The impact of nonverbal cues on mediated tutoring interaction: an experimental study

Federico Tajariol, Jean-Michel Adam, Michel Dubois

2

The impact of nonverbal cues on mediated tutoring interaction: an experimental study

Video-mediated settings are more and more used in several mediated human activities, as teaching. Nevertheless, we do not deeply know yet the functions of nonverbal cues that these communication technologies make available to distant participants. The aim of this paper is to enrich the field of video-mediated communication studies, focusing on the effect of nonverbal cues on a video-mediated tutoring situation. We will first state communication theoretical underpinnings and then we will focus on the role of kinesic cues and ostensive-inferential cues in communication and mediated activities. After that we will describe our experimental method, explain in details the coding scheme we used to analyze the tutoring dialogue and the measures we collected. We will finally conclude with a discussion of our outcomes to highlight some convergences in regard to the previous literature and to the design of user interfaces in video-mediated tutoring environments.

2.1. INTRODUCTION

Communication technologies are more and more used in mediated human activities, as in teaching activities [COL96]. Audio-video technologies allow distant participants (e.g., teachers, tutors, students, tutees, etc) to see one another during the interaction [HEA97] and even to show one another the actions they are doing in their working environment [NAR97]. Nevertheless, even if pedagogical institutions have already introduced these technologies in their education services, by claiming that “the more communication links are available, the more the efficacy of the interaction is” [DAF86], the effects on the cognitive dimensions of the communication between tutors and tutees have not been clearly established yet.

Our research is a part of a large scientific project, called FORMID (FORMation Interactive à Distance, Distance Interactive Training), whose aim is to conceive and develop a web-based software architecture to let tutors and teachers to assist distant tutees in synchronous mediated situations [GUC06, ALM08, GLA09].

In this project, we conducted some experimental studies in order to highlight the function of video-mediated link [TAJ03, TAJ08]. In particular, we focus on the effect of nonverbal cues conveyed by video-mediated technologies, on tutoring interaction.

The chapter is structured as follows: after presenting our field of study, that is video-mediated communication and synchronous mediated tutoring, we will then describe our experimental method and the experimental study we conducted on the affordances of two nonverbal cues, kinesic and ostensive-inferential, on mediated tutoring dialogue. Next, we will show the outcomes, stressing that that the efficiency of tutoring dialogue is different according to the conveyed nonverbal cue. In the discussion section, we discuss our outcomes in regard to the previous literature.

2.2. BACKGROUND: MEDIATED TUTORING DIALOGUE AS A KIND OF JOINT ACTION

As for other human activities, tutoring dialogue is based on communication acts: we use language to act on our environment [AUS62] and to make our addressee believe by saying words [SEA69]. When we communicate with our partner (as a tutor or as a tutee), we cannot directly know her (his) thoughts and intentions, but we can just infer them by interpreting our partner’s utterances [SPE87] and nonverbal behavior [ARG77]. Nevertheless, communication is a joint action as playing a piano duet [CLA96] rather than a one-way human activity, where the speaker sends content and the addressee receives it. In the same way, tutoring dialogue is a cooperative action

[FOX93], where tutor selects the content of his (her) contribution according to the knowledge (s)he supposes tutee owns. Both tutor and tutee coordinate their turns to ground on a mutual understanding to confirm or not that they understand what has been said [CLA86].

When the tutor and the tutee are in a face-to-face setting, they both produce and interpret verbal and nonverbal cues in a multimodal process [GUR76]. In the nonverbal cues, we distinguish kinesic cues (i.e., facial expressions, postures, etc.) and ostensive-inferential cues (i.e., actions and deictic gestures). The kinesic cues ensure the speaker-listener turns transitions and inform the participants about the other person’s thoughts, feelings and intentions [AMA03]; the ostensive-inferential cues facilitate the verbal referring process, allow participants to coordinate their actions and to anticipate the other’s needs [NAR97]. In tutoring dialogue, the tutor infers when and how to help the tutee without disturbing him (her) by just observing his (her) facial expressions [FOX93]. If we imagine tutors and tutees acting in a audio-video mediated setting, we can suppose that they would act more properly if they have both ostensive-inferential and kinesic cues at their disposal rather just one kind of nonverbal cues. Thus, following the richness media framework [DAF86], we can suppose that the richer a communication setting is, the better the efficacy and the quality of the interaction will be.

Nevertheless, this framework is not suitable to foresee media effects nor is it to understand how nonverbal cues affect mediated activities [WHI02]. In other words, we cannot clearly state the effects of nonverbal cue on mediated tutoring interaction and tutees’ learning performance.

