RealTourist – A Study of Augmenting Human-Human and Human-Computer Dialogue with Eye-Gaze Overlay

Pernilla Qvarfordt[1], David Beymer[2], and Shumin Zhai[2]

[1] Department of Computer and Information Science, Linköping University,
SE-581 83 Linköping, Sweden
`perqv@ida.liu.se`

[2] IBM Almaden Research Center, 650 Harry Rd, San Jose, CA 95120, USA
`{zhai, beymer}@almaden.ibm.com`

Abstract. We developed and studied an experimental system, RealTourist, which lets a user to plan a conference trip with the help of a remote tourist consultant who could view the tourist's eye-gaze superimposed onto a shared map. Data collected from the experiment were analyzed in conjunction with literature review on speech and eye-gaze patterns. This inspective, exploratory research identified various functions of gaze-overlay on shared spatial material including: accurate and direct display of partner's eye-gaze, implicit deictic referencing, interest detection, common focus and topic switching, increased redundancy and ambiguity reduction, and an increase of assurance, confidence, and understanding. This study serves two purposes. The first is to identify patterns that can serve as a basis for designing multimodal human-computer dialogue systems with eye-gaze locus as a contributing channel. The second is to investigate how computer-mediated communication can be supported by the display of the partner's eye-gaze.

1 Introduction

In face-to-face conversation, much can be felt from the conversational partners' eye-gaze—whether they are interested or bored, attentive or preoccupied, engaged or unmindful, in doubt or in agreement, wanting to continue or trying to finish the conversation. Indeed, research has confirmed that eye-gaze plays an important role in face-to-face conversation. It enables us to assess a conversational partner's under-standing, what he or she is looking at, and his or her feelings [2].

In this paper we present the design and study of RealTourist, an experimental on-line tourist information service. RealTourist allows a tourist to consult a remote tourist consultant with shared visual information on computer screens. In addition, the system overlays the tourist's gaze locus onto the tourist consultant's view of the shared workspace (see Fig. 1). By superimposing eye-gaze, a direct and accurate indication of the tourist's trajectory of visual attention on the visual spatial stimuli is displayed for the consultant. Potentially, such a paradigm gives rise to opportunities for better awareness of the partner's attention, more assurance that the partner is "on the same page", less communication ambiguity, and another channel for grounding—a fundamental characteristic and need in human conversation [7].

Fig. 1. The tourist consultant's view of the RealTourist system. The red and blue bull's-eye pattern above the photo is the tourist's gaze.

The purpose of this study is two-fold. First, we attempt to contribute to an empirical foundation of future multimodal human-computer interaction systems with eye-gaze as a critical channel of input. RealTourist can be considered a "Wizard of Oz" simulation of an intelligent multimodal system that can see the user's eye-gaze. Second (and compatible with the first purpose), this study contributes to the understanding of RealTourist as a paradigm of computer mediated human-human communication system augmented with eye-gaze overlay. Given the exploratory nature of this work, our research methodology is a combination of system development, literature analysis and qualitative empirical study.

2   Real Tourist – An Experimental System

RealTourist has two views, one for the tourist and one for the tourist consultant. The tourist could talk to the tourist consultant by voice and see a map of a city on an 18" 1024 x 768 pixel TFT display integrated with a Tobii ET-17 eye-tracker (Fig. 2). The

Fig. 2. The "tourist" side of the RealTourist system

eye tracker operated at 30 Hz and its output was sent to the tourist consultant's computer. The tourist consultant's computer ran an application, TGuide, which displayed the interactive map of the city with the tourist's eye-gaze overlaid on the map. The map, driven by the tourist consultant, was shared on the tourist's side. With TGuide the consultant could access tourist information of the city through pull down menus and control a photo display of places such as a restaurant or a museum (Fig. 1). The two computers were connected by local area network (TCP/IP). RealTourist was developed in a combination of Macromedia Director and C++.

The tourist consultant could see the tourist's eye-gaze on the map as a multi-colored dot. We found that superimposing instantaneous eye-gaze was too jittery and hard to follow. A simple recursive filter, which made the eye-gaze location subjectively easier to perceive, was adopted:

$$y(i) = 0.1x(i) + 0.9y(i-1) \ .$$
(1)

where $y(i)$ was the current displayed 2D position of the eye-gaze; $x(i)$ was the current measured gaze position, and $y(i-1)$ was the past (last sample) location.

Tourist information and maps of two imaginary cities, Vapour Bay and Nolia, were created. Each city had a number of hotels (13), restaurants (12), attractions (9), night-clubs (5) and movie theaters (3). Each hotel, restaurant, attraction and nightclub had a small body of information text. In addition, general information about transportation and tourism were made available to the tourist consultant via the TGuide pull down menu.

