

Capturing Distalization

Rose Stamp¹, Lilyana Khatib² & Hagit Hel-Or²

¹Bar-Ilan University, ²University of Haifa

Bar-Ilan University, Department of English Literature & Linguistics, University of Haifa, Computer Science department
rose.stamp@biu.ac.il, lilyanakhatib0@gmail.com, hagit@cs.haifa.ac.il

Abstract

Coding and analyzing large amounts of video data is a challenge for sign language researchers, who traditionally code 2D video data manually. In recent years, the implementation of 3D motion capture technology as a means of automatically tracking movement in sign language data has been an important step forward. Several studies show that motion capture technologies can measure sign language movement parameters – such as volume, speed, variance – with high accuracy and objectivity. In this paper, using motion capture technology and machine learning, we attempt to automatically measure a more complex feature in sign language known as distalization. In general, distalized signs use the joints further from the torso (such as the wrist), however, the measure is relative and therefore distalization is not straightforward to measure. The development of a reliable and automatic measure of distalization using motion tracking technology is of special interest in many fields of sign language research.

Keywords: motion capture, distalization, proximalization, Kinect Azure, Israeli Sign Language

1. Introduction

Sign language users exploit different articulators of their body for linguistic purposes, including the face, head, torso, and the hands. In spoken language research, linguists use a range of sophisticated computer programs in the analysis of speech. However, until relatively recently, sign language researchers lacked the equivalent type of technology for measuring different aspects of visual languages. With the introduction of infra-red motion capture technology to the field of sign language linguistics, researchers can track movement in an automatic way. Motion capture has been used as a tool for analyzing a range of sign language phenomena; for example, distinguishing between verb types (Malaia et al., 2008), lexical signs and constructed action (Stamp et al., 2018a), signs first mentioned and repeated signs (Stamp et al., 2018b), etc.

In this paper, we focus on one specific feature in sign language, described as distalization. Distalization refers to the process of distancing the joint engaged in the movement further from the body (Meier et al., 2008; Poizner et al., 2000). Some signs are produced with joints closer to the body, known as proximalized signs – typically, these joints are the shoulder and the elbow, while distalized signs are produced with joints further from the body, such as the wrist and finger joints (see Figure 1). In some cases, the same sign may exist in two variations: one proximal and one distal.

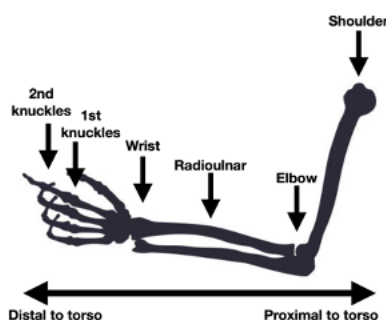


Figure 1: Distalization of joints

For example, the sign meaning ‘understand’ in Israeli Sign Language (Figure 2) can be produced proximally (left) and distally (right). In the distal example, the movement originates at the wrist, while in the proximal example, the movement originates from the elbow. However, distalization is relative, and therefore, while the wrist joint is distal in the sign for ‘understand’, it is proximal for the sign ‘donkey’, in comparison to the distal form produced at the finger joints. The choice of distal or proximal variants has been associated with a number of factors including fluency and sonority (Mirus et al., 2000; Napoli & Liapis, 2019), as well as the indexing of different social identities, such as gender and sexuality (Blau, 2017; Moges, 2020).

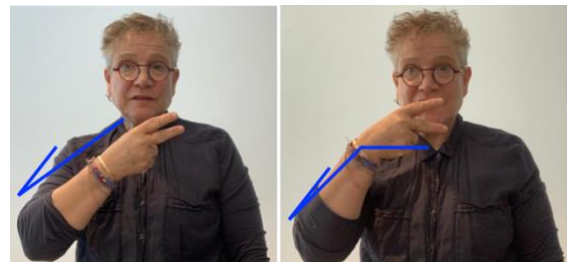


Figure 2: Proximalized sign (left) and distalized sign (right) for ‘understand’

In this study, we implement 3D motion capture technology and computational modelling to automatically detect distal and proximal signs. We hope that this offers linguists a potential alternative to the manual coding of 2D video data, which has often been adopted in previous studies (e.g., Blau, 2017).

