Pattern Analysis and Applications
https://doi.org/10.1007/s10044-021-01027-2

ORIGINAL ARTICLE

Assessing task understanding in remote ultrasound diagnosis via gesture analysis

Edgar Rojas-Muñoz[1] · **Juan P. Wachs**[2]

Received: 3 August 2020 / Accepted: 11 August 2021

Abstract
This work presents a gesture-based approach to estimate task understanding and performance during remote ultrasound tasks. Our approach is comprised of two main components. The first component uses the Multi-Agent Gestural Instruction Comparer (MAGIC) framework to represent and compare the gestures performed by collaborators. Through MAGIC, gestures can be compared based in their morphology, semantics, and pragmatics. The second component computes the Physical Instructions Assimilation (PIA) metric, a score representing how well are gestures being used to communicate and execute physical instructions. To evaluate our hypothesis, 20 participants performed a remote ultrasound task consisting of three subtasks: vessel detection, blood extraction, and foreign body detection. MAGIC's gesture comparison approaches were compared against two other approaches based on how well they replicated human-annotated gestures matchings. Our approach outperformed the others, agreeing with the human baseline over 76% of the times. Subsequently, a correlation analysis was performed to compare PIA's task understanding insights with those of three other metrics: error rate, idle time rate, and task completion percentage. Significant correlations ($p \leq 0.04$) were found between PIA and all the other metrics, positioning PIA as an effective metric for task understanding estimation. Finally, post-experiment questionnaires were used to subjectively evaluate the participants' perceived understanding. The PIA score was found to be significantly correlated with the participants' overall task understanding ($p \leq 0.05$), hinting to the relation between the assimilation of physical instructions and self-perceived understanding. These results demonstrate that gestures an be used to estimate task understanding in remote ultrasound tasks, which can improve how these tasks are performed and assessed.

Keywords Gestures · Task understanding · Human collaboration · Ultrasound training

1 Introduction

The correct use of ultrasound technologies is crucial in the diagnosis of various pathologies. As these technologies became more portable and prevalent, their applications have been extended to areas such as triage and Point-of-Injury care [43]. For instance, portable ultrasound devices have been critical to provide medical assistance in the battlefield [44], disaster triage for mass casualty events [60], and drug delivery procedures [41].

Nonetheless, there is a bottleneck in the way ultrasound tasks are assessed. These assessments are often performed by experienced medics via clinical examinations, or via self-assessment questionnaires [55, 59]. Such approaches, however, are subjective, are limited by the experts' availability, and can become challenging to perform remotely [39]. To develop more reliable assessment mechanisms, research has been focused on generating objective metrics via neural networks and cognitive load measurements [2, 36]. In this context, gestures can be an important avenue to explore. Gestures have been found to be key components of how humans interact, collaborate, and communicate [23, 46]. Particularly for collaborative tasks, the information encompassed in gestures can reveal important insights of how the tasks are being performed and understood [30].

This work explores whether gestures can be used as estimators of task understanding and performance during remote ultrasound tasks. We present a methodology comprised by

✉ Juan P. Wachs
jpwachs@purdue.edu

[1] Department of Visualization, Texas A&M University, 3137 TAMU, College Station, TX 77840, USA

[2] School of Industrial Engineering, Purdue University, 315 N. Grant Street, West Lafayette, IN 47907, USA

Published online: 13 September 2021

two main components. The first component uses the Multi-Agent Gestural Instruction Comparer (MAGIC) framework to represent and compare the gestures performed by the collaborators [48]. Through MAGIC, gestures can be compared based on their morphology, semantics, and pragmatics. The second component computes the Physical Instructions Assimilation (PIA) metric, a score representing how well are gestures being used to communicate and execute physical instructions [49]. By using these two components, an objective estimation of task understanding during remote ultrasound tasks can be obtained.

The contributions of this work include (1) developing a gesture-based approach to estimate task understanding; (2) comparing our MAGIC approach against other gesture assessment methods in the context of remote tasks; (3) comparing our gesture-based approach to objective and subjective estimators of task understanding; and (4) evaluating our gesture-based task understanding estimation approach in a remote ultrasound task.

The paper proceeds as follows: Sect. 2 presents prior work related to ultrasound tasks and the importance of gestures in collaboration. Section 3 discusses the MAGIC framework and the PIA metric. Section 4 describes the remote ultrasound task, details our data collection, and explains the analyses used to evaluate our approach. Finally, Sect. 5 presents and discusses our results, and Sect. 6 concludes the paper.

2 Previous work

Remote ultrasound technologies have been widely adopted through various medical disciplines such as cardiac and pulmonary Point-of-Injury care [17], Apnea and Pneumothorax treatment [39], and Focused Assessed Transthoracic Echo of the heart and pleural space [57]. This widespread adoption has to do, in part, with the development of portable, prevalent, and easy-to-use ultrasound platforms [43]. Studies have shown that novices carried out ultrasound examinations successfully when guided by a remote expert [17, 57]. This positions ultrasound technologies as a crucial component in the diagnosis of various pathologies.

Clinical examinations from expert medics are commonly used to assess performance of ultrasound tasks. The experts evaluate performance based on their own criteria or predefined checklists [1, 35, 55]. Self-assessment questionnaires are also used to evaluate ultrasound skills of novice medics [59]. While widely spread, both approaches are prone to subjective biases and are limited by the experts' availability. Therefore, research has been focused on creating objective assessments of ultrasound tasks. These include kinematic representations of the participants' performances, obtained via neural networks [36], and cognitive load measures such as eye-based physiological indices [2].

This work explores whether gestures can also be used to assess ultrasound tasks objectively. Gestures are key components of how humans interact, collaborate, and communicate. Gestures can help generating novel problem-solving strategies [7], facilitate learning [12], and offload working memory [50]. Moreover, people tend to gesture as they perform a shared physical task [21]. In such cases, gestures are not only used to solve the task (e.g., assembling parts), but as means of inquire or instruction (e.g., requesting clarification) [23]. Additionally, research has shown that gestures can facilitate the exchange of ideas and intentions during teamwork, leading to better task performance [3]. Gestures are so relevant in collaboration that systems to convey gestures remotely have been explored to support better distributed collaborative environments [29–31].

Despite studies on gestures' importance for task performance and understanding, gestures have not been explored as means to quantitatively assess task understanding. Instead, task understanding is usually assessed via objective metrics such as task completion percentage, task completion time, and number of errors [6, 22, 28, 32], or subjective techniques such questionnaires and interviews [58]. While these metrics can be effective to estimate understanding, they do not consider the information encompassed in the collaborators' gestures. Since gestures are so relevant for collaboration [23, 34, 46], ignoring them can lead to incomplete estimations of understanding.

3 Methodology

This work presents our methodology to use gestures as estimators of performance and understanding during remote ultrasound tasks. Our methodology is divided into two main subsections. The first subsection presents an explanation of the MAGIC framework. MAGIC abstracts gestures into data structures representing the gesture's morphology, semantics, and pragmatics. Such data structures are later leveraged to compare the gestures. The second subsection introduces the PIA metric, a score representing how well are instructions being received and executed by collaborators. The creation of this framework relates back to our first contribution: developing a gesture-based approach to estimate task understanding.