When the tourist consultant and tourists talked about a place, the tourist consultant could show a photo of it on the tourist's map. The photo had a line connecting to the location of the places in the map (Fig. 1).

3   Related Work and Research Implications

There is a rich body of literature that bears relevance to the current work. For the sake of brevity, in this section we only review work in HCI and CSCW that sets the general background of this study. We will review the literature in the broader cognitive sciences later when discussing our empirical findings.

Researchers have long explored using eye-gaze both as alternative computer input methods and for computer mediated communication. Eye-gaze as a modality for human-computer interaction has primarily been explored in the tradition of direct manipulation interfaces in which eye-gaze is used as a pointing device, either explicitly [13, 26] or implicitly [28, 29]. In multimodal interaction, eye-gaze has either been used for the purpose of disambiguation [23, 30], or as a complement to the speech channel [15]. We are interested in eye-gaze as a general channel in multimodal human-machine dialogue system beyond the function of pointing. Given the fundamental ambiguity in natural language [27], developing human-computer dialogue system is very challenging. Of particular difficulty is for computer systems to take initiative relevant to the user's need and interest in conversation. "Knowing" where

the user is looking at on its "face" (the interface), a computer system should have more contextual clues on what topics to start and carry a dialogue. In order to develop multimodal dialogue systems with eye-gaze as a contributing channel, we have to understand how eye-gaze patterns are related to discourse in human-human communication.

The benefit of seeing a communication partner's eye-gaze has motivated many well known CSCW design solutions. For example, Buxton & Moran [6] used half-silvered mirror to optically align camera with video screen to enable eye contact ("video tunnelling"). Vertegaal used virtual faces (avatars) that could rotate depending "who is talking to whom" [25]. For tasks that involve visual spatial information, research suggests that a shared view of the workspace is more important than a view of the partner's face /gaze [8, 16]. People tend to look more at the shared workspace than their communication partner [3]. With advances in technology it is possible to bring the view of the partner's face and the view of the shared work pace closer in location. For example by presenting graphical information on a "clear board" which superimposes the partner's image over common work surface in a video tunnel, Ishii and colleagues [12] made both the collaborator and drawings visible in the same visual space. Monk and Gale [18] have shown that the number of turns and numbers of words spoken decreased when the collaborators could see each other's gaze through this type of set-up, although no improvement on task completion time was found.

Instead of displaying partner's whole face, only displaying the eye-gaze is an interesting alternative. While not the focus of their research, Vertegaal [25] displayed all participants' tracked eye-gaze as light spots in a shared document view. As another example, Velichkovsky [24] conducted a pioneering study in which one party's eye-gaze was superimposed onto a computer puzzle game. Velichkovsky was interested in how joint attention may facilitate the performance of a team in which the expert gave suggestions to a novice on how to move the scattered parts around. Since such a puzzle game requires frequent deictic referencing, the improved team performance over voice-only condition is quite plausible.

The literature also indicates that communication (human-human or human-computer) is complex and plastic. In general simple measurements are not sensitive to experimental manipulation and do not give a good sense of what is going on. For example the impact of being able to see a partner's gaze is not easily measurable in task completion time (Ochsman and Chapanis [19]). When the conversational partners could not see each other, they compensated the lack of visual feedback from the partner's face with more verbal feedback [5, 10].

4   An Empirical Study

To observe how user's verbal and gaze behavior were related, we conducted an empirical study in which "tourists" talked with "tourist consultants" using the Real-Tourist system. With the tourist consultant's help and advice, the tourist was asked to plan a conference trip to two cities.
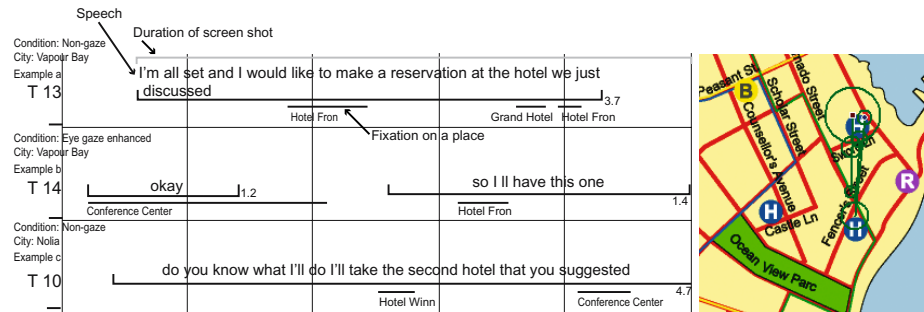
Fig. 3. Examples of the correspondence of eye-gaze and verbal deictic expressions, screen shot for example (a)

4.1 Experimental Design

Seventeen participants volunteered to act as tourists in the study; however, only twelve participants, one woman and eleven men, could be used in the study, since the eye tracker had problem tracking five of the participants. (The most recent model of eye-tracker from the same vendor has improved significantly.) Their ages ranged from 27 to 40, and the median age was 34. Two additional participants, one woman and one man, acted as tourist consultants. The tourist consultants had extensive training on the two imaginary cities used in the tasks and on the TGuide interface.

