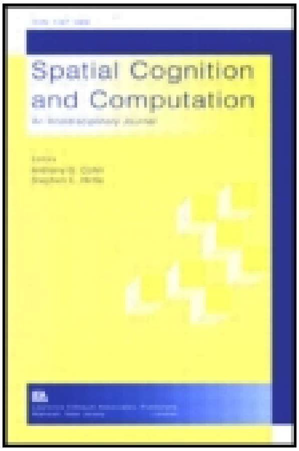


Spatial Cognition & Computation: An Interdisciplinary Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hscc20

Effects of Communication Methods on Communication Patterns and Performance in a Remote Spatial Orientation Task

Wai-Tat Fu [a] [b] , Laura D'Andrea [b] & Sven Bertel [c]

[a] Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[b] Human Factors Division and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[c] Faculty of Media, Bauhaus-Universität Weimar, Weimar, Germany

Accepted author version posted online: 10 Jul 2012.Version of record first published: 08 Mar 2013.

To cite this article: Wai-Tat Fu , Laura D'Andrea & Sven Bertel (2013): Effects of Communication Methods on Communication Patterns and Performance in a Remote Spatial Orientation Task, Spatial Cognition & Computation: An Interdisciplinary Journal, 13:2, 150-180

To link to this article: http://dx.doi.org/10.1080/13875868.2012.701680

Spatial Cognition & Computation, 13:150–180, 2013
Copyright © Taylor & Francis Group, LLC
ISSN: 1387-5868 print/1542-7633 online
DOI: 10.1080/13875868.2012.701680

Effects of Communication Methods on Communication Patterns and Performance in a Remote Spatial Orientation Task

Wai-Tat Fu,[1,2] Laura D'Andrea,[2] and Sven Bertel[3]

[1]Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[2]Human Factors Division and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
[3]Faculty of Media, Bauhaus-Universität Weimar, Weimar, Germany

Abstract: An experiment was conducted to examine the impact of communication methods (text-only, audio-only, and audio-plus-video) on communication patterns and effectiveness in a 2-person remote spatial orientation task. The task required a pair of participants to figure out the cardinal direction of a target object by communicating spatial information and perspectives. Results showed that overall effectiveness in the audio-only condition was better than the text-only and audio-plus-video conditions, and communication patterns were more predictive of errors than individual differences in spatial abilities. Discourse analysis showed that participants in the audio-plus-video condition performed less mental transformation of spatial information when communicating, which led to more interpretation errors by the listener. Participants in the text-only conditions performed less confirmation and made more errors by misreading their own display. Results suggested that speakers in the audio-plus-video condition minimized effort by communicating spatial information based on their own perspective but speakers in the audio-only and text-only conditions would more likely communicate transformed spatial information. Analysis of gestures in the audio-plus-video condition confirmed that iconic gestures tended to co-occur with spatial transformation. Iconic gesture rates were negatively correlated with transformation errors, indicating that iconic gestures more likely co-occurred with successful communication of spatial transformation. Results show that when visual interactive feedback is available, speakers tend to adopt egocentric spatial perspectives to minimize effort in mental transformation and rely on the feedback to ensure that the hearer correctly interprets the information. When visual interactive feedback is not available, speakers will put more effort in transforming spatial information to help the hearer to understand the information. The current result demonstrated that allowing two persons to see and

Correspondence concerning this article should be addressed to Wai-Tat Fu, Department of Computer Science, University of Illinois at Urbana-Champaign, 201 N. Goodwin Avenue, Urbana, IL 61801, USA. E-mail: wfu@illinois.edu

communicate with each other during a remote spatial reasoning task can lead to more errors because of the use of a suboptimal communication strategy.

Keywords: remote spatial communication, communication methods, discourse analysis, gesture analysis

1. INTRODUCTION

Your friend is visiting from out of town but he is lost when driving to your house, so he calls you up on his cell phone. Because your friend does not have a map, he can only describe to you what he sees around him, and because you do not know where he is, you can only guess his location from his description and to locate landmarks that you know based on the map that you have. This kind of a remote spatial orientation task is difficult not only because the two persons are physically separated, but also because they often have different spatial information and perspectives of their surroundings, and thus need to engage in a sequence of information exchanges in order to establish a common spatial representation.

Another reason why this is difficult is that the medium or method of communication often imposes hard constraints on how spatial information can be communicated. In this example, the use of the cell phone demands that spatial information be transformed and packaged into verbal units that are communicated linearly, and the person on the other side has to unpack the transformed information to fit into their own spatial representation in order to make inferences. Other communication methods may impose different constraints that induce different transformation of spatial and nonspatial information during communication. The additional transformation apparently makes the process even more challenging, and may induce different communication patterns and performance in different environments.

Describing and exchanging spatial information is undoubtedly one of the earliest forms of communication: Humans are known to have a long history of developing the ability to communicate to others (by verbal or nonverbal means) where to go for food, and where not to go to avoid danger. In addition to describing the environment based on one's own spatial perspectives, being able to describe an environment from the perspectives of others is an essential ability for effective communication of spatial information. Communication of spatial information is often more challenging than nonspatial information, as the persons not only need to establish common ground on *references* to objects (Clark & Brennan, 1991), but also on spatial frames of reference or *perspectives* to infer the relative locations or directions of objects in the environment (Schober, 1993, 1995; Zacks & Tversky, 2005). Establishing common ground requires two persons to coordinate their communicative (turn-taking, verbal feedback, gestures, etc.) and noncommunicative (reading, reasoning, spatial transformation, etc.) actions to accomplish a common goal.

Research has shown that the coordination of communicative and noncommunicative actions is sensitive to contextual factors such as the relations of objects or task demands (Carlson-Radvansky & Radvansky, 1996; Taylor & Tversky, 1996), differences in spatial abilities (Hegarty & Waller, 2004), or whether the speaker is engaged in a conversation with a real or an imaginary person (Schober, 1993).

Coordination of communicative and noncommunicative actions can occur in many different forms. When two persons with different spatial perspectives are communicating, the speaker may choose to describe the spatial layout based on his or her own egocentric viewpoint and coordinate axes, or they may choose to first transform the information to fit the perspective of the listener before sending the information. In the former case, the listener has to put in extra effort to interpret the untransformed (or partially transformed) information to fit his or her own spatial perspective. Although research has studied the choice of perspective when one is describing the spatial layout without any information about the listener (Levelt, 1982) or when one is describing it to an imaginary or real person at different vantage points (Schober, 1993), there is a lack of research on the choice of perspective when two persons are communicating remotely.

In fact, while research shows that interactive feedback during communication is an important factor influencing the choice of perspectives, these studies have been mostly focusing on face-to-face communication. When communicating spatial information remotely, the methods of communication (e.g., text, audio, or video) often constrain the fidelity of interactive feedback. For example, although video-based methods afford a combination of verbal and nonverbal (such as gestures) feedback, audio-based methods have to rely on verbal only feedback. Texting may further constrain the richness of feedback, as typing not only is slower, but also limited in its ability to express social signals (e.g., approval or understanding). These characteristics of communication methods may therefore induce the use of different strategies in representing, transforming, and communicating spatial information between individuals, and these strategies may lead to differences in communication effectiveness.

The goal of the current article is to investigate how processes and performance of a remote spatial orientation task may be moderated by different communication methods and the complexity of the spatial information. A laboratory study was designed to understand how different communication methods (text-only, audio-only, and audio-plus-video) lead to differences in the processing and communication of spatial information. Given that previous results have shown that different levels of complexity in the relations of objects in a scene influence how two persons communicate and establish common ground, we are also interested in knowing whether spatial information complexity (which is operationalized as the amount of information required to be exchanged between two persons to acquire shared spatial knowledge) impact communication effectiveness, and whether they interact with the effects of communication methods.

On the theory side, our goal is to understand how spatial representations and processing are related to their communication, and how different communication methods may induce different strategies of representing, transforming, and communicating spatial information. On the practical side, the current study will help to understand the impact of different features of communication technologies may have on human spatial information processing, and their effectiveness in facilitating different spatial reasoning tasks.

1.1. Communicating Spatial Information

Communication of spatial information requires the speaker to adopt a *spatial perspective*. A spatial perspective refers to the speaker's physical or imagined point of view when describing locations of objects in a scene. Because the spatial perspectives of the two persons communicating are likely not the same, the linguistic descriptions will likely be different (e.g., a person's right may be another person's left).

