A Vision-Based System for Automated Sign Language Teaching

Kostas Blekos, *Member, IEEE*, Dimitrios Kosmopoulos, *Member, IEEE*, and Klimis Antzakas

***Abstract*—Sign languages are used by millions of individuals who are deaf or hard-of-hearing worldwide, yet learning them can be challenging due to limited access to qualified instructors and real-time feedback. Sign language learning is crucial for inclusion and accessible communication; however, existing e-learning solutions often fail to provide detailed or explainable feedback. To address this gap, we present a novel computer-assisted system that leverages computer vision and machine learning to teach sign languages. In our approach, students follow prerecorded videos of an instructor performing signs, and the system captures the students' own sign production through a conventional webcam. By analyzing both static handshapes and dynamic trajectories, our method provides detailed and explainable feedback at the level of keyframes, pointing out specific misalignments in hand, wrist, and motion path. This differs from many existing tools that offer only binary correct/incorrect evaluations without guidance for improvement. We implemented the system as a prototype with a full curriculum and conducted a pilot usability evaluation with 43 participants, ranging from beginners to skilled signers. Results indicate that users found the platform intuitive, the feedback mechanism informative, and the overall experience motivating for continued practice. The pilot assessed user perceptions of usability and feedback quality; learning outcomes were not measured in this initial evaluation. Notably, our design is language-agnostic, enabling broad application across different sign language communities. By offering instant fine-grained feedback, our system aims to improve self-guided learning of sign languages and promote greater accessibility, ultimately bridging communication gaps and fostering inclusion for the Deaf community.**

***Index Terms*—Computer vision, machine learning, sign language teaching.**

I. Introduction

According to the World Health Organization [1], up to five out of every 1000 children are born who are deaf or hard of hearing. Those children, especially those who may not have easy access to in-person sign language education, could benefit from an affordable system that could provide an accessible and tailored learning experience. Furthermore, some of the key beneficiaries include the family members and friends of individuals who are deaf, educators, sign language interpreters, schools and educational institutions, health professionals, emergency responders, researchers, and linguists.

There have been numerous approaches for teaching sign languages by employing e-learning platforms sparked by high accessibility, ease of use, and the flexibility in learning at their own pace (see, e.g., [2] and [3]). However, as stated in [3], users were demotivated to return to the site due to the lack of a validation mechanism.

Several methods for automated evaluation of students' performance have been presented in the past; however, the feedback that they provide is typically binary, i.e., correct or incorrect [4], [5]. This may be better than no feedback whatsoever, but still not very helpful to the student who may not understand how to improve their signing.

In this work, we aim to cover that specific drawback. We present an innovative and affordable system/service for interactive sign language teaching for everyone interested in learning sign languages. It is based on computer vision and machine learning and provides detailed instructions on how to improve signing. Our system supports learning both static handshapes and dynamic signs with movement trajectories and specific locations. The teacher's signs are prerecorded, and each sign includes a label indicating its meaning. Users start by practicing static handshapes as a foundation and then progress to whole words and phrases, which involve dynamic movements.

One of the key novelties of our system is that it is language-agnostic. We have tested it with datasets in German, Greek, and Portuguese sign languages. The system can be adapted for use with different sign languages without significant modifications.

The system is expected to facilitate the teaching process of sign language courses for the key beneficiaries mentioned earlier, as well as help any individuals interested in learning and mastering their sign language skills. Today, it is often challenging to find qualified tutors in sign languages other than the officially recognized local variety, a situation that significantly impacts learners' opportunities and access to inclusive education [6], [7]. Online and computer-assisted solutions like our proposed system have the potential to partially address this gap by offering multilingual support and more flexible accessible learning opportunities. In the long term, such tools could significantly narrow existing communication and cultural gaps between communities of people who are deaf and those who are hearing, promoting broader inclusivity in education and society [7], [8].

Received 20 March 2024; revised 27 September 2024, 25 March 2025, 12 August 2025, and 20 November 2025; accepted 18 February 2026. Date of publication 4 March 2026; date of current version 26 March 2026. This work was supported by the Greek State Scholarships Foundation under Grant 2024-1-EL01-KA220-HED-000257847 (Visual Interactive System for Teaching and Assessment of Sign Languages). *(Corresponding author: Kostas Blekos.)*

Kostas Blekos and Dimitrios Kosmopoulos are with the Department of Computer Engineering and Informatics, University of Patras, 26504 Patras, Greece (e-mail: mplekos@upatras.gr; dkosmo@upatras.gr).

Klimis Antzakas is with the Department of Educational Sciences and Social Work, University of Patras, 26504 Patras, Greece (e-mail: k.antzakas@upatras.gr).

Digital Object Identifier 10.1109/TLT.2026.3670474

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

More specifically, the contributions of this work are multifold.

1) We present a framework for tracking the student for both static and dynamic signs employing motion by using a conventional web camera.
2) We provide detailed feedback on both handshape (real-time) and motion trajectory, allowing students to practice independently and receive immediate guidance on their signing technique.
3) Our system is language-agnostic, capable of working with multiple sign languages without requiring significant modifications.
4) We employ a novel scoring mechanism that takes into account not just the static handshape but also the dynamics of the sign, including speed, acceleration, and directional changes. This provides a more comprehensive evaluation of sign accuracy.
5) Our approach offers explainable feedback through visual cues (color-coding and skeleton visualization), making it easier for learners to understand and correct their mistakes.
6) We evaluated the proposed methodology and system in a pilot user study with 43 participants and performed detailed technical verification of the results. Our evaluation focused on perceived usability and feedback quality rather than learning outcomes, consistent with the exploratory nature of a pilot study.

Based on the above, we implemented a full curriculum for sign language teaching and integrated it into an e-learning platform.

The rest of this article is organized as follows. In Section II, we discuss the related work to highlight the added value of our approach. In Section III, we describe the proposed methodology. In Section IV, we present the results of user evaluation. Finally, Section V concludes this article.

II. Related Work

The application of computer-assisted technologies in sign language teaching has seen significant advancements, yet it is not without its limitations. This section examines the current methodologies and tools, as well as the challenges and the constraints that they present.

Sagawa and Takeuchi [9] present one of the first approaches using a glove and an avatar for teaching Japanese Sign Language. The system is interesting; however, the gloves can be costly and, unlike visual sensors, they cannot capture the visual cues like body motion or facial expressions.

Zhang et al. [10] present an interesting system for teaching Chinese Sign Language using a mobile phone. The application provides a full environment that allows learning of a set of words in sequential or random fashion. It also allows testing the users' knowledge. The method uses point clouds, which are more lightweight than video streams. The sign is recognized by segmenting the feature stream in the time axis and then by using a neural net. The top-three classes are selected and presented to the student. The higher the sign probability, the higher the grade of the student. However, the method does not explain to the student possible mistakes.

Shao et al. [11] describe an augmented reality (AR) sign language teaching system in which detailed feedback is given to the user about improving their sign production. While their system provides valuable insights, our approach differs by offering explainable feedback through visual cues and being language-agnostic, capable of working with multiple sign languages without significant modifications.

Rho et al. [4] utilize a depth-sensing device (LeapMotion) to extract the handshape in 3-D and use a neural network to classify the handshapes of the New Zealand Sign Language alphabet. The network employs handcrafted features, and the problem is formulated as a classification problem, i.e., a different label for each letter of the alphabet. While the system shows potential, its effectiveness is contingent on the availability of the sensor, which may not be widely accessible due to cost and market availability. In addition, the formulation as a classification problem limits seriously the feedback provided to the user to a binary response (correct or incorrect).

Truong et al. [5] employ a monocular camera and a prosthetic arm to visualize the signs. They use the MediaPipe library to extract key points as we do. Then, a gated recurrent unit-based model is trained to recognize words based on the data. The system has been demonstrated on Vietnamese Sign Language on the alphabet and the numbers from zero to nine. The prosthetic arm, though interesting, cannot be assumed to be available to every student. Furthermore, the dynamic signs are classified with only a label, so the student is not able to receive detailed feedback on potential mistakes.

Alam et al. [12] present recent work using machine learning to recognize signs in a learning context. However, their approach does not focus on providing explainable feedback to learners, which is a key aspect of our system.

Phan et al. [13], [14] present an avatar-based system that has several modes of interaction with the user, one of them being arrows to indicate the wrong position, paths to indicate the full trajectory, and a so-called phantom visualization where two gray arms of the teacher performing the correct sign are superimposed on the student avatar performing the student's recorded movements. The provided feedback is far more detailed and useful than the other approaches. It is not clear, however, how the "phantom" operation actually synchronizes with the user motion, e.g., in case of different sign duration. Furthermore, the system focuses on gross movements and does not provide feedback in real time.

The teaching of sign languages has been often treated as an immersive experience. Wang et al. [15] implement an immersive virtual environment for learning the American sign language (ASL) digits from zero to nine with a Whack-a-Mole type of game. Then, they found out by user validation that the immersive environment was far more appealing for the users.

Quandt et al. [16], [17] develop a system where participants can watch and then replicate sign language gestures demonstrated by a virtual avatar in a fully immersive virtual environment, facilitated by a head-mounted display. To capture the necessary movements for the signs, a sign language instructor donned a motion capture suit that recorded the movements of the body, hands, and fingers. The findings from this study underscored the favorable response of sign language learners to using such technology for their sign language education. In

a similar fashion, Kasparakis et al. [18] describe the creation and assessment of a system for learning and communicating sign language in virtual environments. The system is structured to replicate real-world learning and testing scenarios for sign language students within a virtual reality learning environment. Such immersive environments for learning are very useful for research, but it seems rather hard to become mainstream due to complex setup and relatively high costs.

