Capturing Distalization

Rose Stamp[1], Lilyana Khatib[2] & Hagit Hel-Or[2]

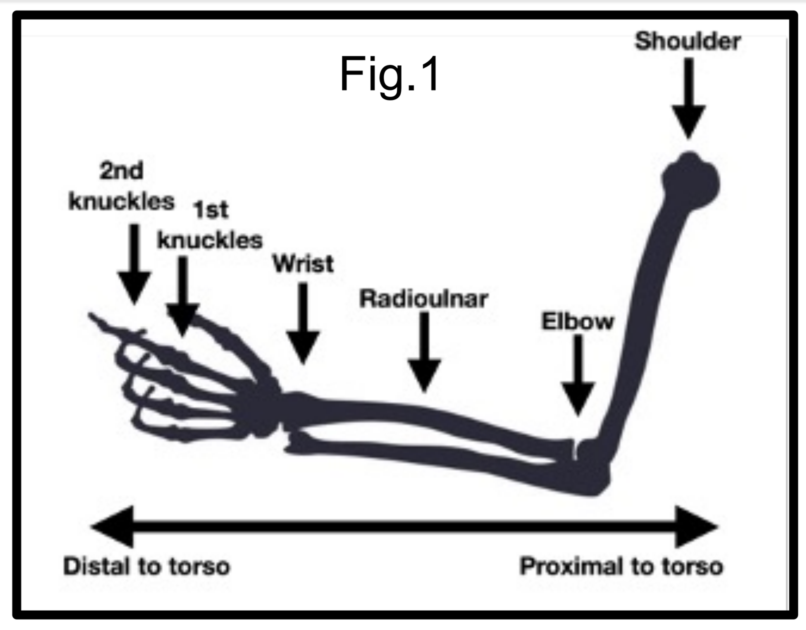
[1]Bar-Ilan University, [2]University of Haifa, Israel

Rose.stamp@biu.ac.il

What is distalization?

- Distalization refers to the process of using joints further from the body (Meier et al., 2008; Poizner et al., 2000).
- The same sign may exist in two variations: proximal and distal.
- Proximalized signs use joints closer to the body such as the shoulder or elbow, while distalized signs are produced with joints such as the wrist and finger joints (see Fig.1).
- For example, the sign 'understand' in Israeli Sign Language (Fig. 2) can be produced proximally (Fig.2a) and distally (Fig.2b).

