




Constructing a Highly Accurate Japanese Sign Language Motion Database Including Dialogue

Yuji Nagashima¹(✉) , Keiko Watanabe¹, Daisuke Hara², Yasuo Horiuchi³, Shinji Sako⁴, and Akira Ichikawa¹

¹ Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan
nagashima@cc.kogakuin.ac.jp, ed13001@ns.kogakuin.ac.jp

² Toyota Technological Institute,

2-12-1 Hisakata, Tempaku-ku, Nagoya 468-8511, Japan

³ Chiba University, 1-33, Yayoicho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan

⁴ Nagoya Institute of Technology,

Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan

Abstract. Sign language is a visual language. Research into Sign language, including in related areas such as linguistics and engineering is lagging behind spoken language. One reason for this is the absence of a common database available to researchers from different areas.

This paper defines how to words are selected for inclusion in the sign language database, how to select informants, what format the data are included in and how to record the data. Also, the database includes three-dimensional behavioral data based on dialogue. The following explains the results.

Keywords: Japanese sign language · Database · Optical motion capture · Dialogue

1 Introduction

Sign language is used by deaf people, and is a natural interactive language different from and independent of oral language. Although it is a language, research on sign language remains underdeveloped in linguistics, engineering, and other related fields when compared to spoken languages. One reason for this is the lack of a multi-purpose database for shared use by linguists, engineers, and researchers in other fields.

However, differences are expected between the data formats desired by linguists, engineers, and those working in cognitive science. To standardize sign language research in many areas, it is desirable to record the same sign language movements in a data format compatible with researchers' needs. Analyzing the same sign language movements from different perspectives and in different ways may multiply opportunities to obtain new insight.

This paper defines how to words are selected for inclusion in the sign language database, how to select informants, what format the data are included in and how to record the data. Also, the database includes three-dimensional behavioral data based on dialogue. The following explains the results.

2 Selecting the Words to Be Included in the Database and the Informants

The selection of JSL expressions to be included in the database was based on data about the frequency of use of Japanese words. The informants, are native signers, were chosen based on a criterion to use a type of Japanese sign language which is ease to read.

2.1 Selecting JSL Words

For JSL data, referring to the corpora of spoken Japanese such as the “Lexical Properties of Japanese [1]” made by NTT and the “Corpus of Spontaneous Japanese [2]”, as well as sign language corpora such as the “NHK sign language news [3]” and the “Japanese-JSL Dictionary [4]” published by Japanese Federation of the Deaf, we have chosen signs to be recorded.

From the list of candidates, Japanese sign language expressions to be included in the database were selected based on the expressions in the Japanese-JSL Dictionary. The Japanese-JSL Dictionary includes more expressions than any other sign language dictionary published in print form. It contains nearly 6,000 expressions.

Furthermore, the following words have been added. They are not included in the dictionary we consulted.

- Words relating to the latest IT
- Words relating the imperial family and Reiwa, the new era name in Japan
- Words relating to sports, as the Tokyo Olympics is right around the corner
- Words relating to disaster
- Words relating to disability and information security

The sign language expressions have been verified in cooperation with persons who use sign language as their primary language.

2.2 Selection of Informants

As well as selecting signs to be recorded, we have discussed requirements about the informants who we will work with. We have decided that the informants must:

- be native signers of JSL born in a Deaf family
- use a type of JSL which is ease to read
- consent to making the JSL data public

Taking these requirements into consideration, we have decided on M (male, 41 yrs.) and K (female, 42 yrs.).

