Deep JSLC: A Multimodal Corpus Collection for Data-driven Generation of Japanese Sign Language Expressions

Heike Brock, Kazuhiro Nakadai

Honda Research Institute Japan Ltd.

Honcho 8-1, Wako-shi, Saitama, Japan

h.brock@jp.honda-ri.com, nakadai@jp.honda-ri.com

Abstract

The three-dimensional visualization of spoken or written information in Sign Language (SL) is considered a potential tool for better inclusion of deaf or hard of hearing individuals with low literacy skills. However, conventional technologies for such CG-supported data display are not able to depict all relevant features of a natural signing sequence such as facial expression, spatial references or inter-sign movement, leading to poor acceptance amongst speakers of sign language. The deployment of fully data-driven, deep sequence generation models that proved themselves powerful in speech and text applications might overcome this lack of naturalness. Therefore, we collected a corpus of continuous sentence utterances in Japanese Sign Language (JSL) applicable to the learning of deep neural network models. The presented corpus contains multimodal content information of high resolution motion capture data, video data and both visual and gloss-like mark up annotations obtained with the support of fluent JSL signers. Furthermore, all annotations were encoded under three different encoding schemes with respect to directions, intonation and non-manual information. Currently, the corpus is employed to learn first sequence-to-sequence networks where it shows the ability to train relevant language features.

Keywords: sign language, deep learning corpus, assistive language technologies

1. Introduction

Research has shown that information conveyed using Sign Language (SL) is much more comprehensible and accessible than written information display for the majority of native SL speakers (Traxler, 2000). However, human translation services are neither always available on demand, nor applicable in certain settings such as internal company meetings. The automatic display of spoken or written content by signing CG avatars or robot agents could therefore offer a cheap possibility to make information more accessible to Deaf or Hard of Hearing (DHH) individuals with lower written language literacy (Huenerfauth, 2008). Unfortunately, automated SL generation technologies have not yet reached a level of quality that would foster their full acceptance by DHH users (Kipp et al., 2011b). This is mainly due to the language's multi-dimensional character: information is not only conveyed using single gestures and finger movements, but also Non-Manual Signs (NMSs) such as body language and facial expressions. Furthermore, lexical items within a signed conversation can be dynamically changed to express spatial and temporal information (Huenerfauth and Hanson, 2009). This creates non-linear, multichannel dependencies that (a) complicate the translation of SL as compared to spoken languages and (b) require specific descriptions for motion synthesis. As a result, temporal and spatial relations of generated signing motion sequences might easily appear artificial and unnatural when applied to three-dimensional avatar animations (Kacorri et al., 2015).

To increase the acceptance of three-dimensional sign animations within affected users, effort is made to develop new methodologies that can better represent the temporal, spatial and multimodal aspects of signed content information (Ebling and Huenerfauth, 2015; Kacorri and Huenerfauth, 2016). To date, all of these methods rely on the concatenation and interpolation of pre-recorded sample sequences and separate lexical items. Systems that can reproduce natural signed utterances by intrinsic learning of their specific lingustic features are not known to be reported yet, commonly restricting existing data sources to non-repetitive corpus collections of signs and short phrases of specific content domains. Considering the recent success of deep neural networks in sequence generation tasks such as text-to-speech translations (Oord et al., 2016), one can expect similar architectures to also be meaningful for the synthesis of SL. To learn and evaluate respective network models, it is essential to have access to a suitable training corpus. In this work, we therefore present a fully annotated Japanese Sign Language (JSL) corpus that was specifically designed for the learning of deep sequence generation models for sign animation synthesis. Here, it is particularly important to ensure the accuracy, density and completeness of the recorded signing data up to subtle finger and facial movements and the repetitive occurrence of NMSs with wide intra-feature variability. Therefore, the presented deep JSL Corpus (JSLC) constitutes a collection of natural JSL sentences recorded with a multi camera optical motion capture system and includes various types of NMSs as well as variations in sentence intonations and signing speed. The corpus is designed to offer possibilities for easy data augmentation and ready to be applied under various deep network models. To bring the quality of automatically generated sign animation display to a next level of realism within the next years, it shall now be used for the development of an appropriate network model.