A number of studies have shown that linguistic descriptions of spatial information can be influenced by a number of factors, such as functional relations of objects and the context (Andonova et al., 2010). When describing more complex configurations involving multiple objects, both local object-to-object configuration and global perspectives of the configurations may impact the choice of description (Tenbrink et al., 2011). The choice of spatial perspectives during communication therefore needs to be coordinated such that the pair can either adopt the perspective of one of the pair, or some common perspective that they both agree on.

Relatively few studies have investigated the choice of spatial perspectives when two persons communicated spatial information. One exception is the study by Schober (1993), who studied how speakers set spatial perspectives differently with imaginary persons than with actual conversational partners. In the study, one participant, the director, described which of the two circles in a display had an X on it for another participant, the matcher. In the pair condition, the matcher was another participant; but in the solo condition, the matcher was an imaginary person. The matcher's task was to mark the circle on his display (which could be rotated compared to the director's display) that corresponded to the marked circle in the director's display.

Schober found that while almost all participants in the solo condition chose the nonegocentric perspective (i.e., they never adopted a spatial perspective based on their own vantage points), choices of spatial perspectives varied widely for participants in the pair condition. Further analysis showed that when the first speaker in a pair chose an egocentric perspective, the partner tended to also use an egocentric perspective. In fact, Schober found a very high correlation ($r = 0.94$) in the proportion of use of an egocentric perspective between directors and matchers. The results suggested that when a speaker describes locations of objects to an imaginary person, he or she tends to transform the spatial information to the frame of reference of the imaginary hearer. On the other hand, when two persons are communicating face-to-face, they will more likely describe *untransformed* spatial information to each other, expecting the other person to interpret the untransformed information in their own spatial perspective.

One intuitive explanation for the difference found by Schober (1993) is that feedback from a real person helps communication. Schober and Clark (1989) showed that collaborative feedback could enhance understanding even after controlling for the amount of information communicated. The idea is that, when engaged in actual conversation, hearers can express difficulties in understanding and request clarification until the speakers and hearers believe that they both understand each other. It is therefore possible that speakers in Schober 's (1993) study chose to communicate untransformed spatial information to minimize mental effort and relied on the interactive feedback to ensure that their information was correctly interpreted. It is, however, also possible that the mere co-presence of the pair could induce some forms of social influence that changed their communication patterns.

To minimize social effects induced by co-presence, the current study focused on how two persons communicate remotely. In a remote spatial orientation task, the speaker and listener not only have different vantage points, but they are also physically separated. The communication media may constrain what kind of interactive feedback is available. For example, while a video call can afford both verbal and nonverbal (e.g., gestures, nodding, frowning, etc.) feedback, a regular phone call (audio-only) can primarily afford only verbal feedback. When people are texting each other, verbal feedback are further limited to writing. Comparing spatial communication across these media in remote spatial communication can therefore help to delineate the effects of different forms of interactive feedback on choices of spatial perspectives and communication performance. We will elaborate on the characteristics of these communication methods next.

1.2. Effects of Communication Methods

There has been much research comparing the effects of communication methods on structures of communication and performance (Chapanis, Ochsman, Parrish and Weeks, 1977; O'Conaill, Whittaker & Wilbur, 1993; Green & Williges, 1995; Olson, Olson & Meader, 1995; O'Malley et al., 1996; Doherty-Sneddon et al., 1997; Daly-Jones, Monk, & Watts, 1998; Veinott, Olson, Olson, & Fu, 1999). For example, because it takes more time to type words than to say them, communication by text-based method is often less efficient compared to audio-based method (Olson & Olson, 2000). In addition, the conversation dynamics, such as how and what information is communicated, may be different because of the different time costs and effort imposed by the methods.

For example, because it takes longer to type text, conversations tend to be more condensed in text-based than audio-based communication, in the sense that there tends to be more information contained in a single sentence when people are typing their communication (Siegel, Dubrovsky, Kiesler, & McGuire, 1986; Straus, 1997). However, this could also be an artifact of the fact that in text-based communication, histories of communication were often retained and can be referred to, but these histories were not retained in audio-based communication. Thus, speakers might communicate less information in a single sentence in audio-based communication, as they wanted to make sure that the information was received correctly. In fact, in audio-based communication, there are often more instances of expression of confirmation of receipt of information or clarification requests than in text-based communication (Straus & McGrath, 1994).

One major difference between audio-based and video-based communication is that speakers can communicate using social cues such as facial expression and gestures in video-based method (such as by video conferencing) but not in audio-based method (such as by a regular phone). For example, comparing communication between pairs who can see each other through video conferencing and those using audio-only conferencing, the audio-only pairs tend to use extra utterances to elicit verbal feedback than the video pairs (O'Malley et al., 1996; Doherty-Sneddon et al., 1997). The lack of visual feedback (e.g., facial expressions, gestures, etc.) apparently leads the speaker to put in extra effort to ensure that the hearer correctly interprets the information, as well as to compensate for the lack of nonverbal communication cues that allow them to communicate affective states such as doubts (e.g., frowning) or approval (e.g., nodding). These could perhaps explain why studies have found that video-based communication was sometimes better for social tasks such as negotiation than audio-based communication (Veinott, Olson, Olson, & Fu, 1999).

To further understand why video-based communication was better, a recent study by Dong and Fu (2012) compared how negotiation was conducted in an appointment-scheduling task when pairs of participants communicated with video-plus-audio, audio-only, and text-only media. Similar to earlier studies, negotiation outcomes were generally better (measured based on the final solutions agreed by the pair) in the video-plus-audio condition than the other two conditions. Meditational analysis showed that communication patterns mediated the impact of communication media on the negotiation outcomes. Specifically, Dong and Fu found that participants in the video-plus-audio condition tended to exchange less information in each utterance, and relied more on interactive feedback to reach a common solution.

The results suggest that the richer fidelity of interactive feedback in video-based communication may induce the speaker to minimize effort in their description and rely more on feedback to verify that the hearer understands the description correctly. The results are in general consistent with those by Schober (1993), who showed that participants in pairs would more likely use egocentric perspectives when exchanging spatial information. However, given that communication of spatial information is inherently differently from the time information in the appointment-scheduling task in Dong and Fu, it is possible that the communication patterns and outcomes will be different.

Among the different types of nonverbal feedback, gestures are found to be important especially when they reflect useful cognitive representations of the objects or the environment layouts (Alibali, 2005; Goldin-Meadow & Wagner, 2005; Hostetter, 2011). Gestures are more likely used when speakers are expressing spatial information, such as when they are providing route directions (Allen, 2003) or performing spatial transformations (Trafton et al., 2006).

Studies have shown that many gestures are intended to be communicative, as demonstrated by findings that show that communication effectiveness is often better when the speakers can use gestures than when they cannot use gestures to communicate (Graham & Argyle, 1975), and the finding that speakers tend to gesture more when they know that others can see them (Krauss, 1998; Alibali, Heath, & Myers, 2001). Similarly, using a picture description task, Melinger and Levelt (2004) showed that speakers (who could see each other) who produced gestures representing spatial relations omitted more required spatial information from their descriptions than speakers who did not gesture. The results suggest that speakers often express or “offload” part of their message via the manual modality (through gestures), which supports the notion that gestures are communicative.

Trafton et al. (2005) and his colleagues also found that speakers are differentially likely to use gestures when expressing different types of spatial information. Specifically, they found that speakers were more likely to gesture when expressing spatial transformation (e.g., mental rotation of externally presented images) than expressing geometric relations (spatial relations and locations) and spatial magnitude. It was also found that when speakers could not use gestures to communicate, the proportion of words used to denote spatial information was increased (Graham & Argyle, 1976), suggesting that information that originally could be communicated through gestures could be transformed into verbal units. Thus, it is possible that the conversations between the speakers could also be different when communicating in audio-based (in which they know their gestures will not be seen) than video-based method.

1.3. Individual Differences in Spatial Ability

Research has found that individuals differ in their ability to perform spatial transformations, and spatial ability predicts performance well in many domains (Hegarty & Waller, 2004; Keehner et al., 2006). Two common kinds of spatial abilities are the ability to mentally rotate objects and the ability to imagine the appearance of objects from different perspectives (orientations) of the observer (Hegarty & Waller, 2004). Given that mental rotation and perspective taking are predictive of spatial task performance (Hegarty & Waller, 2004), one may expect that these spatial abilities will also predict performance in the current remote spatial reasoning task.

For example, people who are better at performing spatial transformation may be better at communicating transformed information to better fit the perspective of the other person, thus facilitating communication efficiency and performance. However, it is also possible that effects of communication methods may be more critical in determining communication patterns, or may interact with spatial ability of the persons to influence communication and performance. The spatial ability of participants was therefore measured to study its role in influencing communication and performance in the current task.

