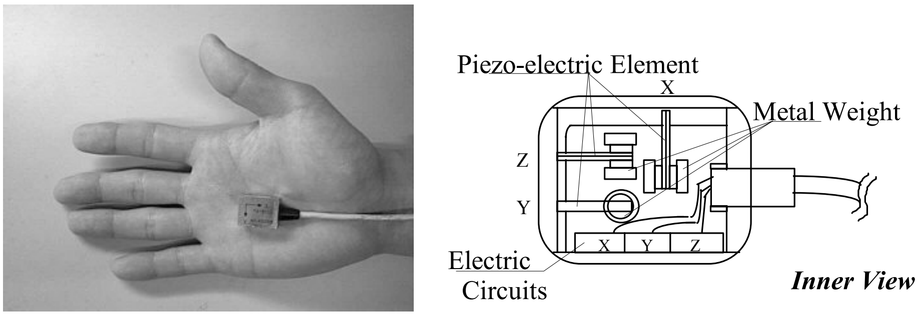


Figure 1. *3D acceleration sensor : its outer view(left) and schematic inner view(right)*

The three acceleration data $a_X(t)$, $a_Y(t)$ and $a_Z(t)$ are independently obtained in realtime, which correspond to the accelerations in *x*, *y* and *z* directions at time *t*, respectively. There is a large variation of acceleration patterns in the same kind of dynamic motion, as in the case of automatic recognition of handwritten characters. Furthermore, the sensor employed in this study does not have a good quantitative reproducibility. Consequently, a high recognition rate cannot be obtained by applying pattern matching method of acceleration vector series. The global features of the gesture motion must be extracted.

In this study, the motion features are extracted from the following three two-dimensional vectors in order to obtain intuitive characteristics of the motion:

$$\mathbf{A}_1(t)=(a_y(t), a_z(t)),\ \mathbf{A}_2(t)=(a_z(t), a_x(t)),\ \mathbf{A}_3(t)=(a_x(t), a_y(t)) \quad (1)$$

$\mathbf{A}_1(t)$, $\mathbf{A}_2(t)$, $\mathbf{A}_3(t)$ are the projection vectors of the acceleration on the *y-z*, *z-x* and *x-y* planes, respectively. In the study, the data acquisition frequency f_{in} was set to 30 Hz. Although this is not sufficient as a sampling frequency in the measurement of motion which includes rapid changes of the direction or the velocity, we are focusing on the realtime recognition with the minimum acceleration data. A succession of fifteen to thirty data set which accords with the duration of one shot gesture is used for gesture recognition. Eleven characteristic parameters shown in Table 1 are extracted from one sequence of each set of projection vectors on the *y-z*, *z-x* and *x-y* planes, for the realtime gesture recognition[2][6].

Table 1. *Characteristic Parameters of Acceleration Sensor.*

Pd :	Change of Force Given as Time Differences of Vectors.
Pg :	Rotating Direction Given as Vector Products $\mathbf{A}(t) * \mathbf{A}(t+1)$.
Pr :	Directional Characteristics Given by Maximum Vector.
Pa0 - 7 :	Characteristics of Directional Distribution of Vectors.

2.2 Characteristic parameters Obtained by Dataglove

The static shapes of hands also contribute to the meanings of the gestures. Therefore a pair of datagloves (Nissho Electrnics SuperGlove) are employed in this system, which allow the realtime analysis of three dimensional hand position and finger geometry. By distributing ten bend sensors along the five finger parts of the glove, the device outputs roughly proportional values to finger joint angles. Furthermore, a magnetic position sensor (Polhemus Sensor) fixed on the back side of the wrist area gives three dimensional positions *x*, *y* and *z* and orientations which are azimuth (θ_{az}), elevation (θ_{el}) and roll (θ_{rl}) attitudes. Fourteen characteristic parameters R_0 - R_{13} as shown in Table 2 are defined. R_3 - R_8 represent the sine and cosine of the azimuth, elevation and roll data. Cosine and sine functions eliminate the disconnection of the degree at $+\pi$ and $-\pi$, and work for the selection of the hand posture at a degree of $(n\pi)$ [n = -1, 0, 1] and $(n\pi)/2$ [n = -1,

+1], respectively. At first before the use of the system, a user inputs the limits of hand positions (up, down, right and left) and the finger curvature (grasping and straight), so that the physical factors of sensor data influenced by a user's physique are normalized. The characteristic parameters are obtained from the normalized data.

