Sign Language Motion Capture database recorded by one device

Pavel Jedlička[1]

1 Introduction

Sign language (SL) is a way of communication which uses the movement of a body. It uses manual, facial, and other body movements to express information. SL is a basic communication system of the deaf persons and it is often their natural way of communication. It is also the deaf community where evolution and development of SL originates.

As SL and spoken language developed separately and therefore linguistic mechanisms differs. According to Gibet et al. (2017), the deaf people are often facing problem using written language (based on spoken language), because it uses different grammatical rules, as well as the nature and the spatial organization of linguistic concept than spoken language. However, most information in media or on the Internet are available in spoken or written form. It leads to difficulties for deaf people to access the information.

2 Motivation and related work

In public television as an example, they use translation in form of video of signing speaker in the window added into the screen. Using artificial animated avatar which uses SL synthesis seems to be a good way to improve this situation. Compared to real SL speaker, artificial avatar technology is more flexible. It allows the content to be edited, can be produced more easily than video and it also preserves anonymity of the signer.

Recently, approaches based on key-frame techniques and procedural synthesis has been developed. This approach provides fine control over the movements of the avatar. This avatars are however poorly accepted by deaf community, because it's lack of human-like motion. There are some works e.g. McDonald et al. (2016) that aims to deal with this problem by adding human-specific noise in the key-frame driven motion. Data-driven synthesis, on the other hand, preserves the motion of original SL speaker. As mentioned in Gibet et al. (2017), this approach of synthesis often benefits from using motion capture (MoCap). The importance of 3D data is also mentioned in Metaxas et al. (2016). Moreover, the animations based on data from real signer provided by MoCap are more human-like.

There are some different approaches to create MoCap-based datasets or corpora for SL. There are also several challenges that need to be addressed. As SL is complex movement of different body parts and face, there is need to record each component for each sign. Problem is that every utterance can be done with both time and spatial variance even by same signer. While most of the recent approaches aims to record each part separately to gain the best precision possible, the need of merge this components together to synthesis one sign demands proper timing and it can lead to flaws or unnatural behavioral of resulting avatar. So it can be said that if it is possible to record all components at once, it is advantage for usability.

[1] Ph.D. student of Aplied sciencies and informatics, subject Cybernetics, specialization Artificial inteligence, e-mail: jedlicka@kky.zcu.cz

There is an approach by Krňoul et al. (2016a) to explore and develop methods to record SL dataset. This approach aims to record all components of SL utterance at one time using different devices. The state-of-the-art optical devices for body and face motion capture was used as well as resistive-based device for finger tracking. Wearing Cyberglove2, finger movement tracker, and VICON's CARA, the facial movement MoCap tracker, both causes discomfort and, therefore, can affect fluency or natural movement. It is also less precise than optical based system and calibration of Cyberglove2 itself is not a simple task, see Krňoul et al. (2016b).

3 New approach

Based on experience from above mentioned work, we tested approach using only one device to record all components at once. Specifically, we used combination of different VICON MoCap cameras and using VICON's MX system. Different shapes and sizes of retro-reflexive markers were used for different body parts, eg. 14mm and 8mm diameter spherical markers were used for arm and head tracking, and 2.5 and 4mm diameter hemispherical markers were used for tracking facial features and finger movement. This approach seems to be very promising while it deals with most of aforementioned problems. The drawbacks of this approach is that quality of recording is dependent on high camera count and their precise layout that is subject-dependent. While all markers are tracked optically, occlusions in some signs can not be avoided and therefore it is to be dealt with during post-processing.

Acknowledgement

The work has been supported by the grant of the University of West Bohemia, project No. SGS-2016-039.

References

Naert, L., Larboulette, C., and Gibet, S., 2017, *Coarticulation Analysis for Sign Language Synthesis.*, Universal Access in Human–Computer Interaction. Designing Novel Interactions", 2017, pages 55–75, ISBN=978-3-319-58703-5.

McDonald, J., Wolfe, R., Wilbur, R., Moncrief, R., Malaia, E., Fujimoto, S., Baowidan, S., and Stec, J., 2016 *A new tool to facilitate prosodic analysis of motion capture data and a data-driven technique for the improvement of avatar motion*, 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).

Metaxas, D., Dilsizian, M., Tang, Z., Huenerfauth, M., and Neidle, C., 2016, *The Importance of 3D Motion Trajectories for Computer-based Sign Recognition*, 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).

Krňoul, Z., Jedlička, P., Kanis, J., and Železný, M., *Toward Sign Language Motion Capture Dataset Building*, Speech and Computer, Springer International Publishing, 2016, Cham, pages 706–713, ISBN=978-3-319-43958-7.

Krňoul, Z., Kanis, J., Železný, M., and Müller, L., *Semiautomatic Data Glove Calibration for Sign Language Corpora Building*, 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).