Revista Romana de Interactiune Om-Calculator 8(4) 2015, 321-340

Understanding perceptual-gestural knowledge in TEL systems with eye-tracking

Ben-Manson Toussaint[1,2], Vanda Luengo[3], Francis Jambon[1]

[1] Université de Grenoble Alpes
Laboratoire d'Informatique de Grenoble, F-38000 Grenoble, France
E-mail: francis.jambon@imag.fr

[2] Ecole Supérieure d'Infotronique d'Haïti
29, 2e ruelle Nazon, Port-au-Prince, Haiti
E-mail: benmanson.toussaint@esih.edu

[3] Université Pierre-et-Marie-Curie
Site Jussieu. 3, rue des Fossés St Bernard, 75252 Paris, France
E-mail: vanda.luengo@lip6.fr

Abstract. This paper presents our methodology to capture and model multimodal interactions in Intelligent Tutoring Systems (ITS). We are specifically interested in perceptual-gestural interactions combining perceptions, gestures and other type of actions. Traces of such interactions are multisource and heterogeneous. The challenge is to foster their representation into sequences that render their multimodal nature. In this work, we want to show how the proposed representation yields the analysis of the influence of visual perceptions on learners' performance. Our case study is the ITS TELEOS, a simulation-based Intelligent Tutoring System dedicated to percutaneous orthopedic surgery. We also conducted an experiment on PILOTE 2, a flight simulation environment, to give a proof of concept of the genericity of our propositions.

Keywords: Intelligent Tutoring Systems, perceptual-gestural knowledge, eye-tracking, multimodal interactions.

1. Introduction

Knowledge is considered as perceptual-gestural when it can be described as a combination of different types of knowledge, specifically: theoretical knowledge, perceptual-knowledge and gestural knowledge. It is underlined in Intelligent Tutoring Systems, by interactions involving actions and/or gestures along with perceptions. Perceptions are used as controls for deciding on these actions/gestures execution or validation (Luengo et al., 2011). We assume that they provide useful insights on the information gathered by learners to support their decisions and thus, are non-negligible in a didactic point of view.

However, this type of knowledge is empirical and often tacit. As a consequence, it is hard to capture and model. In fact, capturing perceptual-gestural knowledge in a learning environment requires the use of complementary sensing devices that generate heterogeneous traces. In this study, we are specifically interested in learner's visualizations as support to executed actions and gestures.

Our case study is TELEOS (Luengo et al., 2011), a simulation-based Intelligent Tutoring System dedicated to percutaneous orthopedic surgery. Knowledge involved in this domain is perceptual-gestural (Ceaux et al., 2009; Mathews et al., 2012). In this type of surgery, visual analyzes require the perfect mental coordination of 2D images (X-Rays) and 3D objects (X-Rays unit, patient's body, surgical tools) to insure the safe trajectory of surgical tools through the targeted anatomical area. The conducted experiment puts the focus on the influence of learners' behavior related to visual analysis on their performance. Also, we propose a second case study to analyze the replicability of our approach.

The rest of the paper is organized as follows. The second section presents related works on capturing and analyzing perceptions in Intelligent Tutoring Systems; the third section describes the methodology to capture multisource traces in our case study; the fourth section presents our proposition to formalize these traces into perceptual-gestural sequences; the fifth section presents the conducted experiment and results; and the sixth section draws our conclusion and perspectives.

2. Related work

The literature reports many prominent pieces of work on Intelligent Tutoring Systems dedicated to domains where perceptual-gestural knowledge is involved. We can mention ITSs that have been proposed for training helicopters (Mulgund et al., 1995) and planes (Remolina et al., 2004) piloting as well as for car driving (de Winter et al., 2008; Weevers et al., 2003).

As one of the most recent related researches, we can also cite CanadarmTutor that was designed to train astronauts of the International Space Station for handling an articulated robotic arm (Fournier-Viger et al., 2011). However, the emphasis is generally carried in these works on actions and gestures and not on the perceptions accompanying these latter. In CanadarmTutor, the manipulation of the robotic arm from one configuration to another is guided by cameras through the operation scenes. Visual perceptions that are likely in play for this guidance would be worth further analysis.

Other works have been conducted on the analysis of perceptions in learning contexts. For example, visual perceptions are captured and analyzed to deduce learners' cognitive abilities (Steichen et al., 2013) or their metacognitive skills in exploratory learning (Conati & Merten, 2007). Some researchers would rather use collected perceptual information for measuring the learners' mental workload or cognitive effort (Lach, 2013), or for inferring their behavior in the learning process (D'Mello et al., 2012; Mathews et al., 2012). In other studies, sensing devices are used for capturing postures, facial expressions and body language as emotional signals (Steichen et al., 2013).

For our part, we believe that perceptions denote knowledge states along with actions they are related to and, therefore, should be analyzed from an epistemic point of view. They can bring more precision to generated pedagogical feedback as experts strongly underline the importance of verifying specific anatomic points on the X-Rays to support decision or validation of surgical gestures (Ceaux et al., 2009).

The aim is to point out the benefits from studying perceptual-gestural knowledge up on its original multimodal characteristics. To realize this, we need first to foster the consistent representation of perceptions-related behaviors and actions/gestures into perceptual-gestural sequences.

