The Influence of Conversational Setting and Cognitive Load on Reference in 2-party Spoken Dialogue

Barbara Howarth

Department of Psychology
University of Glasgow

Submitted for the Degree of Ph.D. to the Higher Degree Committee
of the Faculty of Social Sciences, University of Glasgow.

February 2004

Acknowledgements

First, I would like to thank my supervisor, Professor Anne Anderson, for her enthusiasm, dynamism and guidance throughout the research and writing up of this thesis. Secondly, I would like to express my gratitude to my second supervisor, Professor Simon Garrod, for his wealth and depth of knowledge, and to Jim Mullin for his technical expertise. I would also like to thank the members of the Language and Communication Group at the University of Glasgow for their constructive criticism on the research presented in this thesis.

Finally, I would like to thank my family and friends for their help and support.

Declaration

I declare that this thesis is my own work carried out under normal terms of supervision.

To Colin and my family

Abstract

The main objective of this thesis is to investigate the way in which the conversational setting (video-mediated compared with face-to-face) and cognitive load (as illustrated by time pressure) influence spoken dialogue, with particular emphasis on the way speakers refer to objects in a discourse.

Two studies were carried out which examined dialogues of pairs of participants performing a problem-solving task. Study 1 examined word duration in a video-mediated conversational setting. In Study 2, pairs of participants performed the Map Task (Brown et al., 1984) under time pressure and without the pressure of time. One group of participants performed the task in a face-to-face conversational setting and the other in a video-mediated setting.

Consistent with the Dual Process Model (Bard et al., 2000), cognitive load influenced complex processes, such as task strategy and the establishment of *common ground*, or mutual knowledge. In contrast, automatic processes, such as *articulatory priming* (the faster articulation of repeated mentions of words referring to the same object), occurred irrespective of the setting in which the conversation took place or of any increase in cognitive load. Under time pressure, interlocutors were less collaborative and less co-ordinated in the way they established common ground than without the pressure of time. Time pressure also led interlocutors to adopt a strategy of making fewer references to objects, or landmarks on the map.

While articulatory reduction occurred irrespective of the conversational setting, participants in a video-mediated setting spoke more slowly than participants in a face-to-face setting. Following Lindblom (1995), this suggested that participants adjusted their articulation in order to be understood in the relatively unfamiliar video-mediated environment. Interlocutors in a video-mediated conversational setting were also less collaborative and less co-ordinated in the way they established common ground compared with participants communicating in a face-to-face setting. Speakers may have felt socially distant (Short, Williams and Christie, 1976) from their interlocutor and the communicative situation in a video-mediated setting.

The findings of this research imply a distinction between consciously controlled processes and automatic processes. Characteristics of spoken dialogue, such as the setting in which a conversation takes place or the cognitive load associated with the communicative task or goal, are more likely to impact on consciously controlled processes than automatic processes. Thus, for example, when participants in a dialogue converse in the usual face-to-face manner and where the cognitive demands associated with the communicative task are relatively low, interlocutors tend to be relatively collaborative in their communication (Clark and Wilkes-Gibbs, 1986; Pickering and Garrod, in press). However, when the communicative circumstances are less than ideal, because the conversational setting is unfamiliar, or because time is short, then complex facets of spoken discourse, such as collaborating with one's interlocutor to establish common ground, may be disrupted. An adequate account of spoken dialogue must account for the effect of dynamic aspects of dialogue such as where the conversation takes place and the cognitive demands associated with the communicative task or goal.

Table of Contents

List of Tables

List of Figures

Chapter 1

An Introduction to Spoken Dialogue in Video-mediated and Face-to-face Communication.

1.1 Introduction

One of the most basic forms of human communication is spoken conversation. This can be thought of as “the integrated activity of two or more persons collaborating to make up what is defined as a dialogue” (Bara, 1995:243). Conversation is considered to be a “highly contextualised form of language use” (Levelt, 1989:29), which is subject to variability in terms of the articulation of utterances (Fowler and Housum, 1987), as well as the content and form of language used (Isaacs and Clark, 1987; Anderson and Boyle, 1994). From the point of view of cognitive psychology, systematic variation in spoken output is informative since the final utterance produced by a speaker reflects decisions taken at the highest levels of language processing (Lieberman, 1963).

Clark (1996) argues that the study of language use must be considered in relation to the *setting* or location in which a conversation takes place. Nowadays computer-based technologies provide new conversational settings in which a dialogue can take place. Forms of communication traditionally associated with a face-to-face setting, such as collaborative working, no longer require the participants to be physically present in the same location. The advent of videoconferencing technology, for instance, enables interlocutors to see each other as well as hear each other. Thus, video-mediated communication offers a novel conversational setting, which, at first sight, appears to approximate face-to-face communication.

A second aspect of spoken dialogue is the fact that it is dynamic. That is to say, a conversation occurs in real time. Consequently, a speaker must plan, formulate and produce utterances in real time whilst interacting with another person. The focus of this dissertation centres on spoken dialogue in video-mediated and face-to-face conversational settings, with particular emphasis on the time frame in which a dialogue occurs.

1.2 Face-to-face and Video-mediated Communication

As stated above, the conversational setting in which a dialogue takes place is a fundamental consideration when investigating spoken dialogue. In the context of this dissertation, I will focus on spoken communication in a video-mediated conversational setting as compared with a face-to-face setting. The term *video-mediated communication* will refer to a conversation between two or more participants that is supported by an audiovisual link. The point of interest being the fact that the participants can see each other as well as hear each other.

It was originally thought that the presence of a visual channel in computer-based communication would serve to simulate the nature of face-to-face communication. However, it is not necessarily the case that the ability to see one's interlocutor in video-mediated communication will deliver the same benefits as the ability to see the other person in a face-to-face conversational setting (Anderson, O'Malley, Doherty-Sneddon, Langton, Newlands, Mullin, Fleming and Van der Velden, 1997). In a video-mediated conversational setting, interlocutors are not co-present in the same physical space. Clark (1996) suggests that the lack of certain features associated with face-to-face communication, such as the co-presence of the speakers, may limit and alter the way language is used. What the participants can say, and when, may be restricted. For example, some studies have shown that, in video-mediated communication participants interrupt each other less often and employ more formal handovers of the floor than in face-to-face communication (Sellen, 1995; O'Conaill, Whittaker and Wilbur, 1993). This suggests that interlocutors engage in a less interactive conversational style in a video-mediated setting than in a face-to-face setting. The remote nature of communication in a video-mediated setting may lead interlocutors to feel distant from one another. Social presence theorists (e.g. Short, Williams and Christie, 1976) suggest that different communications media can be distinguished by the salience of the other person and the communicative situation. According to this view, a video-mediated setting may give rise to a sense of *social distance* between the interlocutors and the communicative situation since the interlocutors are not co-present and do not share the same physical space. Interestingly, field studies have shown that participants often opt for face-to-face communication over conversations mediated by technology (Fish, Kraut, Root and Rice, 1993). Observations such as these suggest that speakers' perceptions of the conversational setting may lead to objective differences in spoken output.

1.3 Modelling Interpersonal Communication

A widely held account of interpersonal communication views discourse as a collaborative process. According to the Collaborative Model (Clark, 1992; Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989), participants in a dialogue engage in a joint activity whereby they collaborate with one another to construct a mental model of the dialogue based on their mutual knowledge and beliefs, or *common ground.* Much of the evidence in support of the Collaborative Model stems from studies of *referential communication.* In short, referential communication is concerned with the way words and expressions are used to refer to objects in a discourse. The most compelling evidence in support of the Collaborative Model stems from the observation that individuals participating in a dialogue have a greater understanding of the content than *overhearers* (Schober and Clark, 1989). The term overhearers refers to individuals who listen to a conversation in which they themselves do not take part. The degree to which interlocutors collaborate to achieve successful communication is reflected in the linguistic forms they choose. For example, Anderson and Boyle (1994) showed that question forms play an important role in introducing new information into a dialogue. The use of question forms was found to be effective in eliciting informative responses from listeners and producing more accurate task performance.

A second widely held view of interpersonal communication holds that the ability to take the perspective of the listener into account lies at the heart of successful communication. Hence, speakers adapt their use of language to the perspective of the listener (Schober, 1993). Research of face-to-face communication has demonstrated that speakers systematically vary the way they refer to objects depending on who they are talking to and the knowledge or expertise their listener may have (e.g. Isaacs and Clark, 1987; Fussell and Krauss, 1992). Evidence in support of this view is not limited to the way words and expressions are used. It extends to the way words forming those words and expressions are articulated. For instance, Fowler and Housum (1987) showed that

2nd mentions of words were articulated more quickly and less clearly than 1st mentions of words referring to the same object. It was reasoned that when an object is referred to a second time, it can be taken as *given* within the discourse context. Consequently, the listener would not require as clear an acoustic signal.

Whether the type of articulatory effect observed by Fowler and Housum (1987) does in fact denote adaptation to a particular listener has been questioned. Dell and Brown (1991), for instance, have suggested that this type of effect can be adequately accounted for in terms of generic speech production mechanisms. More recently, it has been suggested that the extent to which the speaker adjusts his or her speech to the listener may depend on the cognitive demands on the speaker's time and attention. For instance, The Dual Process Model proposed by Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon and Newlands (2000) holds that certain processes underlying speech and language production are automatic and occur within a very fast planning cycle. In contrast, other processes are consciously controlled and occur over a longer planning cycle. These latter processes are more likely, therefore, to be influenced by the cognitive demands on the speaker's time and attention. Thus, the cognitive load on the speech and language production resources may be an important consideration in the study of spoken dialogue.

1.4 The Influence of Cognitive Load in Face-to-face Communication

The term *cognitive load* relates to “the amount of information that has to be held and manipulated in working memory” (Roßnagel, 2000:432) and can be thought of as the mental energy required to perform a given task. Cognitive load is formally defined within the framework of Cognitive Load Theory (Sweller, 1988; Sweller, van Merriënboer and Paas 1998). This will be outlined in Chapter 4 of this dissertation.

Interestingly, cognitive load has been shown to influence perspective taking in terms of the use of referring expressions. For example, Roßnagel (2000) showed that under conditions of low cognitive load (illustrated by an easy version of a model construction task) speakers varied the way they referred to parts of the model, depending on whether they were addressing an adult or a child. When the cognitive load was increased (by having participants perform a more difficult version of the task) this adaptation to the addressee was offset. Findings such as these suggest that adaptation to the listener, for instance, may be sacrificed if the cognitive demands on the speaker’s time and attention are increased.

1.5 Research Questions

The main objective of this dissertation is to explore the influence of conversational setting and cognitive load on human communication in spoken dialogue. The overall questions to be addressed are: Whether the conversational setting will influence aspects of spoken communication such as reference, articulation and collaboration; whether these aspects of spoken output will be influenced by the cognitive load on the speaker; and whether any increase in cognitive load will have the same effect on spoken output in different conversational settings?

Research suggests that the ability to see the other person does not necessarily lead interlocutors to behave the same way in face-to-face and video-mediated conversational settings. Previous studies of video-mediated communication have focused on overall characteristics of the dialogue, such as dialogue length, task outcome (e.g. Anderson et al., 1997) and turn taking (e.g. Sellen, 1995), and have often focused on the long-term use of video-mediated technologies in the workplace (e.g. Fish, Kraut, Root and Rice, 1993). Studies of referential communication have been largely limited to studies of face-to-face communication. Thus, it is not known how the conversational setting will influence the way interlocutors refer to objects in a discourse. For instance, will the setting in which a dialogue takes place influence the way words and expressions referring to objects are used and articulated? And if interlocutors feel socially distant from one another (Short, Williams and Christie, 1976), will they be less collaborative in the way they refer to objects in a discourse?

Studies of referential communication have focussed heavily on the role of listener knowledge on the way speakers refer to objects (e.g. Isaacs and Clark, 1987; Fussell and Krauss, 1992). Such observations have given rise to a view of communication as an essentially collaborative process (Clark and Wilkes-Gibbs, 1986) or one in which the speaker adapts references to objects according to the perspective of the listener (Schober, 1993). This view has been applied to the articulation of words forming referring expressions as well as the words and phrases used to refer to objects (e.g. Fowler and Housum, 1987).

However, more recent accounts of spoken dialogue suggest that certain processes, such as those involving adaptation to the listener, may be influenced by the cognitive demands on the speaker's time and attention (Bard et al., 2000; Roßnagel, 2000; Horton and Keysar, 1996). If cognitive load affects perspective-taking (Roßnagel, 2000), will it also influence the way interlocutors collaborate with one another? Will cognitive load affect all aspects of reference, or as Bard et al., (2000) suggest, are controlled processes more likely to be influenced than automatic processes? Furthermore, if cognitive load does influence these aspects of communication will it have the same impact in a video-mediated setting as in a face-to-face setting?

The purpose of this dissertation is to address these questions in order to gain a better understanding of the nature of the processes underlying speech and language production.

1.6 Methodology

In order to address the questions outlined in section 1.4, spoken dialogue was explored in two conversational settings, namely face-to-face and video-mediated. A collaborative problem solving task, the Map Task (Brown, Anderson, Yule and Shillcock, 1984), was performed under varying conditions of cognitive load. A video-mediated conversational setting was chosen for the following reasons: First, in common with face-to-face communication, interlocutors are able to see each other as well as hear each other. Thus to some extent, video-mediation appears to simulate the nature of face-to-face communication. Second, it is believed that "the language of face-to-face conversation is the basic and primary use of language" (Fillmore, 1981:152 cited in Clark, 1996). Thus, face-to-face communication offers a suitable basis for comparison with other types of conversational setting.

With respect to the manipulation of cognitive load, previous studies have either varied the difficulty of the task or have involved participants performing an additional memorisation task. However, it seems reasonable to suppose that performing the same task in less time will also serve to increase the cognitive load associated with a task. For this reason, cognitive load was manipulated by having participants perform two versions of the Map Task under timed and untimed conditions. There are two main reasons for manipulating cognitive load in this way. First, as stated in the introduction, an interesting feature of spoken dialogue is that speakers must produce utterances in real time. Second, the Dual Process Model proposed by Bard et al., (2000) holds that certain processes underlying language production in spoken dialogue can be differentiated by the time course of their planning cycles.

The Map Task was chosen primarily for the collaborative nature of the task. As a problem-solving task, interlocutors must collaborate to achieve a specific goal. This then offers the opportunity to examine the nature of that collaboration. In addition to this, the Map Task has been extensively used in previous studies and is known to elicit naturalistic, spontaneous dialogue involving repeated references to a specific set of objects - namely landmarks on a map. This task offers a suitable means then for examining the use of referring expressions and the articulation of words forming referring expressions. Furthermore, the task itself gives rise to a task outcome that readily lends itself to an objective measure of task performance.

The research presented in this dissertation was carried out in the following way. First, an initial study examined the articulation of words forming referring expressions in video-mediated dialogues. This provided some indication of whether video-mediated communication functions in the same way as face-to-face communication in terms of word articulation, at least. Second, a further study was conducted to collect a body of materials from which to explore the influence of conversational setting and cognitive load in spoken dialogue. The first line of investigation examined broad characteristics of the dialogue, such as dialogue length, task outcome and rate of speech. The second line of investigation primarily explored the influence of cognitive load in video-mediated and face-to-face conversational settings on articulation and use of referring expressions. A third line of investigation explored the influence of cognitive load and conversational setting on the way interlocutors collaborate to introduce objects into a dialogue.

Finally, the research presented in this dissertation is exploratory in nature and as such is limited to a study of 2-party dialogues. This represents the most basic form of communication and therefore provides a useful starting point for this type of research.

1.7 The Implications of the Research

The main objectives of this dissertation are first, to consider the position that certain processes underlying speech and language production in spoken dialogue may be limited by the cognitive demands on the speaker's time and attention. Second, to consider whether these processes will be influenced by the setting in which a conversation takes place.

There is a general sense in which research into the differences between video-mediated and face-to-face communication is important. There are a range of communicative situations, such as remote collaboration, teleconferencing, and distance learning, where video-mediated technologies are considered to be an invaluable tool. If alternatives to face-to-face communication, such as video-mediated communication, have distinctive properties that lead to differences in speech quality and the way speakers use language, then this would have important consequences for the design and use of computer-mediated technologies.

More specifically, a major task for speakers is to model the dialogue as it proceeds. They must keep track of what is happening in the discourse situation, while simultaneously constructing messages for expression (Levelt, 1989). In doing so, speakers build "mental models" of the referents - objects, persons and events that are introduced and referred to in a dialogue. The speaker's record, or model, of the discourse includes knowledge about the content of the discourse as the dialogue proceeds (e.g. Prince, 1981). Consequently, differences in articulatory quality and the choice of appropriate referring expressions reflect the thought processes underlying speech and language production. Thus, the study of spoken dialogue is of major interest within the field of cognitive psychology.

The purpose of the research presented in this dissertation is to gain a better understanding of the higher-level factors underlying spoken dialogue in video-mediated and face-to face conversational settings. The main issue I seek to address relates to how the processes underlying speech production will be affected by varying the demands on the speaker's time and attention, and whether speakers respond to cognitive load in the same way in face-to-face and video-mediated conversational settings. The findings of this research will largely be interpreted within the framework of the Dual Process Model proposed by Bard et al., (2000) and the Collaborative Model proposed by Clark and Wilkes-Gibbs, 1986).

In Chapter 2 of this dissertation I will review the literature relating to referential communication in terms of recent models of interpersonal communication, and discuss how references to objects might be influenced by the conversational setting and by cognitive load. Chapter 3 focuses on the question of whether video-mediated communication functions in the same way as face-to-face communication in terms of word articulation. The results of an initial study (Study 1), which examined word duration in a video-mediated conversational setting, will be presented. Chapter 4 explores the influence of the conversational setting and the cognitive demands on the speaker on spoken output. The results of a second study (Study 2), which examined the impact of cognitive load (as illustrated by time pressure) on articulation and reference in video-mediated and face-to-face conversational settings, will be presented. Chapter 5 explores the issue of whether cognitive load and conversational setting will influence the way interlocutors collaborate with each other when introducing new objects into a discourse. Further results of Study 2 will also be presented in this chapter. Finally in Chapter 6, the main findings of the research will be summarised and brought to a conclusion. The implications of the findings will be discussed in terms of their impact on our understanding of the processes underlying speech production in dialogue, and pointers to future work will be outlined.

Chapter 2

Referential Communication in Spoken Dialogue: A Literature Review

2.1 Introduction

In Chapter one, an overview of the dissertation was presented. To reiterate, the purpose of this dissertation is to examine spoken dialogue in video-mediated and face-to-face communication. From the point of view of cognitive psychology, the study of spoken dialogue is an important topic since spoken output may reflect underlying cognitive processes involved in speech production. Furthermore, the advance of computer-based technologies, such as video-conferencing systems, provides novel conversational settings in which a dialogue may take place. The main objective of this Chapter is first, to review current research on referential communication in the context of psychological models of interpersonal communication. Second, to consider how the processes underlying speech and language production might be influenced by increased cognitive demands on interlocutors and by the setting in which a dialogue takes place.

In section 2.2, I will review studies of referential communication in order to illustrate the ways in which speakers systematically vary their speech and language when referring to objects in a discourse. This variation occurs both in terms of the linguistic forms speakers use to refer to objects and in terms of the way speakers articulate words forming the names of objects. Existing accounts of variation in speech and language in spoken discourse will be discussed in section 2.3. Section 2.4, provides a critique of current models and highlights the limitations of traditional views of interpersonal communication. In Sections 2.5

and 2.6, I will discuss the notion that the way in which speakers refer to objects may be affected by factors such as the cognitive load associated with a given task and the setting in which the conversation takes place. This will lead to the development of a set of hypotheses in section 2.7, which will constitute the bases of this dissertation.

2.2 Referential Communication

Studies of referential communication have shown that speakers vary their spoken output in terms of the words and expressions they use to refer to objects as well as the way words forming those words and expressions are articulated. Several basic observations have emerged from the literature on referential communication, a review of which is provided by Krauss and Fussell (1996). The main findings are described below.

2.2.1 Articulatory Variation in Relation to Reference

One way in which speech can vary is in terms of how clearly words are articulated. It has been shown that the articulatory clarity of a word is influenced by the context in which that word appears. For example, Lieberman (1963) found that words which were highly predictable from their sentence contexts, as in (1), were less intelligible to listeners when presented in isolation than words which were difficult to predict from their sentence contexts, as in (2):

(1) A stitch in time saves *nine*

(2) The word that you will hear is *nine*

The intelligibility of words presented in isolation gives some indication of how clearly words are articulated. When words are excerpted from running speech, subjects cannot make use of cues provided by the sentence context. Thus any difficulty in word identification can be attributed to the articulatory clarity of the word in question. Although Lieberman (1963) used a sample obtained from only three speakers, more reliable results have also been reported. For example, Hunnicutt (1985) also showed that words, which were highly predictable from their sentence contexts, were articulated less clearly than words which could not be readily predicted from their sentence contexts. Although Hunnicutt's (1985) findings were based on text-type, spoken sentences, similar findings have been reported in more naturalistic settings (e.g. Bard and Anderson, 1983; Fisher and Tokura, 1995). Thus the articulatory quality of words can be influenced by the immediate sentence context.

Articulatory clarity can also be influenced by higher-level factors such as the discourse status of a word. Fowler and Housum (1987) compared the duration and intelligibility of "New" and "Old" words occurring in naturalistic settings such as radio recordings. Word duration provides a further measure of articulatory quality. If words are produced quickly, then speakers cannot have time to articulate those words clearly. "New" words were defined as those words which occurred for the first time in a passage and "Old" words were defined as repetitions of words that had previously been mentioned in the passage. Mentions of "old" words were articulated more quickly and were less intelligible to listeners than introductory mentions of "new" words. Furthermore, listeners were able to utilise these differences in articulatory clarity to differentiate between *"old"* or *"new"* items (Fowler and Housum, 1987). It was also found that mere repetition of words was insufficient to induce the effect of reduced articulatory detail. Repeated mentions of words read in lists do not exhibit the effect of reduced articulatory detail (Fowler, 1988). In addition to this, the effect is attenuated for mentions of words produced in monologue as opposed to dialogue (McAllister, Potts, Mason and Marchant, 1994). It is argued that the *given* status of the entity concerned induces the effect. For instance, Bard, Lowe and Altman, (1989) demonstrated that the effect does not occur when the word refers to a *"new"* item of the same kind. Furthermore, Bard and Anderson (1994) showed that introductory mentions of items, which can be taken as given by virtue of physical presence (i.e. situationally given in Prince's 1981 system), were less intelligible than introductory mentions which were truly new to the context. Taken together these studies showed that the discourse status of words in a dialogue, specifically whether or not an entity can be taken as given, influences the articulatory quality of the speech output. The attenuated articulation of words to signal *given* information has also been reported by Hawkins and Warren (1994), Samuel and Troicki (1998), and Robertson and Kirsner (2000).

It has been shown then, that speakers systematically vary the articulatory clarity of their speech in accordance with factors such as sentence context or discourse status. Words which are easily predictable from their sentence contexts, or which can be taken as given within the wider discourse context tend to be articulated less clearly and more quickly than words which cannot be so readily interpreted from their sentence or discourse contexts.

2.2.2 Variability in the Use of Referring Expressions

From the large body of research on the use of referring expressions, several basic observations emerge. First, the phrases used to refer to objects become shorter over successive references. Second, speakers tend to use definite noun phrases such as *the book,* rather than *a book* on repeated reference to an object. Third, referring expressions tend to vary according to the vantagepoint of the addressee or to the knowledge that the addressee can be assumed to possess. And finally, listeners who participate in a dialogue understand more than listeners who simply overhear a conversation. Before describing examples, which illustrate these phenomena, I will first describe several experimental paradigms that have been commonly used to investigate referential communication.

A variety of tasks have been developed to investigate the use of referring expressions in non-conversational as well as more interactive conversational settings. In non-conversational paradigms the listener is largely silent, or even imaginary. A drawback of this method, however, relates to the generalisability of findings to the most commonly used form of communication, namely spoken dialogue. Nevertheless, many tasks have been adapted such that interlocutors are free to interact with each other as much as they like. The most basic type of referential communication task involves describing spatial relationships between simple geometric shapes. A second type of paradigm involves identifying specific referents from an array of objects. In this task, a *director* describes an object to an addressee, the *matcher* who must correctly identify the referent from the array. In a second type of commonly used matching task (e.g. Clark and Wilkes-Gibbs, 1986), a director describes a series of abstract Chinese Tangram figures to a matcher, who must place the figures in the same order as the director. An advantage of this particular task is that the figures have no pre-existing names. This, then, enables the experimenter to assess the way in which interlocutors establish references to objects. A fourth type of task, The Map Task (Brown et al., 1984), is a problem-solving task which was designed to elicit interactive conversation. An *Instruction Giver* describes a route on a schematic map to a partner, the *Instruction Follower*, whose task is to replicate the route on their map. Interlocutors' maps differ in that not all landmarks appear on both maps. Consequently, interlocutors must solve the problem of which landmarks are shared and which are not in order to successfully navigate their way round the map. The task is essentially a role-play in which the content and vocabulary are largely defined by the task itself. However, within these constraints, interlocutors are free to interact with each other and employ a strategy of their choice to navigate their way round the map. A major advantage of this task is that the resulting dialogues are relatively natural and spontaneous.

In an early referential communication task from which the Tangram task was later derived, Krauss and Weinheimer (1964, 1966) found that referring expressions used to describe nonsense figures became shorter over successive experimental trials. Thus, a figure referred to as "the upside down Martini glass on a wire stand" on an initial trial became abbreviated to "the Martini" in later trials. Interestingly the effect was attenuated when the task was performed in a non-conversational setting. If the listener could speak in response to the descriptions, then the effect was observed. Yet when subjects were instructed that descriptions would be recorded and played back later to a future listener, there was much less shortening of descriptions with repeated reference.

Similarly, for participants performing the Tangram Task, Clark and Wilkes-Gibbs (1986) found that referring expressions became shorter and simpler across and within trials. In their experiment, the director was presented with a series of 12 Tangram figures which the matcher had to arrange in the same order. It was found that the decline in the number of words was steepest from Trial 1 to Trial 2, but levelled out over subsequent trials. In addition to this, the number of words and conversational turns required to place an object dropped significantly over trials and also within trials. Interlocutors used fewer words and fewer turns to describe the last figure in the series than to describe the first. Once again, the decline in the number of words and turns was greatest in early trials but diminished across trials. These findings indicated that interlocutors became more efficient from the beginning to the end of a trial and also over subsequent trials.

Another general observation is that speakers tend to use definite noun phrases rather than indefinite noun phrases upon repeated reference to the same object. Clark and Wilkes-Gibbs (1986) noted that in initial trials of the Tangram task, speakers described a figure with an indefinite references, such as *"a person who's..."*. Yet on subsequent trials the same figure was referred to using a definite reference such as *"the ice skater"*. This finding was replicated in a later experiment (Wilkes-Gibbs and Clark, 1992) and has also been demonstrated in a written version of the Tangram task (Hupet and Chantraine, 1992). Definite references, such as those involving the definite article *the*, are thought to indicate that an entity can be taken as given.

A third widely reported phenomenon relates to the finding that referring expressions vary in accordance with the vantage point of the addressee. For example, using a non-conversational paradigm Schober (1993) found that speakers tended to describe simple objects located in different places in terms of addressee-oriented phrases, such as *in front of you"* more often than speaker-oriented phrases such as *"in front of me."* This suggested that speakers took the visual field of an imaginary addressee into account when describing the objects. However, the effect was attenuated when the task was performed in a conversational setting (Schober, 1993). In order to account for this, Schober (1993) concluded that, in a conversational setting, feedback from the addressee reduced the need for speakers to produce a fully communicative message from the outset since any misunderstanding could be rectified with feedback from the addressee.

There is evidence to suggest that speakers go beyond adjusting their referring expressions to the spatial perspective of the addressee, and in fact develop a joint spatial perspective with their addressee. In a study conducted by Garrod and Anderson (1987) pairs of participants played a computerised maze game in which a number of squares with connecting pathways are displayed. Interlocutors can refer to these using terms such as *rows*, *lines* or *columns.* Garrod and Anderson (1987) found that once a speaker chose a reference term such as *"the fifth row"*, this tended to set a precedent that was followed throughout the dialogue, although not over successive trials of the game. This phenomenon, termed *entrainment* by Garrod and Anderson (1987), has also been observed by Isaacs and Clark (1987) and Jefferson (1982). It was thought to be indicative of a co-ordination strategy whereby speakers formulated their utterances on the basis of referring expressions that have been used previously in the dialogue.

Another widely reported finding relates to the observation that a speaker's use of referring expressions is influenced by what the listener can be assumed to know. Isaacs and Clark (1987), for instance, found that speakers giving directions in New York City tended to use landmark names if they believed their addressee was familiar with the city. However, when addressing someone they believed to be unfamiliar with New York, they tended to supplement landmark names with a description that facilitated identification. Similarly, in a referential task in which speakers described familiar and unfamiliar faces to a partner, Fussell and Krauss (1992) found that speakers used more informative referring expressions that were rich in descriptive content when they believed a face to be unfamiliar to their addressee. The findings of these studies suggest that speakers offer more information when they believe their addressee may not have sufficient knowledge to identify an object.

Finally, another important observation of spoken dialogue is that listeners who take part in a conversation understand more than listeners who simply overhear the conversation. In a study conducted by Schober and Clark, (1989) a speaker, addressee and overhearer performed a version of the Tangram task in which speaker and addressee were asked to order a series of Tangram figures, while a third participant overheard their conversation. It was found that addressees participating in the conversation were better at understanding references to objects than overhearers. Schober and Clark (1989) proposed that overhearers' poorer comprehension of referring expressions was a consequence of not being involved in the communicative process.

To summarise this section, several consistent findings emerge from the literature on referential communication. First, upon repeated reference, words referring to the same object are articulated more quickly and are less intelligible to listeners. Second, referring expressions become shorter on successive use. Third, speakers tend to use definite references upon repeated reference to the same object. Fourth, speakers alter the way they refer to objects in accordance with the spatial perspective of the listener and in accordance with what the addressee can be assumed to know. Finally, participants in a dialogue are better able to identify the objects being referred to than listeners who simply overhear a conversation.

2.3 Models of Interpersonal Communication

Models of interpersonal communication can be classified into 2 broad categories: Those which view interlocutors as being essentially autonomous from one another and those which emphasise the interactive nature of communication. Encoder/Decoder Models, Intentionalist Models, Perspective-Taking Models fall into the former category while the Collaborative Model and the Interactive Alignment Model fall into the latter. These are described below.

2.3.1 Autonomous Models

Encoder/decoder, intentionalist and perspective-taking models, have been referred to as *"autonomous"* since they hold in common the view that language production and comprehension operate in isolation and the speaker and listener act as individuals, effectively "decoupled" from one another.

According to encoder/decoder models of communication (e.g. Cherry, 1956), interlocutors, viewed as information processing units, communicate with one another through the transmission of messages via a communication channel. Applied to spoken dialogue, encoder/decoder models of communication hold that a speaker encodes a message into a linguistic representation, or code, which is transmitted via the speech signal to the listener, who recovers the message by decoding the linguistic representation. Evidence in support of encoder/decoder models is limited and they do not explain how the pragmatic functions of language, such as assertions, requests and suggestions are interpreted from the speech signal.

In common with encoder/decoder models, intentionalist models (Austin, 1962; Grice, 1975; Schiffer 1972; Searle, 1969) hold that the listener arrives at an understanding of language by decoding the linguistic code. However, the listener must go beyond a literal decoding of the incoming signal, such as speech to infer the intentions of the speaker. Thus, an advantage of this model is that it is able to explain how pragmatic functions of language, such as assertions, requests and suggestions are interpreted from the speech signal. Communication is viewed as the exchange of communicative intentions, which are conveyed through the selection of word strings that clearly express the intentions of the speaker. When formulating an utterance the speaker refers to a set of rules to convey his or her intentions. The addressee, who possesses the same set of rules, refers to these in order to infer the intentions of the speaker. These rules are based on Searle's (1969) theory of speech acts and Grice's co-operative principle. Grice argues that interlocutors engaged in conversation are essentially co-operative in that they "make their contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which (they) are engaged" (Grice, 1975:45).

Intentionalist models, as well as encoder/decoder models are able to account for the general observation that referring expressions become shorter within a particular trial of an experiment (e.g. Clark and Wilkes-Gibbs, 1986). According to Grice (1975) messages should contain only the information required. In the Tangram task, for example, fewer objects remain in the array as the task proceeds. Hence less information is required to distinguish the target object from the others. Since less information is required, the referring expressions used to identify the target object become shorter. However, intentionalist models could not explain why referring expressions became shorter when the number of objects in the array remained the same (Krauss and Weinheimer (1964, 1966) or why referring expressions shortened in an asymmetrical fashion. For example, Clark and Wilkes-Gibbs (1986) observed that the decline in the number of words used to describe the first tangram figure in an array to the last figure in the array, became shorter over successive trials of the experiment. Intentionalist models did not predict this.

A major limitation of intentionalist models, as well as encoder decoder models, relates to the numerous studies which indicate that speakers adapt their references to objects in accordance with the perspective of the listener (e.g. Schober, 1993; Isaacs and Clark, 1987; Fussell and Krauss, 1992). According to intentionalist models, the speaker does not take the perspective of the listener into account when planning and formulating utterances. Rather, the listener uncovers the intentions of the speaker by applying a set of inferential rules.

As the name suggests, perspective taking models (e.g. Schober, 1993) do take the perspective of the listener into account. Schober (1993) argues that individuals experience the world from different vantage points, and for this reason speakers must consider the perspective of their addressee to ensure successful communication. Applied to spoken communication, the basic notion is that speakers adjust the content of their utterances in accordance with their addressee's point of view. Thus the comprehension of speech depends, in part, on the ability of the speaker to make assumptions about what their partners know, feel, think and believe.

Evidence in support of perspective-taking models stems from the observation that speakers tend to vary the way they refer to objects in a discourse in accordance with the vantage point of their listener (Schober, 1993), or with the knowledge that their listener can be assumed to possess (Isaacs and Clark, 1987; Fussell and Krauss, 1992). Schober (1993) further argued that feedback from the listener would allow the speaker to refine their understanding of the addressee's perspective. In this way, perspective-taking models were able to account for the attenuation of a preference for addressee-based descriptions over egocentric spatial descriptions in a conversational setting (Schober, 1993). The same line of reasoning was offered in explanation of why referring expressions should become shorter on successive use when feedback was provided but not when feedback was unavailable (Krauss and Weinheimer, 1964, 1966). Feedback reduced the pressure on a speaker to create a fully communicative message from the outset. Consequently, where no feedback was received, speakers relied on their own beliefs regarding the perspective of the listener. Yet in a conversational setting, feedback from the listener reduces the need for fully specified references from the outset.

The view that the speaker adapts his or her speech to the perspective of the listener has also been applied to the finding that words referring to the same object are articulated more quickly and are less intelligible to listeners on repeated mention (Fowler and Housum, 1987). Listeners were able to take advantage of the information provided by reductions in word length to retrieve the earlier context of the word (Fowler and Housum, 1987). It was reasoned that where a word refers to old information, articulatory clarity may be reduced since the referent could be taken as given within the context of the discourse. It was inferred from this that the psychological processes underlying speech production must therefore include a speaker's model of what the listener knows and perceives (Bolinger, 1963, 1981; Chafe, 1974; Lindblom , 1990). Thus if a word is partially specified by the immediate sentence context or by the discourse status of the entity to which it refers, the speaker may assume that the listener would not require as clear a signal because the word can be interpreted in the context of that information.

Although perspective-taking models incorporated the role of the listener, it has been argued that autonomous models are not appropriate for dialogue since the processes involved in language production and comprehension must be interrelated, or *coupled* (Pickering and Garrod, in press). For example, Garrod and Anderson (1987) showed that, in The Maze Game, the referring expressions used by one interlocutor influenced those used by the other. This is not reflected in autonomous accounts of dialogue.

2.3.2 Interactive Approaches to Dialogue

Although perspective-taking models place emphasis on the role of the listener, the speaker and listener are still, nevertheless, viewed as largely autonomous from one another. In contrast to autonomous models, interactive models, such as the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark and Shaeffer, 1989; Clark and Marshall, 1981), and more recently, the Interactive Alignment Model (Pickering and Garrod, in press), view communication as a joint activity.

According to the Collaborative Model (e.g. Clark and Wilkes-Gibbs, 1986), spoken dialogue is characterised as a joint communicative activity in which the speaker and addressee(s) co-ordinate their mutual knowledge and beliefs. Collaborative communication proceeds in a systematic fashion through the accumulation of mutual knowledge and beliefs, or *common ground* (Clark, 1992: Gazdar, 1979, Stalnaker, 1978). Common ground is established through an acceptance process wherein one speaker *presents*, or introduces an object which his or her interlocutor must accept or reject as constituting part of the common ground. In short, the Collaborative Model holds that speakers and addressees go beyond autonomous action and collaborate with each other on a moment-by-moment basis to try to ensure that what is said is also understood.

Many of the findings that have been taken as evidence in support of perspective-taking models are also consistent with the Collaborative Model. For example, within the framework of the Collaborative Model (Clark and Wilkes-Gibbs, 1986), the finding that referring expressions become shorter over time has been taken as an indication that the speaker and addressee develop a joint perspective over time. Furthermore, there were more basic acceptances of referents, such as "*okay*" across trials which indicated that references in later trials were based on prior mutually accepted knowledge. The finding that the use of referring expressions changes depending on what the listener can be assumed to know (Isaacs and Clark, 1987; Fussell and Krauss, 1992) has been taken as an indication that interlocutors co-ordinate their activities. Isaacs and Clark (1987), for instance, also demonstrated that not only do knowledgeable speakers tailor their messages to their addressee (as predicted by perspective-taking models), but they also collaborate with their addressee to ensure that the information they present has been adequately grounded.

However, perhaps the most compelling evidence in support of the Collaborative Model stems from the observation that those listeners who actively participate in a conversation understand more than *overhearers* (Schober and Clark, 1989; Wilkes-Gibbs and Clark, 1992). Since overhearers did not take part in the grounding process, their understanding of the referring expressions was impaired. Consequently overhearers were worse than addressees at identifying the figures in the array. Evidence in support of the Collaborative Model is not limited to the way interlocutors use and understand referring expressions. The finding that articulatory reduction occurs in dialogue but not in monologue (McAllister et al., 1994) provides evidence that collaboration between participants in a dialogue impacts on the articulation of speech. If understanding in conversation were an autonomous process then there should be no such difference. Schober and Clark (1989) therefore concluded that understanding is part of a collaborative process.

There is, nevertheless, evidence to suggest that the Collaborative Model is limited in its generalisability to certain conversational situations. For example, Fay (2000) showed that while interpersonal communication within small groups of 5 interlocutors was consistent with the Collaborative Model, communication involving large groups of 10 people was more in line with an autonomous view of communication. In the larger groups, Fay (2000) observed that a dominant speaker emerged, and that an individual's understanding of the topic of discussion was influenced by the dominant speaker rather than by the people they interacted with, as predicted by the Collaborative Model. In line with the encoder/decoder models, the dominant speaker transmitted the greatest amount of information. Consequently, participants' understanding of what was said was influenced by the dominant speaker irrespective of whom individuals spoke with.

Pickering and Garrod (in press) point out that while the Collaborative Model has emphasised the nature of communicative strategies employed in dialogue, the mechanisms underlying these strategies remain unclear.
In order to address this issue, they proposed an Interactive Alignment Model which assumes that, in dialogue, production and comprehension become tightly coupled such that linguistic representations are co-ordinated or *aligned.* Priming is the central mechanism in the process of alignment and can occur at each level of linguistic representation. Furthermore, alignment at one level of representation enhances alignment at other levels. Essentially, this model views the interlocutors as interlinked at various levels of language production. Thus, what one interlocutor says at one level can influence what the other says. The basic interactive alignment process is automatic and largely unconscious. However, when feedback occurs interlocutors can take control of the alignment process via repair mechanisms.

Evidence in support of the Interactive Alignment Model (Pickering and Garrod, in press) can be found in the observation that when speakers were speaking to a silent or imaginary addressee, or to a tape recorder, the shortening of referring expressions over successive references was attenuated (Krauss & Weinheimer, 1966; Schober, 1993). This illustrates the fundamental observation that speaking in a dialogue differs from speaking in a monologue. Furthermore, the Interactive Alignment Model can account for the observation that participants co-ordinated their use of terms, such *as " the fifth row" (rather than column)* in The Maze Game (Garrod and Anderson, 1987). Evidence in support of the notion that interlocutors tend to develop the same set of referring expressions has also been reported for other referential communication tasks (Brennan and Clark, 1996; Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992). In a similar vein, Levelt and Kelter (1982) found that when Dutch speakers asked a question such as *"What time do you close?"* addressees tended to respond with an appropriate response such as *"Five o'clock"*. These findings support the notion that language production and comprehension are coupled such that the linguistic forms used by one interlocutor influence those used by the other.

Further support for the Interactive Alignment Model, is provided by demonstrations of syntactic and articulatory priming in dialogue. For example, using a conversational picture description task, Branigan, Pickering and Cleland (2002), showed that syntactic structures used by a confederate strongly influenced those used by the participants in the experiment. Furthermore, Branigan, Pickering and Cleland (2002) showed that syntactic priming was stronger when interlocutors actively participated in the discourse than when an interlocutor simply overheard the discourse. This showed that syntactic priming was due, in part, to the interaction between participants in a dialogue. Evidence of comprehension-to-production alignment at an articulatory level is provided by the observation that, in naturalistic dialogue, *articulatory reduction* (the attenuated articulation of repeated mentions of words referring to the same object) occurs across speakers (Bard et al, 2000). Observations such as these provide evidence in support the proposal that alignment is underpinned by priming mechanisms which operate at a syntactic and articulatory level of representation, at least.

In summary, studies of referential communication reflect a basic assumption underlying spoken communication; namely that one speaks to be understood. Many argue therefore that any model of communication must account for the listener. Not surprisingly then, it is those models of interpersonal communication which lay emphasis on adaptation to the addressee's perspective (e.g. Schober, 1993) or co-ordination between interlocutors (Clark and Wilkes, 1986; Pickering and Garrod, in press) which have dominated the relevant literature. However, researchers have questioned the assumption that speakers adapt their speech to the addressee's perspective (Dell and Brown, 1991; Brown and Dell, 1987; Bard et al., 2000; Horton and Keysar, 1996). Furthermore, it has been suggested that the processes underlying adaptation to the listener may be limited by cognitive availability (Bard et al., 2000; Roßnagel, 2000).

2.4 Cognitive Limitations on Speech and Language Processing

More recently, discussions of the role of the listener in spoken dialogue have shifted from *whether* speakers take their addressee into account to *when* speakers take their addressee into account. For example, in a route description task, Buhl (2001) found that, when describing routes to an imaginary addressee, more speakers described spatial locations in terms of their own perspective than in terms of their addressee's perspective. This showed that speakers do not always behave in an addressee-oriented manner.

Brown and Dell, (1987), on the other hand, showed that the circumstances under which speakers embellished referring expressions were not limited to issues associated with *audience design*, or adapting one's speech to one's addressee. For example, in a story telling task, Brown and Dell (1987) demonstrated that the use of referring expressions could vary in accordance with the typicality of an object in a given context. Subjects provided more explicit descriptions when describing atypical instruments, such as an *ice pick* used in a stabbing, than when describing typical instruments, such as a *knife* in a stabbing. To account for this finding, Dell and Brown (1991) proposed that certain adaptations of referring expressions need not make recourse to models of the listener but could be adequately accounted for in terms of generic language processing mechanisms.

Following a similar line of reasoning, Branigan and McLean (2003) argued that the adaptation of utterances to one's addressee must be balanced with the need to communicate effectively in a timely fashion. Using a picture verification task, they showed that speakers produced syntactic structures that reflected their own knowledge states rather than those of their addressee. This suggested that speakers do not engage in audience design for subtle aspects of their utterances, such as a choice between an active or passive syntactic structure, which are superfluous to *adequate* communication.

There is evidence to suggest that whether or not speakers take their addressee into account may depend upon the cognitive demands on the speaker's time and attention. Schober (1993), for example, showed that participants found egocentric descriptions, such as *in front of me,* easier to produce than addressee-based descriptions, such as *in front of you.* This suggested that taking the listener’s perspective into account requires cognitive effort. In support of this notion, Horton and Keysar (1996) found that when there was no pressure of time, speakers tended to describe objects moving across a computer screen in relation to contextual information that was shared between the speaker and listener (e.g. “dark” to describe an object moving across a computer screen in the context of a static light-coloured object). Yet when under pressure to initiate utterances quickly, the preference to use shared contextual information was offset. Horton and Keysar (1996) suggested that the pressure of time caused the speaker to fall back on an initial plan, which did not incorporate a model of what the addressee could be assumed to know. Rather, common ground was incorporated into a correction mechanism.

Roßnagel (2000), however, suggested that the findings reported by Horton and Keysar (1996) indicated that cognitive availability rather than common ground may have determined which information was incorporated into the speech plan. Roßnagel (2000) hypothesised that cognitive load may influence controlled processes, such as perspective-taking, that could be disrupted when the cognitive demands on working memory were increased due to increased task difficulty, for instance. In support of this hypothesis, it was found that under conditions of low cognitive load, German speakers performing a model construction task tended to describe component parts of the model (rather than use a technical term) and embellish task instructions when addressing a boy confederate but not when addressing an adult confederate. However, when the cognitive load associated with the task was high (due to increased difficulty or due to a secondary memorisation task) this difference was offset. This finding, which replicated earlier studies reported in Roßnagel (2000), supported the hypothesis that taking the perspective of one's addressee is subject to the cognitive demands on working memory. The notion that the processes underlying speech and language production may be limited by the cognitive demands on a speaker's time and attention has also been expressed more formally in The Dual Process Model proposed by Bard et al., (2000).

2.5 The Dual Process Model

Although a large body of evidence has been taken in support of the view that speakers adapt their speech to their addressee, Bard et al. (2000) criticised previous studies of spoken dialogue for a failure to explicitly manipulate manifestations of the listener's knowledge (e.g. Fowler and Housum, 1987). In order to meet this challenge, Bard et al. (2000) conducted a series of experiments which examined explicit manifestations of the listener's knowledge on articulation.

The starting point for the Bard et al. (2000) study was the objection that maintaining an accurate model of what the listener knows and perceives in spoken dialogue must place unrealistic demands on the speaker (Clark and Marshall, 1981; Stalnaker, 1978). Not only would speakers have to hold an internal account of common ground but also, such an account would need to be constantly up-dated. In order to achieve this, speakers would have to vigilantly observe for evidence of agreement and disagreement of mutual understanding, as well as make appropriate inferences based on such evidence. It seemed unreasonable, therefore, to assume that the speaker could perform such a demanding and costly task during real-time conversations.

In response to this problem, Clark and Marshall (1981) had earlier proposed that any realistic model of on-line dialogue processing must include a default option whereby the speaker could resort to an optimistic working assumption about shared knowledge. In other words, in certain circumstances, a speaker may fall back on their own knowledge as an adequate approximation of what the listener knows and perceives. The purpose of Bard et al.'s study was to investigate when such default took place. Thus a series of experiments were conducted to determine the circumstances under which speakers would resort to a speaker-centred assumption about shared knowledge. Bard et al. (2000) exploited the design of the Map Task (Anderson et al., 1991) to investigate the effect of aspects of the addressee's knowledge (such as what the listener could hear, what the listener could see and what the listener had heard) on articulatory clarity. The duration and intelligibility of first and second mentions of the names of landmarks on a map were examined in order to determine whether speakers would reduce the articulatory clarity of their speech in response to what the addressee did or did not know.

Bard et al. (2000) showed that repeated mentions of words forming referring expressions were both less intelligible and shorter in duration than introductory mentions even though the object being referred to was New to the listener but Given for the speaker. This finding suggested that speakers responded to their own knowledge rather than that of the listener. Furthermore, differences in shared information, such as experience with a map or the ability to see a particular landmark, failed to change speakers' behaviour. Speakers did not mitigate the attenuation of second mentions in these circumstances. This was inconsistent with the notion that speakers tailor their speech to their particular listener by intending each addressee to base his or her inferences not just on any knowledge but upon their mutual knowledge and beliefs or *common ground*, (Clark, 1992). In addition to this, it was found that speakers failed to articulate clearly, even though listeners provided explicit feedback to indicate that they could not see an object. Bard et al. (2000) suggested that instead of mitigating the repetition effect in response to the listener's overt feedback, speakers continued to reduce intelligibility and duration regardless of listener's responses to the initial mention. In other words, feedback from the listener did not result in recourse to a model of the listener's knowledge as suggested by Horton and Keysar's (1996) proposal that common ground be incorporated into a correction mechanism. Overall, the results of the Bard et al., (2000) experiments indicated that, at an articulatory level at least, speakers did not respond to manifestations of the listener's knowledge such as what the listener could see, what the listener had heard, or feedback from the listener.

Nevertheless, Bard et al. (2000) found that repeated mentions were shorter and less intelligible than introductory mentions regardless of which speaker first introduced the entity into the discourse. As stated in 2.3.2 of this Chapter, this finding was also consistent with the Interactive Alignment Model (Pickering and Garrod, in press) and suggested a link between speech production and speech perception.

In order to account for speakers' articulatory control of the intelligibility of words forming referring expressions, Bard et al. (2000) proposed a model of dialogue based on that proposed by Brown and Dell (1987) and Dell and Brown (1991). The basic proposition centres on the notion that speakers proceed from a model of their own knowledge around which they structure their basic message.

Then, gradually, speakers adjust their output to non-prototypical information or to the listener's needs. Bard et al. (2000) extended the model to incorporate two distinct kinds of processing; fast, automatic, priming processes, and slower inferential processes.

Priming processes are deemed to be automatic and depend on the experience of the individual speaker. They operate during any attempt to produce spoken utterances and occur within a very fast planning cycle. At least two types of articulatory priming are distinguished within this model; priming that is dependent on discourse information and priming that is dependent on the immediate linguistic context. Once introduced into a dialogue, entities can be activated (McKoon and Ratcliff, 1980), even if they are not fully grounded as mutual knowledge, by explicit interactions between interlocutors. Activation of the representation of referent objects primes their names (Mitchell & Brown, 1988) which are consequently produced faster (e.g. Balota et al., 1989). The usual result of fast speech is decreased articulatory detail. Bard et al. (2000) further proposed that *given* status per se triggered priming. For this reason, it is unimportant which speaker made the introductory mention, provided it is heard and registered. The second type of priming processes are dependent on the immediate linguistic context. Similarly words which are appropriate syntactic and semantic continuations of a sentence can be primed by the sentence context itself (see e.g. O'Seaghdha, 1997). Again the result of priming is decreased articulatory detail of the primed names.

