



HAKA: HierArchical Knowledge Acquisition in a sign language tutor

Itsaso Rodríguez-Moreno ^{*,1}, José María Martínez-Otzeta, Basilio Sierra

Department of Computer Science and Artificial Intelligence, University of the Basque Country (UPV/EHU), Manuel Lardizabal 1, Donostia-San Sebastián 20018, Gipuzkoa, Spain

ARTICLE INFO

Keywords:
Sign language
Language tutor
Action recognition
Procrustes similarity
Multidimensional scaling

ABSTRACT

Communication between people from different communities can sometimes be hampered by the lack of knowledge of each other's language. A large number of people needs to learn a language in order to ensure a fluid communication or want to do it just out of intellectual curiosity. To assist language learners' needs tutor tools have been developed. In this paper we present a tutor for learning the basic 42 hand configurations of the Spanish Sign Language, as well as more than one hundred of common words. This tutor registers the user image from an off-the-shelf webcam and challenges her to perform the hand configuration she chooses to practice. The system looks for the configuration, out of the 42 in its database, closest to the configuration performed by the user, and shows it to her, to help her to improve through knowledge of her errors in real time. The similarities between configurations are computed using Procrustes analysis. A table with the most frequent mistakes is also recorded and available to the user. The user may advance to choose a word and practice the hand configurations needed for that word. Sign languages have been historically neglected and deaf people still face important challenges in their daily activities. This research is a first step in the development of a Spanish Sign Language tutor and the tool is available as open source. A multidimensional scaling analysis of the clustering of the 42 hand configurations induced by Procrustes similarity is also presented.

1. Introduction

People from interacting communities where different languages are spoken have the need to overcome this communication barrier. Along history these interactions have made languages disappear, evolve or become dominant in a region or cultural domain. In spite of the current trends of increasing globalization and interconnectedness, there are more than 7,000 living languages in the world (Eberhard, Simons, & Fennig, 2022).

Given these figures it is safe to conclude that the knowledge of languages not natively acquired is a must for many people in the world. In addition to those who have that need, there are also people who want to be fluent in another language just out of intellectual curiosity. The demand of individuals as well as the public and private sector made the global market size of the language industry at around 49.6 billion U.S. dollars in 2019 (Mazereanu, 2019).

Information technologies are nowadays a fundamental part of the process of teaching, studying and/or practicing a new language. Even in the traditional teacher-student setting, videoconferencing applications

allow for remote lessons weakening the constraints faced when the teacher and the student needed to share the same spatial location. Recording of the lessons also permits a time managing more suited to the needs of the learner. But the most revolutionary changes have come with the use of software engineering practices and artificial intelligence techniques. Companies who deploy websites with user progress customization, tailored tutoring and automatic voice recognition are competing with companies with a more classic approach based on onsite teaching by native speakers.

In this paper we report on the ongoing work in a Spanish Sign Language tutor which could capture the user image in a standard webcam and recognize the hand configuration or sign she is performing and give feedback about her progress. Sign languages are used by the hearing-impaired community, mainly for communication among their members, given the lack of knowledge outside their community. A sign is composed by a succession of hand configurations, along with body movements or facial expressions. Our approach is bottom-up, building the hand configuration recognizer before and as a part of the sign recognizer. In this way, compared with end-to-end deep learning

* Corresponding author.

E-mail addresses: itsaso.rodriguez@ehu.eus (I. Rodríguez-Moreno), josemaria.martinezo@ehu.eus (J.M. Martínez-Otzeta), b.sierra@ehu.eus (B. Sierra).

¹ ORCID: 0000-0001-8471-9765.

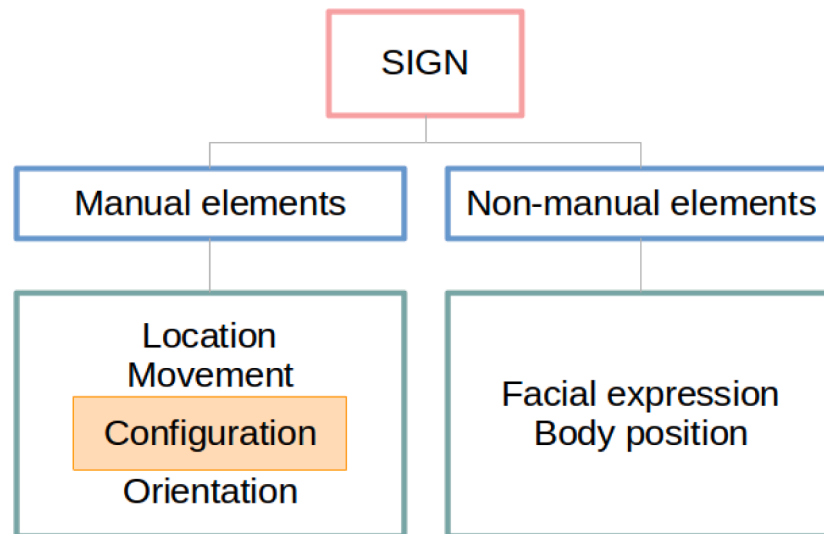


Fig. 1. Elements which compose a sign.

approaches, it is possible to have a better interpretability of the tutor decisions and the learner can be informed of why she has made a mistake and not only that she has made a mistake. The hand configuration recognizer is based in the Procrustes distance between the user configuration and the stored models, and therefore it is very easy to add or remove another model to the stored library.