3. Capturing visual interactions

3.1 Recording visual traces

The simulation interface of TELEOS is composed of sections that represent the main artifacts of a percutaneous operating room. Namely, as illustrated in Figure 1.a, it includes a 3D model section where the patient's model is displayed; the current and previous X-rays sections and the settings panel that embeds three settings sub-sections: the fluoroscope settings panel; the cutaneous marks panel and the trocar manipulation panel.

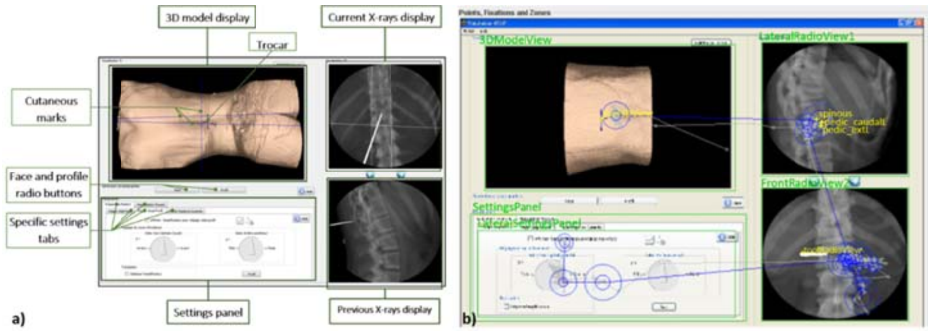

Figure 1. a) TELEOS simulation interface. b) Visual path traced by the scanpath analyzer

The “fluoroscope” is the unit used to generate X-rays throughout the operation; the “trocar” is the surgical tool used to reach the targeted vertebra; the cutaneous marks are lines drawn by the surgeon on the patient’s skin to spot the insertion point of the surgical tools. These sections represent the areas of interest (AOI) of the interface, i.e., areas that are recorded throughout the simulation session when they are set by the user.

AOIs associated to the X-rays embed the points of interest of the targeted vertebra, i.e., specific parts of the vertebra that should be analyzed on X-rays to support decisions on surgical actions and gestures. Figure 2 illustrates the modeling process of vertebrae’ points of interest.

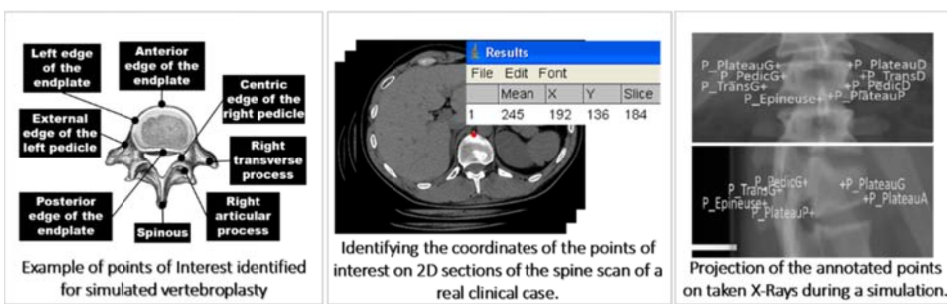

Figure 2. Modeling process of vertebrae’ points of interest

These points of interest have been determined by expert surgeons for each clinical case and integrated manually as annotations to the patients’ models (Luengo et al., 2011). The simulation environment is dynamic. As a consequence, the occurrences of the above-mentioned AOIs are also dynamic. In fact they can move, appear and disappear from the interface based on learners' activity and interaction with the system.

In other words, the presence of an AOI at a given location is neither stable nor predictable because it depends on the co-evolution of the activity, the learner and the state of the simulation. As an illustration in TELEOS, the position of the points of interest that mark the pedicles of a vertebra, changes with respect to the angle of capture of the X-rays. For addressing this challenge, a specific tool has been implemented to capture learners behaviors related to visual perceptions in a dynamic context (Jambon & Luengo, 2012).

3.2 Categorizing visual perceptions

Perceptions should be differentiated based on the cognitive efforts they demand or the intent they underlie. In fact, some perceptions intend a precise analysis of the environment whereas others are simple information gathering on the environment state. Other parameters can also help at differentiating learners' perceptions and inferring their intent, their resolution strategy or their profile. Considering visual perceptions, those parameters are possibly the fixations duration, their frequency, the most gazed areas and points or a combination of all those. We propose that visual perceptions should be considered into two different categories: (1) visual perceptions of verification/validation type and (2) visual perceptions of decision type.

The first kind of perceptions, of verification type, underlies cognitive activity that aims at analyzing the environment by putting into play precise knowledge elements. They target specific points of the environment that reflect the consequences of executed actions or gestures. More than a simple information gathering, their role is to verify and validate those latter. For instance, in vertebroplasty, visual analysis of the position of the spinous of a vertebra as a landmark to decide on the correctness of its centering, underlie perceptions of verification and validation.

The second kind of perceptions, of decision type, is less precise than perceptions of verification/validation type. Their role is limited to general information gathering on some specific artifacts of the environment. Generally, those artifacts are related to the tools that are used to act on the environment. As an illustration from our case study, we can cite the controller used for handling the X-Ray unit (*cf.* Setting panels in Figure 1.a).