2. THE PRESENT STUDY

Based on previous studies (e.g., Schober, 1993; Veinott et al., 1999; Dong & Fu, 2012), it is expected that video-based communication may induce speakers to minimize efforts when describing spatial information, and relied more on interactive feedback to establish common grounds during communication. It is also expected that the use of gestures, one important type of nonverbal visual feedback, may moderate the extent to which the communication is conducted.

For example, it is expected that gestures that are believed to help communication can reduce errors. In contrast, in audio-based or text-based communication, the lack of visual interactive feedback may encourage speakers to perform more transformation of spatial information to increase the chance that the information will be correctly interpreted. This is consistent with findings in Schober's (1993) study, in which participants describing spatial information to an imaginary person tended to transform information to fit the imaginary person's spatial perspectives. Presumably, the lack of interactive feedback from the imaginary person induced the speaker to put in extra effort to make the description easier for the hearer to understand. It is also expected that this effect will be stronger when the configuration of the objects is more complex, as complexity will increase the chance that errors will be made in the description and interpretation of information.

The current study focused on the impact of three kinds of communication methods: text-only (Text), audio-only (Audio), and audio-plus-video (Video), on communications of spatial information with different levels of complexity, and how they eventually influence performance. To this end, a remote spatial orientation task was designed in which participants had to communicate and exchange spatial information to (1) identify the target, and (2) infer the cardinal direction of the target (details below).

The task was designed such that only one participant would know the relative location of the target with respect to other objects on the display, but did not know the cardinal directions of the objects (i.e., similar to the situation when one can see objects around them but does not have a map); while the other participant would know the cardinal directions of a subset of the objects, but did not know which one was the target (e.g., similar to the situation when one is giving direction to another person by looking at a map, but does not know where the person is).

Another important feature was that participants had different spatial perspectives and asymmetric information (e.g., it was expected that the participant with the cardinal direction was able to adopt an extrinsic perspective, but the other participant without cardinal directions could only adopt an intrinsic perspective), and they had to communicate and eventually come up with some forms of a shared representation to infer the cardinal direction of the target. Given that the task involved interleaving of spatial transformations and exchange of information based on different spatial perspectives, It was expected that communication methods would influence communication structures, which would in turn influence performance. The task would therefore allow us to unpack the interactions between the communication process and performance in the remote spatial orientation task.

2.1. Participants

Participants were 48 adults recruited from a U.S. university community, and they were paid for their participation. Participants were randomly assigned to a pair who did not know each other before they participated in the experiment to control for any previous communication experiences between the pair. Due to technical difficulties, data from one pair of participants (Audio condition) were lost. Of the remaining 46 adults (mean age = 24.6; mean years of education = 15.7), 29 were female and 17 were male. Participants were screened for normal or corrected-to-normal vision.

2.2. Spatial Ability Measures

Spatial abilities were measured with two paper-pencil tasks. Participants' ability to mentally rotate objects was measured with the Mental Rotation Test (MRT) (Vandenberg & Kuse, 1978). Perspective taking ability (i.e., the ability to imagine how a scene looks from a different location in space) was assessed with the Perspective Taking/Spatial Orientation Task (PTSOT) (Kozhevnikov, Hegarty, & Mayer, 2002; Hegarty & Waller, 2004).

2.3. The Remote Spatial Orientation Task

In the task, two separate displays of information were presented to the participants. One participant was shown the *Responder's* display, and the

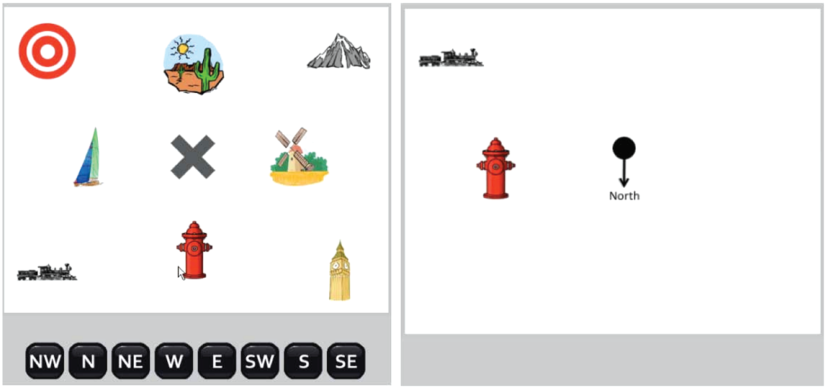


Figure 1. Sample displays for the Responder (left) and the Instructor (right) in the "simple" condition of the spatial orientation task. The target is represented by the red concentric circles in the display of the Responder, who has to report the cardinal direction of the target by clicking on one of the direction buttons at the bottom of the screen. (in this example the correct response is SW). In the "complex" condition, some object icons would appear in multiple locations in the Responder's display. Objects were always presented upright (color figure available online).

other was shown the *Instructor's* display (see Figure 1). The Responder was presented with a 2D array of seven image icons and one target icon (indicated by 2 red concentric circles), all located in one of the eight cardinal directions (North, South, East, West, Northeast, Northwest, Southeast, Southwest). The Responder was responsible for reporting the cardinal direction in which the target was located, relative to the position on his/her own display (indicated by an X in the center of the screen). The Instructor was given a display showing two of the seven image icons seen by the Responder, as well as an arrow indicating the North direction as seen in maps (see Figure 1).

In both the Responder's or Instructor's display, the icons maintained relative spatial relationships that were identical with the other icons. However, the entire configuration of icons was rotated to some degree (in 90° increments), such that the Instructor and Responder's displays were not "pointing" in the same direction. This was created to mimic common situation in which two persons are facing different directions and need to reconcile their spatial orientation. Stimuli of two levels of complexity (simple, complex) were displayed. In the simple condition, each icon on the Responder's display was unique. In the complex condition, the Responder's display contained identical icons that appeared in multiple locations, such that just "naming" an icon was not sufficient to uniquely identify the target. The complex condition therefore demanded more communication of the relative locations and directions of the icons to identify the target icon.

2.4. Procedure

Pairs were randomly assigned to one of the three communication methods. Within each pair, individuals were randomly assigned to their roles (Responder vs Instructor). The orientation task was conducted such that they were seated at computers separated by enough distance and barrier so that participants were out of each other's sight and hearing range during the task. Before performing the remote spatial orientation task, individuals completed the set of demographic and spatial ability tests. They were then shown to their computers and instructed on how to use the particular communication method (i.e., text messaging, audio, or video chat).

On each computer, an instruction screen was presented, which explained the tasks and offered a sample display representative of what each individual would view, depending upon their role (Responder vs. Instructor). Participants were told that their partner's display might not be the same as theirs, but they were not told what exactly their partner saw (i.e., they did not know that the maps was rotated). There was no particular rule of engagement, and they were allowed to say anything during the experiment. The pair then performed one practice task, during which the experimenter would answer any questions that they had. Participants then completed a set of 20 tasks, 10 complex and 10 simple, with their order randomized (simple and complex tasks were not blocked). In each condition, the task workspace took up half of the computer screen; the other half contained the communication tool (Text condition: an IM chat box; Video: Skype video chat interface (see Figure 2); Audio: Skype audio chat interface). Accuracy of task performance and conversations between pair members were recorded.

Across all pairs, the practice trial was identical. However, the icons in each of the 20 experimental trials were randomly generated for each pair. This randomization included: the 7 icons that appeared on the Responder's display (randomly selected from a master set of 18 icons), the target's location on the Responder's display, the direction of North on the Instructor's display, the relative locations of the two icons appearing on both the Instructor's and Responder's displays, the degree of rotational disparity between the Instructor's and Responder's displays (in 90° increments), and the order of the complex/simple conditions (10 trials for each condition) throughout the experiment.