The paper is organised as follows: the next section is devoted to other research related to the current work. Then the proposed approach is presented and the following section presents the functionalities of the developed application. The paper closes with a discussion about the insights gained from Procrustes analysis and outlines further work.

2. Related work

Computer-assisted language learning is an active interdisciplinary field of research which encompasses many topics covering from pedagogy to artificial intelligence (Chen, Zou, & Su, 2021). In (Zhang & Zou, 2022) the authors find that five main types of technology are employed for language learning: speech-to-text and text-to-speech recognition, mobile learning, socialized learning, multimedia learning and game-based learning. They also show that these technologies make it easier the delivery of content and facilitate interactions, while at the same time promote language practising and restructure teaching approaches.

While the majority of languages could benefit from speech-to-text and text-to-speech technologies, sign languages processing needs to focus on gestures instead of speech. Signs are performed mainly with the hands, although general body movement and face expressions may also convey meaning. In order to recognize the sign, some data capture system is needed. The use of electronic gloves that detect or record the hand position (De Marco & Foulds, 2003; Ahmed, Zaidan, Zaidan, Salih, & Lakulu, 2018) or of colored gloves that can be segmented from the whole image (Wang & Popović, 2009; Azar & Seyedarabi, 2020) is being made redundant thanks to the latest advances in machine learning, which permit hand pose estimation without extra equipment. In (Quinn & Olszewska, 2019) the authors present a system designed to detect and recognize British Visual Language signs, using Histogram of Oriented Gradients (HOG) (Dalal & Triggs, 2005) and multi-class SVMs (Cortes & Vapnik, 1995). Hidden Markov Models (HMMs) (Baum & Petrie, 1966) have also been applied in this field of research. In (Azar & Seyedarabi, 2020) a dynamic Persian sign language recognition system is presented, where Gaussian HMMs are used as modeling tools for hand trajectories of signers. The hand segmentation is facilitated by the users wearing white gloves. Another approach based on skin segmentation is described

in (Roy, Kumar, & Kim, 2021), where the Camshift algorithm (Bradski, 1998) is employed to track the hand and the different trajectories when performing ASL signs are classified using HMMs.

Deep learning techniques are widely used nowadays due to their effectiveness in all kind of domains and sign language processing is no exception. Researchers in different sign languages have made use of deep learning to achieve significant progress in sign recognition. A real-time system for recognition of American Sign Language using a convolutional neural network is presented in (Taskiran, Killioglu, & Kahraman, 2018), while in (Boháček & Hruží, 2022) the authors perform a word-level sign recognition for ASL using pose-based transformers (Vaswani, et al., 2017). In (Sevli & Kemaloglu, 2020) the authors also use convolutional neural networks for the recognition of digits in Turkish sign language, while in (Aktaş, Gökberk, & Akarun, 2019) a ResNet architecture (He, Zhang, Ren, & Sun, 2016) is employed to recognize Turkish non-manual signs, as facial expressions and head movements. Other sign languages for which deep learning approaches have been applied include Indian (Sharma, Sharma, Saxena, Singh, & Sadhya, 2021), Chinese (Gao, et al., 2021), Indonesian (Fadlilah et al., 2021) and Brazilian (Rocha, Lensk, Ferreira, & Ferreira, 2020), to name just a few. Comprehensive surveys of the current deep-learning-based research can be found in (Rastgoo, Kiani, & Escalera, 2021) and (Al-Qurishi, Khalid, & Souissi, 2021).

While the performance of deep learning systems is impressive, they often lack the ability to explain their conclusions in human-understandable terms (Fazi, 2021). They are also often presented as universal learners from low-processed data, with less need for feature engineering, although adding experts' insight into a traditional approach could be competitive against them (Jiang, et al., 2018).

Furthermore, applications using some definition of distance between static or dynamic gestures have been proven successful in some domains. In (Ibañez, Soria, Teyseyre, Rodríguez, & Campo, 2017) the authors propose a lightweight approach to gesture recognition by encoding the movements of 3D joints coordinates provided by a Kinect camera into a string. Then they use techniques of string matching to tackle the problem. Another classical statistical technique, Procrustes analysis (Gower, 1975; Dryden & Mardia, 2016), has been employed to detect hand-over-face expressions (Révy, Hadházi, & Hullám, 2022).

The hand gesture recognition module of the Spanish Sign Language tutor system that we are developing builds on the definition of distance between the performed hand configuration and the elements of a database. One of the advantages of this approach is that the system could show the user which is the gesture, from the database, most similar to

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42

Fig. 2. The 42 configurations of the Spanish Sign Language. Blue: front of the side (palm) in the background. Grey: back of the side (knuckles) in the background.

the one she is performing. This might be more useful than just an error message telling the user that she is wrong but not why.

