Interface Development for Computer Assisted Sign Language Learning: Compact Version of CASLL

Saori Tanaka†, Yosuke Matsusaka‡, Kaoru Nakazono††

†MID
Nonoshita, Nagareyama, Chiba,
270-0135, Japan
tanaka@mid-japan.com

‡National Institute of Advanced Industrial Science and Technology
Umezono, Tsukuba, Ibaraki 305-8568, Japan
yosuke.matsusaka@aist.go.jp

††NTT Network Innovation Laboratories
3-9-11, Midori-cho, Musashino-shi, Tokyo,
180-0012, Japan
nakazono@core.ecl.net

Abstract

In this paper, we introduce e-learning system called CASLL and demonstrate the small interface that is applicable for the different types of screen including the laptops and the other types of mobile movie players. In the existing learning program, users learn sign words and then try to select the appropriate Japanese translations in a natural conversation expressed by two native signers. In the proposed program, users try to segment each word from a stream of signing by manipulating a control knob on the bottom of a movie screen, and then do the same tasks in the existing learning model. Ten Japanese learners participated in the evaluation experiments. The mean accuracy rate of the proposed program was higher than that of the existing program. The result has indicated that focusing on transitional movements has an effect for learning JSL as a second-language. Although the segmentation learning method has been shown as an effective learning method, there were some technical problems. Some learners answered that they could not see each JSL movies at once by using their own laptops to conduct the learning programs. Therefore, we modified the interface to match the smallest interface as possible. This paper ends with a future plan for the evaluation of new interface.

1. Introduction

In order to present some contents of native sign language by using the technology of computer networks, a lot of research has been done all over the world in the field of educational engineering. In these works, there are two types for their aims; one is to support the deaf people themselves in order to fill the social gaps between the deaf people and the hearing people, and the other is to disseminate the knowledge of sign language in a society so that the circumstances around the deaf people are improved.

In the case of first group, the research of animation generation by natural language processing technology has been progressing in Greek Sign Language(Efthimiou et al., 2004). In Japan, based on the knowledge that the speed for playing JSL images depends on the level of proficiency for JSL, a system for playing JSL images in five speed level has been developed(Isono et al., 2006).

Meanwhile, in the case of the second type of research, a remote communications system to connect a class of American Sign Language to the terminal of students has been developed(Lehman and Conceicao, 2001). In Japan, the self learning system of finger spelling with the function of feedback(Tabata et al., 2001), the JSL database with search function based on the linguistic knowledge of native Japanese signers(Fukuda, 2005) and the teachware of JSL have been developed(for JSL Learning,).

As one of our series of previous studies for developing human interface by using Japanese Sign Language (JSL) contents(Tanaka et al., 2007b; Nakazono and Tanaka, 2008; Tanaka et al., 2008), we proposed a new learning program and compare it with the existing learning program implemented in the Computer Assisted Sign Language Learning (CASLL) system(Tanaka et al., 2007a). Those studies can be categorised in the second type of research.

2. Goal of the paper

In this paper first, we introduce a more effective learning method compared to the exsisting learning program implemented in the CASLL system. Secondly, we define the size of movie screen as small as possible, and develop user-friendly interface that is applicable for the different sizes of screen.

3. Design of segmentation learning

In this section, we will sort out the problems of the existing learning style compared to the characteristics of continuous sign in natural discourse.

In natural discourse by native signers, it is observed that hand-shape and position changes under the influence of those in the proceeding word, while it didn't occur when each word is carefully pronounced (Fukuda et al., 1961) [1]. It is also reported that hand-shape slightly changes its feature when two words form a phrase (Ichida, 2005). When a signer produces number "1" as a single word, he/she holds the other four fingers tight except for the index finger. However, when it forms a phrase with the word which five fingers open softly, the other four fingers except the index finger are closed lightly. This fact is caused by the fact that the hand shape with five fingers open effects the anteroposterior word, and then the feature of "open" is added to the four fingers except the index finger. This phenomenon seems to be equivalent to the phoneme assimilation in spoken language, the term of "assimilation" is used in sign language research(Ichida, 2005). The phenomenon is related to efficient articulation and is often observed in natural speed discourse in native signers.