3.3 Heterogeneity of recorded traces

Visual traces are recorded in TELEOS along with punctual actions executed on the simulation interface and gestures executed with a haptic arm that simulates the trocar. Traces from these three sources are recorded independently. They are heterogeneous both in their content type and format, but also in their time granularities (Toussaint et al., 2015).

4. Formalization of perceptual-gestural sequences

Learner's interactions involved in a simulation session are captured into three different modalities. From this point, the challenge is to link those latter into sequences that reflect properly all aspects of underlying knowledge elements. These sequences are referred to as perceptual-gestural sequences. We describe their representation in the following sections.

4.1 Characterizing perceptual-gestural sequences

For representing the different modalities with their temporal order, we consider each element of the sequence as an item. Each group of items with similar time of occurrence is considered as an itemset. A sequence is composed of several itemsets.

The formal definition of perceptual-gestural sequences is as follows:

$$S = <(A_{i,\ i=1..p}\ [a_{ij,\ j=1..q}];\ |\ G_{i,\ i=1..r}\ [g_{ij,\ j=1..s}];\ [P_{k,\ k=1..v}\ [q_{kl,\ l=1..w}]]);\ (P_{k,\ k=1..v}\ [q_{kl,\ l=1..w}])>$$

The parentheses "(" and ")" define the itemsets in the sequence. There is no restriction to represent items from different categories in the same itemset if their occurrences are simultaneous. For example, an action will be represented in the same itemset as a visualization item if this action is performed simultaneously with the visual information capture.

- $A_i = A_1 \ldots A_p$ is the set of actions ; $p \geq 1$

$$\begin{cases} \text{if } r >= 1 \text{ then } p >= 0 \\ \quad \text{else}, p >= 1 \end{cases}$$

- $G_i = G_1 \ldots G_r$: is the set of gestures

$$\begin{cases} \text{if } p >= 1 \text{ then } r >= 0 \\ \quad \text{else}, r >= 1 \end{cases}$$

- $|A \cup G| \geq 1$: a perceptual-gestural sequence comprises at least one action or gesture.
- $a_{ij} = a_{i1} \ldots a_{iq}$: set of actions parameters ; $q >= 0$
- $g_{ij} = g_{i1} \ldots g_{is}$: set of gestures parameters ; $s >= 0$
- $P_k = P_1 \ldots P_v$: set of perceptions ; $v \geq 1$
- $q_{kl} = q_{k1} \ldots q_{kw}$: set of perceptions parameters; $w \geq 0$.

In layman's terms, a perceptual-gestural sequence includes actions and gestures with perceptions that support their execution and that occur or not at the same time.

4.2 Characterizing enriched perceptual-gestural sequences

In the perspective of increasing the precision of learners' interactions reported in a perceptual-gestural sequence, we propose to enrich it with information from the system. Specifically, we propose to take into account the "reactions" of the learning environment related to learners' interactions. Typically, we consider two types of interactions from the system: (1) the simulation states and (2) the automatic evaluations based on expert rules.

The simulation states denote the current positions of the different artifacts represented in the simulation interface. The automatic evaluations based on expert rules refer to the evaluations of learners' actions produced by the diagnosis module of the ITS. These evaluations are based on expert rules provided by experts of the domain and are termed as *"situational variables"* (Minh Chieu et al., 2010). The integration of this information produce what we call “enriched perceptual-gestural sequences”. These sequences are formally defined as follows:

$$S_a: < ([\square_i], S_i\, [\square_{ij,\, j=1..\mu}])_{\, i=1..\eta}\ (\square_{q=1..v}[\upsilon_{r=1..w}])>$$

As presented in (2), an enriched perceptual-gestural sequence contains one or several perceptual-gestural sequences. These latter are characterized by information related to simulation states and annotations based on automatic evaluations from expert rules (situational variables). An evaluation can possibly assess several sequences:

- $\square_i$ represents the temporal tag stating the order of the sequence S_i in the enriched sequence ;
- S_i is the set of perceptual-gestural sequences composing the enriched sequence
- $\square_{ij}$ is the set of simulation states recorded in a sequences S_i
- $\square_q$ is the set of situational variables in the enriched sequence
- υ_r are the values of the situational variables

5. Reifying the proposed model

The model described in the previous section is reified with the tools of the framework PeTRA (Toussaint et al., 2015). PeTRA is a framework of treatment developed specifically for handling treatment of multisource and heterogeneous traces and transforming them into perceptual-gestural sequences with respect to the model described above.

Figure 3 gives an overview of the framework and its tools (also referred to as “operator”). After preparation and transformation operations, the obtained base of sequences can be exploited for Learning Analytics and Educational Data Mining purposes. We will not present the framework in extension in this paper. Detailed presentation of the treatment process with PeTRA can be found in Toussaint et al. (2015). We briefly present below the main operators of the framework used to obtain perceptual-gestural sequences on which this study has been conducted, namely, the merger, the annotator and the semantizer.