2. Distalization

Distalization is an important measure in sign language research; the measure appears in studies on sign language production and perception, first and second language acquisition and studies on language variation and change. In studies on sign language production, the movement of the most proximal joint, the shoulder, is shown to exert the

greatest amount of energy and therefore use of distal joints has been associated with ease of articulation. Proximal signs are associated with non-nativity and are engaged in communication from a distance (Sandler & Lillo-Martin, 2006), suggesting that proximalization aids in the process of sign language perception. The choice between distal and proximal joints, therefore, is a balance between ease of articulation and ease of perception (Napoli & Liapis, 2019). There is an important link between ease of articulation and language acquisition and language variation and change. Studies show that learners of a sign language, both children and adults, begin by using proximal signs and then shift to distal signs as they increase their motor control (Gesell, 1929; Jensen et al., 1994; Meier et al., 2008). Therefore, distalization is an indicator of sign language fluency. For example, in a study examining signers of American Sign Language, fluent signers tended to reduce effort through distalization (Napoli et al., 2011).

Furthermore, distalization may influence the overall size of signing. Proximal signs, using the shoulder or elbow joints, are claimed to give an overall impression of larger signing. Movement around the wrist joint, in contrast, can give an impression of smaller signing. As a result, distal signs are often used to communicate something private or whispered (Brentari, 1999), while proximal signs are associated with conveying anger or excitement. Moreover, distalized forms may be used to index different social identities, including sexuality (Blau, 2017) and gender (Moges, 2020). In interviews conducted with female-bodied masculine ASL signers, Moges (2020) found that participants associated proximalization with masculinity, and participants were shown to proximalize their own signing when projecting a more masculine identity. The relationship between sign size and gender has not always been clear in the literature; some research claims that women sign bigger than men, i.e., that women tend to proximalize (De Santis, 1977), and other research suggests the opposite (Eichmann, 2004). In a recent study, implementing motion capture technology, it was shown that women's signing is characterized by a larger signing space than men's signing (Stamp et al., in prep.). The feature of distalization has also been associated with indexing gay identities in several studies (Blau, 2017; Fitzgerald, 2004; Michaels, 2008, 2015; Murray, 2002). In contradiction to this though, some researchers claim that gay-indexed styles of signing are characterized by a larger use of the signing space (Michaels, 2008, 2015), suggesting that distalization may not directly correlate with sign size.

The measurement of distalization however is not straightforward. It involves tracking the movement of several joints (e.g., finger joints, wrist, elbow, and shoulder), as well as measuring the degree of rotation around each joint. Manual coding of distalization is not optimal; data is usually based on 2D videos (often obscuring the observation of rotational movement around the joints), it is often coded subjectively, and it is considerably time-consuming and error prone. Therefore, the development of a reliable and automatic measure of distalization using motion tracking technology is of special interest in the fields of articulation and perception, acquisition, as well as language variation.

In the next section, we outline the tool utilized for tracking movement in this study, Microsoft Kinect Azure.

3. Motion capture in sign language research

Microsoft Kinect is a camera and body-tracking sensor system originally designed for video game play. Kinect (*Microsoft: 'Kinect for Xbox One'*, 2018) uses the time-of-flight (ToF) principle, in which the distance to an object is determined by the time it takes for the light emitted from the infrared light projector to reach the object and return to the camera's sensor (Foix et al., 2011; Hansard et al., 2012; Shotton et al., 2011). This enables the recognition of human bodies in the scene and an estimation of their locations in 3D space (see Figure 3). The Kinect camera also records standard RGB (red-green-blue) videos and audio. The advantages of using Kinect for motion capture in sign language research is that the device is inexpensive and non-invasive (therefore, causing minimal interference with signing).



Figure 3: A depth image representing distance from the camera for every point in the human figure (bright points are more distant)

In addition, when a participant is recorded, the system uses the depth image to extract a skeleton representation of the participant computed per frame (Shotton et al., 2011). The skeleton data is composed of 32 major skeleton joints of the human body, connected by line segments (see Figure 4). For every frame, the system outputs the 3D location of each of the skeleton joints (a triplet x,y,z, in meters, given in the camera's frame of reference).