3. Proposed approach

A sign can be decomposed into the elements shown in Fig. 1.

- Manual elements.
 - o *Location*: the location where signs are performed, including the part of the body, the plane and the contact point. The *plane*

indicates where the sign is performed according to the distance to the body, whereas the *contact point*, if any, refers to the part of the dominant hand that touches another part of the body.

- o *Configuration*: the shape of the hand when performing a sign. In the Spanish Sign Language (*Lengua de Signos Española*, LSE) there are three different types of configurations: the configurations representing the Spanish alphabet performed by the dominant hand, the hand configurations to sign the natural numbers, and the configurations representing the phonemes, similar to the phonological system formed by the distinctive sounds of an oral language.

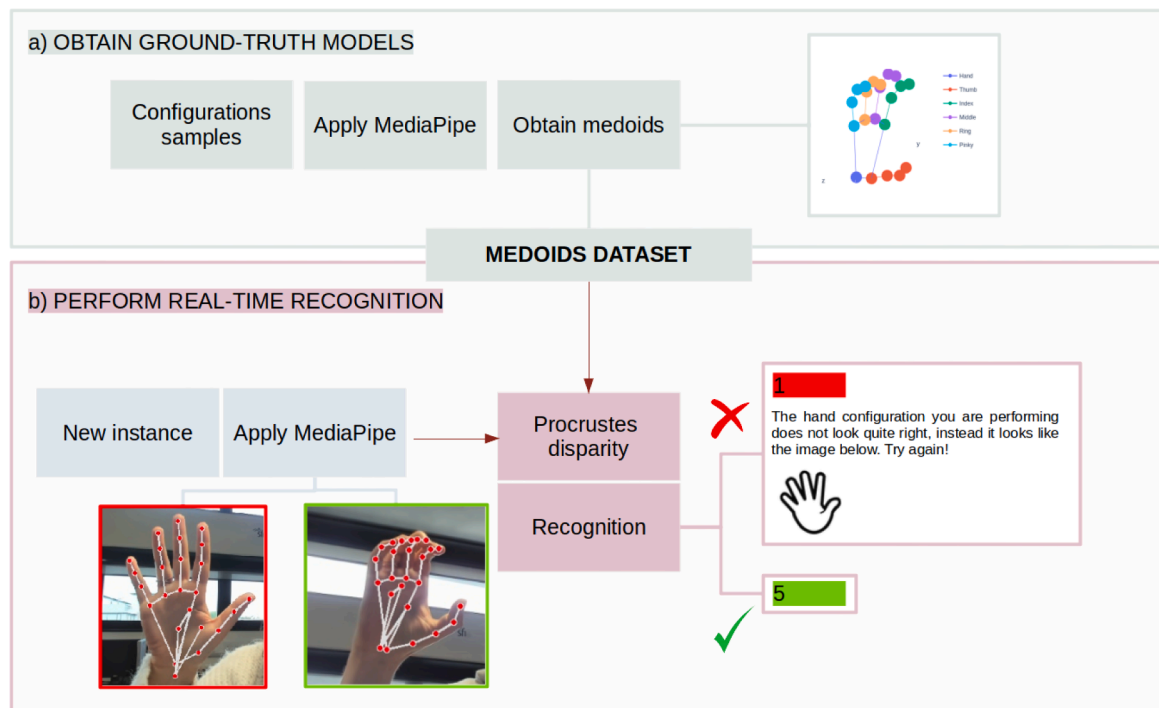


Fig. 3. Followed pipeline.

- o *Orientation*: orientation of the hands involved on the articulation of the sign with respect to the body of the signer, that is, the distinctive relative degree of rotation of the hands when signing.
- o *Movement*: the movement usually involved when performing a sign, the distinctive hand actions that form words.
- Non-manual elements. In addition to the manual elements, there are other non-manual components which are also crucial when defining a sign: the facial expression and the body position.

Since our aim is to help people to start learning the Spanish Sing Language, as a first approach we propose a basic tutor to first learn the different phonological configurations of the Spanish Sign Language, and therefore develop the skills needed to perform the signs of the language. As shown in Fig. 2, 42 different configurations are the basis for the Spanish Sign Language. Their recognition with high accuracy can be a challenging task due to the high degree of visual similarity among some subsets of configurations.

In other approaches it is usual to obtain a set of training examples and process it to create a model, knowing very little about the structure of the possible classes before the start of the learning process. In contrast, the depictions of all the possible hand configurations of the Spanish sign language are known in advance. Therefore, it is possible to create ground-truth models with the hand configurations and then compute the distance from a new input to those models, returning the most similar one. This approach is similar to KNN (more precisely 1-NN), with Procrustes disparity as distance. The process of creating the ground truth could be considered similar to prototype selection in KNN (Garcia, Derrac, Cano, & Herrera, 2012).

The followed pipeline can be seen in Fig. 3, where first we extract the hand-landmarks of the user using MediaPipe (Zhang, et al., 2020) and compare the captured shape to saved pre-defined models using Procrustes analysis in order to predict the label corresponding to the most similar model. It is worth mentioning that the process depicted in the upper part of the diagram, namely the computation of the medoids, is only executed once, as a necessary step before the tutor can be used by the public.