3 How Data Is Recorded and in What Format

We have discussed the best source format, spatio-temporal resolution, format of data files, and storing method for academic fields such as linguistics and engineering. Detailed analyses of the distinctive features, phonemes or morphemes of sign language involve detailed analyses of manual signals and non-manual markers. Not like Audio data which are one-dimensional on a time axis, sign language data are three-dimensional with a spatial spread on a time axis. If we do numerical analyses about sign language motion data, as we do with audio data, it is highly expected to obtain new information useful for sign language research. However, compared with audio data, sign language motion data will have much higher complexity because they involve more dimensions than audio data. Accurately analyzing sign language motion and generating high-quality computer-generated models for sign language will require highly accurate three-dimensional measurements of motions. If highly accurate three-dimensional sign language motion data and two-dimensional video data of sign language can synchronize, then it will be possible to analyze how the three-dimensional motions break down into two dimensions and progress on a time axis in terms of the recognition of sign language including non-manual markers and the analysis of its motions. This will make it possible to obtain a new methodology for understanding and recognizing sign language.

In the database that we built, the following data in three different frame rates are recorded by synchronizing them.

Figure 1 shows the configuration of the recording equipment and synchronization conceptual diagram used in 2018.

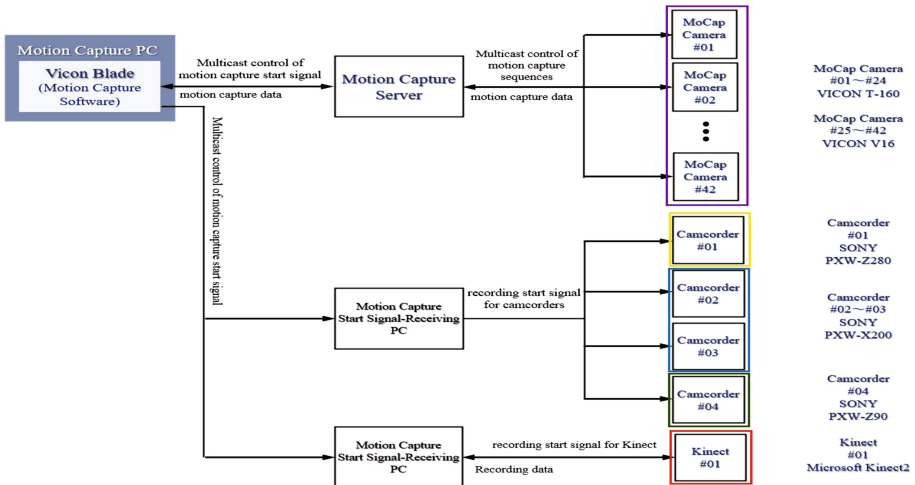


Fig. 1. The configuration of the recording equipment and synchronization conceptual diagram used in 2018

3.1 Three-Dimensional Behavior Measurement

Optical motion capture (hereinafter “MoCap”) was used to measure three-dimensional data with high precision in terms of space and time. Measuring each word involved the use of forty-two T-160 cameras from VICON, each with 16 megapixels to achieve a spatial resolution of 0.5 mm and a temporal resolution of 119.88 fps. A total of 66 cameras, 33 per person, were used to shoot dialogues. MoCap data was recorded in three formats, C3D, BVH and FBX.

The frame rate of each data format is shown below.

- C3D data at 119.88 fps
- BVH data at 119.88 fps
- FBX data at 119.88 fps

3.2 Camera for Shooting

Three SONY camcorders were used to shoot videos of sign language; one located in front of an informant, one to the left and the other one to the right. For 2019, all three camcorders were 4K. The capture frame rate is 59.94 fps.

3.3 Depth Sensor

Less expensive sensors capable of measuring distances with relatively high precision are widely available. We used the Kinect 2 ToF (Time of Flight) format. The maximum frame rate of data is 29.98 fps.

4 Recording Words in the Database

Videos for 4873 words were shot in the same shooting method and data format as described in Sect. 3 above. Table 1 shows, by year, the number of words shot in the videos. Each of the words with different movements and the same meaning (hereinafter “synonyms”) are counted as a single Japanese label in the number of Japanese labels shown in the table. Therefore, each label with a synonym has different movements recorded. For numerical figures, units, the Japanese syllabary and alphabet, about 400 words were recorded in the 2019 version.

