A Plan for Developing an Auslan Communication Technologies Pipeline

Jessica Korte[1[0000−0002−4412−7199]], Axel Bender[2], Guy Gallasch[2], Janet Wiles[1[0000−0002−4051−4116]], and Andrew Back[1[0000−0001−5474−1910]]

[1] The University of Queensland, Australia {j.korte,j.wiles,a.back}@uq.edu.au
[2] Defense Science and Technology, Australia
{Axel.Bender,Guy.Gallasch}@dst.defence.gov.au

Abstract. AI techniques for mainstream spoken languages have seen a great deal of progress in recent years, with technologies for transcription, translation and text processing becoming commercially available. However, no such technologies have been developed for sign languages, which, as visual-gestural languages, require multimodal processing approaches. This paper presents a plan to develop an Auslan Communication Technologies Pipeline (Auslan CTP), a prototype AI system enabling Auslan-in, Auslan-out interactions, to demonstrate the feasibility of Auslan-based machine interaction and language processing. Such a system has a range of applications, including gestural human-machine interfaces, educational tools, and translation.

Keywords: Auslan, Australian Sign Language, sign language recognition, sign language production, sign language processing

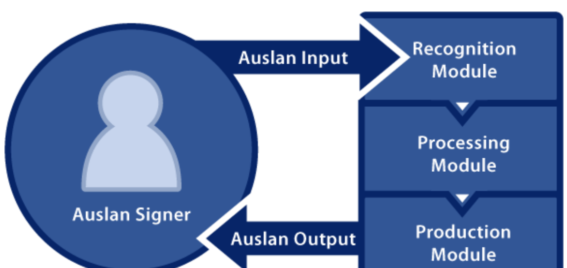

Fig. 1. The proposed Auslan Communication Technologies Pipeline architecture. Each module represents a segment of functionality which requires the development of specific AI approaches: the *Recognition Module*, intended to recognise sign input, requires development of Auslan recognition as a multimodal language; the *Processing Module*, which provides the functionality of the pipeline, requires Natural Sign Language Processing; and the *Production Module*, which will output Auslan signing, requires human-like production of sign output that is acceptable to the Deaf community. Each of these modules will interface with the next module in the Pipeline to deliver a functional application

1 Introduction

While mainstream spoken languages, with large corpora of written and spoken data, are seeing a surge in the development of AI tools for recognition, translation, production and processing, such tools for sign languages are lacking. From a perspective of equity, sign languages should also have access to such AI tools, to support the access of Deaf and other signers to communication-supporting technologies; and the field is on the cusp of having the technical ability to develop such tools. However, the language processing approaches developed to date are largely not suitable for processing visual-gestural languages with relatively limited datasets, as is the case with many sign languages around the world, including Auslan.

Much of the AI research on sign language recognition to date has been in the field of image or gesture classification. Such automatic gesture classification can provide basic gesture recognition, but experiences problems including robustness, performance under "noisy" real-world conditions, and variability from user to user. We propose a new AI framework for multimodal visual-gestural language recognition, incorporating various nuances of lexical signs, visual representation, gestures, body language, and facial expression as sub-language components within the newly developed approach known as entropy-based AI, drawing on both image classification and language processing fields. This project has significant potential for sign language and gesture recognition, and sign language-based human-machine interfaces.

This paper presents a plan for the Auslan Communication Technologies Pipeline (Auslan CTP), a module-based system for recognising, processing and producing Auslan communication. The modular nature of the system allows for flexibility of purpose, and "plug-and-play" replacement of components as improvements are made to each module. Fig. 1 above shows the components of the pipeline.

2 Sign Language Recognition

"Gesture recognition" has been explored in the context of pre-defined non-language gestures [1, 27, 35] (including military-inspired gesture sets [6, 21, 29]), and various sign languages (e.g. Auslan [11, 18, 19]) using techniques including computer vision [5, 21, 27, 28], sensor-based wearables [5, 6, 18, 19, 22, 29], Markov models [11, 22, 24, 28, 31] and neural networks [4, 23, 25, 26]. These approaches require extremely large training sets with multiple examples of a single gesture or sign to learn to recognise input [4]; therefore, the scope of such work has been restricted to date to small vocabularies of strictly lexical signs [5, 11, 18, 19, 22, 28, 31], such as fingerspelling signs [28]. This can result in high accuracies over the known vocabulary and known signers [11, 18, 19], but lacks real-world applicability, as accuracy decreases due to inter-signer variation [17] in expression and sign formation in a wider population of signers [19], and the limited vocabularies only include lexical signs in standard forms, where natural signed discourse relies on both non-lexical signs [15] and situational use of space [3]. Signers will enact information that is not encoded in a particular sign, e.g. miming looking at a phone [15]; or using areas of signing space to represent referents. Such enactment does not use conventionalised signs but conveys meaning in relation to the context of the signed conversation which is easily intelligible to a human collocutor [15], but which is difficult to classify through machine learning (ML).