The specific subtasks for the tourist included finding a suitable hotel that was not too far away from the Conference Centre and within a price limit, a restaurant within a price limit and attractions to visit during the weekend. The tourists were also asked to consider their personal preferences in making his or her plans.

Working with one of the two tourist consultants, each tourist made trip plans for both cities, one with eye-gaze overlaid on the tourist consultant's screen and other without. The tourists did not know whether or how their eye-gaze was displayed before and during the experiment, but were informed after the test. The city and gaze conditions were balanced across participants.

4.2 Set-Up and Procedure

The tourists and the tourist consultant were located in the same room. They were separated by a physical divider so they could hear but not see each other.

The experimental session started with an introduction of the study and eye tracker calibration (Fig. 2). This was followed by the trip planning for the two cities. After completion, the tourist was asked to fill in a questionnaire. The consultant also filled a questionnaire and took notes after each tourist participant completed his or her test.

4.3 Data Collection and Analysis Method

Although the study used a classical experimental set-up, it did not follow the classical quantitative experimental paradigm that compares performance difference between two conditions. The study was inspective in nature, and can be regarded as a "qualita-

tive experiment." It was closer in spirit to ethnographic field studies or inspective studies of problem solving skill analysis in chess playing. The focus was to observe and identify patterns and phenomena rather than performance, which is a necessary phase of scientific inquiry. As Simon argued, "Perhaps we need to add to the textbooks a chapter, or several chapters, describing how to observe the world intently, in the laboratory or outside it, with controls or without them, heavy with hypotheses or innocent of them". [22, p. 394]

During the experiment, the conversation between the tourist and the tourist consultant, the tourist's eye-gaze (in both conditions), and screen events were all recorded. After the experiment, the data were played back in real time by a toolkit we developed which allowed us to play, pause, repeat, transcribe and annotate what happened in the experiment. The main body of the analysis was done on these playbacks. They were inspected, segmented and annotated with regards to different events, such as a change of focus in the conversation. Events or parts of events exemplifying reoccurring patterns found in the inspection were then further segmented, annotated and transcribed. The purpose of the transcriptions was to confirm the findings from the inspections of the playbacks. In the transcribed segments the tourist's fixations were analyzed with a dispersion-based fixation algorithm [21].

For this study we developed a special notation system composed of speech-gaze transcription graphs. In these graphs, verbal dialogue is transcribed and coded on a time line with the tourist's eye-gaze marked (Fig. 3). The dialogues were segmented into 5 second "pages". At occasions, the speech-gaze transcription graphs were supplemented with a still image of the tourist's eye-gaze on the map. The time period shown in the still image is marked with a gray line in the speech-gaze transcription graph (Fig. 3). Due to space constraints, only a few examples of speech-gaze transcription graphs are presented in this paper.

5   Results, Analysis, and Discussion

This section presents findings based on data from our experiment as well as related eye-gaze behavior literature in the cognitive sciences such as psycholinguistics. We organize our findings around a non-exhaustive and often mutually dependent list of plausible impacts of eye-gaze overlay on human-human communication and human-machine conversation.

(1) Eye-gaze carries deictic and spatial reference information; hence, displaying it may reduce effort of frequent referencing.
When communicating about spatial tasks one often needs to make reference to an object, a location or a path. This can be done either by giving complex verbal descriptions, such as "in the north east part of town, close to the Cathedral", or by deictic references in a combination of an utterance and a gesture, such as saying "here" and pointing at a place. Deictic information in fact is also naturally carried in eye-gaze. It is obvious that in explicit deictic pointing, one has to look at an object before gesturing at it. More generally, people look at the object they talk about even when they are not pointing at it [11, 17, 20]. Thus, the eye-gaze locus can serve as an

implicit pointer when a person utters a spatial reference. Velichkovsky's [24] focused on using eye-gaze as an explicit pointer during cooperative problem solving by an expertise-novice team. He showed that when the expert's eye-gaze position was represented by a dot in the puzzle on a computer screen, the number of overall words used decreased.