2.5. Equipment and Software

Stimuli were presented on each participant's computer screen; responses were all made using the mouse. The software *Camtasia* was used to record audio and video feeds in the Audio and Video conditions. In the Text condition, participants communicated by typing to each other using the *AOL Instant Messenger*, which retained all histories of text in a single window throughout

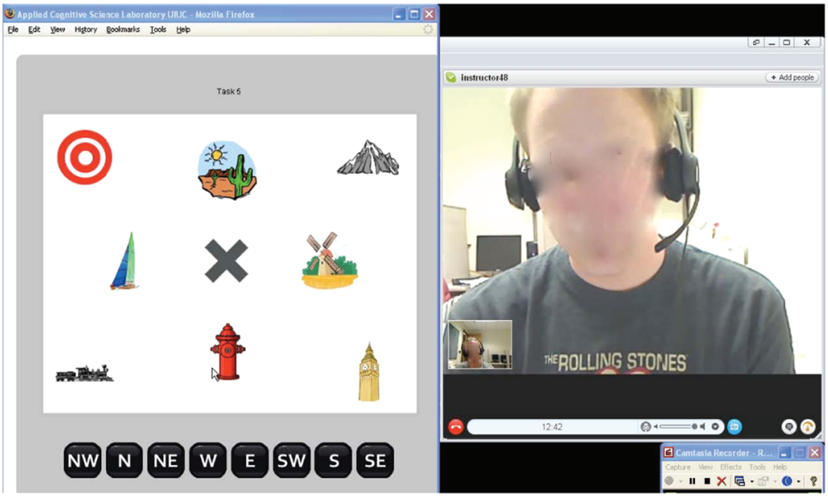

Applied Cognitive Science Laboratory UIUC - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Task 5
NW N NE W E SW S SE
instructor48
+ Add people
THE ROLLING STONES
12:42
Camtasia Recorder - R...
Capture View Effects Tools Help

Figure 2. Screenshot from video condition (Responder's computer screen). The left side of the display showed the icons, and the right side showed the video of the other participant. The face was not blurred in the actual experiment. In the Text condition, the right would show the texting window; in the audio condition, the right side of the screen was blank (color figure available online).

the experiment. In the Audio condition, participants wore high-fidelity headphones and microphones when performing the tasks. Auditory communication was enabled through the use of Skype's auditory calling feature. In the Video condition, participants wore the same headphones and microphones when performing the task. Video chat communication was enabled through the use of Skype's video chat feature. Each computer was supplemented with the Logitech Webcam Pro 9000 cameras mounted on the top of the monitors in order to permit video chatting.

3. RESULTS

3.1. Performance in Reporting Correct Cardinal Direction

A two-way ANOVA examining the effect of communication methods (Video, Audio, Text) and task difficulty (simple, complex) on the mean percentage correct in the reported cardinal direction of the target objects (referred to as accuracies in the rest of the article) was conducted. There was a significant main effect of communication methods ($F(2, 40) = 5.14, p < 0.05$) and task complexity ($F(1, 40) = 3.99, p < 0.05$) on accuracies. The methods by complexity interaction was not significant ($F(2, 40) = 1.18, p = 0.32$) (see

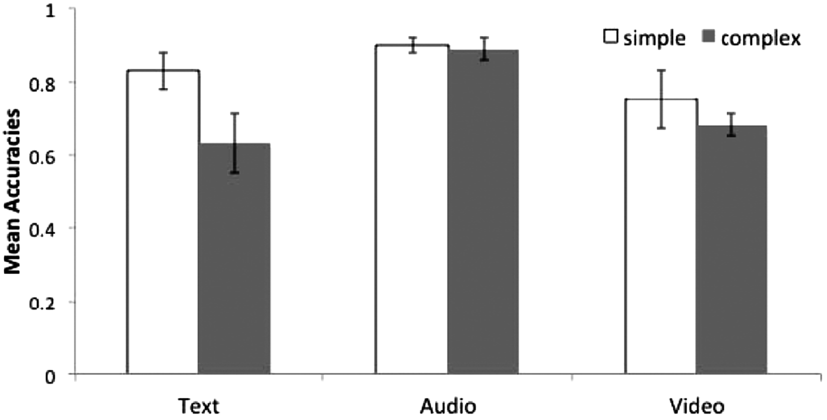


Figure 3. The mean percentage correct in reporting the cardinal direction of the target object (mean accuracies) in each of the three communication methods and in each type of problem. Error bars represent standard errors.

Figure 3). All mean accuracies were significantly above chance (i.e., 1 out of 8) ($p < 0.01$).

Simple planned comparisons revealed that pairs in the Audio condition outperformed those in the Text condition ($t(6) = 2.61, p < 0.05$) and the Video condition ($t(6) = 3.22, p < 0.05$). However, performance in the Text condition was not different from that in the Video condition ($t(7) = 0.08, p = 0.94$). This demonstrated that pairs in the Audio condition performed the tasks significantly better than pairs in either the Video or Text conditions.

The pattern of results on task performance showed that the Audio condition was significantly better than the other conditions. Consistent with previous research (Olson & Olson, 2000), performance in the Audio condition was better than the Text condition. However, it was also found that performance in the Audio condition was better than the Video condition, which was inconsistent with the notion that communication would be less effective when the pair could not communicate through facial expression or gestures.

To better understand this pattern of results, a series of additional analysis was conducted to understand (1) the error patterns in each communication methods, (2) how different communication methods induced different communication patterns, (3) how individual spatial abilities and communication patterns differentially influence performance, and (4) how gestures in the Video conditions were related to communication patterns and performance. These analyses were then summarized to conclude why performance in the Audio condition was better than the Video and Text conditions, and its implication to communication of spatial information in general.

3.2. Error Analysis

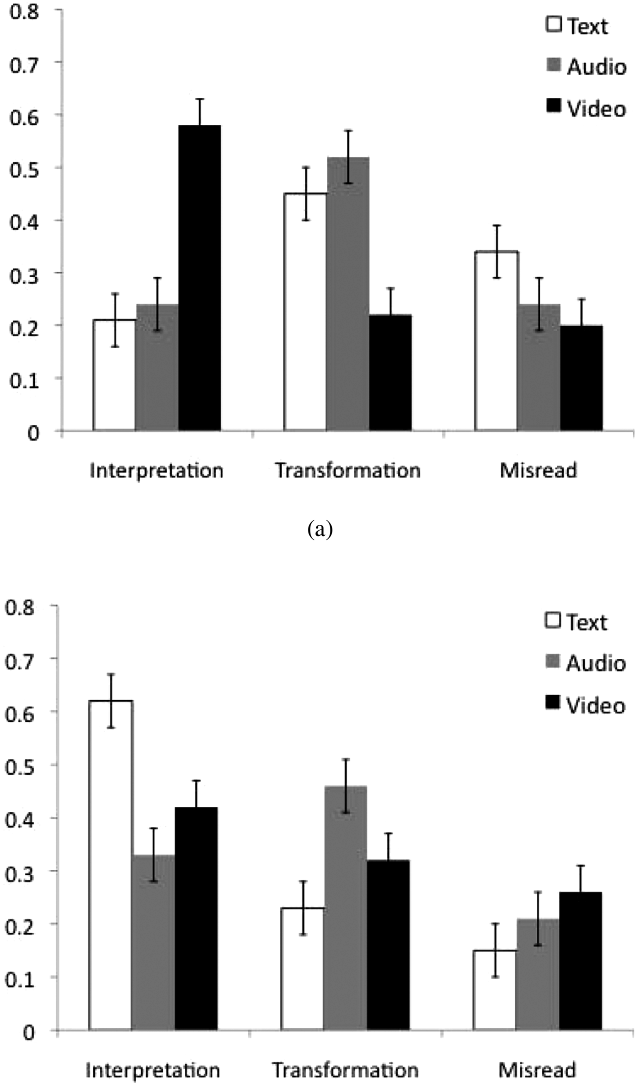
All error trials were extracted from each condition and the communications between participants were inspected to understand how and where errors occurred. All the correct trials were analyzed to identify errors that were made during the trial but were later corrected to examine whether there were differences in detecting and correcting errors between communication methods. The *uncorrected* and *corrected* errors were then compared to understand the differences between the communication methods to understand how performance was impacted by different communication patterns.

Given the current interest in how the different communication media induced differences in the choice of spatial perspectives when describing the objects, the focus was on errors related to the how the hearer interpreted the information and how the speaker described spatial information. These led to two broad categories of errors: Interpretation and Description. The description errors were further divided into whether they occurred when the speaker transformed the information (transformation errors) and when the speaker described the information using their own egocentric perspective. This categorization scheme led to three exhaustive categories of errors in both the error and correct trials: *interpretation*, *transformation*, and *misread*. The categories were general in the sense that they referred to the processes of description and interpretation, which are fundamental in communication.

Interpretation errors were mistakes made by the participant who received a message from their partner but misinterpreted the content of the received message. For example, in Figure 1, the instructor said, "the hydrant is north of the train," the responder said "ok, so the train is north, and the hydrant is south." A transformation error occurred when the participant attempted to either mentally rotate or take a particular perspective to figure out the direction of an object on the display, but the transformation was wrong. For example, in Figure 1, the responder said, "if the hydrant is north of the train then the yacht should be east of the train." A misread error occurred when the participants extracted wrong information from their own display. For example, Figure 1, the instructor said, "the train is north of the hydrant."