There is one Auslan corpus which captures enactment [13], collected and annotated by Johnston and colleagues for linguistic research [14], but it is not fully curated for machine learning, and cannot be made public as it contains sensitive content and signers are identifiable.

We propose that a system for recognising Auslan in real-time must differentiate:

1. Handshape, orientation and location of static signs;
2. Handshape, orientation, location and movement of dynamic signs;
3. Facial expression and body language, and their relevance to a particular sign [16];
4. Fully lexical, semi-lexical and non-lexical signing, and their meanings;
5. Transitions between signs; and
6. Individual variability of sign production.

We are planning to address points 4 and 5 through movement entropy, as this should change between signs or sign components [30, 33], allowing for automated segmentation. Entropy-based AI [2] is a new methodology to address points 4 and 6, which in effect, is using entropy as a measure of the characteristics of the signs.

2.1 Entropy-Based AI

While image based recognition of gesture can be performed well in situations when the visibility and related conditions are ideal, in real world situations there can be significant challenges. Hence, one aspect of this work is to consider information theoretic approaches which seek to combine probabilistic information about the signs in order to assist recognition.

For example, a static image or series of images may be insufficient to differentiate signs, especially if visibility is poor; but when aided by movement, then this can make our recognition task considerably easier. While various forms of image recognition approaches have been adopted in the past based on sequential images over time, in this stage, we are aiming to include a measure of how sign movement can assist in recognition.

The basic idea here is to treat sign language as multimodal language rather than an image classification problem. An example of the potential effectiveness of this approach can be understood as follows.

Suppose that instead of treating gesture recognition as a single static image, we now treat it as a short synthetic narrative, where the “words” are each submovements within each sign. In the same way that humans can recognise meaning from incomplete or slightly incorrect language input (for example, a signer using the wrong handshape, but correct location and movement; or starting location being obscured), we can now consider a sign in terms of these “synthetic words”. For example, suppose we have a set of symbols {A,B,C,D,E,F} - which might correspond to a subset of handshapes and locations - then a given sign might be represented usually by a sequence of [D,A,C,B] with corresponding entropy. For incomplete or obscured input, such as a sequence [D,A,x,B] with corresponding entropy, we can provide an estimate of the most likely sign based on the observed input sequence of synthetic features.

If we compute the probabilistic structure of these “words” for each sign being observed, then in terms of the overall entropy, it is expected that there will be a particular probabilistic structure for each type of sign. Hence, the entropy can be used to assist in identifying signs, and it can also assist in identifying when the movements are “surprising” or “unexpected”, thereby indicating that this may be in fact, a different gesture than the sign being recognized by image recognition. This approach is somewhat similar to the use of short-term prediction used in dictionary based systems like predictive text. Here however, we are coupling it with image recognition to provide a richer framework, either confirming the trust in the image, or perhaps, indicating that there is an issue with the observed image. For example, a gesture might appear to be one thing, but the way in which it was formed over the short time it was “constructed”, indicates that it cannot be conclusively relied on. We believe that this ability to identify “surprising” or “untrustworthy” input can enable us to identify semi-lexical and non-lexical signs, as well as potentially to differentiate non-linguistic gestures.

Furthermore, the approach suggested here can provide additional insight into improving recognition accuracy under conditions of poor visibility or duress. The image recognition part might indicate that the sign could be one of several possible signs, but when we use the ranking of the entropy-based model, then this could indicate the most likely gesture overall.