There two main reasons for this weakness: firstly, mediated tutoring is not widely studied yet as a kind of mediated communication; secondly, the literature in the field of video-mediated communication has produced mixed findings. On the one hand, some studies suggest that nonverbal cues do not always support participants to better perform their mediated activities [AND96, DAL98, DOH97, O’M96]. On the other hand, some studies show that kinesic and ostensive-inferential cues help distant participants to establish a mutual understanding [DOH97, ISA94, WHI93], that ostensive-inferential cues support participants to perform a mediated activity more quickly [FUSS03, GER04] and lead them to appreciate the quality of the interaction better than kinesic cues [FUS03, WHI93].

2.3. RESEARCH PROBLEM

Our main research questions are the following: which type of nonverbal cue improves the efficiency of tutoring dialogue? Would it be better for the quality of the interaction and of the mutual understanding if the tutor and the tutee see each other in a video-mediated setting (kinesic cues)? Or would it be better if the tutor observes the tutee's actions to improve (her) his learning (ostensive-inferential cues)?

To answer these questions, we set one experimental study, in which we compare the affordance of kinesic and ostensive-inferential cues on different measures concerning a mediated synchronous tutoring interaction, that is tutee's learning performance, tutoring dialogue and tutor's and tutees' evaluation on the quality of the interaction.

2.4. EXPERIMENTAL STUDY

In this section, we will describe the subjects, the experimental task and the apparatus, the procedure, the dependant measures we collected and our hypotheses.

Subjects. We recruited twelve tutors we paid 25 € each for their participation. We also recruited seventy-two undergraduate students in psychology (fifty women and twenty-two men; age M=23.8, S.D.=5.1), all of them unskilled at HTML programming. They received a credit for their social psychology course.

Task. Tutors and students were involved in a sort of practical pedagogical work. One tutor had to help two tutees to learn basic commands of HTML language and to create an easy web page. In fact, they had to edit some sentences of a text in bold and italic, as well as building an internal link. The tutor could help and communicate with only one student at once, as in a dyadic tutoring situation: in fact, when the tutor was talking to a tutee, the other one could not hear their dialogue.

Apparatus. The tutor and the two tutees were in three separate rooms, each one was equipped with a personal computer (central unit and monitor) and another monitor was used to show the person. Anytime the tutor wanted to communicate with a tutee, (s)he chose the student by pressing a button on a WIMP device. Both tutees could ask the tutor to help them anytime, sending him (her) a standard message by means of synchronous chat software. Each computer was equipped with a web browser and a simple text editor.

Experimental conditions. We set the three following conditions:

c1) audio & video-person (kinesic cues): the tutor and the tutees could see each other's face and upper torso on one screen (Fig. 1) and they communicated by means of the audio channel;

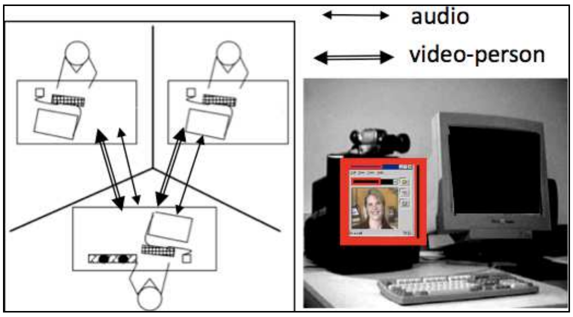

Figure 1. Left: Tutor (bottom) and two tutees (up).
Right: the tutor's PC screen.

c2) audio & video-actions (ostensive-inferential cues): the tutor and the tutees could communicate by the audio channel, and the tutor could observe both tutees' computer screens (by means of VNC software). So, the tutor's computer screen was split in two windows showing the tutees' desktops (Fig. 2) and (s)he could observe their activity and make the decision to help them at will;

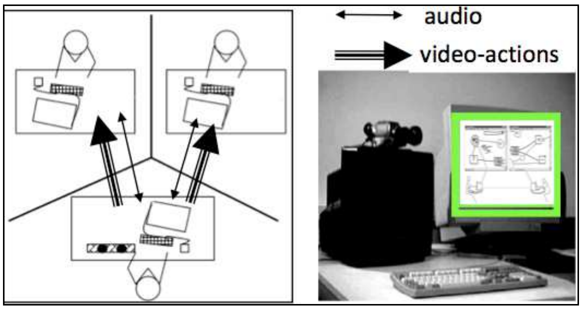

Figure 2. Left: Tutor (bottom) and two tutees (up).
Right: the tutor's PC screen

c3) audio & video-person & video-actions (kinesic & ostensive-inferential cues): the tutor could observe both tutees' face and upper torso and their computer screens (Fig. 3), while the verbal communication between the tutor and each tutee was supported by the audio channel.