Table 2. *Characteristic Parameters of Dataglove.*

R_0 - R_2 :	3D Position of hand.
R_3, R_4 :	$sin(\theta az)$, $cos(\theta az)$.
R_5, R_6 :	$sin(\theta el)$, $cos(\theta el)$.
R_7, R_8 :	$sin(\theta rl)$, $cos(\theta rl)$.
R_9 - R_{13} :	Joint Angles of 5 Fingers.

2.3 Gesture Recognition Algorithm Based on Characteristic Parameters

The gesture recognition uses *P*'s and *R*'s. The realtime recognition is made by comparison with standard data acquired in the learning phase to make the system suited for the individual users. In order to recognize a gesture from the acceleration time series, the starting point of the gesture has to be detected first. Even if the gestures of the same kind are measured by the acceleration sensor, the force pattern may be observed differently depending on the individuals. In this study, the start of the gesture is identified from the magnitude of the acceleration, since the beginning of the motion is accompanied by the local peak of acceleration.

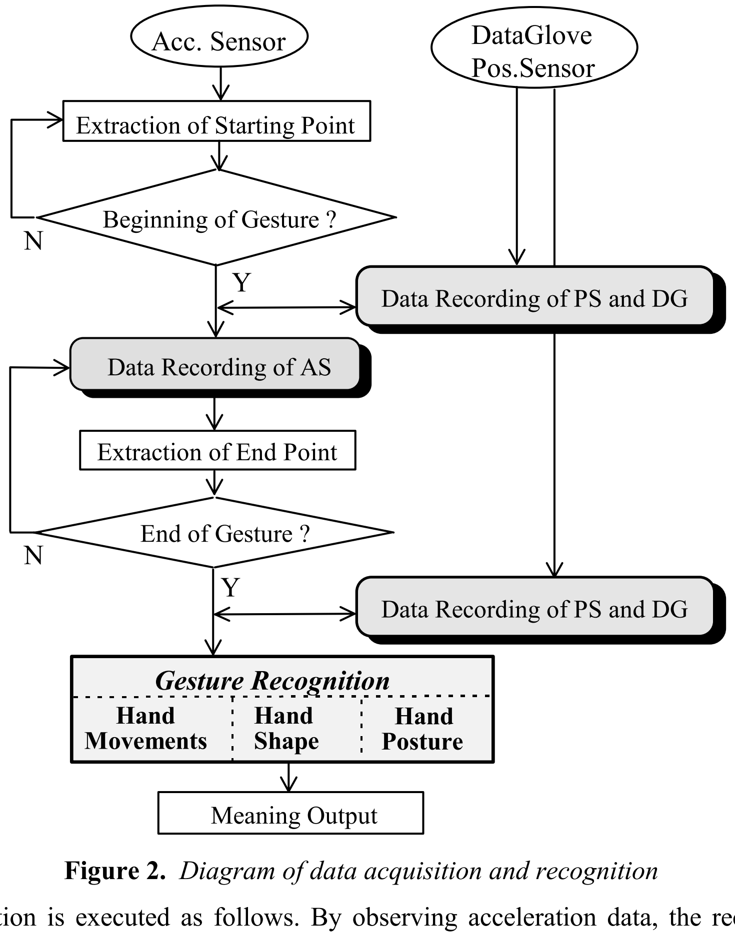


Figure 2. *Diagram of data acquisition and recognition*

The gesture recognition is executed as follows. By observing acceleration data, the recognition algorithm starts its action when acceleration value exceeds a pre-determined threshold. First, data of the datagloves and the position sensors are recorded in the computer memory, which is followed by the acceleration data recording. When the end of the gesture is detected by observing the acceleration data being recorded, data of the datagloves and the position sensors are recorded once again. Next whole recorded data are transformed into characteristic parameters which is handed to the recognition algorithm, then the meaning is determined by referring to the standard pattern data. The flow diagram of the data acquisition and the recognition is described in Figure 2.