Meanwhile in the learning situation of beginners, there are many students who face difficulty reading words in natu-

[1] In American Sign Language, it is also observed that the position of word is getting closer to the proceeding word position when the speed of articulation is increased(Claude, 2003)

ral discourse, even though they have already learned each word. It is not uncommon to see the number of students are decreasing by more than half in a class, as the levels of a course advances. One of the reasons for this problem seems to be the general learning method used in classes or groups of JSL. In some classes, students are often required to memorize all the vocabulary for greetings, introducing their family, and numbers. After the lesson, they start to practice to read and express the words with students in pairs. By such kinds of learning style, it is difficult to read sentences with many phoneme assimilations in natural discourse.

Based on these problems, we defined three points as necessary functions for the design of proposed learning program; (1) students can check the images as many times as they want to, (2) students can learn a word expressed carefully by comparing it to the same word expressed with other words continuously and (3) students can be aware of the segment boundaries including transitional movement.

We also implement the word-learning program that users mainly study JSL vocabulary on CASLL, and then evaluate how much each user can recognize the words in natural discourse. Based on the results, we verify the effectiveness of proposed program including the segment task before the vocabulary task.

Feg.1 shows the relations in each leaning program and continuous sign images in natural discourse. In the word learning that is more conventional in JSL education, users mainly focuses on the lexical movement without transitional movement in the segment: the short arrows in fig.1. Compared with this, we design the segmentation learning which makes users focus on the long arrow in fig.1. We expect that the users can study sign production including the natural assimilations between words, and it will improve their skill for reading as a result. We will describe the details in each learning program.

4. Two learning programs and system

Table.1 show the process in each learning program. The texts in boldface are common tasks in each one. In the segmentation learning, there are three different tasks before the common tasks. Those three tasks are comportments of the segmentation learning, the proposed method in this study.

In the word learning, users study only lexical movements without paying attention to transitional movements, and then memorize the Japanese translations. After that, users challenge the reading task to answer the meaning of word in each question. On the other hand, in the segmentation learning, the users count the stroke of movements in each example of natural discourse. This task is aimed at making them focus attention in each component of movement, and they should become used to the natural speed with many assimilations. After the counting task, they try to correct the order of word images which are randomly-aligned in the order of natural discourse. In the "word images", a signer produces each word separately. At the upper side of the word image, there is discourse image in which a native signer is signing (Fig.2). The re-realignment task is aimed at relating the words and the sequential signs. The users should be conscious of the difference between the words and the segments including transitional movement and lexical movement. After the re-realignment task, users perform the segment task in which they try to match the knob at each segment in the short piece of discourse image(Fig.3). We designed these three tasks so that users can be conscious of each segment including transitional movement. After these three tasks, they perform the tasks for the word learning: the word learning task, and the examination for checking how many word they correctly memorized. In the following section, we will explain about the detail of each learning program.

Table 1: Learning Processes of Two Learning Programs

	Word Learning Program		Segmentation Learning Program
1	**Word Learning Task**	1	Counting Task
2	**Word Test**	2	Realignment Task
3	**Reading Task**	3	Segment Task
4	Realignment Task	4	Word Learning Task
5	Translation Test	5	Word Test
		6	Reading Task
		7	Translation Test