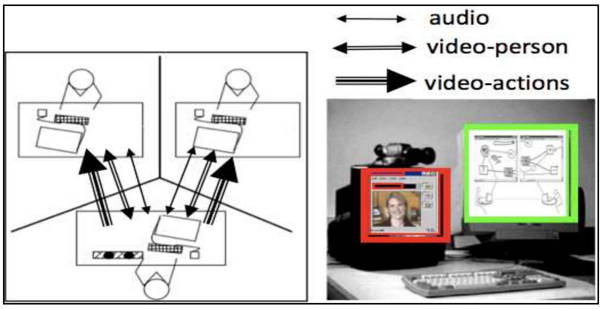

Figure 3. Left: tutor (bottom) and two tutees (up).
Right: tutor's PC screen

Each tutor performed the task in all conditions (within-participants design) and we counterbalanced the order of the tutors' performance to guard against order effects. About tutees, they were equally shared between the three experimental conditions: we controlled the age ($F(2, 69) = 1.73$; n.s.) and the HTML knowledge ($F(2, 69) = 0.16$; n.s.) of each group of tutees.

Procedure. Each experimental session lasted an hour and it was composed of three parts. In the first part, lasting 15 minutes, the researchers introduced the tutor to the two tutees and explained the aims of the experimentation. Then, the tutor and each tutee settled down in three separate rooms. Each subject filled in a consent form and was briefed on the main commands of the communicating technical apparatus. Next, each tutee answered a pre-test to evaluate his (her) HTML knowledge. Then the experimental session started and lasted 35 minutes (second part). At the beginning, each tutee received a four-page HTML manual, which contained some HTML basic commands: however, the students necessarily needed the tutor's help to design the web page. The tutor could help each student spontaneously or s(he) had to ask him (her) when a student asked for help. In the third part of the experimentation, each subject was asked to complete a questionnaire to evaluate how s(he) perceive the quality of the interaction during the experimental session, then s(he) was debriefed and dismissed.

Measures. We videotaped each experimental session (N=24) that we transcribed verbatim. Researches concerning the common grounding processes in communication [CLA96], the affordances of visual information in mediated communication [KRA03, KAR99] and the literature on linguistic and cognitive categories of the face-to-face tutoring dialogue [FOX93, CHI94, CHI97] inspired us the following measures:

a) the proactive behaviour of the tutor : we distinguished the tutors' spontaneous interventions towards the tutees as proactive interventions from the tutors' reactive interventions when they had replied to a tutee's call;
b) the learning score : we collected two measures. First, we measured the rules of HTML that tutees had learned during the session: tutees were asked to answer to four multiple choice questions at the beginning (pre-test) and at the end (post-test) of each session (the highest possible score was of 12 points). Secondly, we scored the quality of the web page that they had realized by means of a specific grid analysis (the possible highest score was 186 points);
c) tutor's and tutees' evaluations on the quality of the interaction: at the end of each experimental session, tutor and tutees are asked to fill in a questionnaire. We built a questionnaire starting from the concept of interpersonal awareness [MCC94, WAT96], which suggests that communication features have an effect on how participants assess the level of

social presence of their own distant partner and the level of mutual understanding they reached during the interaction. We added the dimension of clarity and the dimension of relevance of partner's utterances and, in order to have a deepest insight of the participant's feelings, we operationnalized these dimensions as followed: a) we called centrifugal impressions that each participant thinks about the contribution of his (her) distant partner; we called centripetal impressions that each participant thinks about his (her) own contribution to the interaction. The centrifugal impressions were measured by means of four questions (7-item Likert scale): each tutor was asked to evaluate the precision of tutees' questions and explanations, the relevance of tutees' answers and to assess the level of attention they showed during the interaction; each tutee was asked to evaluate the precision of tutor's questions and explanations, the level of attention during the interaction and the relevance of his (her) suggestions to tutee's questions. The centripetal impressions were measured through four questions (7-item Likert scale): tutors and tutees were invited to assess the precision of their own questions and explanations, the level of their own attention during the interaction and the relevance of their own explanations to the partner's questions;

d) the mutual understanding between the tutor and the tutees: we developed a coding scheme to categorize tutor's and tutees' verbal markers used in the common ground process, as well as the deictic utterances they used to refer to the graphical objects and commands. We also coded deictic expressions (i.e. T: "you move the icon down", S: "here?", T: "yes") and pronouns (i.e., T: "take this and move it on your desktop") (Tab. 2.1);

Table 2.1. Mutual understanding coding scheme

Role	Category	Utterance/Example
Tutor	To accept tutee's utterance	T: "*yes*" "*ok, right*"
	To check tutee's understanding	T: "*is it clear now?*", "*is it ok?*"
Tutee	To accept tutor's utterance	S: "*yes*", "*ok*"
	To check tutor's utterance	S: "*could you repeat, please?*", "*what?*"

e) the intrinsic speech acts concerning tutor-tutee dialogue : we classified the tutors' and tutee's speech acts which are related to the essential nature of the tutoring dialogue (Tab. 2.2) according the most relevant literature [FOX93, CHI94].