In the learning phase, a user inputs gestures to be recognized M times each. Then the average E_α^G and the standard deviations μ_α^G of the characteristic parameters are calculated for each gesture G as shown below, and stored as the standard pattern data.

$$E_\alpha^G = \frac{1}{M}\sum_{i=1}^{M} V_\alpha^{g_i} \quad , \quad \mu_\alpha^G = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(V_\alpha^{g_i} - E_\alpha^G)^2} \qquad \begin{array}{l} V: \text{Parameter Values of Learning Pattern} \\ g_i: i\text{-th Sample of Gesture } g \\ \alpha: P\text{'s or } R\text{'s} \end{array} \tag{2}$$

In the recognition phase, the characteristic parameters V_α' are extracted, and the normalized distance e_G is calculated for each standard pattern data as below,

$$e_G = \sum_\alpha \varepsilon_\alpha^G = \sum_\alpha \frac{(V'_\alpha - E_\alpha^G)^2}{(\mu_\alpha^G)^2} \qquad V': \text{Parameter Values of Input Pattern} \tag{3}$$

Then the minimum e_G is selected as a candidate. In case it is smaller than a predetermined threshold value Th shown below, the result of the gesture recognition is confirmed.

$$Th = \min_{G \neq H} \left\{ \sum_\alpha \frac{(E_\alpha^H - E_\alpha^G)^2}{(\mu_\alpha^G)^2} \right\} \qquad \begin{array}{l} \text{for all } G, H \\ G, H: \text{Gesture } G, H \end{array} \tag{4}$$

Table 3 shows the results of recognition experiments for the 10 kinds of gestures which are:

vertical swing, horizontal swing, diagonal swing, sharp single swing,
clockwise rotation, counterclockwise rotation, star shape motion, triangular shape motion,
heart-shaped motion, pointing of direction, pause.

Gesture recognition was repeated 10 times each for two persons who are the particular individual for whom the standard pattern was constructed, and another individual. 100 percent gesture recognition is achieved for the individual for whom the standard pattern was constructed. And relatively good results were also obtained from the gestures by another individual. Mis-recognition arises for the “diagonal swing”, “counterclockwise rotation” and “triangular shape”, where the direction of motion/rotation is almost the same, as well as for “pointing of direction” and “sharp single swing”, where only the direction of the motion are different. Thus, it is obviously better that the user prepares the standard patterns of the individual before operating the system and enters the recognition phase. But the recognition is also possible to some extent even if the parameters are used to recognize the gestures of other individuals.

In the gestures used in our daily life, single operations such as vertical swing, horizontal swing, diagonal swing, rotation, sharp single swing, and pointing of direction, as well as their combinations, are commonly used. In the proposed system, “star shape”, “triangular shape” and “heart-shaped motion” can be recognized from the three dimensional acceleration patterns, even though their real-time recognition by image processing and other techniques seems quite difficult.