Figure 2: User's Screen for Reorder Task

4.1. Outline of two learning programs

- **Word Learning Program**

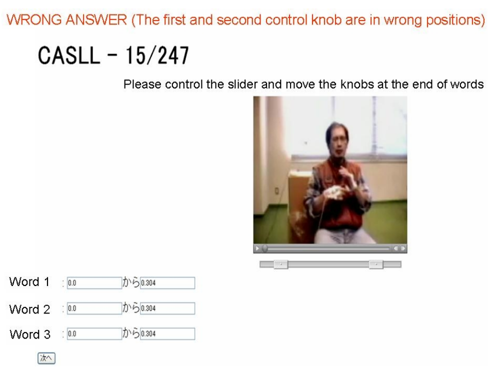

Figure 3: User's Screen for Segmentation Task

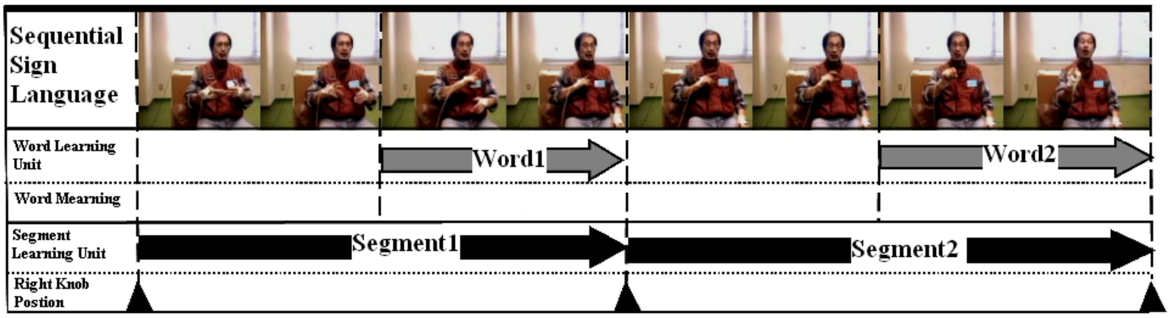

Figure 1: Learning Units for Segmentation and Word Learning Programs

1. Word learning task

   Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

2. Word test

   Users are required to select the correct Japanese translation for each which is randomly-presented word. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

3. Reading task

   Users are required to read each word presented in the image of natural discourse which length of about 5-10 seconds and select the meaning of it.

4. Realignment task

   Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. Each image has Japanese subtitles.

5. Translation test

   Users are required to select a correct translation that is appropriate for the example image by a native signer.

- **Segmentation Learning Program**

1. Counting task

   Users are required to count the stroke of movements in each example of natural discourse. They can select the number of stroke from choices.

2. Realignment task

   Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. In order to match the order of word images to the sentence in discourse image, they need to manipulate the buttons for switching the images without Japanese subtitles. At this point, they have not learned the meaning of each word, so they can only check how each stroke structures the word in natural discourse.

3. Segment task

   Users are required to select the length of each segment in a short piece of image cut from natural discourse. As in Fig.3, they need to manipulate the knobs on the slider with referring to the discourse image, and set the position of knob at at the last frame when the hands start to change the shape and translate to the next word. If the user doesn't set it at the right position, the arrow message is shown up saying "you have wrong answer (the second knob is at the wrong position)". After the user corrects the position, he/she can move to the next task[2].

4. Word learning task

   Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

5. Vocabulary test

   Users are required to select the correct Japanese translation for each word which is randomly-presented. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

6. Reading task

   Users are required to read each word presented in the image of natural discourse that length is about 5-10 seconds and select the meaning of it.

7. Translation test

   Users are required to select a correct translation that is appropriate for the example image by a native signer.

5. System overview

Fig.4 shows the system overview of CASLL.
Once the scripts with the procedure of two learning programs has been sent to the server, the server finishes the preparation for judging the answers; generation of hints; and generation of the forms of buttons, knobs, and images.

[2]The correct answer was set with the 2 frames margin in the vicinity.