2. Sign Language Synthesis and JSL

Starting with the beginning of the last decade, various research efforts were made to enhance communication between hearing and non-hearing individuals, and to make information more accessible to DHH users. This led to a constant increase of assistive language tools, such as sign

translation software (Sun et al., 2013), dictionaries (Pro Deaf, 2017) or even commercial devices like the Motion-Savvy tablet app (Motion Savvy, 2017). However, most of these technologies are not capable of processing continuous and natural sentence utterances, and even less are concerned with the synthesis of signed expressions from a given set of text or speech data. Current SL generation methods are either fully synthesized from form notation such as VRML (Kipp et al., 2011a) or semi-automatically synthesized using a corpus of pre-captured gestures (Gibet et al., 2016; Ebling and Huenerfauth, 2015). Most recent neural network models that achieved superior results in speech generation are not known to be reported for applications with SL. One of the main reasons for this leeway is the multi-dimensional character of continuous signing with its inherent use of NMSs and spatial and temporal references that impede the use of machine translation algorithms. Another reason is the necessity of an extensive, highly detailed and accurate motion data set that is very costly to obtain: respective data is best obtained by using a multi-camera optical motion capture system for body, finger and facial capturing, and requires specialized post-processing procedures to exclude marker occlusions and mislabeling. Besides, existing SL motion capture data sets are commonly not publicly available. This imposes the need to acquire new data for system development.

Most current sign animation technologies were developed for the processing of comparatively well-researched SLs such as American Sign Language (ASL) (Lu and Huenerfauth, 2011), German Sign Language (DGS) (Ebling and Huenerfauth, 2015) or French Sign Language (LSF) (Gibet et al., 2016). JSL in contrast remains a relatively unexplored language under both syntactic and semantic aspects. Grammatical structures and lexical items can be undefined or subject to regional and demographic variations. For this reason, the generation of JSL sequences also constitutes a rare research domain. To date, sign animations were created for the following two fields of application: medical content translation for enhanced doctor-patient communication (Nagashima et al., 2016) and the broadcasting of weather forecasts in the national TV program (Umeda et al., 2016). Both animations are based on semi-automatized synthesis methods that utilize data from a problem-specific corpus of optical motion capture data. These corpora contain a variant number of single lexical items and few short complete sentences that were each signed once by one single speaker. Such sparsity in the vocabulary domain and the absence of NMSs is likely unsuitable for the deep learning of signed sequence connections. Therefore, we built the present deep JSLC in a different way that is not known to be reported similarly by any other research yet.

3. Corpus Definition

A large amount of data is necessary to train deep sequence generation networks that could cover the full vocabulary spectrum of JSL. However as compared to speech or text data, the acquisition of detailed, accurate and complete SL expressions is a very expensive and time-consuming task. To evaluate the potential benefit of deep sequence generation techniques for natural sign display, it was consid-

ered sufficient to first train networks on a smaller subset of JSL vocabulary that could be extended for better generalization later. Therefore, we defined the proposed deep JSLC as a collection of sentence expressions and multi-word phrases within a dense vocabulary domain. Apart from the choice of repetitive vocabulary and content, this was achieved by the utilization of compound words built from lexical items and space and size classifiers (e.g. hospital signed as MEDICAL/ BUILDING and doctor signed as MEDICAL/ MAN, or to swim signed as SWIM and pool signed as SWIM/ PLACE).

Furthermore, the corpus should incorporate a combination of the most important and variant linguistic features of JSL that can hardly be reproduced by conventional generation methods in a satisfying way (Figure 1):

- **Direction:** Directional modification of lexical items conveys spatial and grammatical content information, such as the passive voice of words (e.g. BEING SPOKEN AT versus TO SPEAK) or the center of actions (e.g. TO GO and TO COME, TO SEE, TO BEG)

- **Syntax:** Intonation conveyed by syntactic NMSs (e.g. a raised eyebrow or head shaking) provides information on the sentence structure such as negation, past tense or question

- **Adjective inflection:** Positive, comparative and superlative forms of adjectives are expressed using a varying amount of facial NMSs (e.g. pressing eyes together, open mouth)