The second group of processes are slower and more complex. These include, updating memory for dialogue events, and determining which beliefs and goals must be attributed to interlocutors on the basis of what they say and the feedback they provide. Complex processes are deemed too slow to impact on running speech on a word-to-word basis (Brown and Dell, 1987; Horton and Keysar, 1996). Furthermore, these types of processes must compete with dialogue planning for time and attention (cf. Beattie, 1981). Consequently their completion will depend on the complexity of the process itself and the time at the speaker's disposal.

In terms of the Dual Process Model, articulatory reduction occurs because priming is a fast, automatic process which takes place at the early stages of propositional construction. Models of the listener are not available at this stage. Thus the on-line control of production is sensitive to what the listener mentioned. In contrast, production control is not necessarily sensitive to listener feedback or to information relating to common ground. This needs to be interpreted via slower inferential processes. Consequently, if the speaker does not have sufficient time to complete these processes s/he would not be able to adjust speech output in response to listener knowledge, for example.

Further evidence in support of the Dual Process Model stems from the observation that what the speaker could see and whom the speaker was addressing influenced the syntactic form of referring expressions, but not the articulation of individual words forming those referring expressions (Bard and Aylett, 2001). It is known that referring expressions become syntactically simpler the more readily they can be interpreted (Ariel, 1990; Fowler, Levy & Brown, 1997; Gundel, Hedberg and Zacharski, 1993). Bard and Aylett (2001) examined the influence of listener and speaker knowledge on the form of referring expression and compared the results with the findings of their previous study on the articulation of repeated mentions of words (Bard et al., 2000).

There was no indication that aspects of addressee knowledge, such as feedback from the listener or what the listener could be inferred to know, influenced either the design of referring expressions or the articulation of individual words. In order to account for this observation, Bard and Aylett (2001) suggested that the demands of a problem-solving task, such as the Map Task, might be sufficient for controlled processes to suffer. There may not have been sufficient cognitive resources available to run the processes involved in consulting listener models and run the necessary computation to adjust the syntactic form of referring expressions. In contrast, the form of referring expression *was* found to be more sensitive than articulation to basic aspects of dialogue, such as whom the speaker was addressing and what the speaker could see. This finding was consistent with the Dual Process Model which predicts that task and memory load, should affect the design of referring expressions, but that neither should influence the articulation of individual words.

2.6 Referential Communication in a Video-mediated Conversational Setting

Much of the research on referential communication has been carried out in a face-to-face conversational setting. Little work has been conducted to investigate the impact of video mediation on referential communication. Nevertheless, there are several reasons why a video-mediated conversational setting might influence the way participants communicate with one another. This in turn, could impact on the way speakers refer to entities, or objects, in a discourse.

First, the presence of a video-link will not necessarily simulate the nature of face-to-face communication. Consequently, this could give rise to differences in the way language is used in a video-mediated conversational setting compared with a face-to-face setting. Second, the remote nature of video-mediated communication may lead speakers to feel socially distant from one another and from the communicative situation (Short, Williams and Christie, 1976). This could affect the way participants in a dialogue interact and collaborate with one another. Third, the unfamiliarity of video-mediated communication relative to face-to-face communication may lead interlocutors to engage in distinct modes of communication, such as more careful speech. Finally, it may be more difficult for interlocutors to interpret visual cues such as the direction of eye gaze, facial expressions and body posture. This may have important consequences for the process of grounding, or the establishment of mutual knowledge.

2.6.1 The Benefits of Seeing One's Partner in a Video-mediated Conversational Setting.

According to Media Richness Theory (Daft and Lengel, 1984), the "*richness*" of a communication medium is determined by the ability of that medium to support certain characteristics of communication. These include the availability of instant feedback, the use of natural language, and the capacity to transmit non-verbal information. It is not surprising then, that the addition of a visual channel to mediated modes of communication, such as videoconferencing, was originally expected to simulate the nature of face-to-face communication.

It has been argued that speech is often adequate for many communicative situations. Consequently, there may be little advantage of adding a visual channel to good quality, audio-only, communication (see Whittaker, 2003 for a review). Generally speaking, studies of mediated communication have shown little or no advantage of adding visual information to the audio channel (e.g. Ochsman and Chapanis, 1974; Sellen, 1995; Fish, Kraut, Root and Rice, 1993; Doherty-Sneddon, Anderson, O'Malley, Langton, Garrod, and Bruce, 1997). It should be noted, however, that the addition of a visual channel in mediated communication may be of greater benefit for *social* tasks, such as negotiating and bargaining, than work-oriented tasks such as problem-solving (Williams, 1977). Nevertheless, for the purpose of this thesis I will focus primarily on collaborative problem solving rather than socially oriented tasks.

With respect to co-present communication (where participants in a conversation are situated in the same physical location) certain benefits of seeing one's partner over simply hearing one's partner have been reported for structural aspects of dialogue, such as dialogue length and turn-taking. For example, Boyle, Anderson and Newlands (1994) found that co-present participants performing a collaborative problem-solving task (The Map Task) interrupted each other less often, took fewer conversational turns, and exchanged less verbal feedback to achieve the same level of task performance when they could see each other than when they could not see each other. In contrast, studies of participants performing The Map Task in mediated conversational settings (where participants in a conversation were based at remote locations) have shown that the addition of a visual channel to audio only communication did not deliver the same benefits as the ability to see one's partner in co-present communication (Doherty-Sneddon et al., 1997; O'Malley, Langton, Anderson, Doherty-Sneddon and Bruce, 1996). Similar findings were reported for a second collaborative task

(The Travel Game), in which participants interacted with a confederate "travel agent" to plan a holiday itinerary (Anderson, Newlands, Mullin, Fleming, Doherty-Sneddon and Van der Velden, 1996). When interlocutors were co-present, dialogues contained fewer words and more optional changes of plan when participants could see each other compared with when they could not. Yet no such advantage was observed for video-mediated dialogues over corresponding audio-only dialogues. A limitation of these studies, however, is that they were based on cross-study comparisons. Video-mediated and face-to-face conversational settings were not directly compared. Nevertheless, these findings suggested that communication in a video-mediated setting, compared with a face-to-face setting might be less interactive and spontaneous, with interlocutors requiring more talk to achieve the same level of performance.

2.6.2 The Relative Unfamiliarity of Video-mediated Communication

Research on video-mediated communication has largely focussed on aspects of conversation such as dialogue length, task performance and turn-taking. Yet there is some evidence to suggest that the relative novelty of conversing in a video-mediated conversational setting may influence the way words referring to objects are articulated. For example, Blokland and Anderson (1998) conducted a study to compare the articulation of initial mentions of words forming the names of landmarks on a map in video-mediated and audio only conversational settings. It was found that word tokens uttered in a video-mediated setting were more intelligible to listeners than word tokens uttered in a corresponding audio only setting. This indicated that, in mediated communication, speakers articulated the names of landmarks more clearly when they could see the other person than when they could not. This finding was surprising since it had been shown that, for co-present communication, speakers articulated landmark names less clearly, rather than more carefully, when they could see the other person compared with when they could not (Anderson, Bard, Sotillo, Newlands, Doherty-Sneddon, 1997). Blokland and Anderson (1998) suggested that the clearer articulation of words could have been a response to the novelty of video-mediated communication. It should be noted, however, that in this particular study the refresh rate for the video image was only 5 frames per second. At this rate, asynchronicity is observed between audio and the speakers lip movements. Thus, as Blokland and Anderson (1998) themselves pointed out, the effect could have been due to the poor quality of the video-link, despite the fact that the audio signal was excellent. Differences in the quality of video-mediated settings have been shown to affect the nature of communication (e.g. O'Conaill, Whittaker and Wilbur, 1993; O'Malley et al., 1996).

Further evidence in support of the notion that the novelty of the conversational setting may influence the nature of communication stems from studies of computer-mediated communication. For example, Newlands, Anderson and Mullin (2003) found that, for a computer-mediated, text-based version of The Map Task, performance was initially poor compared with a face-to-face version of the task. However, performance improved as participants gained experience with this mode of communication. Since the range of communication channels is severely restricted in text-based communication, it is not clear whether these findings will generalise to video-mediated, spoken communication. Nevertheless, the finding that differences observed between computer-mediated and face-to-face communication can be reduced or offset with greater experience of using computer-based technologies has also been reported by Walther (1994) and Kelly and McGrath (1985). This suggests that users of computer-mediated technology can adapt and overcome the limitations imposed by this type of communications technology.

2.6.3 A Sense of Social Distance

Clark (1992) identified a set of salient features of spoken conversation such as gestures, eye gaze, the fact that the participants can see and hear each other and their common surroundings without interference, and the fact that participants formulate and execute their utterances in real time. He suggested that in some conversational settings certain features, such as the ability to read eye-gaze and gestures may be restricted. This, in turn, could affect the way interlocutors communicate with one another.

In a video-mediated setting, some of these features may be less salient. For example, in many studies of video-mediated communication the video image displays only a head and shoulders view of one's interlocutor (e.g. Anderson et al., 1996; Doherty-Sneddon et al., 1997). Consequently, it may be difficult to interpret the body posture of one's conversational partner. Furthermore, in order to make eye contact with one's interlocutor, participants may need to look into a camera, or computer monitor, rather than stare directly at their partner (e.g. Anderson et al., 1996). The restricted availability of such non-verbal cues may contribute to a sense of *social distance* between interlocutors.

Social Presence Theorists (Short, Williams and Christie, 1976) hypothesised that different communications media can be distinguished by their sense of *social presence*, or salience of the other person(s) and the communicative situation. The observation that the visual channel has been shown to be more beneficial in socially oriented tasks (Williams, 1977) has been taken as an indication that visual information plays an important role in conveying a sense of social presence (Short, Williams and Christie, 1976). Conversations over the telephone, for instance, are thought to convey less social presence than face-to-face conversations (Short, Williams and Christie, 1976; Williams, 1977). Interestingly, in a field study to investigate the use of video-mediated technologies in the workplace, Fish, Kraut, Root and Rice (1993) showed that videophones were perceived to be more similar to the telephone than to face-to-face communication. Furthermore, users of the technology preferred face-to-face interactions over video-mediated interactions for certain types of tasks. This suggested that video-mediated technologies convey less social presence than face-to-face communication.

Although social presence was originally characterised as a property of the communication medium itself, the concept has been extended to incorporate social aspects of conversations such as intimacy (Biocca, 1997 cited in Tu, 2000). Intimacy is conveyed, in part, through features of body language, such as the proximity of conversational partners, smiling and leaning forward, as well as eye contact (Burgoon, Buller, Hale and de Turck, 1984). Eye gaze has also been found to play an important role in establishing engagement and interaction with another person (Heath, 1986; Kendon, 1990). Furthermore, the use of gaze is thought to be important in regulating turn taking behaviour in spoken conversation (Argyle and Cook, 1976; Kendon, 1967). Research has shown that turn taking is influenced by video-mediation. For example, in a study of multiparty conversations, Sellen (1995) found that in video-mediated dialogues participants interrupted each other less often and employed more formal handovers of turn than in face-to-face conversations. This was taken as evidence of a less interactive and more formal mode of communication in a video-mediated conversational setting.

The attenuation of turn-taking cues has also been taken as an indication of more socially distant discussions (e.g. Rutter, 1987; Sellen, 1995; Heath and Luff, 1992). Heath and Luff (1992), for example, argued that the use of non-verbal cues, such as gaze and body orientation might be less effective in establishing engagement in video-mediated communication than in face-to-face communication. Consequently, if participants in a video-mediated conversational setting feel somewhat disconnected from the communicative situation, they might compensate by using more formal techniques for turn taking.

A problem with interpreting turn-taking behaviour in video-mediated and face-to-face communication is that technical problems can also disrupt the regulation of turn taking. For example, O'Conaill, Whittaker and Wilbur (1993) conducted a study to compare multiparty face-to-face conversations with video-mediated conversations conducted over poor quality and high quality systems. The poor quality video-mediated setting suffered from transmission lags, half duplex (allowing transmission of only one speaker at a time), and poor quality video. In contrast, the high quality video-mediated setting benefited from negligible transmission delays, full duplex (two-way) audio, and broadcast quality video. It was found that, compared with face-to-face communication, conversations over the poor quality system were characterised by longer conversational turns, fewer interruptions, less overlapping speech, and fewer backchannel responses (e.g. *uh hum, right, etc.*). However, many of these differences were attenuated in the high quality video-mediated setting (compared with face-to-face conversations). Thus some of these differences could be attributed to factors such as transmission delay and poor quality audio and visual information. As Clark (1992) suggests, the less control participants have over the formulation, timing, and meaning of their actions, the more specialised their techniques may be. Nevertheless, there were still fewer backchannel responses and more formal handovers of turn, even in the high quality video-mediated conversational setting compared to face-to-face communication.

O'Connaill, Whittaker and Wilbur (1993) suggested that the greater formality in turn taking behaviour could have been due to non-directional sound which made it more difficult to locate speakers. While this may have been an important factor in group communication, it is not clear that the ability to locate a particular sound source, such as a specific speaker, would have the same effect on turn taking behaviour in two-party dialogues. Alternatively, since neither video-mediated system supported mutual gaze, this may also have accounted for the difference in turn taking between conversations in the high quality video-mediated and face-to-face settings.

A second factor, which has been found to influence more social aspects of communication, is the perception of physical distance between participants. If the participants are not co-present in the same physical space, the remoteness of the communicative situation may contribute to a sense of social distance between them. In a study to examine the effect of the perception of physical distance on social aspects of communication, Bradner and Mark (2002) showed that participants communicating via a video-link were more likely to deceive, be less persuaded by, and initially co-operate less, with someone they believed to be in a distant city than with someone they believed to be in the same city. This observation could not have been due to the medium itself, since the same technical set up was used in both the "same city" and "different city" conditions. The finding that physical proximity influences the way people collaborate with one another has also been shown by Kraut, Egido and Galegher, (1990 cited in Bradner and Mark, 2002) and reported by Olson and Olson (2000). Furthermore, Li and Mantei (1992) showed that the degree of collaboration between participants varies directly with the physical proximity of co-workers. These findings suggest that a sense of social distance could have important consequences for the way interlocutors collaborate with one another when they refer to objects in a conversation.

2.6.4 The Principle of Least Collaborative Effort

It has been suggested that a speaker in a conversation is able to reduce their individual processing effort by relying on the other speaker. Clark and Wilkes-Gibbs (1986) proposed that interlocutors follow a principle of *least collaborative effort* whereby they minimise their joint processing effort. For example, if one speaker indicates difficulty recalling a particular word, they can ask their addressee to complete the utterance. Clark and Brennan (1991) argue that, the more readily participants can monitor each other's faces, gestures and their shared environment, the more efficient the process of grounding will be. In a face-to-face conversational setting, the grounding process may be relatively easy. However, as discussed in section 2.6.3 above, the restricted use of features, such as eye gaze and body posture, in a video-mediated setting may influence the way interlocutors collaborate, which is an important aspect of the establishment of mutual knowledge, or *common ground.*

In spoken conversation interlocutors can indicate understanding by drawing on a variety of non-verbal information including smiles, gaze and head nods (Kendon, 1967; Schegloff, 1982; Yngve, 1970). If the usefulness of these cues is limited, then interlocutors will experience greater difficulty monitoring the listener for signs of misunderstanding (Clark, 1992). As stated in section 2.6.3 above, in many video-mediated set ups participants must look at the camera rather than at the image of their interlocutor to make eye contact. The restricted availability of visual signals, such as eye contact, in video-mediated conversational settings could have important consequences for the process of establishing mutual knowledge.

In support of this notion, in face-to-face communication, it has been shown that the ability to see the other person leads to shorter dialogues containing fewer words (e.g. Boyle, Anderson and Newlands, 1994). More specifically, Doherty-Sneddon et al., (1997) showed that, for co-present pairs of participants performing The Map Task, dialogues contained more verbal feedback to elicit and check understanding when participants were unable to see each other compared with when they could see each other. Doherty-Sneddon et al. (1997) argued that when interlocutors could not see each other, they compensated for the lack of visual cues by using explicit verbal devices to check for understanding. However, when dialogues performed in a high quality video-mediated conversational setting (similar to that used by O'Connaill, Whittaker and Wilbur, 1993) were compared with corresponding audio only dialogues, there was no difference in the number of words, or in the number of conversational turns. Surprisingly, when eye contact was simulated, video-mediated dialogues contained *more* words as well as *more* turns than video-mediation without eye contact, or audio-only dialogues. Nevertheless, in both video-mediated conversational settings, the Instruction Givers elicited significantly less feedback from their interlocutors than in the audio only setting. Taken together, these findings suggested that although video-mediated communication did go some way to offering a means for checking mutual understanding, visual information, such as eye contact, did not fulfil the same function as it appeared to do in face-to-face communication. One can infer from such cross study comparisons that the ability of participants to collaborate with one another to reduce processing effort will be impaired in a video-mediated setting relative to a face-to-face setting.

Differences observed in turn-taking behaviour between face-to-face and video-mediated communication (e.g. Sellen, 1995; O'Connaill, Whittaker and Wilbur, 1993) also provide evidence to suggest that the grounding process will be less efficient in a video-mediated conversational setting than in a face-to-face setting. As discussed in section 2.6.3 above, differences in turn-taking behaviour have been attributed to the reduced effectiveness of visual signals in video-mediated conversational settings. Turn-taking behaviour reflects the degree of co-ordination between participants in a conversation, with respect to determining who should speak and when (Sacks, Shegloff and Jefferson, 1974). Thus, if turn taking is disrupted in a video-mediated setting compared with a face-to-face setting, this offers a further indication that the grounding process may be more difficult and require more effort.

2.7 Summary

Studies of referential communication have indicated several basic findings with respect to the way speakers refer to objects in a conversation. First, upon repeated reference, words referring to the same object are articulated more quickly and are less intelligible to listeners (Fowler and Housum, 1987; Hawkins and Warren, 1994; Samuel and Troicki, 1998; Robertson and Kirsner, 2000). Second, referring expressions become shorter with successive use (Krauss and Weinheimer 1964, 1966; Clark and Wilkes-Gibbs, 1986). Third, speakers tend to use definite references upon repeated reference to the same object (Clark and Wilkes-Gibbs, 1986; Wilkes-Gibbs and Clark, 1992; Hupet and Chantraine, 1992). Fourth, speakers alter the way they refer to objects in accordance with the spatial perspective of the listener (Schober, 1993; Garrod and Anderson, 1987) and in accordance with what the addressee can be assumed to know (Isaacs and Clark, 1987; Fussell and Krauss, 1992). Finally, participants in a dialogue are better able to identify the objects being referred to than listeners who simply overhear a conversation (Schober and Clark, 1989).

These findings have been largely interpreted in terms of the basic assumption that, in spoken communication, one speaks to be understood. Consequently, it has been argued that any model of interpersonal communication must account for the role of the listener in spoken dialogue. Not surprisingly, then, it is those models of interpersonal communication which lay emphasis on adaptation to the addressee's perspective (e.g. Schober, 1993) or co-ordination between interlocutors (Clark and Wilkes, 1986; Clark, 1992) which have dominated the relevant literature. Yet researchers have questioned the assumption that speakers adapt their speech to the addressee's perspective (Dell and Brown, 1991; Brown and Dell, 1987; Bard et al., 2000; Horton and Keysar, 1996).

More recently it has been suggested that the processes involved in modelling what the listener can be assumed to know, or modelling the perspective of the listener may be limited by cognitive availability (Bard et al., 2000; Roßnagel, 2000). This has been formally expressed within the framework of the Dual Process Model (Bard et al., 2000). The model suggests a distinction between automatic processes, such as articulatory priming, and more complex processes, such as task planning or modelling the addressee's knowledge. Crucially, complex processes occur within a slower planning cycle and are subject to the cognitive demands on the speaker's time and attention. While the Dual Process Model has been applied to face-to-face spoken dialogue, it is not clear whether the Model will generalise to dialogue in other conversational settings such as video-mediated communication. Nevertheless, there is reason to suspect that in a video-mediated conversational setting, a feeling of social distance, the relative unfamiliarity of the technology, or the restricted availability of eye gaze and body language may influence the nature of communication relative to a face-to-face setting. This in turn may impact on the way interlocutors refer to objects in a discourse.

The purpose of this dissertation is to consider the implications of the Dual Process Model for speech and language production in spoken discourse.

Question 1: How will increased cognitive load impact on spoken output in face-to-face and video-mediated communication?

The Dual Process Model (Bard et al., 2000) implies that complex processes, such as task planning or dialogue modelling, will be influenced by the cognitive resources available to the speaker. In contrast, automatic processes, such as articulatory priming, should occur irrespective of the cognitive demands on the speech production system.

With respect to face-to-face communication, there is evidence to indicate that increases in cognitive load, due to the difficulty of a task or performing tasks simultaneously, impacts on aspects of dialogue modelling such as taking the perspective of one's addressee. This is reflected in the way speakers refer to objects within a discourse (Roβnagel, 2000). Similar findings have been reported for speakers initiating utterances under time pressure (Horton and Keysar, 1996). These findings are consistent with the Dual Process Model, which holds that the completion of complex processes involved in dialogue modelling is dependent on the time at the speaker's disposal and the demands on the speaker's attention. The model then predicts that complex processes, which are thought to include making decisions, task planning and naming objects, are likely to be influenced by increased cognitive load. In contrast, processes which are deemed to be automatic should occur irrespective of any increase in cognitive load. For example, articulatory reduction (the reduced articulatory clarity of repeated mentions of words referring to the same object) is believed to be the result of priming processes. Consequently, any increase in cognitive load should have little or no impact on articulatory reduction since priming processes are deemed to be automatic and occur within a fast planning cycle.

Question 2: How will spoken output be influenced in a video-mediated conversational setting (compared with a face-to-face setting)?

There are reasons to suspect that speakers may communicate differently in a video-mediated conversational setting compared with a face-to-face setting.

First, the relative unfamiliarity of a video-mediated conversational setting compared with a face-to-face setting could lead to differences in certain aspects of articulation. Blokland and Anderson (1998) showed that initial mentions of words referring to objects (landmarks on a map) were articulated more clearly in a video-mediated setting compared with a corresponding audio only setting. This suggested that the speed of articulation of individual words may differ according to the conversational setting. Furthermore, in face-to-face communication, speakers tended to speak *less* clearly rather than *more* clearly when they could see their partner compared with when they could not (Anderson, Bard, Sotillo, Newlands, Doherty-Sneddon, 1997). This offers some suggestion that the relative novelty of video-mediated communication compared with face-to-face communication may influence articulatory control. Nevertheless, an investigation of the effects of priming requires an examination of *repeated* mentions of words relative to introductory mentions of words

A second possibility is that interlocutors may experience a sense of social distance in a video-mediated setting (Short, Williams and Christie, 1976). Evidence suggests that the limited availability of non-verbal cues, such as eye-gaze and body language, together with the remoteness of the communicative situation, may give rise to a sense of social distance between interlocutors (e.g. Doherty-Sneddon et al., 1997; Bradner and Mark, 2002). The restricted availability of non-verbal cues in a video-mediated setting may cause difficulty in establishing common ground (Clark and Brennan, 1991). Eye-gaze is also thought to play a useful role in regulating turn taking. Research has shown that turn taking behaviour, which reflects the degree of co-ordination between participants, can be disrupted in a video-mediated conversational setting (e.g. Sellen, 1995; O'Connaill, Whittaker and Wilbur, 1993). This suggests that interlocutors in a video-mediated conversational setting may experience more difficulty co-ordinating their behaviour relative to interlocutors conversing in a face-to-face setting. Thus difficulty in co-ordinating behaviour may lead speakers to become less collaborative and less aligned in the way they refer to objects in a discourse.

In addition to this, the Dual Process Model (Bard et al., 2000) suggests that complex processes, such as collaborating with one's partner, as well as processes such as naming an object, and task planning, will be more sensitive to conversational setting than automatic processes, such as articulatory priming. According to Bard et al, (2000) priming processes are deemed to operate within a short planning cycle and are dependent on the experience of a single individual. Since they are not dependent on the demands of the task and the time at the speaker's disposal, priming process could be robust to any effect of video-mediation (as well as cognitive load). In contrast, complex processes are governed by slower inferential processes, which operate over a longer planning cycle. Thus there is more time for factors, such as a response to the conversational setting, to make their impact felt.

Question 3: Will cognitive load have an equal impact on spoken output in face-to-face and video-mediated communication?

It is possible that attributes of video-mediated technology may serve to increase the demands on the speaker. For example, if participants perceive the mode of communication as novel, or experience difficulty in interpreting non-verbal signals from their interlocutor, then this could add to the cognitive demands on the speaker. If this is the case, then any effect of cognitive load would be expected to be greater in a video-mediated conversational setting than in a face-to-face setting. However, following the line of reasoning implicit in the Dual Process Model, any difference in the impact of cognitive would be expected to be observed for those processes which are complex in nature and place greater demands on the speaker's time and attention. Thus any increase in cognitive load is expected to influence processes, such as taking decisions regarding task planning, the process of grounding, and naming an object. Yet those processes which are deemed to be automatic and occur too quickly for other factors to set in, should occur irrespective of any combined effect of cognitive load and conversational setting.

Exploration of the questions outlined above will constitute the main theme of the following chapters.

Chapter 3

The Articulatory Quality of Words in Video-mediated Communication

3.1 Introduction

The main objective of this dissertation is to examine spoken output in video-mediated and face-to-face dialogues in order to gain a better understanding of the underlying cognitive processes involved in speech production. The study of spoken dialogue presents a challenging area of research since the advance of computer-based technologies in today's society presents speakers with novel conversational settings such as those presented by video-conferencing systems. This raises interesting questions regarding the role of the setting in which a dialogue takes place. Research in face-to-face communication has shown that speakers tend to attenuate the articulatory quality of repeated mentions of words referring to the same object. Will speakers behave in the same way when communicating in a video-mediated conversational setting? This chapter aims to address this question through the presentation of an initial study which investigated word duration in video-mediated dialogues. In section 3.1, I will outline the background and motivation for Study 1. In section 3.2, I will describe the methodology used. In section 3.3, I will present the results of Study 1 and discuss the implications of the findings in section 3.4.

3.1.1 Articulatory Control in Face-to-face Communication

It is well known that speech is subject to enormous variability to the extent that no two utterances are ever articulated in quite the same way. In the relevant literature, two measures of articulatory quality are often reported. The first is a psychological measure based on the intelligibility of words to listeners. In short, words extracted from running speech are presented to listeners for identification (but see Lieberman, 1963 for details). The number of correctly identified words provides a measure of *intelligibility*. For example, Pollack and Pickett (1963) showed that words uttered in isolation (and with no contextual cues therefore) tended to be fully recognisable, while only 50% of listeners could identify individual words taken from running speech. Since no clues could be obtained from the context, it can be inferred that words which are less intelligible to listeners have been articulated less clearly than words that are more easily identifiable. A second measure of articulatory quality is provided by the temporal *duration* of word tokens. While intelligibility is often taken as a measure of how *clearly* words are articulated, *duration* provides a measure of how *quickly* words are articulated.

The articulatory quality of words has been shown to depend on information beyond the acoustic signal itself. The identity of a word can be partially specified by grammatical and semantic information derived from the sentence context or by the discourse status of the entity being referred to (Bard and Anderson, 1983; Fisher and Tokura, 1995; Fowler, 1988; Fowler and Housum, 1987; Hawkins and Warren, 1994; Hunnicutt, 1985; Lieberman, 1963; Samuel and Troicki, 1998). Fowler and Housum (1987) further demonstrated that listeners were able to make use of the reduction of words as a cue for determining that the object being referred to had already been introduced into the discourse. In addition to this, listeners were faster at making judgements when presented with second occurrences of words in a text compared with first occurrences. Observations such as these have led to the assumption that an account of speech production in spoken dialogue must include a speaker model of what the listener can be assumed to know (e.g. Bolinger, 1965, 1981; Chafe, 1974; Lindblom, 1990). If a word can be partially specified by sentence or discourse context, then the speaker can assume that the listener would not require as clear a signal as would otherwise be the case. Consequently, the speaker produces words which are articulated less clearly and more quickly in more redundant contexts.

3.1.2 Building Common Ground

One interpretation of the attenuated articulation of repeated mentions of words (articulatory reduction) is based on the traditional belief that the speaker adapts their speech to the perceptual needs of the listener (Schober, 1993). Extending this view, it has been suggested that speakers tailor their speech to the comprehension needs of the addressee by taking into account not just the addressee's knowledge but the knowledge and beliefs that can be assumed to be shared between interlocutors - their *common ground* (Clark, 1992). The building of common ground relates to the way entities and objects in a discourse become mutual knowledge. It has been suggested that entities may become mutually known or grounded by virtue of previous mention. Once mentioned, the entity can then be taken as *given* within the context of the discourse. According to Prince (1981), *given* information has been primed, or readied for use in discourse. Extending this line of reasoning, the attenuated articulation of repeated mentions of words has been attributed to priming processes, which result in faster articulation and reduced articulatory detail. Consistent with this view, is the observation that this *givenness effect* does not occur when the 2nd word refers to a *new* item of the same sort (Bard, Lowe and Altmann, 1989). This suggests that, while previous mention of given entities does degrade the articulatory quality of words, previous mention without givenness does not. Fowler and Housum (1987) claim that words which mention given information can be interpreted by the listener in the context of that information. Consequently, the speaker can articulate words denoting given entities more quickly and less clearly.

3.1.3 Modelling Spoken Discourse

Dell and Brown (1991) suggested that in order to understand the processes underlying speech production, it is necessary to distinguish between two types of variation. While certain types of variation in spoken output did appear to denote genuine adjustment to a particular listener, others types of variation were more ambiguous. Thus, for example, the observation that adults adjust the linguistic and acoustic features of their speech when addressing young children (e.g., Newport, Gleitman, and Gleitman, 1977; Snow, 1972; Snow and Ferguson, 1977) appears, by definition, to implicate some kind of model of a particular listener (but see Bard and Anderson, 1983 for an alternative explanation of parental speech). Yet other speech phenomena, such as the articulatory reduction of words that can be partially specified by their sentence or discourse contexts (e.g. Lieberman, 1963; Fowler and Housum, 1987), could be accounted for in terms of generic language processes.

Dell and Brown (1991) do not reject outright, adaptation to the listener as an account of phenomena observed in spoken output. In fact, they provide evidence in support of this position. In a series of experiments relating to the way instruments of actions (e.g. a knife used in a stabbing) are specified when retelling a story, Brown and Dell (1987) showed that speakers were more explicit in their descriptions of atypical instruments of an action than typical instruments. For example, an ice pick used in a stabbing was mentioned more often than a knife, when subjects retold the story to a confederate. They further demonstrated that the effect of typicality was dependent on structural and pragmatic information including knowledge of the listener (but see Brown and Dell, 1987 for details). This indicated that the speaker did consult a model of what the listener knows.

Although an explanation based on consultation with a model of the listener could account for certain separate clause specifications, it did not provide an adequate explanation of observations of within-clause structures. Brown and Dell (1987) showed that when the instrument was specified in the same clause as the verb (e.g. The robber stabbed the man with a knife), the language production system rigidly adhered to its tendency to explicitly mention atypical instruments more than typical instruments, regardless of knowledge from the listener. Control experiments had indicated that an informative picture aids readers in overcoming comprehension difficulties. Thus it would seem reasonable to assume that the speaker's model of the listener should have included what the listener knew from the picture. In the light of a picture of a robber with an ice pick, the listener could readily infer that the ice pick was the instrument used in the stabbing. If the speaker consulted a model of the listener, the typicality effect should have attenuated when the speaker knew that the listener had an informative picture of the instrument. In order to account for their results, Dell and Brown (1991) proposed that, for within-clause structures, some feature of the language production system itself brought about the typicality effect. Consequently, beliefs about listener-knowledge would have no bearing on speakers' tendency to mention atypical instruments over typical instruments. They hypothesized that the language production system possesses its own devices whose operations may roughly approximate that of consultation with listener-knowledge. Thus generic listener adaptations are deemed to be "automatic products of the production system" (Dell and Brown, 1991:107). Dell and Brown (1991) objected that Fowler and Housum's (1987) experiments did not provide a clear demonstration of active consultation of a model of the listener. For example, the articulatory clarity of words has been shown to degrade for repeated mentions of words denoting given entities (e.g. Fowler and Housum, 1987) and for words in highly redundant sentence contexts, such as the word *nine* in the context "*A stitch in time saves ______*" (Lieberman, 1963). It had been suggested that these observations could be accounted for in terms of priming processes analogous to those proposed to account for intralexical priming (Balota, Boland & Shields, 1989). Intralexical priming is deemed to be the result of automatic spreading activation between related lexical items. The rate of production is related to the activation level of the items. Thus an item receiving extra activation from a related one in the same sequence can be produced faster. It has been assumed that such an effect denotes an adaptation to the listener on the grounds that words in primed contexts are easier to understand. Thus the speaker does not need to produce as clear a signal. Brown and Dell (1991), however, suggested that automatic spreading activation was effected by the production system itself, independently of the speaker's beliefs about the listener.

3.1.4 The Dual Process Model

Dell and Brown's proposal that generic processes may underpin articulatory reduction effects has been taken up by other researchers. Bard, Anderson, Sotillo, Aylett, Doherty-Sneddon and Newlands (2000) criticized previous research for a failure to explicitly manipulate listener knowledge. To investigate the impact of listener knowledge on the intelligibility and duration of words in spoken dialogue, Bard et al., (2000) conducted a series of experiments that explicitly manipulated several aspects of listener knowledge. They found that while speakers responded to what the listener had mentioned, they appeared to be insensitive to what the listener had heard or to feedback from the listener. Repeated mentions of words were articulated less clearly than introductory mentions of words regardless of which speaker introduced the item. Yet speakers also attenuated the articulation of words, even when the listener could not be assumed to have known the object being referred to - either because they could not see it, or because they had not heard it mentioned. This type of information should have entered a model of the listener. Bard et al., (2000) reasoned that if speakers articulate clearly enough to meet the perceptual needs of the listener, they should have mitigated the attenuation of repeated mentions of words in these cases.

In order to account for their findings, Bard et al., (2000) extended the model put forward by Dell and Brown (1991). They proposed a "Dual Process Model" of speech production that distinguished two basic types of process. The first of these are priming processes (Balota, Boland & Shields, 1989; Mitchell and Brown, 1988) which are deemed to be fast, automatic, and the result of the speaker's own recent experience. A second group of processes, which I will call *complex processes* for convenience, occur over a longer planning cycle and are deemed too slow to precede every attempt at speech production. These processes, which include, constructing a model of the listener, drawing inferences, and making decisions, compete with each other for time and attention. In running speech there may not be the necessary time for these processes to make their impact felt on a word-to-word basis. The processes involved in modelling what the listener can be assumed to know must compete for system resources with computations which support planning a dialogue or tracking a shared task, Thus, in running speech (where utterances must be formulated and produced in real-time) inferential processes may suffer, leaving the speaker with only cost-free defaults in the form of his or her own experience. Speakers articulated repeated mentions of words referring to the same entity more quickly and less clearly regardless of which speaker introduced the entity into the dialogue. This prompted Bard et al., (2000) to suggest that priming could be triggered by *given* status per se. If given status was the prime, it was therefore unimportant which speaker introduced the new entity into the discourse.

The Dual Process Model was also able to explain why speakers responded to what the listener said but not to information, such as feedback from the listener, that should have entered a speaker model of the listener. Listener-mention may confer given status on an entity and consequently trigger priming. Feedback from the listener, on the other hand, needs to be interpreted via complex processes, which are planned over a longer planning cycle. As a result, there may not be the time or available resources to run the processes involved in interpreting feedback and make the necessary articulatory adjustment to mitigate priming.

The Dual Process Model had been proposed on the basis of observations of face-to-face communication, but the purpose of this dissertation is to explore the influence of different conversational settings on spoken output. How would the processes underlying speech production be affected by video-mediation? Will video-mediated communication function in the same way as face-to-face communication, or will the control of word articulation pattern differently in a video-mediated conversational setting?

3.1.5 Video-mediated Communication

There were several reasons to suspect that interlocutors might communicate differently in a video-mediated conversational setting compared with a face-to-face setting. These were discussed in section 2.6 of Chapter 2. A brief summary is provided here for convenience. Since video-conferencing systems enable speakers to see each other, it was originally hoped that video-mediated technologies would approach face-to-face interaction in effectiveness and efficiency. Yet, Social Presence Theorists (e.g. Short, Williams and Christie, 1976) suggest that communications media can be distinguished by their degree of *social presence*, or salience of the other person and the communicative situation. Thus, computer-based communication should only function in the same way as face-to-face communication insofar as that medium is able to mimic the features of face-to-face communication in terms of being expressive, interactive and focussing attention on personal attributes. In a video-mediated conversational setting, the use of gaze and body posture may be restricted, particularly where a head and shoulders image of the interlocutor is presented. This may lead speakers to become less engaged with one another. Clark and Brennan (1991) further suggest that if the set of features associated with a conversational setting are somewhat reduced compared with a face-to-face setting, participants may experience difficulty in establishing common ground. Alternatively, the technology itself may induce some kind of novelty effect (Blokland and Anderson, 1998), and this may impact on the communication process.

Consistent with such views, studies of video-mediated communication have found that the ability to see the other person does not appear to offer the same benefits as the ability to see the other person in face-to-face communication (Doherty-Sneddon et al., 1997; O'Connaill, Whittaker and Wilbur, 1993; O'Malley et al., 1996; Sellen, 1995). Generally speaking, face-to-face dialogues tend to be shorter and contain fewer words when compared with a corresponding audio-only condition, yet the same length advantage was not observed for video-mediated dialogues over corresponding audio only dialogues (e.g. Doherty-Sneddon et al., 1997).

Much of the research, aimed at drawing comparisons between face-to-face and video-mediated communication, has focussed on aspects of communication such as dialogue length, turn-taking and interruptions (e.g. Sellen, 1995;) with very little research on the articulatory quality of words. Nevertheless, a study of word articulation in video-mediated and audio-only contexts would seem to suggest that the ability to see the other person in a video-mediated context is not sufficient to produce the same pattern of behaviour observed in face-to-face communication. In studies of co-present communication (as opposed to remote communication mediated by technology) speakers tended to produce word tokens that were less intelligible and shorter in duration when participants could see each other compared with when they could not (Anderson, Bard, Sotillo, Newlands, Doherty-Sneddon, 1997). Yet Blokland and Anderson (1998) found that introductory mentions of words forming the names of landmarks on a map were more intelligible to listeners and of the same duration when uttered in a video-mediated conversational setting compared with an audio only setting. This would seem to suggest that, in terms of word articulation, speakers may articulate more carefully in a video-mediated conversational setting than in a face-to-face setting.

However, in the Blokland and Anderson (1998) study, only *introductory* mentions of word tokens were examined. Consequently, it remains unclear how the articulatory quality of *repeated* mentions of words will be influenced in a video-mediated conversational setting. Will speakers shorten repeated mentions of words as in face-to-face communication? Or, will communication in a video-mediated setting lead speakers to mitigate articulatory reduction? In face-to-face communication, it has been shown that speakers reliably shortened repeated mentions of word tokens except in response to gross aspects of listener knowledge, such as the introduction of a new listener to the task (Bard et al., 2000). However, the effect seemed resistant to more subtle aspects of listener knowledge, such as what the listener could see or what the listener offered in the way of feedback. Of particular relevance to the present study, speakers did however, respond to what the listener mentioned. Articulatory reduction occurred irrespective of whether the same speaker or different speakers articulated repeated mentions of word tokens. It is this latter effect which is of particular interest here. Will speakers exhibit the same pattern of behaviour in a video-mediated conversational setting?

Video-mediated communication may differ from face-to-face communication in that aspects associated with this conversational setting, such as a sense of social distance or any novelty value, may lead speakers to be less sensitive to their listeners needs. If speakers feel socially distant from their interlocutor and from the communicative situation, they may resort to an egocentric model of the dialogue and attribute some special status to their own discourse. In this case, speakers may shorten their own repetitions more than those introduced by the other person. Speakers may anticipate difficulty on the part of the listener due to the novelty of a video-mediated conversational setting, or in response to social distance and mitigate the shortening of repeated mentions of word tokens.

If on the other hand, video-mediated communication functions in the same way as face-to-face communication, then we would expect speakers to shorten repeated mentions of word tokens regardless of which speaker introduced the item (Bard et al., 2000). If the Dual Process Model is correct, then given status (achieved by previous mention) should trigger priming processes and result in the greater speed of articulation regardless of which speaker introduced an entity into the dialogue. The Dual Process Model predicts that priming should be unaffected by higher level cognitive factors, such as a feeling of social distance or a novelty effect, since priming processes are deemed to be automatic and occur too quickly for such factors to make their impact felt on a word-to- word basis. Thus we would expect to observe the same pattern of results observed by Bard et al., (2000) in face-to-face communication.

The purpose of this present study is to determine whether, in common with face-to-face communication, speakers communicating via an audio-visual computer link would be equally sensitive to listener- and speaker mention.

3.2 Methodology

Before describing the methodology used for this study, I will first deal with several methodological issues.

3.2.1 Methodological Issues

Intelligibility v Duration

As stated in section 3.1.1, studies of spoken dialogue often report two measures of articulatory quality, namely, *intelligibility* and *duration.* It is often thought that the *intelligibility* of words provides a measure of the articulatory clarity of words while *duration* provides a measure of the speed of articulation. Yet these aspects of word articulation are not unrelated. It takes time for the speaker to make the articulatory movements involved in speech production. Thus, as pointed out by Lieberman (1963), differences in duration reflect, in part, the preciseness of articulation.

Nevertheless, measures of intelligibility and duration do not always yield the same pattern of results. For example, Blokland and Anderson (1998) found that introductory mentions of words uttered in a video-mediated conversational setting were more intelligible to listeners than words uttered in a corresponding audio only setting. Yet there was no indication that these words were articulated more quickly since there was no significant difference in duration between "video-mediated" word tokens and "audio only" word tokens. One reason for this may reside in the fact that duration is an acoustic measure of speech whereas intelligibility is a perceptual measure of speech (since word tokens are presented to listeners for identification). The relationship between speech production and speech perception is complex. In addition to this, a contributory factor to intelligibility is amplitude. In fact, the experimenter in the Blokland and Anderson (1998) study informally observed that participants in the experiment appeared to be shouting at the video-window. Amplitude was significantly greater in the video-mediated setting than in the audio only setting indicating that the speakers did indeed speak more loudly in the video-mediated condition. Some studies of the phonetics of speech suggest that the amplitude of the speech segment, particularly that associated, with the central vowel in a syllable, contributes to the intelligibility of a word (e.g. Lively, Pisoni, Vansummers and Bernacki, 1993). This would explain why duration and intelligibility measures do not always yield the same pattern of results (e.g. Blokland and Anderson, 1998).

In short, measures of intelligibility and duration both reflect articulatory clarity to some extent. While duration offers a measure of the speed of articulation, intelligibility offers a perceptual measure of articulatory clarity which can be influenced by the amplitude of the speech signal. For the purpose of the current research, I am specifically interested in speech output rather than the perception of speech and also in how quickly speakers articulate. Thus duration is the preferred measure here.

Technical Set-up

When interpreting the results of video-mediated communication the technological set-up must be taken into account. Studies have shown that communication can be affected by the quality of the videoconference link, the frame rate, audio and video synchronization, or different configurations of videoconferencing systems (Barber and Laws, 1994; O'Conaill, Whittaker, Wilbur, 1993; O'Malley et al., 1996). This highlights the need to use high quality video-mediated links in this type of work. For the purpose of the research reported in this doctoral thesis, a high quality video-link, which supports full duplex audio will be employed.

Spontaneity of Spoken Dialogue

Several studies reported in this chapter have in fact been based on read speech (e.g. Lieberman, 1963; Fowler and Housum, 1987) rather than on spoken dialogue. A disadvantage of these studies is that the addressee is not engaged in an interaction and read speech does not offer the ecological validity of natural speech. Spontaneity is an important feature of spoken dialogue. For this reason, The Map Task (Brown et al., 1984; Anderson et al., 1991) will be used for the studies reported in this and in subsequent chapters. The Map Task is a collaborative problem-solving task which elicits natural speech in spontaneous and unconstrained dialogue. The maps are schematic and show a start point and a number of landmarks. The participants' maps are not identical and include shared and unshared landmarks. It is suitable for examining the duration of introductory and repeated mentions of word forming the names of landmarks. The task itself elicits a large data set of repeated references to landmarks in relatively similar sentence contexts. The goal of the task is for one participant, an Instruction Giver to instruct the other participant, an Instruction Follower, to accurately reproduce the route on their map.

3.2.2 Method

Materials

The materials for this study were drawn from a subset of data collected for a previous experiment (Anderson, et al., 1999). The original experiment included 48 two-party dialogues of participants performing The Map Task via a high quality videoconference link. The 48 undergraduates, recruited from the University of Glasgow and the University of Nottingham, did not know each other prior to taking part in the experiment. The participants performed 2 versions of The Map Task, changing roles as Instruction Giver and Instruction Follower. Each participant sat at a workstation which displayed an image of the map (6.3" x 7.85") and a video window of the other participant (3.5" x 4.5"). The workstations processed and sent video images across the network to the other workstations. An asynchronous Transfer Mode (ATM) network directly connected the workstations. The connection between Glasgow and Nottingham was made via the SuperJANET ATM network and the video images were delivered across the network at 25 frames per second.

In the present study 40 of the 48 two-party dialogues originally recorded were used for analysis. Four pairs of dialogues were rejected for the following reasons. The recording levels were very low for one pair of dialogues. For a second pair of dialogues, part of the conversation had not been recorded. For a third pari of dialogues, introductions to landmarks were made before the start of the task. For a fourth pair of dialogues one of the speakers was thought to be a non-native speaker of English.

Design

The design used in this study is similar to that used by Bard et al., (2000). However, in the present study, the materials were drawn from fewer dialogues, thus in order to maximise the sample, words forming the names of both shared and unshared landmarks were analysed. In the Bard et al., (2000) study only shared landmarks were analysed. In the present study, all word tokens were uttered as part of the names of landmarks on the map. All occurred in both the first and the second mentions of the same landmark within a single dialogue. First mentions of word tokens were compared with second mentions of words in *same-speaker* and *different- speaker* repetition conditions. In the *same-speaker* condition, word pairs were introduced and repeated by the same speaker, as illustrated in dialogue extract A. Landmark names are shown in italics.

Dialogue extract A

Speaker A: So the start point is just north of the *telephone booth*

Speaker B: Aye got it

Speaker A: OK Go slightly to the west of the *telephone booth*

In the *different-speaker* condition, the landmark was introduced by one speaker and repeated by the other speaker, as illustrated in dialogue extract B.

Dialogue extract B

Speaker A: Keep going until you reach the *stone circle*

Speaker B: mhm I don't actually have a *stone circle* here

Speaker A: You don't

Speaker B: No

Of the 649 repeated mentions, 364 pairs of word-tokens were repeated by the same speaker and 285 were repeated by a different speaker. The dependent variable was word duration measured in milliseconds.

Procedure

The dialogues were recorded onto a Viglen personal computer and analysed using the Syntrillium speech analysis software package, "Cool Edit". Recordings were digitised at a sampling rate of 16Hz. Word onsets and offsets were determined by examining spectrogram and time-amplitude waveform displays and by listening to the results of the recordings. The duration of the words was measured in milliseconds from the onset to the offset of the word.

Dependent variable

In Bard et al., (2000) the raw duration of words were normalised following a technique devised by Campbell and Isard (1991). This involves comparing words in a context with a "citation form". However, in the materials used in this experiment, no citation forms were available. Previous studies have been based on the raw duration of word tokens and the effect of articulatory reduction was observed (Fowler and Housum, 1987). This therefore is the method employed here. Thus the raw duration of words is used for analysis.

3.3 Data Analysis

3.3.1 Articulatory Reduction

The data were organised by item and by speaker. The results for the data organised by speaker are presented first. The mean duration (in milliseconds) for first and second mentions of lexical items uttered by the same and by different speakers are shown in Figure 3.1.

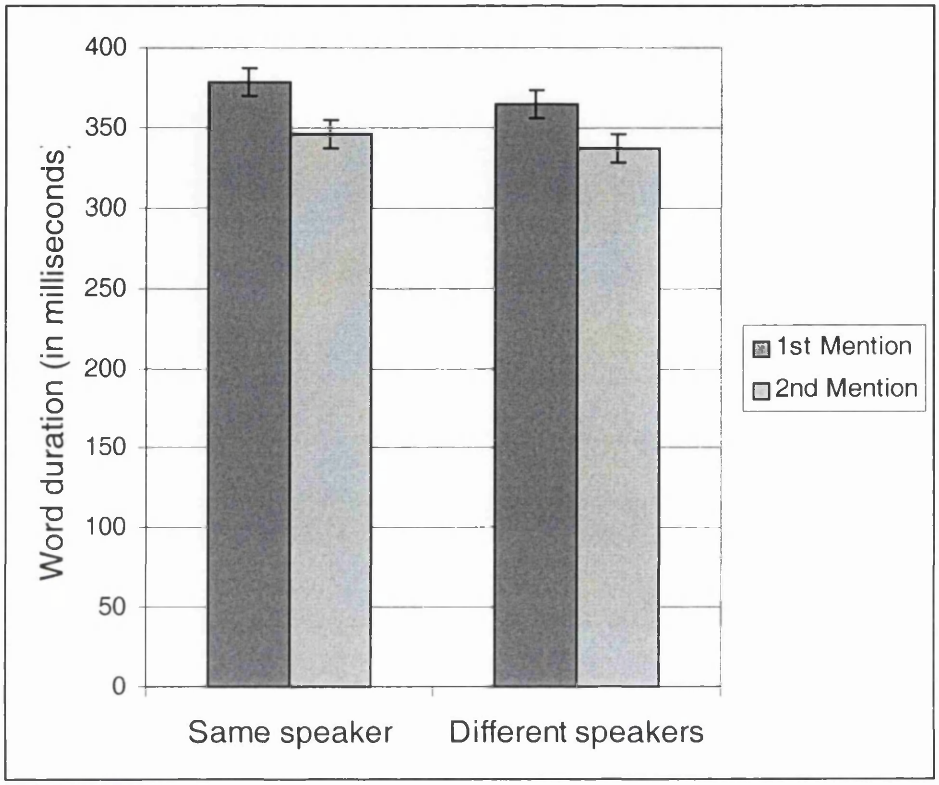


Figure 3.1. Mean word duration (±SE) for 1st and 2nd mentions of word tokens uttered by the same speaker (N = 80) and by different speakers (N = 80).

As can be seen from figure 1, the mean word duration for 2nd mentions (M2) of word tokens was shorter than 1st mentions (M1) for pairs of word tokens uttered by the same speaker (mean M2 duration = 346; mean M1 duration = 379) and by different speakers (mean M2 duration = 337; mean M1 duration = 365). The results of the analysis of variance indicated that the differences observed in word duration were reliable.

A 2 x (2) analysis of variance (by speaker) was carried out on the data with *mention* treated as a within-subjects factor and *speaker repetition* treated as a between-subjects factor. The results of the analysis of variance are shown in table 3.1 below.