This entropy-based AI approach calls for a symbolisation of the input space into a “synthetic” entropy-based language. In current work to date, this does not seem to require an optimal process, provided the entire input space is accounted for. The symbolization will segment various sign elements (lexical signs and sign segments, visual representation, handshapes, body language and facial expression) into micro-features, for example, small movements over time. These small features become the basic symbolic building blocks, i.e. “synthetic letters” within our new synthetic language framework. These synthetic letters will then be used to form synthetic words with particular probabilistic structure, for example, adopting a multidimensional N-gram model, and hence it becomes possible to develop this richer approach to not only gesture recognition, but also to introduce a degree of robustness, enabling identification through a multimodal, information theoretic mechanism. This is anticipated to provide more stable communication.

2.2 Open Questions

Treating Auslan as a full, visual-gestural language with meaning encoded in lexical, semi-lexical and non-lexical signing raises the following questions:

1. How can entropy-based AI methods support accuracy and trustworthiness of sign classification, especially considering inter-signer variability?
2. How can entropy-based AI approaches be used to segment sign movements at all levels of lexicality?
3. How can time- or sequence-based ML approaches inform Auslan recognition?
4. What are the requirements for a dataset of Auslan data for use in machine learning to recognise signs at all levels of lexicality? This includes considerations of:
   (a) file formats, size and resolution;
   (b) approaches to encoding data for machine learning use;
   (c) sourcing data in a post-COVID19 world; and
   (d) transferability of lessons from spoken language ML corpora and/or linguistics corpora.
5. How can an Auslan processing system recognise and respond appropriately to body language communication?
6. How can incremental learning, zero-shot learning, or other similar machine learning paradigms allow for extensibility of Auslan recognition?

2.3 Research Approach

Developing a new framework of trustworthy Auslan recognition requires a comprehensive data set. Hence, the first step is to obtain a sign language dataset. This will be done by collecting Auslan data in video and depth formats from expert and native signers. Deaf signers are as expressive and individual as speakers of any language, and they are experts in the use of visual-gestural communication. Their involvement in the project places them, as sign language experts, in the position of deciding what to communicate, and how to "write" their language into technology, including in terms of determining the approach to encoding. It is expected that data encoding will involve some combination of: handshape; hand orientations; start, end and/or key locations; movements; expressions; mouthing and facial movements; linguistic glosses; HamNoSys or SignWriting encoding; English translation; dialect; signer fluency; and clarity of signing. Once coding has begun, we plan to create machine learning sub-modules to automate some elements of encoding, such as a machine learning model which attempts to identify and recommend handshapes.

Once coded, the dataset will be used to develop the experimental system for evaluating proposed models. In the first instance and for baseline reference purposes, a machine learning model will be implemented for sign language classification. The main aspects of the proposed model will then be developed using the entropy-based AI framework where the first step is as follows:

1. Formulate the sign elements,

2. Determine the entropy characteristics of the sign language,
3. Examine the probabilistic characteristics of Auslan data at all levels of lexicality,
4. Develop the architectural framework of the Entropy-based AI system for sign language and gesture recognition.

A possible machine learning architecture is shown in Fig. 2.

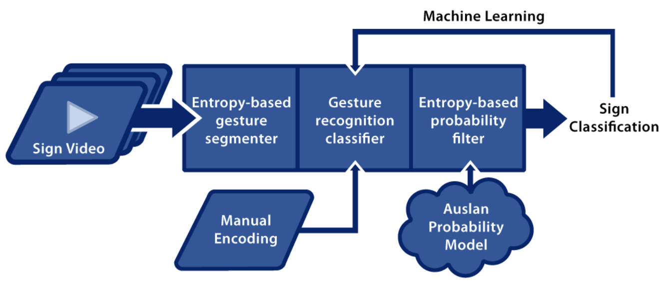

Fig. 2. Possible architecture of entropy-based Auslan recognition machine learning system

3 Virtual Sign Language Production

Virtual production of sign languages is of significant strategic importance as a basis for two-way communication between humans and machine agents. Typically sign production uses one of 3 approaches:

1. encoding of sign representation for automated, usually avatar-based, production e.g. [8, 36, 38];
2. pre-rendering video or animation [20] of fixed messages; or
3. AI techniques, e.g. Generative Adversarial Networks [32].

Each approach can have problems in clarity, comprehension and trustworthiness of produced sign. Automated avatar production (approach 1) can lack natural expressions, body movements and non-manual sign aspects [20, 38], and is discouraged by the World Federation of the Deaf and the World Association of Sign Language Interpreters for live interpretation [39]. Pre-rendered video or animation (approach 2) is fixed and cannot be altered quickly or inexpensively [12, 37, 38]. GAN production (approach 3) may resemble a real signer from the dataset [32], which could raise reputational issues.