- **Content separation:** Lexical items are separated by a subtle nod of the head (e.g. my mother signed as ME/ MOTHER versus me and my mother signed as ME (+NOD)/ MOTHER)
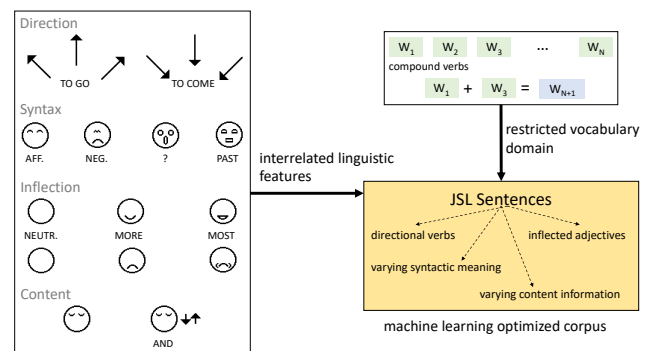

Figure 1: Signed expressions are difficult to convey in a natural way using virtual avatars. The fundamental structure of the corpus was designed specifically to enable a more reliable and accurate machine learning of contextual interrelation within JSL sentence expressions.

For use in deep neural network models, the previous constraints were used to define JSL sentences of large variations within every single linguistic feature and in between all features. In concrete, the main corpus was built as follows: to account for the dense domain criteria, 69 daily

conversation samples of 5-10 lexical items each were chosen as fundamental corpus sentence patterns from a set of intermediate-level SL practice sentences. These 69 sentence patterns were then modified to define 4 to 6 pattern variations by a random combination of dense-domain compound verbs and the previous linguistic features. The following sentences A, B, C and D for example constitute variations of one sentence pattern, whereas PT1 denotes a reference to oneself, PT2 denotes a pointing to the opponent person (respectively conversation partner) and PT3 denotes a pointing to a third person, object or place:

A PT3/ MOVIE/ BUILDING/ PT3/ INTERESTING/ MOVIE/ WATCH/ PAST (translating as 'I watched an interesting movie in this cinema.')

B MOVIE/ BUILDING/ PT3/ VERY INTERESTING/ MOVIE/ WATCH/ PAST/ PT2? (translating as 'You watched a very interesting movie in the cinema?')

C PT3/ MOVIE/ BUILDING/ PT3/ PT1/ SATO/ MAN/ MOVIE/ WATCH/ NO (translating as 'I did not watch an interesting movie in this cinema with Mr. Sato.')

D MOVIE/ BUILDING/ PT3/ PT1/ SATO/ WOMAN/ PRETTY INTERESTING/ MOVIE/ WATCH/ PAST (translating as 'I watched a pretty interesting movie in the cinema with Ms. Sato.')

Utilizing this sentence pattern variation strategy, a collection of 430 JSL sentences were defined to build the basic framework for corpus collection. These sentences constitute natural JSL expressions as expressed by native signers and therefore contain a very unbalanced word distribution: especially lexical items that support the semantic understanding within a signed sentence such as PT3 (providing contextual references to objects and persons) occur frequently within a JSL sentence. To reduce their impact on the network learning, we defined an additional set of 260 short training phrases. These phrases were built from the basic sentences and constituted a semantically meaningful succession of 3 to 6 less frequent words. Phrase variations taken from the sample sentences A, B, C and D were for example INTERESTING/ MOVIE/ WATCH/ PAST, MOVIE/ WATCH/ NO and SATO/ WOMAN/ PRETTY INTERESTING/ MOVIE/ WATCH.

4. Corpus Acquisition

All basic sentences and phrases were signed between one to three times in varying speed and sentence intonation by one Child Of Deaf Adults (CODA), leading to 931 corpus sentences and 502 corpus phrases with varying spatial and semantic content information. The data acquisition was furthermore assisted by a deaf native JSL speaker who supervised the grammatical correctness and naturalness of the signed corpus content.

During signing, the displacement of 123 markers attached to the signer's body was captured using an optical Vicon camera system of 48 cameras with a sampling rate of 120Hz. Additional data for extensive corpus annotation or the learning of sentence recognition networks was acquired using a Microsoft Kinect (sampling rate 30Hz) and

a consumer video camera (sampling rate 60Hz). All capture modalities were synchronized via an external trigger and recorded a total of 10.384 signed utterances within a vocabulary of 197 lexical items.