MediaPipe Hand Tracking solution, part of MediaPipe, is able to

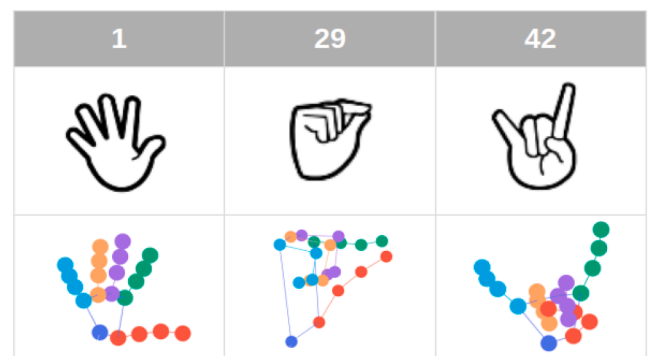


Fig. 4. Configurations and their corresponding saved medoids models.

estimate the spatial location of 21 landmarks for each hand. Each hand landmark is composed of three values, the coordinates (x,y,z), where z represents the depth with respect to the wrist. These data will be used to represent the ground truth configuration models as well as the hand configurations performed by the users.

The steps of the pipeline are explained in the following subsections.

3.1. Ground-truth models

To build the ground-truth models a person with basic formal education in Spanish Sign Language has performed 50 times each configuration in front of a Logitech BRIO 4 K Ultra HD Webcam and in good lighting conditions. The medoid of each set of 50 repetitions of configurations with respect to the Procrustes disparity has been computed and saved as ground-truth. In Fig. 4 some examples of the saved models are shown, along with the configuration they belong to. These are the models which are used as ground-truth to compare with the configuration the user is performing. As it can be seen, the saved pre-defined models are very similar to the actual configurations.

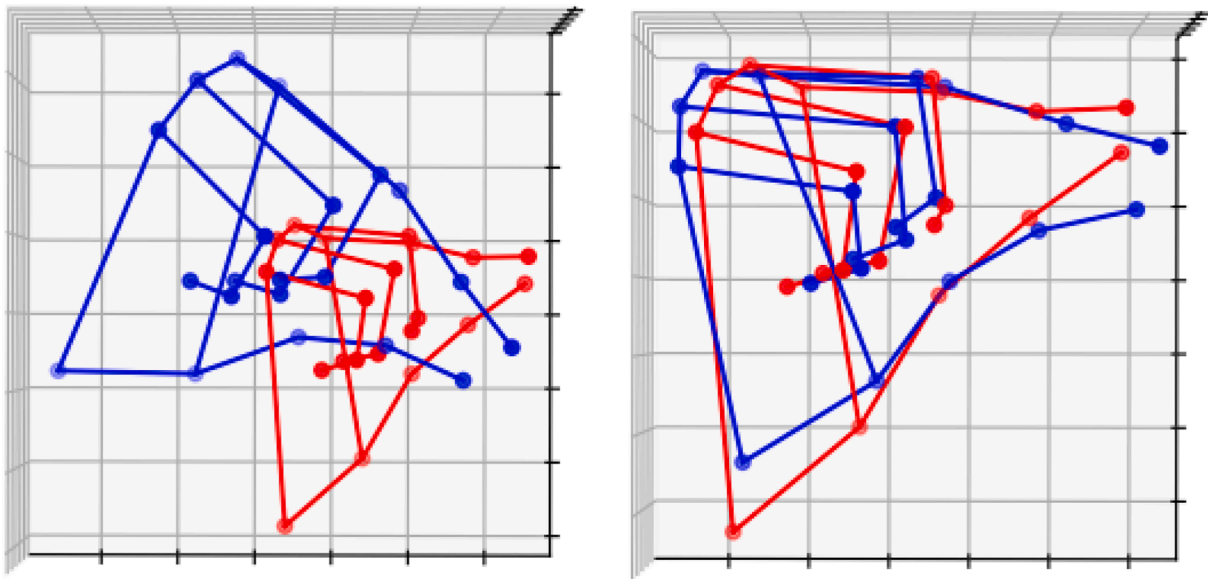


Fig. 5. Example of the transformations applied in Procrustes disparity. On the left, two original hand shapes as obtained by MediaPipe. On the right, the same hand shapes after applying the transformations to minimize their differences.

3.2. Procrustes analysis

The similarity between the configurations performed by the user and the saved models is calculated according to the Procrustes analysis, a statistical technique applied in areas ranging from microbiology (Tremblay, et al., 2015) to social robotics (Zabala, Rodriguez, Martínez-Otzeta, Irigoien, & Lazkano, 2021). Having two different shapes, the Procrustes analysis consists in performing a combination of transformations, including scaling, rotation and reflections, in order to minimize the difference between both shapes. After applying the transformations, the difference is calculated as indicated in Equation (1).

$$diff = \sqrt{(data_1 - data_2)^2} \tag{1}$$

In Fig. 5 an example of the hand shapes before and after the transformations can be seen. While on the left, the shapes are shown exactly as obtained after applying MediaPipe to the recording, on the right, the shapes have been transformed trying to minimize their difference. In particular, after centering both shapes around the origin, the optimal transformations are applied to the second shape (the blue shape in this case) to try to achieve as much similarity as possible with the other shape.