Table 3.1. Analysis of variance (by speaker) for word duration

Source	df	F	Probability
Between Subjects			
Speaker Repetition (S)	1	0.96	0.33
S within-group error	78	(5457.17)	
Within Subjects			
Mention (M)	1	22.98	*<0.01
M x S	1	0.20	0.65
M x S within-group error	78	(1556.23)	

Note. Values enclosed in parentheses represent mean square errors.

As can be seen from table 3.1, there was a main effect of mention [F(1, 78) = 22.98, p < 0.01], no effect of speaker repetition (F < 1) and no interaction (F<1). Overall, second mentions of word tokens were shorter in duration (342 ms) than first mentions (372 ms) regardless of whether the introducer and repeater were the same speaker or different speakers.

The data were also analysed by item. Following Bard et al., (2000), individual word tokens were employed as the unit of analysis. The mean duration (in milliseconds) for first and second mentions of words forming landmark names uttered by the same and by different speakers are shown in Table 3.2 below.

Table 3.2. Mean duration (in milliseconds) with standard deviation of 1^{st} and 2^{nd} mentions of words forming landmark names uttered by the same speaker and by different speakers (for data organised by item).

		Mention			
		M1		*M2*	
Speaker Repetition	N	M	SD	M	SD
Same	364	378	146	342	132
Different	285	364	130	330	124

As can be seen from table 3.2, the mean word duration for 2^{nd} mentions (M2) of word tokens was shorter than 1st mentions (M1) for pairs of word tokens uttered by the same speaker (mean M2 duration = 342 ms; mean M1 duration = 378 ms) and by different speakers (mean M2 duration = 330 ms; mean M1 duration = 364 ms). The results of the analysis of variance indicated that the differences observed in word duration were reliable.

A 2 x 2 analysis of variance (by item) was carried out on the data with speaker repetition and mention treated as between-subjects variables. The results of the analysis of variance are shown in table 3.3 below.

Table 3.3. Analysis of variance (by item) for word duration

Source	df	F	Probability
Between Subjects			
Speaker Repetition (S)	1	1.68	0.20
S within-group error	647	(31511.03)	
Within Subjects			
Mention (M)	1	88.49	*<0.01
M x S	1	0.09	0.76
M x S within-group error	647	(4398.38)	

Note. Values enclosed in parentheses represent mean square errors.

There was a main effect of mention [F(1,647) = 88.49, p < 0.01]. Overall, 2nd mentions of word tokens were shorter in duration (336 ms) than 1st mentions (371 ms). There was no significant effect of speaker repetition, and no interaction (F < 1).

In order to ensure reliability of measure, word duration was re-measured for 10 randomly selected dialogues (5 face-to-face and 5 video-mediated). The two sets of measures were found to be highly correlated (Pearson r = 0.988; N = 122). This indicated that the duration of words had been measured reliably.

3.3.2 Articulatory Reduction: Word-related and Discourse-related Factors

Multiple regression analyses were also carried out on the data to explore the effect of both word-related and discourse-related variables on articulatory reduction.

Dependent variable

The dependent variable was articulatory reduction, measured as the mean difference in duration between 1st and 2nd mentions of word tokens. This allowed for the exploration of a series of factors on articulatory reduction rather than on word duration per se.

Independent variables

The independent variables can be grouped into characteristics of the target words, *word-related variables,* and features pertaining to the dialogue itself, *discourse-related variables.*

Word-related variables

M1 duration: This interval scale variable is a measure in milliseconds of the duration of word tokens, as uttered in the first instance. Longer words should have more scope for shortening. Thus, the length of a word should be a good predictor of articulatory reduction, which is expected to be greater for longer words.

Monosyllabicity: This is a dichotomous dummy variable, which codes the fact that a word is monosyllabic. Polysyllabic words constitute the reference category. Linguists hold that certain syllables in words are marked for stress, or prominence. Chafe (1974) suggests that this is achieved, in part, by lengthening certain syllables in a word relative to others. It follows from this that articulatory reduction should be greater for monosyllabic words relative to polysyllabic words. Where a word contains only one syllable, it must be stressed and longer, therefore, in duration.

Discourse-related variables

Given that the analysis of variance indicated that articulatory reduction occurred irrespective of speaker-repetition (whether both mentions of word tokens were uttered by the same speaker of by different speakers) this variable was considered to be irrelevant and was not included in the model.

Task familiarity: This variable is a dichotomous dummy variable which codes the fact that speakers are conducting the task for a second time. Instances where speakers complete the task for the first time constitute the reference category. When listening to the dialogues, it was felt that the participants appeared to "settle" when performing the Map Task for the second time. The first time participants performed the task, they appeared to behave as though their maps were the same, even though they had been clearly instructed that the Instruction Giver's map and the Instruction Follower's maps differed in that not all landmarks appeared on both. On the basis of this observation, one might speculate that when performing the task for the second time, speakers would articulate words more slowly on repetition, as they become aware that the landmarks do not appear on both maps. If this is the case, then task familiarity should have a negative effect on articulatory reduction, effectively reducing the difference in duration between first and second mentions of word tokens.

Turn distance: This is an interval scale variable that was obtained by counting the number of speaker turns between the first and second mentions of word tokens. There is some evidence to suggest that articulatory reduction can be offset under certain circumstances, such as a change in topic for instance (Robertson and Kirsner, 2000). Thus, the distance, in terms of the number of conversational turns, between first and second mentions of word tokens might be expected to have a negative effect on articulatory reduction.

Speaker role: This variable refers to whether the introductory mentions of word tokens were uttered by the Instruction Giver or the Instruction Follower. The "Instruction Follower" serves as a reference category which is compared with the category "Instruction Giver". The Instruction Giver plays a more dominant role in the Map Task. Anderson and Boyle

(1994), for example, found that Instruction Givers tend to use more questions to introduce landmarks into the discourse than Instruction Followers. One might speculate, then, that the Instruction Giver may be more attentive than the Instruction Follower to the introduction of new information or objects into the discourse. Consequently, task role might have a positive effect on articulatory reduction. In other words, articulatory reduction may be stronger for Instruction Givers than for Instruction Followers.

Gender conflict: In Study 1, pairs of participants were either the same gender or mixed gender. The category "mixed gender" refers to words in dialogues where one speaker was male and the other female and is compared with the reference category "same gender" where the speakers were both male or both female. Sociolinguists have hypothesised that men and women differ in their conversational styles (e.g. Tannen, 1994). This might lead one to speculate that conversational styles may conflict where the genders of the participants differ. It is possible that in this case, speakers may be less attentive to introductory mentions by the listener given a conflict in conversational styles. If this were the case, then articulatory reduction might be attenuated for *mixed gender* pairs compared with *same gender* pairs.

The results of the multiple regression analyses are shown in table 3.4 below.

Table 3.4. Regression of word-related and discourse-related factors on articulatory reduction.
Dependent variable: M1/M2 duration difference (in milliseconds)

	Model 1		Model 2	
Independent variables	t-ratio	Pr|t>0|	t-ratio	Pr|t>0|
Constant	-3.76	<0.01	-6.59	<0.00
Word-related factors				
M1 duration	12.73	<0.00	14.27	<0.00
Monosyllabicity			6.33	<0.00
Discourse-related factors				
Task familiarity	-2.82	<0.00	-2.78	<0.01
Turn distance (M1-M2)			-0.14	<0.89
Speaker role			0.09	0.93
Gender conflict			0.74	0.46
N = 649	$r^2 =$	0.205	$r^2 =$	0.253

Model 1 in table 3.4, indicates that *M1 duration* and *task familiarity* (i.e. completion of the task for the second time) were found to be good predictors of articulatory reduction (as measured by the difference in duration between first and second mentions of word tokens). Model 1 is significant (F = 83.32, $p < 0.001$) and accounts for approximately 20% of the variance in the data ($r^2 = 0.205$). At first sight, it may seem that the value of r-squared for both Model 1 and Model 2 are relatively small (0.205 and 0.253 respectively). Nevertheless, it is generally accepted that speech is subject to enormous variability, to the extent that the same word uttered by the same speaker will differ from one occasion to the next. Thus even values of r-squared as low as 0.2 are of empirical and theoretical interest since this indicates that the variability in speech is not entirely random.

The co-efficient estimate for M1 duration indicated an increase in articulatory reduction, for words which were longer in duration, as uttered in the first instance. This was significantly different from zero ($p < 0.01$) and the 95% confidence interval indicates that the effect of M1 duration was positive (lower bound = 0.255; upper bound = 0.345). The co-efficient estimate for task familiarity also differed significantly from zero ($p < 0.01$) and 95% confidence interval indicated that task familiarity had a negative effect on articulatory reduction (lower bound = -31.5; upper bound = -5.7). This indicated that, as expected, performing the task a second time had the effect of attenuating articulatory reduction.

The addition of the variables; *monosyllabicity*, *turn distance*, *speaker role* and *gender conflict* in Model 2 did not significantly improve on Model 1 (F = 0.04, $p > 0.05$). Model 2 indicated that monosyllabicity was also found to be a good predictor of articulatory reduction. In effect, monosyllabic words offered greater scope for articulatory reduction than polysyllabic words. However, since monosyllabicity and M1 duration are not independent of one another, the variable monosyllabilicty was excluded from Model 1.

The effect of increasing the turn distance by one dialogue turn was not significant. The co-efficient estimate did not differ significantly from zero. Nevertheless, before drawing conclusion as to the relevance of turn distance on articulatory reduction, it is worth noting that a problem arose in that the relationship between M1 – M2 duration difference (articulatory reduction) and turn distance was not linear. This constitutes a violation of the assumptions underlying linear regression. Several attempts were made to transform the variable *turn distance* by squaring the number of turns, using the log or using the exponent of the number of turns. However, none of these techniques resulted in a linear relationship. One solution to this problem might be to measure the distance between first and second mentions of word tokens in terms of time, or duration. However, this solution may not be appropriate since the relevant factor, as indicated by Robertson and Kirsner (2000) may be a change in topic. They showed that, while word duration decreased for second mentions of words referring to the same object, when a topic change occurred in a dialogue, the pattern was reversed. This suggests that, at the level of the discourse, articulatory reduction may be influenced by discrete factors, such as a change in conversational topic, rather than by continuous variables such as the time between uttering first and second mentions of word tokens. With respect to the current study, the Map Task did not involve changes in conversational topic.

The co-efficient estimate for speaker role did not differ significantly from zero. This indicated that whether the speaker was in the role of Instruction Giver or Instruction Follower was irrelevant from the point of view of articulatory reduction. With respect to the influence of gender conflict, contrary to expectation, mentions of words taken from dialogues where the participants are of mixed gender had a positive, rather than negative effect on duration difference. Nevertheless, the co-efficient estimate was not significant (p= 0.46). On this basis one might conclude that gender was also an irrelevant variable and reject it from inclusion in the model.

With respect to gender, Tannen (1994) argued that women are more co-operative in their conversational styles than men. Thus, although gender conflict was not a good predictor of articulatory reduction, it is nevertheless possible, that women may be more attentive than men in noting introductory mentions, particularly those by the other speaker. One way to test this in future studies would be to use a dichotomous variable male/female and exclude those cases where the gender of the participants is mixed. A disadvantage of this approach, however, might be that reducing the number of cases may affect comparability between models, if the effect of excluded cases masks the effect of relevant variables. An alternative approach would be to choose three categories for gender male, female, and mixed. In this case, the relationship between three nominal categories would not be linear and therefore violate the assumption of linearity underlying the regression model. To overcome this problem, the category mixed gender could serve as a reference category to which the categories male and female could be compared in turn.

It was noted in section 3.2, that in Bard et al.'s (2000) study the raw duration of word tokens had been normalized to account for any difference in duration due to variation in speaking styles, for example. It could be objected that the raw difference between first mentions and second mentions does not provide a good measure of articulatory reduction, even though Fowler and Housum (1987), employed raw duration measures in their seminal study. In order to go some way to addressing this issue, the multiple regression analyses reported in table 3.4 were also carried out using the proportional difference in duration of second mentions of word tokens relative to first as the dependent variable. However, this did not improve on Model 1 in table 3.4 ($p < 0.05$) when the two models were compared. Furthermore, it was found that the raw duration difference between first and second mentions of word tokens and the proportional duration difference were highly correlated (Pearson $r = 0.93$, $n = 649$, $p < 0.001$). This was true of pairs of word tokens uttered by different speakers (Pearson $r = 0.934$, $p < 0.001$, $n = 285$) as well as of those uttered by the same speaker (Pearson $r = 0.925$, $p < 0.001$, $n = 364$). This observation raises the question of the necessity to normalize word duration measures in this type of study. In fact, by their own admission, the normalization technique used in Bard et al.'s (2000) study was less than ideal.

A final point regarding the multiple regression analysis presented in table 3.4 above relates to the nature of the variables themselves. Several of the variables employed were dummy dichotomous variables rather than variables that can be measured on an interval scale. Although this is not an ideal situation from a statistical point of view, from a linguistic point of view, those factors which were good predictors of articulatory reduction, such as the *turn distance*, are more appropriately described by variables which are discrete in nature, such as a change in conversational topic.

3.4 General Discussion

This study examined the duration of repeated mentions of words in video-mediated dialogues for word pairs uttered by the same speaker and by different speakers. Although the results of this study are not directly comparable with those obtained by Bard et al., (2000), the same pattern of results was nevertheless observed. In common with face-to-face communication, repeated mentions of word tokens were shorter in duration than introductory mentions regardless of which speaker introduced the item. Remote communication via a video-link did not lead speakers to resist the pressure to articulate repeated mentions of words more quickly than introductory mentions.

According to Social Presence Theorists (Short et al., 1976) participants communicating remotely via a video-link should have experienced a sense of social distance from their surroundings and from the other person. If this were the case, speakers might have assigned some kind of special status to their own mentions over those introduced by the listener. However, there was no evidence to suggest that speakers discriminated between their own introductory mentions and those of the other speaker. Although a measure of social distance was not obtained in this study, it is highly likely that speakers would have experienced a sense of social distance in this particular context given the remote nature of the communication. Participants were based in Glasgow and Nottingham and communicated via the Internet and were informed that this was the case in the task instructions. Nor does it seem that the faster articulation of repeated mentions of words referring to the same object, or *articulatory reduction* occurred as a result of speakers adjusting their speech to the listener's comprehension needs in a novel conversational setting. If this had been the case, speakers might have responded to the novelty of communicating via a video-link by mitigating the shortening of repeated mentions of words. Alternatively, they might have assigned some kind of special status to their own mentions over those introduced by the listener, or may have failed to register introductory mentions of words uttered by the other speaker. In which case, mentions of words uttered by the same speaker should have been articulated more quickly on repetition, but this articulatory reduction effect should have been offset for mentions uttered by different speakers. Yet no indication of this was observed in the present study. Articulatory reduction occurred irrespective of any sense of social distance or unfamiliarity with the medium that their listener may have experienced.

The results of this study are consistent with the Dual Process Model (Bard et al., 2000) which holds that the shortening of repeated mentions of words relative to introductory mentions of words referring to the same object is due to priming. Priming processes operate during any attempt to produce spoken utterances and are deemed to be too fast for other factors, such as a response to the conversational setting, to make their effects felt. Furthermore, priming processes are deemed to be exclusively dependent on the experience of the speaker. Thus, provided the speaker registers that an entity has been mentioned, priming would be expected to occur. For this reason, no difference was observed between same-speaker and different-speaker repetitions.

In addition to this, the Dual Process Model holds that priming is triggered by the given status of an entity. Bard et al., (2000) suggest that given status can be conferred on an entity by either speaker-mention or by listener-mention (Bard et al., 2000). Care should be taken however, with drawing this conclusion on the basis of the results obtained here. The design of the current study allows for the possibility that the speaker could have reduced second mentions because they were *given* by virtue of physical presence. That is, a landmark could have been given for the repeater because they could see it on their map (situationally given according to Prince (1981)) and what the listener mentioned may have been irrelevant. Research has shown that the attenuated articulation of repeated mentions can be induced by the physical presence of an object (Bard and Anderson, 1994). Bard et al, (2000) noted this in their study of face-to-face dialogues and consequently examined different-speaker repetitions for the effects of what the repeater could see while uttering the second mention. Introductory and second mentions of words forming the names of the landmarks on a map were compared in cases where the repeater could, or could not see the landmark being referred to. It was found that words referring to objects that the repeater could not see were less intelligible than words which the repeater could see. It is not clear why intelligibility loss should have been greater when the repeater could not see the landmark on the map than when they could. According to Bard et al., (2000), if both listener mention and the visibility of an object are important is assigning given status, then speakers should have reduced items more when they could see them than when they could not - vice versa.

The results of Bard et al.'s (2000) duration analysis were somewhat contradictory. They reported an overall effect of repetition with second mentions of words being shorter in duration than first mentions. But no overall effect of visibility of the object to the speaker. On the basis of these results, Bard et al., (2000) concluded that an introductory mention by the listener sufficed to assign given status: If speakers had assigned given status to what they could see but not to what the listener mentioned, then they should have reduced intelligibility only when repeater could see the landmark.

As far as the present study is concerned, the question of interest is whether or not video-mediated communication appears to function in the same way as face-to-face communication. In contrast, Bard et al., (2000) specifically addressed the question of how given status could be achieved. Thus for the purpose of this particular study, it is irrelevant whether *given* status is achieved because the listener has mentioned an object or whether given status is achieved because the repeater can see the landmark being referred to. Bard et al., (2000) suggest that priming is triggered by *given status* per se and the results obtained here are consistent with that notion. Since given status triggers priming, it is unimportant which speaker introduces an entity into the discourse.

Bard et al.'s (2000) account appears to suggest that, in terms of word articulation, speakers behave in an egocentric fashion. However, Bard et al., (2000) provided evidence which indicated that speakers do adjust their articulation with respect to whom they are addressing. Speakers articulated references to objects which were given from the point of view of the listener, but were articulated as clearly as words referring to New items when the speaker addressed a New listener (through having performed 2 versions of the task with different partners). Furthermore, the finding that articulatory reduction occurs across speakers implies a connection between production and comprehension which is consistent with the Interactive Alignment Model proposed by Pickering and Garrod (in press). However, for the purpose of future work, a better test of the Interactive Alignment Model proposed by Pickering and Garrod (in press) would be provided by an analysis of unshared landmarks uttered by different speakers. This is because the Interactive Alignment Model predicts that priming can occur via comprehension processes. In other words, priming can occur on the basis of what the speaker hears.

Multiple regression analyses were also carried out on the data to explore the influence of word-related and discourse-related factors. Overall, the results of these analyses indicated that the duration of a word (as uttered in the first instance) and familiarity with the task (as indicated by performing the task a second time) were good predictors of articulatory reduction. It is perhaps not surprising that the duration of initial mentions of word tokens should be a good predictor of articulatory reduction. If articulatory reduction occurs primarily by virtue of the *given* status of the entity being referred to (Bard et al., 2000), then it follows that the difference in duration between first and second mentions of words referring to the same object will be proportional to the duration of the initial mention.

Although certain discourse-related factors, such as the role of the speaker, were irrelevant with respect to articulatory reduction, it is interesting that task familiarity should be a good predictor of the degree of articulatory reduction. It is not clear why this should be the case since the processes underlying articulatory reduction are deemed to be automatic in nature. One possibility is that, when performing a problem-solving task which involves navigating one's way around a map, participants make an initial assumption that the maps should be the same. On receiving information to the contrary, it may take time to adjust. Thus when the task is performed for the second time participants might be more aware of the precise nature of that task. Consequently, they may take more care and hence articulatory reduction is offset to some extent.

To sum up, the findings of Study 1 offer some indication that, in terms of word articulation at least, video-mediated communication seems to function in the same way as face-to-face communication. However, a major limitation of this study is that only words in a video-mediated context were examined. Although the results indicated that video-mediated communication does not lead speakers to mitigate articulatory priming, it is not clear whether speakers articulate word tokens more or less quickly overall in a video-mediated conversational setting than in a face-to-face setting. In order to evaluate the role of conversational setting on articulatory reduction, a more direct comparison of face-to-face and video-mediated contexts is required.

In addition to this, the Dual Process Model holds that distinct processes underpin speech production in spoken dialogue. More complex processes, such as those involved in naming an object for instance, are deemed to be more sensitive to the cognitive demands on the speaker's time and attention. This suggests that while priming processes may be unaffected by the setting in which a conversation takes place, other aspects of speech production such as the design of referring expressions are more likely to be influenced by factors which may increase the demands on the speaker's time and attention. Thus, in the following chapter, the role of time pressure and the influence of the conversational setting on word duration and on referential forms will be explored.

Chapter 4

The Influence of Cognitive Load on Articulation and Reference in Video-mediated and Face-to-face Conversational Settings

4.1 Introduction

The main issue to be addressed by this dissertation concerns the way in which speech and language production processes are influenced by the conversational setting in which a dialogue takes place. In particular, I will focus on the way interlocutors refer to objects, or entities, in a discourse. In Chapter 3, the results of an initial study, which examined word duration in video-mediated dialogues, were presented. It was found that repeated mentions of words forming the names of landmarks on a map were articulated more quickly than 1st mentions irrespective of whether both mentions were uttered by the same speaker or by different speakers. The same pattern of results has been observed in face-to-face dialogues (Bard et al., 2000) which suggests that, in terms of word articulation, video-mediated communication appears to function in the same way as face-to-face communication.

The purpose of this Chapter is to explore the influence of cognitive load in different conversational settings. The results of a second study to investigate the impact of cognitive load on articulation and reference in video-mediated and face-to-face dialogues will be presented.

4.1.1 Cognitive Load Theory

I will begin this Chapter with a brief overview of Cognitive Load Theory (Sweller, 1988, Sweller, van Merriënboer and Paas, 1998; Paas and van

Merriënboer, 1994). Cognitive Load Theory (CLT) is primarily concerned with the limitations of working-memory capacity and has largely been applied to learning. It is assumed that a limited working memory is connected to an unlimited long-term memory (Baddeley, 1986). Demands on working memory can affect the ability to construct schema, or mental representations, and this consequently impedes learning.

Paas and van Merriënboer (1994) describe a range of factors which determine the level of cognitive load. These include characteristics of the subject such as cognitive ability, characteristics of the task such as task complexity, and characteristics of the environment such as noise. According to Paas and van Merriënboer (1994) there are three measurable dimensions of cognitive load; *mental load*, *mental effort* and *performance.* Mental load is the portion of cognitive load that is imposed exclusively by the task and environmental demands. However, this cannot be separated from the mental effort required to perform a given task, or the cognitive capacity that must be allocated to the task. These factors contribute to the overall cognitive load which is reflected in a person's performance.

A distinction is made between; *intrinsic* cognitive load, *extraneous* cognitive load and *germane* cognitive load. Intrinsic cognitive load relates to the content of the material to be learned. Extraneous cognitive load is defined in terms of cognitive activity due to the organisation and presentation of the task. Germane cognitive load occurs when free working memory capacity is used for deeper construction and automation of schemata. For the purpose of this chapter, I am primarily concerned with extraneous cognitive load and how that can be manipulated.

Extraneous cognitive load may be increased for a variety of reasons. For instance, it may be increased because the task is more difficult, in which case more cognitive resources may be required to successfully complete the task goal. Alternatively, cognitive load may be increased because a person is doing several things at once, such as performing a task while retaining a set of digits in memory. If several items need to be maintained in working memory, the increased cognitive load may prevent the formation of internal mental representations, or schemata.

4.1.2 Cognitive Load and Articulation

Research demonstrates that cognitive load is reflected in the speech signal itself. Characteristics of the speech signal, such as pauses and false starts, are thought to reflect the underlying processes involved in speech production such as planning, self-monitoring and repair (Lounsbury, 1954; Clark, 1971; Rochester, Thurnston and Rupp, 1977). Interestingly, pauses within a word or phrase have been shown to reflect the complexity of the unit being planned (Butterworth, 1980; Wheeldon and Lahiri, 1997), while disfluencies in speech tend to be more common in more complex constituents (Clark and Wasow, 1998; Oviatt, 1995) and for more complex response choices (Oviatt, 1995). In a similar vein, studies have shown a positive correlation between task difficulty and speech onset latency, or the time taken to begin a task (Sternberg, Monsell, Knoll and Wright, 1980). It is believed that processes underlying speech such as planning, monitoring and repair, place cognitive demands on the speech production system's resources and contribute, therefore to cognitive load. In line with this notion, Roβnagel (2000), observed shorter speech onset latencies together with a higher proportion of intraphrasal pauses and false starts under conditions of low cognitive load, as implemented in an easy version of a model construction task. Yet under conditions of increased cognitive load, speech onset latencies were significantly longer and the proportion of intraphrasal pauses and false starts was significantly reduced. These observations indicate that increases in cognitive load are reflected in the speech signal itself.

There is further evidence to suggest that cognitive load may directly influence the acoustic parameters of speech. For instance, in a study of male speakers, Brenner, Doherty and Shipp (1994) found significant increases in fundamental frequency[1], vocal intensity, and a marginal increase in speaking rate with increased demands in cognitive workload, or task difficulty. Similarly, in a study of Female, Finnish schoolteachers, Ratala, Vilkman, Bloigu (2002) showed increases in fundamental frequency at the end of the working day. This effect was attributed to increased "loading" rather than vocal strain since the increase was greater for those teachers with "few voice complaints" than those with "many voice complaints." Increases in fundamental frequency have also been attributed to factors such as mental fatigue (Whitmore and Fisher, 1996) and time pressure (Mendoza and Carballo, 1998). Whitmore and Fisher (1996) also found that the influence of mental fatigue extended to other parameters of speech such as word duration.

Nevertheless, not all studies have shown that cognitive load influences the acoustic characteristics of speech. For example, Kopardekar and Mital (1997)

[1] Fundamental frequency relates vocal fold vibration and is associated with voice pitch.

observed no overall articulatory differences with increased cognitive load, although they did find evidence to suggest that articulatory effects of cognitive load may be subject to individual differences. Furthermore, a criticism of studies which investigate the influence of cognitive load is that they have confounded cognitive load with psychological stress (Lively, Pisoni, Vansummers and Bernacki, 1993). Whitmore & Fisher (1996), for example, based their study on pilots involved in stressful situations such as bombing missions. Thus, it is not clear whether their findings were due to increases in cognitive load, or to the high levels of stress pilots would be likely to experience in this type of situation.

In an effort to control for this, Lively et al., (1993) examined the effect of cognitive workload on speech output. Cognitive workload was defined as the "information processing load placed on the speaker while performing the task" and was manipulated by having speakers perform a compensatory visual tracking task while uttering test sentences of the form "Say hVd again". Workload sentences were compared with a corresponding control condition. It was found that, in the workload condition, some speakers produced utterances with increased speaking rate and increased amplitude compared with the control utterances. Furthermore, when utterances from those speakers were presented to listeners for identification, sentences produced under the workload condition were less intelligible than sentences produced under the control condition. It should be noted, however, that the number of speakers involved in this study was small, only 5 speakers. Consequently, the findings are limited in their ability to generalise to larger populations.

On balance, findings such as these provide evidence that the acoustic parameters of speech are influenced by factors such as cognitive load. However, many of the studies presented above have focussed on very narrow acoustic characteristics of speech, such as fundamental frequency (e.g. Mendoza and Carballo, 1998) or on the reading of simple rather meaningless utterances (Lively et al., 1993). Although there is some suggestion in the literature that cognitive load influences the broader aspects of articulation such as rate of speech, it is not clear whether this observation will generalise to a more naturalistic setting such as spoken dialogue.

4.1.3 Cognitive Load and Reference

The influence of cognitive load is not limited to articulation. Cognitive load has also been shown to influence a speaker's choice of referring expression. Horton and Keysar (1996), for example, found that time pressure influenced the way speakers described simple objects moving across a computer screen. When there was no pressure of time, speakers tended to describe the objects in terms of information that was shared between the speaker and the listener. However, when speakers were instructed to initiate utterances quickly, the tendency to rely on shared information was offset. Similarly, Roβnagel (2000) showed that taking the listener's perspective could be influenced by cognitive load. Since the influence of cognitive load on perspective taking is the theme of the following chapter, the results of Roβnagel's study will be described in more detail in Chapter 5. For the present, of particular interest, is the type of referring expressions used under differing conditions of cognitive load. In Roβnagel's (2000) study, 45 native speakers of German instructed a confederate, either a boy or a student, to assemble a machine model. The parts of the model were labelled with pseudo-technical terms, such as *transmission unit, pump holder* etc., and participants performed the task under 3 conditions of cognitive load. In a low load condition, participants had a 2nd assembled model before them while giving instructions. In a high load condition, participants had to recall the assembly of the model entirely from memory (the model was not at hand). In a third "dual task" condition, participants had the assembled model at hand but carried a memory load of seven random digits while giving their instructions.

References to component parts were coded as a *technical term alone*, a *technical term plus a description*, or a *description only.* The use of referring expressions by participants instructing the adult confederate to construct the model is illustrated in table 4.1 below.

Table 4.1 Percentages of types of specification and designation for the student data in Experiment 1 (table reproduced from Roβnagel, 2000)

		Cog.load	
Designation	Low	Dual Task	High
Technical term	56.0	89.3	88.0
Term + Description	38.7	6.7	8.7
Description	44.0	10.7	12.0

It is interesting to note that the level of detailing was reduced when the cognitive load was high. For example, participants used fewer *terms plus description* under the high cognitive load condition (8.7%) and the dual task condition (6.7%) compared with the low load condition (38.7%). Correspondingly, participants tended to use more short referring expressions, such as technical terms alone, when the cognitive load was high (88%) or when a dual task was performed (89.3%) compared with when the cognitive load was low (56%). Although these differences were not tested statistically, descriptive statistics seem to suggest some form of economy. Generalising somewhat, participants used shorter referring expressions under conditions of increased cognitive load.

4.1.4 Cognitive Load in Video-mediated Communication

There are several reasons to suspect that video-mediated communication may function differently to face-to-face communication. These were discussed in Section 2.2 of Chapter 2 and are summarised briefly, for convenience. Generally speaking, video-mediated dialogues have been reported to be longer (in terms of the number of words), less interactive, and contain fewer conversational turns than face-to-face dialogues (Anderson et al., 1997;Doherty-Sneddon et al., 1997; O'Conaill, Whittaker and Wilbur, 1993; O'Malley et al., 1996; Sellen, 1995). Attempts to explain why video-mediated communication should function differently to face-to-face communication include; the limited availability of visual signals, the remote nature of the communication, and the novelty of the technology. For instance, it has been suggested that communication in certain conversational settings, such as video-mediated communication, may be limited by the availability of gestures, eye gaze and facial expressions (Clark, 1996). Furthermore, in a video-mediated conversational setting, interlocutors communicate with each other from remote locations and this may invoke a sense of social distance between the interlocutors and the communicative situation (Short, Williams and Christie, 1976; Williams, 1977). It has also been suggested that differences observed between face-to-face and video-mediated communication may be due to the novelty of the conversational setting (e.g. Blokland and Anderson, 1998). Some evidence in support of this notion is provided by a study to investigate participants' ability to adapt to the new mode of communication. Newlands, Anderson and Mullin (2003) showed that while task performance was initially poor in a computer-based setting, participants' performance improved to match that of face-to-face communication with repeated exposure to the new mode of communication. Evidence in support of the notion that users of computer-based technology become accustomed to and overcome the limitations of that medium has also been reported by McGrath, Arrow, Gruenfeld, Hollingshead and O'Connor (1993) as well as Hollinsghead, McGrath and O'Connor (1993). Similarly, Walther (1995) found that, for participants communicating in small groups of 3, computer-mediated communication was judged to be no less intimate than face-to-face communication. Furthermore, although computer-mediated communication was initially rated as more aggressive than face-to-face communication, with repeated use of the technology this difference dissipated.

It is possible that video-mediation itself may serve to increase cognitive load compared with a face-to-face conversational setting. If this is the case, task performance should be worse in a video-mediated setting compared with a face-to-face setting. Recall, that Cognitive Load Theory (e.g. Sweller, 1988) holds that any increase in cognitive load will be reflected in a person's performance. Comparisons of face-to-face and video-mediated communication have indicated mixed findings in terms of the influence of conversational setting on task performance. On the one hand, several studies report no advantage in terms of task performance of either video-mediated or face-to-face communication over audio only communication (e.g. Anderson et al., 1996; Anderson et al., 1997; Doherty-Sneddon et al., 1997). On the other hand, significant differences in communicative efficiency and task performance have been reported for high quality video-mediated communication compared with face-to-face communication (Kraut, Fussell and Siegel, 2003).

One reason why task performance should be significantly worse in a video-mediated conversational setting is that there are aspects of the technology itself which might serve to increase cognitive load. For example, if information, such as text and graphics, is not adequately integrated, this may cause attention to be split between the two information formats. It has been argued that the integration of different types of representation has associated cognitive costs (Chandler and Sweller, 1991; Chandler and Sweller, 1992; Sweller, Chandler, Tierney and Cooper, 1990). Furthermore, there is some evidence to suggest that cognitive load may be considerable where the presentation of computer-based information leads participants in a conversation to split their attention between different information formats. In a study to assess team-working skills in a computer-based, collaborative knowledge mapping task, Chung, O'Neill and Herl (1999) found little evidence to indicate a correlation between team processes and team outcomes. This was surprising since collaborative working was expected to have a positive effect on team outcomes. In small groups, participants should benefit from active participation in a discussion (Fay, 2000). In the study conducted by Chung, O'Neill and Herl (1999), the groups were small, with only 3-people in each group. While it was found that, overall, the groups were successful in constructing knowledge maps (sometimes commonly referred to as "Mind Maps"), the attentional demands of the task were reported to be heavy. For instance, participants' attention was often drawn away from the cognitive map in order to read messages, presented in a text box, from other group members. Furthermore, low performing groups (i.e. those groups with the most basic maps) generally spent more time reading information presented in the text box than constructing their knowledge map. In addition to this, it was observed that participants in low performing groups seemed unable to benefit from the group discussion. On the basis of these observations, Chung, O'Neill and Herl (1999) concluded that the demands of constructing the knowledge map could have induced too heavy a cognitive load on participants since the large amount of messages they were sending was interfering with their ability to follow the discussion. A similar line of reasoning has been used to account for the observation that, learning from print can be better than learning from web based designs (Eveland and Dunwoody, 2001). In this case, it was suggested that an overuse of hypermedia in the web-based designs could have disoriented learners and thereby increased the cognitive load associated with the learning task (Eveland and Dunwoody, 2001). Although these studies did not involve video-mediated communication, they suggest that configurations on a computer monitor, which involve the presentation of information in several windows for example, may increase the cognitive load associated with the task.

4.1.5 Cognitive Limitations on Speech Production

In section 2.3 of Chapter 2, several models of interpersonal communication, such as the collaborative model (e.g. Clark and Wilkes-Gibbs, 1986) and the perspective-taking model (e.g. Schober, 1993), were presented. A review of the relevant literature revealed a strong emphasis on the investigation of the role of the listener in interpersonal communication. More recently, however, researchers have suggested that other factors, such as cognitive limitations, may play a more crucial role in spoken discourse. Several suggestions have been put forward which center around the notion that the processes underlying speech and language production may be limited by the cognitive demands on the speakers. These are described below.

Horton and Keysar (1996) showed that, under time pressure, speakers tended to refer to objects moving across a computer screen in terms of their own information rather than information they shared with their listener. To account for this observation, they proposed that the initial planning stages of speech production may be dependent on the speakers own knowledge, with models of listener-knowledge implicated in the later stages as part of a correction mechanism. Roßnagel (2000) further suggested that the results of Horton and Keysar's (1996) experiment implied that, for the initial planning stages at least, cognitive availability rather than the listener's information needs determined which information should be incorporated into the speech plan. Under conditions of increased cognitive load, such as time pressure, controlled processes (including modelling what the listener knew, and monitoring and repair) may be disrupted or sacrificed.

Reasoning along similar lines, Bard et al., (2000) proposed a Dual Process Model of dialogue processing, which was outlined in section 2.5 of Chapter 2. The Dual Process Model holds that cognitive availability is more likely to influence those processes which are consciously controlled than those which are automatic. Thus, a basic distinction is drawn between the processes that underpin speech production in spoken dialogue. On the one hand, automatic processes, such as priming, are deemed to depend on the sole experience of the speaker and occur within a very fast planning cycle. On the other hand, consciously controlled processes, such as task planning, evaluating what the listener knows, and making decisions, are deemed to be more complex in nature and operate over a longer planning cycle. Thus, controlled processes are more likely to be influenced by the cognitive demands on the speaker's time and attention than automatic processes. It is this point which is of particular relevance to this Chapter.

4.1.6 Research Questions and Hypotheses

The questions to be addressed in this Chapter are: Will articulation and reference be influenced by cognitive load and by the setting in which a conversation takes place? Will cognitive load have the same impact in video-mediated and face-to-face conversational settings?

With respect to articulation, Lively et al's (1993) findings suggest that increased cognitive load may be reflected by an increase in speaking rate. On the basis of their results, the rate of speech should be faster under conditions of increased cognitive load. On the other hand, aspects of articulation, such as articulatory reduction (i.e. the shortening of repeated mentions of words relative to introductory mentions) is unlikely to be influenced by cognitive load since priming is deemed to be an automatic process which occurs within a very fast planning cycle (Bard et al., 2000). Automatic processes are not subject, therefore, to the demands on the speaker's time and attention, and are not expected to place demands on working memory. Similarly, the Dual Process Model (Bard et al., 2000) predicts that articulatory priming should occur too quickly for facets of video-mediated communication, such as a sense of social distance or the novelty of the medium, to make their impact felt on a word-by-word basis. Although the

Dual Process Model has not been explicitly applied to video-mediated communication, articulatory reduction has been shown for face-to-face dialogues (Bard et al., 2000). Furthermore, the findings of Study 1, reported in Chapter 3, indicated that articulatory reduction also occurs in a video-mediated conversational setting.

In short, in terms of articulation, it is expected that articulatory reduction will occur irrespective of cognitive load or conversational setting, but that the rate of speech will be faster under conditions of increased cognitive load.

With respect to reference, the Dual Process Model (Bard et al., 2000) holds that cognitive load is more likely to influence the way speakers name an object, than the way speakers articulate words forming those names. It is thought that the processes involved in naming an object involve an evaluation of how readily names will be interpreted (Roßnagel, 2000) which suggests that the processes involved in naming an object, may be consciously controlled. Furthermore, according to the Dual Process Model (Bard et al., 2000), complex processes occur over a long planning cycle and are subject to the demands on the speaker's time and attention. This suggests that if time is short and/or the demands on the speaker's cognitive resources are increased, the processes involved in naming an object may be disrupted. The study conducted by Roßnagel (2000) suggests that under conditions of high cognitive load participants will tend to use relatively short referring expressions. Following a similar line of reasoning, the processes involved in naming an object are more likely to be influenced by facets of video-mediated communication, such as a sense of social distance or the novelty of the medium, since there is more planning time for such factors to make their impact felt.

With respect to whether cognitive load will impact on articulation and reference in the same way in a video-mediated conversational setting compared with a face-to-face setting, there is some evidence to suggest that the configuration of information in computer-based technology may contribute to cognitive load (Chung, O'Neill and Herl, 1999; Eveland and Dunwoody, 2001). If this is the case, then any impact of cognitive load should be more pronounced in a video-mediated setting than in a face-to-face setting. Once again, the Dual Process Model (Bard et al., 2000) suggests that consciously controlled processes will be subject to cognitive load effects while automatic processes should occur irrespective of any variation in cognitive load.

In order to test these hypotheses pairs of participants performed a timed and untimed version of the Map Task in either a face-to-face or a video-mediated conversational setting. The procedure used for the collection of data is described in section 4.2 below.

4.2 Data Collection

4.2.1 Design

Previous studies have manipulated cognitive load by increasing task difficulty or having participants perform a secondary memory task. However, in spoken dialogue, one of the most critical features relates to time. A speaker must produce utterances in real time, as well as interact with another person and cope with the demands of a given task. It seems reasonable to suppose that if less time is available to the speaker to do the same work, then cognitive load will be increased. Furthermore, the Dual Process Model predicts that complex processes, as opposed to automatic processes, will be subject to the demands on the speaker's time and attention. Thus an appropriate way of manipulating cognitive load for the purpose of the present study was to impose a time pressure on the task. Pilot work suggested that a 3-minute time limit would be sufficient to put participants under pressure yet be able to complete or nearly complete the task. With a time limit of only 2 minutes, it was noted that participants were unable to complete the task and consequently made no references to a substantial number of landmarks on the map. With a time limit of 4 minutes participants were able to complete the task, but the routes produced by the Instruction Followers in the video-mediated setting did not appear to differ substantially from those produced in the face-to-face setting. Thus cognitive load was manipulated by having pairs of participants perform 2 versions of a problem-solving task, The Map Task. The task was performed once within a 3-minute time limit and once without the pressure of time.

The task was compared in 2 conversational settings; face-to-face and video-mediated. Conversational setting was treated as a between-groups variable since it was likely that the effect of performing the task in a video-mediated setting followed by a face-to-face setting would not be symmetrical with the effect of performing the conditions in the opposite order. For example, any sense of social presence established in a face-to-face setting may be carried through or transferred to the video-mediated setting. Whereas, participants performing the task in a video-mediated setting first may feel socially distance from their interlocutor but then establish a greater degree of social presence in the face-to-face setting. Since asymetrical order effects may cause serious problems in the interpretation of the results of repeated measures design (Roberts and Russo, 1999), conversational setting was run as a between subjects variable. In addition to this, a practical problem arose in that, when setting up the technical equipment, it became apparent that there were insufficient channels to obtain speaker per channel recordings for both the face-to-face and video-mediated dialogues in one experimental session.

A latin-square design was used for this study. The design was counterbalanced for task order and map. Thus, participants taking part in the study performed the tasks in one of the four sequences listed below;

(a) Untimed (Map A) → Timed (Map B)

(b) Timed (Map A) → Untimed (Map B)

(c) Untimed (Map B) → Timed (Map A)

(d) Timed (Map B) → Untimed (Map A).

Participants

80 participants, recruited at the University of Glasgow, took part in the study. All were native speakers of English and were aged between 17 and 41. Participants were recruited in pairs and were assigned to either a video-mediated or a face-to-face conversational setting. 21 pairs of participants performed the task in a face-to-face setting and 19 pairs of participants performed the task in a video-mediated setting. In the face-to-face group, 9 pairs of participants were friends, or had met before and 12 pairs of participants were unknown to each other. 8 participants had done the Map Task before as part of a video-mediated experiment and 34 participants were naive to the task. In the video-mediated group, 4 pairs of participants were friends or knew each other and 15 pairs had never met before. 14 participants had done the Map Task before. Of those 14, all

except 2 had done the Map Task as part of a video-mediated experiment. 24 participants were naïve to the task.

Task

The task used for this study was the Map Task (Brown, Anderson, Yule and Shillcock, 1984). Electronic versions of the Maps were created using Adobe Photoshop©. The maps used in the face-to-face condition were printed versions of the maps used in the video-mediated conditions. Once created, the maps were converted to bit-map files and displayed on a computer monitor using Microsoft Paint. The face-to-face versions of the maps were printed onto A4 paper and the video-mediated versions of the maps were displayed on a computer monitor in a 16cm x 20cm window. Copies of the Maps used in Study 2 are included in appendix 1.

Technical set-up

In the video-mediated condition, participants were located in different rooms and communicated with one another using a high quality desktop videoconferencing system. The computer-based version of the Map Task (6.3" x 7.85") and a video window (3.5" x 4.5") of the other participant were displayed on the computer-monitor. To avoid the possibility of effects arising from audiovisual asynchronisation, the audio and visual signals were transmitted directly via cables of similar type and length to ensure that there was no perceptible audio or visual delay. The video image of the other participant was refreshed at a rate of 25 frames per second and contained a head and shoulders image of the other participant. The audio was full duplex.

4.2.2 Procedure

The face-to-face and video-mediated conditions were run in blocks. Participants in the face-to-face group performed a pen and paper version of the Map Task. They sat across a table with a low level barrier to prevent seeing each other's maps. In the video-mediated condition participants were located in different rooms and performed the computer-based version of the task. Task role (Instruction Giver or Instruction Follower) was randomly assigned and each pair of participants performed a timed and untimed version of the task in one experimental session. In the un-timed condition participants completed the task in their own time. In the timed condition, a 3-minute time limit was imposed. In order to maintain a sense of pressure throughout the task, subjects were interrupted at minute intervals to be told (by the experimenter) that there were 2 minutes to go, 1 minute to go and that the time was up. The dialogues were recorded onto cassette tape during the experimental session. In the face-to-face group, loss of audio signal occurred during 4 experimental sessions. In the video-mediated group, 1 Instruction Follower failed to draw the route on the map. Consequently, the data from these sessions were excluded from subsequent analyses. After the experimental session digital recordings of the dialogues were made on a Viglen PC at a sampling rate of 16 kHz. Transcriptions were made of the remaining 70 dialogues.

4.2.3. Transcription and Coding

Transcriptions of the dialogues were made from audiotape. An example transcript is presented in Appendix 3. Each dialogue was transcribed into 2 Microsoft Word tables; a "summary" table and a "transcript" table (see example transcript in Appendix 3). The summary table contains 11 columns and each reference to a landmark is listed on a new row. Column 1("No") contains reference numbers which were used to locate the references to landmarks in the full transcript. Pronouns were referenced using numbers and letters. For example, an initial reference to the *Chemical Weapons Plant* was number "2" and a subsequent reference to that landmark using the pronoun *it* was numbered 2b. Column 2 ("Time") denotes the time, in minutes and seconds, a landmark occurred in the dialogue. Only times for the names of landmarks included in the duration analysis are given. Column 3 ["Mention (Ref Exp)"] denotes 1st and 2nd references to landmarks, while column 4 ["Mention (Dur)"] denotes 1st and 2nd mentions of words forming the same names of landmarks on the map. 1st and 2nd mentions are cross referenced in column 5 ("M1/M2"). Column 6 ("Sp") indicates whether the 1st and 2nd mentions in column 4 were uttered by the same speaker or different speakers. Column 7 ("Landmark") gives the reference to the landmark as uttered by the speaker. Column 8 ("Role") indicates whether the landmark was uttered by the Instruction Giver (IG) or the Instruction Follower (IF). Referential coding scores are given in column 9 ("Code Short RE"). This coding scheme is described below and presented in table 4.2. Column 10 ("Q-form") indicates whether a landmark was introduced into the dialogue using a question form (Q) or a non-question form (Non-Q) while column 11 ("Response") denotes whether the response to an introduction was informative (IR) or not (Non-IR). The criteria used to categorise introductions to landmarks will be described in section 5.2 of Chapter 5.

In the second table, a full transcription of the dialogues is provided. Each speaking turn is entered on a new row. The role of the speaker, namely whether the Instruction Giver (IG), the Instruction Follower (IF) or the Experimenter (Exp) is given in Column 1 ("Role"). Column 2 ("Text") contains the transcript of the speaker's utterances. Where punctuation, such as a question mark, occurs this may relate either to the syntactic form used, or to a subjective judgement by the transcriber regarding the intonation pattern of that utterance. Overlapping speech is indicated by a forward slash and a hyphen was used to indicate interrupted or disfluent speech. Column 3 ("Number") contains the cross-reference numbers used in the summary table and Column 4 ("Introduction) repeats the introductory forms denoted in Columns 10 and 11 of the summary table.

All 1st and 2nd references to landmarks were coded according to the scheme in table 4.2 below.

Table 4.2. Coding scheme for shortening of referring expressions

Code	Description	Example
0	full landmark name as prescribed on map	Do you have a popular tourist spot?
1	truncated landmark name	I have a tourist spot yes
2	pronoun	Ok go right the way round it

The coding scheme illustrated in table 4.2 takes advantage of the fact that each landmark appeared with a name, usually a complex noun phrase, such as *popular tourist spot,* which thus offered scope for shortening. If the name that was actually written on the map was used, the referring expression was coded as 0. If the name was truncated, this was coded as 1. Occasionally, participants used an alternative name to that prescribed on the map (e.g. *overnight stay place* instead of *overnight accommodation*). In such cases the names were categorised by the length, in terms of the number of syllables, irrespective of the fact that an alternative name was used. If a pronoun was used, this coded as 2. Pronouns included *that, it, mine, those, yours* but excluded relative pronouns.

This scheme is similar to one used by Bard and Aylett, (2001) which is reproduced in table 4.3 for convenience.

Table 4.3. Accessibility scale employed by Bard and Aylett (2001)

Code	Description	Examples
0	(numeral or indefinite article) + noun sequence	one mountain a mountain
1	(definite article or possessive) + nominal	the mountain my one
2	possessive pronoun, deictic pronoun, or deictic adjective + nominal	mine that this mountain
3	other pronoun	it

The coding scheme employed by Bard and Aylett (2001) is based on the relationship between a referring expression and its antecedent. Referring expressions involving indefinite noun phrases are thought to refer to less readily interpretable, or *accessible*, antecedents whereas reduced referring expressions, such as pronouns, are thought to refer to the most accessible referents (Ariel, 1990, Gundel, Hedberg and Zacharski, 1993). While the referential coding scheme employed in the present study appears, at first sight, to be similar to that employed by Bard and Aylett (2001), the scheme was adapted for two main reasons. Namely, the accessibility of a referent is not of prime concern in the present study and second, it is questionable whether pronouns can be mapped onto a scale of accessibility. To deal with the first point, the purpose of Bard and Aylett's (2001) study was to examine the effect of repeated mention on referent accessibility, whereas the purpose of the present study is to examine the influence of cognitive load on the use of referring expressions. Given that cognitive load was manipulated by imposing a time pressure, it seems reasonable to suppose that speakers will shorten the names of landmarks in order to save time. A decision to shorten a name may not necessarily involve how readily an expression will be interpreted in the context of the dialogue. With regard to the second point there is some evidence, based on data in Finnish, to suggest that the determinants of a referent's accessibility are complex in nature and that the properties of different referential forms, such as pronouns, cannot always be mapped directly onto an accessibility scale (Kaiser and Trueswell, 2003). For these reasons a coding scheme which captured the shortening of the names of objects and did not differentiate between pronouns was felt to be more appropriate for the purposes of the present study.

The reliability of the referential coding scheme (illustrated in table 4.2) was determined by having 16 randomly selected dialogues (8 face-to-face and 8 video-mediated) independently coded. Inter-judge agreement was assessed using the Spearman rank correlation coefficient. Since the coding scheme was mapped onto an ordinal scale, a non-parametric correlation coefficient was chosen. Intercoder reliability was highly correlated ($r_s = 0.98$; $p < 0.001$, $N = 158$).

4.3 Data Analysis

In this section the procedure used to organise the data and test each of the hypotheses is outlined. In section 4.3.1, the assumption that time pressure will increase cognitive load is tested. In section 4.3.2, the hypotheses that rate of speech will increase under high cognitive load and that articulatory reduction will occur irrespective of cognitive load and conversational setting are tested. In section 4.3.3, the hypothesis that speakers will shorten referring expressions under conditions of increased cognitive load will be tested.