The fact that eye-gaze could support deictic or referencing functions is also quite evident in our experiment. Fig. 3 shows the speech-gaze transcription graph of three such examples. In example (a) (top row, with screen image on the right), after discussing a number of restaurant choices close to one of the hotels on the map the tourist decided to book "Hotel Fron". During the utterance "I'm all set and I would like to make a reservation at the hotel we just discussed", his eye-gaze fixated on Hotel Fron more frequently (twice) and over longer periods than other objects. Similarly in the next example in the same figure, Tourist 14 also referred to Hotel Fron by an utterance while glancing at it, which again shows that the eye-gaze could have served as implicit deictic reference if the tourist consultant could have seen the eye-gaze locus. These examples show that "pointing with the eye" is a quite natural and subconscious behavior.

The subconscious role eye-gaze plays as a pointer to locations in the map was also demonstrated by the tourist consultants' experiences. They often felt more lost in the non-gaze condition: "I was more lost in the non-gaze condition. The tourist was very inquisitive and asked information about specific places that I did not know their location." (*Tourist consultant B*). In sum, displaying eye-gaze onto the workspace can reduce the need to make explicit and effortful references (either verbal or gesture), since often the partner will know exactly what is being referred to based the eye-gaze.

Beyond supporting deictic and spatial references in human-human conversation, the eye-gaze as a pointer to the current focus of the conversation can also benefit human-computer communication. It can help disambiguating the users' speech both at the speech recognition level and the semantic interpretation level.

(2) Eye-gaze reflects a listener's interest and can be used to judge whether to continue on the current conversation topic.

The partner's eye-gaze can play a role also when the partner is listening. Previous research has shown the eye-gaze tended to correlate with objects related to the verbal message heard. The pioneering research of eye movements during listening was done by Cooper [9]. He found that people look at objects that were relevant to what they listened to. For example, when they heard the word "lion", they looked at the lion in the picture. When they heard the word "Africa", they looked at the lion, the zebra and the snake. Not only the eye-gaze follows the speaker's instruction, it may also anticipate what the speaker is going to talk about next. Altmann and Kamide [1] and Kamide et al [14] showed that when the speaker says "the boy will eat the cake" the listener already looked at the cake when the speaker starts uttering "cake."

In our study, we found that the tourist looked at what the tourist consultants talked about at least for a while. In addition we found a tight coupling between eye-gaze patterns and interest level. How interested the tourist was in a topic (a club, a museum, etc) was reflected by the duration and intensity of the tourist's eye-gaze on a particular place.
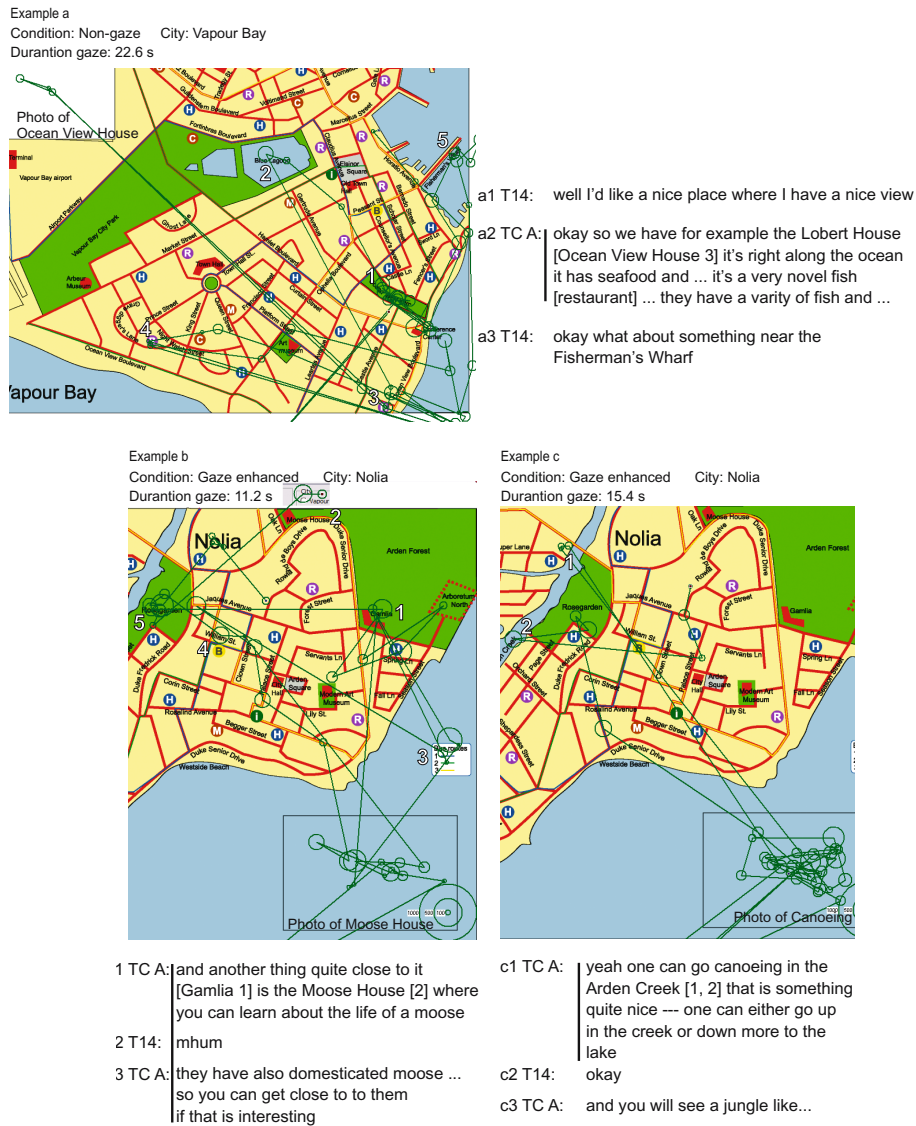
Example a
Condition: Non-gaze   City: Vapour Bay
Durantion gaze: 22.6 s