During coding, the same set of examples described above (together with Figure 1) were given to the coder to explain the categories. We recruited coders who were naïve to the purpose of the experiment. One coder coded all errors, a second coder randomly coded about 50% of the errors. Intercoder agreement was 98% ($\kappa = 0.89$, $p < 0.01$). All errors could be classified into these three categories. There was no significant interaction involving the role (instructor or respondent) on error types. Roles were therefore not included as a factor in subsequent error analysis.

Figure 4a shows the mean proportions of uncorrected errors per participant pair in each communication condition. Over 99% of these errors (either a single or a combination of them) caused the choice of the wrong final response (in a small number of rare cases participants chose the correct response even



(a)

(b)

Figure 4. Mean proportion of (a) uncorrected and (b) corrected errors per participant pair in each communication condition. Error bars represent standard errors.

though the errors were uncorrected). Similar to the analysis of accuracies, 2-way ANOVA on the uncorrected errors showed that all interactions involving complexity were not significant. Thus, simple and complex trials were collapsed in the error analysis. The main effects of communication methods and error types were not significant, but their interaction was significant ($F(4, 40) = 12.32, p < 0.05$).

Tukey's post-hoc tests showed that participants in the Video condition had significantly more interpretation errors than the Text and Audio conditions ($p < 0.05$), those in the Text and Audio conditions had significantly more transformation errors than the Video conditions ($p < 0.05$), and those in the Text condition had significantly more misread errors than the Audio and Video conditions ($p < 0.05$). No other difference was significant. Results showed that the distribution of errors were significantly different across the communication conditions.

Figure 4b shows the mean proportion of corrected errors in each condition. Corrected errors were errors that were corrected before the trial ended by one of the participants. 2-way ANOVA on the corrected errors showed that the main effect of communication methods was not significant. However the main effect of error types was significant ($F(2, 40) = 9.92, p < 0.05$), as well as the interaction between communication methods and error types ($F(4, 40) = 11.12, p < 0.05$). Post-hoc Tukey tests showed that there were significantly more correction of interpretation errors than correction of transformation and misread errors ($p < 0.05$).

In addition, there were significantly more correction of interpretation errors in the Text condition than the Audio and Video condition ($p < 0.05$), and more correction of transformation errors in Audio and Text conditions than the Video condition ($p < 0.05$). No other difference was significant at the $p < .05$ level. Results showed that while participants in the Text condition failed to correct errors caused by misread, a large proportion of the corrected errors were interpretation errors.

The overall pattern of results from the error analysis showed significantly different error patterns across the communication methods. Because it is easier to understand the error patterns by comparing them to communication patterns, the next section will first introduce the discourse analysis that showed how communication patterns differed in each communication method.

3.3. Discourse Analysis

To further unpack how the different communication methods would induce different communication patterns, the text and verbal communications between each pair were transcribed and coded. A coding scheme was developed, which addressed six main categories of utterance types: (1) Untransformed object description, (2) Transformed object description, (3) Untransformed revision/repair, (4) Transformed revision/repair, (5) Request for confirmation, and (6) Request for elaboration. These types are rooted in Clark's work (e.g., Clark & Brennan, 1991) regarding the ways in which pairs of individuals, during conversation, collaborate to reach common ground.

Their categories were refined to highlight how transformed and untransformed information was exchanged, and how often they requested transformed and untransformed information from their partner. Table 1 shows definitions and examples of the categories of utterances. Two independent coders who are naïve to the task categorized the transcribed text and recordings according to the coding scheme with examples as shown in Table 1. One coded all text and the other randomly coded 20% of the transcribed text, and they reached 92% agreement ($\kappa = 0.84, p < 0.01$). The six categories accounted for over 95% of all utterances in all conditions.

The current interest was in whether there were differences across communication conditions in their use of the main utterance types, as well as whether the use of different main utterance types were related to task performance. The mean number of response categories in each condition is shown in Figure 5. A 2-way MANOVA was first performed (3 communication type × 2 difficulty levels) with the 6 utterance types as dependent variables. There were significant main effects of communication types, difficulty levels, and their interaction ($\Lambda = 1.21, p < 0.01; \Lambda = 0.81, p < 0.01; \Lambda = 1.51, p < 0.01$).

To understand their effects on each of the dependent variables, separate ANOVAs was performed on each type of utterances. Results showed that there were significant main effects of communication methods and difficulty levels, as well as significant interactions between communication methods and difficulty levels in each utterance types.

3.3.1. Simple Tasks: Tukey post-hoc tests showed that in the simple tasks, there were significantly more untransformed object description and fewer transformed object description in the Video condition than the Text and Audio conditions ($p < 0.05$). The Audio condition also had more confirmation than the Text and Video conditions ($F(1, 54) = 5.21$, $F(1, 54) = 5.42$; $p < 0.05$). The results were in general supportive of the expectation that speakers in the Video condition would minimize effort in transforming spatial information, while the lack of visual feedback in the Audio condition would induce more verbal confirmation. The higher cost of confirmation in the Text condition also led to less confirmation.

3.3.2. Complex Tasks: For complex tasks, Tukey post-hoc analysis showed that the Video condition had significantly more untransformed object description than the Text and Audio conditions ($p < 0.05$), but it had significantly less transformed object description than the Text and Audio conditions ($p < 0.05$). The same pattern was found for untransformed and transformed revisions, as the Video condition had more untransformed revisions and less transformed revisions than the Text and Audio conditions ($p < 0.05$).

Table 1. Examples of the five categories of utterances in the discourse analysis (see text for details)

Categories	Definitions	Examples
Untransformed object description	Described objects using the same perspective as presented on the participant's own screen. For responders, objects are described based on egocentric terms (e.g., up, down, left, right); for instructors, objects are described based on either egocentric or untransformed cardinal directions.	"The hydrant is below the train on my screen"
Transformed object description	Described objects after some transformation (e.g., mental rotation or perspective taking) of the spatial representation that was different from that on the speaker's own screen.	"When standing on the train facing the hydrant, the target is on the left"
Untransformed revisions	Referred to previous utterances, and attempted to make simple revisions to them based on untransformed representations.	"No, not left, I said right."
Transformed revisions	Referred to previous utterances, and attempted to make revisions based on transformed representations that were different from that shown on their own screen.	"You said the target's above the train on your screen, so it's not east"
Request for confirmation	Requested their partner to confirm and/or verify the content of information without any transformation.	"You said the left, correct?"
Request for elaboration	Requested their partner to provide more information that involves an elaboration or transformation of the information.	"You said the hydrant is below the train, is that north or south?"

There was also significantly more confirmation and elaboration in the Audio condition than the Text and Video conditions ($p < 0.05$) in complex tasks.

3.3.3. Summary of Discourse Analysis: The patterns of results showed that different communication methods did induce different communication structures. Participants in the Video condition exchanged more untransformed

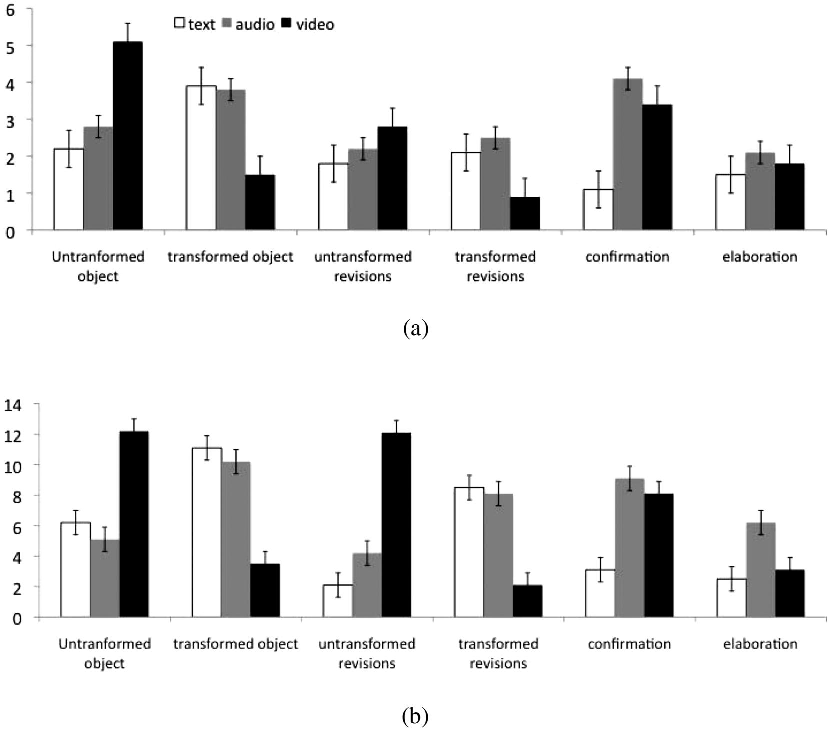


Figure 5. Categories of utterances in each communication conditions in (a) simple and (b) complex tasks.

information to perform the task, while the Text and Audio conditions exchanged more transformed information to finish the task. This pattern was found for both simple and complex tasks. In addition, participants in the Audio condition also performed more confirmation than other conditions in both simple and complex tasks. There were also more untransformed revisions and fewer transformed revisions in the Video condition in the complex tasks, although these differences were not significant in the simple tasks. This was likely due to the fact that, in simple tasks, there were fewer revisions or repairs needed to locate the target.