3.1 Multi-agent gestural instruction comparer

The MAGIC framework represents and obtains similarity metrics between gestures [47, 48]. The framework's goal is to abstract and compare gestures using a representation that considers aspects such as the gestures' shape, movement, meaning, and context. To do this, MAGIC starts by

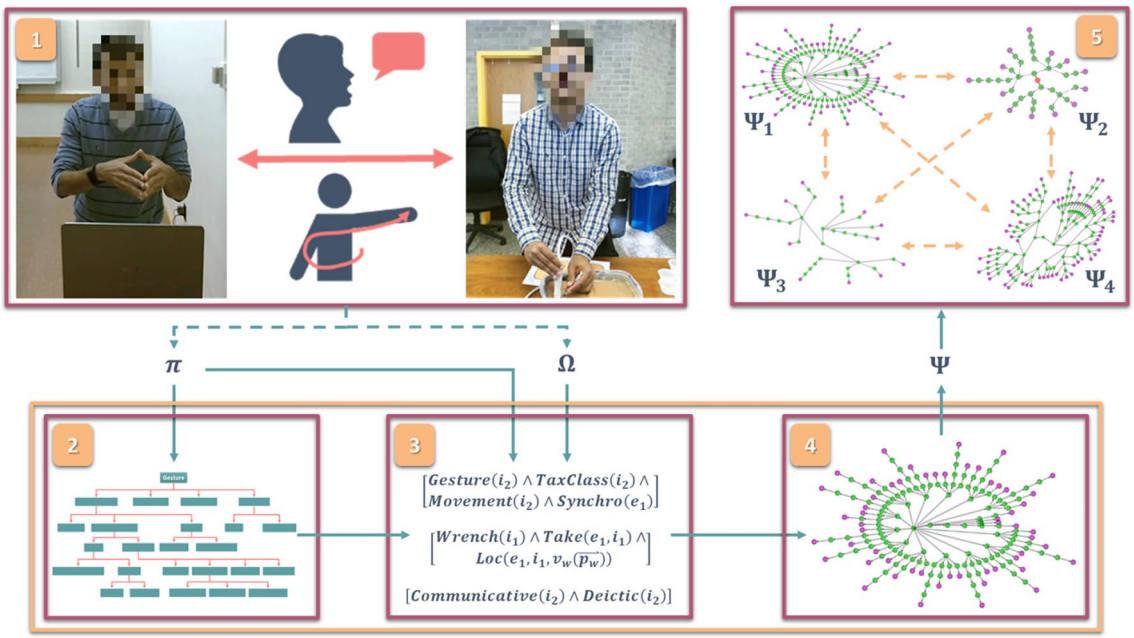

Fig. 1 Schematic of the MAGIC framework. (1) Using speech and gestures, two individuals collaborate to complete a shared ultrasound task. MAGIC gives a formal definition to the elements involved in the collaborative task (e.g., Utterance $\boldsymbol{\pi}$, Context $\boldsymbol{\Omega}$). (2) A taxonomy classification describes aspects of the gesture utterances such as iconicity and expressiveness. (3) A dynamic semantics framework represents each gesture utterance with a logical form containing the gestures' morphology, semantics, and pragmatics. (4) A constituency parsing represents each logical form as a tree data structure (Interpretation Tree $\boldsymbol{\Psi}$). (5) Gestures are matched based on how similar they are. This similarity can be computed based on the time the gestures were performed, or by comparing the Interpretation Trees $\boldsymbol{\Psi}$ representing the gestures

Table 1 Elements and notation of the MAGIC framework

MAGIC element	Definition	Example
Worker $\boldsymbol{\Phi_W}$	Collaborator directly manipulating the environment to perform a task	A person executing actions
Helper $\boldsymbol{\Phi_H}$	Collaborator communicating the commands required to perform a task	A person instructing what actions to perform
Instructions $\boldsymbol{A_H}$	Action communicating how to perform a task	A verbal command saying "Take the probe," accompanied by a gesture pointing at a probe
Executions $\boldsymbol{A_W}$	Action performed to complete a task	A gesture performed to reach for the probe
Utterance $\boldsymbol{\pi}$	The smallest unit of speech or gesture that communicates a complete idea	A gesture to pinpoint a probe; a verbal command saying "Stop!"
Interpretation Tree $\boldsymbol{\Psi}$	MAGIC's data structure representing Utterances $\boldsymbol{\pi}$	See Fig. 1, Sect. 4
Context $\boldsymbol{\Omega}$	Elements and concepts introduced in the current Utterance $\boldsymbol{\pi_i}$, which can be referenced by future Utterances $\boldsymbol{\pi_{i+j}}$	Let "Take the probe" and "Use the probe to find the vessel" be the first and second Utterances ($\boldsymbol{\pi_1}$ and $\boldsymbol{\pi_2}$). The Context $\boldsymbol{\Omega}$ of $\boldsymbol{\pi_2}$ would include all the elements introduced in $\boldsymbol{\pi_1}$, such as the "Probe" concept.

formally defining the elements involved in the collaborative task, summarized in Table 1. Figure 1 presents a schematic of the MAGIC framework.

MAGIC uses a three-stage process to obtain the Interpretation Trees $\boldsymbol{\Psi}$ from the Utterances $\boldsymbol{\pi}$ and the Context $\boldsymbol{\Omega}$. The first stage uses a gestural taxonomy to generate labels describing aspects of the gestures such as iconicity and expressiveness (Fig. 1, section 2). These labels are used as extra information for subsequent stages. The second section uses a dynamic semantics framework to represent each Utterance $\boldsymbol{\pi}$ with a logical form. These logical forms provide an abstraction of the gestures' morphology, semantics, and pragmatics in a single structure (Fig. 1, section 3). Afterward, the third section leverages a

constituency parsing approach to convert the logical forms into tree data structures (the Interpretation Trees $\boldsymbol{\Psi}$; Fig. 1, section 4). These Interpretation Trees $\boldsymbol{\Psi}$ are constructed to quantitatively compare the logical forms from the previous stage.

This three-stage process will be applied to all the gestures, generating an Interpretation Tree $\boldsymbol{\Psi}$ for each of them. These Interpretation Trees $\boldsymbol{\Psi}$ group similar aspects of the gestures together. For instance, all the aspects related to the gesture's shape are going to be contained in the Shape Subtree, while aspects related to the gesture's context are contained in the Context Subtree. This allows to inspect and compare specific aspects of the gestures in isolation. A description of the subtrees that can be obtained is found in [47].

3.2 Gesture matching through integer optimization

The final stage of the MAGIC framework involves comparing the gestures by calculating gesture similarity (Fig. 1, section 5). These gesture matchings reveal which gestures were similar between them and will be used in the subsequent PIA calculation.

These gesture matching processes are performed by solving three integer optimization formulations: MAGIC-based, Time-based, and Hybrid [48]. First, let $\boldsymbol{w_i}$ be a Worker-authored gesture, and $\boldsymbol{h_j}$ be a Helper-authored gesture. Then, let $\boldsymbol{W}$ be a set containing all the Worker-authored gestures $\boldsymbol{w_i}$, and $\boldsymbol{H}$ be a set containing all the Helper-authored gestures $\boldsymbol{h_j}$. Subsequently, let a gesture matching solution be represented with a matrix $\boldsymbol{E}$ of edge weights $\boldsymbol{e_{ij}}$ (size $|W| \times |H|$). Each edge weight $\boldsymbol{e_{ij}}$ will take a value of 1 if the Helper-authored gesture $\boldsymbol{h_j}$ matches (i.e., is the most functionally equivalent) to the Worker-authored $\boldsymbol{w_i}$, and 0 otherwise. Therefore, the goal of the gesture matching stage is to find the gesture matching matrices $\boldsymbol{E}$ that maximize (or minimize, for the Time-based) the cost function of the optimization problems.

To calculate the gesture matching matrices $\boldsymbol{E}$, three cost coefficient matrices (size $|W| \times |H|$) need to be obtained: a matrix $\boldsymbol{B}$ of similarity scores $\boldsymbol{b_{ij}}$ for the MAGIC-based approach, a matrix $\boldsymbol{C}$ of temporal scores $\boldsymbol{c_{ij}}$ for the Time-based approach, and a matrix $\boldsymbol{D}$ of hybrid scores $\boldsymbol{d_{ij}}$ for the hybrid approach. The calculation of these matrices is explained in the next subsections.