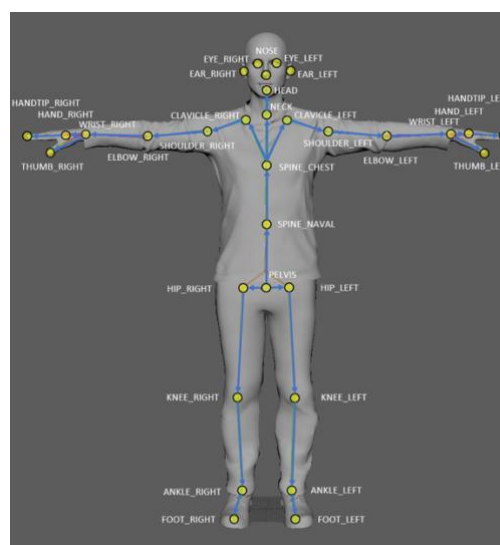


Figure 4: The joints tracked using Kinect Azure (Microsoft Kinect, 2019)

4. Methodology

Two adult female models were recruited to elicit training data (Mean age: 39 years). The models produced a set of Israeli Sign Language signs which are known to vary in terms of distalization (Stamp et al., 2021). The models were recorded using Microsoft Kinect Azure while signing two versions of the same sign (distal & proximal).

The full recording sessions were parsed into segments, comprising of single signs. Each sign was processed and analyzed using specialized code which we developed; the skeleton was extracted per frame and then spatio-temporal features were computed over all frames in the segment. Prior to computing the measurements, the skeletons were normalized to a standard size using the method in Weibel et al. (2016) to eliminate size effects.

The most noticeable features of distalization include the angles of the arms and therefore we focused on extracting movement parameters from four joints: the shoulder, elbow, wrist, and hand joints (see Figure 5). The position of each joint is given in each time frame as 3D coordinates: X, Y and Z. Thirteen features were extracted from these coordinates and used in the training:

- Speed, mean and standard deviation of:
 - Elbow angular change between frames (A)
 - Elbow twist between frames (B)
 - Wrist angular change between frames (C)
 - Hand angular change between frames (D)
- Volume:
 - Fingers

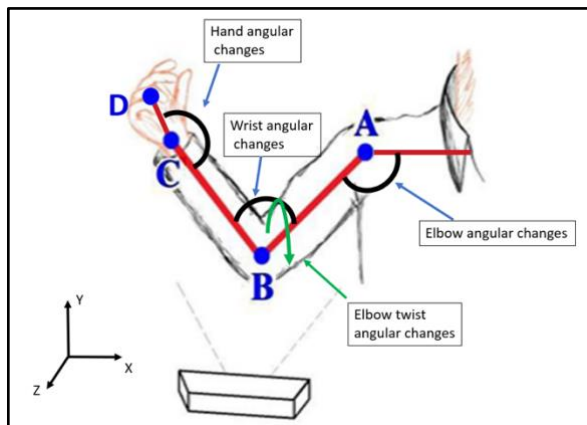


Figure 5: Visual representation of the joint movements

The twist at the elbow measures the angle of rotation of the elbow-wrist bone around the axis of rotation defined by the elbow-shoulder bone. See OSF for the calculation of angle of twist at the elbow joint in MatLab script: <https://osf.io/q3h6r/>. For each feature, the mean, std and average speed were calculated across all frames in the sequence. The set of features computed per video segment, formed a *feature vector* to be used in the machine learning algorithm.

5. Results

The data consisted of 350 samples, which were split into 50% distal and 50% proximal (our classification labels). A feature vector was created for each of the samples as described above. To assess the capability of predicting whether a sign is distally or proximally signed, we trained a machine learning model. We used the Random Forest model (Breiman, 2001; Ho, 1998), which is a collection of decision trees whose weights are learned from examples in the training set. We used 100 trees with unlimited depth. Gini was used as the split criterion at the nodes of the decision trees. We ran the test using a 10-folds validation design. Thus, the data was divided into 10 equal parts, and for each part, the samples were withheld from the rest of the data which were used to train the Random Forest model. The withheld samples were then tested for distalization using the trained model and the accuracy of correct prediction was determined. The process was repeated independently for each of the 10 parts resulting in 10 values of accuracy. Following this approach, we achieved a mean accuracy of 71% (std: 8.0).

In order to enhance the model performance, we reduced the dimensionality of the input feature vector by performing feature ranking (Guyon & Elisseeff, 2006) and removing the least informative features:

1. Hand angle mean
2. Elbow standard deviation
3. Elbow twist angle standard deviation
4. Volume of finger joint

Re-running the model with the remaining nine features using the same 10-fold validation design, we achieved an accuracy ranging between 80%-82%, with a mean accuracy of 81.35%. In other words, the model was able to predict if a new input segment was distal or proximal with 81% accuracy. The 19% of misses were a combination of false positives and false negatives (as displayed below, Figure 6). The data with highlighted misses can be accessed in OSF: <https://osf.io/q3h6r/>.