There have also been several attempts in the past to combine learning of sign languages with games, including some rather basic computer vision and machine learning approaches [19].

The studies discussed so far provide valuable insights into the potential of computer-assisted technologies for sign language teaching. However, they also highlight the need for more comprehensive research to overcome the limitations of current systems. Issues such as the high cost of technology and the need for explainable feedback are yet to be fully resolved. While the state of the art in computer-assisted sign language teaching demonstrates innovative use of technology, the limitations of these studies underscore the necessity for ongoing research and development. Addressing these limitations is crucial for developing more effective, inclusive, and accessible sign language teaching tools and methodologies.

Recent surveys and reviews have provided comprehensive overviews of the challenges and advancements in sign language recognition and translation. Adeyanju et al. [20] critically analyzed machine learning methods applied to sign language recognition, highlighting both progress and persistent challenges. Similarly, Najib [21] discussed the integration of artificial intelligence (AI) in sign language interpretation, emphasizing the need for more robust and interpretable solutions.

Broader surveys, such as [22], underscore the diverse methodologies—from vision-based recognition to translation—while earlier efforts like [23] and [24] laid the groundwork by exploring live gesture feedback and computer-assisted translation. In the realm of feedback, Paudyal et al. have contributed significantly—with work in 2019 [25] and 2020 [26]—by demonstrating the benefits of explainable AI for sign language learning. Prior work on developing error taxonomies for computer-assisted language learning, such as that by Suri and McCoy [27], highlights the importance of categorizing learner mistakes to enhance instructional feedback. More recently, Hassan et al. [28] have investigated learner preferences for automatic feedback systems, and Natarajan et al. [29] have proposed end-to-end deep learning frameworks for sign language processing. Tan et al. [30] have recently reviewed deep-learning-based approaches extensively, highlighting the increasing application of generative AI and large language models in the domain of sign language recognition.

Finally, while most of the existing translation systems focus on professional usage [31], alternative approaches—such as Liu Aiqin's Large Screen Language for the education of students who are deaf [32]—offer insights into innovative classroom-based methods. These studies contextualize the current state of sign language technologies but also reinforce the need for detailed, explainable, and language-agnostic feedback mechanisms, as proposed in our work.

III. Methodology

A. Overall Description

Our system is designed to handle both static handshapes and dynamic movements, capturing the full motion sequence of signs. The keyframe extraction algorithm and trajectory analysis capture both the significant positions and movements in sign language gestures. Specifically, the system analyzes the entire video input, identifying key points in the motion (keyframes) and evaluating the trajectory between these points. This allows us to assess both static handshapes and dynamic movements crucial for sign language interpretation.

The system architecture for the feedback mechanism is as follows (see Fig. 1).

1) *User interaction:* The process begins with user interaction, where the user is able to see the available signs for practice, select one of them, and initiate and terminate the recording of their signing motion using manual controls ("record"/"stop"). The teacher's signs are prerecorded, and each sign includes a label indicating its meaning.
2) *Landmark extraction:* Upon recording, we use Google's MediaPipe framework [33] to extract hand landmarks from the video footage.
3) *Keyframe extraction:* The system then applies the keyframe extraction algorithm to determine keyframes within the recorded video. Keyframes are significant frames where hand gestures are characteristic for sign interpretation. This algorithm is applied to both the user's and the teacher's recordings, with the teacher's keyframes serving as a ground truth reference for the user's performance.
4) *Scoring mechanism:* Each keyframe pair, comprising a user's keyframe and its corresponding teacher's keyframe, is then assessed using the scoring mechanism. This mechanism evaluates the similarity between the user's and the teacher's handshapes, assigning a score based on their congruence.
5) *Feedback presentation:* Finally, the system presents a comparative scoring analysis to the user. This analysis visually demonstrates how to correct the pose or motion to match the teacher.

B. Keyframe Extraction

Our goal is to extract keyframes from the signing videos, as they play a crucial role in our feedback system. Keyframes provide a concise and efficient means to represent videos, eliminating the need for full playback. This allows us to provide user feedback effectively by visually annotating these keyframes.

We start by performing landmark extraction. To this end, we use MediaPipe hand landmark tracking [33]. The hand landmarks are extracted for each frame in the video sequence.

A clustering step is introduced to group frames based on both spatial and temporal characteristics of the hand movements. This is crucial for analyzing not just the hand positions, but also the sequence in which these positions occur over time. The integration of the temporal dimension is achieved by calculating

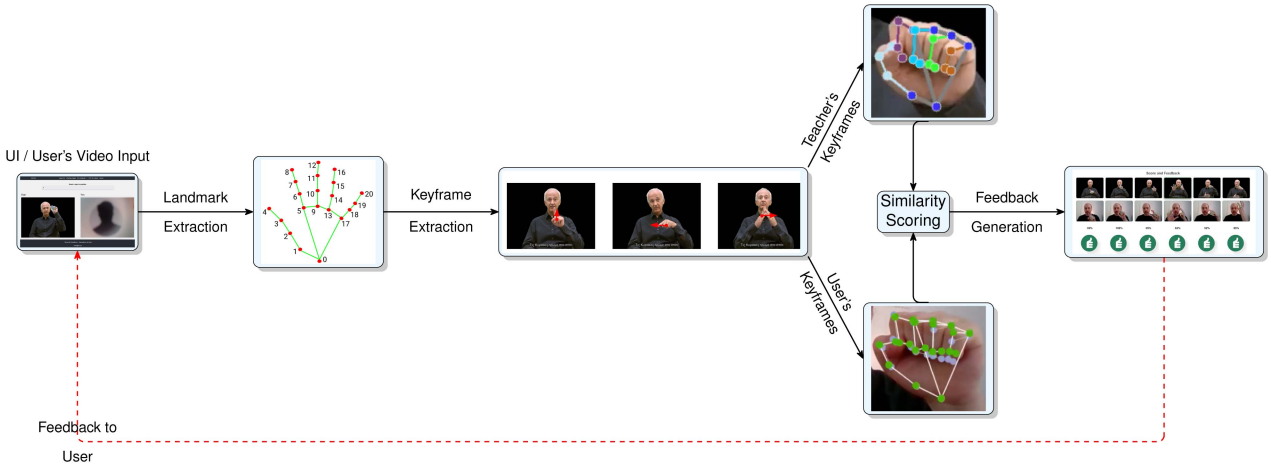


Fig. 1. Architecture of the proposed system. The user interacts with the system via the user interface and video is captured via a camera; the video is processed for landmarks; then, the keyframes of the motion are extracted and compared to the ground truth by the teacher; the respective matching score is extracted.

the centroids of the hand landmarks, where the z-coordinate of each centroid is replaced with the frame's relative position in the video timeline. All vector coordinates are normalized to ensure that none of them will dominate the distance calculated for clustering. This method effectively combines spatial and temporal data, allowing for a more nuanced clustering.

The number of clusters, k, is an estimated parameter and can be adjusted based on the specific requirements of the video or the nature of the signs. Adjusting k alters the granularity of the clustering, with a higher number of clusters providing more detailed segmentation. A simple k-means algorithm is employed for clustering, utilizing the centroids with integrated time dimension. Clustering is an important step as it groups frames into segments based on the similarity of hand movement patterns. This approach aids in identifying distinct gestural phases within the video.

Within each cluster, the frame that exhibits the lowest speed for the hand centroids is selected as the keyframe. This choice is predicated on the assumption that key moments in hand gestures, particularly in sign language, often correspond to points of minimal hand movement [34]. By focusing on the frame with the lowest speed in each cluster, the method effectively captures significant gestural positions or transitions, which are essential for interpreting the sequence of movements.

C. Calculation of Handshape Distance

In our next step, we compare pairs of keyframes by the teacher and the student to find how the depicted handshapes match. To quantify this matching, we need a distance metric.

A popular distance that is typically used for point cloud registration is the Procrustes distance [35], which is often used to optimally superimpose two shapes by translating, rotating, and scaling them, and then measuring their discrepancy. It has been used for handshape analysis (see, e.g., [36]). However, such a distance does not necessarily capture the semantic importance of each finger, as it treats all of them in a similar fashion. A comparison of our scoring with the standard Procrustes distance is shown in Fig. 10.

Here, we mitigate this issue by defining a custom distance using the hand landmarks. These landmarks undergo preprocessing, including alignment, scaling, and rotation, to normalize for various hand sizes and positions. The scoring mechanism then calculates the Euclidean distances between corresponding landmarks, weighted by their anatomical significance in sign language [37]. This method allows us to include semantics by weighting the landmarks, setting limits to the transformations that we regard as acceptable, and by adding additional components that capture the semantics of the thumb finger.

In the following, we describe how the distance is calculated starting from the landmark estimation: Two sets of 21 hand landmarks are acquired using the MediaPipe Hand module. Each landmark is a point in the 3-D space (x, y, z), representing key anatomical locations on the hand.

Preprocessing and normalization:

1) *Alignment:* The sets of landmarks are initially translated for alignment, that is, we translate the landmarks so that their respective wrists (landmark 0 in the MediaPipe hand model) coincide in space.
2) *Scaling:* The landmarks are scaled to normalize for hand size differences. The scaling factor is determined by the distance between the wrist (landmark 0) and the base of the middle finger (landmark 5). This distance is normalized to a standard value, 1 unit, along the y-axis for both hand landmark sets.