a1 T14:   well I'd like a nice place where I have a nice view

a2 TC A: okay so we have for example the Lobert House [Ocean View House 3] it's right along the ocean it has seafood and ... it's a very novel fish [restaurant] ... they have a varity of fish and ...

a3 T14:   okay what about something near the Fisherman's Wharf

Example b
Condition: Gaze enhanced   City: Nolia
Durantion gaze: 11.2 s

1 TC A: and another thing quite close to it [Gamlia 1] is the Moose House [2] where you can learn about the life of a moose

2 T14:   mhum

3 TC A: they have also domesticated moose ... so you can get close to to them if that is interesting

Example c
Condition: Gaze enhanced   City: Nolia
Durantion gaze: 15.4 s

c1 TC A: yeah one can go canoeing in the Arden Creek [1, 2] that is something quite nice --- one can either go up in the creek or down more to the lake

c2 T14:   okay

c3 TC A: and you will see a jungle like...

Fig. 4. Three examples from Tourist 14's conversation with Tourist Consultant A that indicate changes of focus and different interest levels. The vertical lines in transcription indicate the time periods used for the gaze fixation trace.

The three examples in Fig. 4 illustrate the eye-gaze relation to interest level. As we can see example (c) is very different from the other two examples: Tourist 14 looked more intensely on the photo and location of the attraction in example (c) than in example (b). Notwithstanding the longer duration of the segment in example (c) compared to example (b), Tourist 14 hardly looked at anything else other than the

Duration gaze: 2.7   City: Vapour Bay   Condition: Gaze enhanced

TCB: So that is a pretty good choice

[break]

Yeah, you can take the yellow bus line.
It is quite accessible.

Fig. 5. Tourist 3 looked up north and Tourist Consultant B changed the focus accordingly

photo of canoeing or the location of the canoeing. The long time period that Tourist 14 spent looking at the photo of the canoe is best explained by his interest in that activity. Indeed Tourist 14 eventually decided to sign up for a canoe trip.

The tourist consultants also interpreted the tourist's interest as high when they saw a high intensity or long duration of the tourist's eye-gaze on places representing the current conversational focus. For example, when Tourist Consultant A observed that Tourist 14 was no longer looking at the Moose House (example b), she finished off the topic by saying " …if that is something interesting." In contrast, in example (c) she observed Tourist 14's high gaze intensity and continued to talk about canoeing for another 10.3 seconds after Tourist 14 uttered "okay."

Subjectively, the tourist consultants felt that they often could tell what the tourists were interested in based on the eye-gaze display. Typically the tourist consultants interpreted a long dwell time over an object or an area as interest: "It was easy to see if he was interested. He looked very focused on those things" or "He did look around elsewhere quite a bit when I talked about these attractions...he did not seem to be interested." (*Tourist consultant A*)

The fact that a person's interest in the current topic is reflected in the eye-gaze can be used by multimodal systems to adapt the information given to the user. This issue is further explored in [4].

(3) Common focus coordination and topic switching

Closely related to interest detection is the coordination of the common focus and topic switching. Displaying a partner's eye-gaze may enable mutual awareness of each other's loci of focus, hence fostering a common focus when needed. Velichkovsky [24] has argued that eye-gaze tracked and displayed to the partner could support joint attention because it constantly communicates where the conversational partner's attention is.

Eye-gaze may also help the dialogue partners to switch the topic of conversation. We have not found much in the literature in this regard, but it was evident in our experiment. When the tourist was no longer interested in a particular place the tourist consultant was talking about, the tourist started to look at new places (see e.g. Fig. 4, Example (a)). The drifting eye-gaze may give the consultant a clue on either moving

on to other topics, or increasing the effort to get the collaborator's attention to the "right" place. In the gaze enhanced condition of the experiment the consultant could take advantage of the displayed gaze trajectory and predict the change of focus or what the tourist was interested in hearing about next.

