



Construction of a Japanese Sign Language Database with Various Data Types

Keiko Watanabe^{1(✉)}, Yuji Nagashima^{1(✉)}, Daisuke Hara²,
Yasuo Horiuchi³, Shinji Sako⁴, and Akira Ichikawa¹

¹ Kogakuin University,

1-24-2 Nishi-shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan
ed13001@ns.kogakuin.ac.jp,
nagasima@cc.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. We have constructed a sign language database which shows 3D animations. We are aiming at constructing an interdisciplinary database which can be used by researchers in various academic fields. This database helps the researchers analyze Japanese sign language. We have recorded nearly 2,000 Japanese signs to now, and we are planning to record on the database approximately 5,000 signs. Firstly, we decided to pick up frequently used Japanese words on the database. Each sign language expression corresponds to the Japanese words is examined. Secondly, we recorded 3D motion data of the determined sign language expressions. We used optical motion capture to record 3D motion data. The data format obtained through motion capture is C3D data, BVH data and FBX data, and frame rate is 120 fps. In addition, we also recorded a full HD video data at 60 fps, super-slow HD data at 30 fps, and depth data at 30 fps, for use in analysis of sign language.

These are recorded synchronously. In addition, we have developed a new annotation system which can reproduce different types of data synchronously to make the database the most effective. Because it is necessary for data analysis to reproduce synchronously all data, which have been recorded at different frame rates.

Keywords: Sign language · Motion capture · 3D animations

1 Introduction

Sign language is a form of language and a nonverbal means of communication used by people with a hearing impairment. Compared with spoken language, research on sign language is lacking in engineering and linguistics. One reason is the absence of a multipurpose database that can be commonly used by researchers from different areas, such as linguists and engineers. We created a Japanese sign language database to be used in many different areas of research.

We decided to include sign language expressions in the database that are general and used more frequently. Although many sign language dictionaries are published in print form, sign language is inherently expressed by movement. Therefore, the database should be recorded as a set of video data and shows by movement. The database has more than one data format, such as motion capture data and depth data, for each video of a sign language expression. A system has also been developed to play the data for use in analyses of sign language. It is planned to make the sign language data in the database searchable by sign language expression, such as hand gestures, as well as by the meaning of an expression in Japanese spoken language. Ultimately, the database will include nearly 5,000 expressions.

This report explains how the sign language expressions were selected, how the sign language data was recorded, and the system we developed for playing sign language data.

2 Signs to Be Included in the Database

The database consists of sign language expressions that are general and used more frequently in daily life. Frequently-used Japanese words would be selected and a sign language expression would be recorded for each one.

2.1 Selecting Words to Be Included in the Sign Language Database

The selection of JSL expressions to be included in the database was based on data about the frequency of use of Japanese words. The terms of word familiarity are expressions with greater audio density, and those seen more frequently in Lexical Properties of Japanese [1] made by NTT, the Corpus of Spontaneous Japanese [2] and the sign language news [3] on NHK Educational TV were selected as candidates for inclusion in the database. 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 [4]. The Japanese-JSL Dictionary, a publication by the Japanese Federation of the Deaf, includes more expressions than any other sign language dictionary published in print form. It contains nearly 6,000 expressions.

As a result, we chose 3,000 Japanese words, such as *Onaji* (same), *gakko* (school), at first. Furthermore, we plan to add other necessary words such as finger alphabet.

2.2 Discussion About Sign Language Expressions

We discussed how to sign language express each of the Japanese words selected in Sect. 2.1. The sign language expressions have been verified in cooperation with persons who use sign language as their primary language. Japanese words and their sign language counterparts do not correspond perfectly. If one sign language expression cannot be decided for one Japanese word, more than one sign language expressions are recorded.

For example, the word *namae* (name) can be expressed with different expressions. *Onaji* (same) may involve the same movement, but the position of the hands may differ among individuals. In some situations, it may be expressed using only one hand. *Hiraku* (open) can also involve differences—in sign language, it depends on what will be opened. In this situation, two or more sign language expressions were recorded for a single Japanese word.

We also placed importance on consistency in the sign language expressions included in the database. For this reason, the sign language expressions for *hiraku* (open) and *tojiru* (close) correspond in an antonymous manner in the database.

2.3 Recording Data of Signs

To ensure accuracy in the analyses of sign language behavior, all of the selected sign language expressions were shot in video form. The shooting was recorded in three data formats. Recording the 3D behavioral data involved the use of optical motion capture. The 3D behavioral data included in the database are C3D data at 120 fps, BVH data and FBX data at 119.88 fps. Recording depth data involved the use of Kinect 2. Depth and infrared images were recorded at a maximum of 29.97 fps. Furthermore, high-resolution camcorders were used to record video data at a frame rate of 60 fps. It has been also decided that video data will be recorded by three HD camcorders at 59.94 fps and by a super-slow HD camcorder at 119.88 fps (29.97 fps for playing). The data at different frame rates was synchronized before being recorded.



Fig. 1. Appearance of the recording studio

Figure 1 shows how an image was shot and recorded. 42 motion capture cameras were installed and used for recording in detail, including the delicate movements of a hand during an exchange using sign language. A Kinect 2 unit was set in front of the person doing the sign language. Three high-resolution camcorders were placed in front, to the left, and to the right of each person. In addition, full-HD videos at 119.88 fps (29.97 fps for playing) were also recorded as a reference. Until now, 1,400 signs have been recorded using these data formats.

Two people, one man and one woman, worked as sign language models during the shooting. Both are native signers of Japanese sign language. They are the deaf, and were born in a Deaf family. They were also chosen based on a criterion to use a type of Japanese sign language which is ease to read.