Table 3. *Experimental results of gesture recognition.*

	Own Gestures: Recognition Rate %	Another Person’s Gestures: Recognition Rate % (Mis-recognized as)
1. Vertical swing	100	100
2. Horizontal swing	100	100
3. Diagonal swing	100	70(Triangle)
4. Single swing	100	100
5. Clockwise rotation	100	100
6. Counterclockwise	100	80(Triangle)
7. Star shape	100	100
8. Triangular shape	100	80(Counter-clockwise)
9. Heart shape	100	100
10.Pointing of direction	100	70(Single swing)

3. SIGN-LANGUAGE RECOGNITION ALGORITHM BASED ON GESTURE PRIMITIVES

3.1 Gesture Primitives and JSL Database

JSL currently has almost 3,000 words, most of which are represented by the combination of simple motions. It means JSL words seem to be determined by observing both arm motion and hand figures. Strictly speaking, features from both motion and figures should be extracted simultaneously, changes of hand figure during arm motion, however, can be negligible, by considering the visual cognition process of human. In this study, arm motion and hand figure before and after the motion are highly paid attention to. So, gesture primitive blocks as

- Hand figure **-** Motion **-** Hand figure **-**.

are regarded as morpheme, the minimum unit of a gesture, which is used in the first step of JSL recognition procedure. Simple motion patterns are selected as 11 motion primitives and 14 hand figure primitives as listed in Figure 3 and Table 4, respectively, together with the hand posture patterns (**U**p, **D**own, **F**ront, **B**ack, **O**utside and **I**nside direction). We are constructing a JSL database in which sign-language are described by the repetition of primitive blocks.

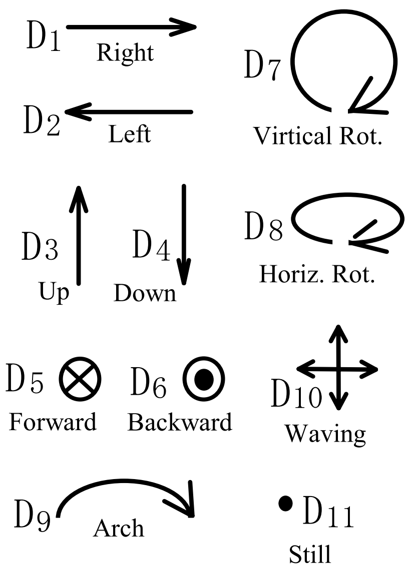


Figure 3. *Hand motion primitives*

Geometry Pattern	Finger States **s**traight/**b**end
F1	bbbbb
F2	sbbbb
F3	bsbbb
F4	bbsbb
F5	Bbbbs
F6	Ssbbb
F7	Bssbb
F8	Sbbbs
F9	Bsbbs
F10	Sssbb
F11	Bsssb
F12	Bbsss
F13	Bssss
F14	Sssss

Table 4. *Hand figure primitives*

	Right	*Left*	*: Supplements*
Happy			: Meaning
S	F3/B	F3/B	: Geometry/Posture
	D3	D4	: Motion
	F3/B	F3/B	: Geometry/Posture
	D4	D3	: Motion
E	F3/B	F3/B	: Geometry/Posture
North			: Meaning
S	F7/O	F7/O	: Geometry/Posture
	D4	D4	: Motion
	F7/O	F7/O	: Geometry/Posture
	D2	D1	: Motion
E	F7/O	F7/O	: Geometry/Posture

	Right	*Left*	*: Supplements*
You			: Meaning
S	F3/O	-	: Geometry/Posture
	D5	D11	: Motion
E	F3/O	-	: Geometry/Posture
Japanese ‘a’			: Finger Spelling
	F2/B	-	: Geometry/Posture
Japanese ‘ta’			: Finger Spelling
	F2/O	-	: Geometry/Posture
Japanese ‘ya’			: Finger Spelling
	F8/B	-	: Geometry/Posture

Figure 4. *Examples of Japanese sign-language database*

30 JSL words often used in the self introduction and in the daily life, and also Japanese 50 finger spellings are currently registered in the database and used for the realtime recognition. Three descriptions of JSL and three finger spellings are listed in Figure 4 as examples. The description of finger spellings consists of the geometry and the posture of hands, on the assumption that they are shown around the shoulder position. In case motion or hand figure has no effect on the meaning, it is marked as -.