Once the cost coefficient matrices are obtained, the gesture matching matrices $\boldsymbol{E}$ can be found by solving Eq. 1 (for the Time-based approach, it changes to a minimization problem). The scores $\boldsymbol{a_{ij}}$ represent either the similarity scores $\boldsymbol{b_{ij}}$, the temporal scores $\boldsymbol{c_{ij}}$, or the hybrid scores $\boldsymbol{d_{ij}}$, depending on the selected approach. This formulation is constrained to each Worker-authored gesture $\boldsymbol{w_i}$ being matched to only one Helper-authored gesture $\boldsymbol{h_j}$.

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{|H|} \sum_{i=1}^{|W|} a_{ij} e_{ij} \\ \text{subject to} \quad & \sum_{j=1}^{|H|} e_{ij} = 1, \forall i \\ & e_{ij} \in \{0,1\} \\ & i = 1,2,\ldots,|W|; j = 1,2,\ldots,|H| \end{aligned} \tag{1}$$

3.2.1 MAGIC-based optimization

The MAGIC-based approach computes similarity from the Interpretation Trees $\boldsymbol{\Psi}$. The goal is to generate matrices $\boldsymbol{B}$ of similarity scores $\boldsymbol{b_{ij}}$ describing how similar is each Worker Interpretation Tree $\boldsymbol{\Psi_W}$ to each Helper Interpretation Tree $\boldsymbol{\Psi_H}$. Each similarity score $\boldsymbol{b_{ij}}$ is computed as the number of nodes in the intersection between two $\boldsymbol{\Psi}$ Interpretation Trees or subtrees representing different gestures, as described in Eq. 2:

$$b_{ij} = \left(\text{num\_nodes}\, \Psi_{h_j} \cap \Psi_{w_i}\right); \begin{matrix} i = 1,2,\ldots,|\boldsymbol{W}| \\ j = 1,2,\ldots,|\boldsymbol{H}| \end{matrix} \tag{2}$$

3.2.2 Time-based optimization

The Time-based approach explores the relevance of time when comparing gestures: Gestures performed closer in time are likely to be related. The time in which gestures were performed is stored in two vectors: $\mathbf{t_W}$ and $\mathbf{t_H}$, respectively, for Worker $\boldsymbol{\Phi_W}$ and the Helper $\boldsymbol{\Phi_H}$. The vectors are then expanded into matrix form by multiplying them by vectors of ones (size $|\boldsymbol{H}| \times \mathbf{1}$ and $|\boldsymbol{W}| \times \mathbf{1}$, respectively). Finally, matrices $\boldsymbol{C}$ of temporal scores $\boldsymbol{c_{ij}}$ are computed with Eq. 3:

$$C = \mathbf{t}_W \mathbf{1}^T - \mathbf{1}\, \mathbf{t}_H^T \tag{3}$$

3.2.3 Hybrid optimization

The Hybrid optimization approach combines the previous approaches to consider both gesture similarity and temporal synchrony. The approach computes matrices $\boldsymbol{D}$ of hybrid scores $\boldsymbol{d_{ij}}$ from the previous $\boldsymbol{B}$ and $\boldsymbol{C}$ matrices. These hybrid scores $\boldsymbol{d_{ij}}$ are calculated by regulating the effect of the similarity scores $\boldsymbol{b_{ij}}$ and temporal scores $\boldsymbol{c_{ij}}$, as depicted in Eq. 4. The constants α, β, and γ, respectively, control importance of the temporal scores $\boldsymbol{c_{ij}}$, when does the function activate, and importance of the similarity scores $\boldsymbol{b_{ij}}$.

$$d_{ij}=\left(-e^{-\alpha c_{ij}}\frac{-e^{-\alpha c_{ij}}-\beta}{|-e^{-\alpha c_{ij}}-\beta|}-e^{-\alpha c_{ij}}\right)+\gamma\frac{b_{ij}}{\sum_{j=1}^{|H|}b_{ij}} \tag{4}$$

3.3 Physical instructions assimilation metric

The PIA score estimates task understanding from how well are physical instructions being assimilated (i.e., given, understood, and executed). This perspective is currently not encompassed by other task understanding estimators. PIA draws inspiration from two theories that describe how individuals understand each other by observing each other's actions and gestures [10, 11, 34, 51, 54]. The approach assumes that correct understanding of a Helper $\boldsymbol{\Phi_H}$ gestural instruction happens whenever the Worker $\boldsymbol{\Phi_W}$ executes the instruction without errors or asking for clarifications.

PIA represents perfect understanding of a task with a score of 100, which happens when every $\boldsymbol{h_j}$ Helper-authored gesture is matched to one and only one $\boldsymbol{w_i}$ Worker-authored gesture. A PIA score less than a 100 means that either multiple Worker-authored gestures w_i matched to the same Helper-authored gesture h_j, or at least one Helper-authored gesture h_j not getting matched. As depicted in Eq. 5, the PIA metric is calculated based on the gesture matching matrix $\boldsymbol{E}$ from the previous step. More details of the PIA formulation can be found in [49].

$$PIA=\frac{1}{|H|}\sum_{j=1}^{|H|}\left(\sum_{i=1}^{|W|}e_{ij}\right)^{-1} \tag{5}$$

4 Experimental apparatus

This section describes the setup to validate our gesture-based approach to assess understanding on a remote ultrasound task. First, we illustrate our ultrasound task. Then, we describe our data collection process, which analyzes: (1) whether our gesture matching approaches could represent human-annotated gesture matchings; and (2) whether the PIA metric is a valid approach to estimate task understanding.

4.1 Remote ultrasound task

Our experimental apparatus relates back to our fourth contribution: evaluating our gesture-based task understanding estimation approach in a remote ultrasound task. Following [4], we created an ultrasound phantom with seven vessels using ballistic gel. The vessels were filled with colored water. Subsequently, the ultrasound phantom was coated with a mixture of silicone and paint that emulated skin and hid the location of the vessels from plain sight. Afterward, toothpicks were inserted into the models to simulate a wooden splinter. The dimensions of the resulting foreign body were 6cm length $\times$ 1.5 cm width $\times$ 1.5 mm height. This model, along with an ultrasound probe (Telemed MicrUs MC10-5R10S), was used by our participants to complete three common tasks in ultrasound training curricula.

The first task involved finding vessels, similar to [4]. Participants were given 10 min to find vessels in the ultrasound phantom using the ultrasound probe. The second task involved using a syringe to extract water from the vessels, as in [56]. Participants used the ultrasound probe to relocate the vessels they found in the previous task. Then, guided by the ultrasound image, participants had to insert the syringe inside the vessels to extract water from them. Participants were also given 10 min to complete this task. The third task involved identifying the position, shape, and orientation of

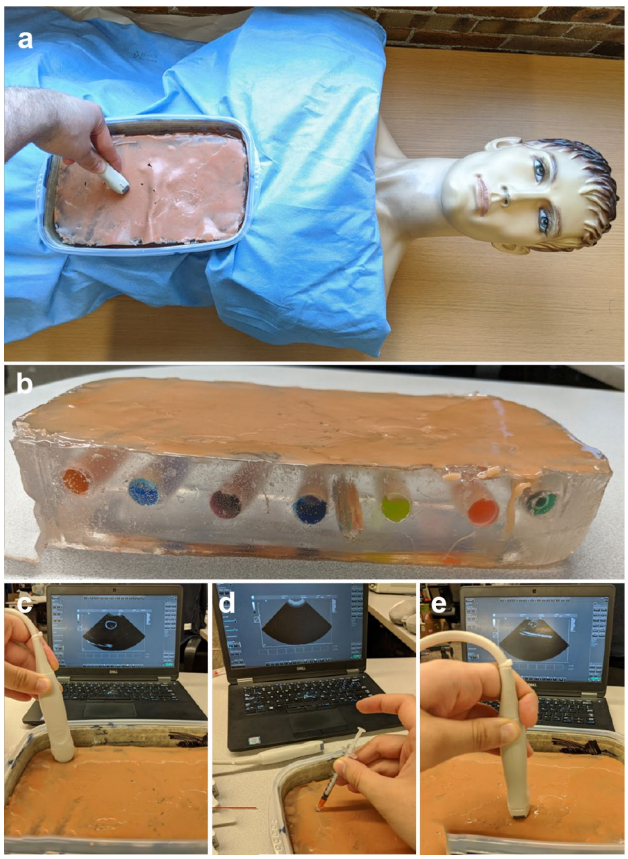

Fig. 2 Ultrasound task. Our ultrasound simulator is shown in (**a**). The created ultrasound phantom is shown in (**b**). The vessels as seen in the ultrasound are shown in (**c**). The process of extracting blood is shown in (**d**). The foreign body as seen in the ultrasound is shown in (**e**)

the foreign body inside the ultrasound phantom, similar to [52]. As in the previous tasks, the participants had to use the ultrasound probe to locate the foreign body, which looked as a solid patch in the ultrasound image. Participants had 5 min to complete this task. Our ultrasound phantom and the tasks participants had to perform are showcased in Fig. 2.