This research proposes to develop, in consultation with the Australian Deaf community, a framework for virtual human-like sign and gesture production based on entropy-based AI, which has the potential to provide natural behaviours. It is expected to extend from one or more of the three known approaches, and to include guidance for usage approved by the Deaf community.

If all three approaches are found unacceptable by the Deaf community, other avenues such as social robots could be explored; or new approaches could be co-designed with Deaf design partners.

3.1 Open Questions

Generating human-like virtual Auslan signs raises the following questions:

1. What approaches to sign production are acceptable to the Australian Deaf community?
2. From a Deaf user perspective, what are the key issues to be addressed in developing a real-time virtual Auslan production system which is human-like?
3. How can an Auslan dataset be used to generate probabilistic symbolic encodings which can be adapted to form the basis for a virtual Auslan production system?
4. How can an Auslan dataset be used to generate Auslan videos in real-time without co-opting the image of a real signer?
5. How can signs with varying degrees of conventionalisation (lexical, semi-lexical, and non-lexical) be encoded for virtual Auslan sign production?
6. What are the notation requirements for a human-like real-time virtual Auslan production system when used in a human-like framework using entropy-based AI?
7. What probabilistic encoding framework or symbolization can be used for capturing natural expressions, emotions, body movements and other similar features?
8. How can human-like Auslan signing be constructed via the proposed framework?

3.2 Research Approach

To address the questions related to Auslan production, resources for producing or generating Auslan signing will be created. This is intended to allow future systems to be able to output signed communication in the form of human-like Auslan, incorporating elements identified by signers as important, including emotion and naturalistic movement.

Sign production (like speech production), can use a range of techniques, which include direct mapping from sign video to avatar production; generative modelling from sign video; a programmable avatar which generates signs based on sign notation; and a modular database of avatar clips or elements (e.g. signs, sign fragments, handshapes, facial expressions, etc) which could be concatenated for real-time generation of Auslan sign. Each has advantages and different requirements for effective use: video mapping techniques are directly usable; sign notation provides for generality; and a library of signs could be of general use in Auslan production. Working with Deaf community members, the feasibility and advantages/disadvantages of these approaches will be explored using the

Auslan dataset and model (developed in addressing questions of sign language recognition) to inform the design of a sign encoding system.

The Auslan production approach chosen will be used in a prototype system, which should be able to generate signs based on annotations, glosses or videos.

4 Natural Sign Language Processing

Natural language processing (NLP) systems typically rely on symbolised written languages [7]. As Auslan has no native writing, there is a need to consider exactly how symbolisation can be done, in terms of notation systems, grammars and other language constructs. Several sign notations have been created (e.g. HamNoSys [10], SignWriting [34]) but each has limitations, and none are widely used by the Australian Deaf community. HamNoSys receives some use by Auslan linguists [14]. Most sign language processing (SLP) to date has focused on translation, relying on a notation or glossing system (e.g. [9, 32]).

4.1 Open Questions

Sign Language Processing raises the following questions:

1. How well can existing NLP approaches be converted to work in SLP?
2. What are the requirements for encodings for an Auslan processing system?
3. Can the use of a video dataset reduce the need for written notation in SLP?
4. How can an Auslan processing system identify and process less conventionalised signed communication, such as enactment?
5. How can sentiment analysis or similar NLP techniques inform the emotional expression and body language of human-like, machine-produced Auslan signs?

4.2 Research Approach

This module is the central processing part of the Auslan Communication Technologies Pipeline, connecting the Auslan Recognition Module with the Auslan Production Module, by developing a method for computational processing of sign language and gesture.

For example, with a question like "How can an Auslan processing system identify and process less conventionalised signed communication, such as enactment?", the research approach would build on prior work around recognising and encoding non-lexical and semi-lexical communication, as well as drawing on existing NLP approaches to draw meaning from context, augmented by entropy-based AI approaches to contextual probability.

The scope and application of the processing system will be determined through consultation with Deaf community members. Options include: an Auslan chatbot, an Auslan digital assistant, an Auslan teaching tool, or a translation system. The choice of application will of course influence the system architecture. A possible chatbot system architecture is shown in Fig. 3.