3.2 Configuration of JSL Recognition System

The recognition system consists of two acceleration sensors, a pair of datagloves and computers. Two PCs are employed for the data acquisition from the sensors: one for the acceleration data, and the other for the datagloves and the position sensors on both hands. Every time motion and hand figure primitives are extracted in each PC, they are transmitted to SGI Indy computer, which are used for the execution of JSL recognition. By wearing a pair datagloves with the acceleration sensor and the position sensor fixed on the back side of the wrist area, user is able not only to execute the recognition of JSL but also register JSL words in the JSL database. The recognizable subjects in this system are the JSL words and finger spellings represented by the sequences of motion primitives and hand figure primitives as described in the JSL database.

Schematic diagram of the JSL recognition system is described in Figure 5. By observing acceleration data, the recognition algorithm starts its action when acceleration value exceeds a pre-determined threshold. First, hand figure primitives and hand postures of both hands are recorded on the gesture table prepared in SGI Indy, which is followed by the motion primitive extraction. Extracted motion primitives as shown in figure 3 are stored in the gesture table as a sequence in temporal order. When the changing point of motion is detected by observing the motion primitive sequence, a hand figure primitive and hand posture data are measured. By the repetition of the procedure, a primitive sequence as

Start: Hand figure - Motion - Hand figure - Motion - ..... - Motion - Hand figure : End

is obtained and stored in the gesture table until the end of gesture is detected by the acceleration data observation. Then the sequence is handed to the JSL recognition algorithm. For the finger spelling recognition, shoulder position of a user is assigned as the display spot. In case the gesture is presented at this area, it is regarded as a finger spelling and its recognition procedure is executed.

JSL recognition is made by comparing gesture primitive sequences stored in the gesture table with JSL database. For the comparison, two algorithms are currently examined. One is an automata algorithm, and the other is HMM (Hidden Markov Model) which is widely used for the time series pattern recognition.

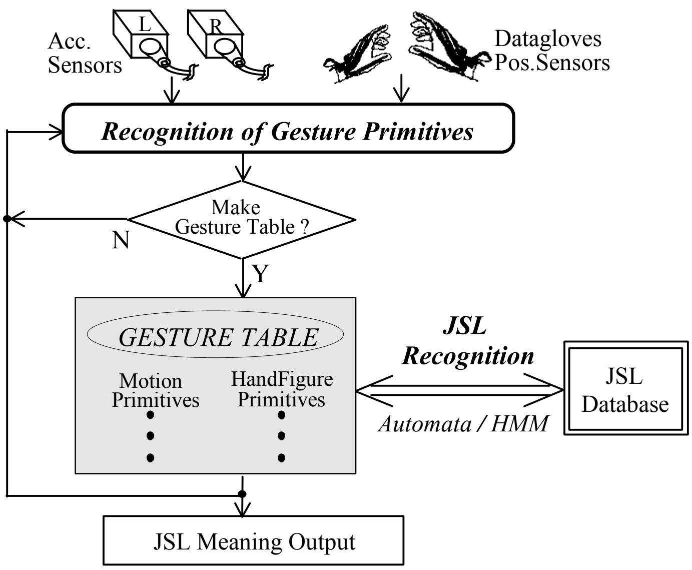
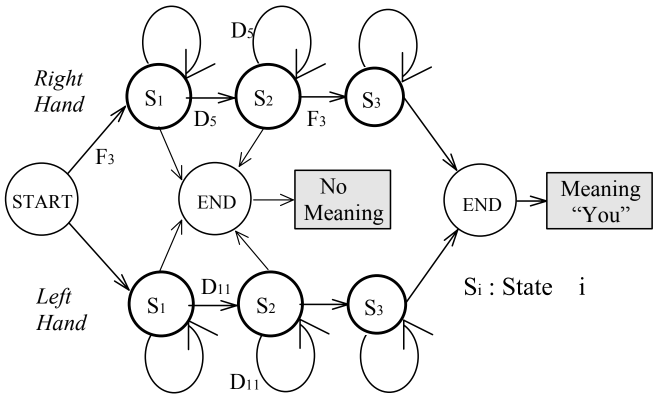