4.2 Data collection

4.2.1 Participants

Three user studies were conducted with a total of twenty participants (graduate students, twelve males, and eight females, age mean 27.2 ± 5.3 years old). The participants were divided into 10 Helper–Worker pairs to collaboratively complete the ultrasound task.

4.2.2 Procedure

First, our gesture comparison approach was evaluated. This evaluation relates back to our second contribution: comparing our MAGIC approach against other gesture assessment methods in the context of remote tasks. We compared the gesture matching matrices $\boldsymbol{E}$ generated by our approaches against two gesture matching baselines: morpho-semantic descriptors (MSD) [37], and naïve time synchronization (NTS). The first baseline, MSD, used Boolean vectors to represent morpho-semantic aspects of the gestures (e.g., Does the gesture have leftward movement? Does the gesture refer to a specific part of the body?). Each gesture was therefore represented with a $\mathbf{48 \times 1}$ vector, where each dimension in the vector represented a morpho-semantic description in [37]. The gesture matching matrices $\boldsymbol{E}$ were generated by comparing the MSD vectors via Hamming distance and cosine similarity.

The second baseline, NTS, compared gestures based on their temporal occurrence. A timestamp $\boldsymbol{t_k}$, in seconds, was assigned to each gesture, where 0 and 1 represented start and end of the video, respectively. Afterward, a time window before and after the execution of each Helper-authored gesture $\boldsymbol{h_j}$ was defined. The Worker-authored gestures $\boldsymbol{w_i}$ were matched to the Helper-authored gestures $\boldsymbol{h_j}$ based on which time window they were contained in. The margins of these time windows are calculated as $\frac{t_{k+1}-t_k}{2}$, where $\boldsymbol{t_k}$ and $\boldsymbol{t_{k+1}}$ are the timestamps assigned to two consecutive Helper-authored gestures $\boldsymbol{h_j}$. This process was performed until all Worker-authored gestures $\boldsymbol{w_i}$ were contain inside a the time window of a Helper-authored gesture $\boldsymbol{h_j}$.

After obtaining these baselines, gesture matching ground truths needed to be created. To generate such ground truth, a member of the research team manually annotated which gestures matched each other. For instance, if the Helper $\boldsymbol{\Phi_H}$ made a gesture to indicate how to properly hold the probe, all the gestures performed by the Worker $\boldsymbol{\Phi_W}$ to fulfill this instruction were matched to that Helper $\boldsymbol{\Phi_H}$ gesture. Through this process, a matrix $\hat{\boldsymbol{E}}$ of ground truth gesture matchings was created for each pair of participants.

All gesture matching matrices $\boldsymbol{E}$ were evaluated based on their percentage of agreement with the matrices $\hat{\boldsymbol{E}}$ of human-annotated gestures matchings. This is done using Eq. 6, which is a variation in the F1-Score formula [24]. In Eq. 6, TP represent True Positives (i.e., $\hat{\boldsymbol{E}}_{ij} = \boldsymbol{E}_{ij} = 1$), FN represent False Negatives (i.e., $\hat{\boldsymbol{E}}_{ij} = 1$; $\boldsymbol{E}_{ij} = 0$); and FP represent False Positives (i.e., $\hat{\boldsymbol{E}}_{ij} = 0$; $\boldsymbol{E}_{ij} = 1$).

$$\text{Matching Score} = \frac{2\text{TP}}{2\text{TP}+\text{FN}+\text{FP}} \tag{6}$$

Additionally, we inspected whether performing the comparisons against different subtrees within the Interpretation Trees $\boldsymbol{\Psi}$ affected the Matching Score. This explored whether specific aspects of the gestures (e.g., shape, meaning) were more relevant in the context of remote ultrasound tasks. Specifically, we inspected comparison against: (1) the subtree representing the context (Context Subtree); (2) the subtree representing whether the gesture exemplifies an object (Exemplify Subtree); (3) the subtree representing shape (Shape Subtree); (4) the subtree encompassing all aspects of the gesture excluding its context (Predicative Subtree); (5) the entire Interpretation Tree; and (6) the combination of subtrees representing meaning (Meaning Subtree). A detailed explanation of these subtrees can be found in [47].

Subsequently, our task understanding estimation approach was evaluated. This evaluation relates back to our third contribution: comparing our gesture-based approach to objective and subjective estimators of task understanding. Our experimental setup evaluated whether the quality of assimilation of physical instructions (in the form of our PIA metric) can estimate task understanding. This was evaluated by comparing PIA against three other metrics for task understanding: error rate, idle time rate, and task completion percentage [28, 38]. Error rate was calculated as the rate between the instructions in which the Worker $\boldsymbol{\Phi_W}$ performed errors and the total number of instructions. An error was counted, for instance, when the Worker $\boldsymbol{\Phi_W}$ inserted the syringe in the wrong way. Idle time rate is defined as the rate between the time in which the Worker $\boldsymbol{\Phi_W}$ did not perform an action related to completing the task (e.g., the time spent thinking or asking questions), and the total task completion time, in seconds. Listening to the Helper $\boldsymbol{\Phi_H}$ instructions was not considered idle time.

The third metric, task completion percentage, represented how much of the entire task did the participants complete. This metric was calculated in two ways: (1) the rate between the number of vessels found and the total number of vessels in the ultrasound phantom (Vessel Detection Completion Percentage; VDCP); and (2) the rate between the number of

vessels from which blood was successfully extracted and the total number of vessels in the ultrasound phantom (Blood Extraction Completion Percentage; BECP). These three metrics were annotated by members of the research team for each pair of participants. Finally, the Pearson product moment correlation (r) was computed to inspect the relationship between the metrics [45]. The PIA score was treated as the dependent variable, while the other metrics were treated as independent variables.

Finally, we compared PIA against human-reported understanding evaluations. Participants were given two post-experiment questionnaires to assess their performance. After compiling their answers, the Pearson product moment correlation (r) was computed to analyze the relation between the participants' answers and their PIA. Correlations were calculated separately for both the Workers $\boldsymbol{\Phi_W}$ and the Helpers $\boldsymbol{\Phi_H}$.

The first questionnaire was an Understanding Assessment Questionnaire (UAQ) consisting of eight Likert scale questions (5 = Strongly Agree to 1 = Strongly Disagree) evaluating the participants' overall understanding during the task. Two out of these eight questions varied depending on the participant's role (i.e., Helper $\boldsymbol{\Phi_H}$ or Worker $\boldsymbol{\Phi_W}$). The questions included: "I was able to understand the task" (*UAQ1*), "I was able to understand the verbal instructions given by the other person" (*UAQ2*); "I was able to understand the gestural instructions given by the other person" (*UAQ3*), "I was able to understand the actions performed by the other person" (*UAQ4*); "I was able to determine if the other person was understanding me" (*UAQ7*), and "I feel I could guide the task again with minimal to no challenge" (*UAQ8*). Questions 5 and 6 differed with respect to the participant's role. For the Helper $\boldsymbol{\Phi_H}$ role, *UAQ5* was "I was able to understand the questions that were asked to me, if any," and *UAQ6* was "I was able to understand when mistakes happened along the task, if any." For the Worker $\boldsymbol{\Phi_W}$ role, *UAQ5* was "I was able to understand which part of the ultrasound phantom was the other person referring to," and *UAQ6* was "I was able to understand how to find the vessels."