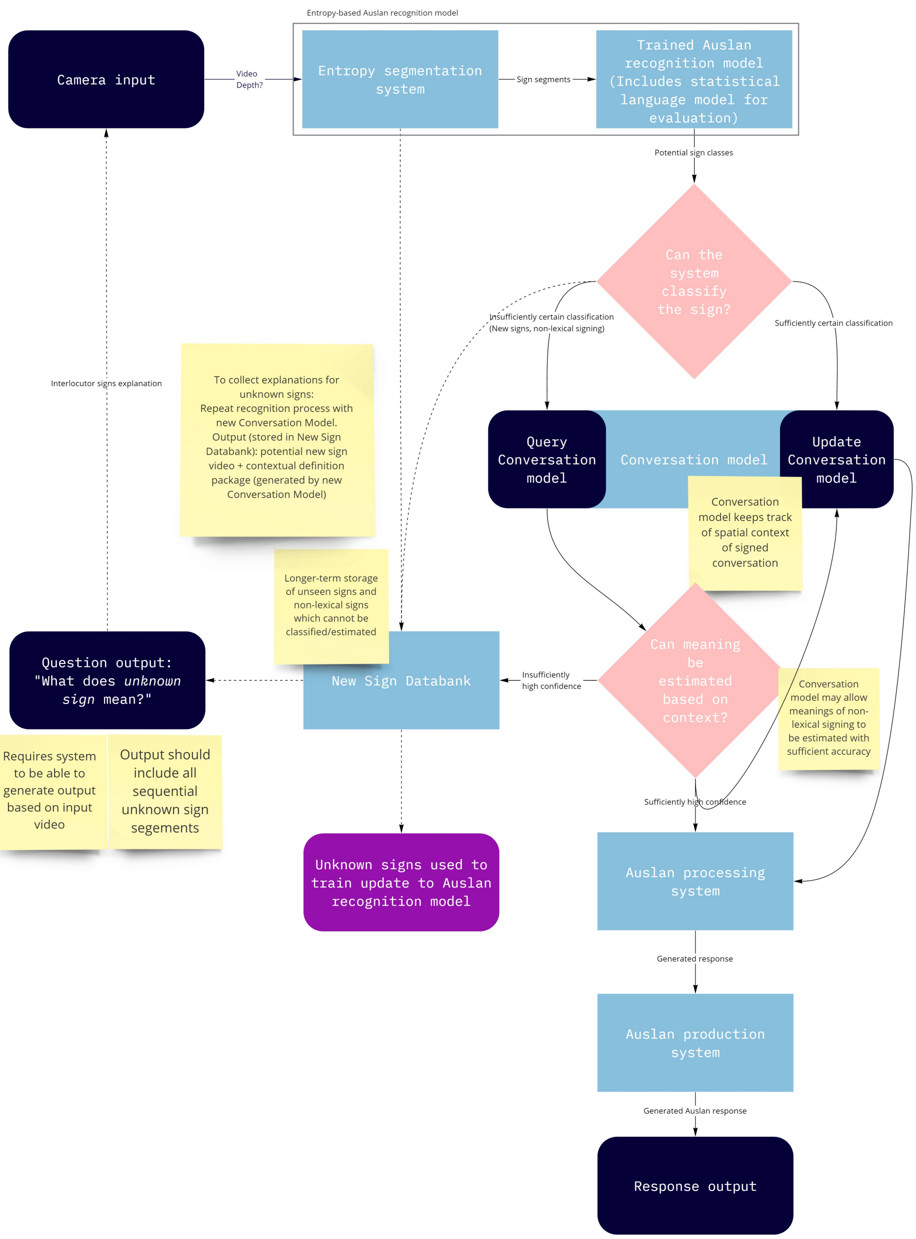

Fig. 3. Possible architecture for a chatbot. Such a system could include capabilities to collect definitions of unknown signs (*dotted arrows on left of diagram*)

5 Connection to communicative gesture research for social and operational robotics

The Auslan CTP research, and gestural human-machine interaction more broadly, has significant potential for use in co-operative and social robotics, through explicit gestural interfaces, and robots with human-like implicit awareness of body language and gesture. This research's focus on interaction via lexicalised and non-lexicalised signing as used by diverse individual signers provides a basis for machine recognition and understanding that could underlie human-robot non-verbal communication, as it may result in an approach to encoding multimodal language such that a robot could understand real-time messages communicated robustly in operational environments, with support for inter-signer variability in message production.