Results indicated that in the Video condition, participants tended to exchange more untransformed information, while those in the Audio and Text conditions tended to exchange more transformed information to perform the orientation task. The results were in general consistent with previous research that shows that communication tends to proceed until it is perceived that there is sufficient common ground established between the speakers (Clark & Brennan, 1991), and the finding that speakers in face-to-face communication tended to communicate more untransformed spatial information (Schober, 1993).

It is interesting to compare results from the current study with those from Schober (1993). Schober showed that participants in the pair conditions performed less mental transformation than those in the solo condition, in which participants were instructed to communicate to an imaginary partner. The difference could be attributed to the availability of verbal interactive feedback in the pair condition. Similar to the pair condition, participants in the Audio condition of the current study could also only provide verbal interactive feedback to their partners. Interestingly, we found that when participants could provide both visual and verbal interactive feedback in the Video condition, they performed less mental transformation than those in the Audio condition.

One way to interpret these results is that the availability of interactive feedback influenced the perception of the shared task environment by the pair, which influenced the structure of the *joint activity* (Clark, 1996) of the pair as they coordinate their communicative (turn-taking, verbal feedback, gestures, and other visual feedback) and noncommunicative (spatial transformation, reasoning, etc.) actions. In particular, the availability of the richer form of interactive feedback (richer in the sense that it is closer to the kind of feedback one receives in face-to-face communication) induces the pair to perform less mental transformation, presumably because they perceive that they can rely on the interactive feedback to establish common ground (i.e., their shared beliefs about their shared knowledge and representations of the objects on their screens). In other words, the visual and verbal interactive feedback induces a shift to rely more on communicative than noncommunicative actions. The current results suggest that this shift led to more errors in the Video condition.

3.3.4. Explaining the Error Patterns: In the error analysis, there were significantly more interpretation errors in the Video condition, more transformation errors in the Text and Audio conditions, and more misread errors in the Text condition (see Figure 4). The discourse analysis showed that the higher interpretation errors could be attributed to the fact that participants in the Video condition were more inclined to communicate untransformed information, which apparently led to more interpretation errors. In contrast, participants in the Audio and Text conditions tended to communicate more transformed information, which led to more transformation errors.

The results suggested that one possible reason why accuracies in the Video condition were worse than the Audio condition was that communication between pairs were suboptimal in term of establishing common ground in the Video condition, in the sense that speakers were overconfident that their untransformed descriptions were interpreted correctly by the listeners. Finally, participants in the Text condition were found to do less confirmation, which could be one reason why there were more misread errors. Additional analysis is presented on the relations between errors and gestures in the video condition in Section 3.5.

3.4. Spatial Abilities, Communication Patterns, and Accuracies

Regression analysis was performed to further investigate the extent to which accuracies was related to the different communication patterns in each condition and the spatial abilities of the participants. To simplify the analysis, the current focus was on two major kinds of communications: the *untransformed* (untransformed object description and untransformed revisions) and *transformed* (transformed object description and transformed revisions) utterances. The sum of the mean scores of the mental rotation test (MRT) and the perspective-taking test (PTSOT) of each pair of participants were calculated as a measure of their *collective spatial abilities* (individual scores, differences, minimum, maximum and other combinations of these were calculated and similar results were found). ANOVA showed no significant difference among the communication conditions on this measure of collective spatial abilities, confirming the random assignment of participants to each condition.

There was no significant correlation between collective spatial abilities and communication patterns. The results were different from the study by Schober (2009), who found that differences of spatial abilities were more predictive of performance. This was possibly due to the differences in the tasks, or because the current study we did not preselect participants and assigned them to different pairs as in Schober (2009). In fact, in the study by Schober, participants were selected such that spatial abilities at the extremes (very high or very low) were paired, and thus the statistical power was larger than the current study.

The analysis first tested how communication patterns and collective spatial abilities contributed to accuracies. Stepwise regression analysis was performed on the relative effects of communication patterns and spatial abilities on accuracies. Two regression models were tested using a pair of participants as the unit of analysis. Given the interest in the relative contributions of spatial abilities and communication, the number of transformed object and transformed revisions utterances were combined to create the variable "transformed utterances," and the number of untransformed object and untransformed revisions were combined to create the variable "untransformed utterances." In the first model, communication patterns (transformed and untransformed utterances) was entered into the model in the first step, then entered spatial abilities. In the second model, the order of entry was reversed. The idea was to test whether the change in regression weights (and R^2) was significant in each step to infer the extent to which accuracies were related to either variable.

Table 2 shows the regression weights and change in R^2 as each variable was entered in each model. In the first model, when communication patterns were entered, the change in R^2 was significant in all conditions in both simple and complex tasks; but when spatial abilities was entered, the change in R^2 was not significant. In the second model, when spatial abilities was entered

Table 2. Results from the regression analysis

		Simple tasks			Complex tasks		
		Text	Audio	Video	Text	Audio	Video
First model	First	_0.56_*	_0.48_*	_0.62_*	_0.63_*	_0.66_*	_0.69_*
	Untransformed utterances	–0.34*	–0.11	–0.42*	–0.38*	–0.16	–0.59*
	Transformed utterances	–0.19	–0.26*	–0.12	–0.56*	–0.69*	–0.24
	Second	_0.12_	_0.16_	_0.13_	_0.23_	_0.24_	_0.22_
	Spatial abilities	0.21	0.23	0.1	0.31	0.38	0.24
Second model	First	_0.36_	_0.24_	_0.18_	_0.46_*	_0.56_*	_0.39_*
	Spatial abilities	0.22	0.21	0.16	0.59*	0.71*	0.39*
	Second	_0.32_*	_0.40_*	_0.57_*	_0.40_*	_0.34_*	_0.52_*
	Untransformed utterances	–0.21*	–0.12	–0.36*	–0.21*	–0.12	–0.48*
	Transformed utterances	–0.25	–0.24*	–0.12	–0.42*	–0.42*	–0.16

Underlined numbers represent change in R^2, other numbers represent the standard regression weight when that variable was entered into the model. * represents values that are statistically significant at $p < 0.05$. (Note that similar results were obtained when spatial ability was replaced by MRT, PTSOT, maximum and/or minimum of MRT and PTSOT, as well as the difference between the measures.)

first, the change in R^2 was significant only for complex tasks; but when the two types of utterances were entered, the change in R^2 was significant in all conditions in both tasks. The results suggested that accuracies depended on the communication patterns more than spatial abilities.

In both models, the regression weights for untransformed utterances were significant in the Text and Video conditions, and those for transformed utterances were significant in the Audio condition. This was consistent with previous results that showed that participants in the Audio condition performed more mental transformation when exchanging and confirming information with each other. Interestingly, in the Text condition, only in complex tasks was the number of transformed utterances related to performance.

Given that in the complex tasks more transformation was needed, spatial abilities were also found to be more strongly associated with performance than simple tasks. The overall pattern of results confirmed that different communication methods induced different communication structures, which was more critical for performance than the measure of spatial abilities of the group. In other words, in contrast to individual spatial reasoning tasks, performance in the current remote spatial orientation task depended more on communication than the collective spatial abilities of the pair. On the other hand, it was also possible that the current stimuli (the use of landmarks) might have induced the use of spatial reasoning in a somewhat larger scale of frame-of-reference, which might have contributed to the lower predictive power of the tests of spatial abilities (mental rotation and perspective-taking) on performance (see Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006).

3.5. Gestures in Video Communication

To further understand how different communication methods induced different communication patterns, all gestures from the participants in the Video condition were extracted and analyzed, based on the coding scheme from Trafton et al. (2005) for the most part (which was based on McNeil, 1992). All personal adjustments (e.g., scratching) were first eliminated and gesture rates per minute for each type of gesture were then calculated for all participants.