Rotation: In the rotation process, each landmark's position is recalculated relative to the wrist. This transformation maintains the distances between landmarks while altering their angles to align with those in a reference set. We calculate the rotation needed to align the student's handshape with the teacher's one. We set a maximum rotation threshold that limits the angle by which any landmark can be rotated in a single adjustment. This threshold is essential for allowing flexibility in comparing hands of different shapes and sizes, while still preserving the general structure and anatomy of the handshape. The threshold value is set by considering the natural range of motion of hand joints and is experimentally set to $\pi/40$ (approximately 4.5°).

Scoring mechanism: In the scoring mechanism, each hand landmark's contribution to the overall similarity score is weighted according to its anatomical significance in sign language [37]. This is crucial because certain parts of the hand, such as fingertips, play a more significant role in sign language gestures than others. On the other hand, some differences should not be penalized as they might arise from natural pose differences, anatomical differences between individual's hands, etc. Our proposed mechanism balances flexibility (that allows matching between different hands) and matching accuracy.

Distance calculation: For each pair of corresponding landmarks (e.g., landmark 1 from both hands), the Euclidean distance in the 3-D space is calculated.

Weighting scheme: Each distance is weighted based on its anatomical significance. For instance, fingertips (landmarks 4, 8, 12, 16, and 20; see Fig. 5) are given higher weights due to their importance in the interpretation of a sign. The weights are predefined constants and are based on our empirical studies. We used a simple formula that proved to work well

$$\mathrm{w}_i = \frac{1}{1 + d^i_{\text{joint}}} \tag{1}$$

where d^i_{joint} represents the distance of the ith joint from the base of the palm, measured in terms of the number of joints between them

$$d^i_{\text{joint}} = \begin{cases} 0, & \text{if } i = 0 \text{ (wrist)} \\ 4 - ((i-1) \bmod 4), & \text{if } i \neq 0 \text{ (other joints).} \end{cases} \tag{2}$$

Special consideration is given to thumb landmarks (indices 3 and 4), where they receive a slightly increased weight (by subtracting 0.5 units from their index). This is due to the high importance of the thumb in signing. Fig. 6 quantifies the weights.

Score computation: The final distance score is the sum of the weighted distances. The weights are normalized so that their sum equals 1, ensuring that the final score is a weighted average of the distances. Furthermore, a penalty is introduced for differences in the general direction of the thumb between the compared sets of landmarks. This penalty is calculated using the dot product of the directional vectors of corresponding thumb landmarks. These vectors represent the direction from one thumb landmark to the next. The dot product measures the similarity in direction between the thumbs of the two handsets. A penalty is then applied based on the angular difference between these directions. The more the directions diverge, the greater the penalty, increasing the overall score and indicating a lesser match. This approach ensures that the overall matching score correctly accounts for the role of thumb orientation.

More specifically, the normalized weight $\hat{w}_i$ of landmark i is given by

$$\hat{w}_i = \frac{\mathrm{w}_i}{\sum_{j=1}^{N} \mathrm{w}_j}. \tag{3}$$

Let us assume that the set of landmarks for the teacher hand is $\mathbf{L}_t = \{\mathbf{x}^i_t, i \in \{0, \ldots, 20\}\}$ and for the student hand $\mathbf{L}_s = \{\mathbf{x}^i_s, i \in \{0, \ldots, 20\}\}$. The distance between the two point sets after normalizations and alignment is given by

$$D_s(\mathbf{L}_s, \mathbf{L}_t) = \sum_{i=1}^{20} \hat{w}_i ||\mathbf{x}^i_t - \mathbf{x}^i_s||_2 \tag{4}$$

i.e., the weighted sum of the Euclidean distances of the respective landmark points.

The thumb direction vectors are defined as

$$\mathbf{v}^i = \mathbf{x}^{i+1} - \mathbf{x}^i, \quad i \in \{0, 1, 2, 3\} \tag{5}$$

and represent the directional vector from one thumb landmark to the next. The "thumb distance" for each of them is calculated as follows:

$$d^i_{\text{th}} = 1 - \frac{\mathbf{v}^i_t \cdot \mathbf{v}^i_s}{\|\mathbf{v}^i_t\| \|\mathbf{v}^i_s\|}, \quad i \in \{0, 1, 2, 3\} \tag{6}$$

i.e., it is based on the angular difference between the thumb directions of the user and the reference, calculated as the cosine similarity between the two vectors

$$D_{\text{th}}(\mathbf{L}_s, \mathbf{L}_t) = \sum_{i=0}^{3} \hat{w}_i d^i_{\text{th}}. \tag{7}$$

The total distance is then given as

$$D(\mathbf{L}_s, \mathbf{L}_t) = D_s(\mathbf{L}_s, \mathbf{L}_t) + D_{\text{th}}(\mathbf{L}_s, \mathbf{L}_t). \tag{8}$$

Distance interpretation: The distance is interpreted on a predefined scale. A lower value (closer to 0) indicates high similarity between the handshapes, whereas a higher distance suggests greater dissimilarity. Values below a certain threshold, identified as 0.35 experimentally, are considered as indicating a "match" in sign recognition (see Figs. 7 and 8). The sensitivity levels for the scoring mechanism were determined empirically with input from sign language teachers. Parameters were adjusted to provide useful feedback to learners, balancing between strictness and flexibility to account for natural variations in handshapes. While the threshold of 0.35 has demonstrated efficacy in various experimental setups in our study, this value cannot be taken as universally optimal; adaptation may be required for different contexts.

Color normalization of distances: The distances for handshapes are normalized to a range, such as 0–1, where 0 indicates identical handshapes and 1 indicates maximum dissimilarity. This normalization aids in setting universal thresholds for recognition accuracy across different instances and users. The initial scoring, based on the distance between corresponding landmarks, produces a range of values from 0 to potentially a very high number, with 0 indicating no distance (perfect similarity) and higher numbers indicating greater dissimilarity. The normalization is achieved by setting a threshold, which represents the upper limit of distance. Any distance above this threshold is capped at the threshold, and values below this are linearly scaled effectively compressing the score range into a 0–1 scale. To visualize these normalized distances for real-time feedback to the user, we used a color-coding system. This system uses a gradient of colors, transitioning from green to red, to indicate varying levels of similarity between handshapes: a lower

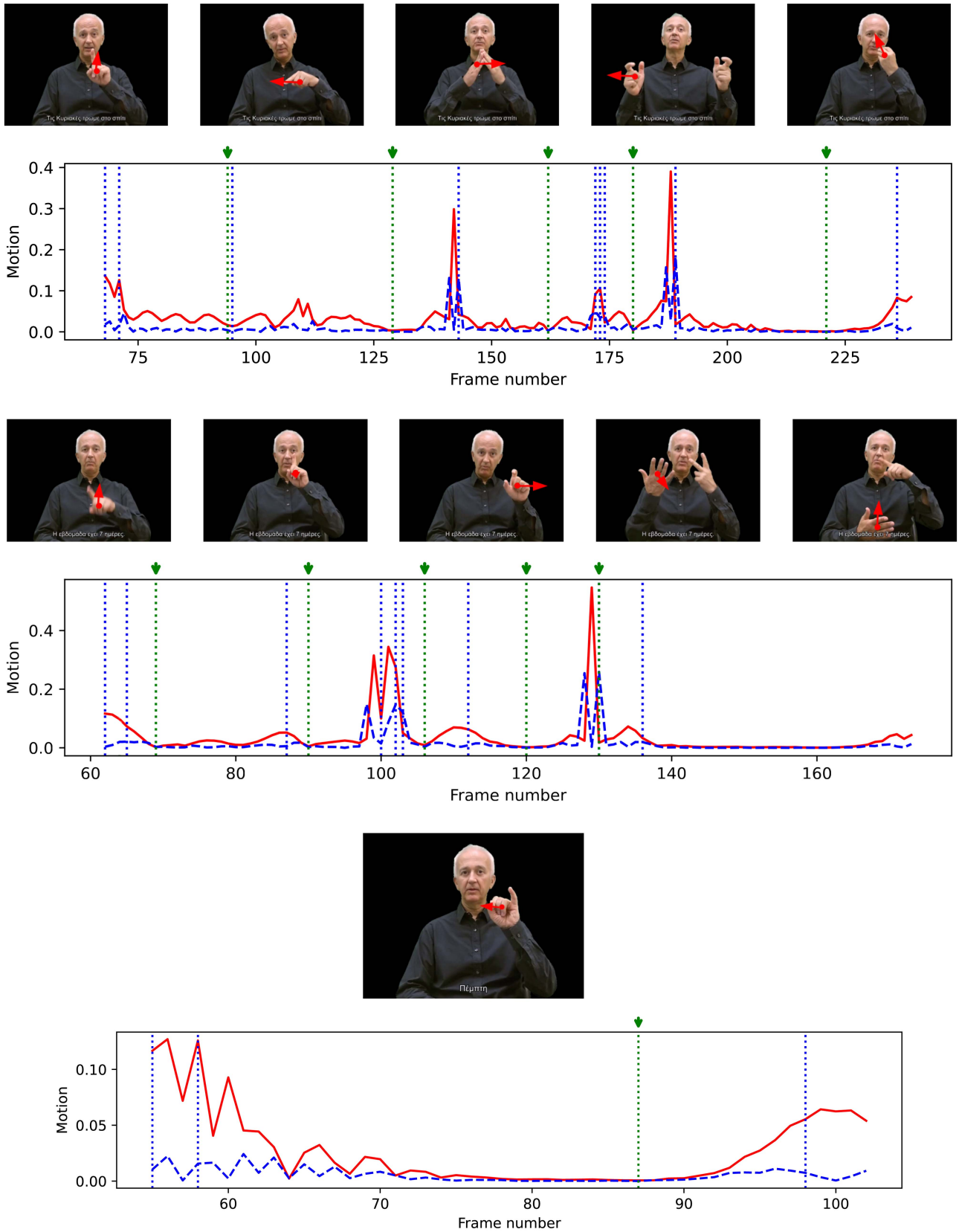


Fig. 2. Keyframe extraction; keyframes (top) trajectory motion data (below)—red arrows signify the current direction of motion; vertical green dotted lines: keyframe identification; vertical blue lines: cluster change; blue/red line: speed/acceleration of centroid. For the final score [see (9)], we compare the keyframes of the teacher and the student and assign a handshape distance (D) and a motion penalty (P) for each keyframe pair.