This way, the Procrustes similarities between the configuration that is being performed and the 42 saved models (one per configuration of

Choose which hand you are using

Right hand Left hand

Choose a configuration to practice

1

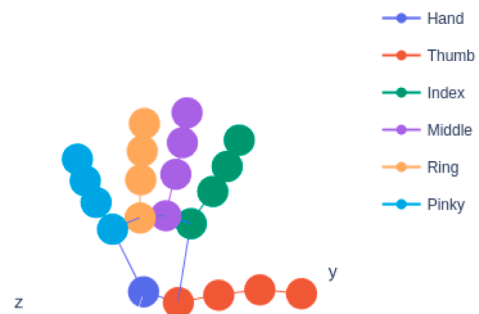




Fig. 6. A screenshot of the web application where, after choosing the configuration to perform and the hand which is being used, the image and medoid (model) for that configuration are shown.

PERFORMANCE.

The 57.69% of the trials were well performed.

These are the ten most confusing configurations in your attempts.

	Chosen configuration	Performed configuration	% Wrong performances
	1	3	23.08
	1	2	19.23

You have selected to practice configuration 1

But you have performed configuration 2



[Continue learning](#)

Fig. 7. Performance of the user, showing the mistakes he/she has made.

the database) are calculated in order to perform the recognition. The configuration with which the lowest difference is predicted, that is, the closest configuration is used to decide the label of the performed configuration.

4. Functionalities

The purpose of the approach is to develop a Spanish Sign Language tutor to help people who want to learn this language. With that in mind, we have built a web application where users are able to learn how to perform the different configurations and practice those that make up different signs. The web application is available as open-source on GitHub https://github.com/rsait/LSE_tutor.

4.1. Learn configurations

In the *learn configurations* functionality, the user has the option to

decide which configurations to practice and with which hand he/she is going to carry them out

Once these details are set, the selected configuration is shown to the user in two different ways:

- An image of the configuration to practice.
- A 3D plot with the hand landmarks obtained by MediaPipe corresponding to the saved model for the selected configuration.

This is expected to be helpful for the user as the visual information is displayed while he/she is performing the configurations. In Fig. 6 a screenshot of the application is shown, where the mentioned elements are shown.

After deciding the configuration and the hand to use, the user can start practicing the selected configuration. If the performance is correct, i.e. the selected configuration corresponds to the configuration the user performs, a green background is displayed to make the user aware that

Choose which hand you are using

Right hand Left hand

Topic:

Adjetivos (Adjectives)

Word:

Pelirrojo (Red-haired)

These are the 3 configurations you must perform:



Fig. 8. A screenshot of the web application where, after selecting a sign, the configurations that compose it are displayed.

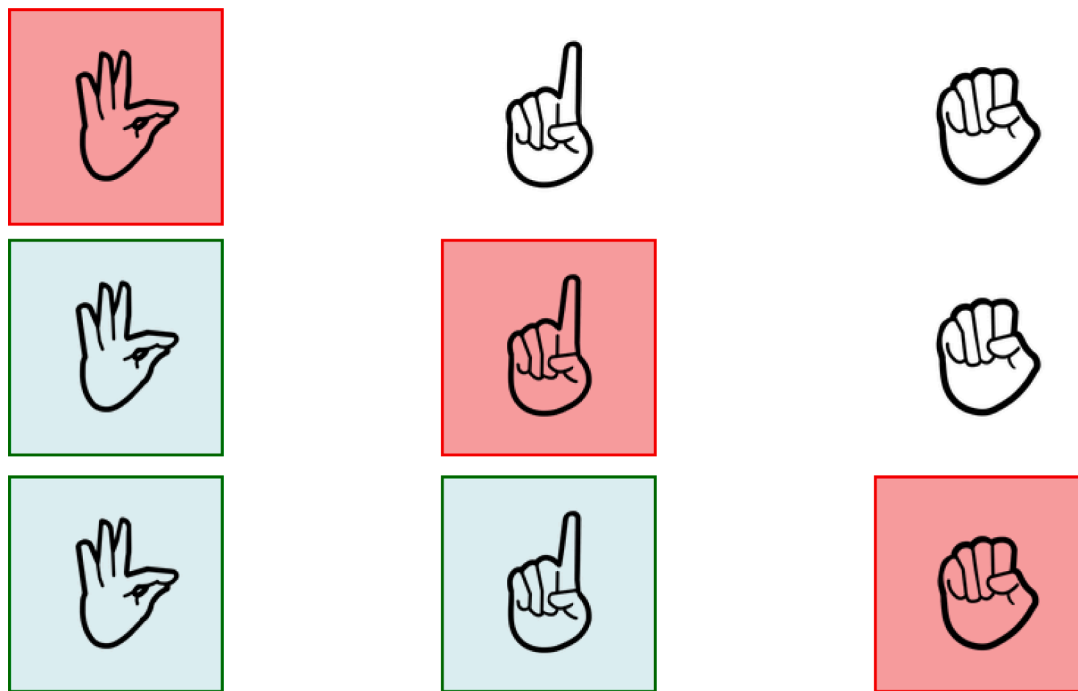


Fig. 9. Sequence of configurations that make up a sign. The red background indicates the configuration to be carried out. Once it has been correctly performed, the green background is set.