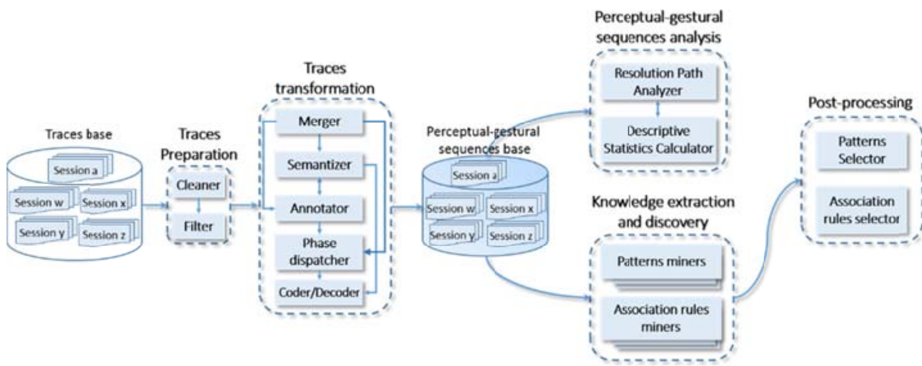

Figure 3. Overview of PeTRA: PErceptual-gestural treatments TRAces framework

Figure 4 summarizes more specifically, the treatment process for generating perceptual-gestural sequences from TELEOS traces with PeTRA. The merger's role is to automatically link the different modalities of learners' interactions into sequences that render their perceptual-gestural nature. In our case study, the merger was used to connect every punctual action executed by the learner with perceptions that support its execution. In other word, this operator produces the perceptual-gestural sequences as formalized in Section 4.1.

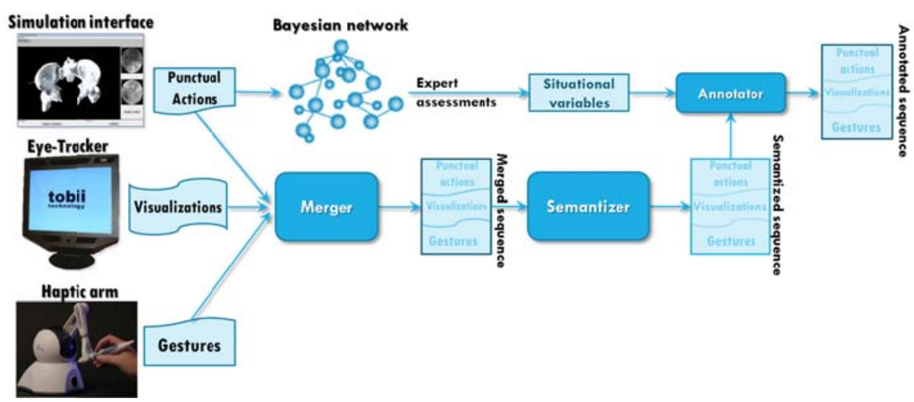

Figure 4. Treatment schema for generating enriched perceptual-gestural sequences from TELEOS traces

Figure 5 illustrates with an example perceptual-gestural sequence generated from traces recorded in TELEOS. This sequence represents the execution of the action *"Impacter_Trocart"*. It reports the actions of putting the surgical tool in contact with the targeted vertebra. The gesture specifies that the trocar has been simultaneously inserted (*trocart_translation_anterieur*), while being moved to the right side of the patient (*trocart_translation_droite*) and in the direction of the lower body of the patient (*trocart_translation_caudale*). The visual perceptions parameters report the areas of interest of the interface that has been gazed. For instance, *O_outil_vue3D* refers to the verification of the position of the trocar on 3D model of the interface. *O_manipReglage_1* refers to the visualization of the setting panel of the fluoroscope.

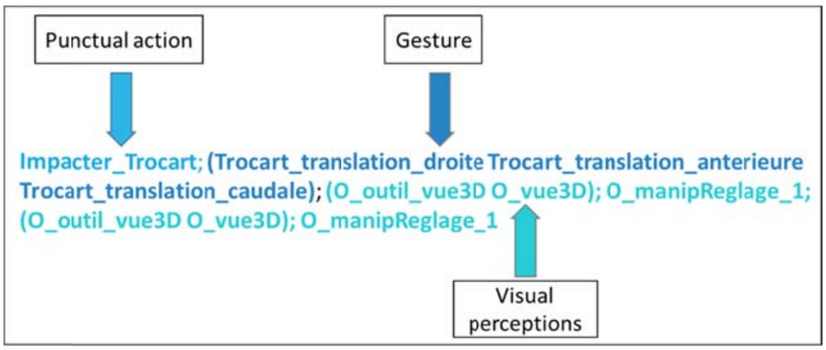

Figure 5. Example of perceptual-gestural sequence

To enrich generated perceptual-gestural sequences, two more operators are involved: the semantizer and the annotator. The semantizer is used to automatically transpose the changes of the positions of the environment objects into semantic states (e.g., *"the trocar has a cranial and left incline"*). The semantic denominations used for this case study refer to the standard anatomical terms of location.

The annotator, for its part, is used to connect expert assessments produced automatically by the knowledge diagnosis module of the Intelligent Tutoring System, to learner's actions that they target. The parameters that carry these expert assessments values are referred to as "situational variables" (Minh Chieu et al., 2010). Figure 6 illustrates the representation of an enriched perceptual-gestural sequence as formalized in Section 4.2.