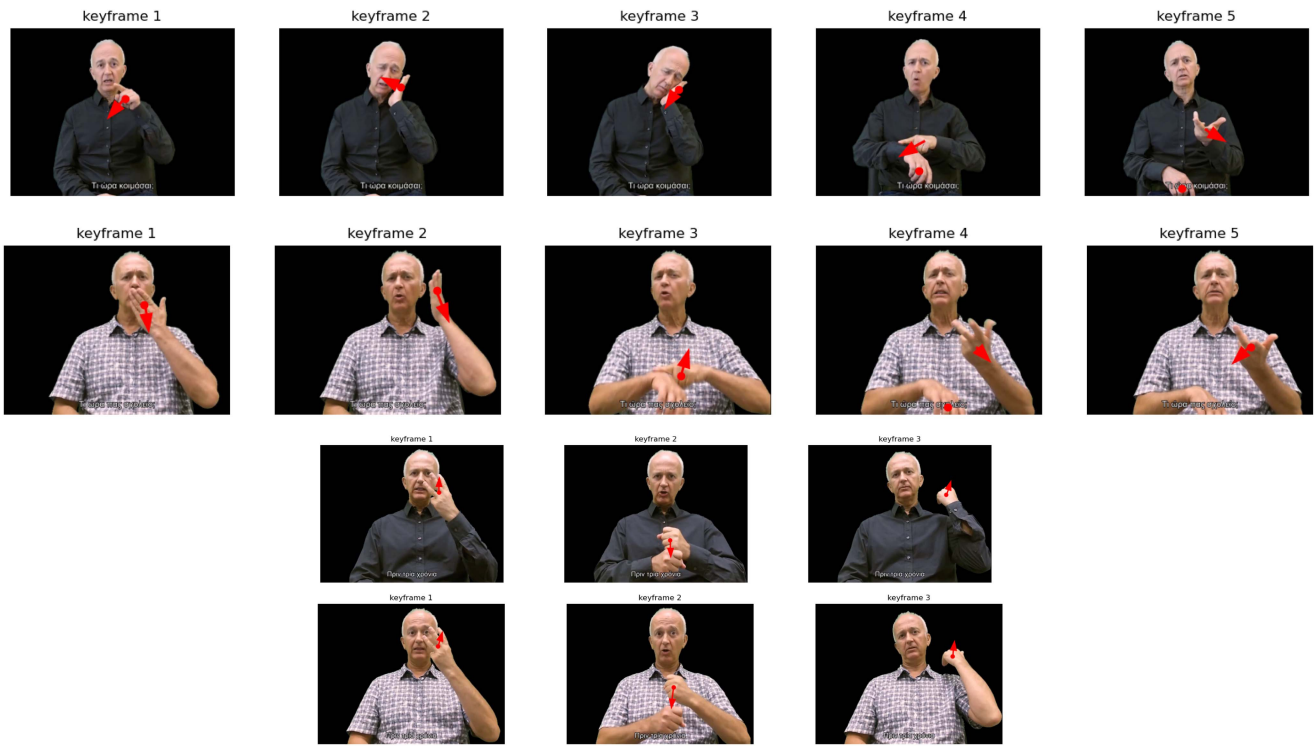


Fig. 3. Example of trajectories (red arrows) in videos as a teacher (black shirt) or simulating a student (gray shirt).

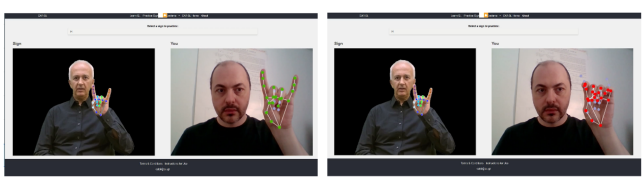


Fig. 4. Scoring example. Red is far (toward score 1); green is close (toward score 0).

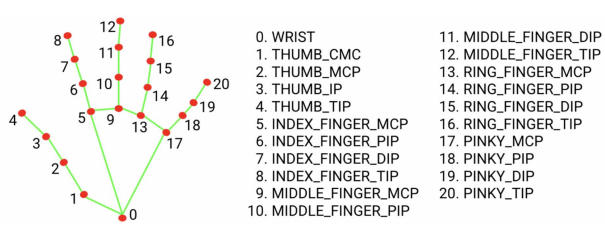


Fig. 5. Detected hand landmarks (from https://developers.google.com/mediapipe/).

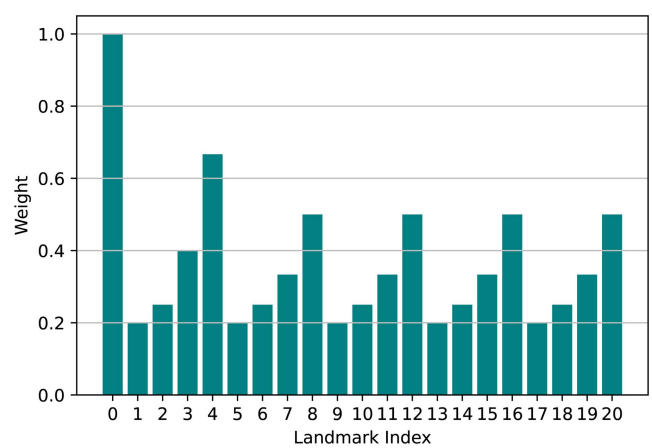


Fig. 6. Weights for each landmark.

distance, which suggests a high degree of similarity, is represented by a greener color, while a higher distance, indicating greater dissimilarity, is depicted with a deep red color.

The feedback is provided to the students in real time through visual cues. The system uses color-coding to indicate the level of correctness of the handshape. The entire hand is colored according to the calculated similarity score, with green indicating a close match and red indicating a significant discrepancy. While individual finger coloring was considered, user testing showed that full hand coloring provided clearer and less confusing feedback. The real-time feedback allows students to adjust their handshape immediately, facilitating quicker learning. An example of the color-based scoring feedback is shown in Fig. 4.

D. Orientation

When the student's hand landmarks are transformed to match the teacher's, only minimal deviations in angles are permitted. This is achieved by setting a maximum rotation threshold for each landmark, which is based on the natural range of motion of hand joints. The threshold is calibrated to allow some flexibility for anatomical differences, yet it is restrictive enough to ensure that significant deviations in orientation result in a noticeable increase in the calculated distance metric. Consequently, a handshape that is significantly misoriented compared to the teacher's will accrue a higher distance, reflecting a lower

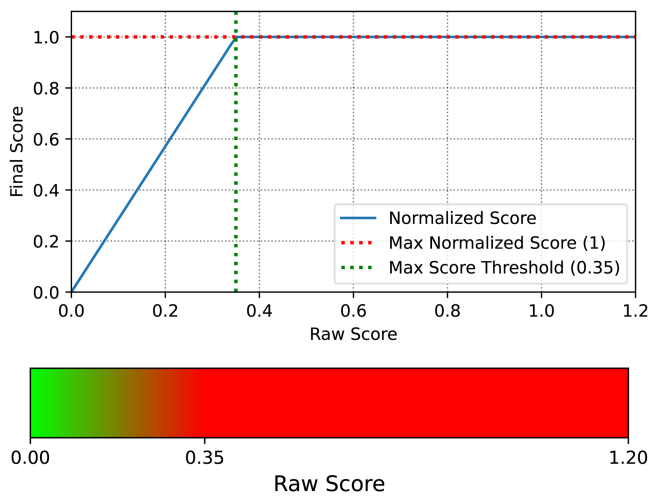


Fig. 7. Normalization of raw score (0 to infinity) to a color range 0–1.

similarity. This method effectively discriminates between minor permissible variations and major critical deviations. Examples of orientation-related scoring differences are shown in Fig. 9 and Fig. 10.

E. Trajectory

The overall motion trajectory of the teacher's and student's hand gestures is evaluated. We use the hand's centroid speed and acceleration data to assess the characteristics of the motion. The centroid is used as a representative point for the hand trajectory, as it also captures changes in the hand shape and orientation when there is no significant overall movement.

A penalty term, denoted as P, is introduced to quantify the alignment between the teacher's and student's hand gestures. Specifically, this term acts as a penalty when the direction vectors of keyframes (red arrows in Figs. 2 and 3) in the teacher's and student's hand gestures exhibit low cosine similarity

$$P_i = 1 - \left| \frac{\vec{v}_{t_i} \cdot \vec{v}_{s_i}}{|\vec{v}_{t_i}||\vec{v}_{s_i}|} \right|$$

where $\vec{v}_{t_i}$ and $\vec{v}_{s_i}$ are the direction vectors of the teacher's and student's hand gestures, respectively, at the ith keyframe. The value of P_i ranges from 0 to 1, with 0 indicating perfect alignment and 1 indicating a complete mismatch in the direction of the hand gestures.