he/she is doing a good job. However, if the prediction does not match the selected configuration, apart from setting a red background, the name and the image of the configuration that the user is performing are shown. This way, the user can analyse which configuration he/she is performing and which one he/she has to perform, making the differences noticeable.

In addition, a record of the mistakes is saved (see Fig. 7). The user is able to see which have been the configurations he/she has performed worst and try to improve them. The mistakes are shown in a table, where the ten most frequent errors are listed. Each record is composed of the configuration that was selected to practice, the configuration that has been performed instead and the percentage of the wrong performances of that confusion. If a record is selected, the images of the corresponding configurations are displayed. The total percentage of the trials that were correctly performed is also indicated.

4.2. Combine signs and configurations

Trying to motivate the learning of the configurations, another functionality has been added to the web application. The opportunity to practice the configurations corresponding to different signs is given. The user can select a sign among 196 different signs which have been added, and the configurations corresponding to the selected sign are shown in the same order in which they are carried out in the sign.

In Fig. 8 a screenshot of the application is displayed, showing the functionality just described. As it can be seen, there are two different dropdowns to select the sign, because these are separated into subjects. The user can choose a topic (first dropdown) and then a sign corresponding to that topic (second dropdown).

As it can be perceived in Fig. 8, when showing the configurations that have to be performed, a red background is set on the first one. The idea is to perform the configurations correctly in the same order as they are performed when the signs are actually executed. Therefore, the sequence of configurations must be performed correctly. The red background is placed on the configuration to be carried out. Once this is correctly performed, the green background is set and the red background is switched to the next one. An example sequence is shown in

Fig. 9.

5. Discussion and further work

To finish, an analysis of the clustering of the 42 hand configurations induced by multidimensional scaling using Procrustes similarity as distance is also presented. In Fig. 10 the plot corresponding to the multidimensional scaling in 2D space of the configurations space is shown. This technique is useful to visualize neighbourhood relationships between objects when only distances are given. Close instances according to the distance appear also close in the graph while less similar instances are farther apart.

In this case, the multidimensional scaling is used to evaluate if the Procrustes disparity agrees with the human perception of similarity between hand configurations. In other words, if a person finds that the cluster distribution of hand configurations looks natural. This seems to be the case after looking at Fig. 10 and the configurations of Fig. 2. For example, configurations 12, 13, 14 and 15 appear very close in the plot. Looking at the shape of these configurations, they really look similar to the human eye, as they all involve the thumb finger, the index finger and the distance between them, maintaining the rest of the fingers stretched. There are other similarities that can be clearly observed in the plot:

- 16, 17, 18 and 19. This group of configurations share the position of the middle finger, the thumb and the distance between them.
- 11 and 39. In these two configurations the fist is closed and the difference is minimal, consisting of the middle finger and how it makes contact with the thumb.
- 33, 34, 35, 36, 37 and 38. These configurations share that the index and the middle finger are stretched, while the ring and the pink fingers are flexed.

Therefore, it can be concluded that the Procrustes disparity is an adequate similarity measure to compare the shape of the hands as the obtained results show that the configurations that are most confusing are considered similar. That is, the system considers similar the same as a user would consider similar.