Table 2.2. Intrinsic tutoring speech acts

Role	Category	Utterance/Example
Tutor	To find out tutee's ongoing task	T: "*Did you try moving it on the red icon?*"
	To help tutee	T: "*Close the window and open the other file*"
	To encourage tutee	T: "*That's good you've nearly finished*"
Tutee	To give tutor information about ongoing task	S : "*I still have to finish this part of the exercise*"
	To ask tutor's help	S: "*Is the I tag in the HEAD part of the text?*"

Three trained coders analyzed the transcriptions of the dialogue. We calculated the reproducibility test [CAR96] to evaluate the reliability of coding: we obtained an average value for the students' speech acts (K = 58) and a high value for the tutors' speech acts (K = 82).

Hypotheses. We expected the tutor to be more proactive when (s)he could observe the tutee's actions (ostensive-inferential cues) rather than s(he) could only observe the tutees' facial expressions and gestures (kinesic cues), because (s)he was always aware of their difficulties and (s)he could take the floor before they asked for help.

Mutual understanding was expected to be easier when the tutor could observe the tutee's actions (ostensive-inferential cues) rather than the tutee's facial expressions (kinesic cues), because they have a same visual environment: in particular, we expected the tutor and tutees to describe less explicitly their actions and all related objects, so they would produce more deictic utterances.

We also predicted tutees' and tutors' centripetal and centrifugal impressions to be more positive when ostensive-inferential cues were available rather than kinesic cues were.

Finally, about tutee's learning performances, we expected tutees to improve them when the tutor could observe them during their practical work session, because the tutor could choose the most appropriate help according to their needs.

2.5. OUTCOMES

We conducted an analysis of variance (ANOVA) and a priori comparisons between the three conditions (Tab. 2.3).

Tutor's proactive behavior. The number of tutor's spontaneous interventions is statistically significantly different between conditions ($F(2, 33) = 4.671$; $p < .02$, $\eta^2=0.22$). In fact, the tutor is statistically more proactive in audio & video-person & video-actions than in audio & video-actions condition (M = 6.6 (1.9) *vs.* 4.6 (2), $t(33)=2.366$; $p < .02$, $d=1.02$) and than in audio & video-person condition (M = 6.6 (1.9) *vs.* 4.2 (2.2), $t(33)=2.859$; $p < .007$, $d=1.16$).

Tutees' learning performance. We first present outcomes for the test on HTML language (i) and then for the practical work (ii).
i) The score to the HTML test is statistically significantly different between conditions ($F(2, 69) = 3.119$; $p < .05$, $\eta^2 = 0.08$). A-priori analysis shows that tutees learn better in audio & video-person & video-actions than in audio & video-person condition (M = 4.3 (0.7) *vs.* 3.7 (1), $t(69)=2.38$; $p < .02$, $d=0.7$). However, results showed that the difference between audio & video-activity and audio & video-person conditions was only statistically tendential (M = 4.2 (1) *vs.* 3.7 (1), $t(69)=1.85$; $p=.07$, $d=0.5$).
ii) The web page score is statistically significantly different between conditions ($F(2, 69) = 5.776$; $p < .005$, $\eta^2 = 0.14$). The tutees perform significantly better in audio & video-person & video-actions than in audio & video-person condition (M = 5.9 (1.6) *vs.* 4.5 (1.5), $t(69)=3.27$; $p<.002$, $d=0.88$) and they do better in audio & video-actions condition than in audio & video-person condition (M=5.5 (1.2) *vs.* 4.5 (1.5), $t(69)=2.4$; $p < .02$, $d=0.73$).

Tutors' and tutees' evaluations of the interaction. We firstly present the outcomes concerning the centrifugal impressions (i-ii) and then the centripetal ones (iii-iv).

i) Tutors' centrifugal impressions: psychometric analysis showed a good reliability value of the four items (Cronbach's α=.77), so that we treated them as one vector only. An ANOVA revealed that tutors' centrifugal impressions are statistically significantly different between conditions ($F(2, 33) = 8.6$; $p < .001$, $\eta^2 = 0.34$). A-priori comparisons show that tutors evaluate tutees' interaction more positively in audio & video-person & video-actions rather than in audio & video-person condition (M=23.7 (2.4) *vs*. 19.2 (4), $t(17.84)=3.32$; $p < .005$, $d = 1.36$); moreover, tutors evaluate tutee's interaction more positively in audio & video-actions rather than in audio & video-person condition (M=23.2 (1.8) *vs*. 19.2 (4), $t(15.42)=3.12$; $p < .01$, $d = 1.29$).