This penalty term ensures that not only the speed and acceleration but also the directional similarity of the gestures are considered in evaluating the student's mimicry of the teacher's motions.

F. Overall Evaluation

In the final stage of the system, we synthesize the key insights from both handshape and motion analysis to estimate the overall performance. This is achieved by first calculating the distance between corresponding keyframe pairs between the student and the teacher. This distance primarily reflects the deviation from an accurate handshape replication. Next, as discussed above, we introduce a penalty factor P_i for discrepancies in the motion trajectory, which considers the dynamic aspects of the sign language gestures, such as speed and directionality. The final score is then obtained by aggregating these components into a weighted sum, where the weights are adjusted to reflect the relative importance of handshape accuracy and motion congruence in the context of sign language interpretation

$$S = w_d \cdot D + w_p \cdot P \tag{9}$$

where D and P are the normalized distance and penalty scores, respectively, and w_d and w_p are the corresponding weights.

For n keyframe pairs, the aggregate scores D and P are computed as

$$D = \frac{1}{n} \sum_{i=1}^{n} D(\mathbf{L}_s^i, \mathbf{L}_t^i) \tag{10}$$

$$P = \frac{1}{n} \sum_{i=1}^{n} P_i \tag{11}$$

where P_i is the penalty for the ith keyframe pair, reflecting any misalignment in the trajectory between the student's and the teacher's gestures.

The weighting scheme can be fine-tuned based on empirical studies to reflect the relative importance of handshape accuracy and motion congruence in the context of sign language interpretation, or specific requirements of the application.

TABLE I
PARTICIPANT DEMOGRAPHICS AND SIGN LANGUAGE EXPERIENCE ($N = 43$)

Characteristic	Category	n	%
Age	15–25	1	2.3
	26–35	17	39.5
	36–45	17	39.5
	46–55	8	18.6
Sign-language knowledge	Excellent	14	32.6
	Very good	14	32.6
	Good	1	2.3
	A little	5	11.6
	Not at all	9	20.9
Motivation	Personal interest	23	53.5
	Professional use	10	23.3
	Research	4	9.3
	Teacher	3	7.0
	Deaf/HoH	3	7.0

IV. PILOT USABILITY EVALUATION

A. Participants and Study Design

Our experimental evaluation consisted of a comprehensive anonymous user study involving $N = 43$ volunteers. Participants reported age in predefined categories (15–25, 26–35, 36–45, and 46–55); the distribution was 26–35: 17 (39.5%), 36–45: 17 (39.), 46–55: 8 (18.6%), and 15–25: 1 (2.3%). Using bin midpoints, age was approximately $M = 37.9$, standard deviation (SD) = 7.9; these summaries should be interpreted with caution. Sex/gender were not collected; we note this as a limitation. Basic demographic and sign language experience data are summarized in Table I.

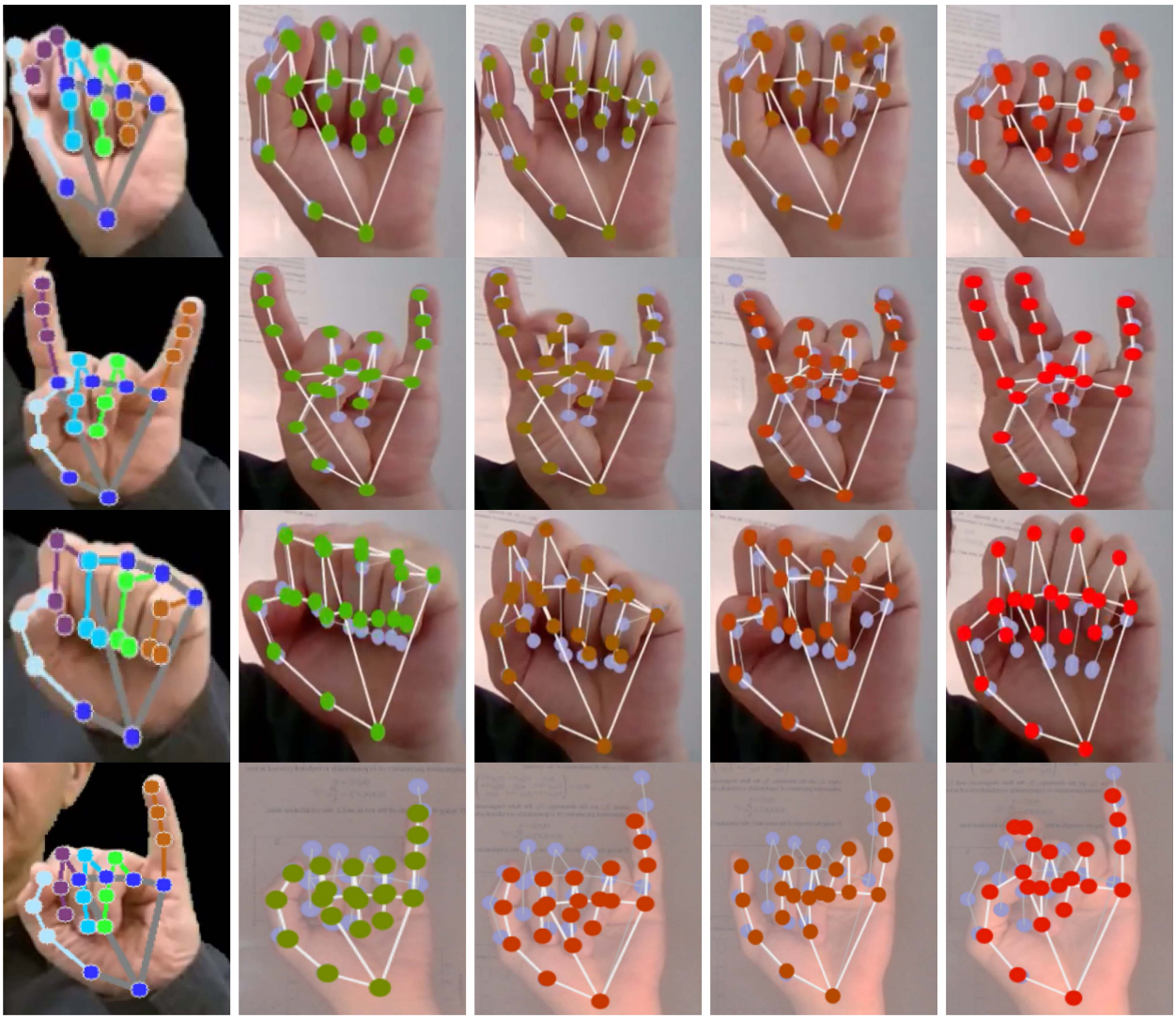


Fig. 8. Variations in handshape configuration and corresponding scoring. This figure illustrates different student handshapes as they attempt to replicate the teacher's standard handshape (shown in the left-most image in each row). Each row presents variations in the handshape configuration of the student, with the resulting scoring depicted through color-coding (green for close match; red for distant match). This underscores the scoring system's ability to detect and assess variations in handshape.

The study employed a mixed-methods approach combining quantitative Likert scale ratings with qualitative feedback opportunities. Demographic proficiency in sign language was assessed on a five-point scale: "Not at all" (1), "A little" (2), "Good" (3), "Very good" (4), and "Excellent" (5). Usability ratings were collected on four-point Likert items with anchors "Not at all" (1), "A little" (2), "Satisfactory" (3), and "Very much" (4). Intention to use the platform in future was assessed on a four-point scale: "Not at all" (1), "A little" (2), "Quite likely" (3), and "Sure" (4). Participants could optionally leave free-text comments about the system and suggestions for improvement.

Participants were recruited through multiple channels including academic institutions, sign language communities, and general public announcements to ensure diversity in background and experience levels.

B. Procedure and Measures

The evaluation protocol included.

1) *System demonstration phase:* Participants were first introduced to the system features through a guided demonstration.
2) *Hands-on interaction period:* Each participant engaged with the system independently for approximately 20 min, practicing both static handshapes and dynamic signs.
3) *Evaluation questionnaire:* Participants completed a structured questionnaire covering system usability, feedback quality, and perceived utility.

User profile: Participants were asked about their existing knowledge of Greek Sign Language and their primary motivations for learning. Table I presents the counts and percentages for each category. Motivations ranged from personal interest to professional or educational reasons, reflecting the diverse uses of signing in everyday communication, research and teaching.