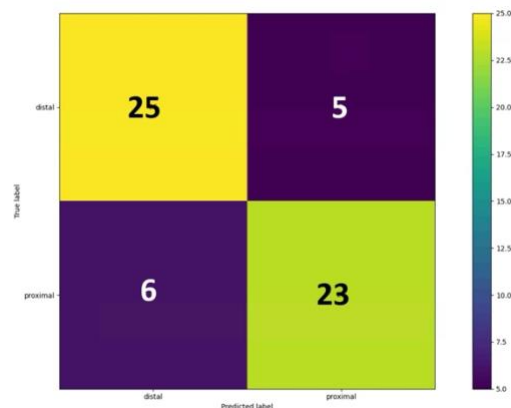


Figure 6: Distribution of the proximal and distal signs, showing 5 false positives (distal signs categorized as proximal) and 6 false negatives (proximal signs categorized as distal)

There may be several reasons why the model classified some of the samples incorrectly. In some cases, the arm joint was not captured well by the Kinect camera. In other cases, the skeleton was tracked well but the distal forms involved finger movement, which generally is not tracked well using Kinect. Finally, although a 10-folds design was used for cross-validation, the dataset is very small and therefore, more training data is required to reach a higher accuracy.

The most predictive features of distalization were standard deviation of the angular changes of the wrist and hand and the least predictive were the features depending on speed (as shown in Figure 7).

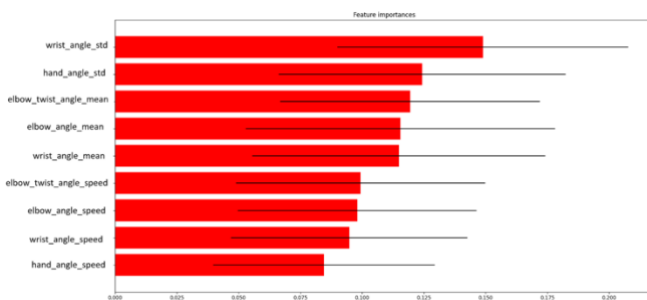


Figure 7: Features which predict distal or proximal signs, in order of their contribution to the prediction.

6. Conclusion

Distalization is a complex measure, in which the features involved are not fully understood. In this paper, we show that motion capture technologies can be implemented to measure distalization in an automatic and objective way. The model reached an accuracy of over 80% in predicting whether a sign is distal or proximal. More work needs to be done to improve the model; however, these preliminary findings suggest that motion capture can be an important tool in the automatic processing of sign language data. In addition, our initial findings point to the importance of the standard deviation of the wrist and hand movements as a predictor of such a movement. Interestingly, our model showed that volume (signing size) was not an important predictor of distal or proximal signs, despite the close relationship between distalization and signing size in the literature. Future studies should test the model on a larger dataset and implement more accurate tracking tools which enable finger joint tracking.

7. Acknowledgements

Thanks to Klil Eden and Nadav Eichler for their help in this project. Funding for this project was supported by an Israeli Science Foundation grant to Dr. Rose Stamp (2757/20; ISL Corpus project).

8. Bibliographical References

Blau, S. (2017). Indexing Gay Identities in American Sign Language. *Sign Language Studies*, 18(1), 5–40.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Brentari, D. (1999). *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.

Eichmann, H. (2004). *Do women sign bigger? A research project investigating gender-specific differences in the articulatory realisation of signs in German Sign Language (DGS)*. [MA thesis]. University of Bristol.

DeSantis, S. (1977). Elbow to Hand Shift in French and American Sign Languages. Paper presented at the Conference on New Ways of Analyzing Variation, Georgetown University, Washington, DC.

Fitzgerald, A. (2004). *Gay sign variation in ISL*. [Diploma in Irish Sign Language /English Interpreting.].

Foix, S., Alenyà, G., & Torras, C. (2011). Lock-in Time-of-Flight (ToF) Cameras: A Survey. *IEEE Sensors Journal*. <https://doi.org/10.1109/JSEN.2010.2101060>

Gesell, A. (1929). Maturation and Infant Behavior Pattern. *Psychological Review*, 36(4), 307–319. <https://doi.org/10.1037/h0075379>

Guyon, I., & Elisseeff, A. (2006). An Introduction to Feature Extraction. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), *Feature Extraction: Foundations and Applications* (pp. 1–25). Springer. https://doi.org/10.1007/978-3-540-35488-8_1

Hansard, M., Lee, S., Choi, O., & Horaud, R. (2012). *Time of Flight Cameras: Principles, Methods, and Applications*. Springer. <https://doi.org/10.1007/978-1-4471-4658-2>