For example, the introduction of robotic and autonomous systems in the military domain has resulted in requirements for human-machine interaction that is robust in harsh operational conditions. In such conditions, it is not affordable and often not possible to communicate verbally; lives depend on the accurate interpretation of environmental cues and the effective and efficient communication with team members (whether human or robot). An important argument why Auslan is a good basis for human-robot interaction in the military domain is, firstly, that Auslan is a full visual-gestural language. Different levels of abstraction (symbolic through to semantic) can be communicated in the language and hence represented in the messages between human and robot. This is especially important in situations where context matters – which is the case in most military tactical settings.

Secondly, Auslan varies from individual to individual, in the same way that speakers of every language are individuals; i.e. every signer has preferred vocabulary, expressions and nuances of sign production. Exploring this variability is important for achieving the aim of natural human-robot interaction in a range of contexts, i.e. allowing humans to use individualised language in their interaction with a robot. Current gesture technology has a tendency to require strict and accurate adherence to a known set of gestures, resulting in a non-robust interaction modality or in the need to train the human in the precise execution of gestures. Breaking this paradigm of "changing the human" to get humans and robots to work together is particularly important in demanding contexts such as military operations where the human has to focus on many things concurrently, and may be under high levels of stress or otherwise distracted from executing precise visual-gestural commands.

Thirdly, use of a multimodal gestural language in social or military human-robot interaction allows for the development of communications technologies that don't rely on computer vision alone, i.e. robustness can be added through the fusion of vision, haptics, audio and other modalities. By extension, the research outlined in this paper may form the basis for more general encoding, able to capture signing, speaking, non-language gestures and abstract representations of the environment, for context-aware processing by autonomous systems.

Further outcomes of this project are a new framework for sign languages based on the newly emerging field of entropy-based AI which has shown significant results in other applications; and an open source Auslan Communication Technologies Pipeline. This will provide significant long term foundational benefits for developing sign language communication technologies. It will increase Deaf people's native-language access to technology and technology-mediated content, including in the vital domains of health and education. This could impact daily life for Auslan signers through educational uses, Auslan user interfaces, and automatic translation of pre-written digital content (pending sufficient levels of accuracy that satisfy the community).

6 SLRTP Relevance

The Auslan Communication Technologies Pipeline project has only recently begun, but the research plan shows promise. The authors of this paper are a multidisciplinary group of researchers and practitioners. We wish to attend the SLRTP workshop to gain feedback on the research plan and proposed software architectures; and importantly, to form connections with and learn from the workshop organisers and attendees.

By the time of the workshop, collection of the Auslan dataset and initial construction of the entropy-based Auslan recognition machine learning system should have commenced. The authors will therefore be able to contribute to technical discussions around dataset collection, curation and design; entropy-based AI; and machine learning approaches, from the perspective of the Auslan CTP project.

7 Conclusions

To date, communication technology pipelines have been developed for spoken languages, and machine learning techniques now have robust performance for speech to text and text to speech for several mainstream spoken languages. It is important for technology developers to recognise that these techniques will not generalise directly to sign languages, but rather that communication technologies for Auslan and other sign languages need new approaches based on the intrinsic properties of full, visual-gestural language with meaning encoded in lexical, semi-lexical and non-lexical signing. The Auslan CTP aims to bring together a multidisciplinary team spanning the Auslan community, practitioners and researchers in design, machine learning, linguistics and social robotics to develop new frameworks for multimodal visual-gestural language recognition and human-like sign production.

8 Acknowledgements

Many thanks to the project mentors, who provided guidance and feedback on this paper: Professor Trevor Johnston, Dr Adam Schembri, Dr Ashfin Rahimi and Associate Professor Marcus Gallaghe.

The research for this paper received funding from the Australian Government through the Defence Cooperative Research Centre for Trusted Autonomous Systems. The DCRC-TAS receives funding support from the Queensland Government.