Intuitiveness: Ease of navigation and intuitiveness of the menus are crucial for students who have all levels of exposure to e-learning environments. Most participants rated the user

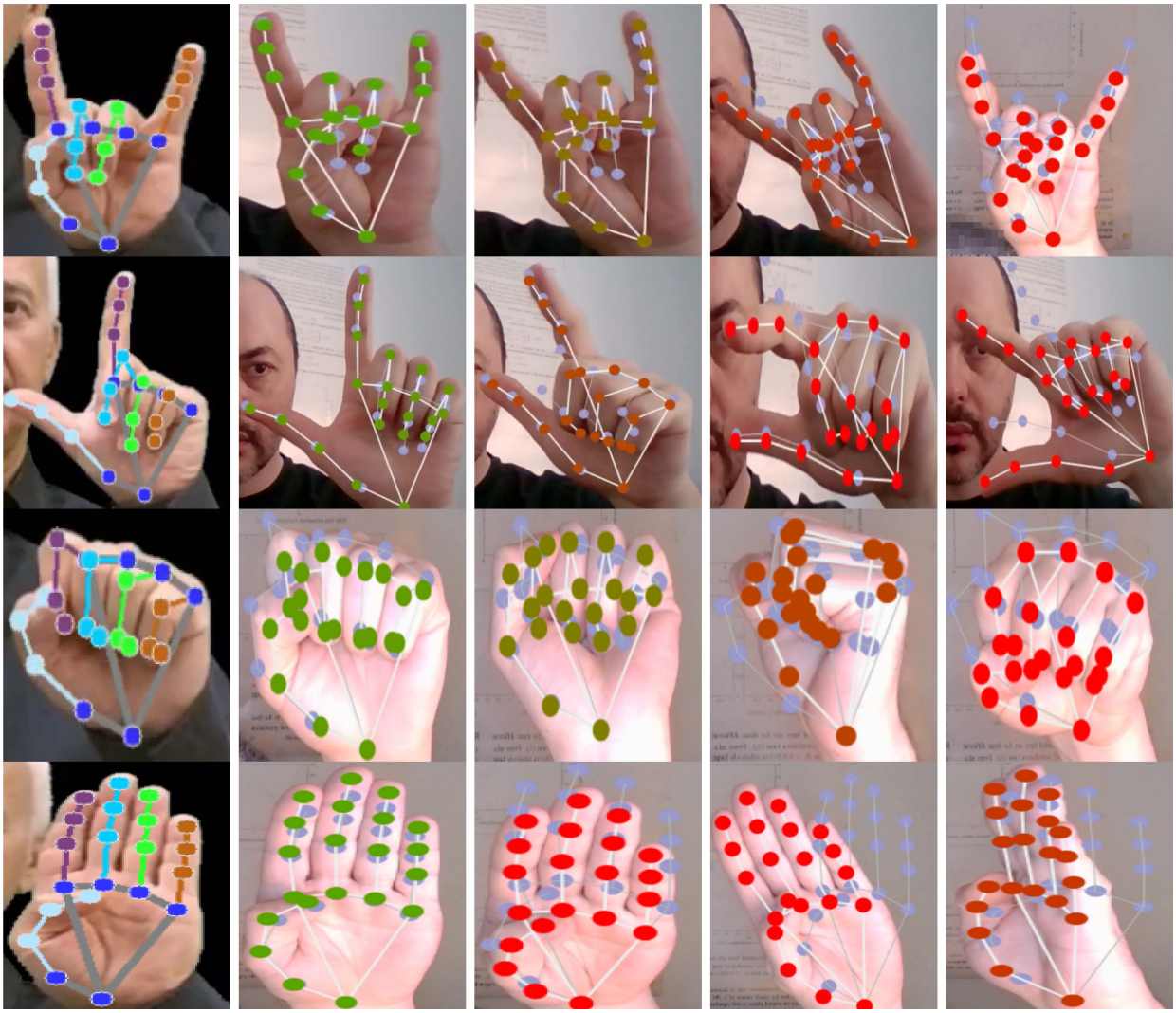


Fig. 9. Variations in handshape orientation and corresponding scoring. This figure illustrates different student handshapes attempting to match the teacher's handshape (left-most image in each row). Each row shows variations in orientation and position of the student's hand, with the resultant scoring depicted through color-coding (green for close match; red for distant match). This highlights the sensitivity of the scoring system to orientation and alignment differences. *Note:* The images are intended to demonstrate the feedback mechanism and do not depict actual student participants.

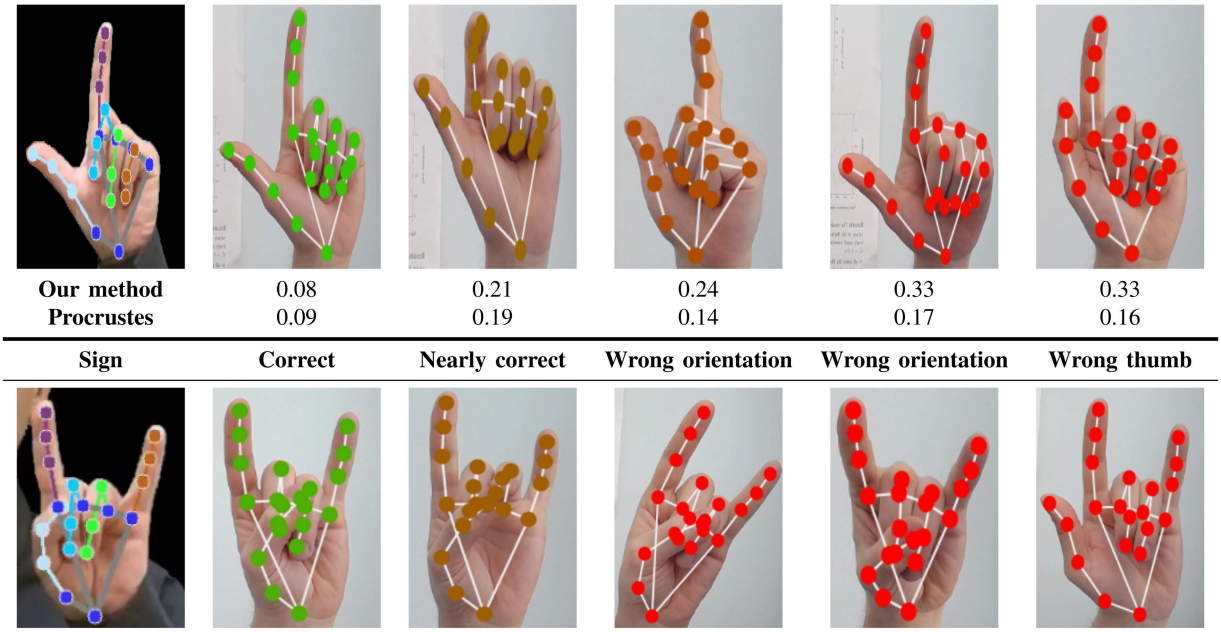
TABLE II
USER FEEDBACK ON SYSTEM FEATURES

Question	**1 Not at all**	**2 A little**	**3 Satisfactory**	**4 Very much**
How well do you think the meanings are conveyed by the image sequences?	4.65%	6.98%	41.86%	46.51%
How understandable is what the teacher is showing?	0%	4.65%	39.53%	55.81%
How useful do you find color as a means of capturing deviation?	0%	4.65%	51.16%	44.19%
How well is your hand tracking and following done?	0%	6.98%	48.84%	44.19%

interface as satisfactory or better on our four-point scale (41/43; 95.3%); 2/43 (4.7%) selected "A little."

Feedback quality: The opinions of the users on the scoring mechanism and feedback mechanism are particularly important for this study. Across these items, between 88% and 95% of participants rated them as at least satisfactory (see Table II). The questions that we addressed here are the following.

1) How well do the keyframes capture the sign videos? 38/43 (88.4%) rated this item as *Satisfactory* or *Very much*; 5/43 (11.6%) rated *A little* or *Not at all*.



Sign	Correct	Nearly correct	Wrong orientation	Wrong orientation	Wrong thumb
Our method	0.08	0.21	0.24	0.33	0.33
Procrustes	0.09	0.19	0.14	0.17	0.16
Sign	**Correct**	**Nearly correct**	**Wrong orientation**	**Wrong orientation**	**Wrong thumb**
Our method	0.11	0.22	0.68	0.44	0.37
Procrustes	0.10	0.18	0.13	0.13	0.13

Fig. 10. Comparison of our scoring method with a standard Procrustes score. This image shows a series of hand signs compared to a teacher's reference sign (first column). The student signs are scored with our method (left number) versus a standard Procrustes distance (right number). From left to right: the teacher's correct sign, a student's correct sign, a nearly correct sign, two signs with wrong orientation, and a sign with the thumb in the wrong position. Our scoring reflects the accuracy of the sign, with a closer match to the teacher's sign resulting in lower scores. The scores increase for signs with incorrect orientations and thumb placements, highlighting the precision of our method. When the sign is correct or almost correct, the two metrics are very similar. Our scoring method demonstrates enhanced sensitivity in penalizing incorrect orientations and thumb placements.

TABLE III
USER EVALUATION RESULTS WITH MEAN AND STANDARD DEVIATION

Question	Mean	Standard Deviation
How well do you think the meanings are conveyed	3.30	0.80
How understandable is what the teacher is showing	3.51	0.59
How useful do you find color as a means of capturing deviation	3.40	0.58
How well is your hand tracking and following done	3.37	0.62
How easy to use did you find the learning platform	3.56	0.59
How likely do you see it as possible to use a similar technique	3.28	0.67

2) How understandable is what the teacher is showing? 41/43 (95.3%) rated *Satisfactory* or *Very much*; 2/43 (4.7%) rated *A little*.
3) How representative is the color/skeleton feedback? 41/43 (95.3%) rated *Satisfactory* or *Very much*; 2/43 (4.7%) rated *A little*.
4) How effective is the hand tracking? 40/43 (93.0%) rated *Satisfactory* or *Very much*; 3/43 (7.0%) rated *A little*.

For transparency with ordinal data, the item medians were either 3 ("Satisfactory") or 4 ("Very much") across all questions; full response distributions are provided in Table II.

Future use: Here, the participants were encouraged to express the likelihood of using the platform for future sign language learning endeavors and to consider it as a supplementary tool in their courses. The results in this segment were promising: 38/43 (88.4%) selected "Quite likely" or "Sure," while 5/43 (11.6%) selected "A little."

Moreover, the feedback received during this evaluation shed light on areas for potential improvement and expansion. Participants expressed a desire for the incorporation of body pose and facial expressions in the platform. This valuable insight highlighted a demand for a more comprehensive learning experience.