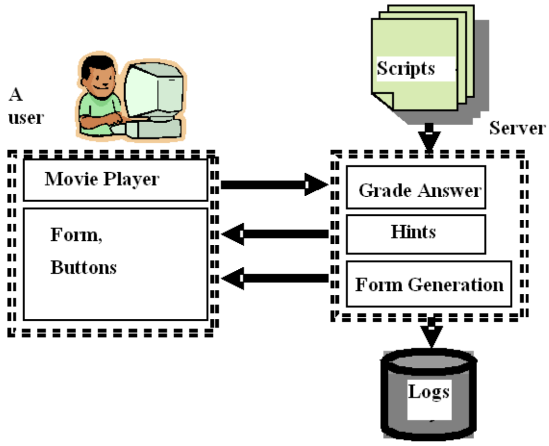

Figure 4: System Overview

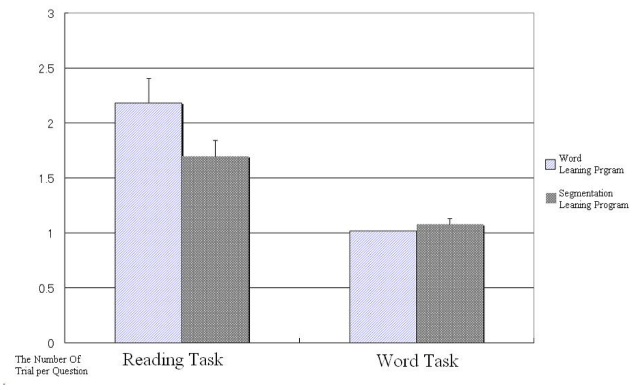

Figure 5: Mean Number of Trials for Reading Task and Word Task

Once the user finished each task with displaying images from the server, each answer will sent to the server. The server will display hints according to the user's answer, and present the form of next task when the user's answer is correct. The hints include the information about the number of correct answers per question, and the wrong position number of the knobs in the segment task. The user's record will be stored in the server in order to evaluate for the next analysis. We designed CASLL based on the assumption that users have Internet Explore 6.0+ (Microsoft), QuickTime 7+ (Apple), and ADSL or cable or large-scale network at universities. From the experiments that we have done before, the users can perform the tasks in CASLL with these environments.

6. Evaluation experiment

6.1. Procedure

The aim of this experiment is to present a self-learning program by which a user can be thoughtful about hihe/sher choice of tasks. For this aim, we compare the scores of two subject groups; the word learning group and the segment learning group. We evaluate the effect of the segment learning for the reading tasks in CASLL. The users can see the question images as much as they want and perform the tasks until they reach each correct answer. In this situation, if the learning effects of two learning programs for the reading task were not so different, we could not say that there was a special learning effect in the segment learning program. If so, the effect of the reading task would be only the result of the word learning. Therefore, for this experiments, we will analyze the proficiencies of vocabulary in two programs. If we could confirm that the proficiencies are high enough in both two programs, we would use it as a baseline and then compare the scores of reading task in both programs. For this comparison, we would use the percentage of number of questions that the users get correct at the first trial, and the number of trials for reaching each correct answer.

6.2. Subjects

The subjects were 10 beginners who have less than two years learning experience at some JSL learning clubs. In these subjects, 5 beginners performed the word learning program and the other 5 beginners performed the segment learning program. The range of their age was from 20 to 38 years old, and the male-female ratio was 3:2 respectively.

6.3. Learning program

Although the learning program at the server can be restructured by writing the script, we used some images cut from a image of natural discourse. The length of each image was about 7 seconds. The total number of images were 6, and those included 53 basic words that were used in a natural discourse. The length of whole image was 49 seconds and the average speed of signing was 1.54 [word/sec]. That speed was quite fast for the users based on the results of Questionnaire that all users answered after the experiment. For those 6 images, we made the script for the two learning programs in Table.1, and implemented the tasks by which each user perform the 6 trials. We also added a function by which the users could answer each question 5 times and after 5 times, they could select "give up" and go next task.

6.4. Result and discussion

6.4.1. Level of proficiency for vocabulary

$$\text{(the number of question - the number of wrong answer)} \div \text{the number of question} \times 100 \quad (1)$$

From the expression (4.1), we calculated the proficiencies of 53 words in both learning programs. These were 98 in both learning programs that were close to perfect. As a result, it was confirmed that the proficiencies of the vocabulary in both learning groups were similarly high.

6.4.2. Analysis for score in reading task

In order to calculate the score of the reading task, we substituted the number of reading questions and the number of wrong answer in the reading task to (4.1). As a result, the average in the word learning group was 67.0 and that in the segment learning group was 54.8. The former average was significantly higher than the later one ($t(8)=1.97, p<.05$).