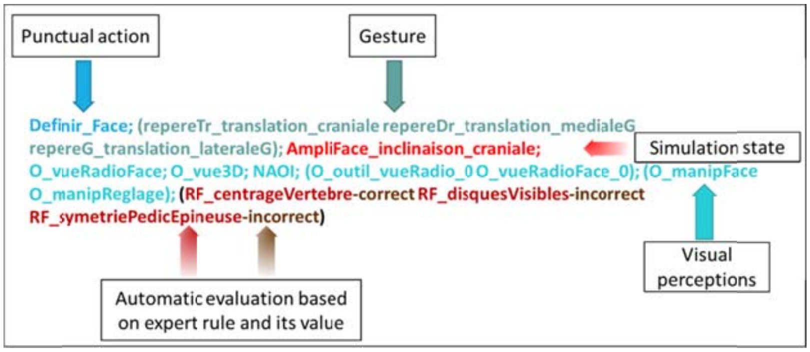

Figure 6. Example of enriched perceptual-gestural sequence

The sequence in Figure 6 represents an action of taking an X-Ray followed by the drawing of cutaneous marks. Cutaneous marks are used to spot the entrance point of the trocar on the patient's skin. The item on the simulation state, *AmpliFace_inclinaison_craniale*, specifies that the X-Ray unit is inclined toward the patient's head. The reported evaluations inform on the correctness of the taken X-Ray with corresponding expert rules. For example, *RF_centrageVertebre-correct* reports that the centering of the targeted vertebra on the front radio is correct. *RF_disquesVisibles-incorrect*, specifies that the visibility of spinal discs is incorrect; and *RF_symetriePedicEpineuse-incorrect*, that the symmetry of the spinous with respect to the vertebra pedicle, is incorrect.

6. Experiment 1: Evaluating the model

6.1 Experiment settings and methodology

Vertebroplasty operations are conducted in three different phases. During the first phase, the surgeon sets the fluoroscope position; the second phase is for cutaneous marks drawing, and the third one is devoted to the trocar insertion through the targeted vertebra. During simulation sessions, the Intelligent Tutoring System does not constrain the evolution of the exercise in a specific direction: the interns can freely jump from one phase to any other.

PeTRA includes an operator that was designed to help analyzing learners' interactions when solving a multi-phases exercises: the resolution path analyzer (*cf.* Figure 3). The resolution path analyzer is a learning analytics operator for eliciting learners' problems solving strategies, specifically, problems that involve a resolution process in several phases. For example, learners' resolution paths can include phases' validation, erroneous actions, modified actions, etc.

Traces used for this experiment were recorded from 9 simulation sessions of vertebroplasty performed by 5 interns and 1 expert surgeon of the University Hospital of Grenoble. The proposed simulation exercises consisted of treating a fracture of the 11th and/or the 12th thoracic vertebra. Each session lasted around one hour. The interns had never used the simulator before but had already assisted to at least one vertebroplasty operation in real life. We integrated an expert in the group so as to define a reference scope for the performance of a simulated vertebroplasty.

Table 1 presents the characteristics of collected and treated data. The enriched perceptual-gestural sequences have been generated from collected raw traces with the framework PeTRA (Toussaint et al., 2015).

Profiles	Session	#Enriched p.-g. seq.	#Fixations	#Incorrect SV	#Validation errors	#Corrective actions
Intern	S01	113	2033	750	9	11
Intern	S02	37	885	178	4	4
Intern	S03	33	690	208	3	5
Intern	S04	128	2482	644	10	39
	S05	41	858	174	6	10
Expert	S06	59	1452	249	4	31
	S07	47	1040	239	5	9
Intern	S08	117	2514	644	20	36
	S09	41	869	193	4	22

Table 1. Data characteristics

In the scope of this study, we wanted to determine in which extent the proposed model for the representation of perceptual-gestural sequences facilitated the analysis of learners' performance. More specifically, we were interested in determining how the proposed representation revealed the influence of visual perceptions on learners' errors during a simulation. For this purpose, we considered the following parameters:

- the number of erroneous validation of actions
- the number of corrective actions applied on these validation errors
- the number and type of visual perceptions associated to these corrective actions
- and the number of situational variables reported as incorrect

As the number of sequences varied greatly from one session to another, we used the average number of visualizations per sequence that better reflects the trend of visual analysis. The same applies for situational variables reported as incorrect.