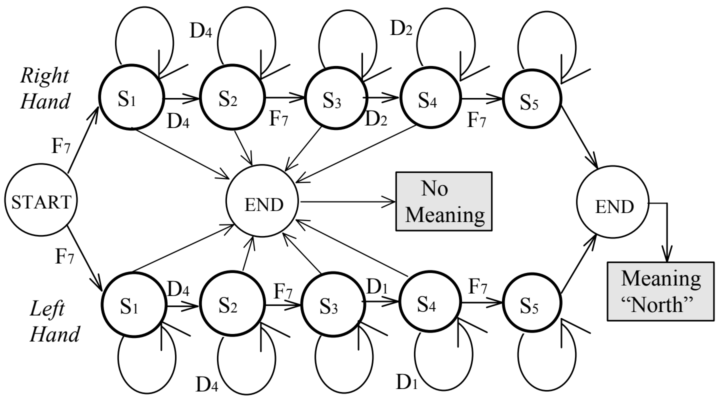
Figure 5. *Japanese sign-language recognition process*

3.3 JSL Recognition by Automata

Automaton is an information processing model consisting of finite number of states which transits deterministically according to an input[7]. Since the transition is determined by all the past input sequences, the final state of automaton is uniquely confirmed by the input sequence. By preparing necessary automata consisting of symbolic sequences of gesture primitives expressing JSL words, computerized recognition can be constructed. Two examples of automata used in the JSL recognition is shown in figure 6.



(a) Recognition of "You"

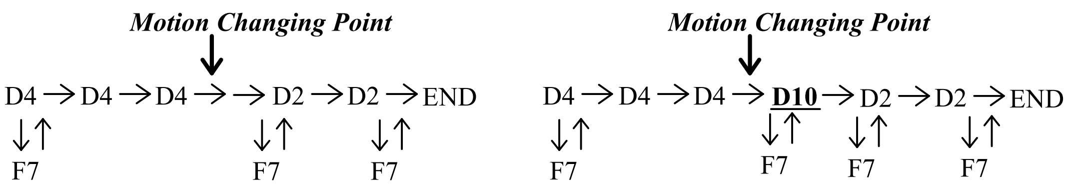


(b) Recognition of "North"

Figure 6. *Automata in sign-language recognition*

Table 5. *Experimental results of sign-language recognition by automata*

JSL words	Recognition Rate (%)	Misunderstood
You	100	-
Like	100	-
Good bye	93	No Meaning
Happy	87	No Meaning
North	83	No Meaning
Japanese 'a'	100	-
Japanese 'ta'	100	-



(a) Example of succeeded recognition *(b) Example of recognition failure at D10*

Figure 7. *Examples of success and failure in recognition of "North"*

Experimental results of recognition by the automata is listed in table 5. For the JSL words consisting of one shot motion such as “You” and “Like”, 100 % recognition rate is obtained. On the other hand, in the recognition of the word “North”, for example, whose motion pattern changes from D4 to D2, the resulting outputs are failed at the changing point in some recognition cases. In this case, the recognition results in “No meaning”. This is caused by the property of automata whose transition is made deterministically. An example of the recognition failure process is schematically described in figure 7.

3.4 JSL Recognition by HMM

Another algorithm employing a HMM is also examined in this study. The HMM is a chain with finite states connected by the probabilistic transitions[8]. Each state is characterized by two sets of probabilities: one is a transition probability among states, and the other is output probability distribution which defines the probability of emitting each symbol from the state. These probabilities can be determined from the learning pattern sets by using learning algorithms such as the Baum-Welch algorithm, the Forward-Backward algorithm and the Viterbi algorithm. Figure 8 shows an example of the HMM having 3 states with the 2 output symbols. Although the learning procedure of the probabilities are necessary for the HMM recognition algorithm, and the recognition ability depends on the learning pattern sets and learning procedure, recognition failures found in the automata algorithm as shown in figure 7 are expected to be avoided.