References

1. Amir, A., Taba, B., Berg, D., Melano, T., McKinstry, J., Nolfo, C.D., Nayak, T., Andreopoulos, A., Garreau, G., Mendoza, M., Kusnitz, J., Debole, M., Esser, S., Delbruck, T., Flickner, M., Modha, D.: DVS128 Gesture Dataset (2017), http://www.research.ibm.com/dvsgesture/
2. Back, A.D., Angus, D., Wiles, J.: Transitive Entropy - A Rank Ordered Approach for Natural Sequences. IEEE Journal on Selected Topics in Signal Processing **14**(2), 312–321 (2019). https://doi.org/10.1109/JSTSP.2019.2939998
3. Bavelier, D., Corina, D.P., Neville, H.J.: Brain and Language: a Perspective from Sign Language. Neuron **21**(August), 275–278 (1998)
4. Bowden, R., Zisserman, A., Kadir, T., Brady, M.: Vision based interpretation of natural sign languages. In: Proceedings of the 3rd International Conference on Computer Vision Systems (2003), http://info.ee.surrey.ac.uk/Personal/R.Bowden/publications/icvs03/icvs03pap.pdf
5. Brashear, H., Henderson, V., Park, K.H., Hamilton, H., Lee, S., Starner, T.: American sign language recognition in game development for deaf children. Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility - Assets '06 p. 79 (2006). https://doi.org/10.1145/1168987.1169002, http://portal.acm.org/citation.cfm?doid=1168987.1169002
6. Ceruti, M.G., Dinh, V.V., Tran, N.X., Van Phan, H., Duffy, L.R.T., Ton, T.A., Leonard, G., Medina, E., Amezcua, O., Fugate, S., Rogers, G.J., Luna, R., Ellen, J.: Wireless communication glove apparatus for motion tracking, gesture recognition, data transmission, and reception in extreme environments. Proceedings of the ACM Symposium on Applied Computing pp. 172–176 (2009). https://doi.org/10.1145/1529282.1529320
7. da Rocha Costa, A.C., Dimuro, G.P.: SignWriting and SWML: Paving the Way to Sign Language Processing. In: Traitement Automatique des Langues Naturelles (TALN). Batz-sur-Mer, France (2003)
8. Efthimiou, E., Sapountzaki, G., Karpouzis, K.: Developing an e-Learning platform for the Greek Sign Language. In: Lecture Notes in Artificial Intelligence. pp. 1107–1113. No. 2000, SpringerVerlag (2004). https://doi.org/10.1.1.100.2567, http://www.springerlink.com/index/KAQ8UGLRQ2TMVYCF.pdf
9. Elliott, R., Glauert, J.R., Kennaway, J.R., Marshall, I., Safar, E.: Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. Universal Access in the Information Society **6**(4), 375–391 (2008). https://doi.org/10.1007/s10209-007-0102-z

10. Hanke, T.: HamNoSys - Representing sign language data in language resources and language processing contexts. In: LREC 2004, Workshop proceedings: Representation and processing of sign languages. pp. 1–6 (2004), http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HankeLRECSLP2004_05.pdf
11. Holden, E.J., Lee, G., Owens, R.: Australian sign language recognition. Machine Vision and Applications **16**(5), 312–320 (2005). https://doi.org/10.1007/s00138-005-0003-1
12. Huawei: StorySign: Helping Deaf Children Learn to Read (2018), https://consumer.huawei.com/au/campaign/storysign/
13. Johnston, T.: Auslan Corpus (2008), https://elar.soas.ac.uk/Collection/MPI55247
14. Johnston, T.: Auslan Corpus Annotation Guidelines. Tech. rep., Macquarie University & La Trobe University, Sydney and Melbourne Australia (2016), http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_February2016.pdf
15. Johnston, T.: Wrangling and structuring a sign-language corpus: The Auslan Dictionary. Presentation at CoEDL Fest 2019 (2019)
16. Johnston, T., Schembri, A.: Australian Sign Language (Auslan): An Introduction to Sign Language Linguistics. Cambridge University Press, Cambridge, UK (2007)
17. Johnston, T., Schembri, A.: Variation, lexicalization and grammaticalization in signed languages. Langage et société **1**(131), 19–35 (2010)
18. Kadous, M.W.: Auslan sign recognition using computers and gloves. In: Deaf Studies Research Symposium (1998). https://doi.org/10.1.1.51.3816, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.3816
19. Kadous, W.: GRASP: Recognition of Australian Sign Language Using Instrumented Gloves. Ph.D. thesis, The University of New South Wales (1995)
20. Kipp, M., Nguyen, Q., Heloir, A., Matthes, S.: Assessing the Deaf User Perspective on Sign Language Avatars. In: The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility. pp. 107–114. ACM, Dundee, Scotland, UK (2011). https://doi.org/10.1145/2049536.2049557
21. Lebron, J.: Recognizing Military Gestures: Developing a Gesture Recognition Interface. Tech. rep., Union College, Schenectady, NY, USA (2013), http://orzo.union.edu/Archives/SeniorProjects/2013/CS.2013/
22. Li, Y., Chen, X., Zhang, X., Wang, K., Wang, Z.J.: A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data. IEEE Transactions on Biomedical Engineering **59**(10), 2695–2704 (2012). https://doi.org/10.1109/TBME.2012.2190734
23. Liao, Y., Xiong, P., Min, W., Min, W., Lu, J.: Dynamic Sign Language Recognition Based on Video Sequence With BLSTM-3D Residual Networks. IEEE Access **7**, 38044–38054 (2019). https://doi.org/10.1109/ACCESS.2019.2904749, https://ieeexplore.ieee.org/document/8667292/
24. Ong, S.C.W., Hsu, D., Lee, W.S., Kurniawati, H.: Partially Observable Markov Decision Process (POMDP) Technologies for Sign Language based Human-Computer Interaction. In: Proceedings of the International Conference on Human-Computer Interaction (2009)
25. Parton, B.S.: Sign language recognition and translation: A multidisciplined approach from the field of artificial intelligence. Journal of Deaf Studies and Deaf Education **11**(1), 94–101 (2006). https://doi.org/10.1093/deafed/enj003
26. Pigou, L., Dieleman, S., Kindermans, P.J., Schrauwen, B.: Sign language recognition using deep convolutional neural networks. In: Lecture Notes in Computer Science. pp. 572–578. Springer, Zürich, Switzerland (2019). https://doi.org/10.1007/978-3-319-16178-5_40