The second questionnaire was the National Aeronautics and Space Administration Task Load Index (NASA-TLX) [27]. The NASA-TLX evaluates perceived workload using six criteria: mental demand (*TLX1*), physical demand (*TLX2*), temporal demand (*TLX3*), perceived performance (*TLX4*), effort required (*TLX5*), and generated frustration (*TLX6*). Each criterion is represented by a 21-level Likert Scale question. Higher TLX scores indicate higher task load.

4.2.3 Apparatus

After signing a written consent form (IRB Protocol #1810021222), participants were randomly assigned to either the Worker $\boldsymbol{\Phi_W}$ or the Helper $\boldsymbol{\Phi_H}$ role. In this setup, the Helper $\boldsymbol{\Phi_H}$ had to remotely guide the Worker $\boldsymbol{\Phi_W}$ through the ultrasound task. The Helper $\boldsymbol{\Phi_H}$ was given an instruction booklet with the steps to perform the ultrasound task. The Worker $\boldsymbol{\Phi_W}$ was instead given the instruments to perform the task: the ultrasound probe, the ultrasound phantom, and a syringe. Afterward, they were directed to different rooms according to their role. Both rooms included color and depth cameras, and a large display connected to a computer hosting a video call with the other room. This video call allowed participants to interact via speech, hand gestures, facial expressions, etc.

Our data were comprised of color and depth video recordings, acquired as the participants performed the ultrasound task. From these recordings, we extracted the audio streams and skeletal information. The skeletal information was used to represent the shape and movement components of the gestures. The audio streams were used to provide the meaning and context components of the gestures. All the gestures performed by the participants were compiled, for a total of 1287 gestures (785 performed by the Helper $\boldsymbol{\Phi_H}$, 502 performed by the Worker $\boldsymbol{\Phi_W}$) acquired over the span of 7 h of video. These included gestures to provide instructions, to ask for clarification, to perform the task, and even those performed involuntarily. Each of these gestures was represented using an Interpretation Tree $\boldsymbol{\Psi}$ and was considered in the PIA calculation. Additionally and prior to starting the task, the Helper $\boldsymbol{\Phi_H}$ received 30 min of side-by-side training on how to use the ultrasound probe. The training showed how to perform the task and explained possible errors and their solutions.

5 Results and discussion

The optimization problems were solved using IBM's CPLX Optimizer from the NEOS Server [14, 16, 25]. Additionally, the hybrid cost coefficients were calculated by setting the constants to $\alpha = 0.01$, $\beta = 1.01$, and $\gamma = 2$. Figure 3 showcases the matching scores of the gesture matching approaches, using different subtrees to compare the gestures. Our results for the ultrasound task are congruent with those obtained in our previous work for a block assembly task and an origami task [48]: The hybrid approach led to higher matching scores.

Both gesture matching baselines (MSD and NTS) were outperformed by the MAGIC-based approach. However, our results fluctuated significantly based on which subtrees were selected to compare the gestures. This is a known limitation of the MAGIC-based approach: comparing between subtrees with unrelated information (e.g., comparing shape against meaning) has a negative impact in the matching scores. The highest matching scores were found when comparing the Context Subtree of the Worker Interpretation Trees $\boldsymbol{\Psi_W}$

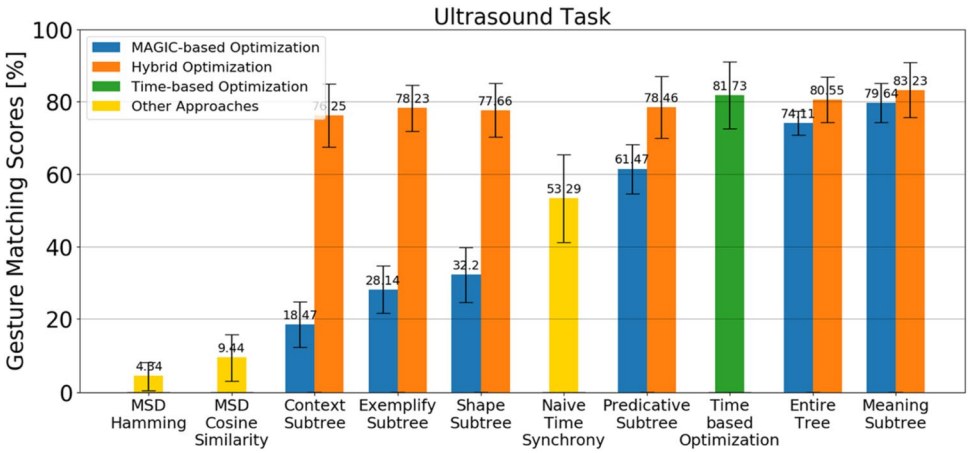

Fig. 3 Gesture matching scores for the ultrasound task. The scores represent the percentage of agreement of the gesture matching approaches with the human baseline

against the Meaning Subtree of the Helper Interpretation Trees $\boldsymbol{\Psi_H}$. This result highlights the relevance of semantics and pragmatics when comparing gestures. Contrarily, comparing only the gestures' physical appearance led to low matching scores since the gestures performed by the Helper $\boldsymbol{\Phi_H}$ were visually distinct to those performed by the Worker $\boldsymbol{\Phi_W}$.

The high scores obtained with Time-based comparison approach confirmed that gestures performed one after the other are likely to be related. However, such comparison should be performed giving higher importance to recently performed gestures, as opposed to the NTS approach where each gesture was given equal importance. Nonetheless, both these approaches are limited to the temporal relations between gestures instead of comparing the gestures comprehensively.

The Hybrid approach alleviates the limitations of the MAGIC-based: no a priori knowledge (i.e., which subtree to select) is required to compare the gestures. Matching scores over 76% for all comparisons demonstrated that the approach agreed with the human baseline without fluctuating based on the selected subtrees. Therefore, the Hybrid approach is a more stable option to compare gestures. Overall, the results demonstrate that our approaches can represent and compare gestures performed in a remote ultrasound task. This addresses our second contribution: Our MAGIC approach outperformed other gesture assessment methods for remote tasks.

The results obtained with the Time-based approach outperformed those obtained with the Hybrid approach for most comparisons. This conflicts with the findings from our prior work [48] for the block assembly and the origami tasks. In such tasks, the Hybrid approach outperformed all the other approaches. Nonetheless, lower scores were expected for the ultrasound task. The ultrasound task was considerably more complex than the other two, as more gestures had to be performed to complete each of subtask. Moreover, the meaning and context of some of these gestures were very similar because the collaborators kept referring to the same instruction in different ways (e.g., explaining how to introduce the syringe to extract water). This increased number of functionally similar gestures in reduced time spans led the Hybrid approach to underperform when compared to the Time-based one. However, the results in [48] confirm the importance of representing the gestures comprehensively instead of only relying on time.

Table 2 summarizes the results of the task understanding estimation metrics, for each Helper–Worker pair (H–W Pair). Better understanding is hinted by higher PIA scores and task completion percentages, and lower error and idle time rates. Additionally, the Pearson correlation matrix in Fig. 4 presents the relationship between the compared metrics. All metrics were significantly correlated with each other ($p \leq 0.04$).

Our results indicate that gestures can be used as estimators of understanding in ultrasound tasks. This reaffirms the idea that task understanding in shared tasks can be increased by observing someone else's actions [15]. The significant correlations between PIA and all other metrics position PIA as an effective gesture-based task understanding estimator.