To ensure sufficiently dense data for subsequent natural signing avatar generation, 92 out of the 123 optical motion capture markers were utilized for the acquisition of detailed finger and facial movements such as the blinking of eyes or the raise of eyebrows (Figure 2). These markers were of 3mm size, with 24 markers placed on each hand up to the wrist and 44 makers attached to the face of the signer. All Vicon recordings were post-processed to ensure correct marker labeling and to eliminate missing frames caused by marker occlusions during data recording.
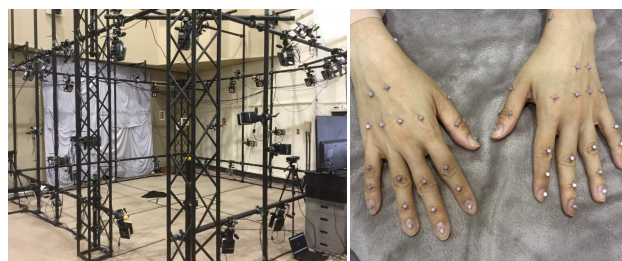

Figure 2: A set of 48 optical cameras was used to record the signed motion sentences. Marker were densely placed on body, finger and face of the sign speaker to obtain a highly-dimensional collection of sign motion data.

The cleaned data was made available in C3D and BVH format, which are two common data formats for the storage and processing of optical motion capture data. Whereas C3D contains the raw three-dimensional point clouds of all marker positions as obtained during the motion capture process, BVH contains kinematic information of a virtual character's body joints. To obtain such higher-level data, the raw C3D data was rigged onto the skeleton of a virtual avatar (Figure 3) and the three-dimensional joint position and rotation of 107 relevant skeleton joints (including finger joints and controllers for facial expression) saved as the file's main motion data streams. For the given problem, the BVH format should be considered particularly useful since newly generated motion streams can be directly transferred onto the corresponding virtual avatar for visual display and evaluation.

5. Corpus Annotation

The fundamental content of the deep JSLC was annotated in gloss notation with the support of the signer and the supervising native speaker during the process of corpus definition. Throughout data acquisition, the gloss notation was actively refined in real-time to ensure that every sentence was expressed in the most natural way of signing. Furthermore, every corpus sentence was annotated as a visual markup with the help of the JS Pad dictionary (Lab, 2016) for the creation of Japanese Sign Writing (Matsumoto et al., 2009) and the additionally captured video data. A sample annotation of the previous sentence variation A is shown in both gloss notation and visual markup in Figure 4.
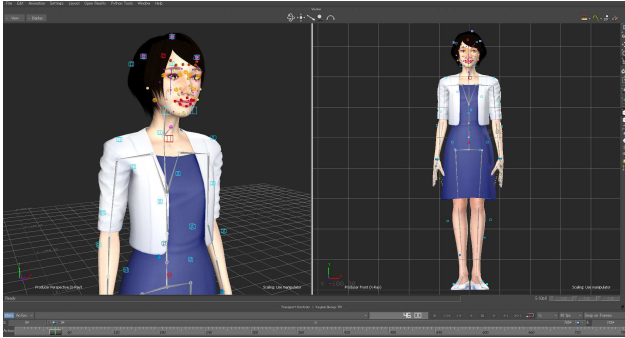
Figure 3: 3D model used to rig the raw motion capture data onto an avatar skeleton to provide joint position and rotation data in BVH format.

A major advantage of deep sequence generation models such as Recurrent Neural Networks (RNNs) is that they do not require the training data to be pre-segmented (Graves, 2013). However for eventual use in baseline networks or automatic sentence segmentation models, additional time annotations for separation of all lexical items in the corpus were determined. These were based on changes in hand and finger shapes as well as motion directions using the synchronized video data.

5.1. Encoding

Three different types of encoding were chosen for subsequent corpus deployment in potential neural network models. Here, the idea was to provide different types of encoding to evaluate whether the presence of specific linguistic feature information could reproduce non-manual signing aspects in a better way.