C. Statistical Analysis of System Features

The quantitative data revealed consistently positive evaluations across all system features. As shown in Table III, mean scores across all evaluated dimensions exceeded 3.0 on our four-point scale, indicating strong overall satisfaction.

Qualitative analysis of free-text comments: Beyond the structured ratings, ten participants (23%) provided free-text comments. We analyzed these comments using a standard six-step thematic analysis procedure (familiarization, coding, theme generation, review, definition, and write-up), following Braun and

Clarke [38]. Two researchers independently open-coded the comments, met to reconcile codes, and grouped related codes into broader themes.

Three salient themes emerged.

1) *User interface clarity and instructional guidance:* Several comments requested clearer guidance (e.g., whether to mirror the instructor, visibility of handshapes under overlays) and suggested adding options like slower playback for first-time learners.
2) *Tracking robustness and content corrections:* A few comments noted that hand tracking could be sensitive under certain conditions (e.g., repeated attempts needed) and pointed out a content correction for one handshape.
3) *Overall positive sentiment:* Many comments expressed that the platform was easy to use and helpful, praising its practicality and usefulness for learning.

Representative comments included (translated where necessary):

> "It is not clear whether we should mirror the signer or follow with a slight delay."
>
> "Sometimes the overlay occludes the handshape; a slower playback option would help."

These themes informed our recommendations for future work (see Section V) and illustrate how qualitative feedback can complement quantitative ratings in mixed-methods evaluations.

Reason for learning sign language and intended future use: To examine whether learners' reason for learning sign language (*"Motivation"*) was associated with their stated "likelihood of using a similar system in the future," we treated *Motivation* as a nominal factor (five categories: Personal interest, Professional use, Research, Teacher, and Deaf/HoH) and cross-tabulated it against the three response levels that actually occurred ("A little," "Quite likely," and "Sure"). A chi-square test of independence showed no significant relationship among the five groups, $\chi^2(8) = 13.57$, $p = 0.09$; the eight degrees of freedom derive from the resulting 5×3 contingency table because no participant selected "Not at all." The effect size, Cramer's $V = 0.40$ (bootstrap 95% CI $\approx$ 0.30–0.55; 500 resamples), still points to at most a moderate association, and in view of the sparse cells and nonsignificant p-value, we refrain from treating this as evidence of a genuine link between motivation type and future uptake.

Effect of prior sign language knowledge on usability ratings: Participants who declared "Not at all/Little" knowledge were classed as *novice* ($n = 14$), whereas those with "Good/Very good/Excellent" were *experienced* ($n = 29$). Because all usability ratings were four-point Likert items, we compared independent groups using the two-tailed Mann–Whitney U test—the nonparametric analogue of the independent-samples t-test. The U statistic is the sum of rank scores; smaller values indicate greater separation between groups.

Independent-samples Mann–Whitney U tests revealed no significant group differences on any of the five usability items (see Table IV). For example, platform ease of use was rated similarly by novices ($M = 3.64$, SD $= 0.63$) and experienced users ($M = 3.52$, SD $= 0.57$), $U = 231.5$, $p = 0.40$. Thus, the small numeric gaps observed are compatible with sampling variability in this pilot study.

TABLE IV
NOVICE VERSUS EXPERIENCED USERS: MANN–WHITNEY U TESTS

Item	M_{nov}	M_{exp}	U	p
Platform ease of use	3.64	3.52	231.5	0.40
Meanings conveyed	3.50	3.21	239.5	0.30
Teacher understandable	3.57	3.48	213.0	0.78
Color feedback useful	3.57	3.31	256.5	0.12
Hand-tracking quality	3.50	3.31	239.0	0.30

Qualitatively, three of the four free-text comments that proposed concrete refinements (e.g., keyframe placement, mirrored playback, and tracking sensitivity) came from experienced signers, with one originating from a novice participant, mirroring the coding patterns summarized in the qualitative analysis subsection. Both groups rated the color-based feedback visualization similarly high (Novice $M = 3.57$, Experienced $M = 3.31$), indicating that the visual feedback was effective for both novices and experienced users.

Effect sizes computed from the associated Mann–Whitney Z statistics remained small (largest $|r| \approx 0.24$), indicating negligible-to-small practical differences.

Ethics statement: This study was reviewed and approved by the Research Ethics Committee of the University of Patras (application `6850/18.10.2023`; https://ehde.upatras.gr/). Written informed consent was obtained from all adult participants ($\geq$18 years). For participants aged 15–17 years, written parental or legal-guardian permission and the participant's written assent were obtained prior to enrollment. All procedures conformed to the Declaration of Helsinki and relevant national regulations, and all data were anonymized in accordance with GDPR 2016/679 before analysis.

V. DISCUSSION AND CONCLUSION

In this article, we proposed a sign language teaching system using a conventional web camera, providing an e-learning environment where students can work independently of tutors at their own pace. The student's real-time motion can be captured, while a video recording their motion can be replayed. Keyframes are extracted, and the system can provide immediate feedback to the student after analysis of the captured motion. The adherence to the motion in each keyframe is observed and graded to improve by focusing on particular parts of the motion during signing. A prototype of our proposed system is implemented and demonstrated as a showcase.

The pilot evaluation with 43 participants yielded several concrete findings. First, perceived usability was consistently high: 95.3% of participants (41/43) rated the user interface as at least satisfactory, with similarly positive ratings for feedback quality across all features (88%–95% rated as satisfactory or better; Table II). Second, 88.4% of participants (38/43) expressed strong intention ("Quite likely" or "Sure") to use the platform for future learning. Third, independent-samples Mann–Whitney U tests revealed no significant differences between novice and experienced signers on any of the five usability items (see Table IV), with all p-values exceeding 0.12. Finally, thematic analysis of free-text comments identified three salient themes:

requests for clearer interface guidance, concerns about tracking robustness under certain conditions, and overall positive sentiment about the platform's ease of use and helpfulness.

Taken together, these uniformly high ratings and strong stated intention to continue using the platform suggest that explainable camera-only feedback can motivate practice even without specialized sensors. This pilot study shows that a webcam-based language-agnostic platform can deliver *fine-grained explainable* feedback on both static handshapes and dynamic trajectories. These results extend earlier camera-only approaches [5], [11] by demonstrating that visualized deviation cues can be perceived as helpful rather than intrusive, echoing Paudyal et al.'s findings on explainable AI feedback [25]. Whereas prior systems such as those by Rho et al. [4] and Zhang et al. [10] provide only binary correct/incorrect feedback, our keyframe-level color-coding allows learners to see *where* their handshape deviates, addressing a gap identified in multiple reviews of sign language learning technologies [22], [28].

The absence of significant differences between novice and experienced signers is noteworthy. One might expect experts to demand more precise feedback or to be more critical of automated systems; instead, both groups found the visual cues equally useful. This suggests that the keyframe-and-trajectory approach provides value across the learning continuum—beginners benefit from immediate correction of gross errors, while experienced signers appreciate fine-grained feedback on subtle deviations. The uniformly high acceptance further indicates that explainable feedback may address the demotivation problem noted in earlier e-learning platforms [3].

We position the proposed system as a complementary tool rather than a stand-alone instructor: its value lies in offering immediate visualized guidance that students can use between human-led sessions. This aligns with Hassan et al.'s findings [28] on learner preferences for systems that supplement rather than replace human instruction. Unlike sensor-based approaches [4], [9] that require specialized hardware, or avatar-based systems [13], [14] with complex synchronization challenges, our webcam-only design lowers the barrier to entry while still providing detailed actionable feedback. The present evidence—high usability, perceived clarity of color/skeleton cues, and positive sentiment in free-text themes—indicates that the approach is acceptable and encouraging for both novice users and sign language experts.

Limitations:

1) *Sample and design:* The study used a convenience sample and a single 20-min practice session; learning gains were not measured. Results, therefore, gauge *perceived* usefulness, not learning outcomes.
2) *Demographic scope:* Sex/gender data were not collected, preventing intersectional analyses of usability perceptions and fairness checks.
3) *Self-selection bias:* More than half the participants reported strong preexisting motivation; less engaged learners may react differently.
4) *Sensor modality:* A lone RGB webcam cannot capture nonmanual facial cues or depth cues, limiting feedback granularity.
5) *Algorithm robustness:* Hand-tracking occasionally failed under low-light conditions, as noted in free-text comments; our current pipeline does not adapt frame-rate or brightness automatically.

Despite these constraints, the platform could serve as: 1) a low-cost homework tool in introductory sign language courses; 2) a formative-assessment plug-in for massive open online courses (MOOCs); or 3) a rapid-prototyping framework for gesture-based human–computer ineteracion in other domains (e.g., dance and physiotherapy rehab).

Future work:

1) *Controlled efficacy trial:* Run randomized study with pre/post-comprehension tests to quantify learning gains.
2) *Multilanguage dataset:* Expand to more sign languages to validate the language-agnostic claim and retrain the scoring weights on diverse hand morphologies.
3) *Nonmanual features:* Integrate face-mesh and upper-body pose (MediaPipe Holistic) plus prosodic timing to capture grammatical facial expressions and posture shifts.
4) *Adaptive feedback logic:* Use reinforcement learning to modulate feedback frequency and modality based on learner error history.
5) *Robustness and fairness audit:* Evaluate performance across camera qualities, skin tones, and lighting conditions to detect systematic bias.

The present work demonstrates that detailed algorithm-driven feedback can be provided with commodity hardware while remaining acceptable to both novice and expert signers.

References

[1] Deafness and Hearing Loss, World Health Organization, Geneva, Switzerland, 2024.