ii) Tutees' centrifugal impressions: the reliability value of the items was weak (Cronbach's α=.30), so we separately treated each of the four items. An ANOVA revealed that tutees' centrifugal impressions are statistically significantly different between conditions ($F(8, 134) = 3.38$; $p < .001$, $\eta^2 = 0.17$). MANOVA results reveal a general effect of conditions on the four items ($F(8, 134) = 3.38$; $p < .001$, $\eta^2 = 0.17$). ANOVA detailed analyses were performed on each of the four items, and they revealed a statistically significant difference on the tutee's evaluation of the tutor's questions ($F(2, 69) = 3.854$; $p < .03$, $\eta^2 = 0.10$), tutor's explanations ($F(2, 69) = 4.424$; $p < .02$, $\eta^2 = 0.11$), tutor's level of attention during the interaction ($F(2, 71) = 3.867$; $p < .03$, $\eta^2 = 0.10$) et on the relevance of tutors' explanations ($F(2, 71) = 4.22$; $p < .02$, $\eta^2 = 0.11$). A-prori comparisons reveal that tutees perceive the tutor as more attentive (M = 4.8 (1.1) *vs*. 4.1 (1.1), $t(69) = 2.116$; $p < 0.4$, $d = 0.63$), they perceive the tutor's questions as clearer (M = 4.2 (0.9) vs. 3.5 (1.4), $t(39.79) = 2.049$; $p < .05$, $d = 0.59$) , the tutor's explanations as clearer (M = 5.4 (1.5) *vs*. 4.5 (1.3), $t(69) = 2.297$; $p < .03$, $d = 0.64$) and the tutor's explanations more proper (M = 5 (1.5) *vs*. 3.9 (1.2), $t(69) = 2.74$; $p < .009$, $d = 0.81$) in audio & video-person rather than in audio & video-activity condition. No statistically significant difference were found between audio & video-person and audio & video-actions conditions.

iii) Tutors' centripetal impressions. The psychometric analysis revealed a good reliability value of the four items (Cronbach's α=.78), so that we could shrink the four items into one vector. ANOVA show a statistically tendential difference between conditions ($F(2 , 33) = 2.979$; p =.06, η^2= 0.15). This outcome suggest that video-activity does not seem to have an effect on tutor's on centripetal impressions.
iv) Tutees' centripetal impressions. Psychometric analysis showed a low reliability of the answers to the four items (α=.50), and ANOVA did not reveal any statistically significant effect on the whole four items ($F(8, 134) = 0.376$; n.s., η^2= 0.22).

Mutual understanding. The number of deictic utterances is statistically significantly different between conditions ($F(2, 33) = 5.453$; p <.01, η□= 0.99). A-priori comparisons show that the tutors and the tutees produce significantly more deictic utterances in audio & video-actions than in audio & video-person condition (M=125.2 (21.6) *vs*. 90.9 (29.3), t(20.2)=3.27; p <.005, d=1.33) and more in audio & video-person & video-actions than in audio & video-person condition (M=134.3 (46.2) *vs*. 90.9 (29.3), t(18.6)=2.75; p <.01, d=1.12). About the tutors' and tutees' verbal markers targeted to mutual understanding, outcomes do not show any statistically significant differences between conditions.

Tutoring speech acts. We first present outcomes for tutors (i) and then for tutees (ii).
i) Tutors. One-way ANOVA shows that the number of tutor's speech acts to find tutee's ongoing task is statistically significantly different between conditions ($F(2,33)=4.529$; p <.02, η^2= 0.21): tutors produce fewer speech acts oriented to the tutee's task in audio & video-actions than in audio & video-person condition (M = 27.4 (7.6) *vs*. 35.3 (8.1), t(22)=2.8; p <.01, d=1) and fewer in audio & video-person & video-actions than in audio & video-person condition (M = 26.6 (6.8) vs. 35.3 (8.1), t(22)=2.6; p <.005, d =1.16). The number of tutor's speech acts to help the tutee is statistically significantly different between conditions ($F(2, 33) = 15.248$; p <.000, η^2=0.48): there are fewer tutor's help speech acts in audio & video-person than in audio & video-actions condition (M=14.8 (10.4) *vs*. 40.2 (12.9), t(33)=4.99; p <.000,

d=2.18) and fewer in audio & video-person than in audio & video-person & video-actions condition (M=14.8 (10.4) *vs*. 37.9 (13.8), t(33)=4.55; *p* <.000, *d*=1.91). The number of tutor's encouraging speech acts was statistically significantly different between conditions ($F(2, 33) = 4.813$; $p < .01$, $\eta^2 = 0.22$): tutors more often encourage the tutees in audio & video-person than in audio & video-actions condition (M=3.2 (2.3) *vs*. 0.8 (1.1), t(33)=2.69; *p*<.01, *d*=1.33) and they do more often in audio & video-person & video-actions condition than in audio & video-actions condition (M=3.2 (2.8) *vs*. 0.8 (1.1), t(33)=2.69; *p* <.01, *d*=1.12).