The first encoding constitutes a simple one-hot encoded representation of all occurring lexical items and does not contain any additional information on the linguistic features incorporated within the deep JSLC. The second encoding constitutes the simple one-hot encoded representation plus additional information on all linguistic features as additional elements of the one-hot encoded vectors for each respective sign. In concrete, these additional vector elements represent information on:

1. Use of left or right hand to convey primary sentence meaning

2. Start location of the primary hand within a position segment defined in relation to the upper body

3. End location of the primary hand within a position segment defined in relation to the upper body

4. Signing in active or passive form (if applicable, else none)

5. Stage of adjective inflection (if applicable, else non)

6. Inclusion of interrogation (if applicable, else non)

7. Inclusion of negation (if applicable, else non)

The third encoding was based on the visual markup annotation, following the representation of Sign Writing Markup Language (SWML) (Costa et al., 2001). Here, every lexical item was encoded as the combination of its Sign Writing components in SWML. For example the lexical item MOVIE is built by one head icon with the SWML index 04-01-001-01-01-01, two handshape icons (left and right hand) with the SWML indices 01-05-001-01-01-03 and 01-05-001-01-01-11 and two directional icons (for left and right hand) with the SWML indices 02-03-006-01-01-13 and 02-03-006-01-02-01.

6. Corpus Deployment

We evaluated the corpus usability for the learning of JSL sentence structure with a straightforward modification of the sequence to sequence model (Seq2Seq) for English-French translation (Sutskever et al., 2014). However, the determination of generated sequence quality is a difficult task that is commonly performed by rigging the generated sequences on a virtual character, and by subsequently assessing their naturalness and understandability in user studies. For this reason, we used the corpus in a reversed recognition scenario here: the acquired motion data streams were first encoded by a RNN cell and then passed to a decoder RNN cell providing an output expectation of the expressed sentence. Since the Seq2Seq model is bidirectional, evaluation of corpus efficiency can be expected to also hold valid for generation scenarios.

Using a smaller subset of 810 of the full corpus sentences only, we first learned several variations of a basic Seq2Seq network with 1 to 3 hidden layers and a varying number of cells per layer for all three encodings. Here, it should be noted that deep networks are commonly trained on much larger data sets. However since SL data collections cannot be acquired as easily as text or image data, the number of available training data can already be considered numerous for the given data content. The optimizer used during network training was an Adagrad optimizer, and both Gated Recurrent Units (GRUs) and Long Short-Term Memory Units (LSTMs) were used as cell types. Results indicate a constant decrease of training loss over time within all network models, whereas the Sign Writing based encoding performed slightly better than the two simpler, gloss based encodings. Best results with maximal test accuracy of $\approx 20\%$ after 2000 training epochs were achieved with 1 hidden unit of 256 LSTM cells (Figure 5). However, all network architectures showed significant overfitting and did not generalize well on unknown test data: especially rare words were easily misclassified and labeled as frequent words of little discriminative character (e.g. PT3, MAN or WOMAN). Given the unbalanced word distribution within JSL sentences, this is not surprising, and a better balanced corpus should improve network quality considerably.

To test eventual effects of an enlarged data collection, we included the 502 additional phrases in the training data and learned a new recognition network using the same encoding and network parameter in the next step. Test accuracy of the network raised to $\approx 40\%$ after 2000 training epochs, while overfitting was significantly decreased. The subset of underrepresented word phrases that can be freely added to