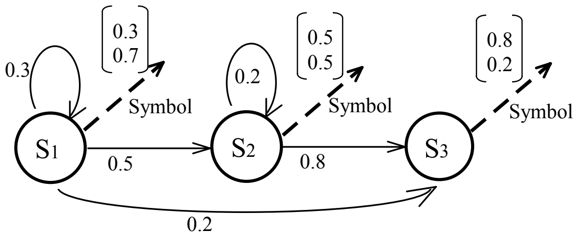


Figure 8. *Example of HMM*

HMM used for the JSL recognition is a left-to-right model having 3 states with a skip transition. And the Baum-Welch algorithm is adopted for the learning. Output symbols from each state are 24 symbols which consist of 10 motion primitives and 14 hand figure primitives. Table 6 shows several examples of experimental results. Comparing with the automata algorithm, JSL words with one shot motion tend to cause the misunderstood results in the HMM. The reason is that the mis-recognition of gesture primitives seems to be hard to be absorbed in the calculations of transition probabilities.

Average of 94 % recognition rates was obtained for the 30 JSL words stored in the database, on the other hand, automata algorithm resulted in the 89 % recognition rates. Although recognition ability was improved by adopting the HMM, the HMM outputs misunderstood results when it fails the recognition. On the other hand, the automata ends with the answer of “No meaning”, to allow the system to interactively ask the user to input the gesticulation again.

Table 6. *Experimental results of sign-language recognition by HMM*

JSL words	Recognition Rate (%)	Misunderstood
You	100	-
Like	88	Interesting
Good bye	96	Happy
Happy	91	Good bye
North	100	-
Sad	100	-
Interesting	94	Like
Father	96	Interesting
Mother	100	-

4. CONCLUSIONS

The automata algorithm and the HMM are introduced for the recognition of JSL together with the construction of JSL database in which words are described as sequences of gesture primitives obtained from acceleration patterns and hand figures. A compact acceleration sensor which is considered to reduce the constraints of motion is introduced for the motion primitive extraction. In the current study of the JSL recognition, the HMM has better performance rather than the automata algorithm. The algorithms have to be further examined on their performance and reliability by increasing the JSL database scale. Although learning procedure by the repetition of pattern inputs in advance is necessary in the HMM method, it is considered to give robust recognition with the integration of gesture sensing devices.

5. REFERENCES

[1] T. Kurokawa, "Non-verbal Interfaces", IEICE (ed.), Ohm Co., 1994

[2] H. Sawada and S. Hashimoto, "Gesture Recognition Using an Acceleration Sensor and Its Application to Musical Performance Control", Trans. Electronics and Communications in Japan, Vol.80, No.5, pp. 9-17, 1997

[3] T. Starner and A. Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models", Proc. of ISCV'95, 1995

[4] J. Kim, W. Jang, and Z. Bien: "A Dynamic Gesture Recognition System for the Korean Sign Language (KSL)", IEEE Trans. on Systems, Man and Cybernetics -Part B:Cybernetics, Vol.26, No.2, April, 1996

[5] H. Sagawa, M. Takeuchi, M. Ohki: "Description and Recognition Methods for Sign Language Based on Gesture Components", 1997 Int'l Conference on Intelligent User Interface , pp.97-104, 1997

[6] H. Sawada, N. Onoe and S. Hashimoto, "Acceleration Sensor as an Input Device for Musical Environment", Proc. ICMC, pp 421-424, 1996

[7] Arto Salomaa, "Theory of Automata", Pergamon Press, 1969

[8] R.J. Elliott, L. Aggoun and J.B. Moore, "Hidden Markov Models, Estimation and Control", Springer-Verlag, New York, 1995