Table 2.3. Main outcomes of the experimental study

Measures	Categories	(c1) audio & video-person	(c2) audio & video-actions	(c3) audio & video-person & video-actions	Comparisons
Tutor's proactive behaviour	Tutor's spontaneous interventions	M = 4.2 (2.2)	M = 4.6 (2)	M = 6.6 (1.9)	C3 > C1 (*p*<.007) C3 > C2 (*p*<.02)
Tutees' learning performance	Tutee's HTML test	M = 3.7 (1)	M = 4.2 (1)	M = 4.3 (0.7)	C3 > C1 (*p*<.02) C2 > C1 (*p*=.07)
	Tutee's web page score	M = 4.5 (1.5)	M = 5.5 (1.2)	M = 5.9 (1.6)	C3 > C1 (*p*<.002) C2 > C1 (*p*<.02)
Subjective Impressions	Tutors' centrifugal impressions	M = 19.2 (4)	M = 23.2 (1.8)	M = 23.7 (2.4)	C3 > C1 (*p*<.005) C2 > C1 (*p*<.01)
Mutual understanding	Deictic terms	M = 90.9 (29.3)	M = 125.2 (21.6)	M = 134.3 (46.2)	C2 > C1 (*p*<.005) C3 > C1 (*p*<.01)
Intrinsic tutoring speech acts	Tutors' acts				
	To find out tutee's ongoing task	M = 35.3 (8.1)	M = 27.4 (7.6)	M = 26.6 (6.8)	C1 > C2 (*p*<.01) C1 > C3 (*p*<.005)
	To help tutee	M = 14.8 (10.4)	M = 40.2 (12.9)	M = 37.9 (13.8)	C2 > C1 (*p*<.000) C3 > C1 (*p*<.000)
	To encourage tutee	M = 3.2 (2.3)	M = 0.8 (1.1)	M = 3.2 (2.8)	C1 > C2 (*p*<.01) C3 > C2 (*p*<.01)
	Tutees' acts				
	To give tutor information about ongoing task	M = 43.8 (14.4)	M = 27.7 (18.4)	M = 26.5 (14.1)	C1 > C2 (*p*<.005) C1 > C3 (*p*<.01)

ii) Tutees. The number of the tutees' speech acts to give tutor information about their ongoing task is statistically significantly different between conditions ($F(2, 33) = 4.529$; $p < .02$, $\eta^2=0.21$): this implies that tutees less often give their tutors information about their ongoing task in audio & video-person & video-actions than in audio & video-person condition (M=26.5 (14.1) *vs*. 43.8 (14.4), $t(22)=2.5$; $p<.01$, $d=1.21$) and less often in audio & video-actions than in audio & video-person condition (M=27.7 (18.4) *vs*. 43.8 (14.4), $t(22)=2.6$; $p<.005$, $d=0.97$). All other tutees' speech acts are substantially equivalent between conditions, no statistically significant differences were found.

2.6. DISCUSSIONS

This study wanted to understand the effect of nonverbal cues on a video-mediated synchronous tutoring interaction. We found that both ostensive-inferential (video-action) and kinesic (video-person) cues can be indistinguishable useful in tutoring interaction. First, we found that the tutors act proactive more often when they can both watch tutees' facial expressions (kinesic cues) and tutees' actions (ostensive-inferential cues). This outcome would to corroborate the richness media theory [DAF86], claiming that if we want to create the benefits of a face-to-face setting at distance we should make available as many cues as possible. However, we stress that other studies show the effect of ostensive-inferential cues alone, rather than both nonverbal cues [KAR99, KRA03].

Second, tutees' learning performance increase when both kinesic and ostensive-inferential cues are available. In particular, both HTML test and web page scores increase when ostensive-inferential cues are available, in a standalone setting or coupled with kinesic cues. These outcomes confirm the mixed findings on the effect of ostensive-inferential cues on task performance: some studies show the positive effect in reducing time performance rather than the quality of the performance [FUS03, GER04, KRA03], whereas other studies did not find any effect [KAR99, KRA03].

Third, nonverbal cues have different effects on tutors' and tutees' evaluations of the quality of the interaction. We found that nonverbal cues have no effect on centripetal impressions, that is, how the participants evaluate his (her) contribution to the interaction. In contrast, both tutors' and tutees' evaluations differ according to the nonverbal cues available: tutors think that tutees are more attentive and their contribution are more interesting when ostensive-inferential cues were available, with or without kinesic cues. From a theoretical side, this outcome would corroborate the idea that the more nonverbal cues are available during the interaction, the less confusing the interpretation of nonverbal cues is. From the side of tutees, outcomes show that they evaluate better the tutor's contribution when kinesic cues are available to them, with or without ostensive-inferential cues.