Fig.1

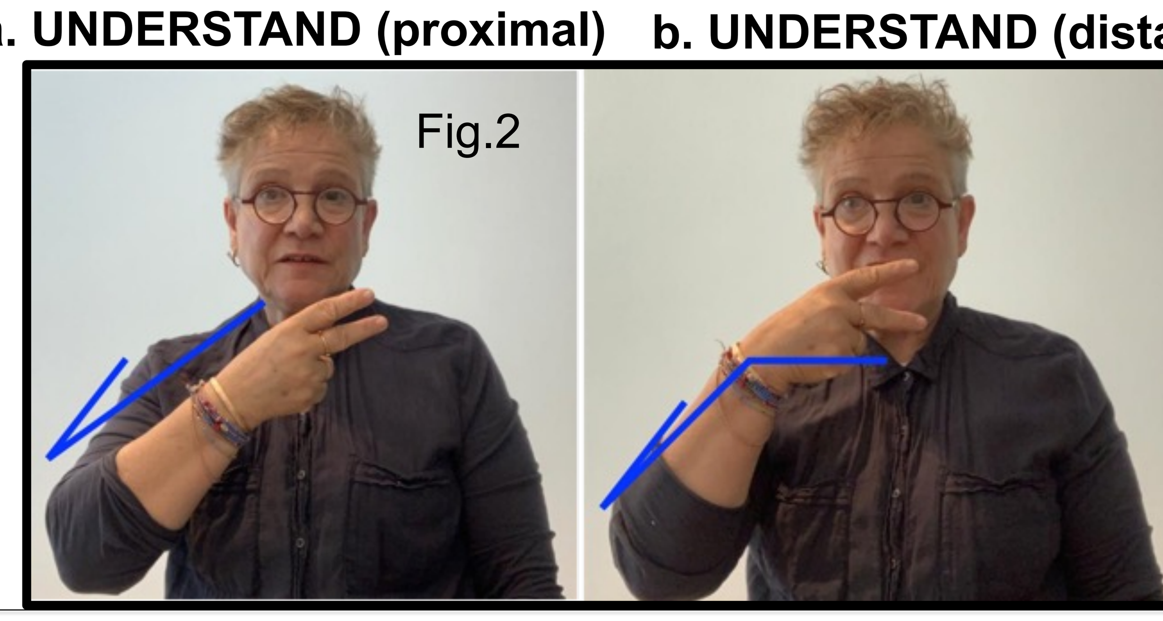


- Distalization is an important measure in the fields of sign language articulation and perception (Napoli & Liapis, 2019), acquisition (Mirus et al., 2000) and sociolinguistics (Blau, 2017).
- The measurement of distalisation is not straightforward; usually, it is coded manually and subjectively.

The goal of this study is to attempt to automatically detect distal and proximal signs, using 3D motion capture technology and computational modelling.

a. UNDERSTAND (proximal) **b. UNDERSTAND (distal)**

Fig.2

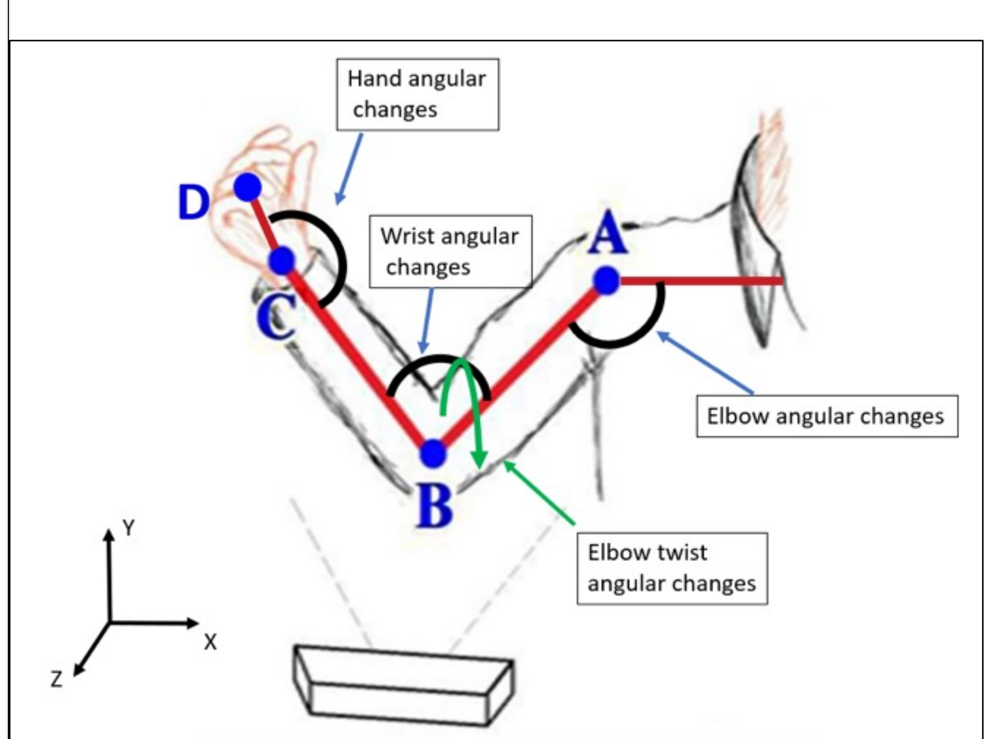


Motion Capture Technology

- 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 (Malaia et al., 2008; Stamp et al., 2018a; Stamp et al., 2018b), objectively and accurately.
- In this study, we use Microsoft Kinect Azure to track the movement of signs.
- Kinect uses time-of-flight (ToF) principle, in which the distance to an object is determined by the time it takes for the light 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.
- The advantages of using Kinect is that the device is inexpensive and non-invasive.

Methodology

- Two adult females (Mean age: 39 years) produced a set of ISL signs in two versions (distal & proximal): n = 350.
- Data were parsed into segments, comprising of single signs.
- Using a specialized code, the skeleton was extracted per frame and spatio-temporal features were computed over all frames in the segment.
- A feature vector was created for each of the samples and the data was used to train a machine learning model to distinguish between distal and proximal signs.