Table 2 Task understanding proxy metrics on a ultrasound task

H–W pair	PIA	VDCP	BECP	Error rate	Idle time rate
1	52.86	28.57	28.57	35.14	18.31
2	65.00	85.71	85.71	27.50	10.44
3	75.24	100.00	100.00	16.42	5.13
4	62.22	100.00	100.00	30.91	5.55
5	60.71	42.86	14.29	30.91	12.36
6	55.71	42.86	0.00	37.29	8.93
7	52.24	57.14	42.86	38.30	6.30
8	79.69	100.00	100.00	10.53	2.11
9	72.53	85.71	71.43	13.73	2.66
10	47.69	14.29	14.29	41.86	13.71

Results indicate percentages over 100 [%]

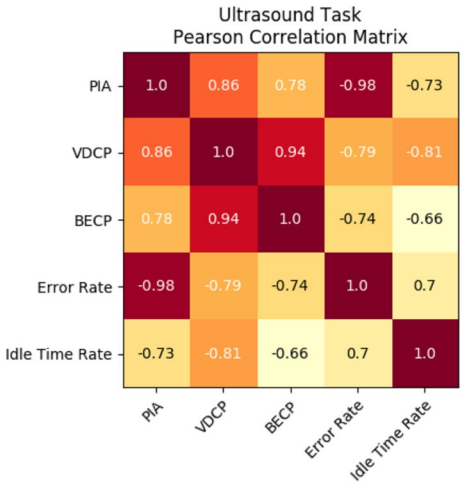

Fig. 4 Pearson correlation matrix for the task understanding metrics. All metrics were significantly correlated ($p \leq 0.04$). The cells' numbers represent the strength of the correlation

Elaborating, whenever the PIA score of a task was low, the task completion percentage also tended to be low, and error and idle time rates tended to be high. These lower PIA scores were linked to those participants that performed a large number of unnecessary gestures while completing the task. Overall, these correlations can have positive impact in how collaborative tasks are assessed. For instance, Helper–Worker pairs that perform a lot of unnecessary gestures while completing a task should be corrected, as they will likely incur in more errors, more idle time and less task completion percentages [19, 26]. These results confirm our third contribution: Our gesture-based approach successfully served as an objective estimator of task understanding.

Additionally, the PIA metric addressed certain inconsistencies in the participants' performance that the other metrics could not express. For instance, although the error rate and PIA score of the fourth Helper–Worker pair were average, they had perfect task completion percentages. A closer inspection of this pair's performance revealed a critical mistake: This pair reported finding nine vessels, even though the ultrasound phantom only had seven vessels. This mistake led this pair to perform several unnecessary gestures and errors trying to unsuccessfully relocate these extra vessels during the blood extraction task.

Additionally, PIA offers interesting insights regarding the overall quality of the understanding. For example, distinctions between "good," "decent," and "bad" assimilation of physical instructions (hence, task understanding) could be obtained by setting thresholds based on the PIA scores. Elaborating, if an empirical threshold of 70 was to be established as an indicator of "decent" task understanding, only three Helper–Worker pairs would have achieved this goal. Another advantage of the PIA metric is its generalizability: PIA is agnostic to the framework utilized to create the gesture matching matrices E required for its computation. In this work, we used the MAGIC framework to generate the matrices. Nonetheless, the scores could also be generated from the matrices E generated from other gesture matching approaches, such as MSD or NTS.

Figure 5 presents the normalized UAQ answers for the ultrasound task, according to the participants' role. Likewise, Fig. 6 presents the normalized TLX answers for the task. Finally, Fig. 7 presents the correlations between the PIA metric and all the UAQ and TLX questions, divided based on the participants' role.

The correlations between the participants' answers to the questionnaires and the PIA metric revealed interesting insights. First, significant positive correlations were found between the Workers $\boldsymbol{\Phi_W}$ answers to half of the UAQ questions and the PIA metric. These correlations represent a link between the assimilation of physical instructions (in the form of PIA) and task understanding. Specifically, Workers $\boldsymbol{\Phi_W}$ that received higher PIA scores tended to report an overall better understanding of the task. Furthermore, significant negative correlations were found between the Workers $\boldsymbol{\Phi_W}$ answers to half of the TLX criteria and the PIA metric. Specifically, the PIA scores were

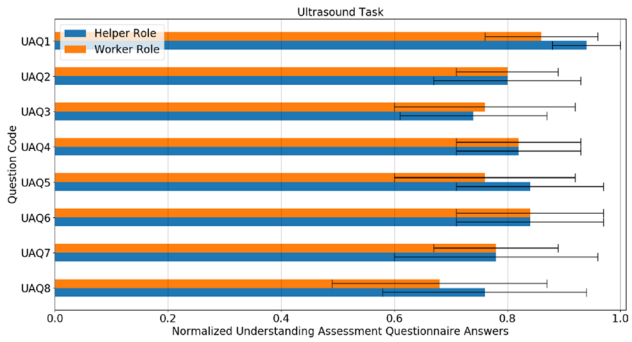

Fig. 5 Normalized UAQ answers, divided based on the participants' role (Helper $\boldsymbol{\Phi_H}$ or Worker $\boldsymbol{\Phi_W}$)

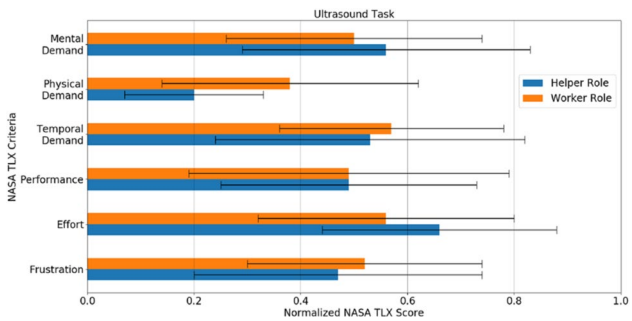

Fig. 6 Normalized NASA-TLX answers, divided based on the participants' role (Helper $\boldsymbol{\Phi_H}$ or Worker $\boldsymbol{\Phi_W}$)

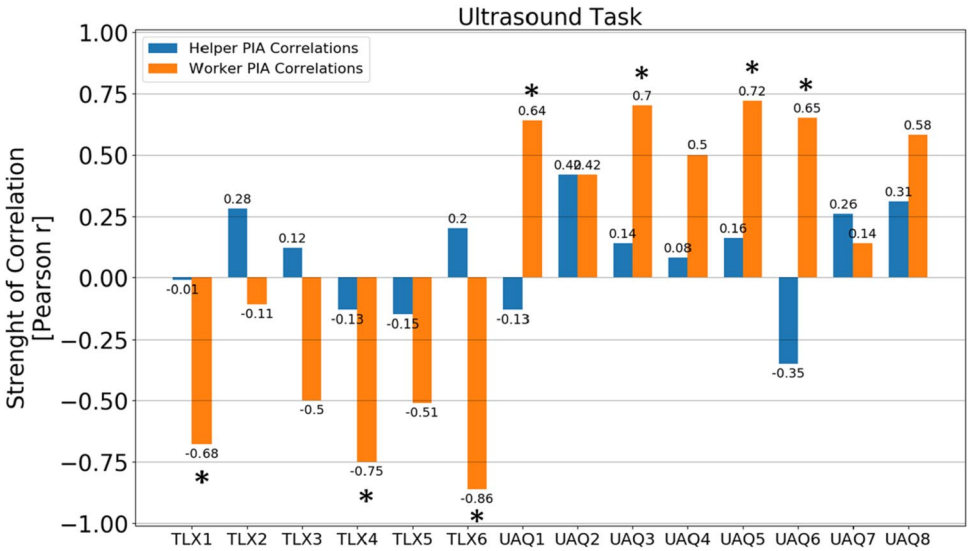

Fig. 7 Correlations between the PIA metric and all the UAQ and TLX questions, divided based on the participants' role (Helper $\boldsymbol{\Phi_H}$ or Worker $\boldsymbol{\Phi_W}$). Asterisks represent statistical significance between PIA and the respective criterion ($p \leq 0.05$)

correlated to the Workers $\boldsymbol{\Phi_W}$ *TLX1* (mental demand), *TLX4* (perceived performance), and *TLX6* (frustration) criteria. Elaborating, these correlations mean that Workers $\boldsymbol{\Phi_W}$ that received lower PIA scores tended to report higher mental demand, being unsatisfied with their performance, and feeling more frustrated. A possible explanation for these trends is how performing errors and correcting accordingly can increase a person's frustration levels [18]. Nonetheless, no correlations were found between the Helpers $\boldsymbol{\Phi_H}$ questionnaire answers and their PIA scores. Reduced levels of engagement due to the remote nature of the task could be a possible explanation for the lack of correlations [20].