6.2 Results

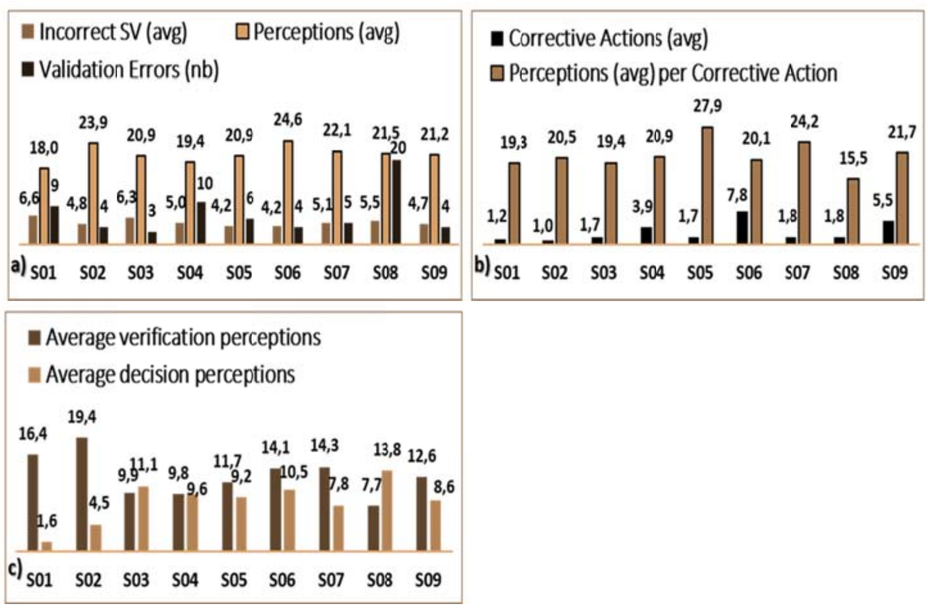

Figure 7. a) Histogram of average incorrect situational variables, average visual perceptions and number of validation errors per session. b) Histogram of corrective actions and average associated visual perceptions per session. c) Histogram of average verification perceptions and average decision perceptions

The graph of Figure 5.a summarizes the distribution of visual perceptions, incorrect situational variables and validation errors for each session. Pearson correlation indicates strong negative relationship (-0.62) between visual perceptions and incorrect situational variables. On the other hand, correlation between visual perceptions taken as a whole and validation errors is rather moderate (-0.31). However, strong negative correlation is noticed between visual perceptions of verification type and validation errors (-0.53).

6.3 Discussion

The session with the highest rate of perceptions (24.6) reports 19% fewer incorrect situational variables than the others, in average. The same is observed between visual analysis and validation errors for all the sessions, except for S08.This can be explained by the fact that the subject performed few corrective actions and visual analysis to support these actions. In fact, in the graph b of figure 7, we can notice that this session has one of the lowest averages of corrective sequences (1.7) along with the lowest average of visual perceptions (15.5) associated to these sequences. As a comparison, session S02 reported the lowest average of corrective actions (1.0, see Figure 7.b) but numerous visual analyzes (20.5) for supporting validation decisions and consequently limiting the number of errors (4. *Cf.* Figure 7.a). Moreover, it can be seen in Figure 7.c that few visual analysis in S08 are of verification type (7.7 against 13.8 of exploration perceptions).

Session S09 was performed by the same intern. Conversely, in that session, less validation errors and fewer incorrect situational variables were observed even with approximately the same rate of visual perceptions. This is the consequence of the reversal of behavior related to visual perceptions and the execution of more corrective actions (*Cf.* Figure 7.b).

As a conclusion, the model proposed for the representation of perceptual-gestural sequences of learners' interactions so as they render their multimodal nature, is congruent. In other words, the proposed representation this type of sequences fosters analysis of learners' interactions that take into account their different modalities. In this case, it was possible to determine whether or not interns' visual perceptions influence their performance conducting vertebroplasty simulation.

7. Experiment 2: Evaluating the genericity of our proposition

From this point, we wanted to verify from which extent our propositions could suit other domains involving interactions that underlie the same type of knowledge. For this purpose, we applied our model for the representation of multimodal interactions involving visual behavior in the domain of aviation, and we gathered traces from a new simulation environment: PILOTE2.

7.1 Multimodal and heterogeneous nature of traces in PILOTE 2

Aviation apprenticeship combines theoretical and procedural knowledge. Theoretical part of knowledge in the domain refers to abstract concepts related to the operation of an aircraft (e.g., the necessary speed for taking off given the weight of the aircraft and the length of the runway). Procedural knowledge of the domain consists of cognitive processes related to decision-making and maneuvers execution. These are based on the state of the aircraft and the state of the world (the environment of the aircraft), and they involve motor skills necessary to the manipulation of the aircraft commands (joystick or steering).

Gathering information on the aircraft and the world states involves visual analysis of indicators on the dashboard and the aircraft surrounding area. Further, in case of instrument flight due to bad weather, piloting an aircraft requires specific knowledge about reading indicators for decision-making when the visibility of the environment is reduced or absent. Thus, in this case study, activity traces to be recorded include visual perceptions traces and procedures execution traces. In the simulation environment PILOTE2, this includes traces captured by an eye-tracker and traces of actions execution with the simulator's commands and from the simulation software interface. Figure 8 gives an insight of the traces produced from simulation sessions in PILOTE 2. The "event" category are traces recorded when an actions is executed with the aircraft commands and the "variables" category provides information on the state of the aircraft on a regular time interval basis or when an action is executed. The state of the aircraft includes the state of its main component (e.g., engine, flaps, fuel, brake, etc.) and its state relatively to the world (e.g., altitude, speed in the air, vertical speed, incline, cape, etc.).