27. Pisharady, P.K., Saerbeck, M.: Recent methods and databases in vision-based hand gesture recognition: A review. Computer Vision and Image Understanding **141**(December), 152–165 (2015). https://doi.org/10.1016/j.cviu.2015.08.004
28. Sahoo, A.K., Mishra, G.S., Ravulakollu, K.K.: Sign language recognition: State of the art. ARPN Journal of Engineering and Applied Sciences **9**(2), 116–134 (2014)
29. Sathiyanarayanan, M., Azharuddin, S., Kumar, S., Khan, G.: Gesture Controlled Robot for Military Purpose. International Journal For Technological Research In Engineering **1**(11), 2347–4718 (2014), www.ijtre.com
30. So, C.K.F., Baciu, G.: Entropy-based motion extraction for motion capture animation pp. 225–235 (2005). https://doi.org/10.1002/cav.107
31. Starner, T., Pentland, A.: Visual recognition of American Sign Language using Hidden Markov Models. In: Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition. pp. 189–194. Zurich, Switzerland (1995)
32. Stoll, S., Cihan Camgoz, N., Hadfield, S., Bowden, R.: Text2Sign: Towards sign language production using neural machine translation and Generative Adversarial Networks. International Journal of Computer Vision (2019). https://doi.org/10.1007/s11263-019-01281-2
33. Suh, I.H., Lee, S.H., Cho, N.J., Kwon, W.Y.: Measuring motion significance and motion complexity. Information Sciences **388-389**, 84–98 (2017). https://doi.org/10.1016/j.ins.2017.01.027
34. Sutton, V.: What is SignWriting?, https://www.signwriting.org/about/what/what02.html
35. Twenty Billion Neurons GmbH: twentybn (2019), https://20bn.com/datasets/jester/v1
36. University of East Anglia: Virtual Humans Research for Sign Language Animation, http://vh.cmp.uea.ac.uk/index.php/Main_Page
37. University of Hamburg: eSign Overview, https://www.sign-lang.uni-hamburg.de/esign/overview.html
38. Verlinden, M., Zwitserlood, I., Frowein, H.: Multimedia with Animated Sign Language for Deaf Learners. In: Kommers, P., Richards, G. (eds.) World Conference on Educational Multimedia, Hypermedia and Telecommunications. Montreal, Canada (June 2005), https://www.learntechlib.org/p/20829/
39. World Federation of the Deaf, World Association of Sign Langauge Interpreters: WFD and WASLI statement on use of signing avatars. Tech. Rep. April, Helsinki, Finland / Melbourne, Australia (2018), https://wfdeaf.org/news/resources/wfd-wasli-statement-use-signing-avatars/