Table 1. Number of words shot in videos, by year

Year	2017	2018	2019	Total
Number of Japanese labels	400	600	3,873	4,873

The 2019 version also includes data from dialogues between two informants discussing different topics. For one of the topics (about 7 min long), post-processing (creation of BVH data, C3D data and FBX data) was performed to make the data three dimensional. Figure 2 shows how the dialogues were shot. Figure 3 shows 3D computer graphics from the FBX data of the dialogue data which underwent post processing .



Fig. 2. Dialogue shot in a video

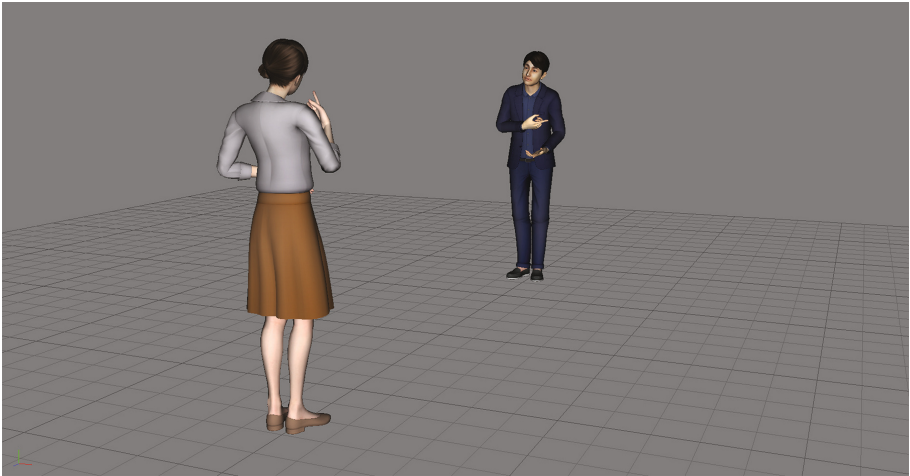


Fig. 3. 3D computer graphic of the dialogue data (drawn by FBX Reviewer)

5 Summary and Future Actions

This paper is about how the JSL database was built, how the data was recorded, and the actual recordings of words and dialogues. To date, 4873 words and dialogue data including the Japanese syllabary and alphabet have been recorded. A single set of dialogue data is transformed into three-dimensional behavioral data and can be synchronized with 4K videos and Kinect data for analysis. Three-dimensional behavioral data are important in the analysis of sign language. Currently, we are building an annotation support system which will be capable of synchronization and analyzing the data recorded in the database. Details about the annotation support system will be reported at another time. Incidentally, we have already completed the development of the viewer [5] that will be needed for the annotation support system.

From here onward, we will concentrate on efforts to make the database publicly available. After being opened to the public, the database will become available in linguistics, engineering and other interdisciplinary domains. In that regard, the usefulness of DB will be immeasurable. This will make it possible for different researchers with different perspectives to analyze the same sign language expressions and will hopefully help to obtain new insights. Moreover, we will complete the annotation support system and open it to the public.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number 17H06114.

References

1. Amano, S., Kasahara, K., Kondo, K. : NTT database series [Lexical Properties of Japanese] volume 9: word familiarity database (addition ed.). Sanseido Co., Inc, Tokyo (2008)
2. Corpus of Spontaneous Japanese. <http://pj.ninjal.ac.jp/corpus.center/csj/>. Accessed 29 Nov 2018
3. Katou, N.: NHK Japanese sign language corpus on NHK news. In: 2010 NLP (the Association for Natural Language Processing) Annual Meeting, pp. 494–497 (2010)
4. Japan Institute for Sign Language Studies (ed.): Japanese/ Japanese Sign Language Dictionary. Japanese Federation of the Deaf, Tokyo (1997)
5. Nagashima, Y. : Constructing a Japanese sign language database and developing its viewer. In: Proceedings of the 12th ACM PETRA 2019, pp. 321–322 (2019). <https://doi.org/10.1145/3316782.3322738>