Event
```
UsybusMsg: type=fsx:event | from=FlightSimConnect@binoclar | tc=1432027519987 |
name=ELEV_TRIM_UP | device=simulator
```

Variables
```
UsybusMsg: type=fsx:variables | from=FlightSimConnect@binoclar | FLAPS HANDLE
INDEX=0 | GENERAL ENG FUEL PUMP SWITCH:1=0 | RUDDER POSITION=0 | GENERAL ENG
THROTTLE LEVER POSITION:1=96 | data=controls | GENERAL ENG MIXTURE LEVER
POSITION:1=99 | GENERAL ENG PROPELLER LEVER POSITION:1=99 | tc=1426523947953 |
ELEVATOR POSITION=0 | BRAKE RIGHT POSITION=0 | FLAPS NUM HANDLE POSITIONS=3 |
BRAKE LEFT POSITION=0 | device=simulator | AILERON POSITION=0 | RECIP ENG LEFT
MAGNETO:1=0 | RECIP ENG RIGHT MAGNETO:1=0 | ENG ANTI ICE:1=0

UsybusMsg: type=fsx:variables | from=FlightSimConnect@binoclar | STALL WARNING=0
| data=dashboard | VERTICAL SPEED=0 | HEADING INDICATOR=89 | ATTITUDE INDICATOR
BANK DEGREES=28 | AIRSPEED INDICATED=0 | tc=1426523947953 | TURN COORDINATOR
BALL=0 | TURN INDICATOR RATE=0 | MAGNETIC COMPASS=89 | device=simulator | ENG
FUEL FLOW GPH:1=0 | ATTITUDE INDICATOR PITCH DEGREES=-19 | GENERAL ENG RPM:1=0 |
INDICATED ALTITUDE=1305
```

Figure 8. Traces recorded on PILOTE 2 at command manipulation (Event) and on the state of the aircraft (Variables)

We will not discuss in detail nature of all traces captured in PILOTE2 environment. However, we present eye-tracking traces gathered in the simulation environment.

7.2 Eye-tracker traces in PILOTE 2

Actions executed by pilots are supported by visual verification of the state of the aircraft. The tools states of the aircraft are displayed on the indicators of the dashboard. Figure 8 shows the visual path of the pilot on the dashboard (in blue) as it is captured during a flight simulation. Each visualization is also reported, with respect to their time of occurrence, on the monitoring screen (circled in red). The reported information sets the name of gazed areas of interest and duration of each visualization.

Areas of interest considered in PILOTE 2 are the following:

- ASI (Air Speed Indicator): speed of the aircraft in the air
- AH (Artificial Horizon): indicates the incline of the aircraft
- ALT: Altitude
- TC (Turn Coordinator): indicates the turning state of the aircraft and the symmetry of its trajectory
- DG (Directional Gyro) : indicates the current cape of the aircraft

- VSI (Vertical Speed Indicator) : indicates the vertical speed of the aircraft as a complementary information with the air speed indication
- RPM (Round Per Minute) : indicates the rotation speed of the engine and hence the power that it provides
- FF (Fluel Flow) : amount of fuel consumed by the aircraft

Figure 8. Visual behavior captured on flight simulation environment

Raw traces recorded by the eye-tracker on PILOTE2 are illustrated by an example in Figure 9. The main information reported by the traces is: the timecode of the visualization as the value of the variable “tc”, the fixed area or point of interest as the value of the variable “name” and the duration of the visualization.

```
UsybusMsg: type=eyetracking:fixinzone | from=EyeTrackingAnalyzer@hydre |
duration=269 | tc=1432029792954 | name=VSI | device=simulator
```

Figure 9. Raw eye-tracking traces recorded in PILOTE 2

7.3 Proof of concept

The experiment on the genericity of our proposition was conducted with enough traces to provide a proof on concept of their suitability for multimodal traces related to a different domain but not for analyzing apprentice pilot's learning.

We showed our model for the representation of perceptual-gestural sequences fosters the representation of pilot's visual behavior with respect to executed actions during flight simulations (Toussaint, 2015). Further, as illustrated in Figure 10, we could link, in enriched perceptual-gestural sequences, pilot's visualizations to learners' errors and successes reported by the system.

```
AH;AH;RPM;ASI;RPM;(incremente_gaz);(volets_sortis vitesse_stable vz_stable
altitude_stable);(incremente_vitesse);(volets_sortis vz_stable gaz_stable
altitude_stable);TC;AH;(AH ALT);(incremente_vitesse);(volets_sortis vz_stable
gaz_stable altitude_stable);ASI;(incremente_vitesse);(volets_sortis vz_stable
gaz_stable altitude_stable);AH;(incremente_vitesse);(volets_sortis vz_stable
gaz_stable altitude_stable);(AH ALT);(incremente_vitesse);(volets_sortis vz_stable
gaz_stable altitude_stable);ASI;(incremente_vitesse);(volets_sortis vz_stable
gaz_stable altitude_stable);(TOF_volets-correct TOF_vitesse-incorrect TOF_gaz-
correct);(TOF_volets-correct TOF_vitesse-correct TOF_gaz-correct);take-off
```

Figure 10. Enriched perceptual-gestural sequence generated from flight simulation traces.

8. Conclusion and future work

We presented in this paper our proposition for addressing the challenge of processing multi-source heterogeneous traces from Intelligent Tutoring Systems dedicated to domains involving perceptual-gestural knowledge. The aim is to consistently connect perceptions to actions and gestures they support so as to foster the consideration of all aspects of involved knowledge. We demonstrated the efficiency of our proposition for processing traces recorded on TELEOS, a simulation-based Intelligent Tutoring System dedicated to percutaneous orthopedic surgery. We showed that our proposition for the formalization of perceptual-gestural sequences fosters the analysis of the influence of visual perceptions on interns' performance in simulation sessions of vertebroplasty.