4.3.1 Cognitive Load Response

In order to check that participants had responded to time pressure, the influence of cognitive load on task performance and dialogue length (in terms of duration and number of words) was examined. If time pressure served to increase cognitive load then, according to Cognitive Load Theory (e.g. Sweller, 1988), this would be reflected in task performance. Thus task performance should be significantly worse under time pressure than without the pressure of time.

It has been shown that characteristics of the speech signal itself, such as pauses and disfluencies, can be useful indicators of cognitive load (e.g. Roßnagel, 2000). However, such measures require detailed acoustic analysis and are beyond the scope of this current work to employ. Nevertheless, given that cognitive load is manipulated by imposing a time limit, it seemed reasonable to suppose that if a 3-minute time limit was sufficient to impose a considerable cognitive burden on participants, then the time allocated should be significantly less than the time it would take for participants to complete the task without the pressure of time. According to this line of reasoning, if participants responded to time pressure, then the resulting dialogues should be shorter in length (in terms of duration and the number of words) compared with dialogues performed without the pressure of time.

In short, if the manipulation of cognitive load was effective, then, under time pressure, task performance would be worse and dialogues would be significantly shorter in duration and contain fewer words compared with no pressure of time.

Dependent Variables

Three dependent variables were used to assess whether or not participants had responded to time pressure; *task performance*, *dialogue duration* and the *number of words per dialogue.*

A measure of task performance was obtained by following a procedure devised by Anderson, Bader, Bard, Boyle, Doherty-Sneddon, Garrod, Isard, Kowtko, McAllister, Sotillo and Thompson (1991). The original route on the Instruction Giver's Map was copied onto acetate, then superimposed on the maps produced by the Instruction Followers. The degree to which the Instruction Follower's route deviated from the route on the Instruction Giver's map is indicated by the difference in area (in squared centimetres) between the two routes. Low route deviation scores indicate better task performance through more accurate reproduction of the map. High route deviation scores indicate poor task performance. Inter-judge reliability of the route deviation scores was determined by having 16 randomly selected routes (8 face-to-face and 8 video-mediated) independently measured. Inter-judge agreement was assessed using the Pearson's r statistic. Inter-judge reliability was found to be highly correlated ($r = 0.951$, $p < 0.001$, $N = 16$).

The duration of the digitised dialogues was measured in seconds. Measurements were taken from the time the participants began to speak following the signal, “You may begin” until the time was up (timed condition) or until the task was completed (untimed condition). Participants had been instructed to indicate they had finished the task by using a phrase such as “that’s it”.

The number of words per dialogue was counted using the *tool command* on Microsoft Word. Since all utterances take time, interjections such as *erm, uh huh* etc were included in the word count.

The data for analysis was drawn from 64 dialogues, with 16 dialogues in each cell of the experimental design. According to the collaborative model of communication (Clark and Wilkes-Gibbs, 1986; Clark, 1992) a dialogue represents the joint product of each pair of participants in terms of linguistic output. For this reason, the analyses of dialogue duration and number of words per dialogue, employ the dialogue itself as the unit of analysis. From a theoretical point of view, task performance can also be viewed as the result of a joint activity. From statistical point of view, however, the route itself was produced by the Instruction Follower. Hence the analysis of task performance is based on a conventional analysis with data from a single participant constituting the unit of analysis.

Given the manipulation of time pressure, it was possible that the data would not meet with the Homogeneity of Variance and Normality of Distribution assumptions underlying analysis of variance. Consequently, before proceeding with the analysis it was necessary to check whether the distributions associated with the dependent variables were normally distributed and whether the variance associated with the test samples scores was homogenous. The normality of distribution scores was determined across the untimed and timed conditions for both the video-mediated and the face-to-face conversational settings. Kolmogorov-Smirnov tests were carried out on the data and the results of these tests are presented in table 4.4 below. The Kolmogorov-Smirnov Z statistic is provided along with the significance of the p-value.

Table 4.4 Kolmogorov-Smirov tests for task performance, dialogue duration and number of words per dialogue

		Cognitive Load			
		Un-timed		*Timed*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
		Task performance			
Face-to-face	32	0.57	ns.	0.46	ns.
Video-mediated	32	0.76	ns.	0.48	ns.
		Dialogue duration			
Face-to-face	32	1.22	ns.	0.59	ns.
Video-mediated	32	1.20	ns.	0.74	ns.
		Number of words per dialogue			
Face-to-face	32	0.79	ns.	0.72	ns.
Video-mediated	32	0.69	ns.	0.84	ns.

As can be seen from table 4.4, the results of the Kolmogorov-Smirnov test indicate that the distributions of all test samples were normal. The samples were also tested for equality of variance. Levene's test of equality of variance revealed that the variance of the dependent variable, task performance was equal across groups [$F(3,60) = 1.6$, $p = 0.20$]. However, the variance of the dependent variable, *duration*, was not equal across groups [$F(3,60) = 8.84$, $p < 0.01$] nor was the variance associated with the *number of words* [$F(3,60) = 8.27$, $p < 0.01$]. In both cases, the difference in variance was large. For example, in the case of *duration*, the largest variance was 24336 (for the video-mediated untimed condition) and the smallest variance was 676 (for the face-to-face timed condition). This raises the question of whether the data should be subjected to analyses of variance. According to Roberts and Russo (1999) ANOVA is a robust test and, provided the cell sizes in the design are equal, even major deviations from the assumptions underlying analysis of variance are unlikely to result in Type I or Type II errors. Thus, analyses of variance were carried out on the data.

Task performance

The mean values for task performance (as measured by route deviation) with standard deviations are shown in Table 4.5 below.

Table 4.5. Mean task performance scores (route deviation in cm²) for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	32	33.0	16.0	52.9	20.3
Video-mediated	32	43.9	18.9	54.2	30.2

As can be seen from table 4.5, the mean task performance scores suggest that Instruction Follower's performance was worse under increased cognitive load. In the face-to-face group, mean route deviation was 52.9cm² in the timed condition compared with 33cm² in the untimed condition. In the video-mediated group, the difference was less pronounced. The mean route deviation was 54.2cm² in the timed condition and 43.9cm² in the untimed condition. The results of the analysis of variance indicated that, overall, the difference in route deviation scores between timed and untimed conditions was reliable.

A 2 x (2) analysis of variance was carried out on the data with conversational setting (video-mediated versus face-to-face) as a between-subjects factor and cognitive load (untimed versus timed) as a within-subjects factor. The summary table is presented in table 4.6 below.

Table 4.6. Analysis of variance for task performance

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	0.95	0.34
S within-group error	30	(620.10)	
Within subjects			
Cognitive Load (L)	1	10.53	*<0.01
S x L	1	1.04	0.31
S x S within-group error	30	(348.23)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of cognitive load [F(1,30) = 10.53, p < 0.01]. Overall, Instruction Followers deviated more from the original route in the timed condition (mean route deviation = 53.59cm^2) than in the un-timed condition (mean route deviation = 38.45 cm^2). There was no effect of conversational setting (F<1) and no significant interaction.

Post hoc analyses were also carried out on the data. The results of Newman Keuls pair-wise comparisons indicated that while the effect of cognitive load was statistically significant for the face-to-face conversational setting (p < 0.01), the difference in task performance did not reach significance in the video-mediated setting. With respect to conversational setting, although there was a numerical difference in task performance between the face-to-face and video-mediated settings in the untimed condition, the results of the Newman-Keuls pair-wise comparison was not statistically significant (at the 0.05 level of significance).

Overall, the results of these analyses indicated that time pressure served to increase cognitive load. There was no indication that video-mediation made a significant contribution to cognitive load.

Two versions of the Map Task were used in Study 2. In order to ensure that the maps were functionally equivalent a 2 x 2 independent measures of analysis of variance was carried out on the data to examine the effect of map version (Map A versus Map B) on route deviation in the untimed and timed conditions. If Map A and Map B were functionally different then route deviations would be significantly different for Map A compared with Map B. The results of the analysis of variance showed that, not surprisingly, the effect of cognitive load was significant [$F(1,60) = 7.86$, $p < 0.01$]. There was no effect of Map Version ($F < 1$) and no interaction between map version and cognitive load ($F < 1$). The results of Newman-Keuls pair-wise comparisons revealed no significant difference (at the 0.05 level of significance) between Map A and Map B in either the untimed or timed conditions.

Dialogue duration

The mean values for dialogue duration with standard deviations are shown in Table 4.7 below.

Table 4.7. Mean dialogue duration (in seconds) for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	32	307	147	167	26
Video-mediated	32	292	156	154	30

As can be seen from table 4.7, the untimed dialogues were longer in duration than the timed dialogues for both the face-to-face setting (5mins 7 seconds cf. 2mins 47 seconds) and the video-mediated setting (4 mins 52 secs cf. 2 mins 34 secs). The results of the analysis of variance indicate that this difference in duration was significant.

A 2 x (2) analysis of variance by dialogue was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The summary table is shown in table 4.8 below.

Table 4.8. Analysis of variance for dialogue duration

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	0.22	0.64
S within-group error	30	(13191.53)	
Within subjects			
Cognitive Load (L)	1	29.25	* < 0.01
S x L	1	0.002	0.97
S x S within-group error	30	(10624.46)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of cognitive load [F (1,30) = 29.25, p < 0.01]. Overall, timed dialogues were shorter in duration (mean duration = 2mins 40 secs) than untimed dialogues (mean duration = 5mins). There was no interaction (F < 1) and no effect of conversational setting (F < 1). Post Hoc analyses were also carried out on the data. The results of Newman Keuls pair-wise comparisons indicated that the effect of cognitive load was significant in both the face-to-face conversational setting ($p < 0.01$) and the video-mediated setting ($p < 0.01$). In both settings, dialogues were significantly shorter in duration in the timed condition than in the untimed condition. There was no significant effect of conversational setting, however, for either the untimed or timed conditions.

The results of these analyses indicated that the manipulation of time pressure served to increase cognitive load.

Number of words per dialogue

The mean number of words per dialogue with standard deviations for the untimed and timed face-to-face and video-mediated dialogues is shown in table 4.9 below.

Table 4.9. Mean number of words per dialogue for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	32	962	467	509	93
Video-mediated	32	835	508	446	106

As can be seen from table 4.9, there were more words in the untimed dialogues than the timed dialogues for both the face-to-face setting (962 cf. 509) and the video-mediated setting (835 cf. 446).The results of the analysis of variance indicated that the difference in number of words between the untimed and timed conditions was reliable. The summary table for the analysis of variance is presented in table 4.10 below.

Table 4.10. Analysis of variance for number of words

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	1.02	0.32
S within-group error	30	(141525.30)	
Within subjects			
Cognitive Load (L)	1	26.62	*0.01
S x L	1	0.15	0.70
S x S within-group error	30	(106505.30)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of cognitive load [$F(1,30) = 26.62$, $p < 0.01$]. Overall, timed dialogues contained significantly fewer words (477) than untimed dialogues (897). There was no significant effect of conversational setting and no interaction ($F < 1$). Post hoc analyses were also carried out on the data. The results of Newman-Keuls pair-wise comparisons indicated that the effect of cognitive load was significant in both the face-to-face ($p < 0.01$) and video-mediated ($p < 0.01$) conversational settings. In both cases, dialogues contained fewer words under time pressure. There was no significant effect of conversational setting, however, in the untimed or timed conditions.

Once again, the results of these analyses indicate that the manipulation of time pressure served to increase cognitive load.

Summary

Taken together, the results of the analyses presented above indicated that participants responded to increased cognitive load (as illustrated by time pressure). Under time pressure, task performance was significantly worse, and dialogues were significantly shorter (in terms of duration and number of words) than without the pressure of time. Since cognitive load is reflected in performance, the finding that Instruction Follower's deviated more from the original route supports the assumption that time pressure increased cognitive load. However, the difference in task performance between the timed and untimed conditions was not reliable for the video-mediated setting. It is possible that this could be due to some kind of ceiling effect. In the case of the Map Task, the route is constructed around the landmarks on the map. Thus, it is unlikely that Instruction Followers will make extreme deviations from the original route since they are guided, to some extent, by the position of the landmarks on their own map.

The finding that participants took significantly longer and used more words to complete the task when they were instructed to perform the task in their own time indicates that the time limit of 3 minutes was sufficient to put participants under considerable pressure. This then provides further evidence to indicate that the manipulation of cognitive load was effective.

Although there was some suggestion in the relevant literature that video-mediation itself may contribute to cognitive load, there was no indication that task performance was influenced by the conversational setting. Nor was there any reliable effect of conversational setting on dialogue duration or the number of words used. Thus the notion that video-mediated itself may serve to increase cognitive load is not supported by the results of these analyses. This is perhaps not surprising when one considers the nature of the Map Task more carefully. In fact, studies which have examined patterns of eye gaze indicate that in a video-mediated conversational setting, participants tend to spend approximately 80% of the time looking at an on-screen version of the Map Task and only 15% of the time looking at the video-window of their participant (Clayes, 2000). This indicates that participants attend primarily to the task at hand. In the present study, the presentation of information in the video-mediated version of the task was similar to that employed by Clayes (2000) in her video-mediated versions of the Map Task. The Map Task was presented in one window and the video image of the other person was presented in a smaller window to one side of the task. In Clayes (2000) study, there was an additional window which displayed an image of a confederate. Given that participants spent most of their time attending to the Map Task, it was unlikely that this type of configuration of information would have induced a split-attention effect and thereby produced a significant increase in cognitive load.

4.3.2 Articulation

Rate of speech

In order to test the hypothesis that speaking rate would be faster under increased time pressure, the effect of cognitive load on rate of speech was examined in face-to-face and video-mediated dialogues. A measure of the rate of speech was derived by dividing the number of words in the dialogue by the duration of the dialogue. Thus giving the rate of speech as the number of words per second for each dialogue. Since spoken dialogue is the object of investigation here, the rate of speech for each dialogue was calculated rather than the speaking rate per person. In line with the Collaborative Model (Clark and Wilkes-Gibbs, 1986), the dialogue was used as the unit of analysis.

The normality of distribution scores was determined across the untimed and timed conditions for both the video-mediated and the face-to-face groups. Kolmogorov-Smirnov tests were carried out on the data and the results of these tests are presented in table 4.11 below. The Kolmogorov-Smirnov Z statistic is provided along with the significance of the p-value.

Table 4.11. Kolmogorov-Smirov tests for rate of speech

		Cognitive Load			
		Un-timed		*Timed*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
Face-to-face	32	0.55	ns.	0.76	ns.
Video-mediated	32	0.41	ns.	0.85	ns.

As can be seen from table 4.11, the results of the Kolmogorov-Smirnov test indicate that the distributions of the test samples were normal. All values of Z were non-significant. The samples were also tested for equality of variance. Levene's test of equality of variance revealed that the error variance of the dependent variable was equal across groups [F(3,60) = 0.80, p = 0.50].

The mean rates of speech for untimed and timed dialogues in face-to-face and video-mediated conversational settings are shown in figure 4.1.

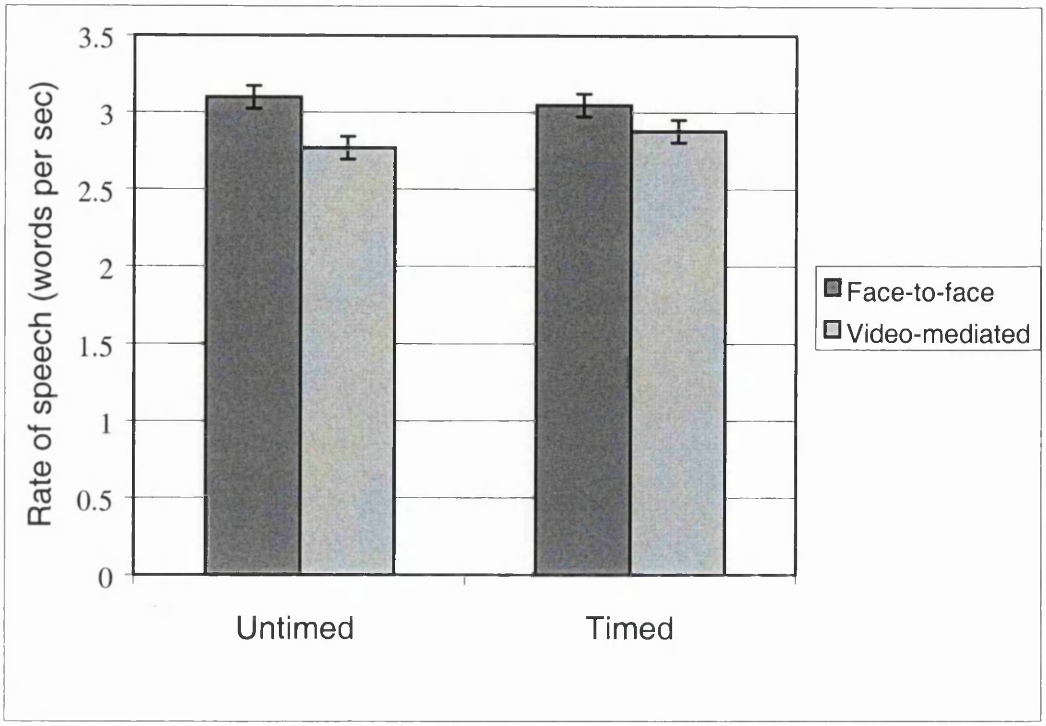


Figure 4.1. Mean speaking rate (± SE) for face-to-face (n = 32) and video-mediated (n = 32) conversational settings in untimed and timed conditions.

Surprisingly, as indicated in figure 4.1, there was a tendency towards a slower rate of speech in the video-mediated setting compared with the face-to-face setting. In the untimed conditions, mean rate of speech was 2.77 words per second for the video-mediated setting compared with 3.1 for the face-to-face setting. In the timed conditions, the mean rate of speech in the video-mediated setting was 2.88 words per second compared with 3.05 in the face-to-face setting. The results of the analysis of variance indicated that, overall, the difference in rate of speech between the video-mediated and the face-to-face conversational settings was reliable.

A 2 x (2) analysis of variance with conversational setting as a between-groups factor and cognitive load as a within-groups factor was carried out on the data. The summary table is presented in Table 4.12 below.

Table 4.12. Analysis of variance for rate of speech

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	5.84	0.02*
S within-group error	30	(0.176)	
Within subjects			
Cognitive Load (L)	1	0.27	0.61
S x L	1	1.83	0.19
S x S within-group error	30	(0.056)	

Note. Values enclosed in parentheses represent mean square errors

There was a main effect of conversational setting [F(1,30) = 5.84, p = 0.02]. Overall, the rate of speech was slower in the video-mediated conversational setting (mean words per second = 2.82) than in the face-to-face group (mean words per second = 3.08). Contrary to expectations, there was no effect of cognitive load (F < 1), nor was there a significant interaction.

Post hoc analyses were also carried out on the data to ascertain whether the effect of conversational setting was significant in both timed and untimed conditions. Newman-Keuls pairwise comparisons revealed that while the difference in rate of speech between the video-mediated and face-to-face settings was significant in the untimed condition (p > 0.05), the difference did not reach significance (at the 0.05 level) in the timed condition. There were no significant effect of cognitive load in either the face-to-face or video-mediated conversational settings.

A problem with this analysis, however, relates to the fact that the measure of speaking rate used here is rather crude. The rate of speech was calculated by dividing the number of words per dialogue by the duration of the dialogue. Furthermore, the unit of analysis was the dialogue itself rather than the individual speaker. This does not take into account the fact that speakers may talk simultaneously. Consequently, the effect of conversational setting could be due to more overlapping speech in a face-to-face setting compared with a video-mediated setting (cf. Sellen, 1995). However, the following analysis will help to clarify this issue.

Articulatory Reduction

To test the hypothesis that articulatory reduction would occur irrespective of conversational setting or cognitive load, the effect of repeated mention and cognitive load on word duration was examined in face-to-face and video-mediated settings. It was expected that 2nd mentions of words (forming the names of landmarks of the map) would be shorter in duration than 1st mentions regardless of cognitive load or conversational setting. Furthermore, if the results of the previous analysis were due to speed of articulation rather than overlapping speech, then overall, words uttered in a video-mediated setting should be longer in duration than words uttered in a face-to-face setting.

The data was first analysed by subject. From a total of 64 speakers, 33 failed to produce examples of 1st and 2nd mentions of word tokens in both the timed and untimed conditions (17 from the video-mediated group and 16 from the face-to-face group). Of the remaining speakers, 7 were rejected at random in order to preserve the balance of the original design, which was balanced for order of timing condition and the version of map used (Map A or Map B). Thus, 24 speakers were included in the analysis. 1st and 2nd mentions of word tokens were compared in untimed and timed, face-to-face and video-mediated conversational settings. The word tokens were all uttered as part of the same landmark name.

Only 1st and 2nd mentions of word tokens uttered by the same speaker were included in the analysis.

Word duration was measured in milliseconds using speech analysis software (Syntrillium waveform editor *Cool Edit*) following the procedure outlined in Section 3.2.2 of Chapter 3. Reliability of measurement was assessed using Pearson's correlation coefficient. Twelve dialogues (6 face-to-face and 6 video-mediated) were selected at random and the duration of word tokens was measured a second time by the experimenter. Reliability of measure was highly correlated ($r = 0.98$; $p < 0.001$, $N = 120$).

The normality of distribution scores was determined across the untimed and timed conditions for both the video-mediated and the face-to-face settings. Kolmogorov-Smirnov tests were carried out on the data and the distribution of the dependent variable was found to be normal for the test samples. All values of Z were non-significant. The samples were also tested for equality of variance. Levene's test of equality of variance revealed that the error variance of the dependent variable was equal across groups [$F(3,92) = 0.11$, $p = 0.95$].

Mean word durations for 1st and 2nd mentions of words in face-to-face and video-mediated conversational settings (collapsed across cognitive load conditions) are illustrated in figure 4.2 below.

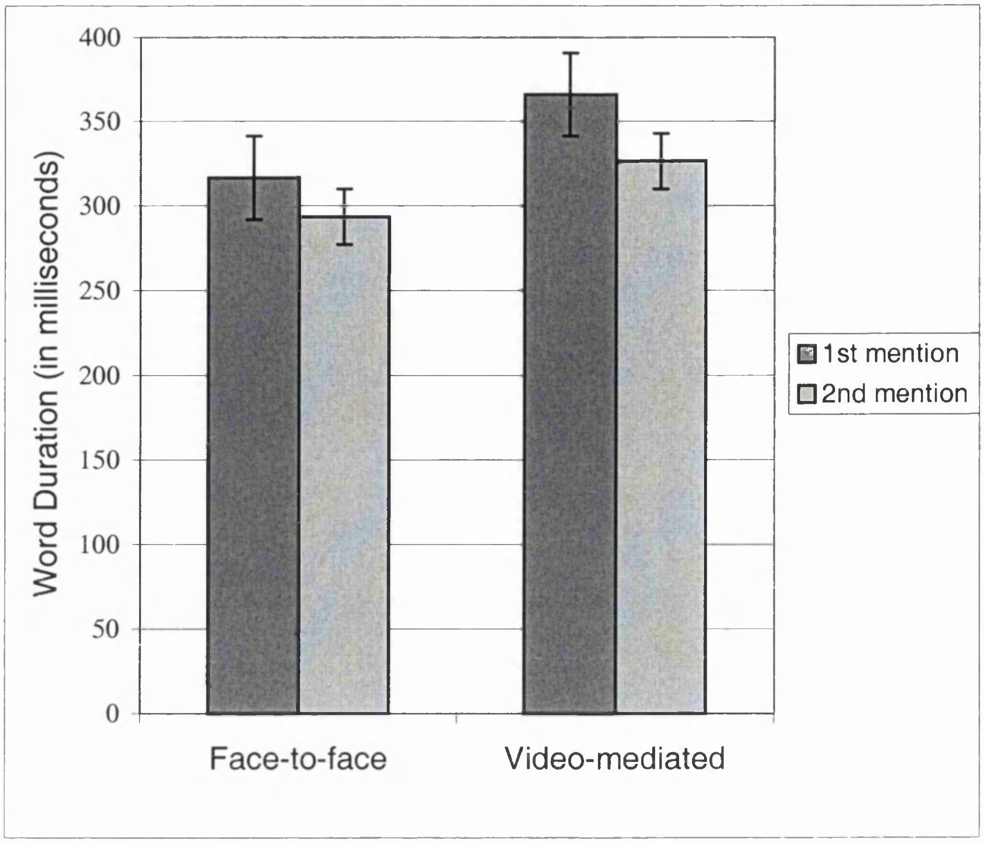


Figure 4.2. Mean word duration (±SE) for 1st and 2nd mentions of word tokens in face-to-face (n = 24) and video-mediated conversational settings (n = 24), collapsed across untimed and timed conditions.

As can be seen in figure 4.2, word duration was shorter for 2nd mentions of word tokens than for 1st mentions in the face-to-face conversational setting (mean M1 duration = 316ms; mean M2 duration = 293ms) and in the video-mediated setting (mean M1 duration = 366ms; mean M2 duration = 326ms). Figure 4.2 also indicates a tendency for words of longer duration in the video-mediated setting compared with the face-to-face setting. The results of the analysis of variance confirmed that the overall differences in duration between 1st and 2nd mentions of word tokens and between word tokens uttered in a video-mediated compared with a face-to-face conversational setting were reliable.

A 2 x (2) x (2) analysis of variance (by-speaker) was carried out on the data with conversational setting as a between-subjects factor, and cognitive load and mention (1st versus 2nd) as within-subjects factors. The summary table is shown in table 4.13.

Table 4.13. Analysis of variance for articulatory reduction

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	6.11	*0.02
S within-group error	22	(6638.85)	
Within subjects			
Cognitive Load (L)	1	2.11	0.16
Mention (M)	1	20.47	*<0.01
S x L	1	0.98	0.33
S x M	1	1.39	0.25
M x L	1	2.25	0.15
S x L x M	1	2.31	0.14
Error L x S	22	(5266.46)	
Error M x S	22	(1149.70)	
Error LM x S within-group error	22	(689.39)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of mention [F(1,22) = 20.47; p < 0.01]. Overall, 2nd mentions of word tokens were articulated more quickly (mean word duration = 310ms) than 1st mentions (mean word duration = 341ms). There was also a main effect of conversational setting [F(1,22) = 6.11; p = 0.02]. Word tokens were articulated more slowly in the video-mediated setting (mean word duration = 346 ms) than in the face-to-face setting (mean word duration = 305 ms). There was no significant effect of cognitive load and there were no significant interactions.

Post hoc analyses were also carried out on the data. The results of Newman-Keuls pairwise comparisons indicated that the effect of mention was significant for both the face-to-face conversational setting ($p < 0.05$) and for the video-mediated setting ($p < 0.01$). While the effect of conversational setting was significant for 1st mentions of word tokens ($p < 0.05$), the effect did not reach significance for 2nd mentions of word tokens. Newman-Keuls revealed no significant effect of cognitive load at either level of conversational setting or mention.

An analysis (by-item) was also carried out on the data. First, the test samples were tested for normality of distribution. Kolmogorov-Smirnov tests were carried out on the data and the results of these analyses are shown in table 4.14 below.

Table 4.14. Kolmogorov-Smirov tests for duration of words forming landmark names (data organised by item)

		Mention			
		1st Mention		*2nd Mention*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
Face-to-face untimed	211	1.12	ns.	1.41	*0.04
Face-to-face timed	116	1.47	*0.03	1.61	*0.01
Video-mediated untimed	157	1.23	ns.	1.37	*0.05
Video-mediated timed	88	1.08	ns.	0.63	ns.

As can be seen from Table 4.14, the results of the Kolmogorov-Smirnov test indicated that the distribution of several test samples (FTF timed M1; FTF untimed and timed M2, and the VM untimed M2) was not normal. Levene's test of equality of variance was also carried out on test samples and the error variance associated with the test samples was found to be homogeneous [$F(7,1136) = 1.11$, $p = 0.35$].

It is widely considered that analysis of variance is a robust test, where even major deviations from the underlying assumptions are unlikely to reduce the power of the test. However, according to Roberts and Russo (1999), this robustness is only true as long as a design with equal cell sizes is used. Consequently, given that the data did not meet with the normality of distribution assumption underlying analysis of variance, items were rejected at random using SPSS in order to obtain test samples of equal size for each cell of the design. 352 word pairs (88 word pairs in each cell of the design) were included in the analysis.

Mean word durations with standard deviation for 1st and 2nd mentions of words in face-to-face and video-mediated conversational settings (collapsed across cognitive load conditions) are illustrated in table 4.15 below.

Table 4.15. Mean word duration (in milliseconds) for 1st and 2nd mentions of word tokens in face-to-face and video-mediated conversational settings (collapsed across untimed and timed conditions).

		Cognitive Load			
		1st Mention		*2nd Mention*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	176	335	131	307	119
Video-mediated	176	384	143	360	146

As can be seen from table 4.15, word duration was shorter for 2nd mentions of word tokens than for 1st mentions for both face-to-face (mean M1 duration = 335ms; mean M2 duration = 307ms) and video-mediated (mean M1 duration = 384 ms; mean M2 duration = 360 ms) conversational settings. Furthermore, table 4.15 also indicates that the mean word duration was longer in the video-mediated setting compared with the face-to-face setting for both 1st and

2^{nd} mentions of word tokens. The results of the analysis of variance confirmed that the overall differences in duration between 1^{st} and 2^{nd} mentions and between word tokens uttered in a video-mediated compared with a face-to-face conversational setting were reliable.

A 2 x 2 x (2) analysis of variance (by-item) was carried out on the data with conversational setting and cognitive load treated as between-subjects factors and mention as a within-subjects factor. Following the procedure used by Bard et al., (2000) individual items were used as the unit of analysis. The summary table for the analysis of variance is shown in table 4.16 below.

Table 4.16. Analysis of variance for articulatory reduction (by item)

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	14.27	*<0.01
Cognitive Load (L)	1	0.60	0.44
S x L	1	0.45	0.50
S within-group error	348	(32355.61)	
Within subjects			
Mention (M)	1	28.33	*<0.01
S x M	1	0.20	0.66
L x M	1	1.44	0.23
S x L x M	1	2.53	0.11
Error L x S	348	(4277.47)	

Note. Values enclosed in parentheses represent mean square errors

In common with the by-speaker analysis, there was a main effect of conversational setting [F(1,348) = 14.27; p < 0.01]. Overall, words uttered in a video-mediated setting were longer in duration (mean duration = 372msec) than words uttered in a face-to-face setting (321msec). As expected, there was a main effect of mention [F(1,348) = 28.33, p < 0.01]. Second mentions of word tokens were shorter in duration (333msec) than first mentions (360msecs). There was no effect of cognitive load (F <1) and there were no significant interactions.

Post hoc analyses were also carried out on the data. The results of Newman-Keuls pairwise comparisons indicated that the effect of mention was significant for both the face-to-face conversational setting ($p < 0.01$) and for the video-mediated setting ($p < 0.01$). The effect of conversational setting was significant for 1st mentions of word tokens ($p < 0.05$) and, in contrast to the by-speaker analysis, also reached significance for 2nd mentions of word tokens ($p< 0.01$). This could have been due to the larger number of items in the by-items analysis. There was no significant effect of cognitive load at either level of conversational setting or either level of mention.

Overall, the results of the by-items analysis replicated the results of the by-speaker analysis.

Summary

The results of the rate of speech analysis suggested that, contrary to expectation, speakers did not speak more quickly under time pressure. There was no significant difference in rate of speech between dialogues performed under time pressure and dialogues performed without the pressure of time. Interestingly, however, the rate of speech was found to be slower in the video-mediated conversational setting than in the face-to-face setting. On the basis of the rate of speech analysis alone it was possible that the effect could have been due to more overlapping speech in the face-to-face setting than in the video-mediated setting. This has been observed in previous comparisons of face-to-face and video-mediated communication (e.g. Sellen, 1995). However, the results of the word duration analysis indicated that, overall, word tokens representing initial mentions (1st and 2nd) of words forming landmark names were articulated more slowly in a video-mediated conversational setting than in a face-to-face conversational setting. Although this was found to be the case for 1st mentions of words forming landmark names, with respect to 2nd mentions the effect was significant in the by-items analysis but did not reach significance in the by-speaker analysis.

Overall, the results of the articulatory reduction analysis indicated that the observed effect of conversational setting on rate of speech was likely to be due to rate of articulation rather than to overlapping speech. Taken together, the results of the analyses of rate of speech and articulatory reduction support the interpretation that participants in a video-mediated setting spoke more slowly than participants in a face-to-face setting.

The results of the word duration analysis support the hypothesis that articulatory reduction occurs irrespective of the conversational setting or cognitive load. Speakers reliably articulated 2nd mentions of words more quickly than 1st mentions of words referring to the same object. This finding is consistent with the Dual Process Model (Bard et al., 2000) which holds that articulatory reduction is due to automatic priming processes that occur within a very fast planning cycle. For this reason, articulatory priming is unlikely to be affected, even if the demands on the speaker's time and attention are increased.

4.3.3 The Shortening of Landmark Names

In order to test the hypothesis that participants would shorten referring expressions under time pressure, the effect of cognitive load on referential forms was examined in face-to-face and video-mediated dialogues. It was expected that, under time pressure, participants would use shorter names to refer to landmarks on the map. A higher proportion of truncated landmark names and pronouns in the timed condition compared with the untimed condition would reflect this. Only 2nd references to landmarks were included in this analysis on the basis that 2nd references would have the greatest scope for shortening.

In order to yield a dependent variable that would be measurable on an interval scale, the proportion of 2nd references to landmarks whose names were shortened (either truncated or substituted with a pronoun) was employed as the dependent variable. The data were organised by speaker and by item. The data organised by speaker is presented first. Eight speakers (4 from the video-mediated group and 4 from the face-fo-face group) were excluded from the analysis since they produced fewer than 2 examples of repeated references to landmarks in at least one condition of cognitive load. 56 speakers were included in the by-speaker analysis. The sample was drawn from references to both shared and unshared landmarks.

Kolmogorov-Smirnov tests were carried out on the data to determine whether the distribution associated with the dependent variable was normal across the untimed and timed, face-to-face and video-mediated groups. The results of the Kolmogorov-Smirnov tests indicated that the distribution of the test samples was normal. All values of Z were non-significant. Levene's test of homogeneity of variance, carried out on the data, indicated that the error variance of the dependent variable was equal across groups [F(3,108) = 0.47, p = 0.71]. The mean proportions of 2nd references to landmarks that had been shortened from the name on the map (either truncated or substituted with a pronoun) with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 4.17 below.

Table 4.17. Mean proportion (by-speaker) of shortened 2nd references to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	28	0.52	0.26	0.45	0.32
Video-mediated	28	0.49	0.28	0.55	0.30

As can be seen from table 4.17, in the face-to-face setting, the proportion of shortened landmark names was slightly higher in the untimed condition (0.52) than in the timed condition (0.45). In the video-mediated conversational setting, the pattern is reversed. The proportion of shortened referring expressions was slighly higher in the timed condition (0.55) than in the untimed condition (0.49). However, the results of the analysis of variance indicated that there were no reliable differences between the untimed and timed conditions.

A 2 x (2) analysis of variance (by-speaker) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factors. The summary table is shown in table 4.18 below.

Table 4.18. Analysis of variance (by-speaker) for the proportion of shortened 2nd references to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	0.39	0.53
S within-group error	54	(0.09)	
Within subjects			
Cognitive Load (L)	1	0.01	0.98
S x L	1	1.43	0.24
C x S within-group error	54	(0.08)	

Note. Values enclosed in parentheses represent mean square errors

As can be seen from table 4.18, there was no effect of conversational setting (F < 1), no effect of cognitive load (F < 1), and no significant interaction.

The data were also analysed by-item. The data for the landmark, *broken gate*, was excluded from the analysis since there was only a single reference to this landmark in the video-mediated untimed condition.

Kolmogorov-Smirnov tests carried out on the data indicated that the distribution of the test samples was normal. All values of Z were non-significant. Levene's test of homogeneity of variance indicated that the error variance of the dependent variable was equal across groups [F(3,108) = 0.58, p = 0.63]. The mean proportions of 2nd references to landmarks that had been shortened from the name on the map (either truncated or substituted with a pronoun) with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 4.19 below.

Table 4.19. Mean proportion (by-item) of shortened 2nd references to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	28	0.46	0.26	0.48	0.27
Video-mediated	28	0.47	0.28	0.57	0.22

As can be seen from table 4.19, in the video-mediated conversational setting, the proportion of shortened referring expressions was higher in the timed condition (0.57) than the untimed condition (0.47). In the face-to-face setting, there was little difference between the timed condition (0.48) and the untimed condition (0.46). However, the results of the analysis of variance indicated that there were no reliable differences between the untimed and timed conditions.

A 2 x 2 analysis of variance (by-item) was carried out on the data with conversational setting and cognitive load treated as between-subjects factors. The summary table is shown in table 4.20 below.

Table 4.20. Analysis of variance (by-item) for the proportion of shortened 2nd references to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	0.96	0.33
Cognitive Load (L)	1	1.60	0.21
S x L	1	0.58	0.45
S within-group error	108	(0.07)	

Note. Values enclosed in parentheses represent mean square errors

As can be seen from table 4.20, there was no effect of conversational setting (F < 1), no significant effect of cognitive load, and no interaction (F < 1).

The results of the by-speaker and by-item analyses did not support the hypothesis that time pressure led to a greater use of short referring expressions.

Nevertheless, descriptive statistics indicated that more pronouns were used in repeated references to landmarks in the video-mediated timed condition compared with other conditions. The percentages of referring expressions used in each coding category for timed and untimed, video-mediated and face-to-face dialogues are shown in table 4.21 below.

Table 4.21. Percentages of 2nd-reference referring expressions for referential coding categories in untimed and timed, face-to-face and video-mediated conversational settings

		Face-to-face		Video-mediated	
Code		*Untimed*	*Timed*	*Untimed*	*Timed*
0	Name on map	56.1	54.7	55.7	44.2
1	Truncated name	4.2	8.6	8.5	5.8
2	Pronoun	39.7	36.7	35.7	50

As can be seen from table 4.21, in the video-mediated setting, 50% of 2nd references to landmarks were pronouns in the timed condition compared with 35.7% in the untimed condition. Could it be the case that increased cognitive load led to a greater use of pronouns in a video-mediated setting? In order to test this hypothesis, an analysis of the use of pronouns was conducted.

The data were first analysed by-speaker. The proportion of pronominal 2nd references to landmarks was calculated for each speaker and used as the dependent variable. Eight speakers (4 from the video-mediated group and 4 from the face-fo-face group) were excluded from the analysis since they produced fewer than 2 examples of repeated references to landmarks in at least one condition of cognitive load. 56 speakers were included in the by-speaker analysis. The sample was drawn from references to both shared and unshared landmarks.

Kolmogorov-Smirnov tests carried out on the data indicated that the test samples were normally distributed. All values of Z were non-significant. Levene's test of homogeneity of variance indicated that the error variance associated with the dependent variable was equal across groups [F(3,108) = 0.35, p = 0.79].

The mean proportions of pronouns with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 4.22 below.

Table 4.22. Mean proportion of pronouns (by-speaker) for face-to-face and video-mediated conversational settings in untimed and timed conditions.

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	28	0.49	0.27	0.40	0.32
Video-mediated	28	0.43	0.27	0.51	0.30

As can be seen from table 4.22, in the video-mediated setting, the proportion of pronouns was higher in the timed condition (0.51) than in the untimed condition (0.43). In the face-to-face conversational setting, however, the opposite pattern was observed. The proportion of pronouns was lower in the timed condition (0.40) than the untimed condition (0.49). However, analyses of variance revealed that the observed differences between the untimed and timed conditions were not reliable. The summary table is shown in table 4.23 below.

Table 4.23. Analysis of variance (by-speaker) for the use of pronouns in 2nd references to landmarks

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	0.30	0.59
S within-group error	54	(0.11)	
Within subjects			
Cognitive Load (L)	1	0.02	0.89
S x L	1	1.76	0.19
S x S within-group error	54	(0.10)	

Note. Values enclosed in parentheses represent mean square errors

As can be seen from table 4.23, there was no effect of conversational setting (F < 1) or cognitive load (F < 1) and no significant interaction.

The data were also analysed by-item. Table 4.24 shows the mean proportions of pronouns with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions.

Table 4.24. Mean proportion of pronouns (by-item) for face-to-face and video-mediated conversational settings in untimed and timed conditions.

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	28	0.44	0.26	0.45	0.28
Video-mediated	28	0.42	0.29	0.51	0.26

As can be seen from table 4.24, for the video-mediated conversational setting, the proportion of pronouns was higher in the timed condition (0.51) than in the untimed condition (0.42). However, analyses of variance carried out on the data indicated that this difference was not reliable.

A 2 x 2 analysis of variance (by-item) was carried out on the data with conversational setting and cognitive load treated as between-subjects factors. The summary table is shown in table 4.25 below.

Table 4.25. Analysis of variance (by-item) for the proportion of pronouns for 2nd references to landmarks

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	0.11	0.74
Cognitive Load (L)	1	1.01	0.32
S x L	1	0.77	0.38
S within-group error	108	(0.08)	

Note. Values enclosed in parentheses represent mean square errors

As can be seen from table 4.25, there was no effect of conversational setting (F < 1) or cognitive load (F < 1) and no significant interaction.

The results of the by-speaker and by-item analyses did not support the hypothesis that time pressure led to an increased use of pronouns in the video-mediated conversational setting.

4.3.4 The Quantity of References to Landmarks

In section 4.3.1, it was found that participants used significantly fewer words to complete the Map Task under time pressure compared with no pressure of time, in both the face-to-face and video-mediated conversational settings. This raises the question of how participants economised in terms of the words they used. One way in which participants could have made savings, was simply by mentioning fewer landmarks and by making fewer references to those landmarks. In order to test these hypotheses, the effect of conversational setting and cognitive load on the number of landmarks mentioned and on the number of references per landmark was examined. It was expected that, under time pressure, fewer landmarks would be mentioned with fewer references to each of those landmarks. This was expected to be the case in both the face-to-face and video-mediated conversational setting.

The maximum number of landmarks on a map was 15. Under time pressure, not all participants finished the task. In such cases, those landmarks occurring towards the end of the route may not have been referred to because the participants had not reached that stage on the route. Consequently, an adjustment factor was calculated for those cases where the task had not been completed. The adjustment factor was calculated by expressing (unmentioned) landmarks after the point reached on the map, as a ratio of the total number of landmarks on the map. The number of landmarks mentioned in the dialogues was then multiplied by the adjustment factor. The mean number of references per landmark was also calculated for each dialogue.

Before proceeding with the analysis of variance the samples were tested for homogeneity of variance and normality of distribution for each of the dependent variables. Kolmogorov-Smirnov tests indicated that the test samples were normally distributed. All values of Z were non-significant for both dependent variables. Levene's test of equality of variance was also carried out on the data. The error variance of the samples was equal for the dependent variable *landmarks mentioned* [F(3,60) = 0.57, p = 0.64] but not for the dependent variable *references per landmark* [F(3,60) = 5.65, p = 0.02]. However, given that the sample sizes were equal in each cell of the design, it was unlikely that a violation of the homogeneity of variance assumption, would lead to type I or type II errors (Roberts and Russo, 1999). Thus analysis of variance was carried out on the data. In accordance with the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992), the dialogue was used as the unit of analysis.

Number of landmarks mentioned

Figure 4.3 shows the mean number of landmarks mentioned for face-to-face and video-mediated conversational settings in untimed and timed conditions.

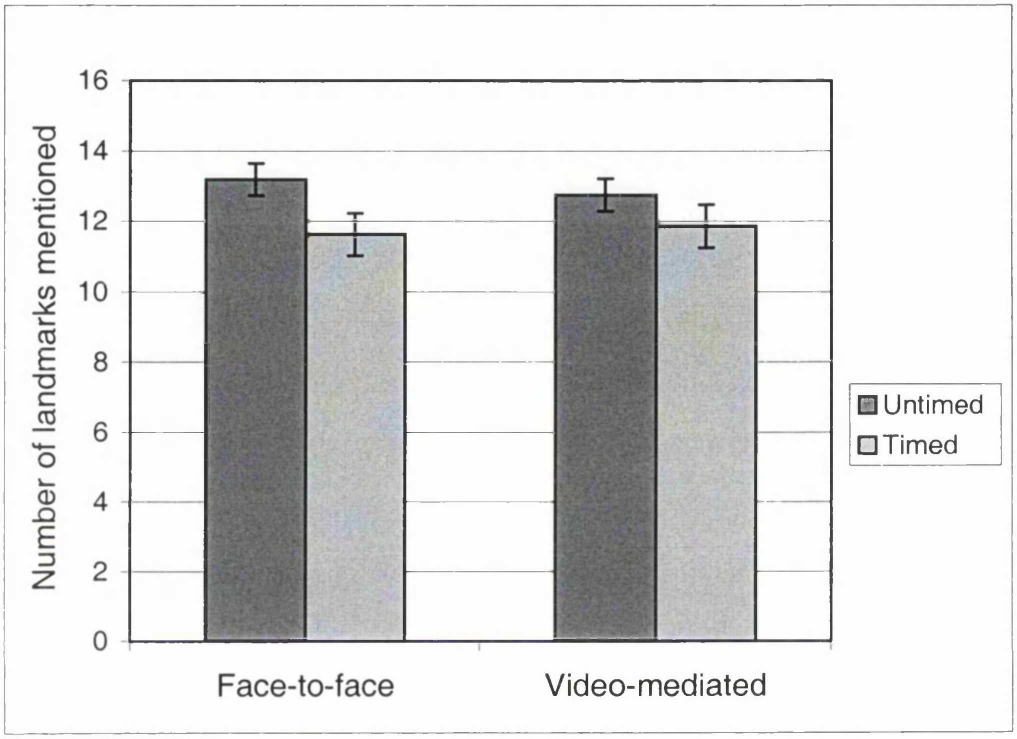


Figure 4.3. Mean number of landmarks mentioned (± SE) for face-to-face (n = 16) and video-mediated (n = 16) conversational settings in untimed and timed conditions.

As can be seen from figure 4.3, in the face-to-face setting the mean number of landmarks mentioned was lower (11.6) under time pressure than without the pressure of time (13.2). This pattern was also observed for the video-mediated setting. The mean number of landmarks mentioned in the timed condition was 11.8, compared with 12.75 in the untimed condition. The results of the analysis of variance indicated that these differences were reliable.

A 2 x (2) analysis of variance (by-dialogue) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The number of landmarks mentioned was the dependent variable. The summary table is presented in Table 4.26 below.

Table 4.26. Analysis of variance (by dialogue) for the number of landmarks mentioned

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	0.03	0.87
S within-group error	30	(6.94)	
Within subjects			
Cognitive Load (L)	1	10.43	*<0.01
S x L	1	0.77	0.38
S x S within-group error	30	(2.29)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of cognitive load [F(1,30) = 10.43, p < 0.01]. Overall, fewer landmarks were mentioned in the timed dialogues (11.75) than in the untimed dialogues (12.97). There was no effect of conversational setting (F < 1) and no interaction (F < 1). Given the expectation that cognitive load would affect the number of landmarks mentioned in both conversational settings, an analysis of simple main effects was carried out on the data irrespective of a non-significant interaction. There was a significant effect of cognitive load for the face-to-face setting [F(1,30) = 8.43, p = 0.007, MS error = 2.29] and a marginal effect of cognitive load in the video-mediated setting [F(1,30) = 2.77, p = 0.10, MS error = 2.29]. This indicated that participants mentioned significantly fewer landmarks under time pressure although this difference was marginal for the video-mediated setting.

Number of references per landmark

Figure 4.4 illustrates the effect of cognitive load on the number of references per landmark in the face-to-face and video-mediated conversational settings.

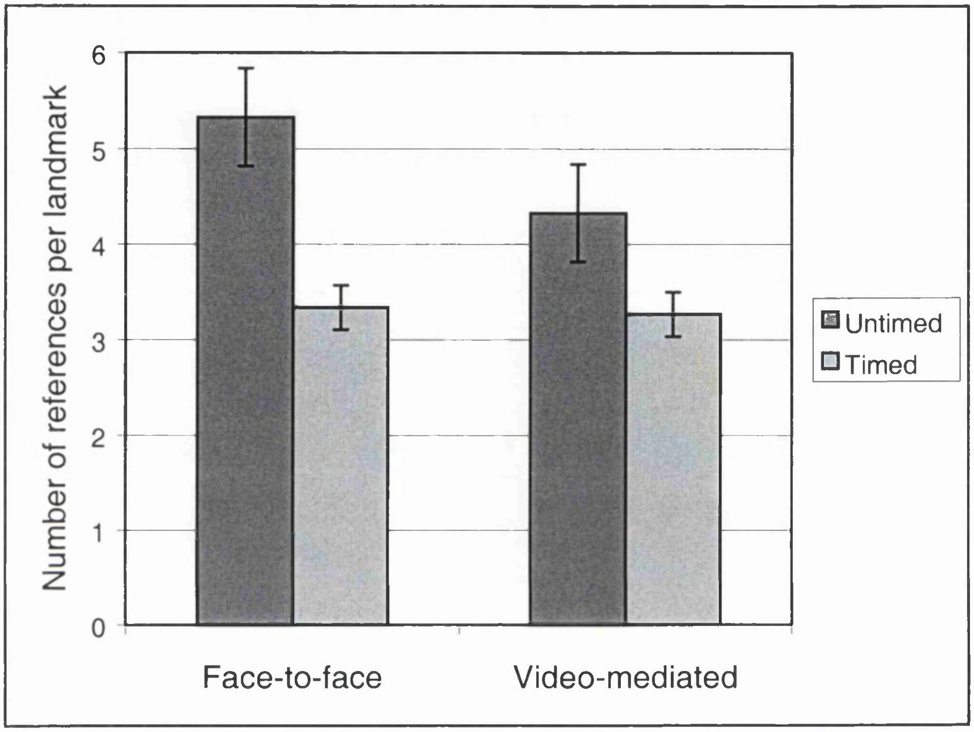


Figure 4.4. Mean references per landmark (± SE) for face-to-face (n = 16) and video-mediated (n = 16) conversational settings in untimed and timed conditions.

As can be seen from figure 4.4, there were fewer references per landmark under time pressure in both the face to face conversational setting (mean untimed references = 5.33; mean timed references = 3.34) and video-mediated conversational setting (mean untimed references = 4.33; mean timed references = 3.37). These observations were supported by the results of the analysis of variance.

A 2 x (2) analysis of variance (by-dialogue) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The number of references per landmark was the dependent variable. The summary table is presented in Table 4.27 below.

Table 4.27. Analysis of variance (by dialogue) for references per landmark

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	1.68	0.20
S within-group error	30	(2.76)	
Within subjects			
Cognitive Load (L)	1	15.94	*<0.01
S x L	1	1.47	0.23
S x S within-group error	30	(2.34)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of cognitive load [F(1,30) = 15.94, p < 0.01]. Overall, participants made fewer references to the landmarks under time pressure (3.31) than without the pressure of time (4.83). There was no significant effect of conversational setting and no significant interaction. An analysis of simple main effects was also carried out on the data irrespective of a non-significant interaction. There was a simple main effect of cognitive load in both the face-to-face setting [F(1,30) = 10.19, p < 0.01, MS error =2.34] and the video-mediated setting [F(1,30) = 6.62, p = 0.01, MS error = 2.34].