Fig. 5 shows an example of this behavior. While the Tourist Consultant talked about a restaurant in the northern part of Vapour Bay, he noticed that Tourist 3 looked at the Conference Center. When Tourist 3 next roughly followed the yellow bus route, Tourist Consultant B told the tourist that it was possible to reach the restaurant by bus. This episode shows that the eye-gaze overlay enabled the Tourist Consultant to adapt the dialogue to the Tourist's changing interest. Subjectively, the tourist consultants noticed that they could use the eye-gaze to infer what the tourist was going to ask or talk about next: "[With eye-gaze] I could see what he was interested in and what he was going to talk about." and "I could be better prepared on what the person was interested in…the communication seems to flow better in the second task [with eye-gaze]" (*Tourist consultant A*).

In contrast, for the non-gaze condition, the tourist consultant was more driven by the tourist information tool and the map: "With the non-gaze condition I was more driven by the menu, going down the list since I do not know where the tourist was looking at." and "In the non-gaze condition I was more driven by what ever happened to pop up to me, e.g. I see Museum of Modern Art and I will talk about it." (*Tourist consultant B*)

The extent to which eye-gaze information could be used to steer the conversation also depended on the tourist's personality and style. Some tourists let the consultant drive the conversation, while others wanted to get very specific information themselves. With both these styles, the eye-gaze helped the tourist consultant but in different ways.

When the tourist consultant drove the conversation, the tourists' eye-gaze was used to determine what the tourist consultant should talk about. "I could see places he was interested in and talk about them. He was not talking a lot so I had to drive the conversation." (*Tourist consultant A*) "I tried to use his eye-gaze more this time to drive our conversation. I could use it sometimes effectively." (*Tourist consultant B*)

When the tourist drove the communication, the tourist consultant did not need to rely on the eye-gaze as much for topic switching, but the eye-gaze still could be used for its deictic and reference function discussed earlier: "I could see which places he was talking about. I lacked that in the second city (without eye-gaze). He switched often between different hotels when making up his mind. The eye-gaze would have been very helpful in the second city." (*Tourist consultant A*)

These findings indicate that an eye-gaze overlay can play an important role in the process of grounding [7] in human-human communication. In addition, these findings show that eye-gaze information can provide additional information in human-computer communication for selecting information to present to the user.

(4) Reduced ambiguity and increased redundancy in communication.
While face-to-face communication in natural language is often ambiguous, tele-communication or human-computer communication can be more so. Since people often look at what they talk about, displaying eye-gaze may increase redundancy and reliability in communication.

In the experiment, the tourist's eye-gaze was tightly connected to what was said about objects on the map. The utterance could be a proper name or a referring expression. At occasions these utterances were ambiguous. Instead of referring a hotel by name, a tourist might say "the hotel we just discussed" or "this one" (Fig. 3), which could be interpreted in more than one way if one attended to these words alone. Sometimes what said verbally was not what the tourist really meant. For example in Fig. 3 example (c), Tourist 10 said "you know what? I'll take the *second* hotel you suggested" while he actually meant the *third* hotel. In all these cases, the tourist's eye-gaze could have provided information to either resolve the ambiguity, or to warn the consultant to disambiguate a disparity between eye-gaze and verbal utterance. In all these cases, additional verbal information was needed to resolve the ambiguities.

(5) Completion time, assurance, engagement, confirmation, understanding and confidence in communication.

Task completion time in the experiment did not differ significantly between the gaze conditions ($F_{1, 10}$ =.001, p=.982). This mirrors other results in communication or collaboration studies between different interfaces, such as the study by Monk and Gale [18]. Individual communication style was a more dominant factor on completion time in our study. The conversations with Tourist Consultant B lasted on average longer than with Tourist Consultant A ($F_{1, 10}$ =21.596, p=.001). Tourist Consultant B spent more time introducing the cities and gave more alternatives, while Tourist Consultant A was more targeted. Tourist Consultant A also knew the cities and the TGuide tool better than Tourist Consultant B.

Rather than task completion time, the major impact of an eye-gaze overlay was on confirmation, assurance, understanding and confidence of the tourist consultants when they saw the eye-gaze overlay. Comparing the two conditions used in the experiment, the tourist consultants felt that their communication with the tourists was qualitatively different. They found the gaze-enhanced condition more engaging: "With eye-gaze I was more engaged with the tourist" (*Tourist consultant B*) "It is easier to feel connected with the person when I see the eye-gaze." (*Tourist consultant A*). On the question of how much the eye-gaze helped the communication, they rated on average across all sessions, 72.7, on a 100-point scale (with 50 being neutral).