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844. <https://doi.org/10.1109/34.709601>

Jensen, J. L., Schneider, K., Ulrich, B. D., Zernicke, R. F., & Thelen, E. (1994). Adaptive Dynamics of the Leg Movement Patterns of Human Infants: I. The Effects of Posture on Spontaneous Kicking. *Journal of Motor Behavior*, 26(4), 303–312. <https://doi.org/10.1080/00222895.1994.9941686>

Malaia, E., Borneman, J., & Wilbur, R.B. (2008). Analysis of ASL motion capture data towards identification of verb type. STEP '08: Proceedings of the 2008 Conference on Semantics in Text Processing, 155–164.

Meier, R. P., Mauk, C. E., Cheek, A., & Moreland, C. J. (2008). The Form of Children's Early Signs: Iconic or Motoric Determinants? *Language Learning and Development*, 4(1), 63–98. <https://doi.org/10.1080/15475440701377618>

Michaels, P. A. (2008). *The cultural and linguistic study of gay sign variation in the deaf gay community*. [PG Dip thesis.]. Durham University.

Michaels, P. A. (2015). *A study of the identity, culture and language of a sample of the deaf gay male community in Britain*. [MA thesis]. Durham University.

Microsoft Kinect (Version Azure). (2019). [[Computer software.]. <https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints>

Mirus, G., Rathmann, C., & Meier, R. P. (2000). Proximalization and Distalization of Sign Movement in Adult Learners | Semantic Scholar. *Psychology*. <https://www.semanticscholar.org/paper/Proximalization-and-Distalization-of-Sign-Movement-Mirus-Rathmann/33b71e49670ab10dd006a41b64caf4570b66209e>

Moges, R. (2020). The Signs of Deaf Female Masculinity: Styles of Gendering/Queering ASL. In K. Hall & E. Barrett (Eds.), *The Oxford Handbook of Language and Sexuality*.

- <https://doi.org/10.1093/oxfordhb/9780190212926.013.64>
- Murray, E. (2002). *Gay sign variation in Irish Sign Language*. [Diploma in Irish Sign Language Teaching. Dublin]. University of Dublin, Trinity College.
- Napoli, D. J., N. Sanders, and R. Wright. 2011. *Some aspects of articulatory ease in ASL*. Presented at Stony Brook University.
- Napoli, D. J., & Liapis, S. (2019). Effort reduction in articulation in sign languages and dance. *Journal of Cultural Cognitive Science*, 3(1), 31–61. <https://doi.org/10.1007/s41809-019-00027-3>
- Poizner, H., Brentari, D., Tyrone, M. E., & Kegl, J. (2000). The structure of language as motor behavior: Clues from signers with Parkinson's disease. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 432–452). Lawrence Erlbaum.
- Sandler, W., & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*.
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., & Moore, R. (2011). IEEE Computer 46 Society Conference on Computer Vision and Pattern Recognition. In J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, & R. Moore (Eds.), *Proceedings CVPR: Vol. 56(1)* (pp. 1297-1304).
- Stamp, R., Hel-Or, H., Cohn, D., Raz, S. & Sandler, W. (2018a). Kinect-ing the dots: Motion capture technologies and the sign language lexicon. Paper presented at the International Congress of Linguists (ICL) conference. University of Cape Town, South Africa. July 3rd, 2018.
- Stamp, R., Dachkovsky, S., Hel-Or, H., Cohn, D., Raz, S. & Sandler, W. (2018b). A Kinematic Study of Phonetic Reduction in a Young Sign Language. Paper presented at the 8th Conference of the International Society for Gesture Studies (ISGS). Cape Town, South Africa. July 5th, 2018.
- Stamp, R., Ben-Israel, A., Hel-Or, H., & Raz, S. (2021). It's just a gay thing: A comparative kinematic analysis of gestures and signs in Israel. Accepted as a paper at the 27th Lavender Languages & Linguistics conference. San Francisco, 21-23rd May 2021.
- Stamp, R., Khatib, L., Eden, K., Ben-Israel, A., Cohen, D., Hel-Or, H., Karpova, V., & Raz, S. (in prep.). Embodiment of gender in sign language: A comparative kinematic analysis.
- Weibel, N., Hwang, S.-O., Rick, S., Sayyari, E., Lenzen, D., & Hollan, J. (2016). Hands That Speak: An Integrated Approach to Studying Complex Human Communicative Body Movements. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 610–619. <https://doi.org/10.1109/HICSS.2016.82>