Fourth, to ensure their mutual understanding, tutors and tutees produce an equivalent number of verbal markers whatever nonverbal cues they have at their disposal. This outcome suggests that, as in face-to-face tutoring [VAN03], the tutor verbally checks the tutee's comprehension even if nonverbal cues are available. We would like to stress that tutor and tutees know when they can apply a deictic strategy. In fact, deictic expressions increase when ostensive-inferential cues are available to the tutor and the tutees: the latter are aware that the former can observe them anytime, and this help all participants to build their common ground [CLA91]

Finally, tutor and tutees produce different intrinsic tutoring speech acts according to the nonverbal cues they can use. When ostensive-inferential cues are available, the verbal exchange about tutee's task decrease and tutors' helping acts increase, confirming previous researches [KAR99, KRA03]. This suggests that ostensive-inferential cues are necessary to that tutoring interaction be successful, especially when procedural knowledge is to be learnt. In contrast, when kinesic cues are available, tutors' encouraging acts increase. These two outcomes suggest that ostensive-inferential cues would shape the tutoring dialogue on the task to accomplish (functional side), whereas kinesic cues would tailor it rather to the tutee's feelings (relational side).

What we found can have implications for the design and the use of video-mediated settings for tutoring interaction. For instance, if we wanted the tutor to be proactive, it would better to let kinesic cues available to the tutor during the interaction; or, if we preferred the tutees to improve learning, it would be better to let the tutor observe the tutees' ongoing actions.

We would like to stress that these outcomes should be taken with regard to the limits of our study: we wonder whether these outcomes would change by increasing participants' experience of video-mediated settings, or by increasing the number of tutees for each experimental session and, by choosing another learning content, that is declarative topics rather than procedural ones.

Despite the weight of these limits, the data we presented and the comparison we made with the most relevant literature in the field, suggest that it is necessary to set up a proper methodology that consider all dimensions of human activity, that is verbal interaction, task performance and subjective impressions. Otherwise, if we considered each of these dimensions alone, we would not take a right picture of the mediation.

REFERENCES

[ALM08] Adam, J.M., Lejeune, A., Michelet, S., David, J.P., Martel, C.: *Setting up on-line learning experiments: the LearningLab platform.* In P. Diaz, Kinshuk, I. Aedo & E. Mora (Eds), Proceedings of the 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008). Los Alamitos, CA: IEEE Computer Society.

[AMA03] Amado, G. & Guittet, A.: *Dynamique des communications dans les groupes*. Paris : Colin, 2003.

[AND96] Anderson, A.H., Newlands, A., Mullin, J., Fleming, A.M., Doherty-Sneddon G. & Van Der Velden, J.: *Impact of video-mediated communication on simulated service encounters*. Interacting with computers, 8,193-206, 1996.

[ARG77] Argyle, M. & Graham, J.: *The Central Europe Experiment : looking at persons and looking at things.* Journal of Environmental Psychology and Nonverbal Behaviour, vol.1, 6-16, 1977.

[AUS62] Austin, J.L.: *How to do things with words.* Oxford : Oxford University Press, 1962.

[CAR96] Carletta, J., Isard, A., Isard, S., Kowtko, J.C., Doherty-Sneddon, G., & Anderson, A.H.: *The reliability of a dialogue structure coding scheme.* Computational Linguistics, 23(1), 13-31, 1996.

[CHI94] Chi, M.T.H., De Leeuw, N., Chiu, M.H. & Lavancher, C.: *Eliciting self-explanations improves understanding*. Cognitive Science, vol.18, n.3, 439-477, 1994.

[CHI97] Chi, M.T.H.: *Quantifying qualitative analyses of verbal data : a practical guide*. The Journal of the Learning Sciences, vol. 6, n.3, 271-315, 1997.

[CLA86] Clark, H.H. & Wilkes-Gibbs, D.: *Referring as a collaborative process.* Cognition, vol. 22, 1-39, 1986.

[CLA91] Clark, H. H., & Brennan, S. E.: *Grounding in communication.* In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), Perspectives on Socially Shared Cognition, 127-149. American Psychological Association, 1991.

[CLA96] Clark, H.H.: *Using languages*. Cambridge : Cambridge University Press, 1996.

[COL96] Collis, B.: *Tele-learning in a digital world: The future of distance learning*. London: International Thomson Publications, 1996.

[DAF86] Daft, R.L., & Lengel, R.H.: *Organisational information requirements, media richness and structural design*. Management Science, vol. 32, 554-571, 1986.

[DAL98] Daly-Jones, O., Monk, A.F. & Watts, L.A.: *Some advantages of video-conferencing over high-quality audio conferencing : fluency and awareness of attentional focus.* International Journal of Human Computer Studies, 49, 21-58, 1998.

[DOH97] Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S. & Bruce, V.: *Face-to-face interaction and video mediated communication : a comparison of dialogue structure and co-operative task performance.* Journal of Experimental Psychology: Applied, vol. 3, 105-125, 1997.

[FOX93] Fox, B.A. *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems.* Hillsdale NJ: Lawrence Erlbaum, 1993.

[FUS03] Fussell, S.R., Setlock, L. D. & Parker, E.M.: *Where do helpers look? Gaze targets during collaborative physical tasks.* In Proceedings of CHI 03 Extended Abstracts, 768-769. NY: ACM Press, 2003.

[GER04] Gergle, D., Kraut, R.E., & Fussell, S.R.: *Language efficiency and visual technology: minimizing collaborative effort with visual information.* Journal of Language and Social Psychology, vol. 23, 491-517, 2004.