1 TC B:	yeah is this the one [you asked for] ^a
2 T8:	no ah ^b
3 TC B:	no the other one okay ^b

Fig. 6. Tourist Consultant B uses the tourist's eye-gaze to confirm his understanding of the tourist's request

One could be more sure if his or her partner really "got

the point" by looking at the partner's eye-gaze response. If the speaker's statement about a particular object was followed by the listener's eye-gaze on or around the same object, the speaker can be more confident that the statement is heard. We found that the tourist consultants repeatedly used the eye-gaze overlay to confirm either their own understanding of what the tourist meant, or to confirm that the tourist was following their instructions. Fig. 6 shows an example of how Tourist consultant B first used Tourist 8's eye-gaze in addition to his spoken response to determine that the alternative (Restaurant 2) he showed to the tourist was wrong; and then he used it to determine which was the right alternative (Restaurant 1).

When commenting on the general effects of eye-gaze overlay, the consultants said: "Mostly ensuring me that we are on the same page" (*Tourist consultant B*) "I could make sure that the person followed and understood what I was talking about" (*Tourist consultant A*) "The eye-gaze gives me confidence that the tourist is with me. I felt more presence of the tourist." (*Tourist consultant B*)

Although eye-gaze overlay was a novel and synthetic phenomenon, the tourist consultants were overall positive towards it and found it helpful. Similar to what was found by Velichkovsky [24], they did not find the overt display of eye-gaze confusing, disturbing, or unnatural. Note we used a processed (filtered) display of the eye-gaze in the study. On the other hand, we have observed in some cases the eye-gaze could attract too much attention. For example, consultant A felt that sometimes she had to focus on finding information needed by the tourist and deliberately ignored the tourist's eye-gaze for a while.

6   General Discussion and Conclusions

Being a "window to the mind", human eye-gaze is tightly coupled with mental processes and plays important roles in face-to-face communication [2]. Focusing on the functions of eye-gaze in dialogue, we investigated RealTourist both as a computer-mediated communication paradigm and as a study (simulation) of intelligent multimodal interaction with the user's eye-gaze as a critical channel of input. Our analysis of data and literature identified various functions that the eye-gaze plays including: (1) The overlaid eye-gaze can serve natural deictic and referencing functions that help one partner to keep track of what the other partner talks about, hence reducing the need and effort of frequent explicit verbal or gesture referencing; (2) The eye-gaze information reveals the partner's interest, which helps one to determine how much to talk about a particular topic; (3) Eye-gaze display can help synchronize the two partners' attention and form a common task focus. One may also use the eye-gaze information to switch topics and steer the conversation to the partner's need; (4) Information carried in eye-gaze can help to increase communication redundancy and resolve some of the ambiguities in verbal expressions; (5) Eye-gaze overlay gives one increased assurance or confidence that his or her conversational partner is engaged and indeed getting the information communicated.

Although using RealTourist is a novel experience, the tourist consultants were able to take advantage of it and reacted positively. They felt that it made a qualitative difference in communicating about spatial tasks. Aiming at an inspective pattern identification of new phenomenon rather than an evaluative performance comparison

of well established constructs, the research presented here was not necessarily definitive or complete. As a computer mediated communication paradigm, the eye-gaze augmentation was simplex in the RealTourist system. Although this did not prevent us from observing some of the functions afforded by eye-gaze overlay, a full duplex experiment may reveal more complex and more interesting behavior.

As a Wizard of Oz experiment of multimodal human-computer dialogue system in which the computer may communicate with the user based on speech and eye-gaze information, this study provides a rich source of empirical information on how eye-gaze can be used in the future. In fact, we have taken some of the patterns observed and knowledge gained in this study and embodied them in a prototype system, iTourist, which replaced the human tourist consultant with eye-gaze pattern-based interaction algorithms and databases. We found that iTourist users could successfully accomplish the same trip planning task as in this study [4].