3 Development of a System for Playing Videos

The data synchronously filmed as written in Sect. 2 differs in terms of frame rate. For analysis, however, the video data must be synchronized and played. To this end, we developed a system for synchronizing and playing the video data.

3.1 Function of the System for Playing Videos

The annotation system that is being developed will consist of a viewer part and an analysis support part. The viewer part has two subparts: a viewer for 3D motion data developed by Unity and another viewer for video data developed by .NET Framework. Major functions of the viewer include the following.

- Divide the screen into a maximum of four portions to synchronize and play arbitrary data recorded in the database;
- Data in the BVH file format to draw 3D computer graphics;
- Use of C3D data to draw marker points;
- Drawing motion capture data from arbitrary perspectives and view angles;
- Background drawing of motion capture data is possible with arbitrary data;
- Drawing in the BVH file format involves the selection from two male models and three female models; and
- Recording replay screens.

The screen of the viewer can be divided into a maximum of four portions, each of which can display data. Allowing data to be synchronized and played at different frame rates makes it possible to check and analyze multiple data in accordance with a time axis.

The animation of sign language involves female and male models that can be altered arbitrarily. Figure 2 shows a screen divided into four portions to display data: BVH data (a female model, a male model and a skeletal animation) and C3D data. Figure 3 shows a screen playing video data using .NET Framework. This two-screen show same scene.

Details of the system will be reported later.

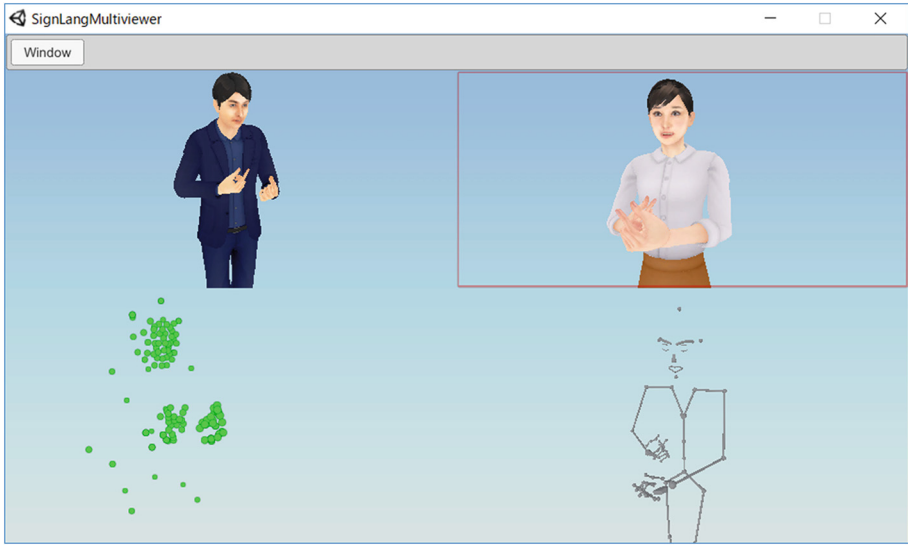


Fig. 2. A screen divided into four portions to display data using UNITY

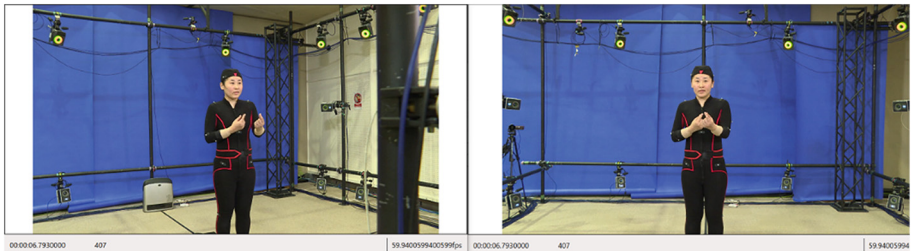


Fig. 3. Video data playing unit using .NET Framework

4 Conclusion

This paper is about the Japanese sign language database that is currently being created.

We considered the Japanese words to be included in the database. With the cooperation of native sign language speakers, we selected the sign language expressions corresponding to the Japanese words. Each sign selected was recorded in video form. Three types of data, including 3D behavioral data, depth data, and high-resolution data were synchronously recorded. To date, nearly 2,000 signs corresponding to 1,400 Japanese words have been recorded. We have planned to record nearly 5,000 signs. In the future, the remaining data will be recorded with an aim to complete the database.

A system has also been developed for synchronizing and playing the data. The system will be able to simultaneously play data at different frame rates.

The system for playing videos currently involves selecting data names. In the future, it will be designed in such a way that a set of sign language information can be entered into the corresponding signs already recorded. The sign language information will include the part of speech and type of word, the morpheme structure of sign language and sign language movement, such as hand shapes. Through the entry of this information, the database allows for searching by sign language expression, such as hand shapes and moves, whereas a regular dictionary only allows a person to search for sign language according to the meaning of the expression in Japanese language.

From our perspective, facilitating analyses of sign language data in greater detail requires sign language to be recorded on a sentence basis as well as on a word basis.

Acknowledgements. This work was *partly supported* by grant-in-aid from the *Ministry of Education, Culture, Sports, Science and Technology (MEXT)* of Japan (No. (S)17H06114).

References

1. Amano, S., Kasahara, K., Kondo, T. (eds.): NTT Database Series [Lexical Properties of Japanese], vol. 9, Word Familiarity Database (Addition), Sanseido, Tokyo (2008)
2. Corpus of Spontaneous Japanese. http://pj.ninjal.ac.jp/corpus_center/cs/j/. Accessed 29 Nov 2018
3. Katou, N.: 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)