[GLA09] Guéraud, V., Lejeune, A., Adam, J.M., Dubois, M. & Mandran, N. : *Flexible Environment for Supervising Simulation-Based Learning Situations.* AIED 2009, 14th International Conference on Artificial Intelligence in Education, July 2009, Brighton, England, 2009.

[GUC06] Guéraud, V., Cagnat, J.M.: *Automatic semantic activity monitoring of distance learners guided by pedagogical scenarios*, in proceedings of EC-TEL, volume 4227 of LNCS, pp 476-481, 2006

[GUR76] McGurk, H. & MacDonald, J.: *Hearing lips and seeing voices*. Nature, n.264, 746-748, 1976.

[HEA97] Heath, C., Luff, P. & Sellen, A.: *Reconfiguring media space : supporting collaborative work*. In K.E. Finn, A.J. Sellen & S.B. Wilbur (Eds.) Video-mediated communication, 323-349. Lawrence Erlbaum, 1997.

[ISA94] Isaacs, E., & Tang, J.C.: *What video can and cannot do for collaboration.* Multimedia Systems, vol. 2, 63-73, 1994.

[KAR99] Karsenty, L.: *Cooperative work and shared visual context : an empirical study of comprehension problems in side-by-side and remote help dialogues.* Human Computer Interaction, n. 14, 283-315, 1999.

[KRA03] Kraut, R.E., Fussell, S.R., & Siegel, J.: *Visual information as a conversational resource in collaborative physical tasks*. Human Computer Interaction, vol. 18, 13-49, 2003.

[MCC94] McCarthy, J.C., & Monk, A.F.: *Measuring the quality of computer mediated communication.* Behaviour & Information Technology, vol.13, 311-319, 1994.

[NAR97] Nardi, B., Kuchinsky, A., Whittaker, S., Leichner, R. & Schwarz, H.: *Video-as-Data: Technical and social aspects of a collaborative multimedia* *application*. In K.E. Finn, A.J. Sellen & S.B. Wilbur (Eds.) Video-mediated communication, 487-519. Lawrence Erlbaum, 1997.

[O'M96] O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V.: *Comparison of face-to-face and video-mediated communication.* Interacting with computers, vol.8, n.2, 177-192, 1996.

[SEA69] Searle, J.: *Speech Acts, An Essay in the Philosophy of Language*. Cambridge: Cambrige University Press,1969.

[SPE87] Sperber, D., & Wilson, D.: *Relevance: Communication and cognition.* Oxford: Blackwell, 1987.

[TAJ03] Tajariol, F., Adam, J.M., & Dubois, M.: *A study on the effect of communication and monitoring tools in web-based tutoring.* International Journal of Computers and Applications, vol.25, n.3, 206-211, 2003.

[TAJ08] Tajariol, F., Adam, J.M. & Dubois, M.: *Seeing the face and observing the actions: the effects of nonverbal cues on mediated tutoring dialogue.* In B.P. Woolf, E. Aïmeur, R. Nkambou & S. Lajoie (Eds.) Intelligent Tutoring Systems 2008 - 9th International Conference, 480-489. Montreal: Canada. Lecture Notes in Computer Science 5091 : Springer Verlag, 2008.

[TAN93] Tang, J.C. & Isaacs, E.: *Why do users like video video? Studies of multi-media supported collaboration*. Computer supported cooperative work, vol.1 n.3, 163-196, 1993.

[VAN03] Van Lehn, K., Siler, S., Murray, C., & Bagget, W.B.: *Human tutoring : why do only some events cause learning ?* Cognition and Instruction, vol. 21, n.3, 209-284, 2003.

[WAT96] Watts, L.A., Monk, A.F., & Daly-Jones, O.: *Interpersonal awareness and synchronization : assessing the value of communication technologies.* International Journal of Human-Computer Studies, n. 44, 849-875, 1996.

[WHI02] Whittaker, S.: *Theories and methods in mediated communication*. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), Handbook of discourse processes, 123-164. Hillsdale, NJ: Lawrence Erlbaum, 2002.

[WHI93] Whittaker, S., Geelhoed, E. & Robinson, E.: *Shared workspaces: how they work and when are they useful?* International Journal of Man-Machine Studies, vol. 39, 813-842, 1993.

AUTHORS NAMES AND AFFILIATIONS

Provide per each author the following data:

Dr. Federico Tajariol
University of Franche-Comté
Besançon, France
tajariol@gmail.com

Dr. Jean-Michel Adam
Pierre Mendès France University, Laboratory of Informatics Grenoble (LIG)
Grenoble, France
Jean-Michel.Adam@imag.fr

Pr. Michel Dubois
Pierre Mendès France University, Inter-academic Laboratory of Psychology (LIP)
Grenoble, France
Michel.Dubois@upmf-grenoble.fr