Following Trafton et al., gestures were first coded as *simple* or *complex*. Simple gestures were brief and motoric, and are further divided and coded as *beats* and *deictic*. Beats included rhythmic gestures such as hand flicks and waves, and deictic gestures included pointing to an item on the display. Complex gestures were divided into *iconic* or *noniconic*. Iconic gestures had a relationship to what the utterance referred to. For example, a participant might gesture two objects in space to represent their spatial relations. Noniconic gestures were a mix of metaphoric gestures and noncodable gestures (not in any of the categories described here), such as when the gesture had no clear connection to what was being said.

3.5.1. Gestures and Communication: The analysis was then focused on the relationship between the gesture types and the communication types that were identified earlier. In particular, the current interest was in the correlations between transformed and untransformed categories of utterances. For this purpose, the transformed object description and transformed revision categories were merged into transformed description, and the untransformed object description and untransformed revision categories were merged into untransformed description. The general correlational strength of relations between each of these two categories of description was then calculated with the four types of gestures in simple and complex tasks. Significant correlation was found between iconic gestures and transformed description, but the correlation was much stronger in complex tasks than in simple tasks (simple: $r = 0.21$; complex: $r = 0.48$, $p < 0.01$). In addition, it was found that the correlation between deictic gestures and untransformed description was significant in the complex tasks ($r = 0.19$, $p < 0.01$), but not in simple tasks. No other significant correlation was found.

To further confirm the relations between gestures and communication, the co-occurrences of deictic and iconic gestures was counted with the transformed and untransformed descriptions in each trial of the simple and complex tasks (see Figure 6). Then, a 3-way ANOVA shows significant main effect of task types ($F(1, 6) = 3.12$, $p < 0.05$) and interaction effect between utterance and gesture types ($F(1, 6) = 4.12$, $p < 0.05$). Although in general there were significantly more co-occurrences in complex than simple tasks, the co-occurrences of iconic gestures and transformed description and those of deictic gestures and untransformed description was much higher ($p < 0.05$). The results confirmed the results from the correlational analysis—deictic gestures co-occurred more often with untransformed description, while

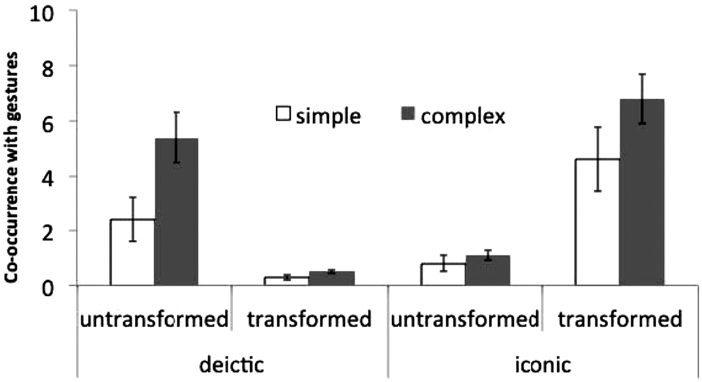


Figure 6. Co-occurrence of deictic and iconic gestures with untransformed and transformed description in simple and complex tasks in the video condition. Error bars represent standard errors.

iconic gestures co-occurred more often with transformed description, and this difference was much larger in complex than simple tasks.

3.5.2. Gestures and Errors: To understand how gestures were related to uncorrected errors in the Video condition, all trials in which gestures or errors were observed was first extracted, and the correlation between the number of deictic and iconic gestures and number of each type of errors was then calculated as shown in Section 3.2. Figure 7 shows the correlation between each type of gesture and each type of errors. Correlations were significant between deictic gestures and interpretation errors ($p < 0.05$), iconic gestures and interpretation errors ($p < 0.05$), iconic gestures and transformation errors ($p < 0.05$), and deictic gestures and misread errors ($p < 0.05$).

As shown in Figure 7, deictic gestures were positively correlated with interpretation but negatively correlated with misread errors, yet iconic gestures were negatively correlated with interpretation errors and transformation errors, and the correlations were almost twice as strong in complex than simple tasks. In other words, when there were more deictic gestures, there was a higher tendency that interpretation errors were observed, but there was a lower tendency that misread errors would occur; when there were more iconic gestures, there was a lower tendency that transformation errors were observed. Combining with earlier findings, one explanation was that because deictic gestures tended to co-occur more often with untransformed description (see Figure 6), information communicated through the verbal, untransformed description and the deictic gestures would more likely be associated with interpretation errors.

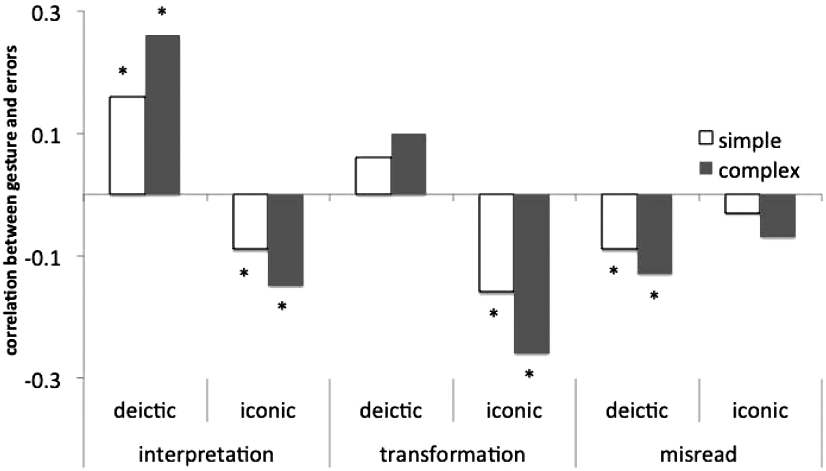


Figure 7. Correlation between gestures and errors in simple and complex tasks. Asterisks indicate significance at $p < 0.05$ level.

The results were consistent with the notion that, compared to the Audio condition, participants in the Video condition were less efficient in establishing a common shared spatial representation during communication, and they tended to rely more on the different, untransformed representations during communication. As a consequence, there were more interpretation errors in the Video condition than in the Audio condition. Given the high correlation, it was possible that deictic gestures not only did not help communication, but might have made the communication of spatial information less efficient. On the other hand, deictic gestures seemed to help individuals to process information on their own display and thus they tended to reduce misread errors.

In contrast to deictic gestures, iconic gestures were negatively correlated with interpretation and transformation errors. Rrevious findings showed that iconic gestures tended to co-occur more often with transformed description. The high negative correlation between iconic gestures and transformation errors was consistent with the notion that iconic gestures helped spatial transformation, and thus the more iconic gestures, the fewer the number of transformation errors. The finding that iconic gestures were negatively correlated with interpretation errors was consistent with the notion that iconic gestures helped communication of spatial information, in the sense that they allowed the listener to better interpret the spatial transformations performed by the speaker.

4. CONCLUSIONS AND DISCUSSION

Results from the current study provided strong evidence that the impact of communication method on both communication patterns and performance in a remote spatial orientation task is stronger than that induced by individual differences in spatial abilities. Specifically, results from the current experiment showed that performance in the Audio condition was better than that in the Text and Video conditions. Previous research has consistently shown that video often helps (or at least not hurts) communication and task performance (e.g., Veinott et al., 1999). However, in our study, we found that performance in the Video condition was worse than the Audio condition. This might appear to be a puzzling result as visual information in the Video condition was believed to help establishing common ground between the persons.

The current results showed that the reason why participants in the Video condition performed worse in this task was that the speakers performed fewer mental transformation and relied on visual interactive feedback to ensure that information was correctly interpreted. This was confirmed by our analysis of the communication patterns, which showed that, compared to the other two conditions, speakers in the Video condition tended to describe spatial information based on their own egocentric spatial perspectives. The availability of visual interactive feedback apparently induced the speakers to "offload" the effort required to establish common ground (or common spatial perspective) to the hearer. When visual feedback was not available in the Audio condition, speakers put significantly more effort in transforming the spatial information to ensure that the hearers could correctly interpret the information.

Consistent with the conclusions by Melinger and Levelt (2004), results from the current study showed that participants might have viewed information conveyed in gestures as a form of shared knowledge when establishing common ground. The current results showed that when participants communicated by combining verbal and manual (gestures) modalities, the perceived shared knowledge expressed by gestures could induce participants to perform less spatial transformation when communicating verbally. In the current task, less transformation led to more errors. This is in contrast to results from previous studies, in which gestures were found to in general help communication (Hostetter, 2011). On the other hand, the difference could be an artifact of the specific design of the task—spatial transformations were essential for the task given that the perspectives were always different between the pair. Further experiments are needed to address the relations between gestures and the perceived common ground established during communication.