6.5. Analysis for number of trial

Although this system allows the users to skip the question with the giving up function after the fifth trials, they can also answer the questions as many as they want without the giving up function. Therefore, at least in principle, they can continue the trials to answer a question until they reach the correct answer. If he/she answers correctly at the first trial, he/she can have higher score, so we analyzed the number of trials they had in both the reading task and the word task. If the segmentation learning had more effect for the reading task than the word learning, the average number of trials of segmentation users would be less than that of the word users. The result was shown in Fig.5.

In Fig.5, the two bars on the left side shows the result of reading task. The mean trial number of reading task in the segmentation users was less than that in the word users. This means that the segmentation users reached correct answers with less trials than those by the word users. The mean trial numbers of reading task were 2.18 in the word learning group, and 1.69 in the segmentation learning group; SD were 0.49 and 0.32 respectively (t(8)1.84, p=.051). The less trial numbers needed by the segmentation users seems to reflect the higher score in the reading task.

In Fig.5, the two bars on the left side show the result of word task. Both groups of users reached the correct answers mostly at each first trial. The mean trial numbers of word task were 1.02 in the word learning group, and 1.07 in the segmentation learning group; SD were 0 and 0.11 respectively. From this result, we found that all user in both learning groups answered almost perfectly for the random questions in the word task.

The difference of learning program in both learning groups were the following; the segmentation users had the counting task, the realignment task, the segment task, the word learning task, the vocabulary test and then the reading task. The first three tasks are the original component of the segmentation learning. On the other hand, the word users had the word learning task, the vocabulary test and then the reading task. The word users seems to find it difficult to read the meaning of each word in natural discourse immediately after the word learning task and the vocabulary test. Additionally, the word users had a perfect vocabulary skill in 53 words, so even if they had more trials for the word learning task, the score of the reading task would be same. Compared with them, the segmentation users saw the natural discourse image over and over again, and had the counting task, the realignment task and the segment task, so the users seems to find it easy to recognize the relation between movement and meaning in the early stages. From the fact that the levels of proficiency for vocabulary are highly enough in both learning programs, the segmentation learning showed the effectiveness in reading the signing words in natural discourse.

6.5.1. Questionnaire

From the above analyses, we found the effectiveness of the segmentation learning, but there might be room for improvement in the program which was a first edition. We studied the possible way for improvement from the analysis of questionnaires. Fig.6 shows the result of questionnaires that all user answered after their learning programs. They were asked to rate their impression based on five questions from Q.1 to Q.2 on the grade of five scale each. Q.1 was "Were the questions difficult for you?" and almost all user rated more than 4 (difficult). It indicated that the challenge levels were high for all user in both programs. Q.2 was "How did you feel about your workload?". For this question, two word users rated 3 (right quantity) and three rated 4 (it was heavy). On the other hand, one segmentation users rated 3 and four rated 5(it was too heavy). The extra duration of the segmentation leaning seems to make the users feel workload to be heavier. Q.3 was "Do you want to learn by this same learning program again?". Four word users rated 4 (yes, when I have time), and one rated 5 (definitely yes), while three segmentation users rated 1 (no, never), one rated 4 (yes, when I have time) and one rated 5(definitely yes). The evaluation result varied greatly in the segmentation learning. Q.4 was "How was the efficiency of this program compared to other learning methods by which you had learned before". One segmentation user rated 1 (highly inefficient) but more than half of users answered "efficient" and "highly efficient". Q.5 was "How was the operability of your computer for this program ?", and more than half users rated "good" and "very handy". The user who answered "a little bit difficult to operate" described her impressions that her computer screen was too small to see all word image when she attempted the realignment task. From the above results of operability, we found that the problems for the interface would be improved easily if we changed it to the adjustable display system with user's screen size. Additionally, we found some room for improvement in the curriculum of the segmentation learning. For the purpose of comparative experiments, we needed to design the segmentation learning program with a number of questions, but the evaluation of Q.3 seems to be improved if we control the number of trials that the segmentation users study at once. Furthermore, presenting a kinder explanation of each of task and meaning at the beginning of the program would be help to reduce the user's mental burden for the number of questions.