Thus, the results of these analyses indicate that, in both conversational settings participants responded to time pressure by mentioning fewer landmarks and by making fewer references to those landmarks.

4.4 General Discussion

4.4.1 Summary of Findings

Before discussing the implications of the findings reported in this chapter, I will first summarise the main results. The manipulation of cognitive load by imposing a time pressure was effective. Task performance was worse under time pressure compared with when there was no pressure of time. Furthermore, the findings indicated that a 3-minute time limit was appropriate to put participants under considerable pressure. Dialogues performed under time pressure were significantly shorter in duration and contained significantly fewer words than dialogues performed without the pressure of time.

With respect to articulation, contrary to expectation, there was no indication that participants spoke more quickly under time pressure. Surprisingly, it was found that the rate of speech was slower for dialogues in a video-mediated conversational setting than in a face-to-face setting. This was consistent with the finding that specific words forming the names of landmarks on a map were articulated more slowly in a video-mediated conversational setting than in a face-to-face setting. With respect to articulatory reduction, as expected, it was found that 2nd mentions of words were articulated more quickly than 1st mentions of words irrespective of cognitive load or conversational setting. The finding that 2nd mentions of words were shorter in duration than 1st mentions of words replicates the articulatory reduction effect reported in Chapter 3. These findings were consistent with the Dual Process Model (Bard et al., 2000) which holds that articulatory reduction is underpinned by fast, automatic priming processes that are unaffected by the cognitive demands on the speaker's time and attention.

In relation to reference, the findings of Study 2 did not support the hypothesis that time pressure would lead speakers to shorten the names of landmarks on the map (from complex noun phrases to pronouns, for instance). Nevertheless, the way interlocutors referred to landmarks on the map was influenced by cognitive load in a more general way. In light of the observation that time pressure led to dialogues containing significantly fewer words than dialogues performed without time pressure, the question was posed as to how participants might have economised in terms of the way they referred to objects. It was found that, overall, participants mentioned fewer landmarks on the map and made fewer references to those landmarks under time pressure than without the pressure of time. This reflects one way in which a saving was made (in terms of the number of words) in relation to reference.

4.4.2 Cognitive Load and Articulation

It was hypothesised that increased cognitive load may be reflected by an increase in speaking rate. There was some evidence in the literature that, for some speakers at least, cognitive load led to an increase in rate of speech (Lively, et al., 1993). Furthermore, in the present study, cognitive load was manipulated by imposing a time pressure. It seemed reasonable to suppose that one way in which speakers might respond to time pressure would be to speak more quickly. Yet the results of Study 2 revealed no indication that cognitive load (as implemented by time pressure) influenced the rate of speech, or the duration of specific words.

Nevertheless, it is not too surprising that the findings of the present study appear to contradict the findings reported by Lively et al., (1993). First, the study conducted by Lively et al., (1993) was based on a small sample of only 5 male speakers. The data for each speaker was analysed separately and an effect of cognitive workload was exhibited by 4 of the five speakers, with one speaker showing the opposite pattern. Second, the study was based on single speakers uttering the rather meaningless phrase “Say hVd again.” Thus, it was not necessarily the case that the finding that cognitive load led to an increase in speaking rate would generalise across speakers communicating in a naturalistic conversational setting.

Interestingly, the present results indicated that speakers responded to the conversational setting by speaking more slowly in a video-mediated setting than in a face-to-face setting. Why should this be the case? One possible explanation makes recourse to Lindblom’s (1990) theory of Hypo- and Hyper-articulation and Lindblom’s (1995) proposal that speakers adjust their articulation in response to the demands of the environment. Such modifications are designed to maximise intelligibility. In other words, speakers attempt to balance the demands on the articulatory system with the need to communicate efficiently under different environmental conditions. When the demands on the system are low, speakers devote relatively few resources to system—oriented control. This results in *hypospeech,* or speech that is more economical in terms of articulatory effort. When the demands on the system are high, then discriminability and intelligibility must be maximised and more articulatory effort is required. This is known as *hyperarticulation*. Applying this framework to the present results, one might speculate that speakers adjusted articulation in response to a video-mediated environment. It could have been the case that speakers perceived a potential communication difficulty in this conversational setting. Previous researchers have suggested that a video-mediated conversational setting may be perceived as novel (e.g. Blokland and Anderson, 1998). Alternatively, speakers may perceive a communication difficulty because they feel a sense of social distance (Short, Williams, Christie, 1976) from the other person and the communicative situation in a video-mediated conversational setting. Consequently, the speaker could have adjusted his or her articulatory effort by hyperarticulating to ensure that s/he was understood in this unfamiliar and remote setting.

In support of this notion a similar finding has been reported in the literature. Blokland and Anderson (1998) compared introductory mentions of words forming the names of landmarks on a map in video-mediated and audio only conversational settings. Although they observed no difference in the duration of words, it was found that introductory mentions of words excerpted from a video-mediated setting were more intelligible to listeners than words uttered in a corresponding audio only setting. This observation offers some suggestion that the setting in which a dialogue takes place can affect the degree to which a speaker may hypo- or hyperarticulate. However, as pointed out by Blokland and Anderson (1998), the results of their study could have been due to the poor quality of the video image. This could not have accounted for the results of the present study, however, since a high quality videoconferencing set up was used.

A problem with Lindblom's model, resides in the assertion that signal-complementary processes are dynamic and cause changes in the balance of system-oriented control and output-oriented control (Lindblom, 1990). This implies that speakers make a running estimation of the contribution that signal complementary processes will make. In the context of the present findings, it seems unreasonable to suppose that, in a video-mediated setting, a speaker would continue to estimate the cost of such an assessment throughout the course of the dialogue. A criticism of Lindblom's H& H theory is that continuous updating of a dialogue model must place untenable computational costs on the speech production system. However, one might speculate that when faced with an unfamiliar and remote video-mediated environment, the speaker would make a conscious, default decision to articulate more slowly. If this were the case, then one might expect the rate of speech in a video-mediated conversational setting to reliably differ from the rate of speech in a face-to-face setting at difference stages of the dialogue such as the beginning, middle and end. This suggests an interesting line of investigation for future work.

With respect to articulatory reduction, the finding that speakers articulated words more quickly on repeated mention regardless of cognitive load or conversational setting is consistent with the Dual Process Hypothesis (Bard et al., 2000). According to the Dual Process Model, articulatory reduction is due to priming processes, which can be triggered by previous mention. These processes are automatic and occur within a very fast planning cycle. For this reason, they are deemed too fast for other factors, such as time pressure or any communication difficulty arising from the conversational setting, to make their influence felt on a word-by-word basis.

In terms of articulation then, the findings of Study 2 suggest a distinction between automatic processes and consciously controlled processes (Bard et al., 2000; Roßnagel, 2000) at the level of articulation. As suggested above, in contrast to articulatory reduction, the slower articulation in a video-mediated setting may represent a consciously controlled default decision to articulate more slowly. Interestingly, according to Cognitive Load Theory (Sweller, 1988, Sweller, van Merriënboer and Paas,1998), controlled processes can become automated by incorporating production rules into schemata. Schemata can integrate information elements and production rules, which then become automated, and require less storage and controlled processing. Thus, one might further speculate that once a decision is taken, to control articulation, subsequent regulation of the rate of speech may become a relatively automated process from that point on in the dialogue.

4.4.3 Cognitive Load and Reference

Although no evidence was found to suggest that cognitive load influences articulation, cognitive load did influence the way speakers referred to objects. Increased cognitive load led speakers to refer to fewer landmarks on the map and make fewer references to those landmarks. Why should cognitive load influence the way references to objects are made but not articulatory reduction? The Dual Process Model, goes some way to explaining why this might be the case. According to the Dual Process Model (Bard et al., 2000), articulatory reduction is thought to be underpinned by priming processes which are deemed to be automatic and occur too quickly for other factors, such as increased cognitive load or any response to the conversational setting, to make their influence felt. In contrast, more complex processes, such as those involved in task planning, are thought to be controlled and are therefore subject to the cognitive demands on the speaker's time and attention. In the case of references to objects, it seems reasonable to suppose that referring to fewer objects and making fewer references to those objects is indicative of a strategy, presumably consciously controlled, for coping with the pressure of time. For this reason, the number of references to landmarks was influenced by cognitive load but not by the conversational setting. One might further suppose that, if this observation is indicative of a strategy for coping with time pressure, then the same type of strategy may not be appropriate for dealing with other types of increases in cognitive load, such as an increase in task difficulty for instance.

4.4.4 Face-to-face and Video-mediated Conversations

The results of Study 2 suggest that the main difference between video-mediated communication and face-to-face communication lies in the rate of articulation. Speakers in a video-mediated conversational setting spoke more slowly than speakers in a face-to-face setting. In section 4.4.2, it was speculated that interlocutors may perceive a potential communication difficulty in a video-mediated environment and articulate more slowly in order to ensure they are understood. It has been suggested that the influence of conversational setting on articulation might be due to the novelty of the medium (Blokland and Anderson, 1998). Alternatively, or in addition, speakers may have experienced a sense of social distance from their interlocutor and the communicative situation in a video-mediated conversational setting. Could this have led to the articulatory differences observed in the present study?

In the case of Study 2, there are several problems with the *social distance* explanation (Short, Williams and Christie, 1976). First, on the one hand, comparisons of face-to-face and video-mediated communication have revealed differences in turn-taking (e.g. Sellen, 1995) suggesting that video-mediated dialogues tend to be less interactive than face-to-face dialogues. This is in line with the social distance explanation. However, in the present study, it was found that the duration of specific words was longer in the video-mediated conversational setting than in the face-to-face setting. It was inferred from this that the increased rate of speech observed in the video-mediated setting must have been due to faster articulation rather than more overlapping speech. Furthermore, O'Conaill, Whittaker and Wilbur (1993) showed that differences in turn-taking behaviour can be attenuated where high quality video-mediated technology is used – as was the case in the present study. It seems unlikely then, that communication in the video-mediated setting was less interactive (thereby indicating social distance between participant) than in the face-to-face conversational setting.

A second problem relates to the fact that, in section 2.6.3 of this thesis, it was suggested that the physical distance between participants may have given rise to a sense of social distance between participants. Yet a study conducted by Rutter, Stephenson and Dewey (1981), which explicitly aimed to disambiguate co-presence and visibility, found no indication that a lack of co-presence influenced the content or style of communication. What is interesting about this particular study is the wide variety of measures used. These included; the number and word length of utterances, the number of floor changes, or changes of turn, simultaneous speech, and mutual silence. They found that remote communication was no less spontaneous or personal than co-present interactions. For the reasons stated above, the *social distance* explanation now seems somewhat less plausible.

An alternative explanation rests on the novelty of the technology. This was suggested as a possible explanation of a previous demonstration of the effect of conversational setting on articulation (Blokland and Anderson, 1998). The participants in this study were undergraduates at the University of Glasgow. Only a few reported having used video-mediated technology before. It is highly unlikely that the participants in this study were familiar with video-mediated communication. In addition to this, the findings that over time, users of computer based-technologies can adapt to computer-based technologies (Newlands, Anderson, and Mullin, 2003; Eveland and Dunwoody, 2001) adds credence to the notion that unfamiliarity with the medium invokes some kind of novelty effect. Furthermore, it has been shown that many of the differences observed between face-to-face and video-mediated communication are attenuated with extended use of technology (Walther 1992,1994; Kelly, Futoron and McGrath, 1990; Kelly and McGrath, 1985). Thus, it seems more plausible to speculate that the novelty of the video-mediated technology and the unfamiliarity of communicating via a video-link may have contributed to the slower rate of speech in a video-mediated conversational setting compared with a face-to-face conversational setting.

With respect to cognitive load, the findings of Study 2 suggest that, in terms of the way interlocutors responded to time pressure, video-mediated communication appeared to function in the same way as face-to-face communication. There was no evidence to suggest that the conversational setting influenced the overall characteristics of the dialogue such as the dialogue length, or the number of words (cf. O'Conaill, Whittaker and Wilbur, 1993; Sellen, 1995). Nor did the conversational setting impact on task performance. Previous studies of face-to-face and video-mediated or computer-mediated communication have revealed mixed findings in terms of the effect of video-mediation on broad characteristics of the dialogue. On the one hand, not all comparisons of face-to-face and video-mediated dialogues have revealed differences in terms of dialogue length and task outcome (e.g. Anderson et al., 1996). Yet, on the other hand, several comparisons of face-to-face and computer based communication have observed effects of conversational setting on these aspects of dialogue (Anderson et al., 1997; Kraut, Fussell and Siegel, 2003; Newlands, Anderson and Mullin, 2003). Why should this be the case?

One reason may lie in terms of comparability. For example, the study conducted by Anderson et al., (1996) compared face-to-face with a corresponding audio only setting, then compared video-mediated dialogues with a corresponding audio only setting. Thus, video-mediated and face-to-face dialogues were not directly compared in this case. Similarly, although Newlands, Anderson and Mullin (2003), found that task performance in a computer-based setting was poor compared with task performance in a face-to-face setting, their study was not based on comparisons of spoken dialogue, but rather on text-based communication. The observed differences were also offset with exposure to computer-based communication. This observation was more in line with previous studies of mediated communication which have shown that collaborative tasks are not usually affected by variations in the conversational setting (Williams, 1977; Chapanis, Ochsman, Parrish and Weeks, 1972,1977).

More recently, however, Kraut, Fussell and Siegel (2003) directly compared video-mediated communication with face-to-face communication and reported reliable differences in task performance for a collaborative bicycle repair task. They found that the participants who were co-located, or side-by-side completed the task more quickly and accurately. Yet in the present study there was no significant difference in the task performance or in the time taken to perform the task. There were, however, several fundamental differences between the nature of the video-mediated set up used in the study conducted by Kraut, Fussell and Siegel (2003) and that used in the present study, which may account for the contradictory findings.

In the study conducted by Kraut, Fussell and Siegel (2003) the *repair worker* wore a head-mounted camera with a monitor mounted in front of the right eye. In addition to this, various types of information were displayed on the participants' computer monitors. This information was rather complex and varied in nature. For example, information was displayed to the helper in 4 windows and included a repair manual, information about what the worker was looking at, as well as information about where the worker was pointing. Given such an array, it could have been the case that assimilating a variety of information increased the cognitive load on the interlocutors. This could have accounted for the poor task performance in the video-mediated setting compared with the face-to-face setting. In contrast, in the present study, the presentation of information was much less complex. On their computer monitors, participants viewed only 2 windows. One window displayed the Map Task, while the other window, contained a head and shoulders view of the other participant. Thus there was less information to assimilate. It is less likely that the technology itself produced disorientation for the participants which would have increased cognitive load. This was consistent with the lack of a significant difference, in terms of task performance, between the video-mediated and face-to-face settings.

A second factor of interest relates to the establishment of common ground. It could have been the case that, in the study conducted by Kraut, Fussell and Siegel (2003), participants experienced difficulty in establishing common ground as a result of the complex nature of the technical set-up. For example, Kraut, Fussell and Siegel (2003) point out that queries from the helper about the worker's point of view suggested that participants in the video-mediated setting experienced difficulty establishing which visual information was shared and which was not. As the worker moved around, objects were often out of view of the head-mounted camera. Furthermore, the workers face was not visible to the helper, who would have been unable to monitor facial expressions. In the present study, although not all information on the map was shared between interlocutors, the images of the Map remained static and consequently, points of reference such as the positions of shared landmarks, remained static. Furthermore, both participants could see their interlocutor. It seems then that the participants in the present study would have shared more common ground and have had greater access to facial expressions than in the study conducted by Kraut, Fussell and Siegel (2003). Interestingly, interpersonal communication has been shown to be more efficient when people share greater amounts of common ground, or mutual knowledge, beliefs, goals and attitudes (Clark and Marshall, 1981; Clark and Wilkes-Gibbs, 1986). Factors such as these could account for the apparent contradictory findings between the present study and the study conducted by Kraut, Fussell and Siegel (2003).

In conclusion, the results of this study suggest that, on the whole, video-mediated communication functions in the same way as face-to-face communication, with the exception of the rate at which interlocutors spoke. No differences were observed between face-to-face and video-mediated conversational settings in terms of broad characteristics of the dialogues such as the time taken to perform the task, the number of words used to achieve the task goal, or the outcome of task performance. Cognitive load appeared to influence those aspects of communication that are likely to place demands on working memory, such as a decision regarding a particular strategy to employ. No difference was observed between conversational settings in terms of the degree to which cognitive load influenced specific strategies to cope with time pressure, such as referring to fewer objects in the discourse.

A picture emerges then wherein dialogue can be viewed as multifaceted in nature. On the basis of the present findings, it seems reasonable to suppose that some processes, such as priming, are automatic whereas other processes, such as a decision to talk more slowly, or a strategic decision, may be consciously controlled.

4.4.5. Future Directions

A problem with the present study relates to the fact that more pairs of participants were unfamiliar to each other than were familiar with each other in the video-mediated setting compared with the face-to-face setting. There is some evidence to suggest that in the case of the Map Task, the familiarity of participants can influence aspects of non-verbal communication such as gaze behaviour. For example, Boyle et al., (1994) found that instruction followers gazed more than twice as much at partners who were familiar to them than at partners who were unfamiliar to them. This suggests that conversational partners who are familiar with one another may be more attentive to non-verbal behaviour than conversational partners who are unfamiliar to one another.

Roßnagel (2000) suggests that naming an object involves making an evaluation about how readily one's partner will understand a name. Furthermore, Isaacs and Clark (1987) showed that familiarity with New York city influenced the way speakers referred to city landmarks. In their study pairs of participants, who were either New Yorkers or non New Yorkers, performed a picture naming task which involved arranging postcards of New York in a prescribed order. When the speaker believed that their addressee was familiar with the city they tended to use landmark names, such as *Times Square* and the *United Nations*. Yet when they believed their addressee was not a native New Yorker, they tended to describe the landmarks rather than use landmark names. On the basis of these findings Isaacs and Clark (1987) suggested that conversational partners are able to accommodate to each other by assessing one another's level of expertise based on assumed shared knowledge and on their partner's responses.

In the present study, participants were instructed that their maps were different. In other words, they were explicitly informed that they may have different knowledge to that of their partner. However, it could have been the case that, speakers who were familiar with each other were better at assessing their partner's knowledge. How might this have been reflected in terms of naming landmarks on the map? One possibility is that interlocutors who were familiar with each other may have been more likely to use short referring expressions such as a truncated name or pronoun. If they were more attentive to each other's non-verbal cues as suggested by Boyle et al's (1994) findings, they may have assessed that, on repeating a landmark name, their partner would be able to identify that landmark. In contrast, participants who were unfamiliar with one another may have been more conservative in assessing that their addressee would know which landmark was being referred to. Consequently, they could have been more explicit in their naming of objects by using more full names (as labelled on the map). Alternatively, people who are friends or familiar with each other may be more likely to be considerate of each other's ability to understand. They may want to be seen to care whereas people who are unfamiliar with one another may be more likely to attribute a lack of success to the failings of their partner. If this is the case, then participants who are familiar with each other may be less inclined to shorten the names of landmarks than participants who are familiar with one another. An examination of the data indicated that the later seemed to be the case, but only under time pressure.

Table 4.28 shows the mean proportions of shortened referring expressions (i.e. truncated landmark names or landmarks names substituted by a pronoun) for participants who were familiar with their partner and those who were unfamiliar with their partner in untimed and timed, face-to-face and video-mediated dialogues.

Table 4.28. Mean proportion of shortened referring expressions (with standard deviations) for unfamiliar and familiar participants, in untimed and timed, face-to-face and video-mediated dialogues.

Face-to-face

	Unfamiliar			Familiar	
N	Untimed	Timed	N	Untimed	Timed
11	0.53 (0.29)	0.56 (0.31)	17	0.51 (0.25)	0.39 (0.32)

Video-mediated

	Unfamiliar			Familiar	
N	Untimed	Timed	N	Untimed	Timed
21	0.46 (0.24)	0.62 (0.26)	7	0.56 (0.40)	0.35 (0.33)

As can be seen from table 4.28, the most notable difference between familiar and unfamiliar participants is that the proportion of shortened referring expressions is greater where the participants are *unfamiliar* rather than *familiar* with each other. In the timed conditions, the mean proportion of shortened landmark names was higher for unfamiliar participants than for familiar participants in both the face-to-face (0.56 and 0.39 respectively) and video-mediated conversational settings (0.62 and 0.39 respectively). In the untimed conditions there is little difference between familiar and unfamiliar participants in the face-to-face group. However, the opposite pattern emerges in the video-mediated group. The proportion of shortened landmark names was higher for familiar participants (0.56) than for unfamiliar participants (0.46).

Unfortunately, the present data did not lend itself to complex analysis of variance (ANOVA) to investigate the role of conversational setting, cognitive load and participant familiarity on the shortening of referring expressions. This would have required a three-way ANOVA. Yet on rejecting speakers to maintain the balance of the design there would have been only 4 speakers in the video-mediated cells of the design for familiar participants. This would have been unsatisfactory since a three-way ANOVA requires a minimum of 8 speakers per cell. Nevertheless, a series of unrelated t-tests were carried out on the data in order to explore whether the differences observed in table 4.28 between familiar and unfamiliar participants were reliable. In order to maximise the number of speakers, the data were collapsed across conversational settings. 56 speakers were included in the analyses, of which 32 were unfamiliar with their addressee and 24 were familiar with their addressee. The data for the untimed and timed conditions were analysed separately. The results of the t-test on the data for the timed condition indicated that the effect of familiarity was significant [t(54) = 2.85, p(1-tailed) < 0.01]. When the task was performed under time pressure, the proportion of shortened referring expressions was greater for participants who were unfamiliar with their partner (0.60) than for participants who were familiar with their partner (0.37). The results of the t-test carried out on the data for the untimed condition did not reveal a significant effect of familiarity. Although conducting multiple t-tests on the data is not ideal since this does not control for a family wise error, these findings nevertheless suggest that the familiarity of the participants may influence the way objects in a discourse are named.

The observations described above suggest that where conversational partners are communicating under time constraints, interlocutors who are unfamiliar with one another are more likely to shorten the names of repeated references to objects than participants who are familiar with one another. This would seem to suggest that, under time pressure, participants who are familiar with their addressee become more considerate of their partner and what they are likely to know. Rather than using a pronoun to save time, they assess that their partner may not have the same landmark on a map and so explicitly name the object. In contrast, conversational partners who are unfamiliar with one another may be less likely to make an assessment about what their partner may or may not know and simply use short referring expressions, such as pronouns, to save time. An exploration of questions such as these offers an interesting line of investigation for future work.

Chapter 5

The Influence of Cognitive Load on Collaboration in Video-mediated and Face-to-face Dialogues

5.1 Introduction

The main objective of this dissertation is to explore the influence of conversational setting and cognitive load on speech and language processes in 2-party spoken dialogues. The overall questions to be addressed are: Does the conversational setting, namely a video-mediated setting compared with a face-to-face setting, influence aspects of spoken communication such as reference, articulation and collaboration? Are these aspects of spoken dialogue influenced by increased cognitive load, as illustrated by time pressure? Does an increase in cognitive load have the same effect on spoken output in a video-mediated conversational setting as in a face-to-face setting?

The results of an initial study, which examined word articulation in video-mediated dialogues, were presented in Chapter 3. It was found that repeated mentions of words forming the names of landmarks on a map were articulated more quickly than 1st mentions irrespective of which speaker introduced the object into the discourse. The same pattern of results has been observed in face-to-face dialogues (Bard et al., 2000). This suggested that, in terms of word articulation at least, video-mediated communication appeared to function in the same way as face-to-face communication.

In Chapter 4, the results of a second study, which examined the influence of cognitive load (as illustrated by time pressure) on articulation and reference in video-mediated and face-to-face communication, were presented. With respect to articulation, it was found that repeated mentions of words forming the names of landmarks on a map were articulated more quickly than introductory mentions irrespective of conversational setting or cognitive load. This replicated the overall results of Study 1 and is consistent with the hypothesis that articulatory reduction, that is the shortening of repeated mentions of words, is underpinned by priming processes which are automatic and occur within a very fast planning cycle (Bard et al., 2000). Consequently, they are unaffected by factors such as the demands on a speaker's time and attention, or by the setting in which a dialogue takes place. It was also found that, in a video-mediated conversational setting compared with a face-to-face setting, words were longer in duration overall, and that the rate of speech was slower. This indicated that speakers spoke more slowly in a video-mediated conversational setting than in a face-to-face setting. This finding could not be accounted for in terms of priming processes. Following Lindblom et al. (1995), it was suggested that the slower articulation observed in video-mediated dialogues may reflect a response to the communicative environment. Perhaps the unfamiliar nature of a video-mediated setting leads speakers to discern a possible communication difficulty. Consequently, the speaker adjusts his or her articulatory effort to ensure that s/he is understood. The findings of Study 2 seemed to suggest a distinction, at the level of articulation, between consciously controlled processes and automatic processes.

Although there was no indication that cognitive load influenced articulation, cognitive load did impact on the way speakers referred to objects. It was found that participants referred to fewer landmarks and made fewer references to those landmarks under time pressure. It was suggested that this type of effect may be indicative of a specific strategy appropriate for coping with time pressure. There was some indication that, for certain speakers, the processes involved in naming an object were influenced by increased cognitive load. There was a tendency for some speakers to use more pronouns for repeated references to landmarks under time pressure. The Dual Process Model goes some way to explaining why cognitive load should influence the way participants refer to landmarks but not on the way they articulate words forming referring expressions. According to the Dual Process Model (Bard et al., 2000), automatic processes should not be affected by the cognitive demands on a speaker's time and attention. In contrast consciously controlled processes which make demands on working memory are more likely to be influenced by cognitive load.

The purpose of this chapter is to examine the influence of cognitive load and conversational setting on the way interlocutors collaborate with each other when introducing entities, or objects, into a discourse.

5.1.1 Cognitive Load and Collaboration

Traditionally, dialogue has been viewed as a collaborative process in which interlocutors collaborate with each other and wherein speakers produce utterances that are designed to meet the comprehension needs of the listener. In Chapter 2 this view of dialogue was discussed in the context of the relevant literature. A brief summary is provided here for convenience. According to the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992), interpersonal communication is viewed as a joint activity in which interlocutors construct a model of the dialogue based on their common knowledge and beliefs, or *common* *ground* (Clark and Marshall, 1981). This process is achieved, in part, through the establishment of definite reference, or the building of mental representations of the objects or referents involved in the discourse. The notion of dialogue as an essentially co-ordinated activity has also been expressed more recently by Pickering and Garrod (in press). According to their Interactive Alignment Model of interpersonal communication, interlocutors co-ordinate or *align* their mental representations via automatic alignment channels. Rather than model, or dynamically update every aspect of their interlocutor's mental state, interlocutors align their mental models at different levels of representation. In other words, interlocutors may influence one another through the words, sounds and meanings they use. A related view of interpersonal communication holds that a speaker must consider the perspective of their addressee to ensure the successful transfer of meaning. Schober (1993), for example, holds that this lies at the heart of successful communication. With respect to dialogue, the basic notion that speakers tailor their spoken output to their addressee has been widely accepted (Krauss & Fussell, 1996).

More recently, however, it has been suggested that, in on-line spoken conversation, interlocutors must balance this adjustment to the listener with the need to produce utterances in real time (Branigan and McLean, 2003). Furthermore, certain levels of language production may not rank highly in a speaker's priorities to balance these needs. For example, Branigan and McLean (2003) found that, at a syntactic level, a speakers choice of active or passive structure reflected their own knowledge rather than the knowledge states of their addressees. Similarly, at an articulatory level, Bard et al., (2000) found that while speakers adjusted their articulation in response to gross characteristics of the dialogue context (such as who they were talking to or what they could see), they did not make the same articulatory adjustments in response to more subtle aspects of the dialogue (such as what the listener could be inferred to know or to feedback from the listener). Bard et al., (2000) argued that whether or not the speaker adjusts articulation to the perspective of the listener may be limited by the cognitive demands on the speaker's time and attention. Curiously, while many studies have investigated the issue of adaptation to the listener, few studies have investigated the influence of cognitive demands on adaptation to the listener, or audience design.

Nevertheless, in one such study, which investigated the influence of time pressure on the use of referring expressions, Horton and Keysar (1996) showed that speakers described simple objects moving across a computer screen in the context of shared information when there was no pressure of time. In contrast, when speakers were instructed to initiate utterances quickly, the same objects were described in the context of information that was available only to the speaker. In order to explain these findings, Horton and Keysar (1996) proposed a modular division between the initial formulation of utterance and subsequent monitoring of the utterances produced. According to the model, the initial formulation of utterances is based on a model of the listener's knowledge, or more precisely on common ground. Under the pressure of time, however, a speaker can resort to a model of their own knowledge in the first instance, since utterances can be revised in response to requests for clarification, or to feedback from the listener. An advantage of this model is that it takes into account the need to balance accurately produced utterances that the listener will understand, with the need to produce utterances quickly and efficiently.

Roßnagel (2000) proposed an alternative explanation of these findings. He suggested that information to be incorporated into the speech plan may depend on cognitive availability rather than the need to adjust one's speech to the meet the comprehension needs of the listener. Under conditions of increased cognitive load the speaker may need to devote more attention to his or her own speech and language behaviour. Consequently, he or she may be unable to devote much attention to the task of drawing inferences about what his or her partner can be assumed to know. Cognitive load is defined in terms of the limitations on working memory and is described within the framework of Cognitive Load Theory (Sweller, 1988; Sweller van Merrienboer and Paas,1998) which was outlined in section 4.1.1 of Chapter 4. In short, cognitive load can be thought of as the mental energy required to perform a given task.

Roßnagel (2000) hypothesised that adjustment to the addressee's perspective is a consciously controlled process which is limited by cognitive availability. Consequently, adjustment to the listener and the monitoring it is based upon might be sacrificed under conditions of increased cognitive load. In order to test this hypothesis, Roßnagel (2000) conducted a study in which participants instructed two confederates, an adult and a child, to construct a Fishertechnik© assembly model in either a *low, high* or *dual task* condition of cognitive load. Participants in the low load group were able to see a pre-assembled version of the model while they gave their instructions. In the high load group, the task was more difficult since participants had to recall the assembly of the model from memory. In the dual task condition, participants had the model at hand, but the demands on working memory were increased by instructing participants to carry a memory load of seven digits while giving instructions. The experiment was conducted in two phases. There was a training phase, during which the participants learned technical terms associated with the components parts of the model and the step-by-step procedure for constructing the model. In the second phase, participants performed the model construction task with the child and with the adult addressees.

Roßnagel (2000) found that, under conditions of low cognitive load, participants varied in the way they referred to the component parts of the model and the detailing of the instructions depending on whether they were addressing the child or the adult confederate. When addressing the boy, participants' instructions were richer in detail with a greater number of fully specified steps. Fewer technical terms were used, as well as fewer technical terms with a description - both being replaced by descriptions of the component parts of the model. Under conditions of high cognitive load and the dual task condition, there was no difference in the detailing of instructions or references to component parts when addressing either the child or the adult.

According to Roßnagel (2000), the results of this study suggested that the information to be incorporated into the utterance plan was dictated by how readily that information could be assimilated and retrieved rather than by the addressee's information needs. Only under conditions of low cognitive load did speakers adapt their speech to the perspective of the listener. Roßnagel (2000) argued that this was probably because, in the high load and dual task conditions the demands on working memory were increased since the information held and recalled from memory was greater. Consequently, the conscious processes involved in providing more descriptive information to a younger addressee were sacrificed. Referring expressions for the parts of the model and step-by-step instructions had been learned during the training phase. Thus, the retrieval of these well-learned terms may have become automated. In other words, as the cognitive load increases, the priorities for the speaker alter and the retrieval of information becomes more important than the comprehension needs of the listener. Or, as Roßnagel (2000) puts it, conscious processes are sacrificed but that the pattern of results reflect a greater reliance on automated components of utterance planning such as information retrieval. This suggests that conscious processes may be influenced by increases in cognitive load.

5.1.2 Conversational Setting and Collaboration

In Chapter 2, a review of the literature relating to the effect of conversational setting on collaboration was presented and discussed. To recap, there are several reasons to suspect that interlocutors may be less collaborative in a video-mediated setting compared with a face-to-face setting. First, a speaker may feel more distant from their interlocutor and from the communicative situation in a video-mediated setting than in a face-to-face setting (Short, Williams and Christie, 1976). In support of this notion there is evidence to indicate that interlocutors in a video-mediated conversational setting are less interactive and more formal in terms of turn taking behaviour than interlocutors in a face-to-face setting (e.g. Sellen, 1995; O'Conaill, Whittaker and Wilbur 1993). Turn taking reflects co-ordination between speakers with respect to who should speak and when. Less interactive and spontaneous turn taking behaviour observed in video-mediated communication suggests that interlocutors are less co-ordinated in this conversational setting. In addition to this, it has been shown that, for text-based communication via a video-link, interlocutors were initially less collaborative when communicating with someone they believed to be in a distant city, as opposed to someone located in the same city (Bradner and Mark, 2002). These observations are consistent with the notion that the process of grounding, or establishing mutual knowledge will be influenced by the conversational setting (Clark and Brennan, 1991).

According to the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986) interlocutors will try to minimise the work they do during the grounding process. The process of grounding may require more effort in conversational settings, such as video-mediated communication, where communicative cues such as gestures and facial expressions are more difficult to use. This proposal is supported by the recent observation that interlocutors conducting in a bicycle repair task communicated more efficiently in a face-to-face setting than in an audio only setting. However, the same efficiency gain was not observed in a video-mediated setting compared with the audio only setting (Kraut, Fussell and Siegel, 2003). These findings suggest that interlocutors may be less collaborative in a video-mediated setting than in a face-to-face setting.

This question has been investigated in text-based computer-mediated communication. Newlands, Anderson and Mullin (2003) showed that participants were less collaborative in a computer-based conversational setting than in a face-to-face conversational setting. Dialogues of participants performing a computer-based text version of The Map Task were compared with dialogues drawn from the HCRC Map Task corpus of co-present participants performing The Map Task. It was found that, initially at least, task performance was poor in the computer-based setting compared with the face-to-face setting. Differences were also observed in the structure and content of the dialogues using the technique of

Conversational Games Analysis (Kowtko, Isard and Doherty-Sneddon, 1992). In the computer-based setting, participants initiated *Instruct Moves* more often but used fewer *Check Moves*. An analysis of conversational games revealed further differences between the use of certain types of moves. For example, computer-based dialogues contained fewer Initiating Moves than the spoken dialogues. The main differences between the two contexts were the increased initiating of Instruct Moves, such as giving instructions, and decreased use of Check Moves, such as questions for clarification. Check moves are important in establishing common ground, or mutual understanding. Since this type of move was less frequent in a computer-based conversational setting, this might suggest that participants were less engaged with one another with respect to the grounding process. It is uncertain, however, whether the findings of the Newlands, Anderson and Mullin (2003) study can be generalised to video-mediated communication. In text-based computer-mediated communication the participants were unable to see each other. Furthermore, Newlands, Anderson and Mullin (2003) also showed that participants adapted their communicative processes with repeated exposure to the medium.

5.1.3 Research Questions

The question to be addressed in this chapter is whether conversational setting and cognitive load will influence collaboration in spoken dialogue?

With respect to cognitive load, previous research has examined the effect of increased task difficulty and time pressure on adaptation to the perspective of the listener in face-to-face communication. It has been shown that adaptation to the listener's perspective is offset under conditions of increased cognitive load due to task difficulty (Roßnagel, 2000). Similarly, under time pressure speakers were found to design references in terms of their own knowledge rather than mutual knowledge, at least in the initial stages of utterance planning (Horton and Keysar, 1996). On the basis of these findings it seems reasonable to hypothesise that, where the demands on the speakers' time and attention are minimal, interlocutors will collaborate with each other to establish common ground. However, if the demands on working memory are increased, due to time pressure for instance, then the desire to collaborate will have to be balanced with the need to produce communicatively adequate utterances in real time. Consequently, interlocutors may be less collaborative under time pressure.

With respect to conversational setting, previous research suggests that interlocutors may experience a feeling of social distance in a video-mediated conversational setting compared with a face-to-face setting. Research has shown that the nature of communication is less interactive (Sellen, 1995) and less efficient (Kraut, Fussell and Siegel, 2003) in a video-mediated setting than in a face-to-face setting. Furthermore, it has been shown that, for The Map Task, interlocutors were less collaborative in a text based computer-mediated setting than in a face-to-face setting (Newlands, Anderson and Mullin, 2003). Taken together, these findings suggest that, in a video-mediated conversational setting, interlocutors are less engaged with one another and are therefore likely to be less collaborative than in a face-to-face setting. In addition to this, if the process of establishing common ground is more difficult in a video-mediated setting, then one might hypothesise that interlocutors will exhibit a less collaborative conversational style in a video-mediated conversational setting compared with a face-to-face setting.

One way to examine these questions is to consider the linguistic forms used to introduce objects, such as landmarks on a map, into a dialogue. One reason for examining this aspect of spoken discourse is that the way objects are introduced into a discourse is critical to the establishment of common ground. In the case of The Map Task, this involves establishing which landmarks are shared and which are not. Second, a consideration of the choice of linguistic forms to introduce objects into a discourse takes into account the function of the forms chosen (e.g. whether or not to use a question form to directly question the existence of an object). It seems reasonable to assume that this is likely to be a conscious decision, and thus likely to be a target for cognitive load effects. Recall, that certain levels of speech and language production, such as articulatory reduction, may be automatic (Bard et al., 2000) or with respect to syntactic structure, may not be a high priority for communicative effectiveness. A speaker may not be consciously aware of syntactic subtleties of particular forms (Branigan and McLean, 2003). Third, previous research has been conducted on the use of linguistic forms to introduce landmarks into the dialogue (Anderson and Boyle, 1994). This study showed that question forms were instrumental to communicative effectiveness, or task success. Thus introductions to landmarks involving question forms should be indicative of successful collaboration in establishing common ground.

In order to test the hypothesis that interlocutors would be less collaborative under conditions of increased cognitive load, and in a video-mediated setting compared with a face-to-face setting, the linguistic forms used to introduce landmarks on a map into the dialogue were examined in face-to-face and video-mediated, timed and untimed dialogues.

5.2 Coding of Introductions to Landmarks

The materials used for the analyses presented in this chapter were taken from those collected for Study 2. The collection of the materials was described in section 4.2 of Chapter 4.

According to the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992), a conversation proceeds in a systematic fashion through the accumulation of common ground. This is established through an acceptance process wherein one speaker "presents", or introduces a referent which his or her interlocutor must accept, or reject, as constituting part of the common ground. Both interlocutors must accept responsibility for each definite reference by trying to establish the mutual belief that the reference has been understood before they let the conversation go on. The two basic elements in the acceptance process are (a) a *presentation* and (b) an *acceptance*. The referential coding described below was intended to capture these 2 basic elements.

Accordingly, all introductory presentations of references to landmarks were coded according to whether the landmark was introduced with a question form or a non-question form. Responses to presentations were also coded according to whether the response was informative or not. Thus, the landmarks on the map could be introduced into the dialogue in one of 4 ways, as illustrated in table 5.1. References to landmarks are shown in italics.

Table 5.1. Categories of introductions to landmarks

Type of Introduction	Example
Question + Informative Response	IG: Have you got an *iron bridge?* IF: Yes, I do
Question + Non-Informative Response	IG: So you're heading right and do you have *armoured vehicles?* IF: Oh erm I've gone maybe I've gone too far
Non-question + Informative Response	IG: ......then turn right towards the *ghost town* IF: I don't have a *ghost town*
Non-question + Non-informative Response	IF: er I'm going through a *military base* IG: going through a *military base*? (The Instruction Giver does not have a *military base* and the landmark is not discussed further)

5.2.1 Introductory Presentations

Following Anderson and Boyle (1994), introductory presentations of landmarks were coded according to whether the speaker used a question form or a non-question form.

Question Presentations

Introductions to landmarks on the map which indicated that the speaker was questioning the listener's awareness or knowledge of the landmark being introduced were categorised as *questions*. These were interrogative linguistic forms which constituted a first reference to a landmark. Interrogative forms included utterances which could be distinguished in form from a declarative or imperative, even if part of the initial auxiliary or subject had been omitted. Examples of question forms are illustrated in (1) to (7) below. References to landmarks are shown in italics.

1. Right you got a *palm tree*?
2. Do you see where the *waterfall* is?
3. See the *iron bridge* up there?

4. Do you have a *beach hut*?

5. Have you got *sunflower fields*?

6. I've got an *overnight accommodation.* Do you have that?

7. I don't know if you've got this *balloon rally* on your map?

Names of landmarks marked by rising intonation as in example (8) were also included in the question category.

→ ↗

8. *camp site*?

In the present study, it was noted that sometimes landmarks were introduced by declarative statements but then, following a brief response by the listener, the Introducer followed up with a question. This type of exchange was included in the question category. An example is shown in (9) below.

9.

IF:	…….as far down as the *ship*?
IG	erm
IF	Or do you not have a *ship* which is actually in the *black sea* where the sand is just on top of the sand?
IG	Yeah but it's a lot further west is it?

Some introductory references were in the interrogative form but the information being sought was not about the listener's knowledge of the newly introduced feature. This type of reference was excluded from the question category. An example is shown in (10) below

10. Is your start just west of the *chemical weapons plant*?

In (10), the "chemical weapons plant" is introduced in a question about the location of the start and so would not be included in the question category.

Non-question presentations

Anderson and Boyle (1994) classified introductory references to landmarks which were part of the instruction-giving process rather than enquiries about the listener's knowledge in the non-question category. These included declarative statements as in (11) to (14),

11. I've got a *palm tree.*
12. You should have a *waterfall* there.
13. There's a *coach park* about halfway down the page.
14. Just straight along the top and then west to underneath the *black gold.*

and imperatives as in (15) and (16),

15. Go round to where there's a *waterfall.*
16. Go right underneath the *winter bay.*

References to the finish were excluded from the coding and analysis since Anderson and Boyle (1994) observed that speakers appeared to treat the finishing point as a generic reference to a type of place on a route rather than a landmark in itself.

5.2.2 Responses to Presentations

In Anderson and Boyle's (1994) study, only responses involving references to unshared landmarks were coded. In the present study, however, references to shared and unshared landmarks were coded.

Difficulty was encountered when coding responses to presentations of landmarks. According to the collaborative model, when an object is presented it can be accepted by either continuing on to the next contribution or by asserting an acceptance with forms such as; *yes, right, I see.* However, in the case of the Map Task, it was noted that utterances such as *yes right I see* could be ambiguous with respect to whether or not they constituted an informative response. One problem arose with respect to forms such as *uh hum,* which seem to take on different meanings depending on the context. This is illustrated in examples (17) and (18). The names of landmarks are shown in italics.

17.

IG	Have you got the *precious stones* yeah?
IF	Uh hum

18.

IG	OK head for the *precious stones*
IF	Uh hum

In (17), *uh hum* is uttered in the context of a question form and seems to function as an acceptance of the existence of the landmark *precious stones.* Thus the response given by the Instruction Follower (IF) is informative in this case. In (18) it is not clear whether the utterance *uh hum* acknowledges the instruction to "head for the precious stones" or functions as an acceptance of the landmark itself. In other words, it is not clear whether the response given by the Instruction

Follower (IF) is informative with respect to the existence of the *precious stones* or not.

A second problem encountered relates to the proposition that participants can build common ground by allowing the conversation to go on (Clark and Wilkes-Gibbs, 1986). The implication is that they implicitly accept that an object is part of the common ground. However, in the Map Task there are cases where the Instruction Follower, for instance, will allow the Instruction Giver to continue the conversation even if the Instruction Follower does not have a landmark on their map. This problem has been noted previously in the literature (e.g. Bard et al., 2000). An example is illustrated in (19) below.

19.

IG	And once we're past that landmark we're going straight down and then turning right just above the *camp site*
IF	Right

In (19) the Instruction Follower responds by saying "right". According to the Collaborative Model (Clark and Wilkes-Gibbs, 1986), this should constitute an instance of an acceptance of the object *camp site.* However, the Instruction Follower does not have a camp site on their map. For these reasons only those forms which constituted an explicit acceptance or denial of a landmark such as *Yes, I've got that* or *I don't have that* were categorised as Informative Responses. This corresponds to the category of responses termed "Fully Informative" by Anderson and Boyle (1994). It should be noted that, in their study Anderson and Boyle (1994) coded responses to introductions to unshared landmarks only. In the present study however, the method for coding the responses to presentations differs. Responses to introductory presentations of landmarks were categorised as either *Informative* or *Non-informative.*

Informative Responses

A response was categorised as an Informative Response (IR) if the responder explicitly indicated that they either did, or did not have the relevant landmark on their map. This could be done by responding positively or negatively to an introductory presentation, as in (20) and (21) below;

20.

IG	Do you have a *chemical weapons plant*?	
IF	I don't, no.	IR

21.

IG	Have you got an *iron bridge*?	
IF	Yes, I do.	IR

A response could be informative irrespective of whether the initial presentation was a Question or a Non-Question.

Non-informative Responses

In their study, Anderson and Boyle (1994) further categorised Non-informative Responses into *Inadequately Informative Responses* and *Uninformative Responses.* The term Inadequately Informative referred to those responses which were relatively informative even though they did not contain an explicit acceptance or denial of a particular landmark. Responses where no relevant feedback or inappropriate feedback was given were categorised as Uninformative Responses. In the present study an initial attempt was made to categorise the responses in this way. However, when a subset of dialogues were presented to a 2nd coder, this lead to a conflict in coding. It was felt that only Fully Informative Responses offered a clear example of collaboration between participants since this type of response consisted of an explicit acceptance or denial of the existence of a landmark. In the present study the question of interest was whether participants would be less co-operative under time pressure and in a video-mediated setting. Consequently, Fully Informative Responses were of primary interest and so all other responses were simply classified as Non-informative.

Non-informative responses included cases where the responder questioned a landmark or indicated a problem in interpreting a presentation, as in (22) and (23) below.

22.

IF	OK I've got *overnight accommodation*	
IG	Is that say about a third down the page on the far right hand side?	NIR
IF	Errr maybe about a third down the page	

23.

IF	So you're heading right towards the right and do you have *armoured vehicles?*	
IG	erm oh I've gone maybe I've gone too far.	NIR

Responses such as *right, yeah, uh hum* were classified as non-informative if interpretation was ambiguous, as in (24)

24.

IG	Keep coming straight across past the *broken gate*	
IF	Right	NIR

Cases where a landmark was not referred to again once presented were categorised as non-informative as in (25)

25.

IG	Have you got a *ship* in the *black sea*?	
IF	Oh oh wait a minute erm	NIR
IG	Cool. Sorry	
	(*ship* not mentioned again)	

Inaccurate or inappropriate responses were also categorised as non-informative. In (26) for example, the "military air base" and the "ghost town" are not the same landmark but are different features in different locations on the map.

26.

IF	Between the *iron bridge* and the *black sea* there's a *military air base*	
IG	Right, I've got that as a *ghost town*	NIR
IF	Right. OK	

5.3 Data Analysis

In this section the procedure used to organise the data and test the hypothesis that interlocutors would be less collaborative under time pressure (than with no pressure of time) and in a video-mediated conversational setting (compared with a face-to-face setting) is outlined. In order to test this hypothesis, the influence of conversational setting and cognitive load on the use of linguistic forms to introduce landmarks into the dialogue was examined.

The data for analysis were drawn from 64 dialogues of pairs of participants performing a collaborative, problem-solving task, The Map Task, in varying conditions of cognitive load, namely, with and without time pressure. One group of participants had performed the task in a face-to-face conversational setting and the other group in a video-mediated conversational setting. The procedure for collecting the materials was described in section 4.2 of Chapter 4.

The Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992) suggests that where interlocutors are engaged in a joint activity, this should lead to a more co-ordinated understanding of those referents which constitute common ground and those which do not. Consequently, one would expect that the more engaged interlocutors are with each other, the better their co-ordination and hence, the better their task performance. This should be the case in the Map Task, where accurate reproduction of the map route depends, in part, on knowing which objects, or landmarks, form part of the common ground. Given the importance of knowing whether or not the Instruction Follower, in particular, does or does not possess a landmark, the degree to which interlocutors collaborate should be reflected in their use of linguistic forms. This notion is supported by the findings of the study conducted by Anderson and Boyle (1994) which showed that, in the Map Task, the use of Question forms to introduce entities was instrumental to successful task outcome. They found a negative correlation between the use of Question-form introductions and the degree to which the route drawn by the Instruction Follower deviated from the Instruction Giver's route. In other words, greater use of Question-form Introductions led to better task performance. Thus it was expected that introductions to landmarks involving Question forms, would correlate reliably with task performance. Furthermore, on the basis of previous research (e.g. Roßnagel, 2000) it was expected that cognitive load would reduce the extent to which interlocutors collaborate to achieve a specific communicative goal. Similarly, there was evidence (e.g. Kraut, Fussell and Siegel, 2003) to suggest that Interlocutors would be less collaborative in a video-mediated conversational setting compared with a face-to-face setting. In section 5.3.1 the types of introductions to landmarks will be presented and their association with task performance will be tested. Section 5.3.2 focuses on the use of Question-form Introductions and examines the influence of cognitive load and conversational setting on their use. In section 5.3.3 the hypothesis that cognitive load and conversational setting will influence the degree to which interlocutors collaborate with each other will be tested. Section 5.3.4 focuses on the use of Informative Responses and examines the influence of cognitive load and conversational setting on the likelihood of Informative Responses elicited by Question Presentations.

5.3.1 Types of Introductions to Landmarks

Landmarks on the Map could be introduced into the dialogue in one of the four ways illustrated in (a) to (d) below;

a. Question Presentation followed by Informative Response (Q & IR)
b. Question Presentation followed by Non-informative Response (Q & NIR)
c. Non-question Presentation followed by Informative Response (NQ & IR)
d. Non-question Presentation followed by Non-informative Response (NQ & NIR)

The mean proportions (by dialogue), expressed as a ratio of 1, for each type of introduction in face-to-face and video-mediated, untimed and timed dialogues are shown in Table 5.2 below.

Table 5.2 Mean proportions of introductions to landmarks for face-to-face (n = 32) and video-mediated (n = 32) dialogues in untimed and timed conditions

Type of Introduction	***Face to-face***		***Video-mediated***		
	Untimed	Timed	Untimed	Timed	**Overall**
Q & IR	0.51	0.44	0.38	0.25	**0.39**
Q & NIR	0.06	0.06	0.02	0.04	**0.05**
Total Questions	**0.57**	**0.50**	**0.40**	**0.29**	**0.44**
NQ & IR	0.15	0.16	0.25	0.26	**0.20**
NQ & NIR	0.27	0.33	0.34	0.45	**0.35**
Total Non-Questions	**0.42**	**0.49**	**0.59**	**0.71**	**0.55**

As illustrated in table 5.2, the mean proportion of introductions to landmarks involving Questions was lower overall (0.44) than the proportion of introductions involving Non-question forms (0.55). With respect to the use of Question forms, the overall proportion of Question Presentations followed by an Informative Response was much higher (mean proportion Q & IR = 0.39) than Question Presentations followed by a Non-Informative Response (mean proportion Q & NIR = 0.05). In contrast, for introductions to landmarks involving Non-Questions, the overall proportion followed by an Informative Response was lower (0.20) than those followed by a Non-Informative Response (0.35).