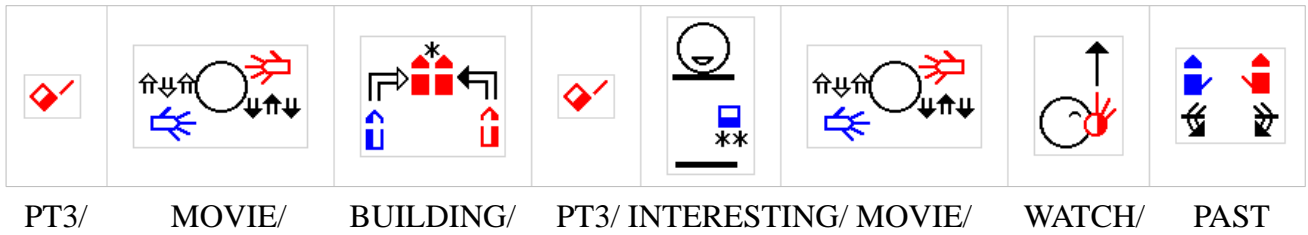| PT3/ | MOVIE/ | BUILDING/ | PT3/ | INTERESTING/ | MOVIE/ | WATCH/ | PAST |

Figure 4: Visualization of previously introduced sentence variation A in Japanese Sign Writing and its corresponding gloss annotation.
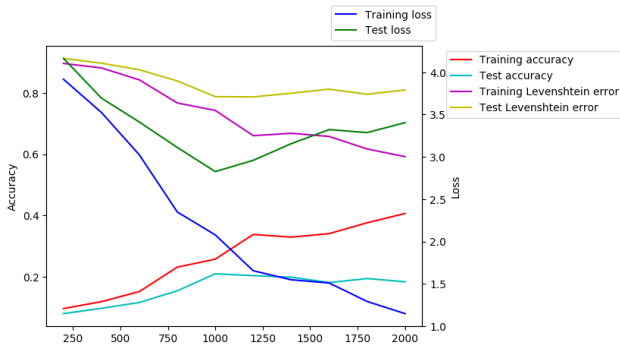

Figure 5: Evolution of training and testing accuracy and loss for sentence recognition as obtained with a standard sequence to sequence model and 1 hidden layer of 256 LSTM cells for a smaller subset of the full corpus sentences only. Over 2000 epochs, a recognition network of $\approx 20\%$ test accuracy was learned. The network is strongly overfitting.
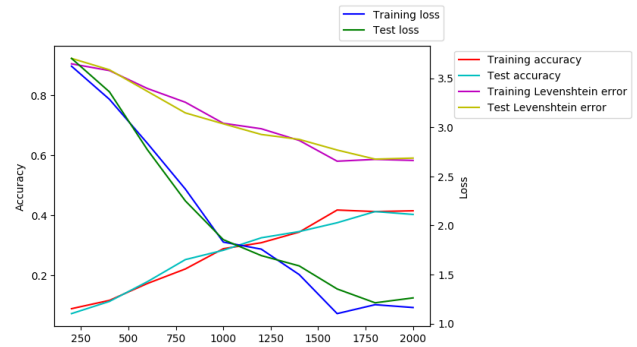

Figure 6: Evolution of training and testing accuracy and loss for sentence recognition as obtained with a standard sequence to sequence model and 1 hidden layer of 256 LSTM cells using a larger number of corpus data for network training. Over 2000 epochs, a recognition network of $\approx 40\%$ test accuracy was learned that generalizes better to unknown data.

the main training data should therefore be considered as an useful extension of the full sentence collection.

7. Discussion

Previous results suggests that the presented corpus is generally capable to train a Seq2Seq network that understands common multi-modal interrelation within JSL utterances. Test accuracies of $\approx 40\%$ do not appear sufficient for application in real-life scenarios yet, but reach the best obtained accuracies of similarly continuous and weakly supervised sentence recognition scenarios (Koller et al., 2016). To date, no specific modifications of the model parameters were performed, and we expect to achieve better results of improved accuracy and smaller loss by adding a suitable data embedding and an attention model. Moreover, it was shown that a higher number of training data is beneficial for network learning. Better recognition and generation networks should therefore be achieved by further augmenting the corpus size and balancing out the general word distribution. Thanks to the corpus design with its repetitive occurrence of identical sentences and phrases, respective data can be synthesized relatively easily from the existing data in the following, using sequence alignment methods such as squeezing, stretching or undersampling.

All in all, we believe that the specific characteristics of its corpus design and content make the present deep JSLC a very valuable collection of JSL motion data. It shall now

be used to define suitable network parameters and variations such as attention models, and to subsequently learn a wide variety of sequence generation networks. In a last step, the usability and eventual benefit of the trained generation networks shall be evaluated with respect to realism and naturalism of the resulting animations.

8. Conclusion

We presented a new corpus of JSL sentence expressions for application in advanced data-driven deep neural networks. This corpus was defined so that it can easily be applied to advanced sequence generation models for the synthesis of Sign Language animations. As opposed to previous SL corpora of similar application purpose, the corpus was built from randomized variations of pre-defined sentence patterns only. It incorporates many spatial and temporal references as well as non manual signs to intrinsically learn interrelations of relevant linguistic features within signed expressions or conversations. The corpus is extensively annotated in gloss and visual mark up, and its signed data content made available using three different motion sensing modalities (motion capture, depth images and video images) that can be utilized in various additional corpus works. First experiments showed the general applicability of the presented corpus in sequence to sequence networks for sentence recognition. In the following, these networks

shall be enhanced and modified to provide intelligent networks that can help to generate naturally signing avatars in the future.