Our approach presents some limitations. First, PIA does not consider verbal-only utterances in its calculation. This happens because verbal-only utterances lack the morphological aspects of gestures (e.g., shape, movement) required to generate Interpretation Trees $\boldsymbol{\Psi}$. For instance, a verbal instruction indicating to rotate the ultrasound probe will be ignored in the PIA calculation if it is not accompanied with a gesture. This implies that our approach will not act as an adequate metric to measure understanding in tasks where the Helper $\boldsymbol{\Phi_H}$ decides not to accompany the instructions with gestures, or in tasks that do not involve physical instructions. Examples of such tasks can include mathematical problem-solving tasks or memory tasks, where the performance does not necessary depends on the physical actions (although embodiment theories indicate that even in those cases, physical action leads to better task performance [5, 61]). Nonetheless, the verbal utterances are not completely ignored, as their information is included as part of the context of the Interpretation Trees $\boldsymbol{\Psi}$ when comparing between gestures.

Moreover, our manual annotation approach needs to be improved. Although we annotate the data semi-autonomously, annotating each gesture can take up to 1 min. Machine learning techniques can address this bottleneck. For example, context-aware image captioning routines could be used to annotate the gestures' pragmatics [13], and speech-to-text techniques can be used to retrieve verbal information autonomously [9].

More avenues of future work include evaluations with a larger and more diverse population, as a population comprised only of graduate students can introduce biases [8, 53]. A larger subject pool should also consider participants from diverse cultures, as gestures are known to be culture-dependent [42]. Moreover, novice medical personnel should be consulted for further specialization of the platform in the medical domain.

Overall, the assessment of remote ultrasound tasks can be improved by considering the insights provided by this work. As ultrasound devices are becoming more relevant to provide remote medical assistance [33, 39, 40], the way in which these procedures are assessed needs to improve. Our gesture-based criteria could capture task understanding aspects ignored by other metrics, which in turn could lead to more reliable assessments of understanding and performance during ultrasound tasks. This evaluation relates back to our fourth contribution: Our gesture-based successfully estimated task understanding in a remote ultrasound task.

6 Conclusion

This work presented a gesture-based approach to estimate task understanding and performance during a remote ultrasound task. The approach uses the Multi-Agent Gestural Instructions Comparer (MAGIC) framework to represent and compare the gestures performed by collaborators. Afterward, the Physical Instructions Assimilation (PIA) metric is

used to obtain task understanding insights from the gestures used to complete the task. Twenty participants performed a remote ultrasound task consisting of three subtasks: vessel detection, blood extraction, and foreign body detection. MAGIC's gesture comparison approaches were compared against two other gestures comparison approaches based on how well they replicated human-annotated gestures matchings. Our approach outperformed the others, agreeing with the human baseline over 76% of the times. Subsequently, a correlation analysis was performed to compare PIA's task understanding insights with those of three other metrics: error rate, idle time rate, and task completion percentage. Significant correlations were found between PIA and all the other metrics for task understanding estimation. This trend positions PIA as an effective gesture-based task understanding estimator, as it successfully complemented the task understanding insights obtained with the other commonly used metrics. Finally, post-experiment questionnaires were used to subjectively evaluate the participants' perceived understanding. The PIA score was found to be significantly correlated with the participants' overall task understanding, hinting to the relation between physical knowledge assimilation and self-perceived understanding. Overall, the results indicate that a gesture-based metric can be used to estimate task understanding during ultrasound tasks, which can have a positive impact in these techniques are performed and assessed.

References

1. Abuhamad A, Minton KK, Benson CB, Chudleigh T, Crites L, Doubilet PM, Driggers R, Lee W, Mann KV, Perez JJ et al (2018) Obstetric and gynecologic ultrasound curriculum and competency assessment in residency training programs: consensus report. Am J Obstet Gynecol 218(1):29–67
2. Aldekhyl S, Cavalcanti RB, Naismith LM (2018) Cognitive load predicts point-of-care ultrasound simulator performance. Perspect Med Educ 7(1):23–32
3. Alibali MW, Kita S, Young AJ (2000) Gesture and the process of speech production: we think, therefore we gesture. Lang Cognit Process 15(6):593–613
4. Amini R, Kartchner JZ, Stolz LA, Biffar D, Hamilton AJ, Adhikari S (2015) A novel and inexpensive ballistic gel phantom for ultrasound training. World J Emerg Med 6(3):225
5. Aussems S, Kita S (2019) Seeing iconic gestures while encoding events facilitates childrens memory of these events. Child Dev 90(4):1123–1137
6. Barmby P, Harries T, Higgins S, Suggate J (2007) How can we assess mathematical understanding. In: Proceedings of the 31st conference of the international group for the psychology of mathematics education, vol 2. ERIC, pp 41–48
7. Broaders SC, Cook SW, Mitchell Z, Goldin-Meadow S (2007) Making children gesture brings out implicit knowledge and leads to learning. J Exp Psychol Gen 136(4):539
8. Cheng J, Zhou W, Lei X, Adamo N, Benes B (2020) The effects of body gestures and gender on viewer's perception of animated pedagogical agent's emotions. In: International conference on human–computer interaction. Springer, pp 169–186
9. Chung YA, Weng WH, Tong S, Glass J (2019) Towards unsupervised speech-to-text translation. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 7170–7174
10. Clark HH, Brennan SE et al (1991) Grounding in communication. Perspect Soc Shared Cognit 13(1991):127–149
11. Clark HH, Schaefer EF (1987) Collaborating on contributions to conversations. Lang Cognit Process 2(1):19–41
12. Cook SW, Yip TK, Goldin-Meadow S (2012) Gestures, but not meaningless movements, lighten working memory load when explaining math. Lang Cognit Process 27(4):594–610
13. Cornia M, Baraldi L, Serra G, Cucchiara R (2018) Paying more attention to saliency: image captioning with saliency and context attention. ACM Trans Multimed Comput Commun Appl (TOMM) 14(2):1–21
14. Czyzyk J, Mesnier MP, Moré JJ (1998) The neos server. IEEE J Comput Sci Eng 5(3):68–75
15. Dekker S (2017) The field guide to understanding'human error'. CRC Press, London
16. Dolan ED (2001) The neos server 4.0 administrative guide. Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National. Laboratory
17. Dufurrena Q, Ullah KI, Taub E, Leszczuk C, Ahmad S (2020) Feasibility and clinical implications of remotely guided ultrasound examinations. Aerosp Med Human Perform 91(7):592–596
18. Évain A, Argelaguet F, Strock A, Roussel N, Casiez G, Lécuyer A (2016) Influence of error rate on frustration of bci users. In: Proceedings of the international working conference on advanced visual interfaces, pp 248–251
19. Fillauer JP, Bolden J, Jacobson M, Partlow BH, Benavides A, Shultz JN (2020) Examining the effects of frustration on working memory capacity. Appl Cogn Psychol 34(1):50–63
20. Fruchter R, Cavallin H (2011) Attention and engagement of remote team members in collaborative multimedia environments. In: Computing in civil engineering (2011). American Society of Civil Engineers, pp 875–882
21. Fussell SR, Setlock LD, Yang J, Ou J, Mauer E, Kramer AD (2004) Gestures over video streams to support remote collaboration on physical tasks. Hum Comput Interact 19(3):273–309
22. Gergle D, Kraut RE, Fussell SR (2004) Action as language in a shared visual space. In: Proceedings of the 2004 ACM conference on Computer supported cooperative work, pp 487–496
23. Goldin-Meadow S, Beilock SL (2010) Actionś influence on thought: the case of gesture. Perspect Psychol Sci 5(6):664–674
24. Goutte C, Gaussier E(2005) A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: European conference on information retrieval. Springer, pp 345–359
25. Gropp W, Moré JJ (1997) Optimization environments and the neos server. In: Buhman MD, Iserle A (eds) Approximation theory and optimization. Cambridge University Press, Cambridge, pp 167–182
26. Haraldsen HM, Solstad BE, Ivarsson A, Halvari H, Abrahamsen FE (2019) Change in basic need frustration in relation to perfectionism, anxiety and performance in elite junior performers. Scand J Med Sci Sports 30(4):754–765
27. Hart SG (2006) Nasa-task load index (nasa-tlx); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, vol 50. Sage publications, Los Angeles, pp 904–908
28. Hoffman G (2019) Evaluating fluency in human–robot collaboration. IEEE Trans Hum Mach Syst 49(3):209–218
29. Huang W, Kim S, Billinghurst M, Alem L (2019) Sharing hand gesture and sketch cues in remote collaboration. J Vis Commun Image Represent 58:428–438