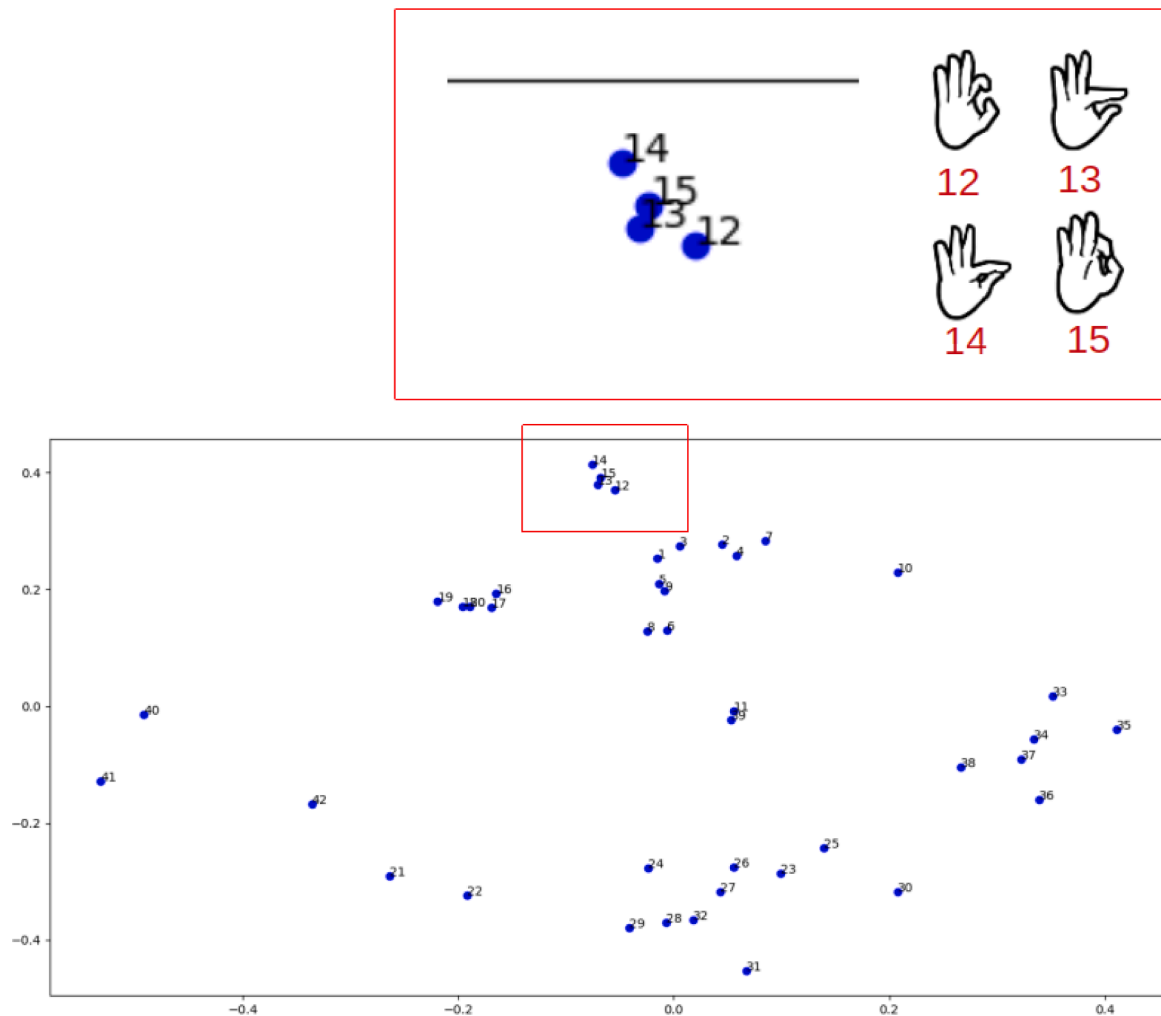


Fig. 10. Multidimensional scaling — example of similar configurations.

The addition of a new hand configuration is straightforward. The only needed modification is the addition of a new model in the model database, and the Procrustes disparity will be computed over the updated database. In this sense this approach shares similarities with K-NN or other so-called “lazy” methods, where updating knowledge does not imply retraining. However, the addition of a new model may require some work, as it must be a faithful representation of the hand configuration to be recognized. We address this issue by computing the medoid of a video sequence of a person performing 50 repetitions of each configuration. This need could represent a limitation in some contexts. Another limitation of the presented approach is that the running time is sensitive to the number of classes (the different hand configurations), since more classes will take longer to perform the Procrustes analysis. This limitation is typical in KNN-related methods and could be an issue with a big number of hand configurations. However, in the particular case of Spanish Sign Language, the number of possible configurations is 42, which can be computed in real time.

Performing the Procrustes analysis implies computing a Singular Value Decomposition, which is, for a matrix $m \times n$, an operation of time complexity $O(m^2n + n^3)$, with constant of proportionality ranging from 4 to 10 (or more) depending on the specific algorithm (Golub & Van Loan, 2013). But, as the dimension of the feature space does not grow larger, the dimension of the matrices stays constant even when more hand configurations are added. Therefore, the computing time grows only linearly when adding more configurations, as only linearly more distances have to be computed and compared.

Preliminary testing with people not involved in the development of the application has shown that they find it easy to use, with their performance on their past attempts helpful to let them know which configurations they find more challenging. For convenience we include in the repository the files needed to build a docker image, so the user might not need to install additional software.

The next step would be to recognize whole signs consisting of a succession of hand configurations in different spacial locations, as well as adding two-hands support and facial expressions. To this end, any approach suitable for time series could be applied, from HMMs to deep learning techniques.