In a second experiment, we also provided proof of concept of the genericity of our proposition based on traces from flight simulation sessions and a simulation environment dedicated to aviation: PILOTE2.

We plan to go further by extending our analyses to the measure of the effective gain from taking into account perceptual aspect of multimodal knowledge compared to treatment that discard either facet of this type of knowledge.

References

Ceaux, E., Vadcard, L., Dubois, M., & Luengo, V. (2009). Designing a learning environment in percutaneous surgery: models of knowledge, gesture and learning situations. *Paper presented at the EARLI Symposium "Simulation-Based Learning: Analyzing and Fostering Complex Skills in the Context of Medical Education*". Amsterdam.

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. Knowl.-Based Syst., 20(6), 557–574

D'Mello, S. Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. Int. J. Hum.-Comput. Stud., 70(5), 377–398

Fournier-Viger, P., Nkambou, R., Mayers, A., Mephu Nguifo, E., & Faghihi, U. (2011). A hybrid expertise model to support tutoring services in robotic arm manipulations. In Proceedings of the 10th Mexican International Conference on Artificial Intelligence (pp. 478–489). Springer, LNAI 7094.

Jambon, F., & Luengo, V. (2012). Analyse oculométrique « on-line » avec zones d'intérêt dynamiques : application aux environnements d'apprentissage sur simulateur. In Actes de la Conférence Ergo'IHM sur les Nouvelles Interactions, Créativité et Usages. Biarritz, France.

Lach, P. (2013). Intelligent Tutoring Systems: Measuring student's effort during assessment. O. R. Zaïane, & S. Zilles, (Eds.), Advances in Artificial Intelligence (pp. 346–351). Springer, LNCS 7884

Luengo, V., Larcher, A., & Tonetti, J. (2011). Design and implementation of a visual and haptic simulator in a platform for a TEL system in percutaneous orthopedic surgery. In J. D. Westwood et al. (Eds.), Medicine meets virtual reality (vol. 18, pp. 324–328). IOS Press.

Mathews, M., Mitrovic, A., Lin, B., Holland, J., & Churcher, N. (2012). Do your eyes give it away? Using eye-tracking data to understand students' attitudes towards open student model representations. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), Intelligent Tutoring Systems (pp. 422–427). Springer, LNCS 7315.

Minh Chieu, V., Luengo, V., Vadcard, L. (2010). Student modeling in orthopedic surgery training: Exploiting symbiosis between temporal bayesian networks and fine-grained didactical analysis. International Journal of Artificial Intelligence in Education, 20(3). 269–301.

Mulgund, S.S., Asdigha, M., Zacharias, G. L., Ma, C., Krishnakumar, K., Dohme, J.A., & Al, R. (1995). Intelligent tutoring system for simulator-based helicopter flight training. Flight Simulation Technologies Conference. American Institute of Aeronautics and Astronautics. Baltimore, USA.

Remolina, E., Ramachandran, S., Fu, D., Stottler, R., a Howse, W.R. (2004). Intelligent simulation-based tutor for flight training. In: Interservice/Industry Training, Simulation, and Education Conference (pp. 1–13). Orlando, FL, USA.

Ríos, H.V., Solís, A.L., Aguirre, E., Guerrero, L., Peña, J., & Santamaría, A. (2000). Facial expression recognition and modeling for virtual Intelligent Tutoring Systems. In: O. Cairó, L. E. Sucar, & F. J. Cantu (Eds.), Advances in Artificial Intelligence (pp. 115–126). Springer, LNCS 1793.

Steichen, B., Carenini, G., & Conati, C. (2013). User-adaptive information visualization: Using eye gaze data to infer visualization tasks and user cognitive abilities. Proc. Int. Conf. Intelligent User Interfaces 2013 (pp. 317–328). New York, ACM.

Toussaint, B.-M., Luengo, V., Jambon, F., & Tonetti, J. (2015). From heterogeneous multisource traces to perceptual-gestural sequences: The PeTra treatment approach. In C. Conati, N. Heffernan, A. Mitrovic, A., & M. F. Verdejo (Eds.), Proc. 17th Int. Conf. Artificial Intelligence in Education (AIED 2015)(pp. 480–491). Springer, LNCS 9112.

Toussaint, B.-M. (2015). Apprentissage automatique à partir de traces multi-sources hétérogènes pour la modélisation de connaissances perceptivo-gestuelles, Université Grenoble Alpes, PhD Thesis, Chapter 10.

Weevers, I., Kuipers, J., Brugman, A.O., Zwiers, J., van Dijk, E.M.A.G., & Nijholt, A. (2003). The virtual driving instructor, creating awareness in a multi-agent system. In Y. Xiang, & B. Chaib-Draa, B. (Eds.), Proc. 16th Conf. Canadian Society for Computational studies of intelligence (pp. 596–602). Springer.

de Winter, J. C. F., de Groot, S., Dankelman, J., Wieringa, P.A., van Paassen, M.M., & Mulder, M. (2008). Advancing simulation-based driver training: Lessons learned and future perspectives. Proc. 10th Int. Conf. Human-Computer Interaction with Mobile Devices and Services (pp. 459–464). New York, ACM