30. Kim S, Lee G, Huang W, Kim H, Woo W, Billinghurst M (2019) Evaluating the combination of visual communication cues for hmd-based mixed reality remote collaboration. In: Proceedings of the 2019 CHI conference on human factors in computing systems. ACM, p 173
31. Kirk D, Crabtree A, Rodden T (2005) Ways of the hands. In: ECSCW 2005. Springer, pp 1–21
32. Kirk D, Rodden T, Fraser DS (2007) Turn it this way: grounding collaborative action with remote gestures. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 1039–1048
33. Kirkpatrick AW, McKee JL, McBeth PB, Ball CG, LaPorta A, Broderick T, Leslie T, King D, Beatty HEW, Keillor J et al (2017) The damage control surgery in austere environments research group (dcsaerg): a dynamic program to facilitate real-time telementoring/telediagnosis to address exsanguination in extreme and austere environments. J Trauma Acute Care Surg 83(1):S156–S163
34. Knoblich G, Sebanz N (2006) The social nature of perception and action. Curr Dir Psychol Sci 15(3):99–104
35. Laurent DABS, Niazi AU, Cunningham MS, Jaeger M, Abbas S, McVicar J, Chan VW (2014) A valid and reliable assessment tool for remote simulation-based ultrasound-guided regional anesthesia. Reg Anesth Pain Med 39(6):496–501
36. Liu R, Holden MS (2020) Kinematics data representations for skills assessment in ultrasound-guided needle insertion. In: Hu Y, Licandro R, Noble JA, Hutter J, Aylward S, Melbourne A, Turk EA, Barrena JT (eds) Medical ultrasound, and preterm, perinatal and paediatric image analysis. Springer, pp 189–198
37. Madapana N, Wachs J (2017) Zsgl: zero shot gestural learning. In: Proceedings of the 19th ACM international conference on multimodal interaction. ACM, pp 331–335
38. Martinez-Moyano I (2006) Exploring the dynamics of collaboration in interorganizational settings. In: Creating a culture of collaboration: the international association of facilitators handbook, vol 4, p 69
39. McBeth PB, Crawford I, Blaivas M, Hamilton T, Musselwhite K, Panebianco N, Melniker L, Ball CG, Gargani L, Gherdovich C et al (2011) Simple, almost anywhere, with almost anyone: remote low-cost telementored resuscitative lung ultrasound. J Trauma Acute Care Surg 71(6):1528–1535
40. McBeth PB, Hamilton T, Kirkpatrick AW (2010) Cost-effective remote iphone-teathered telementored trauma telesonography. J Trauma Acute Care Surg 69(6):1597–1599
41. Mitragotri S (2005) Healing sound: the use of ultrasound in drug delivery and other therapeutic applications. Nat Rev Drug Discovery 4(3):255–260
42. Molnar-Szakacs I, Wu AD, Robles FJ, Iacoboni M (2007) Do you see what i mean? Corticospinal excitability during observation of culture-specific gestures. PLoS One 2(7):e626
43. Nelson BP, Melnick ER, Li J (2011) Portable ultrasound for remote environments, part i: feasibility of field deployment. J Emerg Med 40(2):190–197
44. Olszynski P, Heslop C, Atkinson P, Lewis D, Kim DJ, Pham C, Ritcey B (2020) Ultrasound at the point of care-grown up and moving out! Canad J Emerg Med 22(1):1–2
45. Pearson K (1895) Vii. note on regression and inheritance in the case of two parents. Proc R Soc Lond 58(347–352):240–242
46. Ping RM, Goldin-Meadow S, Beilock SL (2014) Understanding gesture: is the listener's motor system involved? J Exp Psychol General 143(1):195
47. Rojas-Muñoz E, Wachs JP (2019) MAGIC: a fundamental framework for gesture representation, comparison and assessment. In: 2019 14th IEEE international conference on automatic face and gesture recognition (FG 2019). IEEE, pp 1–8
48. Rojas-Muñoz E, Wachs JP (2020) Beyond MAGIC: matching collaborative gestures using an optimization-based approach. In: 2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)(FG), pp 296–303
49. Rojas-Muñoz E, Wachs JP (2020) The MAGIC of e-health: a gesture-based approach to estimate understanding and performance in remote ultrasound tasks. In: 2020 15th IEEE international conference on automatic face and gesture recognition (FG 2020)(FG), pp 314–318
50. Roth WM (2001) Gestures: their role in teaching and learning. Rev Educ Res 71(3):365–392
51. Sacks H, Schegloff EA, Jefferson G (1978) A simplest systematics for the organization of turn taking for conversation. In: Studies in the organization of conversational interaction. Elsevier, pp 7–55
52. Schlager D, Sanders AB, Wiggins D, Boren W (1991) Ultrasound for the detection of foreign bodies. Ann Emerg Med 20(2):189–191
53. Schubotz L, Özyürek A, Holler J (2019) Age-related differences in multimodal recipient design: younger, but not older adults, adapt speech and co-speech gestures to common ground. Lang Cognit Neurosci 34(2):254–271
54. Sebanz N, Knoblich G, Prinz W (2003) Representing others' actions: just like one's own? Cognition 88(3):B11–B21
55. Tang Y, Cheng S, Yang Y, Xiang X, Wang L, Zhang L, Qiu L (2020) Ultrasound assessment in psoriatic arthritis (psa) and psoriasis vulgaris (non-psa): which sites are most commonly involved and what features are more important in psa? Quant Imaging Med Surg 10(1):86
56. Thorn S, Gopalasingam N, Bendtsen TF, Knudsen L, Sloth E (2016) A technique for ultrasound-guided blood sampling from a dry and gel-free puncture area. J Vasc Access 17(3):265–268
57. Vatsvåg V, Todnem K, Næsheim T, Cathcart J, Kerr D, Oveland NP (2020) Offshore telementored ultrasound: a quality assessment study. Ultrasound J 12(1):1–12
58. White R, Gunstone R (2014) Probing understanding. Routledge, London
59. Wong I, Jayatilleke T, Kendall R, Atkinson P (2011) Feasibility of a focused ultrasound training programme for medical undergraduate students. Clin Teach 8(1):3–7
60. Wydo S, Seamon M, Melanson S, Thomas P, Bahner DP, Stawicki SP (2016) Portable ultrasound in disaster triage: a focused review. Eur J Trauma Emerg Surg 42(2):151–159
61. Yeo LM, Tzeng YT (2020) Cognitive effect of tracing gesture in the learning from mathematics worked examples. Int J Sci Math Educ 18(4):733–751