It is interesting to note that the proportion of Question-form Introductions was lower in the video-mediated (VM) setting (VM untimed = 0.40, VM timed = 0.29) than in the face-to-face (FTF) setting (FTF untimed = 0.57, FTF timed = 0.50). Table 5.2 also indicates that the mean proportion of Questions was lower in the timed conditions than in the untimed conditions. Obviously (since they are related) the converse pattern was observed for the use of Non-question forms. The mean proportion of Non-question forms was higher in the video-mediated conversational setting (VM untimed = 0.59, VM timed = 0.71) than in the face-to-face setting (FTF untimed = 0.42, FTF timed = 0.49). Furthermore, the proportion of Non-question introductions was higher in the timed conditions than in the untimed conditions.

The data presented in table 5.2 suggested that overall, fewer Question forms and more Non-question forms were used under time pressure than without the pressure of time, and in a video-mediated conversational setting compared with a face-to-face setting. Interestingly, as noted by Anderson and Boyle (1994), Question forms were almost always followed by an Informative Response. In the present study 89% of Question forms were followed by an Informative Response. This suggests that Question form presentations are highly effective in eliciting Informative Responses.

However, in order to better understand how listeners responded to the presentation of new information, the data were further examined. The probabilities of an Informative Response having been elicited by a Question or Non-question Presentation in face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.3 below.

Table 5.3. Probability of Informative Responses elicited by Question and Non-question presentations of landmarks

Presentation Type	***Face-to-face***		***Video-mediated***	
	Untimed	Timed	Untimed	Timed
Question	0.76(0.51)	0.67 (0.44)	0.59 (0.38)	0.45 (0.25)
Non-Question	0.24 (0.15)	0.33 (0.16)	0.41 (0.25)	0.55 (0.26)
(Total IR)	(0.66)	(0.60)	(0.62)	(0.52)

A measure of the probability of an Informative Response elicited by a Question was obtained by calculating the probability of a Q&IR Introduction, given an Informative Response (i.e. PQ & IR/IR). Similarly the probability of an Informative Response elicited by a Non-question was calculated as PNQ & IR/IR, or 1 – PQ&IR/IR. The corresponding mean proportions of Informative Reponses are shown in brackets for convenience.

As can be seen from table 5.3, the probability of an Informative Response having been elicited by a Question Presentation was greater than the probability of an Informative Response having been elicited by a Non-question for the face-to-face untimed (PQ & IR/IR = 0.77 cf. PNQ & IR/IR = 0.22) and the timed (PQ & IR/IR = 0.74 cf. PNQ & IR/IR = 0.26) conditions. In the video-mediated untimed condition, the likelihood of a Question Presentation over a Non-question presentation was somewhat reduced (PQ & IR/IR = 0.59 cf. PNQ & IR/IR = 0.41). Most interestingly, in the video-mediated timed condition, an Informative Response was less likely to have been elicited by a Question Presentation (PQ & IR/IR = 0.45) than by a Non-question Presentation (PNQ & IR/IR = 0.55).

For the purpose of completeness, the probabilities of Non-informative Responses elicited by Question and Non-question Presentations in face-to-face and video-mediated, untimed and timed conditions are shown in table 5.4 below.

Table 5.4. Probability of Non-Informative Responses elicited by Question and Non-question Presentations of landmarks

Presentation Type	***Face-to-face***		***Video-mediated***	
	Untimed	Timed	Untimed	Timed
Question	0.18 (0.06)	0.15 (0.06)	0.08 (0.03)	0.08 (0.04)
Non-Question	0.82 (0.27)	0.85 (0.33)	0.92 (0.34)	0.92 (0.45)
(Total NIR)	(0.33)	(0.39)	(0.37)	(0.49)

The corresponding mean proportions of Informative Reponses are shown in brackets for convenience. As can be seen from table 5.4, Non-informative Responses were more likely to have been elicited by a Non-question than by a Question Presentation. This was the case in the face-to-face setting for both the untimed condition (PNQ & NIR/NIR = 0.82 cf. PQ &NIR/NIR = 0.18) and the timed condition (PNQ & NIR/NIR = 0.85 cf. PQ & NIR/NIR = 0.15). The likelihood of a Non-informative Response elicited by a Non-question over a question was slightly higher in the video-mediated untimed and timed conditions (PNQ & NIR/NIR = 0.92 cf. PQ & NIR/NIR = 0.08).

The data presented in tables 5.3 and 5.4 indicates that, most notably, the likelihood of an Informative Response elicited by a Question Presentation is reduced under time pressure (compared with no pressure of time) and in a video-mediated conversational setting (compared with a face-to-face setting).

In order to assess the effectiveness of the use of introductory references to landmarks on communicative success, the proportion of each type of introduction (e.g. Q & IR) was correlated with the route deviation scores (described in Section 4.3 of Chapter 4) which measured task performance. It was expected that introductions to landmarks involving questions would lead to better task performance. The results of Pearson's correlation are shown in table 5.5 below.

Table 5.5. Types of Introductions to landmarks correlated with task performance (as measured by route deviated in cm2)

Type of Introduction to Landmark	N	Pearson R	Sig (2-tailed).
Q & IR	64	- 0.401	0.001
Q & NIR	64	-0.039	n.s.
NQ & IR	64	-0.086	n.s.
NQ & NIR	64	0.566	<0.001

As can be seen from table 5.5, there was a significant negative correlation between Question and Informative Responses (Q & IR) and route deviation scores. As the proportion of Q & IRs decreases, deviation from the original route on the map increases. In other words, greater use of Q & IRs was associated with better task performance. There was no significant correlation between Question & Non-Informative Responses (Q & NIR) and route deviation scores. It should be noted, however, that the proportion of Q & NIRs was very low across all conditions (<= 0.06). Nor was there a correlation between route deviation and Non-question and Informative Responses (NQ &IR). However, there was a highly significant positive correlation between Non-Question and Non-informative Responses (NQ & NIR). Greater use of NQ & IR introductions was associated with poor task performance.

The correlations presented in table 5.5 suggested that the use of Q & IRs was instrumental to good task performance. It can be inferred from this that introductions to landmarks involving a Question Presentation followed by an Informative Response was indicative of a collaborative conversational style. According to the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992), the building of common ground is brought about, in part, through the establishment of definite reference. In order to reproduce the map in the Map Task, it is necessary to accurately establish common ground, namely which landmarks are shared and which landmarks are different. Thus accurate task performance should reflect effective establishment of common ground. The fact that there was a negative correlation between the use of Q&IR responses and task performance seems to suggest that this type of response indicates that the interlocutors are engaged with each other and are collaborating to build common ground. In contrast, introductions to landmarks involving Non-questions followed by a Non-Informative Response seemed to have a detrimental effect on task performance. This would suggest that the use of this type of introduction is indicative of a breakdown in the process of building common ground.

To summarise this section, landmarks on the map could be introduced into the discourse by using Question or Non-question Presentations followed by an Informative or Non-informative Response. An analysis of descriptive statistics suggested that time pressure lead to a reduction in the proportion of Question-form Introductions. Similarly, there were fewer introductions involving Question forms in the video-mediated conversational setting than in the face-to-face setting. Conversely, there were more introductions to landmarks involving Non-question Presentations in a video-mediated setting (than in a face-to-face setting), and under time pressure (compared with no pressure of time).

When responses to Presentations were considered, it was observed that Question-form Presentations were nearly always followed by an Informative Response. Nevertheless, it was observed that the likelihood of an Informative Response elicited by a Question (as opposed to a Non-question) was somewhat reduced or offset in a video-mediated conversational setting (cf. a face-to-face setting) and under time pressure (cf. no pressure of time).

In their study, Anderson and Boyle (1994) found a significant negative correlation (r = -0.506) between the number of Question forms used by the dominant speaker (the Instruction Giver) and the degree to which the Instruction Follower deviated from the original route on the map. In other words, the greater the number of Question forms used to introduce landmarks, the better the task performance. In Anderson and Boyle's (1994) study, only introductory forms by the Instruction Giver were considered. In the present study, however, I am primarily interested in the contribution of both speakers to the discourse. Thus the use of introductory Question-forms by the Instruction Giver and the Instruction Follower was correlated with the route deviation scores reported in section 4.3.1 of Chapter 4. Taking into account introductory forms by both speakers (Instruction Followers and Instruction Givers), there was a significant negative correlation between the ratio of question forms used to introduce landmarks and route deviation scores (r = – 0.391). Not surprisingly, the correlation was lower than that reported by Anderson and Boyle (1994) since observation of the dialogue transcriptions suggests that Instruction Followers rarely use question forms to introduce a landmark. It should also be noted that in Anderson and Boyle's (1994) study the raw number of questions was used as the dependent variable. In the present study, however, the proportion of Question forms was used as the dependent variable. This helps to explain the difference in the degree of correlation observed by Anderson and Boyle (1994) and that observed in the present study.

5.3.2 Question-form Introductions

In order to test the hypothesis that interlocutors would be less collaborative under time pressure (cf. no pressure of time) and in a video-mediated setting (cf. a face-to-face setting), the use of linguistic forms that were thought to be indicative of a collaborative conversational style were examined in face-to-face and video-mediated dialogues in untimed and timed conditions. As an initial step, the influence of cognitive load and conversational setting on the use of Question forms was examined. Anderson and Boyle (1994) had shown that Question forms were instrumental to communicative success. Furthermore, the observations made in section 5.3.1 indicated that the proportion of introductions to landmarks involving Question forms would be lower under time pressure compared with no pressure of time and lower in a video-mediated conversational setting compared with a face-to-face conversational setting.

The data for the Question forms were organised by dialogue, by item and also by speaker. The data organised by dialogue are presented first. Before proceeding with the analysis, it was necessary to check whether the distributions associated with the dependent variable were normal and whether the error variance associated with the dependent variable was equal across groups. The normality of distribution scores was determined across the untimed and timed conditions for both the face-to-face and video-mediated settings. Kolmogorov-Smirnov tests were carried out on the data. The results of these tests are presented in table 5.6 below. The Kolmogorov-Smirnov Z statistic is provided along with the significance of the p value.

Table 5.6. Kolmogorov-Smirov tests for the proportion of Question-form Introductions per dialogue for face-to-face and video-mediated conversational settings in untimed and timed conditions.

		Cognitive Load			
		Un-timed		*Timed*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
Face-to-face	32	0.76	ns.	0.91	ns.
Video-mediated	32	0.66	ns.	0.66	ns.

As can be seen from table 5.6, the results of the Kolmogorov-Smirnov tests indicated that the distributions of the test samples for the proportion of Question forms were normal. The test samples were also tested for homogeneity of variance. The results of Levene's test of equality of variance indicated that the error variance associated with the dependent variables was equal across conditions [F(3,60) = 0.67; p = 0.57].

The mean proportions of Question-form Introductions to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in figure 5.1 below.

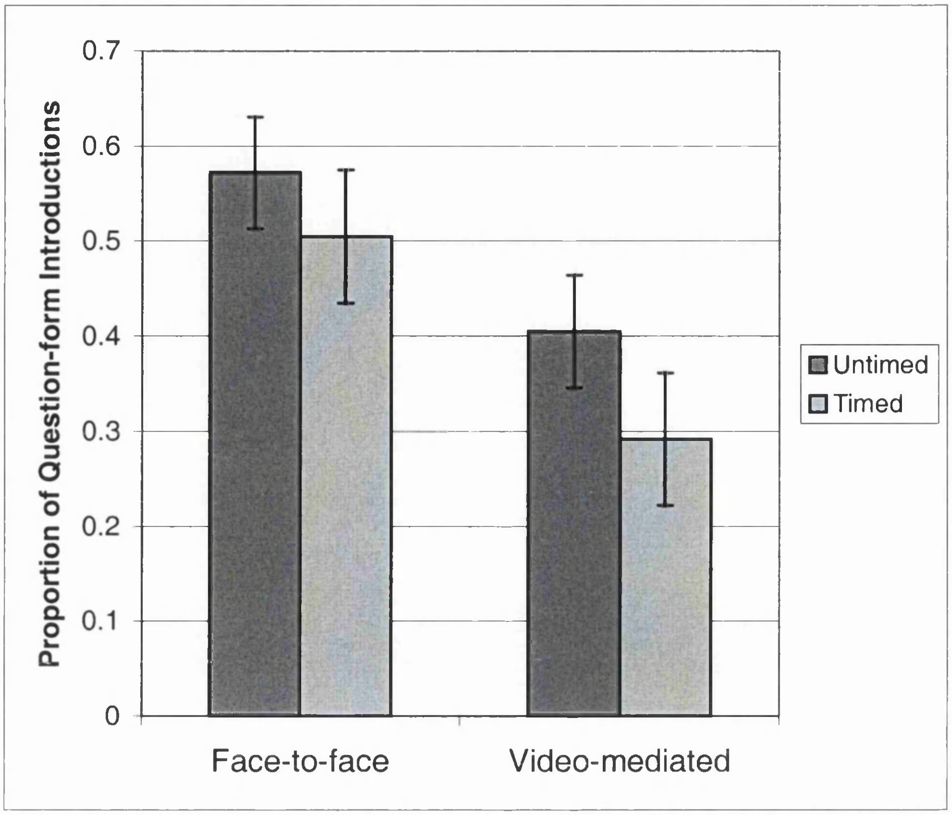


Figure 5.1. Mean proportion of Question-form Introductions (± SE) to landmarks for face-to-face (n = 32) and video-mediated (n = 32) conversational settings in untimed and timed conditions.

Figure 5.1 illustrates that, as expected, the proportion of Question forms used to introduce landmarks into the dialogue was lower in the timed condition than in the untimed condition for both the face-to-face setting (0.50 cf. 0.57) and for the video-mediated setting (0.29 cf. 0.40). Figure 5.1 also illustrates that the proportion of Question forms for the untimed and timed conditions was lower in the video-mediated conversational setting than in the face-to-face conversational setting. The results of the analysis of variance indicated that, overall, the differences observed between the untimed and timed conditions and between video-mediated and face-to-face conversational settings were reliable.

A 2 x (2) analysis of variance by-dialogue was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the proportion of Question forms (expressed as a ratio of 1) used to introduce landmarks into the dialogue. The summary table is presented in Table 5.7 below.

Table 5.7. Analysis of variance (by dialogue) for the use of Question-form Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	5.52	0.02*
S within-group error	30	(0.10)	
Within subjects			
Cognitive Load (L)	1	4.33	0.05*
S x L	1	0.29	0.60
S x S within-group error	30	(0.03)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of conversational setting [F(1,30) = 5.52, p = 0.02]. Overall, the proportion of Question-form Introductions was significantly lower in the video-mediated conversational setting (0.35) than in the face-to-face conversational setting (0.54). There was also a main effect of cognitive load [F(1,30) = 4.33, p = 0.05]. Overall, the proportion of Question form Introductions was significantly lower in the timed condition (0.40) than in the untimed condition (0.49). There was no interaction (F < 1).

Post hoc analyses were also carried out on the data. Newman-Keuls pairwise comparisons indicated that the effect of conversational setting did not reach significance in either the untimed or timed conditions. Nor did the effect of cognitive load reach significance in either the face-to-face or video-mediated conversational settings.

The data were also analysed by-speaker. One speaker in the face-to-face setting and 2 speakers in the video-mediated setting did not introduce any landmarks into the dialogue in at least one condition of cognitive load. Consequently, these speakers were not included in the analysis. In order to maintain the balance of the experimental design (which was counterbalanced for order of cognitive load condition and for version of map used) 3 speakers from the face-to-face group and 2 speakers from the video-mediated group were rejected at random and excluded from the analysis. Thus, 56 speakers (28 in each group of conversational setting) were included in the by-speaker analysis. The mean proportions of Question-form Introductions with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.8 below.

Table 5.8. Mean proportion of Question-form Introductions (by-speaker) for face-to-face and video-mediated conversational settings in untimed and timed conditions.

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	28	0.50	0.34	0.43	0.40
Video-mediated	28	0.40	0.37	0.25	0.34

Table 5.8 shows that, in common with the by-dialogue analysis, the proportion of Question-form Introductions was lower in the timed condition than in the untimed condition for the face-to-face setting (0.43 cf. 0.50) and for the video-mediated setting (0.25 cf. 0.40). Table 5.8 also indicates that the proportion of Question-form Introductions for the untimed and timed conditions was lower in the video-mediated conversational setting than in the face-to-face

conversational setting. The results of the analysis of variance indicated that the overall differences observed between the proportions of Question-form Introductions in the timed condition compared with the untimed condition, and in the video-mediated conversational setting compared with the face-to-face setting were reliable.

A 2 x (2) analysis of variance by-speaker was carried out on the data with conversational setting as the between-subjects factor and cognitive load as the within-subjects factor. The dependent variable was the mean proportion of Question-form Introductions for each speaker. The summary table is shown in table 5.9 below.

Table 5.9. Analysis of variance (by-speaker) for the use of Question-form Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	2.91	0.09
S within-group error	54	(0.18)	
Within subjects			
Cognitive Load (L)	1	4.34	0.04*
S x L	1	0.49	0.49
S x S within-group error	54	(0.08)	

Note. Values enclosed in parentheses represent mean square errors

There was a significant main effect of cognitive load [F(1,54) = 4.34, p = 0.04]. Overall, the mean proportion of Question-form Introductions was lower in the timed condition (0.34) than in the untimed condition (0.45) There was a marginal effect of conversational setting [F(1,54) = 2.91, p = 0.09]. Overall, the mean proportion of Question-forms was lower in the video-mediated setting

(0.32) than in the face-to-face setting (0.46). There was no interaction (F < 1) between conversational setting and cognitive load.

Post hoc analyses were also carried out on the data. Newman-Keuls pairwise comparisons indicated that the effect of conversational setting did not reach significance in either the untimed or timed conditions. Nor did the effect of cognitive load reach significance in either the face-to-face or video-mediated conversational settings.

The data were also analysed by-item whereby the mean proportion of Question forms was calculated for each landmark. The mean proportion of Question-form Introductions with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.10 below.

Table 5.10. Mean proportion of Question-form Introductions (by-item) for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	29	0.56	0.24	0.47	0.29
Video-mediated	29	0.38	0.29	0.28	0.21

Table 5.10 shows that the mean proportion of Question-form Introductions was lower in the timed condition than in the untimed condition for the face-to-face setting (0.47 cf. 0.56) and for the video-mediated setting (0.28 cf. 0.38). Furthermore, as can be seen from table 5.10, the mean proportion of Question-form Introductions for the untimed and timed conditions was lower in the video-mediated conversational setting than in the face-to-face conversational

setting. The results of the analysis of variance indicated that the overall differences observed in the proportion of Question-form Introductions were reliable.

A 2 x (2) analysis of variance by-item was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The proportion of Question form Introductions for each landmark was the dependent variable. The summary table is presented in table 5.11 below.

Table 5.11. Analysis of variance (by-item) for the use of Question-form Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	10.28	<0.01*
S within-group error	56	(0.09)	
Within subjects			
Cognitive Load (L)	1	6.49	0.01*
S x L	1	<0.01	0.95
S x S within-group error	56	(0.04)	

Note. Values enclosed in parentheses represent mean square errors

There was a significant main effect of conversational setting [F(1,56) = 10.28, p< 0.01]. Overall, the mean proportion of Question-form Introductions was lower in the video-mediated setting (0.33) than in the face-to-face setting (0.52). There was also a significant main effect of cognitive load [F(1,56) = 6.49, p = 0.01]. Overall, the mean proportion of Question-form Introductions was lower in the timed condition (0.38) than in the untimed condition (0.47). There was no interaction (F < 1) between conversational setting and cognitive load.

Post hoc analyses were also carried out on the data. Newman-Keuls pair-wise comparisons indicated that the effect of conversational setting was reliable in both the untimed condition ($p < 0.05$) and the timed condition ($p < 0.05$). The effect of cognitive load, however, did not reach significance in either the face-to-face or video-mediated setting.

Overall, the results of the by-dialogue analysis were replicated by the results of the by-items analysis and partially replicated by the results of the by-speaker analysis (where the effect of conversational setting was marginally significant). In the case of the by-speaker analysis, the marginal effect could be due to the fact that the speakers included in the analysis performed the task in different roles. Anderson and Boyle (1994) found that Instruction Followers were less likely to introduce a landmark using a question form than Instruction Givers. This is one possible reason why the effect of conversational setting was marginal in the by-speaker analysis.

In the case of the Map Task there are different categories of landmark. For example, some of the landmarks are shared in that they appear on both the Instruction Follower's map and on the Instruction Giver's Map. Other landmarks appear on only one map and several landmarks are shared but are named differently. Yet the result of the by-item analysis indicated robust effects of conversational setting and cognitive load on the use of Question forms which would seem to suggest that the effect is not limited to a particular category of landmark.

Overall, the results of the analyses of the use of Question forms to introduce landmarks into the dialogue supported the hypothesis that the proportion of Question-form Introductions would be lower under time pressure than without the pressure of time, and in a video-mediated conversational setting compared with a face-to-face setting. This finding is consistent with the notion that speakers were less collaborative in the way new entities are presented into a discourse under these circumstances.

One problem with the analyses of Question-form Introductions is that the response to the Question-form is not taken into account. According to the Collaborative Model (Clark and Wilkes-Gibbs, 1986), there are two stages to the establishment of definite reference. The first stage consists of a Presentation of the new entity and then the second stage of the process involves an acceptance or rejection of that Presentation. Although the use of Question-form Presentations has been shown to correlate with task performance (Anderson and Boyle, 1994), the findings of the present study suggest that, more specifically, it is Question Presentations followed by an Informative Response (Q & IR) which are instrumental to successful task performance. In section 5.3.1 it was found that while Q & IR Introductions correlated with task performance, Question Presentation followed by Non-informative Responses did not. This would seem to suggest that it is the combined effort of both interlocutors, through explicitly eliciting information and explicitly responding informatively to that information which is instrumental to task success.

Thus, an examination of Question Presentations followed by an Informative Response should provide a better test of whether interlocutors were less collaborative under time pressure than without the pressure of time and in a video-mediated conversational setting compared with a face-to-face setting.

5.3.3 Question (and Non-question) Presentations Followed by Informative Responses

In order to further test the hypothesis that time pressure and a video-mediated conversational setting would lead interlocutors to be less collaborative, the influence of cognitive load and conversational setting on the grounding process was examined. Introductions to landmarks involving Question Presentations and Informative Responses (Q & IR) were examined in untimed and timed, video-mediated and face-to-face conversational settings.

Before proceeding with the analyses, Kolmogorov-Smirnov tests were carried out on the data in order to determine the whether the dependent variable was normally distributed across the untimed and timed conditions for both the video-mediated and the face-to-face settings. The results of the Kolmogorov-Smirnov tests are presented in table 5.12 below. The Kolmogorov-Smirnov Z statistic is provided along with the significance of the p value.

Table 5.12. Kolmogorov-Smirov tests for the proportions of Q & IR Introductions to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Un-timed		*Timed*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
Face-to-face	32	0.23	*0.02	0.14	ns.
Video-mediated	32	0.12	ns.	0.16	ns.

As can be seen from table 5.12, the dependent variable for the face-to-face untimed test sample was not normally distributed. The test samples were also tested for equality of variance. The results of Levene's test of equality of variance indicated that the error variance associated with the dependent variable was equal across conditions [F(3,60) = 0.79; p = 0.50]. Given that the data met with the homogeneity of variance assumption underlying analysis of variance and the number of dialogues in each cell of the experimental design were equal, the data were subjected to analysis of variance. According to Roberts and Russo (1999) deviation from the normality of distribution was not likely to result in Type I or Type II errors under these circumstances.

The data were first organised by-dialogue. The mean proportions of Q & IR Introductions for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in figure 5.2.

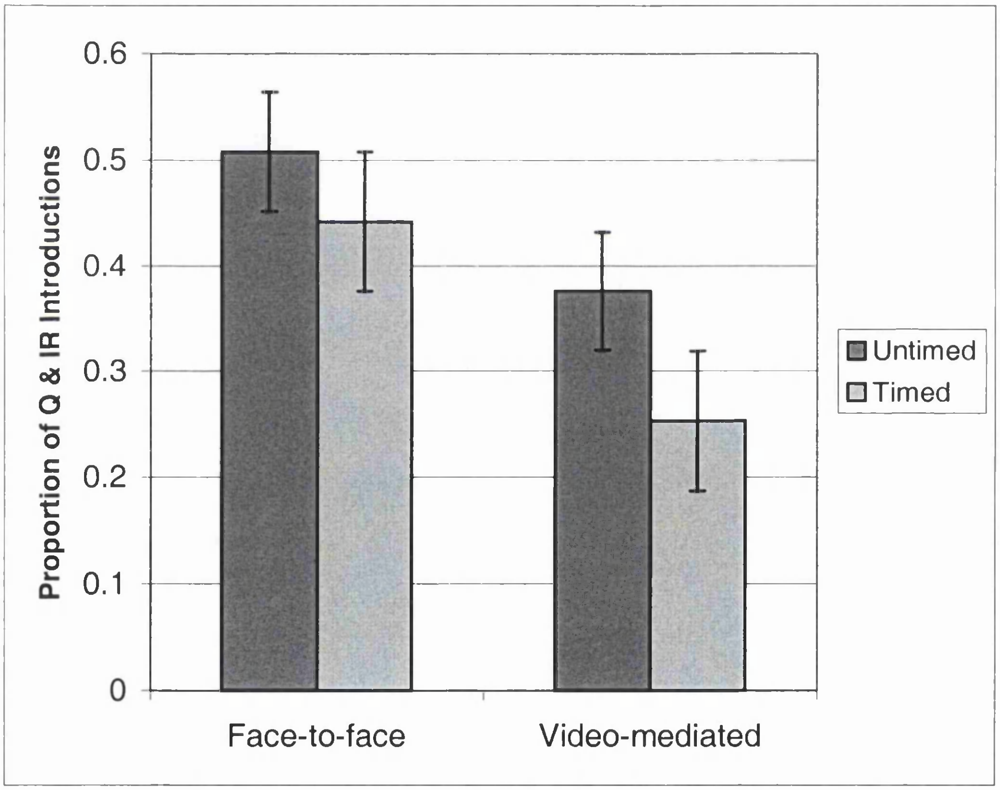


Figure 5.2. Mean proportion of Q & IR Introductions (± SE) for face-to-face (n = 32) and video-mediated (n = 32) conversational settings in untimed and timed conditions.

As can be seen from figure 5.2, the proportion of Q & IR Introductions to landmarks was lower in the timed condition than in the untimed condition for the face-to-face (0.44 cf. 0.51) and video-mediated conversational settings (0.25 c.f. 0.38). Figure 5.2 also indicates that the proportion of Q & IR Introductions in the untimed and timed conditions was lower in the video-mediated conversational setting than in the face-to-face setting. The results of the analysis of variance indicated that the overall differences observed in the use of Q & IR Introductions to landmarks were reliable.

A 2 x (2) analysis of variance (by dialogue) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the proportion of Q & IR Introductions to landmarks expressed as a ratio of 1. The summary table is presented in table 5.13 below.

Table 5.13. Analysis of variance (by dialogue) for the use of Q & IR Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	4.79	0.04*
S within-group error	30	(0.09)	
Within subjects			
Cognitive Load (L)	1	4.17	0.05*
S x L	1	0.38	0.54
S x S within-group error	30	(0.03)	

Note. Values enclosed in parentheses represent mean square errors

As expected, there was a main effect of conversational setting [F(1,30) = 4.79, p = 0.04]. Overall, the proportion of Q & IR Introductions was lower in the video-mediated conversational setting (0.31) than in the face-to-face conversational setting (0.47). There was also a main effect of cognitive load

[F(1,30) = 4.17, p = 0.05]. Overall, the proportion of Q & IR Introductions was lower in the timed condition (0.35) than in the untimed condition (0.44). There was no interaction between conversational setting and cognitive load (F<1).

Post hoc analyses were also carried out on the data. Newman-Keuls pairwise comparisons indicated that the effect of conversational setting did not reach significance in either the untimed or timed conditions. Nor did the effect of cognitive load reach significance in either the face-to-face or video-mediated conversational settings.

The data were also analysed by-item whereby the mean proportion of Question forms followed by an Informative Response (Q & IR) was calculated for each landmark. The mean proportion of Q & IR Introductions to landmarks with standard deviations for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.14 below.

Table 5.14. Mean proportion of Q & IR Introductions (by item) for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	29	0.50	0.24	0.42	0.27
Video-mediated	29	0.36	0.29	0.26	0.21

Table 5.14 shows that the mean proportion of Q & IR Introductions was lower in the timed condition than in the untimed condition for the face-to-face setting (0.42 cf. 0.50) and for the video-mediated setting (0.26 cf. 0.36). Furthermore, table 5.14 indicates that the proportion of Q & IR Introductions was lower in the video-mediated conversational setting than in the face-to-face

conversational setting for the untimed and timed conditions. The results of the analysis of variance indicated that the overall differences observed between untimed and timed conditions and between a video-mediated and a face-to-face conversational setting were reliable.

A 2 x (2) analysis of variance by-item was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The proportion of Q & IR Introductions was the dependent variable. The summary table is shown in table 5.15 below.

Table 5.15. Analysis of variance (by-item) for the use of Q & IR Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	6.81	0.01*
S within-group error	56	(0.09)	
Within subjects			
Cognitive Load (L)	1	6.02	0.02*
S x L	1	0.12	0.73
S x S within-group error	56	(0.04)	

Note. Values enclosed in parentheses represent mean square errors

There was a significant main effect of conversational setting [F(1,56) = 6.81, p = 0.01]. Overall, the proportion of Q & IR Introductions was lower in the video-mediated setting (0.31) than in the face-to-face setting (0.46).There was also a significant main effect of cognitive load [F(1,56) = 6.02, p = 0.02]. Overall, the proportion of Q & IR Introductions was lower in the timed condition (0.34) than in the untimed condition (0.43). There was no significant interaction (F < 1).

Post hoc analyses were also carried out on the data. Newman-Keuls pair-wise comparisons indicated that the effect of conversational setting was reliable in the timed condition ($p < 0.05$) but not in the untimed condition. The effect of cognitive load, however, did not reach significance in either the face-to-face or video-mediated setting. Overall, the results of the analysis by item replicated the results of the by-dialogue analysis. There were fewer Question Presentations followed by Informative Responses under time pressure (than without the pressure of time) and in a video-mediated conversational setting (compared with a face-to-face setting). This pattern of results is the same as that observed by Question-form Introductions overall. These findings support the hypothesis that interlocutors would be less collaborative under time pressure than without the pressure of time, and less collaborative in a video-mediated setting compared with a face-to-face setting.

In section 5.3.1, it was noted that the proportion of Question-form Introductions was lower in the video-mediated setting (compared with the face-to-face setting), and under time pressure (compared with no pressure of time). Conversely there were more Non-question Introductions in the video-mediated setting (compared with the face-to-face setting), as well as in the timed condition (compared with no pressure of time). This stands to reason, since if interlocutors are not using Question forms to introduce landmarks into the dialogue, they must be using Non-question forms. However, the speaker responding, may compensate for a less effective landmark presentation (i.e. a Non-question form) by offering an Informative Response. Anderson and Boyle (1994) observed that, on average, 10 landmarks were introduced by the Instruction Giver and 4 were introduced by the Instruction Follower. Typically, then, the speaker presenting landmarks will be the Instruction Giver and the speaker responding will be the Instruction Follower.

In order to investigate whether or not the Responder (typically the Instruction Follower) compensated for the reduction in Question-form Presentations in a video-mediated setting and under time pressure, the use of Non-question Presentations followed by Informative Responses (NQ & IR) was examined. Descriptive statistics presented in section 5.3.1 suggested that the proportion of NQ & IR Introductions would be significantly higher in the video-mediated setting than in the face-to-face setting, with little or no difference between the untimed and timed conditions.

Before proceeding with the analysis, Kolmogorov-Smirnov tests were carried out on the data in order to determine the whether the dependent variable was normally distributed across the untimed and timed conditions for both the video-mediated and the face-to-face settings. The results of the Kolmogorov-Smirnov tests are presented in table 5.16 below. The Kolmogorov-Smirnov Z statistic is provided along with the significance of the p value.

Table 5.16. Kolmogorov-Smirov tests for the proportion of NQ & IR Introductions to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Un-timed		*Timed*	
Conversational Setting	N	K-S Z	$p < 0.05$	K-S Z	$p < 0.05$
Face-to-face	32	0.25	0.01*	0.24	0.01*
Video-mediated	32	0.27	<0.01*	0.15	ns.

As can be seen from table 5.16, the dependent variable for several of the test samples was not normally distributed. The test samples were also tested for equality of variance. The results of Levene's test of equality of variance indicated that the error variance associated with the dependent variable was equal across conditions [$F(3,60) = 0.79$; $p = 0.50$]. Given that the data met with the homogeneity of variance assumption underlying analysis of variance and the number of dialogues in each cell of the experimental design was the same, the data were subjected to analysis of variance. According to Roberts and Russo (1999) deviation from the normality of distribution was not likely to results in Type I or Type II errors under these circumstances.

The mean proportions of NQ & IR Introductions to landmarks for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in figure 5.3.

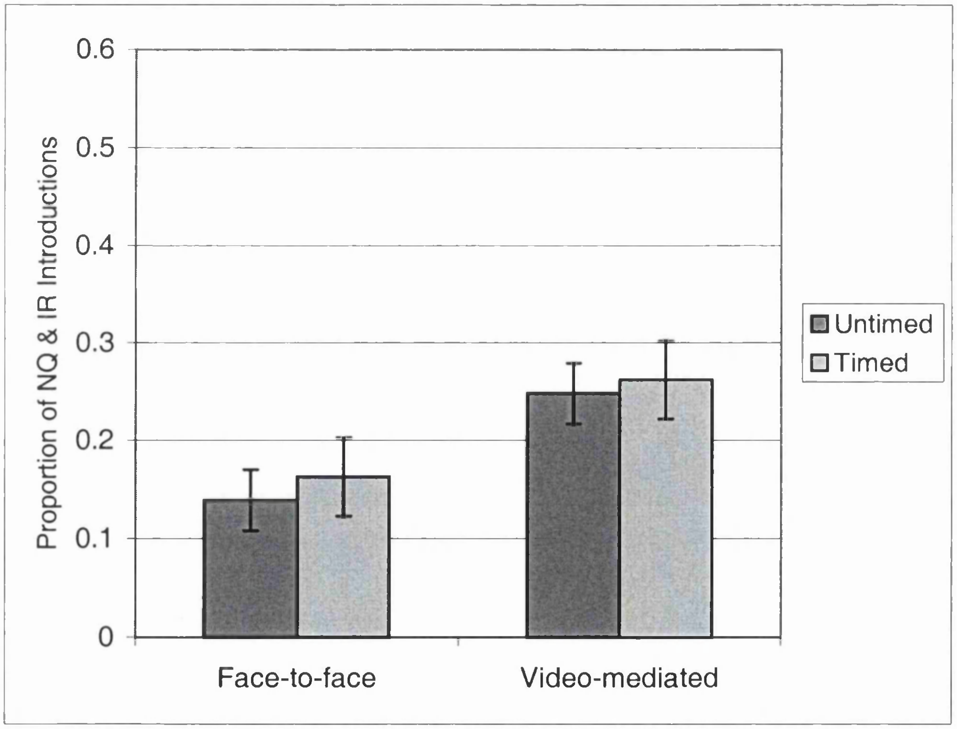


Figure 5.3. Mean proportion of NQ & IR Introductions to landmarks (± SE) for face-to-face (n = 32) and video-mediated (n = 32) conversational settings in untimed and timed conditions.

Figure 5.3 illustrates that the mean proportion of NQ & IR forms used to introduce landmarks was higher in the video-mediated conversational setting (mean untimed proportion = 0.25; mean timed proportion = 0.26) than in the face-to-face conversational setting (mean untimed proportion = 0.15; mean timed proportion = 0.16). The results of the analysis of variance indicated that the difference in the proportion of NQ & IR Introductions between the face-to-face and video-mediated conversational settings was reliable.

A 2 x (2) analysis of variance (by-dialogue) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the proportion of NQ & IR Introductions. The summary table is presented in table 5.17 below.

Table 5.17. Analysis of variance (by-dialogue) for the use of NQ & IR Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	4.65	0.04*
S within-group error	30	(0.03)	
Within subjects			
Cognitive Load (L)	1	0.08	0.78
S x L	1	<0.01	0.99
S x S within-group error	30	(0.04)	

Note. Values enclosed in parentheses represent mean square errors

There was a main effect of conversational setting F(1,30) = 4.65, p = 0.04]. Overall, the proportion of NQ & IR Introductions to landmarks was higher in the video-mediated setting (0.25) than in the face-to-face setting (0.15). There was no effect of cognitive load (F < 1) and no interaction (F < 1). Post hoc analyses were also carried out on the data. However, Newman Keuls pairwise comparisons revealed no significant effects.

The data was also analysed by-item whereby the mean proportion of Non-question forms followed by an Informative Response was calculated for each landmark. The mean proportions of NQ & IR introductions for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.18 below.

Table 5.18. Mean proportions of NQ & IR Introductions to landmarks (by-item) for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	29	0.14	0.16	0.15	0.18
Video-mediated	29	0.25	0.22	0.28	0.22

Table 5.18 shows that the mean proportions of NQ & IR Introductions for the untimed and timed conditions were higher for the video-mediated conversational setting (mean untimed proportion = 0.25; mean timed proportion = 0.28) than for the face-to-face setting (mean untimed proportion = 0.14; mean timed proportion = 0.15). The results of the analysis of variance indicated that the difference observed between the face-to-face and video-mediated conversational settings was reliable.

A 2 x (2) analysis of variance by-item was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the mean proportion of NQ & IR Introductions. The summary table is shown in table 5.19 below.

Table 5.19. Analysis of variance (by-item) for the use of NQ & IR Introductions to landmarks

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	8.70	<0.01*
S within-group error	56	(0.05)	
Within subjects			
Cognitive Load (L)	1	0.07	0.79
S x L	1	0.26	0.61
S x S within-group error	56	(0.03)	

Note. Values enclosed in parentheses represent mean square errors

There was a significant main effect of conversational setting [F(1,56) = 8.70, p < 0.01]. Overall, the proportion of NQ & IR Introductions was higher in the video-mediated setting (0.27) than in the face-to-face setting (0.15). There was no effect of cognitive load (F<1) and no interaction (F < 1). Post hoc analyses were also carried out on the data. Newman Keuls pairwise comparisons indicated that the effect of conversational setting was reliable in the timed condition, but did not reach significance in the untimed condition. There was no effect of cognitive load in either the face-to-face or video-mediated setting conversational setting. The overall results of the by-item analysis replicated the results of the by-dialogue analysis.

To summarise this section, the use of Question Presentations followed by an Informative Response (Q & IR) to introduce landmarks into the dialogue was influenced by cognitive load and by the conversational setting. Time pressure led to a reduction in the use of Q & IR Introductions and there were also fewer Q & IR Introductions in the video-mediated setting compared with the face-to-face setting. What is important to remember about Q & IR Introductions is that in section 5.3.1 of this chapter it was found that greater use of this type of Introduction was associated with good task performance. This suggested that collaboration between two interlocutors engaged in a dialogue is reflected in Question Presentations followed by Informative Responses. Since the use of Q & IR Introductions was influenced by cognitive load, this indicated that, under time pressure, interlocutors were less collaborative. Similarly, the finding that conversational setting influenced the use of Q & IR Introductions suggested that interlocutors were also less collaborative in a video-mediated setting compared with a face-to-face setting. These findings were consistent with the notion that, with respect to the way new entities (namely landmarks on a map) are introduced into a discourse, the degree to which interlocutors collaborate with each other to build common ground is influenced by the cognitive demands on the interlocutors and by the setting in which the conversation takes place.

The analysis of Non-question Presentations followed by Informative Responses suggested that, in a video-mediated conversational setting (compared with a face-to-face setting), the Responder, typically the Instruction Follower, compensated to some extent for the reduction in Question-forms by offering significantly more Informative Responses to Non-question Presentations. Yet, there was no indication that the Responder compensated for the effect of time pressure in this way.

Since the use of NQ & IR Introductions does not correlate with task performance, this particular type of introduction does not reflect effective collaboration between interlocutors. Thus, although the Responder may be compensating for a less collaborative Presenter, this does not necessarily offer an indication of how engaged interlocutors were with one another.

5.3.4 Informative Responses

One way to explore the issue of how engaged interlocutors were with one another was to consider the dialogue from the point of view of the Responder. The crucial question was how did the Responder behave when new entities were presented into the dialogue. According to the Interactive Alignment Model (Pickering and Garrod, in press) the responsibility for certain aspects of high-level planning, such as formulating the speaker's intentions, is distributed between interlocutors. For example, if a speaker produces a question, then the goal for the listener is already specified. If the interlocutors are interacting in a co-ordinated manner, then the listener should respond appropriately by answering a Question with an Informative Response. In their study, Anderson and Boyle (1994) observed that Question forms were more successful in eliciting Informative Responses than Non-question forms. However, in section 5.3.1 of this chapter, it was noted that the likelihood of an Informative Response having been elicited by a Question Presentation was reduced in the video-mediated conversational setting (compared with the face-to-face setting) and under time pressure (compared with no pressure of time).

In order to test whether these observations were reliable, the probability of an Informative Response elicited by a Question (as opposed to a Non-question Presentation) was examined in face-to-face and video-mediated, untimed and timed conditions. A measure of the likelihood of an Informative Response elicited by a Question (Q & IR) was obtained by calculating the conditional probability of a Q & IR Introduction, given an Informative Response (PQ&IR/IR). In this analysis I am specifically interested in the use of Informative Responses, as opposed to Non-informative Responses.

The data were first organised and analysed by dialogue. Before proceeding with the analysis, it was necessary to determine whether the distributions associated with the dependent variable were normal. Kolmogorov-Smirnov tests were carried out on the data and the results of these tests are shown in table 5.20 below.

Table 5.20. Kolmogorov-Smirov tests for the probability of an Informative Response elicited by a Question for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	K-S Z	p < 0.05	K-S Z	p < 0.05
Face-to-face	32	0.28	<0.01*	0.24	0.01*
Video-mediated	32	0.22	<0.03*	0.19	ns.

As can be seen from table 5.20, the distribution of the dependent variable was not normal across the experimental conditions. The test samples were also tested for equality of variance. The results of Levene's test of equality of variance indicated that the error variance associated with the dependent variables was equal across conditions [F(3,60) = 1.15, p = 0.34]. Furthermore, the sample sizes in each cell of the design were equal. Thus, following Roberts and Russo (1999), it was felt that a deviation from the normality of variance assumption underlying analysis of variance was unlikely to lead to a Type I or Type II error.

The mean probabilities of an Informative Response elicited by a Question Presentation for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in figure 5.4 below.

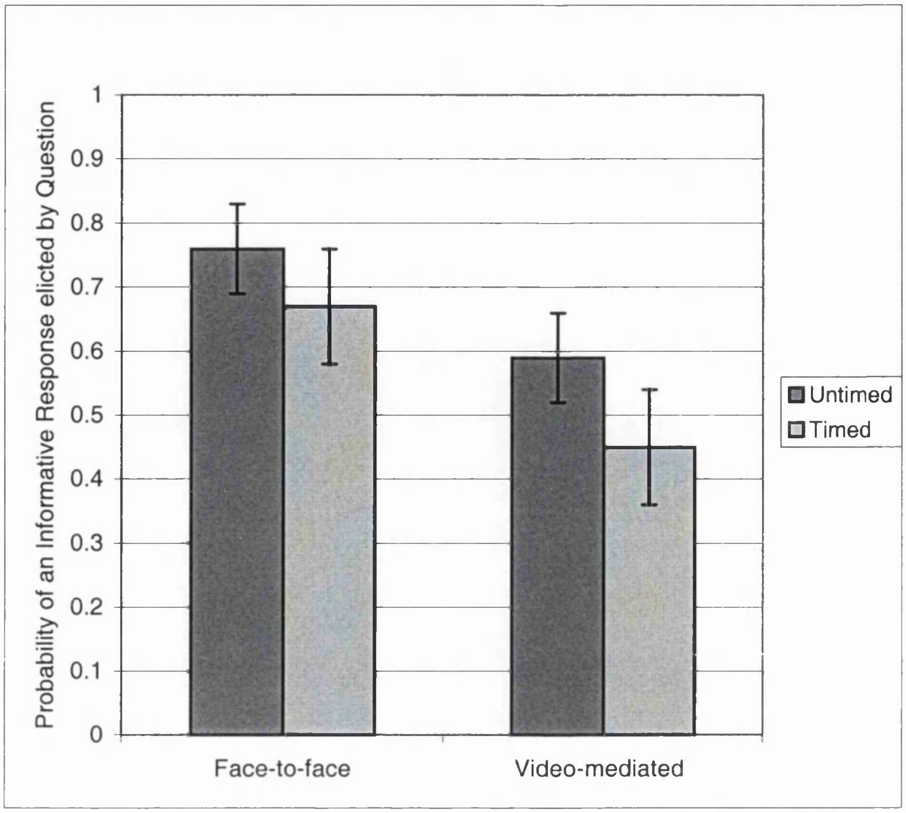


Figure 5.4. Mean probabilities (by-dialogue) of an Informative Response elicited by a Question Presentation (± SE) in face-to-face (n = 32) and video-mediated (n = 32) conversational settings in untimed and timed conditions.

As can be seen from figure 5.4, the probability of an Informative Response elicited by a Question Presentation was lower under time pressure than without the pressure of time in both the face-to-face conversational setting (mean timed probability = 0.67 cf. mean untimed probability = 0.76) and the video-mediated setting (mean timed probability = 0.45 cf. mean untimed probability = 0.59). Figure 5.4 also illustrates that the probability of an Informative Response elicited by a Question Presentation was lower in the video-mediated conversational setting than in the face-to-face setting. The results of an analysis of variance indicated that the differences observed in figure 5.4 were marginally significant.

A 2 x (2) analysis of variance (by dialogue) was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the probability of an Informative Response elicited by a Question (given an Informative Response). The summary table is shown in table 5.21 below.

Table 5.21. Analysis of variance (by-dialogue) for the probability of an Informative Response elicited by a Question Presentation

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	3.82	0.06
S within-group error	30	(0.16)	
Within subjects			
Cognitive Load (L)	1	3.25	0.08
S x L	1	0.20	0.66
S x S within-group error	30	(0.07)	

Note. Values enclosed in parentheses represent mean square errors

The effect of conversational setting was marginally significant [F(1,30) = 3.82, p = 0.06]. Overall, the probability of an Informative Response elicited by a Question Presentation was lower in the video-mediated conversational setting (0.52) than in the face-to-face conversational setting (0.72). The effect of cognitive load was also marginally significant [F(1,30) = 3.25, p = 0.08]. Overall, the probability of an Informative Response elicited by a Question Presentation was lower under time pressure (0.56) than without the pressure of time (0.67). There was no interaction between setting and load (F < 1).

Newman Keuls pairwise comparisons were also carried out on the data to determine whether the effect of conversational setting was significant at either level of cognitive load and whether the effect of cognitive load was significant at either level of conversational setting. However, the results of the pairwise comparisons did not indicate any reliable effects (at the 0.05 level of significance).

The data were also analysed by-item. The data for 10 landmarks were rejected from the analysis since there was only 1 instance of an Informative Response in a least one condition of the experimental design. Before proceeding with the analysis, Kolmogorov-Smirnov tests were carried out on the data to determine whether the distribution associated with the dependent variable was normal. All values of Z were non-significant indicating that the distribution of the dependent variable was in fact normal across the test samples. The samples were also tested for equality of variance. The results of Levene's test of equality of variance indicated that the error variance associated with the dependent variable was equal across conditions [F(2,72) = 2.57, p = 0.06]. Although the result of this test was marginally significant, the sample sizes in each cell of the design were equal. It was felt that this minor deviation from the homogeneity of variance assumption underlying analysis of variance was unlikely to lead to a Type I or Type II error (Roberts and Russo, 1999).

The mean probabilities with standard deviations of an Informative Response elicited by a Question Presentation for face-to-face and video-mediated conversational settings in untimed and timed conditions are shown in table 5.22 below.

Table 5.22. Mean probability (by-item) of an Informative Response elicited by a Question Presentation for face-to-face and video-mediated conversational settings in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	19	0.77	0.20	0.70	0.32
Video-mediated	19	0.54	0.35	0.50	0.21

As can be seen from table 5.22, the probability of an Informative Response elicited by a Question Presentation was lower for the video-mediated (VM) conversational setting than for the face-to-face (FTF) setting in both the untimed condition (mean VM probability = 0.54 cf. mean FTF probability = 0.77) and in the timed condition (mean VM probability = 0.50 cf. mean FTF probability = 0.70). Table 5.22 also illustrates that the mean probabilities of an Informative Response elicited by a Question Presentation were slightly lower in the timed conditions than in the untimed conditions. The results of an analysis of variance indicated that the overall difference observed between the face-to-face and video-mediated conversational settings were reliable.

A 2 x (2) analysis of variance by-item was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The dependent variable was the probability of an Informative Response elicited by a Question Presentation (given an Informative Response). The summary table is presented in table 5.23 below.

Table 5.23. Analysis of variance (by-item) for the probability of an Informative Response elicited by a Question Presentation

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	10.89	<0.01*
S within-group error	36	(0.08)	
Within subjects			
Cognitive Load (L)	1	0.76	0.39
S x L	1	0.04	0.84
S x S within-group error	36	(0.07)	

Note. Values enclosed in parentheses represent mean square errors

There was a significant main effect of conversational setting [F(1,36) = 10.89, $p < 0.01$]. Overall, the probability of an Informative Response elicited by a Question Presentation was lower for the video-mediated setting (0.52) than for the face-to-face setting (0.73). In contrast to the by-dialogue analysis, there was no effect of cognitive load ($F < 1$). Nor was there an interaction between conversational setting and cognitive load ($F < 1$). Post hoc analyses were also carried out on the data. The results of Newman-Keuls pairwise comparisons indicated that the effect of conversational setting was significant in both the untimed condition ($p < 0.05$) and in the timed condition ($p < 0.05$). There was no effect of cognitive load for either the face-to-face or the video-mediated conversational setting.