[2] W. Vicars, "ASL online: The design and implementation of a web-based American sign language course," Ph.D. dissertation, Lamar Univ., Beaumont, TX, USA, Aug. 2003.

[3] S. P. Alexander, M. Vale, and R. McKee, "Learn NZSL: An E-learning resource for New Zealand sign language," vol. 23, no. 2, pp. 60–79, 2017.

[4] E. Rho, K. Chan, E. Varoy, and N. Giacaman, "An experiential learning approach to learning manual communication through a virtual reality environment," *IEEE Trans. Learn. Technol.,*vol. 13, no. 3, pp. 554–566, Jul.–Sep. 2020.

[5] N.-N. Truong, T.-D. Do, T.-S. Le Nguyen, M.-T. Duong, T.-H. Nguyen, and M.-H. Le, "A vision-based hand-sign language teaching system using deep neural network: Methodology and experiments," in *Proc. Int. Workshop Intell. Syst.*, 2022, pp. 1–6.

[6] W. Aljedaani, R. Krasniqi, S. Aljedaani, M. W. Mkaouer, S. Ludi, and K. Al-Raddah, "If online learning works for you, what about deaf students? Emerging challenges of online learning for deaf and hearing-impaired students during COVID-19: A literature review," *Univ. Access Inf. Soc.*, vol. 22, no. 3, pp. 1027–1046, Aug. 2023.

[7] M. Glaser and E. Van Pletzen, "Inclusive education for deaf students: Literacy practices and South African sign language," *Southern Afr. Linguistics Appl. Lang. Stud.*, vol. 30, no. 1, pp. 25–37, Jul. 2012.

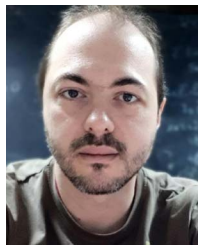
[8] G. Tang, "Sign language and inclusive deaf education: An Asian perspective," *Deafness Educ. Int.*, vol. 26, no. 1, pp. 1–5, Feb. 2024.

[9] H. Sagawa and M. Takeuchi, "A teaching system of Japanese sign language using sign language recognition and generation," in *Proc. 10th ACM Int. Conf. Multimedia*, 2002, pp. 137–145.

[10] Y. Zhang, Y. Min, and X. Chen, "Teaching Chinese sign language with a smartphone," *Virtual Reality Intell. Hardware*, vol. 3, no. 3, pp. 248–260, Jun. 2021.

[11] Q. Shao et al., "Teaching American sign language in mixed reality," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 4, pp. 1–27, Dec. 2020.

[12] M. S. Alam et al., "ASL champ!: A virtual reality game with deep-learning driven sign recognition," *Comput. Educ.: X Reality*, vol. 4, 2024, Art. no. 100059.
[13] H. D. Phan, K. Ellis, and A. Dorin, "MIC, an interactive sign language teaching system," in *Proc. 30th Aust. Conf. Comput.-Hum. Interact.*, Dec. 2018, pp. 544–547.
[14] H. D. Phan, K. Ellis, A. Dorin, and P. Olivier, "Feedback strategies for embodied agents to enhance sign language vocabulary learning," in *Proc. 20th ACM Int. Conf. Intell. Virtual Agents*, Oct. 2020, pp. 1–8.
[15] J. Wang, I. Ivrissimtzis, Z. Li, Y. Zhou, and L. Shi, "Exploring the potential of immersive virtual environments for learning American sign language," in *Responsive and Sustainable Educational Futures*, O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, and T. Papathoma, Eds., Cham, Switzerland: Springer, Aug. 2023, pp. 459–474.
[16] L. Quandt, "Teaching ASL signs using signing avatars and immersive learning in virtual reality," in *Proc. 22nd Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2020, pp. 1–4.
[17] L. C. Quandt, A. Willis, M. Schwenk, K. Weeks, and R. Ferster, "Attitudes toward signing avatars vary depending on hearing status, age of signed language acquisition, and avatar type," *Front. Psychol.*, vol. 13, Feb. 2022, Art. no. 730917.
[18] S. V. Vlasios Kasapakis, E. Dzardanova, and A. Agelada, "Sign language in immersive virtual reality: Design, development, and evaluation of a virtual reality learning environment prototype," *Interact. Learn. Environ.*, vol. 32, no. 10, pp. 6657–6671, 2023.
[19] D. Economou, M. G. Russi, I. Doumanis, M. Mentzelopoulos, V. Bouki, and J. Ferguson, "Using serious games for learning British sign language combining video, enhanced interactivity, and VR technology," *J. Univ. Comput. Sci.*, vol. 26, no. 8, pp. 996–1016, 2020.
[20] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intell. Syst. Appl.*, vol. 12, Nov. 2021, Art. no. 200056.
[21] F. M. Najib, "Sign language interpretation using machine learning and artificial intelligence," *Neural Comput. Appl.*, vol. 37, no. 2, pp. 841–857, Jan. 2025.
[22] M. Alaghband, H. R. Maghroor, and I. Garibay, "A survey on sign language literature," *Mach. Learn. Appl.*, vol. 14, Dec. 2023, Art. no. 100504.
[23] G. MacMaster, "Sign language translation using machine learning and computer vision," Univ. Vermont Honors College Senior Thesis, Burlington, VT, USA, 2020.
[24] D. Kelly, J. McDonald, and C. Markham, "A system for teaching sign language using live gesture feedback," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2008, pp. 1–2.
[25] P. Paudyal, J. Lee, A. Kamzin, M. Soudki, A. Banerjee, and S. Gupta, "Learn2Sign: Explainable AI for sign language learning," in *Proc. IUI Workshops*, 2019.
[26] P. Paudyal, A. Banerjee, and S. Gupta, "On evaluating the effects of feedback for sign language learning using explainable AI," in *Companion Proc. 25th Int. Conf. Intell. User Interfaces*, 2020, pp. 83–84.
[27] L. Z. Suri and K. F. McCoy, "A methodology for developing an error taxonomy for a computer assisted language learning tool for second language learners," Univ. Delaware, Newark, DE, USA, Tech. Rep. TR-93-16, Jan. 1993.
[28] S. Hassan, S. Lee, D. Metaxas, C. Neidle, and M. Huenerfauth, "Understanding ASL learners' preferences for a sign language recording and automatic feedback system to support self-study," in *Proc. 24th Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2022, pp. 1–5.
[29] B. Natarajan et al., "Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation," *IEEE Access*, vol. 10, pp. 104358–104374, 2022.
[30] S. Tan, N. Khan, Z. An, Y. Ando, R. Kawakami, and K. Nakadai, "A review of deep learning-based approaches to sign language processing," *Adv. Robot.*, vol. 38, no. 23, pp. 1649–1667, Dec. 2024.
[31] M. Kaczmarek and M. Filhol, "Computer-assisted sign language translation: A study of translators' practice to specify CAT software," *Mach. Transl.*, vol. 35, no. 3, pp. 287–309, Aug. 2021.
[32] A. Liu, "A new classroom teaching language in schools for the deaf—large screen language," *Chinese J. Special Educ.*, no. 5, pp. 31–35, 2005.
[33] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.
[34] E. Sartinas, E. Psarakis, K. Antzakas, and D. Kosmopoulos, "A 2-D wrist motion based sign language video summarization," in *Proc. Brit. Mach. Vis. Conf.*, 2021, Art. no. 227.
[35] C. Goodall, "Procrustes methods in the statistical analysis of shape," *J. Roy. Statist. Soc. B (Methodol.)*, vol. 53, no. 2, pp. 285–339, 1991.
[36] N. Arvanitis, E. Sartinas, and D. Kosmopoulos, "Procrustes-DTW: Dynamic time warping variant for the recognition of sign language utterances," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops*, 2023, pp. 1–5.
[37] D. Brentari and P. Eccarius, "Handshape contrasts in sign language phonology," *Sign Languages*, D. Brentari, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2010, pp. 284–311.
[38] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jul. 2008.



Kostas Blekos (Member, IEEE) received the B.Sc. and Ph.D. degrees in theoretical and computational physics from the University of Patras, Patras, Greece.

He has held research positions with the Industrial Systems Institute (Athena Research Center), the Computation-based Science and Technology Research Center, The Cyprus Institute, and the University of Patras. His research interests include quantum computing, hybrid quantum–classical algorithms, machine learning, and computer vision.

Dimitrios Kosmopoulos (Member, IEEE) received the B.Sc. and Ph.D. degrees in electrical and computer engineering from the National Technical University of Athens, Athens, Greece.

He was with the Technological Educational Institute of Crete, Heraklion, Greece; Rutgers University, New Brunswick, NJ, USA; the University of Texas at Arlington, Arlington, TX, USA; and NCSR Demokritos, Agia Paraskevi, Greece. He is currently a Professor with the Department of Computer Engineering and Informatics, University of Patras, Patras, Greece. He has authored or coauthored more than 120 papers in leading journals and conferences in the fields of computer vision and machine learning. He has 20 years of teaching experience. He has served as principal investigator and coordinator in several EU-funded and national projects.

Klimis Antzakas received the Ph.D. degree in language and communication science from City University London, London, U.K., in 2008.

He is currently an Assistant Professor of Special Education with emphasis on Deaf Education with the Department of Educational Sciences and Social Work, University of Patras, Patras, Greece, where he heads the Unit of Deaf Education. His research interests include sign language linguistics, deaf education, bilingual education and literacy of deaf and hard-of-hearing students, and accessibility technologies for deaf individuals.

The Open Access publication of this article was financially supported by HEAL-Link.