CRedit authorship contribution statement

Itsaso Rodríguez-Moreno: Conceptualization, Software, Writing – original draft. **José María Martínez-Otzeta:** Methodology, Validation, Writing – review & editing, Supervision. **Basilio Sierra:** Validation, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work has been partially funded by the Basque Government, Spain, under Grant number IT1427-22; the Spanish Ministry of Science (MCIU), the State Research Agency (AEI), the European Regional Development Fund (FEDER), under Grant number PID2021-122402OB-C21 (MCIU/AEI/FEDER, UE); and the Spanish Ministry of Science, Innovation and Universities, under Grant FPU18/04737. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

- Ahmed, M. A., Zaidan, B. B., Zaidan, A. A., Salih, M. M., & Lakulu, M. M. (2018). A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017. *Sensors*, *18*, 2208.
- Aktaş, M., Gökberk, B., & Akarun, L. (2019). Recognizing non-manual signs in Turkish sign language. *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, (pp. 1–6).
- Al-Qurishi, M., Khalid, T., & Souissi, R. (2021). *Deep learning for sign language recognition: Current techniques, benchmarks, and open issues*. IEEE Access.
- Azar, S. G., & Seyedarabi, H. (2020). Trajectory-based recognition of dynamic Persian sign language using hidden Markov model. *Computer Speech & Language*, *61*, Article 101053.
- Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The annals of mathematical statistics*, *37*, 1554–1563.
- Boháček, M., & Hrz, M. (2022). Sign Pose-based Transformer for Word-level Sign Language Recognition. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (pp. 182–191).
- Bradski, G. R. (1998). Real time face and object tracking as a component of a perceptual user interface. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No. 98EX201)*, (pp. 214–219).
- Chen, X., Zou, D., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology*, *25*, 151–185.
- Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine Learning*, *20*, 273–297.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, *1*, pp. 886–893.
- De Marco, R. M., & Foulds, R. A. (2003). Data recording and analysis of American Sign Language. *2003 IEEE 29th Annual Proceedings of Bioengineering Conference*, (pp. 49–50).
- Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis: with applications in R* (Vol. 995). John Wiley & Sons.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2022). *Ethnologue: Languages of the world*. *Ethnologue: Languages of the world*. Dallas, TX: SIL International.
- Fadlilah, U., Handaga, B., et al. (2021). The development of android for Indonesian sign language using tensorflow lite and CNN: An initial study. *Journal of Physics: Conference Series*, *1858*, Article 012085.
- Fazi, M. B. (2021). Beyond human: Deep learning, explainability and representation. *Theory, Culture & Society*, *38*, 55–77.
- Gao, L., Li, H., Liu, Z., Liu, Z., Wan, L., & Feng, W. (2021). RNN-transducer based Chinese sign language recognition. *Neurocomputing*, *434*, 45–54.
- Garcia, S., Derrac, J., Cano, J., & Herrera, F. (2012). Prototype selection for nearest neighbor classification: Taxonomy and empirical study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*, 417–435.
- Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations*. JHU press.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*, 33–51.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 770–778).
- Ibañez, R., Soria, Á., Teyseyre, A., Rodríguez, G., & Campo, M. (2017). Approximate string matching: A lightweight approach to recognize gestures with Kinect. *Pattern Recognition*, *62*, 73–86.
- Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M., . . . Biswas, G. (2018). Expert feature-engineering vs. deep neural networks: which is better for sensor-free affect detection? *International conference on artificial intelligence in education*, (pp. 198–211).
- Mazereanu, E. (2019). Market size of the global language services industry, 2009–2022. *Market size of the global language services industry, 2009–2022*.
- Quinn, M., & Olszewska, J. I. (2019). British sign language recognition in the wild based on multi-class SVM. *2019 federated conference on computer science and information systems (FedCSIS)*, (pp. 81–86).
- Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, *164*, Article 113794.
- Révy, G., Hadházi, D., & Hullám, G. (2022). Towards Hand-Over-Face Gesture Detection. *29th Minisymposium of the Department of Measurement and Information Systems*, (pp. 58–61).
- Rocha, J., Lensk, J., Ferreira, T., & Ferreira, M. (2020). Towards a tool to translate brazilian sign language (libras) to brazilian portuguese and improve communication with deaf. *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, (pp. 1–4).
- Roy, P. P., Kumar, P., & Kim, B.-G. (2021). An efficient sign language recognition (SLR) system using Camshift tracker and hidden Markov model (HMM). *SN Computer Science*, *2*, 1–15.
- Sevli, O., & Kemaloglu, N. (2020). Turkish sign language digits classification with CNN using different optimizers. *International Advanced Researches and Engineering Journal*, *4*, 200–207.
- Sharma, A., Sharma, N., Saxena, Y., Singh, A., & Sadhya, D. (2021). Benchmarking deep neural network approaches for Indian Sign Language recognition. *Neural Computing and Applications*, *33*, 6685–6696.
- Taskiran, M., Killioglu, M., & Kahraman, N. (2018). A real-time system for recognition of American sign language by using deep learning. *2018 41st international conference on telecommunications and signal processing (TSP)*, (pp. 1–5).
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., . . . Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in microbiology*, *6*, 771.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.
- Wang, R. Y., & Popović, J. (2009). Real-time hand-tracking with a color glove. *ACM Transactions on Graphics (TOG)*, *28*, 1–8.
- Zabala, U., Rodriguez, I., Martínez-Otzeta, J. M., Irigoien, I., & Lazkano, E. (2021). Quantitative analysis of robot gesticulation behavior. *Autonomous Robots*, *45*, 175–189.
- Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.-L., & Grundmann, M. (2020). Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214*.
- Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, *35*, 696–742.