9. References

Costa, A. R., Dimuro, G. P., et al. (2001). Supporting deaf sign languages in written form on the web. In *10th World Wide Web Conference, Hong Kong*. Home Page of Web and Society Track, `http://www10.org/cdrom/posters/frame.html`.

Ebling, S. and Huenerfauth, M. (2015). Bridging the gap between sign language machine translation and sign language animation using sequence classification. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*.

Gibet, S., Lefebvre-Albaret, F., Hamon, L., Brun, R., and Turki, A. (2016). Interactive editing in french sign language dedicated to virtual signers: requirements and challenges. *Universal Access in the Information Society*, 15(4):525–539.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Huenerfauth, M. and Hanson, V. (2009). Sign language in the interface: access for deaf signers. *Universal Access Handbook. NJ: Erlbaum*, 38.

Huenerfauth, M. (2008). Generating american sign language animation: overcoming misconceptions and technical challenges. *Universal Access in the Information Society*, 6(4):419–434.

Kacorri, H. and Huenerfauth, M. (2016). Selecting exemplar recordings of american sign language non-manual expressions for animation synthesis based on manual sign timing. In *Proc. SLPAT 2016 Workshop on Speech and Language Processing for Assistive Technologies*, pages 14–19.

Kacorri, H., Huenerfauth, M., Ebling, S., Patel, K., and Willard, M. (2015). Demographic and experiential factors influencing acceptance of sign language animation by deaf users. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '15, pages 147–154, New York, NY, USA. ACM.

Kipp, M., Heloir, A., and Nguyen, Q. (2011a). Sign language avatars: Animation and comprehensibility. In *Intelligent Virtual Agents*, pages 113–126. Springer.

Kipp, M., Nguyen, Q., Heloir, A., and Matthes, S. (2011b). Assessing the deaf user perspective on sign language avatars. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '11, pages 107–114, New York, NY, USA. ACM.

Koller, O., Ney, H., and Bowden, R. (2016). Deep hand: How to train a cnn on 1 million hand images when your data is continuous and weakly labelled. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3793–3802.

Lu, P. and Huenerfauth, M. (2011). Data-driven synthesis of spatially inflected verbs for american sign language animation. *ACM Transactions on Accessible Computing (TACCESS)*, 4(1):4.

Matsumoto, T., Kato, M., and Ikeda, T. (2009). Jspad: A sign language writing tool using signwriting. In *Proceedings of the 3rd International Universal Communication Symposium*, IUCS '09, pages 363–367, New York, NY, USA. ACM.

Motion Savvy. (2017). UNI sign language translation tablet. `http://www.motionsavvy.com/index.html`. Accessed: 2017-09-21.

Nagashima, Y., Watanabe, K., Terauchi, M., Kato, N., Uchida, T., Umeda, S., Miyazaki, T., Azuma, M., and Hiruma, N. (2016). A support tool for analyzing the 3d motions of sign language and the construction of a morpheme dictionary. In *International Conference on Human-Computer Interaction*, pages 124–129. Springer.

Oord, A. v. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*.

Pro Deaf. (2017). Libras sign language translator. `http://prodeaf.net/en-us`. Accessed: 2017-09-21.

Sun, C., Zhang, T., Bao, B.-K., Xu, C., and Mei, T. (2013). Discriminative exemplar coding for sign language recognition with kinect. *IEEE Transactions on Cybernetics*, 43(5):1418–1428.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Traxler, C. B. (2000). The stanford achievement test: National norming and performance standards for deaf and hard-of-hearing students. *Journal of deaf studies and deaf education*, 5(4):337–348.

Umeda, S., Uchida, T., Azuma, M., Miyazaki, T., Kato, N., and Hiruma, N. (2016). Automatic production system of sign language cg animation for meteorological information.

10. Language Resource References

Matsumoto Lab. (2016). *Japanese Sign Writing Dictionary*. Gifu University, JSPad executable, 0.3.1.