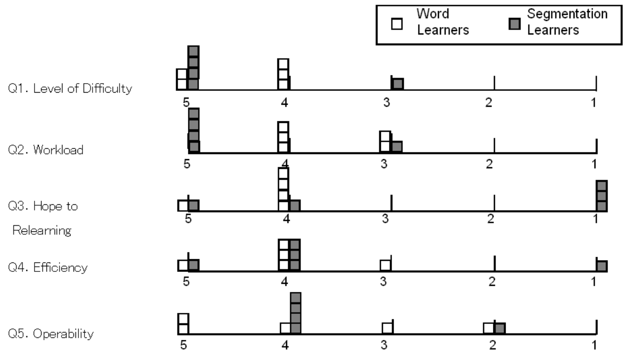

Figure 6: Results of Questionnaire

7. Interface for Small Screen

7.1. Requirements

As we described in the previous section, our interface for *realighment task* has the problem to overview the words. In order to realize learning with mobile device which has small screen display, we have to squeeze the interface to more smaller size. Although, at the same time, because we have to keep learning efficiency and usability, the interface have to satisfy the following requirements.

The interface has to be in small size: As in our main objective.

The user is easy to overview visual words at a glance: Words in sign language is a visual information, and visual information requires certain size on the screen to tell correct information to the user. On designing the interface, we have to care not to make overlaps between the videos nor show the video in too small size to recognize sign language.

The user is easy to pickup words using small interaction devices: Because the user have to interact with the interface using through small and unprecise interaction devices (e.g. touch panel of the mobile phone), the word selection has to be done without any precise control.

7.2. Design

After a concern on the above requirements, we have made following 3 prototype designs.

Type 1: Original interface with a scroll bar. Same as original interface described in previous study with a scroll bar to scroll the screen to browse the words went over the screen (Fig. 2).

Type 2: Overlapped words. Overlap the 1/2 size of the words with the other words. When the mouse overlaps a word, the word will be zoomed and the user can see the word without overlapping (Fig. 7).

Type 3: Slanted words. The words are slanted and shown in small view without overlapping. When the mouse overlaps a word, the word will be zoomed and justified, and the user can see the word without slanting (Fig. 8).

7.3. Implementation

The movie is 320 x 240 pixels and encoded in H.264 codec. The interface is implemented using Adobe Flex3 SDK[3] with papervision3d library[4].

Because the interface runs on *flash* platform, user can open the interface on their any preferred web browser.

7.4. Evaluation Plan

Alghouth we have just started to develop the new interface and classify the visual styles, we are going to evaluate the usability of the new types of interface from the view point of mental workload. As the one of subjective workload assesment methods, NASA-TLX has been developed by NASA (National Aeronautics and Space Administration) Ames Research Center(Hart and Bittner, 1988).

The evaluation procedure consisted of the following three steps.

[3]Adobe Flex3 SDK:`http://www.adobe.com/products/flex/sdk`

[4]papervision3d library: `http://papervision3d.org/`

Figure 7: Overlap the 1/2 size of the words with the other words. When the mouse overlaps a word, the word will be zoomed and the user can see the word without overlapping(Type2).

Figure 8: The words are slanted and shown in small view without overlapping. When the mouse overlaps a word, the word will be zoomed and justified, and the user can see the word without slanting(Type3).

1. Evaluation of workloads on six subscales: Evaluate the workload on six subscales using values from 0 to 100

2. Comparison of subscale weights: Evaluate the weights of the subscales by the pair comparison method

3. Unification of workloads: Calculate the weighted workload (WWL) by taking the average of the weighted values of the subscales

The six subscales are:

1. Mental Demand

2. Physical Demand

3. Temporal Demand

4. Performance
5. Effort
6. Frustration

The one of the authours has developped the JSL version of NASA-TLX(Nakazono et al., 2006).The original NASA-TLX, the explanations of the procedure and evaluation items are written in English. Descriptions in the original NASA-TLX were so stiff and technical that it was assumed to be hard to understand them completely and carry out fair assessments for ordinary people. With this JSL version of NASA-TLX, communication quality of intermediated on-line conversation by interpreter (Nakazono and Tanaka, 2008) have been already evaluated. We will use this methodology for the evaluation of new interface.