Acknowledgement

References

1. Altmann, G.T., and Kamide, Y. Incremental interpretation at verbs: Restricting the domain of subsequent references. *Cognition*, *73*, (1999). 247-264.
2. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. Cambridge University Press (1976).
3. Argyle, M. and Graham, J. The central Europe experiment - Looking at persons and looking at things. *Journal of Environmental Psychology and Nonverbal Behaviour*, *1*, (1977). 6-16.
4. Blank-for-blind-review. Conversing with the User Based on Eye-Gaze Patterns. *blank for review* (to appear).
5. Boyle, E.A., Anderson, A. H., and Newlans, A. The effect of visibility on dialogue and performance in a cooperative problem solving task. *Language & Speech*, *37*, 1 (1994). 1-20.
6. Buxton, W.A.S. and Moran, T.P. EuroPARC's integrated interactive intermedia facility (iiif): Early experience. in Gibbs, S. and Verrijn-Stuart, A.A. (ed). *Multi-User Interfaces and Applications,*, Elsevier, Amsterdam (1990), 11-34.
7. Clark, H.C. and Schaeffer, E.F. Contributing to discourse. *Cognitive Science*, *13*, (1989). 259-294.
8. Clark, H.H. and Krych, M.A. Speaking while monitoring addresses for understanding. *Journal of Memory and Language*, *50*, (2004). 62-81.
9. Cooper, R.M. The control of eye fixation by the meaning of spoken language - a new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, (1974). 84-107.
10. Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrod, S. and Bruce, V. Face-to-face and video mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied*, *3*, (1997). 105-125.

11. Griffin, Z.M. and Bock, K. What the eye says about speaking. *Psychological Science*, *11*, 4 (2000). 274-279.

12. Ishii, H. and Kobayashi, M., ClearBoard: A Seamless Media for Shared Drawing and Conversation with Eye-Contact. *Proc. ACM CHI Conference on Human Factors in Computing Systems* (*1992*), 525-532.

13. Jacob, R.J.K. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems,*, *vol. 9*, no. 3 (1991). 152-169.

14. Kamide, Y., Altman, G. T. M., and Haywood, S. L. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, (2003). 133-156.

15. Kaur, M., Tremaine, M., Huang, N., Wilder, J., Gacovski, Z., Flippo, F. and Mantravadi, C.S., Where is "it"? Event synchronization in gaze-speech input systems. *Proc. Fifth International Conference on Multimodal Interfaces* (*2003*), 151-157.

16. Kraut, R.E., Gergle, D. and Fussell, S.R., The use of visual information in shared visual spaces: Informing the development of virtual co-presence. *Proc. ACM Conference on Computer Supported Cooperative Work (CSCW)* (*2002*), 31-40.

17. Meyer, A.S., Sleiderink, A.M. and Levelt, W.J.M. Viewing and naming objects: Eye movements during noun and phrase production. *Cognition*, *66*, (1998). B5-B33.

18. Monk, A. and Gale, C. A look is worth a thousands word: full gaze awareness in video-mediated conversation. *Discourse Processes*, *33*, 3 (2002). 257-278.

19. Ochsman, R.B. and Chapanis, A. The effects of 10 communication modes on the behaviour of teams during co-operative problem-solving. *International Journal of Man-Machine Studies*, *6*, (1974). 579-619.

20. Richardson, D.C. and Dale, R., Looking to understand: The coupling between speakers' and listmeners' eye movement and its relationship to discourse comprehension. *Proc. the 26th Annual Meeting of the Cognitive Science Society* (*2004*).

21. Salvucci, D.D. and Goldberg, J.H., Identifying fixations and saccades in eye-tracking protocols. *Proc. ACM Eye Tracking Research & Application Symposium (ETRA)* (*2000*), 71-79.

22. Simon, H.A. The scientist as problem solver. in Klahr, D. and Kotovsky, K. (ed). *Complex information processing: The impact of Herbert A. Simon*, Lawrence Erlbaum, Hilsdale, NJ (1989), 376-398.

23. Tanaka, K., A robust selection system using real-time multi-modal user-agent interactions. *Proc. 4th International Conference on Intelligent User Interfaces* (*1999*), 105-108.

24. Velichkovsky, B.M. Communicating attention-gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, *3*, 2 (1995). 99-224.

25. Vertegaal, R., The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. *Proc. CHI'99: ACM Conference on Human Factors in Computing Systems* (*1999*), 294-301.

26. Ware, C. and Mikaelian, H.H., An evaluation of an eye tracker as a device for computer input. *Proc. CHI+GI: ACM Conference on Human Factors in Computing Systems and Graphics Interface* (*1987*), 183-188.

27. Winograd, T. and Flores, F. *Understanding computers and cognition*. Ablex Publishing Corp, Norwood, NJ, (1986).

28. Zhai, S. What's in the Eyes for Attentive Input *Communications of the ACM* (2003), 34-39.

29. Zhai, S., Morimoto, C. and Ihde, S., Manual and gaze input cascaded (MAGIC) pointing. *Proc. CHI'99: ACM Conference on Human Factors in Computing Systems* (*1999*), ACM Press, 246-253.

30. Zhang, Q., Imamiya, A., Go, K. and Gao, X., Overriding errors in speech and gaze multi-modal architecture. *Proc. 9th International Conference on Intelligent User Interfaces* (*2004*), 346-348.