To summarise this section, the results of the analyses of variance offer some indication that participants in the video-mediated conversational setting were less likely to respond informatively to a Question Presentation than participants in the face-to-face conversational setting. Similarly, the pressure of time reduced the likelihood that participants would offer an Informative Response when a landmark on the map was introduced with a question. However, the results of the analysis of variance were only marginally significant and the effect of cognitive load was not replicated in the by-items analysis. It is not clear why this was the case. It is possible that some of the landmarks on the map were less likely to have been introduced by questions, because they were strategically less important. However, informal observations of the data did not seem to bear this out and none of the landmarks emerged as outliers. Nevertheless, the results of the analysis of Informative Responses offered some indication that interlocutors were less aligned, or co-ordinated in a video-mediated conversational setting (compared with a face-to-face setting) and under time pressure (compared with no pressure of time).

5.3.5 Issues Regarding the Categorisation of Introductions to Landmarks

Several issues arise with respect to the classifcation of introductions to landmarks in the present study. With respect to question forms, these were categorised according to the scheme used by Anderson and Boyle (1994). This method was intended to reflect the language function of questioning the existence of landmarks on the map. Where a speaker uses a syntactic question form, as in examples (27) to (29), the function of questioning the existence of a landmark is clear.

27. Do you see where the *waterfall* is?
28. Do you have a *beach hut*?
29. Have you got *sunflower fields*?

When considering declarative statements, however, interpretation of the intended function of a phrase or sentence may be ambiguous. For example, declarative statements such as (30) were categorised as non-questions in the present study and in that of Anderson and Boyle (1994).

30. I've got a *palm tree*

However, one might argue that the utterance in (30) is functionally equivalent to utterances such as (31).

31. I've got an *overnight accommodation.* Do you have that?

Arguably, a speaker may intend to use a declarative such as (30) as a truncated form of (31), in which case, the question, "*Do you have that?*" would be implicitly understood. If one accepts this line of reasoning, then a speaker who says, "*I've got a palm tree*" effectively invites their interlocutor to explicitly state whether or not they have a *palm tree* on their map. On this basis, declarative statements such as (30) could have been included in the question category rather than the non-question category.

The method of categorisation employed in the present study was based on the theoretical model proposed by Clark and Wilkes-Gibbs (1986). They argue that interlocutors engaged in a discourse collaborate with one another to establish a joint model of the discourse as it proceeds. One aspect of this collaborative strategy relates to the way new information is introduced into a discourse, which is the object of investigation here. However, it is not always clear what type of utterances reflect collaborative behaviour. For example, whether or not an utterance such as (30) is indicative of collaborative behaviour depends on whether or not one believes that the speaker invites an explicit acceptance or denial of the landmark in question.

Difficulty was also encountered when categorising responses to introductions of landmarks. For example, utterances such as *yeah*, *right*, *uh hum* are difficult to interpret since their meaning may change depending on the context in which they are uttered. This difficulty was discussed in section 5.2.2. Initially, following Anderson and Boyle (1994) an attempt was made to employ three categories of response labelled “fully informative”, “inadequately informative” and “uninformative”. Anderson and Boyle (1994) coded only responses to unshared landmarks. In the present study, however, responses to both shared and unshared landmarks were coded. Clark and Wilkes-Gibbs (1986) argue that interlocutors engaged in a discourse can be collaborative by allowing the conversation to proceed. For example, responses such as *yeah, right, uh hum* may be used to implicitly accept that an object constitutes part of the common ground. Thus, from a theoretical point of view, these responses could be categorised as *fully informative*. However, such utterances can be difficult to interpret within the discourse context. This problem was illustrated in example (19) of section 5.2.2 and is repeated in (32) for convenience.

32.

IG	And once we’re past that landmark we’re going straight down and then turning right just above the *camp site*
IF	Right

Following, the collaborative model (Clark and Wilkes-Gibbs, 1986), the Instruction Follower (IF) in (32) allows the conversation to proceed, by saying, “right”. By implication then, they have accepted the *camp site* as part of the common ground. Such a response could be considered *fully informative* on this basis. However, in their conversational turn, the Instruction Giver (IG) not only presents the *camp site* as new information but also gives directions. Consequently, it is ambiguous whether the Instruction Follower is acknowledging that they have understood the directions, or whether they have implicitly accepted the camp site as part of the common ground, or whether they have acknowledged both. On this basis, such as response could be classified as *inadequately informative.* In fact, in this particular case, the Instruction Follower did not have a *camp site* on their map and so must have been acknowledging that they understood the directions. Since they failed to point out that they did not have the *camp site* on their map, the response *"right"* could be categorised as *uninformative* even though this is inconsistent with Clark and Wilkes-Gibbs' (1986) model.

In the present study problems such as that exemplified in (32) caused difficulty in achieving reliability of coding when 3 categories were employed (even though the 2nd Coder was informed when the speaker did or did not have the landmark in question on their map). In order to resolve this difficulty, a decision was taken to employ only 2 broad categories. Accordingly, responses which were explicitly informative were classified as such and all other responses were classified as non-informative. This approach constitutes a compromise between achieving reliability of coding and a theoretically motivated means of classifying introductions to landmarks in accordance with the model proposed by Clark and Wilkes-Gibbs (1986).

In example (32) it was noted that the Instruction Giver incorporated the introduction of new information and the function of giving directions into one conversational turn. Interestingly, utterances categorised as non-questions often involved incorporating new information into the process of giving directions, as illustrated in examples (33) and (34).

(33) There's a *coach park* about halfway down the page.

(34) Just straight along the top and then west to underneath the *black gold.*

This would appear to result in sentences that are longer (in terms of the number of words) than questions regarding the existence of landmarks as in examples (35) and (36).

(35) Do you have a *beach hut*?

(36) Have you got *sunflower fields*?

In section 5.3.1, it was found that interlocutors used significantly fewer question forms (and more non-question forms therefore) in the video-mediated setting (compared with the face-to-face setting) and in the timed condition (compared with the untimed condition). This reduction in question forms, and more specifically in question forms followed by an informative response, was taken as an indication of a less collaborative communicative strategy under time pressure (compared with no pressure of time) and in a video-mediated conversational setting (compared with a face-to-face setting). These findings, together with the observations made in examples (33) to (36), suggest that conversational turns relating to the introduction of landmarks will be longer in a video-mediated setting (cf. a face-to-face setting) and under time pressure (cf. with no pressure of time).

In addition, by definition the Instruction Giver (IG) plays a more dominant role than the Instruction Follower (IF) in the Map Task. As noted by Anderson and Boyle (1994), Instruction Givers initiated significantly more introductions than Instruction Followers. Furthermore, responses to introductions typically require shorter utterances than presentations of landmarks. This is illustrated in examples (37) to (39).

37.

IG	Have you got the *precious stones* yeah?
IF	Uh hum.

38.

IG	Do you have a *chemical weapons plant*?
IF	I don't, no.

39.

IG	Go round the *waterfall* then start heading south.
IF	Right, got that. Yeah.

These observations suggest that, with respect to the introduction of new information, conversational turns uttered by the Instruction Giver will contain more words than those uttered by the Instruction Follower.

In section 5.3.4 there was some evidence to suggest that interlocutors were less aligned, or co-ordinated in the way they introduced entities into the dialogue. It was found that, in the video-mediated conversational setting (cf. the face-to-face setting) and under time pressure (cf. no pressure of time) the responder, typically the Instruction Follower, was less likely to offer an informative response when questioned about the existence of a landmark. Although these effects were only marginally significant, this offered some indication that interlocutors in the video-mediated setting (cf. the face-to-face setting) and in the timed condition (cf. the untimed condition) were less co-ordinated with respect to the way new entities, namely landmarks on the map were introduced into the dialogue.

If interlocutors were less co-ordinated with one another throughout the dialogue, then this should also be reflected in turn-taking behaviour. Turn-taking behaviour relates to the way interlocutors use cues, such as eye gaze and gesture, to regulating and synchronize each other's participation in the dialogue (e.g. Argyle and Cook, 1976). The reduced effectiveness of non-verbal cues in video-mediated conversational settings compared with face-to-face settings, has been shown to influence the regulation of turn-taking, with video-mediated leading to less interactive and more formal turn-taking behaviour (e.g. Sellen, 1995; O'Conaill, Whittaker and Wilbur, 1993). Furthermore, if, as suggested above, conversational turns become longer in a video-mediated setting (compared with a face-to-face setting) and under time pressure (compared with no pressure of time), then this could also lead to fewer turns under these conditions. In order to test these hypotheses an analysis of the length of conversational turns relating to introductions to landmarks and the number of turns per dialogue was carried out on the data.

5.3.6 Turn-taking behaviour

Conversational turns were examined in untimed and timed, face-to-face and video-mediated dialogues in order to test the hypotheses that:

(a) Dialogues performed in a video-mediated setting (cf. a face-to-face setting) and under time pressure (cf. no pressure of time) would contain fewer conversational turns;

(b) The length of conversational turns relating to the introduction of landmarks would be shorter for the Instruction Followers than for Instruction Givers; and that turn length would be shorter in a video-mediated setting (cf. with a face-to-face setting) and under time pressure (cf. no pressure of time).

The analysis of the length of turns relating to the introduction of landmarks is presented first.

Turn length relating to the introduction of landmarks

Turn length was measured in terms of the number of words in a conversational turn relating to the introduction of a landmark into the discourse. A converational turn was operationalised as a point in the dialogue were one speaker ceased speaking and the other began. It is typically, the Instruction Giver, who *presents* (or introduces) new information into the discourse while the Instruction Follower is typically the one who *responds* to the new information. Thus, for the pupose of analysis conversational turns by the Instruction Giver were those which incorporated the presentation (i.e. introduction) of a landmark and turns by the Instruction Follower were those which included a response to a landmark. Introductions which involved intervening turns between the presentation of a landmark and the corresponding response were excluded from the analysis. Conversational turns which contained introductions to 2 landmarks were included once in the analysis and were attributed to the first landmark. Utterances such as *er, uh hum* were included in the word count and abbreviations such as *I've* were counted as one word.

Kolmogorov Smirnov tests carried out on the data indicated that several of the test samples were not normally distributed [face-to-face, untimed, IG, Z = 0.22, p = 0.03; video-mediated, untimed, IG, Z = 0.26, p < 0.01; video-mediated, untimed, IF, Z = 0.23, p = 0.02; video-mediated, timed, IF, Z = 0.02]. Levene's test of equality of variance indicated that the error variance associated with the test samples was equal across groups [F(3,120) = 1.40, p = 0.22]. Although the test samples were not normally distributed, the error variance associated with the test samples was homogeneous and there was an equal number of subjects for each cell of the design. Following Roberts and Russo (1999) it was felt that major deviations from the normality-of-distribution assumption underlying analysis of variance would not result in a type I or type II error. Thus the data were subjected to an analysis of variance.

The mean turn lengths (in terms of number of words) for introductions to landmarks uttered by the Instruction Giver and the Instruction Follower for untimed and timed, face-to-face and video-mediated dialogues are shown in figure 5.5.

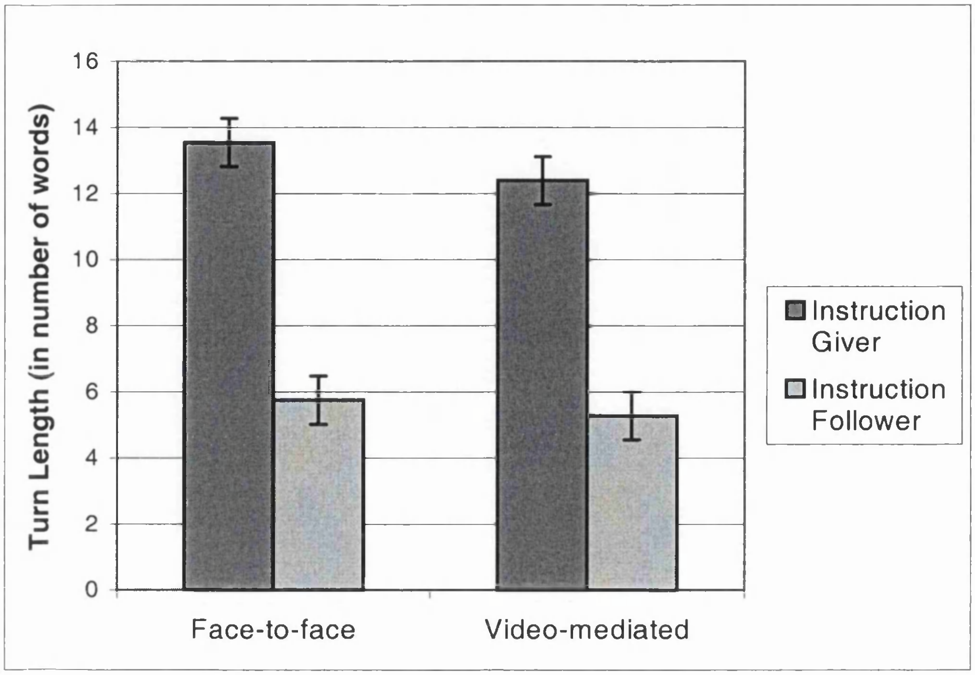


Figure 5.5. Mean turn length for introductions to landmarks for Instruction Givers (± SE) and Instruction Followers in face-to-face (n = 32) and video-mediated (n = 32) conversational settings (collapsed across untimed and timed conditions).

As can be seen from figure 5.5, conversational turns were longer for the Instruction Giver (IG) than for the Instruction Follower (IF) in both the face-to-face (mean IG turn length = 13.5 words; mean IF turn length = 5.7 words) and video-mediated (mean IG turn length = 12.4 words; mean IF turn length = 5.2 words) conversational setting. Analysis of variance carried out on the data showed that the differences observed in figure 5.5 were reliable.

A 2 x 2 x (2) analysis of variance was carried out on the data with conversational setting and task role (Instruction Giver and Instruction Follower) as between-subjects factors and cognitive load as a within-subjects factor. Turn length in terms of the number of words was the dependent variable. The summary table is shown in table 5.24 below.

Table 5.24. Analysis of variance for turn length

Source	df	F	Probability
Between subjects			
Conversational Setting (S)	1	1.22	0.27
Task Role (R)	1	102.83	*<0.01
S x R	1	0.22	0.64
S within-group error	60	(17.30)	
Within subjects			
Cognitive Load (L)	1	0.51	0.48
L x S	1	0.01	0.95
L x R	1	1.18	0.28
S x R x L	1	0.24	0.62
Error C x S	60	(11.67)	

Note. Values enclosed in parentheses represent mean square errors

As expected there was a main effect of task role [F(1,60) = 102.83, p < 0.01]. The mean turn length was significantly longer for the Instruction Givers (12.97 words) than for the Instruction Followers (5.52). There was no significant effect of conversational setting, no effect of cognitive load (F<1), and no significant interactions. The results of the analysis of variance supported the hypothesis that conversational turns incorporating introductions to landmarks by the Instruction Givers would contain more words than turns incorporating responses to introductions by the Instruction Follower. There was no evidence to suggest that conversational turns were significantly longer in the video-mediated setting compared with the face-to-face setting, or under time pressure compared with no pressure of time.

The number of conversational turns per dialogue

An analysis of the number of turns within the dialogue was also carried out on the data. Previous analysis of the data presented in section 4.3.1 of Chapter 4 indicated that timed dialogues were significantly shorter in duration and contained fewer words than untimed dialogues. In order to control for this, the number of turns per minute was used as the dependent variable. Kolmogorov Smirnov tests carried out on the data indicated that the distribution of the test samples was normal (all values of Z were non-significant). Levene's test of homogeneity of variance indicated that the error variance for the test samples was equal across groups [F(1,60) = 0.94, p = 0.42].

The mean number of turns per dialogue for the untimed and timed dialogues are shown in table 5.25 below.

Table 5.25. Mean number of turns per minute for face-to-face and video-mediated dialogues in untimed and timed conditions

		Cognitive Load			
		Untimed		*Timed*	
Conversational Setting	N	M	SD	M	SD
Face-to-face	32	22.02	3.06	22.93	3.73
Video-mediated	32	21.30	4.13	22.29	4.12

As can be seen from table 5.22 there was little difference between the timed and untimed conditions for either the face-to-face or video-mediated dialogues. The results of the analysis of variance indicated that there were no reliable differences between the face-to-face and video-mediated conversational settings or between the untimed and timed conditions.

A 2 x (2) analysis of variance was carried out on the data with conversational setting as a between-subjects factor and cognitive load as a within-subjects factor. The number of turns per minute was the dependent variable. The summary table is presented in table 5.26 below.

Table 5.26. Analysis of variance (by-dialogue) for the number of turns per minute

Source	df	F	Probability
Between subjects			
Conversational setting (S)	1	0.38	0.54
S within-group error	30	(19.10)	
Within subjects			
Cognitive Load (L)	1	1.49	0.23
S x L	1	<0.01	0.96
S x S within-group error	30	(0.96)	

Note. Values enclosed in parentheses represent mean square errors

As can be seen from table 5.26, there was no effect of conversational setting (F<1), no significant effect of cognitive load, and no interaction (F< 1). The hypothesis that dialogues would contain fewer turns in a video-mediated setting (cf. a face-to-face setting) and under time pressure (cf. no pressure of time) was not supported.

A consideration of the nature of conversational turns relating to introductions to landmarks suggests a complexity within the turns themselves which reveals a flaw in the reasoning that motivated the analyses presented in this section. The issue is illustrated in example (40) which is taken from a face-to-face, untimed dialogue.

40.

IG	Oh right well the next whatever it is icon thing south of the *chemical weapons plant* if you go to just the south west of that and then go east. Have you got *armoured vehicles*?

Example (40) illustrates that conversational turns which incorporate introductions to new information can be complex in nature. For example, the reference to *armoured vehicles* is a question containing only 5 words. However, this is embedded in a long conversational turn of 35 words which involves giving directions as well as establishing new information. Thus, while introductions involving questions may tend to be shorter in length than non-questions, this might not be reflected in the length of conversational turn. Consequently, any shift in focus from establishing definite reference to giving directions could be masked in the conversational turns relating to the introduction of new information.

5.4 General Discussion

The Collaborative Model of interpersonal communication (Clark and Wilkes-Gibbs, 1986; Clark, 1992) holds that when interlocutors participate in a conversation they collaborate with each other by constructing a joint model of the dialogue based on their mutual knowledge, or common ground. This is done in part, through the establishment of definite reference, or the building of knowledge of the entities under discussion. The notion of dialogue as an essentially co-ordinated activity has also been expressed more recently. Pickering and Garrod (in press) propose an Interactive Alignment Model of interpersonal communication which holds that interlocutors co-ordinate their mental representations through a largely unconscious process of “alignment”. In essence, the words, sounds and meanings used by one interlocutor may influence the words, sounds and meanings used by the other. The purpose of this chapter was to explore the influence of conversational setting and cognitive load on the way interlocutors collaborate with each other to introduce entities into a dialogue. In order to test the hypothesis that increased cognitive load and video-mediation would lead interlocutors to be less collaborative, the use of linguistic forms to introduce entities into a discourse was examined in face-to-face and video-mediated dialogues of participants performing the Map Task in timed and untimed conditions.

5.4.1 Summary of Results

New entities, namely landmarks on a map, could be introduced into the dialogue via a Question or Non-Question Presentation to which the Responder could provide either an Informative or Non-informative Response. It was found that Question-form Introductions and, more specifically, Question-form Presentations followed by an Informative Response (Q & IR), as illustrated in (41) below, were the most effective type of introduction in achieving successful task outcome.

41.

IG:	Have you got an *iron bridge*?
IF:	Yes, I do

Greater use of Q & IR Introductions to landmarks was associated with good task performance. Yet there was no correlation between task performance and Q & NIR (Question Presentation followed by a Non-informative Response) Introductions or NQ & IR (Non-question Presentation and Informative Response) Introductions. However, introductions to landmarks involving a Non-question Presentation followed by a Non-informative Response (NQ & NIR), as illustrated in (42) below, were found to be associated with poor task performance. An increase in the use of NQ & NIR Introductions seemed to lead to greater inaccuracy when drawing the route on the Map.

42.

IG	Keep coming straight across past the *broken gate*
IF	Right

With respect to the use of Introductions involving Question forms, it was found that increased cognitive load, due to time pressure, and a video-mediated conversational setting (compared with face-to-face setting) led to a reduction in the proportion of Question Presentations used to introduce landmarks into the dialogue. Furthermore, this pattern was repeated when a subset of Question forms, namely those followed by an Informative Response, was examined. These findings supported the hypothesis that interlocutors would be less collaborative under time pressure than without the pressure of time, and in a video-mediated conversational setting compared with a face-to-face setting.

When Introductions to landmarks involving Non-question Presentations followed by an Informative Response (NQ & IR) were considered, it was found that the proportion of NQ & IR Introductions was significantly higher in a video-mediated conversational setting compared with a face-to-face setting but that there was no effect of cognitive load. This suggested that Responders in a video-mediated setting, typically the Instruction Follower, may have compensated, to some extent, for a less collaborative Presenter, typically the Instruction Giver, by responding informatively even though they were not explicitly questioned about the existence of a landmark.

When the use of Informative Responses was considered, it was found that participants communicating in a video-mediated conversational setting were less likely to respond informatively to a Question than participants communicating in a face-to-face setting. Similarly, increased cognitive load, due to time pressure, also reduced the likelihood that participants would respond informatively to a Question Presentation. It should be noted, however, that the effects of conversational setting and cognitive load were only marginally significant. Nevertheless, these findings offered some indication that in a video-mediated conversational setting (compared with a face-to-face setting), and under time pressure (compared with no pressure of time), interlocutors were less aligned or co-ordinated in the way they introduced new information into the discourse. There was no indication, however, that a video-mediated conversational setting or increased cognitive load led interlocutors to be less co-ordinated in their regulation of turn-taking behaviour. Nevertheless, conversational turns incorporating the presentation of a landmark by the Instruction Giver were longer (in terms of the number of words) than responses from the Instruction Follower to introductions of landmarks. This reflected the more dominant role of the Instruction Giver in the Map Task.

5.4.2 The Influence of Conversational Setting on Collaboration

The conversational setting (in common with cognitive load) was found to influence the way interlocutors collaborated with one another to introduce entities into the dialogue. Participants communicating in a video-mediated conversational setting, used fewer Question form Presentations overall, and fewer Question-form Presentations followed by an Informative Response (Q & IR) in particular, than participants in the face-to-face conversational setting. Q & IR Introductions offered a clear example of the process of establishing common ground (or more specifically, definite reference) as described by the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992). The Presenter questions the existence of an object, or referent, being introduced to which the Responder explicitly accepts or denies its existence. This type of communicative strategy was found to be associated with good task performance. This indicated that the use of Question forms followed by an Informative Response reflected successful collaboration between interlocutors. It is for this reason that a reduction in the use of Q & IR Introductions indicated that Interlocutors communicating in a video-mediated conversational setting were less collaborative than interlocutors communicating in a face-to-face setting.

In section 2.6.4 of Chapter 2 of this thesis it was suggested that interlocutors in a video-mediated conversational setting may experience greater difficulty in establishing common ground than interlocutors in a face-to-face conversational setting, if visual cues, such as gaze and body language, were more difficult to interpret. In the present study, interlocutors communicating in the video-mediated setting viewed a head and shoulders image of their interlocutor and were unable to make eye contact by simply looking at their interlocutor. According to Clark and Brennan (1991), if participants in a conversation are restricted in their ability to monitor each others faces, gestures and their shared environment, this will affect their ability to minimise joint processing effort as expressed by the principle of least collaborative effort (Clark and Wilkes-Gibbs, 1986). A consequence of this is that the process of grounding will be rendered less efficient.

However, there are several problems with this explanation as an account of the present findings. First, in this particular study, the role of visual signals, such as gaze and body posture, could have been relatively trivial with respect to the process of establishing definite reference. The introductions to the landmarks on the map were categorised according to clear linguistic functions, such as "asking a question" and "giving an informative response". Informative Responses were, by definition, clear indications that an interlocutor either did or did not have an object. In other words, the linguistic information (e.g. *Do you see the red sea*? - *No I don't)* was clear and should have been sufficient to establish definite reference. It does not seem reasonable to suppose, therefore, that restricted visual signals would have led interlocutors to use this strategy less often? Second, it has been shown that participants in the Map Task rarely look at each other, but spend most of their time looking at the map itself. For example, in a video-mediated version of the Map Task, Clayes (2003) reported that participants in a 3-party conversation spent around 80% of the time looking at the on-line map and only 15% of the time looking at the video-images of their interlocutors. Furthermore, participants in the Map Task rarely exploit mutual gaze. For example, Doherty-Sneddon et al., (1997) found that interlocutors spent less than 5% of the time looking at each other. Even in face-to-face conversations speakers can spend little time (around 5%) looking at each other in referential communication tasks (Argyle, 1988; Argyle and Cook, 1976). Third, it has been suggested that, where mutual understanding cannot be conveyed through visual signals (because the interlocutors cannot see each other for instance), speakers will compensate by using more explicit verbal devices (Doherty-Sneddon et al., 1997). However, in the present study, there was no evidence to indicate that interlocutors used significantly more words in the video-mediated setting than in the face-to-face setting. This suggests that interlocutors did not need to compensate for the restricted use of visual signals.

For the reasons stated above, it seems unreasonable to account for the findings of the present study in terms of the principle of least collaborative effort. However, as discussed in section 4.4.4 of Chapter 4 the restricted use of visual signals could account for less collaborative behaviour in more complex video-mediated communication compared with co-present communication. For example, in the study conducted by Kraut, Fussell and Siegel (2003) participants in the video-mediated setting were engaged in a "hands on" bicycle repair task in which a remote "worker" wore a head mounted camera. Consequently, the movement of the worker radically affected the field of view. Objects under discussion were not always visible and the "helper" was unable to see the "worker's" face. Thus, in this particular case, the less effective nature of communication in the video-mediated setting compared with co-present communication may well have affected the process of establishing mutual understanding.

An alternative account of the findings of the present study stems from the suggestion that participants may have experienced a sense of social distance. In section 2.6.3 of Chapter 2, it was suggested that factors, such as the limited availability of visual cues and physical distance, contribute to a feeling of social distance (Short, Williams and Christie, 1976). For example, seeing a head and shoulders image of one's interlocutor should be less effective in creating a sense of intimacy between interlocutors than seeing one's interlocutor in the flesh (Burgoon et al., 1984; Biocca 1997cited in Tu, 2000). Furthermore, there is some evidence to indicate that a perception of physical distance can influence social aspects of communication such as collaboration (Bradner and Mark, 2002; Li and Mantei, 1992). In the present study, participants in the video-mediated conversational setting were located in different rooms.

The notion that a sense of social presence could have led to a less collaborative communicative style in a video-mediated setting, is consistent with the finding that interlocutors in the video-mediated setting were less likely to respond informatively to a question than interlocutors communicating in a face-to-face setting. According to the Interactive Alignment Model (Pickering and Garrod, in press) when speakers are aligned or co-ordinated in their communication, an interlocutor should answer a question if asked, because the communicative goal is already defined by the question. However, in the video-mediated conversational setting, there was some evidence to indicate that participants were less likely to respond informatively when asked a question (*e.g. Have you got waterfall? – Yes I have*). This suggested that interlocutors in the video-mediated setting were more egocentric or autonomous in the way they communicated than interlocutors in the face-to-face setting. It seems reasonable to suppose that this may have been due to a feeling of social distance.

A second reason why "social distance" offers a plausible account of the present findings relates to a consideration of the nature of the Map Task itself. While Introductions involving Question forms were more prevalent in face-to-face communication, Introductions involving Non-question Presentations were more prevalent in video-mediated communication. In the face-to-face conversational setting approximately 53% of Introductions to landmarks involved a Question form compared with 35% in the video-mediated setting. Non-question form Presentations, on the other hand, accounted for approximately 65% of Introductions in the video-mediated dialogues compared with 46% in the face-to-face dialogues. It is interesting to note that Non-question Introductions frequently, but not exclusively, involve incorporating the presentation of a new entity, or landmark, into the process of giving directions, as illustrated in (43) below.

43. So you're heading right towards the right and keep going past the *armoured vehicles.*

These observations suggested a difference in conversational style between the video-mediated and face-to-face settings. For interlocutors communicating in a face-to-face setting, there seemed to have been greater emphasis on questioning the existence of the landmarks, whereas interlocutors communicating in a video-mediated setting seem to have been more preoccupied with giving directions. Although, establishing which landmarks appear on both maps is an important aspect of the Map Task, the main function is, nevertheless, that of giving directions. If interlocutors were more egocentric in the way they communicated, they may have been more focussed on giving directions than on collaborating with their partner to establish which of the landmarks were shared and which were not. There is some evidence to indicate that people communicating in a video-mediated setting are more task oriented than people communication in a face-to-face setting (Rutter and Robinson, 1981). In the present study, however and contrary to expectation, this was not reflected in the length of conversational turns relating to the introduction of landmarks. Nevertheless, as pointed out in section 5.3.6, conversational turns in the Map Task often contain more than one function, as well as references to more than one landmark. Thus any shift in task focus may not have been evident in the length of conversational turns.

A problem with the "social presence" explanation arises in that an analysis of the number of turns in a dialogue offered no indication that interlocutors were less co-ordinated with respect to the regulation of turn-taking behaviour. Previous research has shown that interlocutors in a video-mediated setting are less interactive and more formal in terms of turn-taking behaviour than interlocutors in a face-to-face setting (e.g. Sellen, 1995; O'Conaill, Whittaker and Wilbur, 1993). The regulation and synchronization of turn taking reflects co-ordination between speakers with respect to who should speak and when. If speakers in a video-mediated setting had felt socially distant from one another, this should have been reflected in a reduced ability to co-ordinate turn-taking behaviour. In the present study, however, there was no evidence that this was the case, which is inconsistent with the social distance explanation. It is not clear why this was so. One possible explanation may lie in the suggestion any sense of social distance may impact differently on different aspects of co-ordination. The regulation of turn taking may be relatively straight forward where only 2 interlocutors are involved. Previous studies of turn taking behaviour have often been based on group conversations rather than dialogues between 2 interlocutors (e.g. Sellen, 1995; O'Conaill, Whittaker and Wilbur, 1993). Furthermore, analyses of turn-taking behaviour in 2-party dialogues of participants performing the Map Task has been based on comparisons between video-mediated communicaiton and audio only (or face-to-face communication compared with audio only) rather than a direct comparison between face-to-face and video-mediated conversations. Thus, any impact of a sense of social distance may be limited to more complex aspects of co-ordinated behaviour, such as the way interlocutors collaborate with one another to introduce and establish new information into a discourse. In the case of the Map Task this is a relatively complex aspect of interpersonal communication given that not all the landmarks on the map are shared between both speakers.

With respect to turn taking, it was found that conversational turns relating to the introduction of landmarks were longer for the Instruction Giver when presenting new information than for the Instruction Follower when responding to new information. This observsation is consistent with Anderson and Boyle's (1994) finding that, in the Map Task, the Instruction Giver introduces significantly more landmarks into the dialogue than the Instruction Follower. This reflects the more dominant role of the Instruction Giver in this particular task. The notion of a dominant speaker in a dialogue has also been put forward by McGrath (1990). According to McGrath (1990) participants in a conversation are viewed as a *social system* that carries out multiple functions. Group members contribute to the system as a whole but the contribution of one group member may differ from that of another. For instance, in a given conversational situation one member of the group may act as a "default" speaker. Thus in the present study, the observation that the reduction in Question-form Presentations was not paralleled by as dramatic a reduction in Informative Responses is indicative of the asymmetry in terms of the contribution of each speaker to the "social system". For example, in the face-to-face setting approximately 53% of Introductions involved Question Presentations and this dropped to 35% in a video-mediated setting (a decrease of 18%). In the case of Informative Responses, 63% of Introductions to landmarks involved an Informative Response and this dropped to 57% (only a 6% reduction in the number of Informative Responses). Thus it seems that, to some extent, the Responder can compensate for the reduction in Question forms by attempting to maintain the level of Informative Responses. Put another way, the Responder, who is typically the Instruction Follower, can still offer an Informative Response even though the Presenter did not explicitly ask for one. This is illustrated in (44) below

44.

IG	Keep coming straight across past the *broken gate*
IF	I haven't got a *broken gate*

Consistent with this line of reasoning, it was found that participants in the video-mediated conversational setting used significantly more NQ & IR (Non-question and Informative Response Introduction) as in (44), than participants in the face-to-face setting. These observations are consistent with McGrath's (1990) view of interlocutors as members of a social system wherein the contribution of one speaker may differ from that of another.

5.4.3 The Influence of Cognitive Load on Collaboration.

The findings reported in this chapter indicated that, with respect to the introduction of new information, the degree to which interlocutors collaborated with one another was also influenced by the cognitive load associated with the task. This was evidenced by the observation that time pressure led to a reduction in the use of Introductions to landmarks involving a Question Presentation followed by an Informative Response. As stated in section 5.4.2 above, this type of landmark offered a clear example, of the process of establishing definite reference since only Responses which comprised an explicit acceptance or denial of the landmark in question were included in this category. Moreover, a further indication that collaboration between interlocutors was reflected in this type of response stems from the observation that Q & IR Introductions were instrumental to successful task performance. Successfully establishing which entities form part of the common ground and which do not is an important aspect of the Map Task. Thus the more engaged interlocutors were in a collaborative interaction, the more successful they should have been at achieving the task goal.

The finding that interlocutors were less collaborative under time pressure is similar to that reported by Roßnagel (2000). He showed that increased cognitive load offset the way speakers adapted the language they used to the perspective of their listener. Similarly, in the present study, increased cognitive load led interlocutors to be less collaborative in their use of linguistic forms to introduce entities into a discourse. In the study conducted by Roßnagel (2000) participants instructed a confederate, either a boy or an adult, to construct a toy model. Under conditions of low cognitive load, speakers tended to substitute technical terms with descriptions of component parts of the model and elaborated their instructions when addressing the boy, as opposed to the adult. Yet under conditions of high cognitive load, due to increased task difficulty for instance, speakers relied more heavily on the use of technical terms and more basic step-by-step instructions irrespective of whom they were addressing.

In order to account for these findings, Roßnagel (2000) suggested that adapting one's speech to the listener, by embellishing instructions and describing component parts of a model (rather than using technical terms), required considerable cognitive effort. As a result, under conditions of increased cognitive load speakers relied more heavily on technical terms and less elaborate instructions because these had been well learned during the training phase of the experiment. In other words, ease of retrieval, rather than the needs of the listener dictated which information was incorporated into the speech plan. Roßnagel (2000) further speculated that the use of well-learned technical terms and step-by-step instructions reflected the operation of relatively automated components of utterance planning.

In the present study however, ease of retrieval from memory cannot offer a plausible explanation of why participants were less collaborative under time pressure since participants had not pre-learned the names of landmarks or directions for the map route. Nevertheless, the notion that complex processes, such as collaborating with one's interlocutor and task planning, are subject to the cognitive demands on the speaker (Bard et al., 2000; Roßnagel, 2000) is an interesting one. In line with Roßnagel's suggestion that the processes which underpin perspective taking may be sacrificed when the cognitive load in increased, the Dual Process Model (Bard et al., 2000) holds that complex processes compete with each other for the speaker's time and attention. Thus, if time is short some processes may not be completed, or may not be run at all.

In relation to the issue of taking the perspective of one's addressee, Horton and Gerrig (2002) suggest that it is important to consider the nature of speakers' experiences of interacting in a particular situation. In the previous section, it was suggested that the nature of the Map Task is such that the processes underlying the establishment of common ground may conflict with task planning since the main function of the task is to give directions. Following a parallel line of reasoning to that proposed by Bard et al., (2000) and Roßnagel (2000), one might suppose that under conditions of increased cognitive load, the process of establishing common ground is sacrificed to some extent in order to devote more attention to the task at hand, namely giving directions. In other words, under time pressure interlocutors must balance the desire to establish definite reference with the need to give directions in a timely fashion. This notion was not supported, however, by an analysis of the length of conversational turns relating to the introduction of landmarks on the map. It was expected that a shift in focus from establishing common ground to giving directions would be reflected in longer conversational turns under time pressure compared with no pressure of time. There was no indication that this was the case in the present study. Nevertheless, as pointed out in section 5.4.2 above, an analysis of conversational turns may not be an appropriate way of examining this issue.

Interestingly, informal observation of the dialogues indicated that some interlocutors treated giving directions and establishing which landmarks were shared, as distinct processes. For example, several pairs of participants in the present study adopted the strategy of questioning all the landmarks first before giving directions. This type of strategy would not have been reflected in an analysis of conversational turns relating to the introduction to landmarks but might be reflected through an analysis of the language functions used within the dialogue as a whole. Conversational Games Analysis (Kowtko, Isaard, Doherty-Sneddon, 1992) offers a suitable technique for undertaking this type of dialogue analysis. However, this technique involves categorising utterances in a discourse according to the language function expressed by each conversational “move” or “game”. Although this type of analysis was beyond the scope of this present work, a study of the influence of cognitive load on language function offers an interesting line of investigation for future research. Work carried out by Kelly and McGrath (1985) on group communication suggests that, under time pressure, groups will eliminate certain aspects of interpersonal communication, such as the evaluation of one another’s task strategies, which are unrelated to achieving the task goal. This suggests that the proportion of utterances relating to the main purpose of a task would be greater under time pressure than without the pressure of time.

Although there was some evidence to suggest that, under time pressure, interlocutors were less co-ordinated with respect to the way they introduced landmarks into the discourse, there was no indication that interlocutors were less co-ordinated with respect to the regulation of turn taking. While there was no difference in the number of turns per minute in the timed condition compared with the untimed condition, there was some evidence, that the likelihood of responding informatively to a question was reduced under time pressure. Although this effect was only marginally significant (by dialogue and not by item), it is nevertheless of theoretical interest for the following reason. A problem arises with respect to the suggestion that cognitive load influences alignment or co-ordination between interlocutors. According to the Interactive Alignment Model (Pickering and Garrod, in press), alignment occurs via automatic priming processes. If this is correct, then automated processes should not have been influenced by cognitive load. One possible explanation may reside in how attentive the Responder was to the Presenter. A pre-requisite for priming processes to bring about alignment is that interlocutors be attentive to what is said. It could have been the case that, under time pressure, the responder was less attentive and failed to note when they were being asked a question. With respect to the regulation of turn-taking behaviour, McGrath (1990) suggests that the timing of conversational turns may be regulated automatically. If this is the case, then the rate of conversational turns in a dialogue should be unaffected by any increase in cognitive load. This would explain why there was no indication that the number of conversational turns was influenced under time pressure (compared with no time pressure) in the present study.

Finally, the findings reported in this chapter may be limited in their generalisability to other conversational situations. For example, there is evidence to suggest that cognitive load effects can be offset where participants are more motivated to collaborate with each other. For example, in an adaptation to his original experiment Roßnagel (2000) showed that the effects of cognitive load could be offset when participants were more motivated to engage in the task. In a second version of his original experiment, Roßnagel (2000) instructed one group of participants that, following the model construction task, they would be video-taped and their performance would be reviewed in order to investigate the strategies they had used. The purpose of the instruction was to motivate participants to perform well in the task. It was found that without motivation, as previously observed, increased cognitive load led to an increase in the use of technical terms over descriptions of the component parts of the model and more basic step-by-step instructions rather than elaborate embellished instructions.

Interestingly, however, the effect of cognitive load was largely attenuated for the group who had been motivated to perform the task well.

It could also be the case that interlocutors vary in their motivation to collaborate with each other. Consistent with this notion, Anderson and Boyle (1994) identified 3 groups of interlocutors with distinct conversational styles. One group of interlocutors were more collaborative in their conversational style and used Question-form Introductions from the outset of the dialogue. A second group appeared to begin the task by assuming shared knowledge of the landmark features, but on feedback from the Instruction Follower to the contrary, Instruction Givers switched to a more collaborative style with greater use of Question forms. A third group of interlocutors appeared to be less collaborative in their conversational style, using few Question-form Introductions throughout the dialogue. An analysis of the route deviation scores indicated significant differences in task performance between the 3 groups. Those interlocutors who adopted a more collaborative communicative style produced better map routes than interlocutors who switched style or continued to use Non-question Introductions throughout the task. In the present study, informal observations of the face-to-face dialogues do suggest differences in communicative style between interlocutors. Unfortunately, however, there were too few dialogues to test these observations formally. For instance, it is generally recommended that multi-factorial analysis of variances requires a minimum of 8 speakers per cell of the design. Yet in the present study, there were 16 dialogues in each cell of the design. To identify 3 groups of interlocutors would have resulted, at best, in only 5 or 6 dialogues per group. This is before rejecting any speakers in order to maintain the balance of the original design. Nevertheless, the role of social presence, motivation, and individual differences in relation to cognitive load presents an interesting line of investigation for future work. On the basis of Anderson and Boyle's (1994) observations, it seems reasonable to suppose that interlocutors may vary in their motivation to collaborate with their interlocutor. The finding reported by Roßnagel (2000) that increased motivation may offset the effects of cognitive load further suggests that different types of communicators may vary in their response to cognitive load. Similarly, any effect of social distance might be attenuated where interlocutors are motivated to collaborate in order to get the task done.

In conclusion, the findings reported in this chapter suggest that the conversational setting and cognitive load combine to influence the degree to which interlocutors collaborate with one another. In a face-to-face setting and where there was no pressure of time, interlocutors were relatively more collaborative in the way they established common ground and seemed to be relatively more co-ordinated in their use of language. This type of behaviour is consistent with the Collaborative Model (Clark and Wilkes-Gibbs, 1986; Clark, 1992) as well as the Interactive Alignment Model (Pickering and Garrod, in press). However, where interlocutors communicated in a video-mediated conversational setting (compared with a face-to-face setting) and when under the pressure of time (compared with no time pressure), they were less collaborative in the way they established common ground and were less influenced by one another in their use of language.

These findings were consistent with the Dual Process Model (Bard et al., 2000) which holds that when the cognitive load associated with the task is increased, due to time pressure for instance, complex processes, such as collaborating with one's interlocutor and task planning, compete with each other for the speakers' time and attention. Thus, the desire to collaborate with one another to effectively establish common ground may be sacrificed in favour of getting the task done. It is suggested that conversational setting and cognitive load impact on the collaborative process for distinct, yet not unrelated reasons. For example, interlocutors communicating in a video-mediated setting may have felt less engaged with their interlocutor and less motivated to collaborate with them, due to a sense of social distance attributed to video-mediated communication as compared with face-to-face communication.

Chapter 6

General Discussion

6.1 Introduction

The purpose of this thesis was to explore the influence of cognitive load and conversational setting on referential communication in 2-party spoken dialogues. The research presented in this dissertation was motivated firstly, by the suggestion that spoken conversation is a "highly contextualised form of language use" (Levelt, 1989;29) which should be considered in relation to the setting, or location in which a conversation takes place (Clark,1996). Secondly, the dynamic nature of dialogue, namely that utterances must be produced in real time, suggests that certain processes underlying speech and language production may be subject to the cognitive demands on the speaker's time and attention (Bard et al., 2000). The principle questions of interest were: How would the conversational setting influence the way interlocutors referred to objects in a discourse? Would cognitive load impact on referential communication? If so, would the impact of cognitive load depend on the setting in which the conversation takes place? In order to address these issues, the influence of cognitive load, as illustrated by time pressure, was examined in face-to-face and video-mediated dialogues of pairs of participants performing The Map Task (Brown et al., 1984).

An initial study (Study 1) was conducted of video-mediated dialogues to examine the duration of repeated mentions of words forming the names of landmarks on the map. The results of Study 1 were reported in Chapter 3. A second study (Study 2) was carried out to examine the impact of cognitive load and conversational setting on articulatory and linguistic aspects of referential communication, as well as broad characteristics of the dialogue such as rate of speech, dialogue length and task performance. The results of Study 2 were reported in Chapter 4 and Chapter 5. A summary of the main findings is presented in section 6.2 below.

6.2 Summary of Results

Consistent with the Dual Process Model (Bard et al., 2000), the results of Study 1 indicated that, in a video-mediated conversational setting, speakers articulated repeated mentions of words forming object names more quickly than introductory mentions of words referring to the same object, irrespective of which speaker introduced the entity into the dialogue. This finding supported the hypothesis that *articulatory reduction* (the shortening in duration of repeated mentions of words referring to the same object) is underpinned by priming processes that are triggered by the *given* status (by virtue of previous mention) of an entity, within the discourse context. The same pattern of results had been observed in face-to-face dialogues (Bard et al., 2000) which suggested that, in terms of articulation at least, video-mediated communication functioned in the same way as face-to-face communication.

In Study 2, video-mediated dialogues were directly compared with face-to-face dialogues and cognitive load was increased by imposing a time limit in which to complete the task. The results of Study 2 (Chapter 4) indicated that the manipulation of cognitive load was effective. Consistent with Cognitive Load Theory (Sweller, 1988; Sweller, van Merriënboer and Pass, 1998) task performance was significantly worse under time pressure than without the pressure of time. This indicated that time pressure served to increase the cognitive load associated with the task. Furthermore, participants took significantly longer and used significantly more words to complete the task in their own time compared with performing the task under time pressure. This indicated that the time limit of 3 minutes was sufficient to put participants under considerable pressure.

The effects of conversational setting and cognitive load on articulation and reference are summarised in table 6.1 below.

Table 6.1. The effect of conversational setting and cognitive load on articulation and reference.

Aspect of Communication	Conversational Setting	Cognitive load
Articulation		
Articulatory reduction	No	No
Overall word duration	Yes	No
Rate of speech	Yes	No
Reference		
Quantity of references	No	Yes
Establishing definite reference	Yes	Yes

With respect to articulation (Chapter 4), table 6.1 shows that articulatory reduction occurred irrespective of the setting in which the conversation took place or by an increase in cognitive load (due to time pressure). This finding replicated the articulatory reduction effect observed in Study 1 with respect to untimed video-mediated dialogues. In contrast, conversational setting influenced the overall duration of 1st and 2nd mentions of words forming landmark names, and the rate of speech. This indicated that interlocutors communicating in a video-mediated setting spoke more slowly than interlocutors communicating in a face-to-face setting. Following Lindblom (1990) and Lindblom et al., (1995), this finding suggested that speakers may have adapted to a video-mediated setting by increasing articulatory effort to ensure their speech would be intelligible to listeners.

In relation to reference, cognitive load influenced the quantity of references to landmarks (Chapter 4) with interlocutors reducing the number of references to landmarks under time pressure. It was suggested that this finding reflected a specific strategy for coping with time pressure, hence there was no indication that the quantity of references was affected by conversational setting. Consistent with the Dual Process Model (Bard et al., 2000) increased cognitive load influenced complex decisions relating to task planning, but did not impact on the automatic priming processes thought to underpin articulatory reduction.

In relation to the way interlocutors introduced entities into the dialogue, the conversational setting and cognitive load combined to impact on the way interlocutors collaborated to establish definite reference. Observation of communication in a face-to-face conversational setting (compared with a video-mediated setting) and with no pressure of time (compared with time pressure) was consistent with the Collaborative Model (Clark and Wilkes-Gibbs, 1986) and with the Interactive Alignment Model (Garrod and Pickering, in press). For example, the proportion of Question-forms followed by an Informative Response was relatively high, indicating a greater degree of collaboration between interlocutors to establish common ground. Furthermore, there was some indication that the linguistic forms used by one interlocutor were influenced by those used by the other. Participants were highly likely to respond informatively when asked a question.

In contrast, observations of communication in a video-mediated setting (compared with a face-to-face setting) indicated that interlocutors were less collaborative with one another and there was some indication that they were less influenced by one another in their use of language (e.g. responding informatively to a question). These observations were not entirely consistent with the notion that the remoteness of video-mediated communication may have given rise to a feeling of social distance between the interlocutors (Short, Williams and Christie, 1976). Time pressure also led interlocutors to become less collaborative with one another and again, there was some indication that they were less influenced by one another in their use of language. This was consistent with the Dual Process Model (Bard et al.,2000), which holds that where the demands on the speakers are increased (due to time pressure in this case) costly processes, such as those involved in establishing common ground for example, may be sacrificed in favour of task planning.

6.3 Implications of Research

It has been suggested that "conversation is easy" and that human beings are "designed for dialogue" (Garrod and Pickering, in press). Furthermore, the literature on referential communication has placed a heavy emphasis on the collaborative nature of interpersonal communication (e.g. Clark and Wilkes-Gibbs, 1986; Clark, 1992). However, the findings of this dissertation suggest that when the communicative circumstances are less than ideal, because people find themselves communicating in a novel, remote conversational setting, or because the cognitive demands on the interlocutors are increased, then the apparent ease with which people communicate is soon disrupted. Perhaps, then, a view of spoken communication as a joint collaborative activity is best considered an "ideal" mode of communication which occurs when conditions are optimal. An optimal situation may be one in which interlocutors communicate in a familiar social setting, such as face-to-face, where there is relatively little cognitive pressure to get a given task done, for example.

The implications of the findings reported in this dissertation suggest that any theory of interpersonal communication will have to consider the dynamic nature of communication. Basic characteristics of a dialogue, such as where a dialogue takes place and the cognitive demands upon interlocutors affect the way people communicate. In addition to this, factors such as who one is talking to, and how motivated they are to engage with their interlocutor to achieve successful task outcome may also serve to influence the way people communicate in a given situation. Furthermore, on the basis of the findings presented in this dissertation, the picture that emerges is one in which spoken dialogue is best viewed as multi-faceted in nature. Thus when considering the influence of factors such as the conversational setting and cognitive load on spoken output the processes underlying speech and language production may not necessarily be affected in the same way. For example, processes which themselves place demands on working memory, such as task planning and collaborating with another person to establish the objects of conversation, may be influenced but processes which place no or little demand on working memory, such as automatic priming processes, may not (Bard et al., 2000).

The findings of this research also have important implications for designers of computer-based technologies such as videoconferencing. The evidence presented in this thesis indicated that interlocutors in a video-mediated setting become less collaborative and more autonomous in the way they interact with one another compared with a face-to-face setting. This has important consequences for the way people refer to objects in a discourse, at least where a collaborative problem-solving task is concerned: Particularly since the cognitive load associated with the task has also been shown to have a bearing on how well interlocutors interact with one another when referring to objects. Yet the development of more complex systems may not be the answer. It seems that, for goal-based dialogue at least, complex technical set ups, involving head-mounted cameras and complex arrays of textual and visual information may impede, rather than aid, successful communication (Kraut, Fussell and Siegel, 2003). Perhaps, as argued by Whittaker (2003), speech alone is sufficient for effective communication, particularly where problem-solving tasks are concerned.

6.4 Future directions

In the light of the thesis that the nature of communication between two interlocutors is influenced by the cognitive load associated with the task and by the setting in which a conversation takes place, it stands to reason that the findings reported in this dissertation may be limited in their generalisability to other communicative situations. For instance, the results of Studies 1 and 2 indicated that video-mediated communication functions differently from face-to-face conversation. However, this may only hold true where interlocutors are unfamiliar with the technology. For text-based communication at least, there is evidence to suggest that interlocutors are able to adapt to new technology such that communicative efficiency approaches that of face-to-face communication (Newlands, Anderson and Mullin, 2003). This then opens an interesting avenue for investigation. As interlocutors become more familiar with videoconferencing, for instance, will this offset any initial differences observed in the nature of communication, such as the rate of speech or the degree to which interlocutors collaborate with each other?