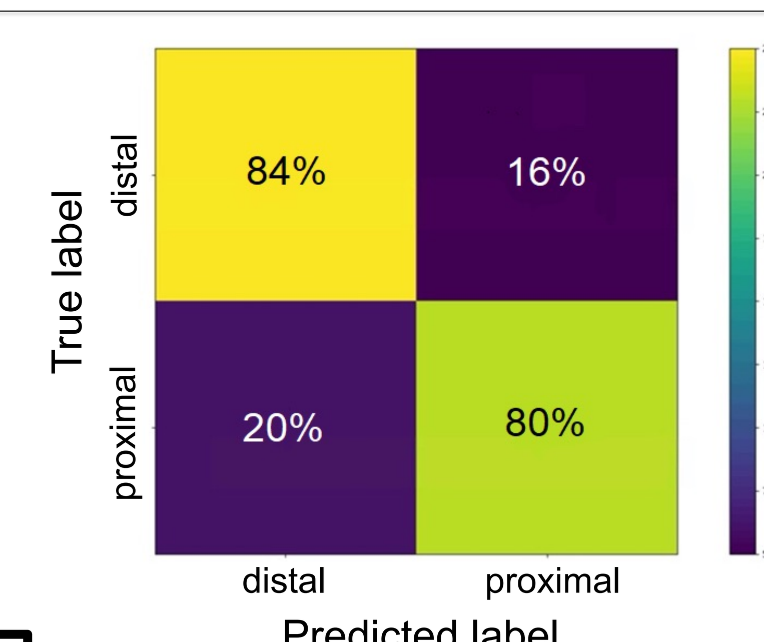
Thirteen features were extracted and used in the training:

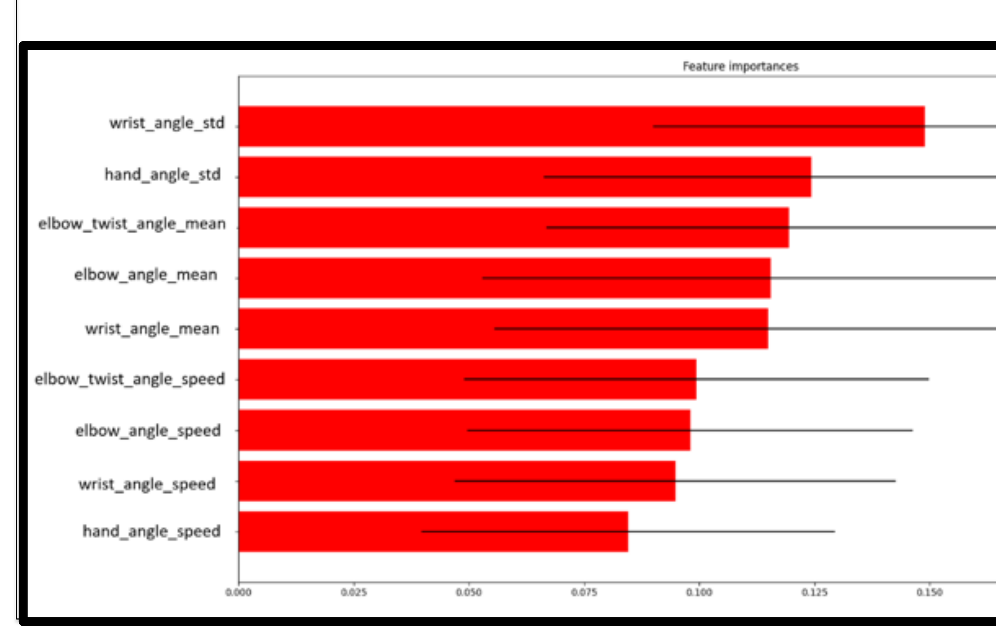
- **Speed, mean & standard deviation of:**
  - Elbow angular change (A)
  - Elbow twist (B)
  - Wrist angular change (C)
  - Hand angular change (D)
- **Volume:**
  - Fingers

Results

- Using Random Forest modelling, a mean accuracy of 71% was achieved.
- To enhance the model performance, we removed the least informative features:

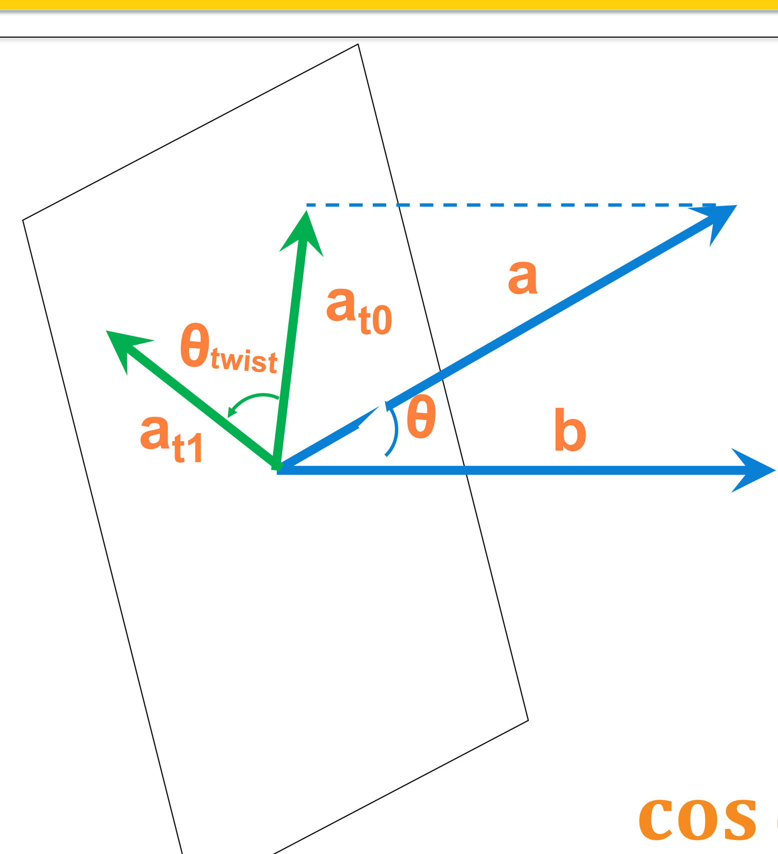
1. Hand angle mean
2. Elbow standard deviation
3. Elbow twist angle st.dev.
4. Volume of finger joint





- Re-running the model, a mean accuracy of 81.35% was achieved.
- 19% misses were a combination of false positives & false negatives.
- The most predictive features were st.dev. of the angular changes of the wrist & hand.
- The least predictive were the features related to speed.

Calculations of angles



To calculate angular change at a joint:

- The angle θ at a joint (e.g., elbow) is calculated based on Equation 1.
- a and b are 3D vectors in size and direction equal to the two bones leaving the joint (e.g., elbow-shoulder and elbow-wrist bones).
- Note, this is a 3D angle.
- Angular change is taken as the difference in θ across a time unit.

$$\cos\theta = \frac{a \cdot b}{|a||b|} \quad [1]$$

To calculate the hand/wrist/elbow twist angle:

- Consider the two vectors a and b leaving a joint.
- Calculate the projection of vector a onto the plane perpendicular to b at two time points. Denoted by a_{t0} and a_{t1}.
- For example, project the elbow-wrist bone onto the plane perpendicular to the elbow-shoulder bone.
- The angle θ_{twist} between the projected vectors is calculated using Equation 1

Conclusions

- Distalization is a complex measure, in which the features involved are not fully understood.
- We show that motion capture technologies can be implemented to measure distalization in an automatic and objective way, to an accuracy of over 80%
- Volume (signing size) was not an important predictor of distal or proximal signs, despite the close relationship with distalization in the literature.
- Future studies should test the model on a larger dataset and implement more accurate tracking tools which enable finger joint tracking.

References

[1] Blau, S. (2017). Indexing Gay Identities in American Sign Language. Sign Language Studies, 18(1), 5–40.
[2] Foix, S., Alenyà, G., & Torras, C. (2011). Lock-in Time-of-Flight (ToF) Cameras: A Survey. IEEE Sensors Journal. [3] 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. [4] 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. [5] Microsoft Kinect (Version Azure). (2019). [[Computer software].]. https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints [6] Mirus, G., Rathmann, C., & Meier, R. P. (2000). Proximalization and Distalization of Sign Movement in Adult Learners | Semantic Scholar. Psychology. [7] Napoli, D. J., & Liapis, S. (2019). Effort reduction in articulation in sign languages and dance. Journal of Cultural Cognitive Science, 3(1), 31–61. [8] 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. [9] 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.). [10] 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. [11] 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.