8. Conclusion

In this paper, we introduced an effective learning method implemented in the e-learning system called CASLL. We also demonstrated the small types of interface which are applicable for the mobile movie players. We will evaluate the usability of the new types of interface by JSL version of NASA-TLX. The most usable interface would be applicable for the website that assumed users have different sizes of screen, cell-phones and other types of mobile movie players.

9. References

M. Claude. 2003. *Undershoot in two modalities: Evidence from fast speech and fast signing*. Ph.D. thesis, Univ. of Texas.

E. Efthimiou, G. Sapountzaki, K. Karpouzis, and S-E. Fotinea, 2004. *Developing an e-Learning platform for the Greek Sign Language*, pages 1107–1113. Springer, 3118 edition.

Compass: Video for JSL Learning. *http://www.inform-inc.co.jp/video.htm*.

Y. Fukuda, H. Kimura, Y. Ichida, K. Fukushima, Y. Izumi, T. Sekine, K. Suzuki, N. Nakajima, A. Kazugai, and M. Norimatsu. 1961. Expression through hand and finger actions used in ″ words in continuous signing ″ : A study in reference to the dialogue in Japanese Sign Language by the deaf persons (in Japanese). *IEICE technical report. Speech*, 97(587):pp.81–85.

F. Fukuda. 2005. Compilations of the electronic dictionary of Japanese Sign Language (a second edition) and its instruction manual (in Japanese). *IEICE technical report. Welfare Information technology*, (66):pp.39–44.

J. Hart and A. Bittner. 1988. Development of NASA-TLX: Results of empirical and theoretical research in human mental workload. *British Journal of Social Psychology*, pages pp.139–183.

Y. Ichida. 2005. Phonology and numeral system in jsl (in Japanese). *Sign Language Communication*, (57):pp.8–16.

H. Isono, Y. Takiguchi, M. Katsumata, and M. Nakama. 2006. Preferred reproduction speed of sign language image and audiovisual system via the internet (in Japanese). pages pp.5–8.

R. Lehman and S. Conceicao. 2001. Involving the deaf community in distance learning, using blended technologies and learning objects. *IEEE Electronic Journal*, pages pp.3–4.

K. Nakazono and S. Tanaka. 2008. Study of spatial configurations of equipment for online sign interpretation service. *IEICE - Transactions on Information and Systems archive*, (6).

K. Nakazono, Y. Takeda, M. Terauchi, and Y. Nagashima. 2006. Evaluation of mental workload of delayed visual communication. *TR of IEICE WIT2006-49*, 106(285):95.

K. Tabata, T. Kurota, M. Murakami, Y. Manabe, and K. Chihara. 2001. Prototype design for sign language education system (in Japanese). In *Proceedings of Japanese Association of Sign Linguistics*, number 27, pages pp.34–35.

S. Tanaka, Y. Matsusaka, and K. Uehara. 2007a. Segmentation learning method as a proposal for sign language e-learning (in Japanese). *Human Interface*, 9(2):61–70.

S. Tanaka, K. Nakazono, M. Nishida, Y. Horiuchi, and A. Ichikawa. 2007b. Skill-nms for an indicator of qualitative skill in the interpreters of japanese sign language. *Proceedings of International Symposium on Skill Science (ISSS)*, pages 178–180.

S. Tanaka, K. Nakazono, M. Nishida, Y. Horiuchi, and A. Ichikawa. 2008. Evaluating interpreter ’ s skill by measurement of prosody recognition. *Transactions of the Japanese Society for Artificial Intelligence*, 23(3):117–126.