The current task was specifically designed such that it demanded exchanges of information based on different spatial perspectives, which forced participants to interleave demanding spatial transformations with communicative actions. Although the current results cannot be generalized to conclude that video-based communication is always worse than audio-based communication, the current set of analysis did indicate that the poorer performance in the Video condition emerged out of the combination of the nature of the spatial task, the task structures (in terms of the availability of different spatial information to each participant in the pair), and communication methods in the experiment. The influence of the task, environment, and communication needs is believed to be guided by the desire to minimize effort (Clark, 1996).

Minimizing efforts can be done either at the *individual* level or at the *collaborative* level. When speakers minimize effort in the individual level, linguistic description may be too ambiguous. The hearer may request clarification, which may eventually lead to more *collaborative* effort for the pair to establish common ground. Minimizing the effort required to establish common ground between a pair can therefore be tricky, as it requires to speaker to assess to what extent the hearer can understand a certain description to determine whether more or less effort is required. Perhaps in our current study, the available of visual interactive feedback induced speakers to overestimate the chance that untransformed spatial information could be correctly interpreted. The overestimation thus contributed to the less than optimal communication. The results again highlighted the complex interactions between cognitive operations and communication that are moderated by the various constraints imposed by the environment (the communication methods, the availability of information, and differences in spatial perspectives).

Another possible reason why pairs in the Video condition performed poorer could be due to the existence of visual interfere, which might have caused extra cognitive effort for the pairs to process the information (that there were two displays in the video condition but only one for the audio). Visual working memory is known to influence spatial thinking, and thus the extra display might have caused participants to switch back and forth between the displays during communication and interfere with their spatial reasoning. The extra chance of producing more errors could have also contributed to the poorer performance, although it could not explain the differences in communication patterns. In other words, although we could not rule out the possible effects of visual interference, but the large set of analysis on the relations between communication patterns, gestures, and errors that we presented suggested that differences in communication structures induced by the communication methods were the major reasons for the differences in performance.

Results from our gesture analysis provided further insight into the relations between communication and performance. Specifically, iconic gestures correlated with successful communication of transformed spatial information. The finding that deictic gestures were positively correlated with interpretation and misread errors was consistent with the idea that participants were overconfident that their partners could correctly interpret the untransformed information. Therefore, at least in the current task, results are in general consistent with the idea that deictic gestures help individual processing of spatial information but they often induce more interpretation errors in communication.

Our results also suggest that not all gestures help communication of spatial information. In fact, deictic gestures were correlated with more interpretation errors in communication. It is possible that, even thought gestures are known to be communicative, the extent that different gestures help depends on the type of information to be processed and communicated. Our results hinted that perhaps iconic gestures are more effective when transformation of spatial information is being communicated, presumably because iconic gestures afford information processing in spatial working memory. On the other hand, deictic gestures may be more useful for self-references or individual cognitive processing rather than communication. Although this is speculative, it is consistent with findings by Trafton et al. that iconic gestures seem to spontaneously occur more frequently than other gestures when spatial transformation is being performed. Future research may be needed to further investigate the communicative nature of different gestures, and how they facilitate processing by different types of working memory in communication.

Results from the current experiment highlighted the complex interactions between communication methods and strategies. In particular, the current results show that when visual interactive feedback is available, speakers tend to adopt egocentric spatial perspectives to minimize effort in mental transformation and rely on the feedback to ensure that the hearer correctly interprets the information. When visual interactive feedback is not available, speakers will put more effort in transforming spatial information to help the hearer to understand the information. The current result therefore demonstrated a somewhat puzzling situation in which allowing two persons to see and communicate with each other during a remote a spatial reasoning task can lead to *more* errors because of the use of a *suboptimal* communication strategy. Although one main motivation for technologies such as video conferencing is to provide more "natural" communication cues (facial expression, hand gestures, etc.) to improve communication effectiveness, the current findings suggest that more research is needed to understand their impact on communication and performance.

REFERENCES

Alibali, M. W. (2005). Gesture in Spatial Cognition: Expressing, Communicating, and Thinking About Spatial Information. *Spatial Cognition & Computation: An Interdisciplinary Journal, 5*(4), 307–331.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language, 44*(2), 169–188.

Allen, G. L. (2003). Gestures accompanying verbal route directions: Do they point to a new avenue for examining spatial representations? *Spatial Cognition & Computation: An Interdisciplinary Journal, 3*(4), 259–268.

Andonova, E., Tenbrink, T., & Coventry, K. R. (2010). Function and context affect spatial information packaging at multiple levels. *Psychonomic Bulletin & Review, 17*(4), 575–580.

Carlson-Radvansky, L. A., & Radvansky, G. A. (1996). The Influence of Functional Relations on Spatial Term Selection. *Psychological Science, 7*(1), 56–60.

Chapanis, A., Ochsman, R. B., Parrish, R. N., & Weeks, G. D. (1977). Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving. *Human Factors, 14*(6), 487–509.

Clark, H. H. (1996). *Using language.* Cambridge, MA: Cambridge University Press.

Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine and S. D. Teasley. (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA Books.

Daly-Jones, O., Monk, A., & Watts, L. (1998). Some advantages of video conferencing over high-quality audio conferencing: Fluency and awareness of attentional focus. *International Journal of Human-Computer Studies, 49*(1), 21–58.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied 3*(2), 105–125.

Dong, W., & Fu, W.-T. (2012). One Piece at a Time: Why Video-Based Communication is Better for Negotiation and Conflict Resolution In *Proceedings of the 2012 ACM conference on computer-supported co-ordinated work* (pp. 167–176). Seattle, WA: ACM Press.

Goldin-Meadow, S., & Wagner, S. (2005). How our hands help us learn. *Trends in Cognitive Sciences, 9*(5), 234–241.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology 10*, 57–67.

Green, C. A., & Williges, R. C. (1995). Evaluation of alternative media used with a groupware editor in a simulated telecommunications environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*, 283–289.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*(2), 151–176.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*(2), 175–191.

Hostetter, A. B. (2011). When do gestures communicate? A Meta-Analysis. *Psychological Bulletin, 137*(2), 297.

Keehner, M., Lippa, Y., et al. (2006). Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology, 20*, 487–503.

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*(1), 47–77.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*(2), 54–60.

Levelt, W. J. M. (1982). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 251–268). Chichester, UK: Wiley.

McNeil, N. M. (1992). *Hand and mind: What gestures reveal about thought.* Chicago, IL: University of Chicago Press.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*(2), 119–141.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8*(4), 389–428.

O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., & Bruce, V. (1996). Comparison of face-to-face and video-mediated interaction. *Interactive Computing, 8*(2), 177–192.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction, 15*(2), 139–178.

Olson, J. S., Olson, G. M., & Meader, D. K. (1995). What mix of video and audio is useful for small groups doing remote real-time design work? In *Proceedings of the CHI'95 conference on Human Factors in Computing Systems* (pp. 362–368). Denver, CO, ACM Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47* (1), 1–24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20*(2), 219–247.

Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K. R. Coventry, T. Tenbrink & J. A. Bateman (Eds.), *Spatial Language and Dialogue (Vol. 1)*. pp. 23–39. Cambridge, MA: Oxford University Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*(2), 211–232.

Siegel, J., Dubrovsky, V., Kiesler, S., & McGuire, T. W. (1986). Group processes in computer-mediated communication. *Organizational Behavior and Human Decision Processes, 37*(2), 157–187.

Straus, S. G. (1997). Technology, group process, and group outcomes: Testing the connections in computer-mediated and face-to-face groups. *Human-Computer Interaction, 12*(3), 227–266.

Straus, S. G., & McGrath, J. E. (1994). Does the medium matter? The interaction of task type and technology on group performance and member reactions. *Journal of Applied Psychology, 79*(1), 87–97.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language, 35*(3), 371–391.

Tenbrink, T., Coventry, K. R., & Andonova, E. (2011). Spatial strategies in the description of complex configurations. *Discourse Processes, 48*, 237–266.

Trafton, J. G., Trickett, S. B., et al. (2006). The relationship between spatial transformations and iconic gestures. *Spatial Cognition & Computation: An Interdisciplinary Journal, 6*(1), 1–29.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional mental rotation. *Perceptual and Motor Skills, 47*(2), 599–604.

Veinott, E. S., Olson, J., Olson, G. M., & Fu, X. (1999). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In *Proceedings of the 1999 ACM Conference on Human Factors in Computing Systems (CHI)*. Pittsburgh, PA: ACM.

Zacks, J. M., & Tversky, B. (2005). Multiple systems for spatial imagery: Transformations of objects and bodies. *Spatial Cognition & Computation: An Interdisciplinary Journal, 5*(4), 271–306.