In addition to this, the effects of cognitive load observed here might be limited to task-oriented conversation. Problem-solving activities, by their very nature, place considerable demands on interlocutors. Thus in more relaxed types of communication, such as an informal social exchange, any increase in cognitive load due to manipulation of the task, may not be sufficient to impact on the communicative process. Furthermore, different types of cognitive load manipulations would not necessarily influence spoken communication in the same way. For example, the observation that speakers made fewer references to objects under increased cognitive load would seem to reflect a specific strategy appropriate for coping with time pressure. It does not seem reasonable to suppose, therefore, that speakers would also do this when faced with other types of cognitive load manipulations, such as increased task difficulty.

A further line of investigation relates to exploration of the distinction between consciously controlled and automatic process. For example, in order to explain why speaking rate was affected by the conversational setting, but not by cognitive load, it was speculated that speakers made an initial, consciously controlled, default decision to speak more slowly. However, subsequent regulation of articulatory pace could have become automated and hence unaffected by increased cognitive load. If this were the case, then it follows that the rate of speech should be consistently slower in a video-mediated setting compared with a face-to-face conversational setting irrespective of the stage of the dialogue, such as the beginning, middle or end.

Finally, as well as the consideration of where a conversation takes place and the cognitive load associated with the task, there may be additional factors which impact on the nature of referential communication in spoken dialogue. Conversations take place between different types of people, who may or may not be motivated to collaborate with one another. Motivation to collaborate may arise because interlocutors do not know each other very well, or because some interlocutors may be more motivated than others to perform a task well. Factors such as these may underpin Anderson and Boyle's (1994) observations that interlocutors vary in their conversational style with respect to the way they introduce entities into a dialogue. Thus interlocutors who know each other well, or who are motivated to perform a task successfully, may be more collaborative in their communication than others, irrespective of the setting in which the conversation takes place. Similarly, interlocutors who are more motivated to do a task well may be more resistant to cognitive load effects than those who are not (Roßnagel, 2000). Thus an interesting line of investigation would be to explore the influence of motivation and conversational styles on the way interlocutors collaborate with one another as they interact with cognitive load and the setting in which a conversation takes place.

References

Anderson, A. H., & Boyle, E. A. (1994). Forms of introduction in dialogues: Their discourse contexts and communicative consequences. *Language and Cognitive Processes, 9*(1), 101 - 122.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics, 59*(4), 580-592.

Anderson, A. H., Newlands, A., Mullin, J., Flemming, A. M., Doherty-Sneddon, G., & Van der Velden, J. (1996). Impact of video-mediated communication on simulated service encounters. *Interacting with Computers, 8*, 193-206.

Anderson, A. H., Mullin, J., Katsavras, E., Brundell, P., McEwan, R., Grattan, E., & O'Malley, C. (1999). *Multimediating multiparty interactions.* Paper presented at the Proceedings of INTERACT 99. IFIP.

Anderson, A. H., O'Malley, C., Doherty-Sneddon, G., Langton, S., Newlands, A., Mullin, J., Fleming, A. M., & Van der Velden, J. (1997). The impact of VMC on collaborative problem solving: an analysis of task performance, communicative process, and user satisfaction. In K. E. Finn & A. J. Sellen & S. B. Wilbur (Eds.), *Video-mediated communication* (pp. 133-172). NJ: Lawrence Erlbaum Associates.

Anderson, A. H., Bader, M., Bard, E., Boyle, E., Doherty-Sneddon, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Sotillo, C., & Thompson, H. (1991). The HCRC Map Task Corpus. *Language and Speech, 34*, 351-360.

Argyle, M. (1988). *Bodily Communication* (2nd ed.). London: Methuen.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze.* London: Cambridge University Press.

Ariel, M. (1990). Accessing NP antecedents. London: Routledge Croom Helm.

Austin, J. L. (1962). *How to do Things with Words*: Oxford University Press.

Baddeley, A. D. (1986). *Working Memory.* Oxford: Clarendon.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern-recognition and onset latency. *Journal of Memory and Language, 28*, 14-36.

Bara, B. G. (1995). *Cognitive Science: A Developmental Approach to the Simulation of the Mind*: Lawrence Erlbaum Associates Ltd.

Barber, P., & Laws, J. (1994). Cognitive Ergonomics- Empowering Designers and Users. *Ergonomics, 37*(11), 1749-1752.

Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language, 10*, 265-292.

Bard, E. G., & Aylett, M. (2001). *Referential form, word duration, and modeling the listener in spoken dialogue.* Paper presented at the Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society.

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language, 21*, 623-648.

Bard, E. G., Lowe, A., & Altman, G. (1989). *The effects of repetition on words in recorded dictations.* Paper presented at the Eurospeech, Paris.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*(1), 1-22.

Beattie, G. W. (1981). A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. *British Journal of Social Psychology, 20,* 243-248.

Blokland, A., & Anderson, A. (1998). Effect of low frame-rate video on intelligibility of speech. *Speech Communication, 26*, 97-103.

Bolinger, D. (1963). Length, vowel, juncture. *Linguistics, 1*, 5-29.

Bolinger, D. (1981). *Two Kinds of Vowels, Two Kinds of Rhythm*: Bloomington: Indiana University Linguistics Club.

Boyle, E. A., Anderson, A. H., & Newlands, A. (1994). The effects of visibility on dialog and performance in a cooperative problem-solving task. *Language and Speech, 37*, 1-20.

Bradner, E., & Mark, G. (2002). *Why distance matters: Effects on cooperation, persuasion and deception.* Paper presented at the CSCW '02, New Orleans, Lousiana, USA.

Branigan, H. P., & McLean, J. F. (2003). *Something old, something new: Addressee knowledge and the given-new contract.* Paper presented at the Twenty-fifth Annual Conference of the Cognitive Science Society, Boston.

Branigan, H. P., Pickering, M., & Cleland, A. A. (2002). Syntactic alignment and participant status in dialogue. *Manuscript submitted for publication.*

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 482-493.

Brenner, M., Doherty, E. T., & Shipp, T. (1994). Speech measures indicating workload demand. *Aviation Space and Environmental Medicine, 65*(1), 21-26.

Brown, P., & Dell, G. (1987). Adapting production to comprehension - the explicit mention of instruments. *Cognitive Psychology, 19*, 441-472.

Brown, G., Anderson, A. H., Yule, G., & Shillcock, R. (1984). *Teaching Talk.* Cambridge: Cambridge University Press.

Buhl, H. M. (2001). Partner orientation and speaker's knowledge as conflicting parameters in language production. *Journal of Psycholinguistic Research, 30*(6), 549-567.

Burgoon, J. K., Buller, D. B., Hale, J. L., & Turck, D. (1984). Relational messages associated with nonverbal behaviours. *Human Communication Research, 10*(3), 351-378.

Butterworth, B. (1980). *Language Production: Vol 1., Speech and Talk*: Academic Press.

Campbell, W. N., & Isard, S. (1991). Segment durations in a syllable frame. *Journal of Phonetics, 19*, 37-47.

Carroll, T. M. (1980). Naming and describing in social communication. *Language and Speech, 23*, 309 - 322.

Chafe, W. (1974). Language and consciousness. *Language, 50*, 111-133.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.

Chapanis, A., Ochsman, R., Parrish, R., & Weeks, G. (1972). Studies in interactive communication: I The effects of four communication modes on the behaviour of teams during cooperative problem solving. *Human Factors, 14*, 487-509.

Chapanis, A., Ochsman, R., Parrish, R., & Weeks, G. (1977). Studies in interactive communication. II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving. *Human Factors, 19*, 487-509.

Cherry, E. C. (1956). *On Human Communication.* Cambridge, MA: MIT Press.

Chung, G. K. W. K., O'Neill, H. F., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15*, 463-493.

Clark, H. H. (1992). *Arenas of Language Use.* Chicago: The University of Chicago Press.

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

Clark, H. H. (1971). The importance of linguistics for the study of speech hesitations. In D. L. Horton & J. J. Jenkins (Eds.), *The Perception of Language* (pp. 69-78). Columbus: Merrill.

Clark, H. H., & Brennan. (1991). Grounding in communication. In L. B. Resnick & J. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* Washington DC: American Psychological Society.

Clark, H. H., & Clark, E. V. (1977). *Psychology and Language: An Introduction to Psycholinguistics.* New York: Harcourt Brase Jovanich.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi & B. Webber & I. Sag (Eds.), *Elements of Discourse Understanding* (pp. 111-222). Cambridge: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*(3), 201-242.

Clayes, E. (2003). *Multimediating Groups.* Unpublished doctoral thesis, University of Glasgow.

Daft , R., & Lengel, R. (1984). Information richness: a new approach to managerial behavior and organizational design. In B. Straw & L. Cummings (Eds.), *Research in Organizational Behaviour.* Greenwich, CT: JAI Press.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the 'model of the listener'. In D. J. Napoli & J. A. Kegel (Eds.), *Bridges Between Psychology and Linguistics*. Hillsdale NJ: Erlbaum.

Doherty-Sneddon, G., Anderson, A. H., 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), 1-21.

Eveland, W. P., & Dunwoody, S. (2001). User control and structural isomorphism or disorientation and cognitive load? Learning from the Web versus print. *Communication Research, 28*(1), 48-78.

Fay, N. (2000). *The Effect of Group Size upon Influence and Process in Group Communication.* Unpublished PhD thesis, University of Glasgow.

Fish, R. R., Kraut, R., Root, R., & Rice, R. E. (1993). Video as technology for informal communication. *Communications of the ACM, 36*(1), 48-61.

Fisher, C., & Tokura, H. (1995). The Given-New contract in speech to infants. *Journal of Memory and Language, 34*, 287-310.

Fowler, C. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 28*, 47-56.

Fowler, C., & Housum, J. (1987). Talkers signalling 'new' and 'old' words in speech, and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489-504.

Fowler, C., Levy, E., & Brown, J. (1997). Reductions of spoken words in certain discourse contexts. *Journal of Memory and Language, 37*, 27-40.

Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378-391.

Garrod, S., & Anderson, A. H. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181-218.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form.* New York: Academic Press.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3). New York: Academic Press.

Gundel, J. K., Hendberg, N., & Zacharski, R. (1993). Cognitive status in the form of referring expressions in discourse. *Language, 69*, 274-307.

Halliday, M. A. K. (1967). Notes on transivity and theme in English: II. *Journal of Linguistics, 3*, 199-244.

Haviland, S. E., & H, C. H. (1974). What's new? Acquiring New information as a process in comprehension. *Journal of Verbal Learning and Verbal Behaviour, 13*, 512-521.

Hawkins, W., & Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics, 22*, 493 - 511.

Heath, C. C. (1986). *Body Movement and Speech in Medical Interaction.* Cambridge: Cambridge University Press.

Heath, C., & Luff, P. (1992). Media Space and Communicative Asymmetries: Preliminary Observations of Video-mediated Interaction. *Human-Computer Interaction, 7*, 315-346.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91-117.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language, 47*, 589-606.

Howell, D. C. (1995). *Fundamental Statistics for the Behavioural Sciences* (3rd edition ed.). Belmont, California: Duxbury Press.

Hunnicutt, S. (1985). Intelligibility versus redundancy - conditions of dependency. *Language and Speech, 28*(1), 47-56.

Hupet, M., & Chaintraine, Y. (1992). Changes in repeated reference: Collaboration or repetition effects? *Journal of Psycholoinguistic Reserach, 21*, 485 - 496.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*, 26-37.

Jefferson, G. (1982). On exposed and embedded correction in conversation. *Studium Linguistik, 14*, 58-68.

Kaiser, E., & Trueswell, J. (2003). *The quest for a referent: Investigating the interpretation of pronouns and demonstratives in real time.* Paper presented at the The Sixteenth Annual CUNY Conference on Human Sentence Processing, Cambridge, MA.

Karttunen, L., & Peters, S. (1975). Conventional Implicature of Montague Grammar. *Berkely Linguistics Society, 1*, 266-278.

Kelly, J., & McGrath, J. (1985). Effects of time limits and task types on task performance and interaction of four person groups. *Journal of Personality and Social Psychology, 49*, 395-407.

Kelly, J., Futoran, G., & McGrath, J. (1990). Capacity and capability: seven studies of entrainment of task performance in task performance rates. *Small Group Research, 21*, 283-314.

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica, 32*, 1-25.

Kendon, A. (1990). *Conducting Interaction: Studies in the Behaviour of Social Interaction*. Cambridge: Cambridge University Press.

Kintsh, W., & van Dyke, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kopardekar, P., & Mital, A. (1997). *Is there any effect of workload on speech parameters of air traffic controllers?* Paper presented at the Annual International Occupational Ergonomics and Safety Conference, Washington D C.

Kowtko, J., Isard, S., & Doherty-Sneddon, G. (1992). *Conversational games within dialogue* (Technical Report HCRC RP-31): HCRC Publications. University of Edinburgh.

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgens & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 655 - 701). New York: Guildford Press.

Krauss, R. M., & Weinheimer, S. (1964). Changes in length of reference phrases as a function of social interaction: A preliminary study. *Psychonomic Science, 1*, 113-114.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343-346.

Kraut, R. E., Fussell, S., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction, 18*(1-2), 13-49.

Levelt, W. J. M. (1989). *Speaking from Intention to Articulation.* Cambridge, Massachusetts: MIT Press.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in questioning answering. *Cognitive Psychology, 14*, 78-106.

Li, J., & Mantei, M. (1992). *Working together, virtually.* Paper presented at the Graphics Interface 92.

Lieberman, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172-187.

Lindblom, B. (1990). Explaining Variation: A Sketch of the H and H Theory. In W. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer Academic.

Lindblom, B., Guion, S., Hura, S., Moon, S., & Wilkerman. (1995). Is sound change adaptive? *Rivista di Linguistica, 7*(1), 5-37.

Lively, S. E., Pisoni, D. B., Vansummers, W., & Bernacki, R. H. (1993). Analyses and Perceptual Consequences. *Journal of the Acoustical Society of America, 93*(5), 2962- 2973.

Lounsbury, F. G. (1954). Transitional probablility, linguisic structure and systems of habit family hierarchies. *Journal of Abnormal and Social Psychology, 49*(Supplement, Osgood C E, Sebeok T A (eds)), 93-101.

McAllister, J., Potts, A., Mason, K., & Marchant, G. (1994). Word duration in monologue and dialog speech. *Language and Speech, 37*, 393-405.

McGrath, J. (1990). Time matters in groups. In J. Galegher & R. E. Krautt & C. Egido (Eds.), *Intellectual Teamwork Social and Technological Foundations of Cooperative Work* (pp. 23-61): Lawrence Erlbaum.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behaviour, 19*, 668-682.

Mendoza, E., & Carballo, G. (1998). Acoustic analysis of induced vocal stress by means of cognitive workload tasks. *Journal of Voice, 12*(3), 263-273.

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 213-222.

Mullin, J., Anderson, A. H., Smallwood, L., Jackson, M., & Katsavras, E. (2001). *Eye-tracking explorations in multimedia communications.* Paper presented at the Human Computer Interaction. People and Computers XV. Interaction without Frontiers., Joint Proceedings of HCI 2001 and IHM 2001.

Newlands, A., Anderson, A. H., & Mullin, J. (2003). Adapting communicative strategies to computer-mediated communication: An analysis of task performance and dialogue structure. *Applied Cognitive Psychology, 17*(3), 325-348.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition.* Cambridge: Cambridge University Press.

O'Conaill, B., Whittaker, S., & Wilbur, S. (1993). Conversation over videoconferences: An evaluation of the spoken aspects of video-mediated communication. *Human-computer Interaction, 8*, 389 - 428.

Olson, G. M., & Olson, J. S. (2000). Distance Matters. *Human-Computer Interaction, 15 (2-3),* 139-178.

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

Osbourne, R. E., & Gilbert, D. T. (1992). The Preoccupational Hazards of Social-life. *Journal of Personality and Social Psychology, 62*(2), 219-228.

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

O'Seaghdha, P. G. (1997). Conjoint and dissociable effects of syntactic and semantic context. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 807-828.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measure of Meaning.* Urbana, IL: University of Illinois.

Oviatt, S. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language, 9*(3), 19-35.

Paas, F. G. W. C., & van Merrienboer, J. J. G. (1994). Instructional-control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351-371.

Pickering, M., & Garrod, S. (in press). Toward a mechanistic psychology of dialogue. To appear in *Behavioral and Brain Sciences.*

Pollack, I., & Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Language and Speech, 6*, 165-171.

Prince, E. (1981). Towards a taxonomy of Given-New information. In P. Cole (Ed.), *Radical Pragmatics.* New York: Academic Press.

Rantala, L., Vilkman, E., & Bloigu, R. (2002). Voice changes during work: Subjective complaints and objective measurements for female primary and secondary schoolteachers. *Journal of Voice, 16*(3), 344-355.

Rice, R. E. (1993). Media appropriateness: Using Social Presence Theory to compare traditional and new organization media. *Human Communication Research, 19*(4), 451-484.

Roberts, M. J., & Russo, R. (1999). *A Student's Guide to Analysis of Variance.* London and New York: Routledge.

Robertson, C., & Kirsner, K. (2000). Indirect memory measures in spontaneous discourse in normal and amnesic subjects. *Language and Cognitive Processes, 15*(2), 203-222.

Roßnagel, C. (2000). Cognitive load and perspective-taking: applying the automatic-controlled distinction to verbal communication. *European journal of social psychology, 30*, 429 - 445.

Rochester, S. R., Thurnston, S., & Rupp, J. (1977). Hesitations as dues to failures in coherence: A study of the thought disordered speaker. In S. Rosenberg (Ed.), *Sentence Production Developments in Research Theory* (pp. 65-87). N J: Erlbaum: Hillsdale.

Rutter, D. R. (1987). *Communicating by Telephone.* New York: Pergamon Press.

Rutter, D. R., & Robinson, B. R. (1991). An experimental analysis of teaching by telephone. *Progress in Applied Social Psychology, 15,* 143-178.

Rutter, D. R., Stephenson, G. M., & Dewey, M. E. (1981). Visual communication and the content and style of conversation. *British Journal of Social Psychology, 20*, 41-52.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696-735.

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*, 175-194.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing Discourse: Text and Talk* (pp. 71-93). Washington DC: Georgetown University Press.

Schiffer, S. R. (1972). *Meaning*. Oxford: Oxford University Press.

Schober, M. F. (1998). Different kinds of conversational perspective-taking. In S. R. Fussell & R. J. Krauss (Eds.), *Social and Cognitive Approaches to Interpersonal Communication* (pp. 145-174). Mahwah, NJ: Erlbaum.

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

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

Searle. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Sellen, A. (1995). Remote conversations: The effect of mediating talk with technology. *Human Computer Interaction, 7*, 347-374.

Short, J., Williams, E., & Christie, B. (1976). *The Social Psychology of Telecommunications*: Wiley.

Snow, C. (1972). Mothers' speech to children learning language. *Child Development, 43*, 549-565.

Snow, C. E., & Ferguson, C. A. (1977). *Talking to Children: Language Input and Acquisition.* Cambridge: Cambridge University Press.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics* (Vol. 9, pp. 315-332). New York: Academic Press.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1980). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In R. A. Cole (Ed.), *Perception and Production of Fluent Speech* (pp. 469-505). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sweller, J. (1988). Cognitive load during problem-solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176-192.

Tannen, D. (1994). *Gender and Discourse.* New York: Oxford University Press.

Tu, C. (2000). On-line learning migration: from social learning theory to social presence theory in a CMC environment. *Journal of Network and Computer Applications, 23*, 27-37.

Walther, J. (1992). A longitudinal experiment on relational tone in computer mediated and face to face interaction. Paper presented and the Hawaii Interactional Conference on Social System Sciences, Los Alamitos, CA.

Walther, J. (1994). Anticipated ongoing interaction versus channel effects on relational communication in computer mediated interaction. *Human Communication Research, 20*, 473-501.

Walther, J, (1995). Relational aspects of computer-mediated communication: experimental observations over time. *Organization Science, 6(2).* 186-203.

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language, 37*, 356-381.

Whitmore, J., & Fisher, S. (1996). Speech during sustained operations. *Speech Communication, 20*(1-2), 55-70.

Whittaker, S. (2003). Things to talk about when talking about things. *Human-Computer Interaction, 18*(1-2), 149-170.

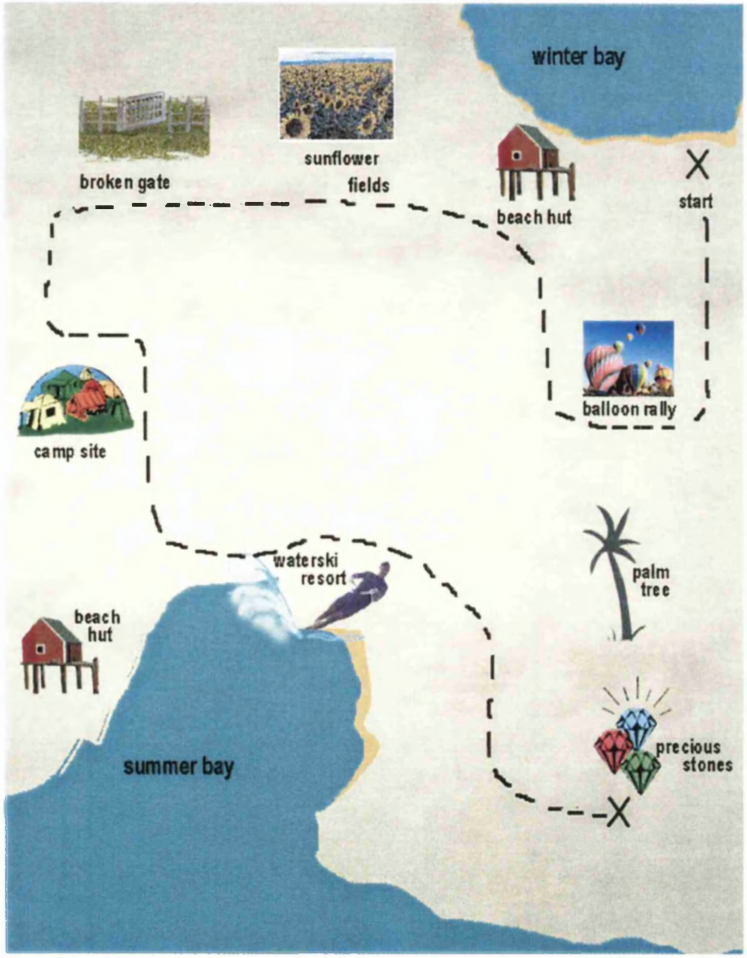
Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183-194.

Williams, E. (1977). Experimental comparisons of face-to-face and video-mediated communication: A review. *Psychological Bulletin, 84*, 963-976.

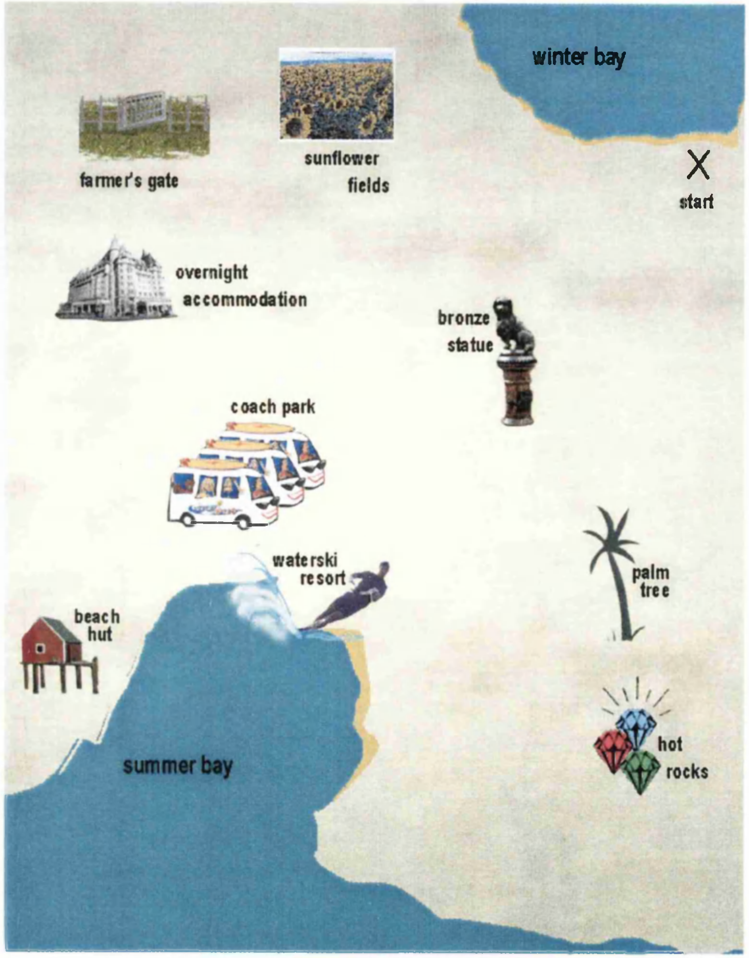
Yngve, V. H. (1970). *On Getting a Word in Edgewise.* Chicago: Chicago Linguistics Society.

Appendix 1: Maps

Map A

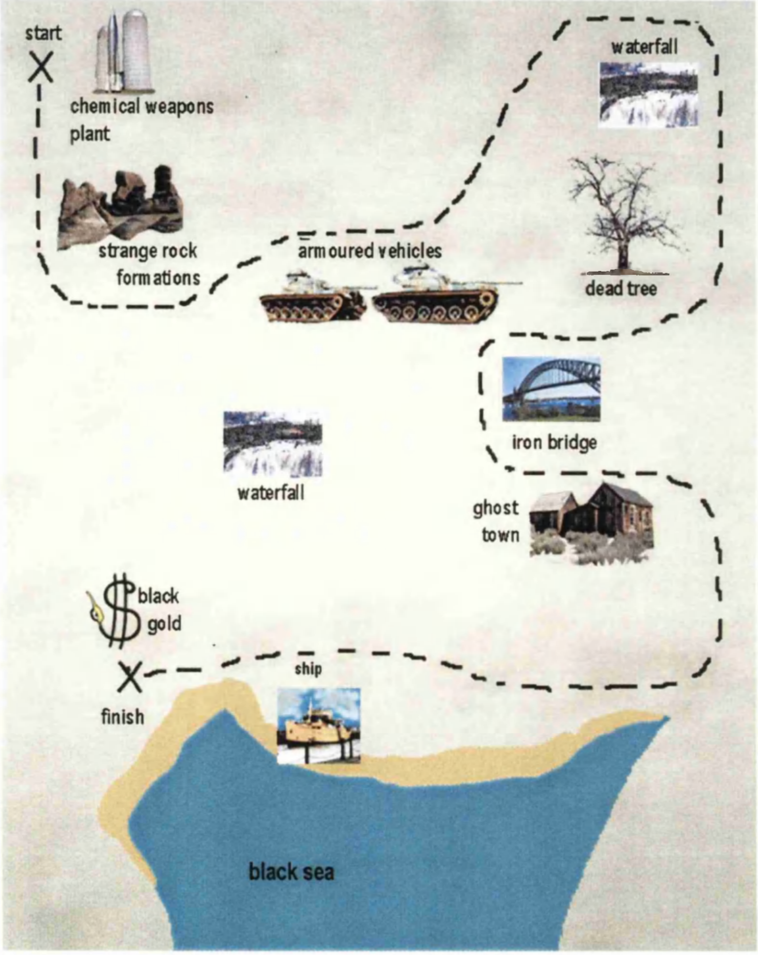


Instruction Giver

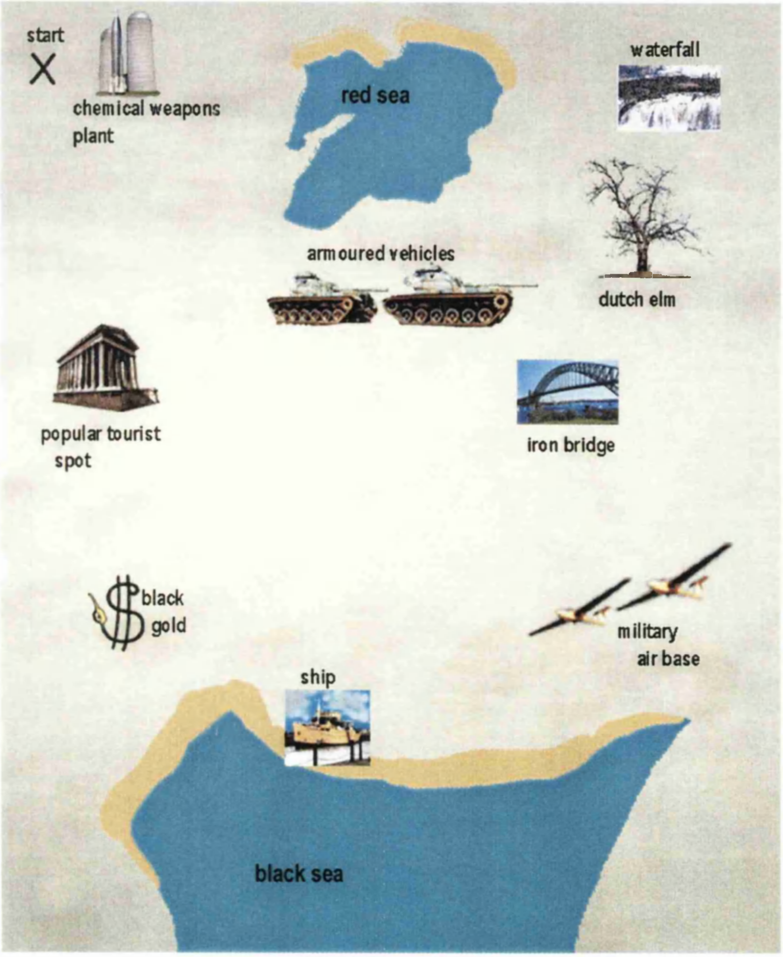


Instruction Follower

Map B



Instruction Giver



Instruction Follower

Appendix 2. Map Task Instructions (Study 2)

FACE-TO-FACE, UNTIMED

Instruction Giver

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map which contains landmarks but no route. To complicate matters, the 2 maps were drawn by different explorers, so some of the landmarks may be different.

Your task is to collaborate with your partner and describe the route as accurately as you can. Your partner must reproduce the route on his or her map. There is no time limit to complete this task.

The most important things to remember during this task are;

- you will complete the task in your own time
- the landmarks on your maps may be different

………..

Instruction Follower

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map with a route drawn on it. Unfortunately the maps were drawn by different explorers so some of the landmarks on the maps may be different.

Your task is to collaborate with your partner by following his or her instructions to reproduce the route on your map. Use the pen and draw the route as accurately as you can. If you make a mistake cross it out using the pen. There is no time limit to complete this task.

The most important things to remember during this task are;

- you will complete the task in your own time
- the landmarks on your maps may be different

FACE-TO-FACE, TIMED

Instruction Giver

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map which contains landmarks but no route. To complicate matters, the 2 maps were drawn by different explorers, so some of the landmarks may be different.

Your task is to collaborate with your partner and describe the route as accurately as you can. Your partner must reproduce the route on his or her map and you have only 3 minutes to complete the task. You will be told;

- when to start the task
- when you have 2 minutes left
- when you have 1 minute left
- when your time is up

The most important things to remember during this task are;

- you have only 3 minutes to complete the task
- the landmarks on your maps may be different

....

Instruction Follower

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map with a route drawn on it. Unfortunately the maps were drawn by different explorers so some of the landmarks on the maps may be different.

Your task is to collaborate with your partner by following his or her instructions to draw the route on your map quickly and accurately. Use the pen to draw the route. If you make a mistake, cross it out. You have only 3 minutes to complete this task. You will be told;

- when to start the task
- when you have 2 minutes left
- when you have 1 minute left
- when your time is up

The most important things to remember during this task are;

- you have only 3 minutes to complete the task
- the landmarks on your maps may be different

VIDEO-MEDIATED, UNTIMED

Instruction Giver

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map which contains landmarks but no route. To complicate matters, the 2 maps were drawn by different explorers, so some of the landmarks may be different.

Your task is to collaborate with your partner and describe the route as accurately as you can. Your partner must reproduce the route on his or her map. There is no time limit to complete this task.

Your partner will be at a different site from you and you will communicate with each other via a video link. If you wish to make eye contact with your partner, look directly into the camera.

The most important things to remember during this task are;

- you will complete the task in your own time
- the landmarks on your maps may be different

………..

Instruction Follower

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map with a route drawn on it. Unfortunately the maps were drawn by different explorers so some of the landmarks on the maps may be different.

Your task is to collaborate with your partner by following his or her instructions to reproduce the route on your map. Use the *pencil* tool and draw the route as accurately as you can. If you make a mistake cross it out using the pencil tool. There is no time limit to complete this task.

Your partner will be at a different site from you and you will communicate with each other via a video link. If you wish to make eye contact with your partner, look directly into the camera.

The most important things to remember during this task are;

- you will complete the task in your own time
- the landmarks on your maps may be different

VIDEO-MEDIATED, TIMED

Instruction Giver

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map which contains landmarks but no route. To complicate matters, the 2 maps were drawn by different explorers, so some of the landmarks may be different.

Your task is to collaborate with your partner and describe the route as accurately as you can. Your partner must reproduce the route on his or her map and you have only 3 minutes to complete the task. You will be told;

- when to start the task
- when you have 2 minutes left
- when you have 1 minute left
- when your time is up

Your partner will be at a different site from you and you will communicate with each other via a video link. If you wish to make eye contact with your partner, look directly into the camera.

The most important things to remember during this task are;

- you have only 3 minutes to complete the task
- the landmarks on your maps may be different

....

Instruction Follower

Your map was drawn by an explorer to provide a route to buried treasure (located at the finish point). Your partner also has a map with a route drawn on it. Unfortunately the maps were drawn by different explorers so some of the landmarks on the maps may be different.

Your task is to collaborate with your partner by following his or her instructions to draw the route on your map quickly and accurately. Use the *pencil* tool to draw the route. If you make a mistake cross it out using the pencil. You have only 3 minutes to complete this task. You will be told;

- when to start the task
- when you have 2 minutes left
- when you have 1 minute left
- when your time is up

Your partner will be at a different site from you and you will communicate with each other via a video link. If you wish to make eye contact with your partner, look directly into the camera.

The most important things to remember during this task are;

- you have only 3 minutes to complete the task
- the landmarks on your maps may be different

Appendix 3. Transcript of Example Dialogues (Study 2)
Face-to-face: Recorded 11/7/01. Typescript 11

Dialogue 1 - Timed - Map B (T0:01 – 3:01)
Dialogue length = 3m
No of words = 540

Summary

No	Time	Mention (Ref Exp)	Mention (Dur)	M1/M2	Sp	Landmark	Role	Code Short RE	Q-form	Response
1	0:09		1			strange rock formation	IG		Q	IR
2	0:12	1	1	M2 @ 3	S	chemical weapons plant	IG		Q	IR
2b		2				it (CWP)	IF	2		
3	0:24		2	M1 @ 2	S	chemical thing (CWP)	IG			
4					S	chemical bit (CWP)	IG			
5	0:45	1	1	M2 @ 7	S	armoured vehicles	IG		Q	IR
5b		2				it (AV)	IF	2		
6	1:03		2	M1 @ 5	S	armed vehicles (AV)	IG			
7						armoured vehicles	IF			
8	1:05	1	1	M2 @ 9	S	red sea	IF		Non-Q	IR
9	1:11	2	2	M1 @ 8	S	red sea	IF	0		
10						red sea	IG			
10b						it (RS)	IG			
11					S	armoured vehicles	IG			
11b						it (AV)	IF			
12	1:34	1	1	M1 @ 13	D	waterfall	IF		Non-Q	IR
13	1:36	2	2	M2 @ 12	D	waterfall	IG	0		

14						waterfall	IF			
15						waterfall	IG			
16						armoured vehicles	IF			
17						waterfall	IG			
18						waterfall	IF			
19						waterfall	IG			
19b						that (WF)	IF			
20					D	waterfall	IG			
21	2:21	1	1	M2 @ 23	S	tree (DT)	IF	1	Non-Q	Non-IR
22	2:24	2				dead tree	IG	0		
23				M1 @ 21	S	tree	IF			
23b						that (DT)	IF			
24	2:30	1	1	M2 @ 25	D	bridge (IB)	IG		Q	IR
25	2:30	2	2	M1 @ 24	D	iron bridge	IF	0		
26						iron bridge	IG			
27	2:37	1	1	M2 @ 33	D	ghost town	IG		Q	IR
28						iron bridge	IG			
28b						it (IB)	IF			
29	2:40	1	1	M2 @ 32	S	military air base	IF		Non-Q	Non-IR
30						iron bridge	IF			
31					D	iron bridge	IG			
32	2:58	2	2	M1 @ 29	S	military base (MAB)	IF	1		
33	3:00	2	2	M1 @ 27	D	ghost town	IF	0		

	Dialogue 1 - Timed - Map B - Transcript		
Role	Text	Number	Introduction
EXP	Ok You may begin now		

IG	Right erm if you go down from straight down from the start position		
IF	Right		
IG	Do you have er a strange rock formation?	1	Q
IF	No		IR
IG	Do you have chemical weapons plant?	2	Q
IF	No it's directly right to the start position		IR
IG	OK right well imagine there's the same shape object like directly below the chemical thing	3	
IF	/Below/ b-		
IG	That's- and you're going round so basically right /you've got/		
IF	/I'm going/ down from the start?		
IG	Yep you're going straight down from the start		
IF	Right		
IG	You're going down about 2 inches and you're curling round		
IF	/Right/		
IG	/Imagining/ there's an extra thing directly below the chemical /bit/ OK	4	
IF	Right /OK/		
IG	Do you have armoured vehicles?	5	Q
IF	Erm yeah it's kind of up to an angle		IR
IG	Up to an angle? right erm		
IF	If I go straight along when I curled round I have to go up a bit		
IG	Right yeah aye go up a bit and over the-		
EXP	2 minutes to go		
IG	Over the /armed vehs/	6	
IF	/Over/ the armoured vehicles?	7	
IG	Yep		
IF	At the red sea's in front of me now so-	8	Non-Q
IG	Oh right ok well you're going up again		
IF	As- through the red sea?	9	

IG	Er well I don't have the red sea /so I don't know/ where it is	10	IR
IF	/Oh right/ erm right /I go/ right?		
IG	Right go along and cross in front of the- along the top of the armoured vehicles /and when you get/	11	
IF	/Along the top/ if it right?		
IG	Yeah and when you get about halfway to- halfway along between the		
IF	Uh huh		
IG	Go up diagonally		
IF	Go up diagonally?		
IG	Yep towards the top right corner		
IF	/Right/ right /that's/ the waterfall up there	12	Non-Q
IG	/OK/ yeah go over the top of the /water/fall	13	Non-IR
IF	/Right/ I have		
IG	And down the side		
IF	And down the side that I've just went up		
IG	/Down the far side/		
IF	/Hang on I've went/ I've went diagonally		
IG	Yeah		
IF	Up to the waterfall so I just go straight down	14	
IG	Aye you're going along the top of the- along the top of the waterfall first	15	
IF	/Oh/ along erm		
IG	Horizontally to the top /right corner/		
IF	/Left hand side?/		
IG	Going to the right hand side		
EXP	1 /minute/ to go		
IF	/I've went/ I've went diagonally from the armoured vehicles straight up to the top right hand corner	16	
IG	Right have you erm- have you not got a waterfall on the top right /corner/?	17	

IF	/I've got/ uh huh the top right hand corner there's a waterfall	18	
IG	Aye right well you're just going over the top of the /waterfall/	19	
IF	/Over the top/ of that?		
IG	Yep and then straight down the side of the waterfall	20	
IF	On the right hand side?		
IG	Yep		
IF	Right and I'm- I'm passing a tree	21	Non-Q
IG	Yep you're going down- now you're going underneath the dead /tree/	22	Non-IR
IF	/Un/ underneath the tree?	23	
IG	Yeah		
IF	Right where I'm going after that		
IG	Now do you have an /ir- a bridge?/	24	Q
IF	/I've got an iron/ bridge yeah	25	IR
IG	Right you're going on down the left hand side of the iron bridge	26	
IF	Right		
IG	And then do you have a ghost town? /below the iron bridge?	27 28	
IF	/No no/ no		
IG	Right /well/		
IF	/I've got/ a military air base just a bit down from it	29	Non-Q
IG	OK well draw a line directly underneath the-		Non-IR
IF	Iron bridge	30	
IG	Iron bridge	31	
IF	Right		
IG	Going towards the right		
IF	Right		
IG	And then go along to the right hand- you're right along to the right hand side now		
IF	Uh huh		

IG	Go straight down		
IF	By the military base or where your ghost town would be?	32 33	
IG	Yep yep		
EXP	OK time up		

Dialogue 2 - Un-timed - Map A (T3:08 – 8:32)

Dialogue length = 5m 24s

No of words = 890

Summary

No	Time	Mention (RE)	Mention (Dur)	M1/M2	SP	Landmark	Role	Code RE short	Q-form	Response
1	3:11	1	1	M2 @ 5	S	winter bay	IG		Non-Q	Non-IR
2	3:16	1	1			balloon rally	IG		Q	IR
3	3:22	1	1	M2 @ 35	S	beach hut	IG		Q	IR
4	3:26	1	1	M2 @ 8	S	sunflower fields	IG		Q	IR
4b		2				they (SF)		2		
5	3:30	2	2	M1 @ 1	D	winter bay	IF	0		
6						winter bay	IG			
7						winter bay	IG			
8	3:51		2	M1 @ 4	S	sunflower fields	IG			
9						sunflower fields	IG			
10	4:05	1	1	M2 @ 52	S	broken gate	IG		Q	Non-IR
11						sunflower fields	IG			
12	4:09	1	1	M2 @ 19	D	farmer's gate	IF		Non-Q	Non-IR
13						sunflower fields	IF			
14						sunflower fields	IG			

12b		2				that (FG)	IG	2		
15						sunflower fields	IF			
15b						it (SF)	IG			
16						sunflower seeds	IF			
17						sunflower seeds-fields	IF			
18	4:21	2				gate (BG)	IG	1		
19	4:22		2	M1 @ 12	D	farmer's gate	IG			
19b						that (FG)	IG	2		
19c						it (FG)	IG			
20	4:26	1	1	M2 @ 53	S	camp site	IG		Q	IR
21	4:32	1				overnight-	IF		Non-Q	Non-IR
22	4:32		1			hotel (OA)	IF			
23						gate (FG)	IF			
24	4:34	2	2			hotel (OA)	IG	1		
25						gate (FG)	IG			
26						hotel (OA)	IG			
27						hotel (OA)	IG			
28						farmer's gate	IF			
29						hotel (OA)	IF			
30						hotel (OA)	IG			
31						farmer's gate	IF			
32						hotel (OA)	IF			
33						hotel (OA)	IG			
33b						it (OA)	IG			
34						hotel (OA)	IG			
35	5:31	2	2	M1 @ 3	S	beach hut	IG	0		
36	5:32	1	1	M2 @ 43	D	summer bay	IG		Q	IR
34b						that (OA)	IG			

34c						that (OA)	IF			
37						hotel (OA)	IG			
38						hotel (OA)	IF			
39						farmer’s gate	IF			
40						farmer’s gate	IF			
41						gate (FG)	IF			
42						beach hut	IF			
43	6:19	2	2	M1 @ 36	D	summer bay	IF	0		
44						hotel (OA)	IG			
45						farmer’s gate	IF			
46						sunflower fields	IG			
47						sunflower see-fields	IF			
47b						it (SF)	IG			
48						farmer’s gate	IF			
49						hotel (OA)	IF			
50						farmer’s gate	IF			
51						farmer’s gate	IF			
52				M1 @ 10	S	broken gate	IG			
53	7:23	2	2	M1 @ 20	S	camp site	IG	0		
54	7:25		11	MF	S	farmer’s gate	IG			
55	7:26		14			hotel (OA)	IG			
56	7:31	1	1	M2 @ 57	D	coach park	IF		Q	IR
57	7:32	2	2	M1 @ 56	D	coach park	IG	0		
58	7:37	1	1	M2 @ 61	S	waterski resort	IF		Q	IR
58a		2	2			that (WR)	IG	2		
59	7:40		3	MF	D	summer bay	IG			
60	7:41		3			coach park	IF			
61	7:45		2	M1 @ 62	S	waterski resort	IF			

62	7:48		3			camp site	IF			
63	7:53		4			beach hut	IF			
64	8:00		4			camp site	IF			
65	8:01		5	MF	S	camp site	IG			
66	8:05		3			waterski resort	IF			
67	8:09		4	MF	D	coach park	IG			
68	8:13		4			waterski resort	IG			
69	8:15	1	1	M2 @ 71	S	precious stones	IF		Q	IR
70	8:20		5	MF	S	waterski resort	IF			
71	8:24	2	2	M1 @ 69	S	precious stones	IF	0		
72	8:28		3	MF		stones	IG			

	Dialogue 2 - Un-timed - Map A - Transcript		
Role	Text	Number	Introduction
IF:	Right		
IG:	Right erm the starting point's at the winter bay	1	Non-Q
IF:	Yeah		Non-IR
IG;	OK if you go down about an inch do you have er the balloon rally?	2	Q
IF:	No		IR
IG:	No erm do you have something there?		
IF:	No		
IG:	Nothing at all?		
IF:	Nothing at all		
IG:	Erm do you have the beach hut?	3	Q
IF:	No		IR
IG:	Do you have the sunflower fields?	4	Q

IF:	Yes I do /they're to the left/ aye to the left of /winter bay/	5	IR
IG:	/Just near the winter bay/ ah hah /right/ if you come down erm about about an inch or a bit more	6	
IF:	Uh huh		
IG:	And you go along just until maybe the end of the winter bay the left hand side	7	
IF:	Right		
IG:	And well kind of go up that way and curve round towards the sunflower fields	8	
IF:	Yeah		
IG:	So it kind of looks like a		
IF:	/It's almost/ following the-		
IG:	Yeah it looks like a "U"		
IF:	Right ok		
IG:	With a a longer tail at the end near the sunflower fields	9	
IF:	Right		
IG:	And if you go straight along er do you have a broken gate near the sunflower fields?	10 11	Q
IF:	I have a farmer's gate to the- directly to the left of the sun/flower fields/	12 13	Non-Q (FG) Non-IR (BG)
IG:	/The sunflower fields/ right go by that	14	
IF:	I'm I going along the top of the sunflower /fields/	15	
IG:	/No/ underneath it underneath it		
IF:	Underneath the sunflower seeds sunflower seeds- (laughs) fields	16 17	
IG:	And along underneath the gate as well the farmer's gate you've got	18 19	Non-IR (FG)
IF:	Right /OK/		
IG:	/And/ then you stop there erm do you have a camp site below that well not directly	20	Q

	below just a bit away from it		
IF:	No I've got overnight- I've got a hotel basically /directly/ below the gate	21 22 23	Non-Q (OA) IR (CS)
IG:	You've got a /hotel/ the gate ok erm go down and round go- go straight down first	24 25	Non-IR (OA)
IF:	Right		
IG:	Until you get to the the hotel and then turn right	26	
IF:	So I'm /going/ underneath		
IG:	/Just so/ no above sorry above above the hotel /it's it's/	27	
IF:	/How can that be/ if I'm going underneath the farmer's gate then it's also above the hotel /at the same time/	28 29	
IG:	/Underneath/ /ah huh/ it's above the hotel	30	
IF:	Yeah so where am I going once I've gone underneath the farmer's gate and above the hotel /cos/ I'm now at the left hand side	31 32	
IG:	/Erm/ is the hotel directly underneath I mean is it are the 2 pictures straight underneath each other?	33	
IF:	More or less with about half a centimetre difference		
IG:	Oh right I've not got that I've got about this much space and er erm well get to above- well- if you're above the hotel just now	34	
IF:	Right		
IG:	Have you got like a beach hut or the summer bay down below that?	35 36	Q
IF:	Yep straight below that		IR
IG:	Right well if you go down the right hand side of the hotel	37	
IF:	The right hand side of the hotel	38	
IG:	The right hand side yeah		
IF:	I won't be underneath the farmer's gate then I won't have gone underneath the farmer's	39	

	gate	40	
IG:	Erm		
IF:	I’m right up against the edge here having gone underneath the gate right up against /the left/-hand side	41	
IG:	/Right/ uh hum /does your- does your/ picture touch the right- the the left hand side?		
IF:	/I’m directly/ no no it’s I’ve just there’s a gap there about a centimetre gap and I’m directly above the beach hut which is in erm just beside summer bay	42 43	
IG:	Uh huh can you go down the right hand side of the hotel no?	44	
IF:	Well I could do but I wouldn’t have gone underneath the farmer’s gate so far right I’ve gone down	45	
IG:	Uh hum		
IF:	From the start		
IG:	Ah hah		
IF:	Along		
IG:	Along		
IF:	To the left /up/ a bit		
IG:	/Up/ a curve		
IF:	Yep		
IG:	It’s like a “U” yeah by the sunflower fields	46	
IF:	Then I’ve gone underneath the sunflower /see- fields/	47	
IG:	/Underneath it/ yeah		
IF:	And I’ve gone underneath the farmer’s gate	48	
IG:	Right		
IF:	And now I’m at the left-hand side of the picture of the whole thing		
IG:	Right well		
IF:	So for me to go down the right-hand side of the hotel I would need to completely back track	49	
IG:	Uh hum		

IF:	And I wouldn't have gone underneath the farmer's gate at all because I would then also be at the right hand side of the /farmer's gate/	50 51	
IG:	/Cos on/ this map you have to pass the- pass underneath the broken gate I've got and then down the right-hand side of the camp site and the- your farmer's gate and the hotel are kind of positioned in the same places I think there closer together on your map /than on mine/	52 53 54 55	
IF:	/Do you have/ a coach park?	56	Q
IG:	A coach park no I don't	57	IR
IF:	With funny looking buses or /something/		
IG:	/No/ I don't		
IF:	Right do you have a waterski resort?	58	Q
IG:	Ah huh that's just at the top of the summer bay	59	IR
IF:	Right well my coach park is almost directly above the waterski resort right	60 61	
IG:	Right		
IF:	Erm is your camp site to the /left/	62	
IG:	To the left of that ah huh		
IF:	And directly above the beach hut	63	
IG:	Dir- yes ah huh that's right		
IF:	Right so your my line's supposed to go above the camp site	64	
IG:	Above the camp site and down	65	
IF:	And down		
IG:	And down		
IF:	And down over the waterski resort	66	
IG:	Right OK so I must be then avoiding the coach park as well so I'm going straight down- I'm now at the waterski resort /OK/	67 68	
IF:	Ah huh you go er do you have the precious stones?	69	Q
IG:	Yes		IR

IF:	Ah hah well just if you're at the waterski resort just kind of curve round and down towards the precious stones	70 71	
IG:	OK right I'm at the stones now	72	
IF:	And that's the that's the finish		
IG:	Right that's us		